Journal of General Virology (2002), **83**, 2575–2585. Printed in Great Britain

On the evolution and molecular epidemiology of the potyvirus *Papaya ringspot virus*

Marion F. Bateson,[1] Rosemarie E. Lines,[1] Peter Revill,[1] Worawan Chaleeprom,[1]† Cuong V. Ha,[2] Adrian J. Gibbs[3] and James L. Dale[1]

[1] School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4000, Australia

[2] Hanoi Agricultural University, Gia Lam, Vietnam

[3] School of Botany and Zoology, Australian National University, ACT 2000, Australia

The potyvirus *Papaya ringspot virus* (PRSV) is found throughout the tropics and subtropics. Its P biotype is a devastating pathogen of papaya crops and its W biotype of cucurbits. PRSV-P is thought to arise by mutation from PRSV-W. However, the relative impact of mutation and movement on the structure of PRSV populations is not well characterized. To investigate this, we have determined the coat protein sequences of isolates of both biotypes of PRSV from Vietnam (50), Thailand (13), India (1) and the Philippines (1), and analysed them together with 28 PRSV sequences already published, so that we can better understand the molecular epidemiology and evolution of PRSV. In Thailand, variation was greater among PRSV-W isolates (mean nucleotide divergence 7·6%) than PRSV-P isolates (mean 2·6%), but in Vietnamese populations the P and W biotypes were more but similarly diverse. Phylogenetic analyses of PRSV also involving its closest known relative, *Moroccan watermelon mosaic virus*, indicate that PRSV may have originated in Asia, particularly in the Indian subcontinent, as PRSV populations there are most diverse and hence have probably been present longest. Our analyses show that mutation, together with local and long-distance movement, contributes to population variation, and also confirms an earlier conclusion that populations of the PRSV-P biotype have evolved on several occasions from PRSV-W populations.

Introduction

The potyvirus *Papaya ringspot virus* (PRSV) is an important pathogen of papaya and cucurbits. The virus is classified into two biotypes that have virions which cannot be distinguished in serological tests, but that differ in their ability to infect papaya (Purcifull, 1984). PRSV-P naturally infects papaya (*Carica papaya*) and is a major limiting factor in papaya production worldwide (Purcifull, 1984). PRSV-W, which has a natural host range within the *Cucurbitaceae* and is unable to infect papaya, has been described as one of the five most important viruses in field-grown vegetables (Tomlinson, 1987). Although PRSV-P can usually be transmitted experimentally to cucurbits, it is not usually found in cucurbits in the field (Gonsalves, 1998). PRSV-P was first described in 1949 after isolation from papaya in Hawaii (Jensen, 1949), and has since been reported from many other countries where it has often devastated papaya production within a few years of the first infection (Gonsalves, 1998). The epidemiology of PRSV-P in papaya is similar to that of other non-persistent, aphid-borne viruses and it is not found in intercropped cucurbits, suggesting that the virus spreads directly from papaya to papaya. There is, however, little information on the epidemiology of PRSV-W in cucurbits. In addition, there is evidence that PRSV-P evolved from PRSV-W (Bateson *et al.*, 1994), presumably by mutation. This was first indicated by the very close sequence similarity of the coat protein (CP)-coding regions of P and W isolates within Australia, and the fact that PRSV-W was found to be

Author for correspondence: Marion Bateson.
Fax +61 7 38641534. e-mail m.bateson@qut.edu.au

†**Present address:** Faculty of Agricultural Production, Maejo University, Samsai, Chiang mai 50290, Thailand.

GenBank accessions for sequences used in this study are AF506840–AF506904.

0001-8419 © 2002 SGM

Table 1. PRSV isolates and sequences used in this study

Acronym*	Host	Geographical origin	Reference	GenBank accession no.
Previous studies				
USP-HW	*Carica papaya*	USA (Hawaii)	Wang *et al.* (1994)	X67673
USW-FL	Cucurbit†	USA (Florida)	Quemada *et al.* (1990)	D00594
USP-FL	*Carica papaya*	USA (Florida)	Davis & Ying (1999)	AF196839
AUP-BD; AUP-BUN; AUP-DAY	*Carica papaya*	Australia (SE Qld)	Bateson *et al.* (1994)	U14736–8
AUW-NT; AUW-DB1	*Cucurbita pepo* (zucchini)	Australia (NT	Bateson *et al.* (1994)	U14744; S89893
AUW-GAT	*Cucurbita maxima* (pumpkin)	Australia (Qld)	Bateson *et al.* (1994)	U14739
THP-X	*Carica papaya*	Thailand	Unpublished	AB044340
THP-11	*Carica papaya*	Thailand	Bateson *et al.* (1994)	U14743
TWP-2	*Carica papaya*	Taiwan	Unpublished	AB044341
TWP-YK	*Carica papaya*	Taiwan	Wang *et al.* (1994)	X78557
VNP-01	*Carica papaya*	Vietnam (north)	Bateson *et al.*, (1994)	U14742
SRP	*Carica papaya*	Sri Lanka	Bateson *et al.* (1994)	U14741
MXP-VC; MXP-CH	*Carica papaya*	Mexico	Silva-Rosales *et al.* (2000)	AJ012649; AJ012099
MXP-COL; MXP-MCJ; MXP-QRF; MXP-YCY; MXP-SP	*Carica papaya*	Mexico	Unpublished	AF309968; AF319468; AF319493; AF319499; AF319502
BZP-2; BZP-9	*Carica papaya*	Brazil	Unpublished	AF344640; AF344647
BZW-10; BZW-11		Brazil	Unpublished	AF344648–9
JAP	*Carica papaya*	Japan	Unpublished	AB044339
CHP	*Carica papaya*	China	Unpublished	AF243496
INP	*Carica papaya*	India (Pune)	Jain *et al.* (1998)	AF063220
INW	*Cucurbita moschata* (pumpkin)	India (New Delhi)	Jain *et al.* (1998)	AF063221
INP-BR	*Carica papaya*	India (Bangalore)	Unpublished	AF305545
MWMV	Cucurbit†	Morocco	Lecoq *et al.* (2001)	AF305545
MWMV-SUDAN	Cucurbit†	Sudan	Lecoq *et al.* (2001)	AF307778
This study				
INU-01	Cucurbit†‡	India		AF506845
PHP-01	*Carica papaya*	Philippines		AF506902
THW-03	*Coccinia indica* (ivy gourd)	Thailand		AF506895
THW-04	*Cucurbita maxima* (pumpkin)	Thailand		AF506894
THW-06; THW-05	*Benincasa hispida* (wax gourd)	Thailand		AF506892–3
THW-07	*Luffa acutangula* (loofah)	Thailand		AF506891
THW-08	*Cucumis sativus* (cucumber)	Thailand		AF506890
THU-10; THU-09	*Cucumis sativus* (cucumber)‡	Thailand		AF506896–7
THP-14; THP-13; THP-12; THP-02; THP-01	*Carica papaya*	Thailand		AF506898–900; AF506901; AF506902
VNP-02 - VNP-29	*Carica papaya*	Vietnam		AF506862–AF506889
VNW-50; VNW-48; VNW-46; VNW-45; VNW-43; VNW-41–35; VNW-32–30	Pumpkin†‡	Vietnam		AF506841; AF506843; AF506903; AF506846; AF506848; AF506850–56; AF506859–61
VNW-42; VNW-47	*Luffa acutangula* (loofah)‡	Vietnam		AF506849; AF506844
VNW-44; VNW-33; VNW-34; VNW-51	Cucumber†‡	Vietnam		AF506847; AF506858; AF506857; AF506840
VNW-49	Cucurbit†‡	Vietnam		AF506842

* Acronyms include two-letter country code, biotype (P or W) followed by isolate definition if necessary.
† Species unconfirmed.
‡ Virus biotype not confirmed.

widespread in Australia at least 20 years before PRSV-P appeared (Greber, 1978; Thomas & Dodman, 1993).

CP sequences have been determined for several isolates of PRSV-P from different parts of the world, together with a smaller number of PRSV-W isolates (Quemada *et al.*, 1990; Bateson *et al.*, 1994; Wang *et al.*, 1994; Jain *et al.*, 1998; Davis & Ying, 1999; Silva-Rosales *et al.*, 2000). Nucleotide and amino acid sequence divergence of up to 14% and 10%, respectively, has been reported between these isolates. Although initial data from the USA and Australia (Quemada *et al.*, 1990; Bateson *et al.*, 1994) suggested that there was little variation in PRSV within these countries, more recent, albeit limited, data from Mexico (Silva-Rosales *et al.*, 2000), Brazil (GenBank) and India (Jain *et al.*, 1998) have suggested there may be greater sequence variation within other countries.

The geographical origin of PRSV is not known. Early confusion in correctly identifying PRSV, especially in cucurbits, and the lack of adequate records from many countries has made studies of the evolution and epidemiology of PRSV difficult. The relative significance of mutation and movement of the virus around the world is not clear even though it may impact on control of the virus. In this study we have determined sequence variability and derived phylogenetic relationships to investigate the distribution and complexity of PRSV populations worldwide and within two countries, Thailand and Vietnam, and we report the results of this work here.

Methods

PRSV isolates. The names and origins of isolates used in this study are listed in Table 1. Thai isolates of PRSV were collected between 1992 and 1994 from fields in different regions of central, north and north-east Thailand where PRSV is prevalent. PRSV was not found in southern Thailand during this survey. The biotype of isolates THW03–THW08 was confirmed by host range tests. Isolates THU09 and THU10 were collected separately from cucurbits but their host ranges were not tested. Cucurbit and papaya samples showing symptoms of PRSV infection were collected throughout Vietnam between 1998 and 2000 and leaf samples were thinly sliced and dried on silica gel. PRSV infection was confirmed by PCR with specific PRSV primers and the PCR products screened by heteroduplex analysis (Chaleeprom, 1997) to provide a preliminary assessment of levels of diversity. Philippines and Indian (INU01) isolates were obtained as fresh leaf material. All cucurbit isolates of PRSV are referred to as PRSV-W; however, those for which the biotype was not confirmed by host range tests are indicated in Table 1.

RT–PCR. RNA was extracted from fresh or lyophilized tissue either as previously described (Bateson *et al.*, 1994) or from 50 mg of dried tissue using an RNeasy kit (Qiagen). cDNA was synthesized from purified RNA using the complementary primer MB12A (5′ GACTCAA-CAAACACACAAGCGCA 3′) as previously described (Bateson *et al.*, 1994). A fragment representing the PRSV CP and partial nuclear inclusion b (NIb) coding regions, was amplified by PCR from Thai, Philippines and Indian (INU01) isolates with primers MB12 (Bateson *et al.*, 1994) and REP4 (5′ GCTTCCGGAGCATCGATTGGAGGC 3′), as previously described (Bateson *et al.*, 1994). The CP-coding region was amplified from Vietnamese isolates in a standard PCR reaction using *Taq* polymerase (Perkin Elmer) and including 30 pmol each of primers MB11 (Bateson *et al.*, 1994) and MB12A. PCR amplification conditions included an initial denaturation cycle of 5 min at 94 °C followed by 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 50 °C and extension for 30 s at 72 °C, with a final extension of 10 min at 72 °C.

Cloning and sequencing. PCR-amplified DNA was gel purified using a Wizard PCR Purification kit (Promega) and then cloned into pGEM-T (Promega) according to the manufacturer's instructions. Vietnam clones were sequenced at the Australian Genome Research Facility (AGRF). Overlapping sequences were obtained using Universal primers (forward and reverse) and specific primers MB11 (Bateson *et al.*, 1994), MB12A, VNCP350 (5′ GTG GTA TGA GGG AGT GAG G 3′) and either CPREV (5′ TCT CGA TAC ACC AAA CCA TCA AGC C 3′), CPREV2 (5′ TCC CCA TCC ATC ATT ACC CAA ACA CC 3′) or CPREV3 (5′ GGG ATA ATC AAC TTG GGT TTC CCC 3′). DNA for sequencing of remaining clones was prepared using the Applied Biosystems Big Dye terminator kit as recommended by the manufacturer, and sequenced using an Applied Biosystems 373A DNA Sequencer. Overlapping sequences were obtained using Universal primers (forward and reverse) and the specific primers MB11, MB12A, MB26, MB6, MB13, MB14 (Bateson *et al.*, 1994), MB28 (5′ TGG ATG GGG AAA CCC AAG TTG A3′), MB29 (5′ TGC CTA AAT GTC GGA GTA GCA TGC 3′) and REP4.

Sequence analysis. Nucleotide and amino acid sequences representing the full CP-coding region of PRSV isolates were aligned using PILEUP (Feng & Doolittle, 1987), available from WEBANGIS (Australian National Genomic Information Service), and CLUSTALX (Thompson *et al.*, 1997). Distances were calculated using the DISTANCES program in WEBANGIS. Neighbour-joining trees were generated using CLUSTALX and TREEVIEW (Page, 1996) using either Indian (INW, INP-BR) or *Moroccan watermelon mosaic virus* (MWMV, MWMV-Sudan) sequences as the outgroup. The robustness of the lineages in the phylogenetic trees was assessed from the internode lengths in the trees, by bootstrapping (Felsenstein, 1985) in CLUSTALX using 1000 resamplings, or by comparing the trees obtained by neighbour-joining with those calculated by the Tree-Puzzle maximum-likelihood method (Strimmer & von Haeseler, 1996; Strimmer *et al.*, 1997).

Each of the aligned sequences was examined for variations in its relatedness to the other sequences throughout its length using the SISCAN program version 2.0 (Gibbs *et al.*, 2000). This detects conflicting relatedness signals that result from recombination or from differential selection, and tests these signals using Monte Carlo randomization procedures so that misleading signals resulting from similarities of composition may be discounted, and 'synonymous', 'non-synonymous' and all differences can be examined separately.

Results

Size and complexity of the PRSV CP

We compared the CP sequences of 93 PRSV isolates, including 13 Thai and 50 Vietnamese isolates, a Philippines isolate of PRSV-P (PHP-01) and an Indian cucurbit isolate of PRSV (INU-01) for which data were generated in this study, as well as 28 other sequences that were already available (Table

```
                 GLNEKLKEKEKQKE KEKEKE KQK DKDNDG ASDGNDVSTSTKTGERDRDVN
VNW-43C          GLNEKLKEKEKQKE KEKEKE KQK DKDNDG ASDGNDVSTSTKTGERDRDVN
VNW-33N          -----------**- ------ --- ------ ---------------G-----
VNW-X[1]         -------------- ------ --- ------ ---------------------
VNW-50N          -------------- ------ E-- ---I-- ---------------------
VNW-44C          --S----------- ------ N-- ------ ---------------------
VNP-X[2]         -------------- -----D --- ------ ---------------------
VNP-09C          D------------- --E--D --- ------ ---------------------
VNP-23C          -------------- -----D --- ------ --G------------------
VNP-28NC         -------------- -----D --- ----G- ---------------------
VNP-02NE         -------------- -----D --- ------ ---------------------
VNP-10NW         -------------D -----D --- ------ ---------G-----------
VNP-16N          -------------- -----D --- ------ -----------R---------
VNP-14NW         -I------------ -----D --- --N--- ---------------------
VNP-11N          -------------- -----D --- ----V- ---------------------
VNP-12NE         -------------- -----D --* N--S-- ----SN---------------
VNP-X[4]         -------------- -----N --- -----E ---------------------

                                              DVNN
                                              ▼
VNW-38C          -------------- ***--D --- --QK-- ---------------------
                                              N
                                              ▼
VNW-48S          -------------- ***--- --- E-Q--- -----N---------------
VNW-37S          -------------- ***--- --- E----- -----N---------------
VNP-07S          -----P-------- ***--D --- --N-E- ---------------------
VNP-06S          ----TP-------- ***--D --- -QN-ES ---------------------
VNP-05S          -------------- ***--D --- --N-E- -----------R---------
VNP-X[3]         -------------- ***--G --- --N-E- ---------------------
VNP-17S          ---------**--- ***--G --- --N-E- -----------R----*----
VNP-04S          -------------- ***--G --- --I--- ---------------------
VNP-26S          -------------- ***G-- --- --I--- ----I----------------
VNP-08C          -------------- ***G-D --- --I--E ---------------------
VNP-25C          --------N----- ***G-D --- N-T--- ---------------------
VNP-24C          -------------- ***--D --- --T--R ---------------------
THW-07           -------------- ***--D --- E-N--V S--------------------
THW03            -------------- ***--D --- YEND-- --------A------------
THW-05           -------------- ***--D --- E----- ---------------------
THW-08           -------------- ***--D --- --VGND ----------I----------
THW-04           ---D---------- ***--- --- ---DE- ---------------------
THW-06           -------------- ***--- --- E-ITK- ---------------------
THP-13, THP-X    ------[D]------ ***--D --- G-E-NE ---------------------
THP-01           ------[D]------ ***--D --- G-E-TE ---------------------
THP-12           ------[D]------ ***--D --- G-E-NE -----N------I--------
THU-10           ---D--[D]------ ***--D --R G-E-NE ---------------------
THU-09           ---D--[D]------ ***--D --- G-E-NE -----H---------------
THP-02, THP-14   ---D--[D]------ ***--D --- G-E-NE ---------------------
THP-11           -F----[D]------ ***--D --R N-E-NE ---------------------
CHP              -------------- ***--D --- N-N--- T--------------------
PHP-01           -------------- ***--D --- --I-E- ---------GI----------
VNW-X[5]         -------------- **---D --- ---*-- ---------------------
VNW-36NW         -------------- **---D --- ----E- ---------------------
VNW-40N          C------------- **---- --- ---D-- -----N---------------
TWP-2, JAP       -------------- **---D --- ------ ---------------------
TWP-YK           -------------- **---D --Q ------ ---------------------
MXP-MCJ          -------------- **I--G --- E-EK-- ------------V--------
USP-HW           -------------- **---- --- E-EK-- ---------------------
BZP-9            --S----------- **--R- --- --EK-- ---------------------
BZP-2            -----R-------- **--E- --- E-EK-- V----------R---------
BZW-10           --------N----- **---- --- E-EK-D ---------------------
BZW-11           -S------------ **--R- --- E-EK-D ---------------------
AUW-GAT          -------------- **---- --- E-EK-D -----------------G---
AUW-NT           -------------- **---- --- E-EK-D -------P-------------
AUP-BD           ------R------- **---- --- E-EK-D ---------------------
AUP-WP           -M------------ **---- --- E-EK-D -----N---------------
AUX[6]           -------------- **---- --- E-EK-D ---------------------
MXP-QRF, YCY     -------------- **---- --- E-EK-D ---------------------
INU-01           -------------- **---- --- E-EK-D -------------------F-
USW-FL           -----F-------- **---- --- E-EK-D ---------------------
USP-FL           -------------- **---- --- E-**-D ---------------------
INP              D------------- **---- -G- E-**-D ---------------------
SRP              ---D---------- **---- --- ERD*-- S---K-A----N---------
INW              --DG------NE-- **-I-- -E- E-K*-N -D------S----V----I---
MXP-SP           -------------- ****-- -L- E-EK-D ---------------------
MXPCH            -------------- ****-- --- E-ERGD ---------------------
MXP-COL          -------------- ****-- --- E-EK-N -----AE--------------
MXPVC1           --D----------- ****-- --- E-EK-N -----------R---K-----
PRP              -------------- ****-- R-R E-EKVD ----KN------I----K---
INP-BR           ---D-------E-- ****** -*- --K*-* ----GY---------------
```

Vietnam W (N, C), P (N, C, S) — VNW-43C to VNP-X[4]

Vietnam W (S), Vietnam P (S, C), Thai P, W, China, Philippines — VNW-48S to PHP-01

Taiwan, Japan, Vietnam W (N) — VNW-X[5] to TWP-YK

Australia, India, Sri Lanka, Americas (USA, Brazil, Mexico, Puerto rico) — MXP-MCJ to INP-BR

[1] 46C;49S;34N;39N;47C;45C;41C;42C
[2] 21C;27C;22C;03N;15NW;13NW;01N
[3] 19S;18S;20S
[4] 29NW;35NW
[5] 30N;31N;32N;51N
[6] P-BUN;P-DAY;W-DB1

Fig. 1. For legend see facing page.

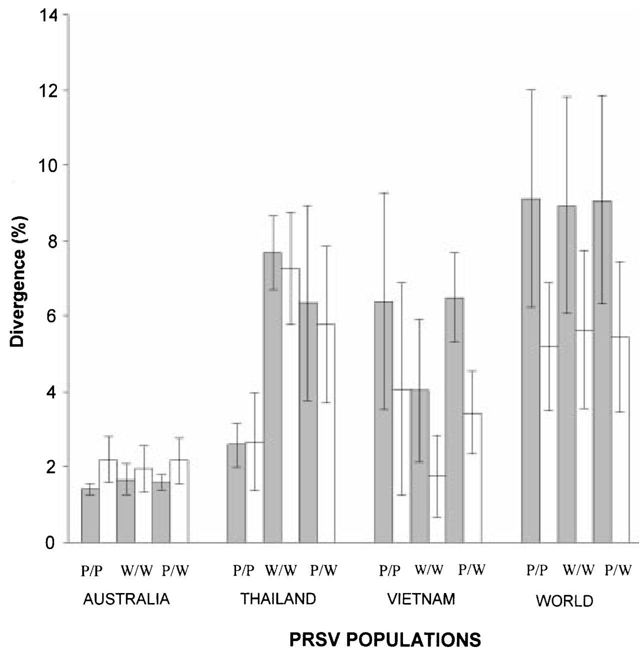


Fig. 2. Comparison of sequence divergence within different PRSV populations. The mean of the percentage pairwise divergence and corresponding standard deviation is shown at nucleotide (grey) and amino acid (white) levels between P isolates (P/P), between W isolates (W/W) and between P and W isolates (P/W) from Australia, Thailand, Vietnam and around the world. Numbers of isolates analysed were Australia: W (3) P (4); Thailand, W (8) P (7); Vietnam, W (22) P (29); World (selected PRSV sequences as shown in Fig. 3), W (14) P (29).

1). The CP-coding region of PRSV isolates ranged in size from 840 to 870 nucleotides. Nucleotide sequence differences between genomes were in multiples of three, maintaining the integrity of the reading frame and resulting in CPs of between 280 amino acids (Indian P isolate, INP-BR) and 290 amino acids (VNW-38 from central Vietnam). Interestingly, the CP-coding region of all Thai PRSV isolates was the same length, 858 nucleotides with 286 encoded amino acids, while those from Vietnam demonstrated considerable heterogeneity in CP length, at 285–290 amino acids. All differences in length were confined to the N terminus of the CP (Fig. 1). A large number of KE repeats, previously reported in the N terminus of the PRSV CP (Silva-Rosales *et al*., 2000), was found in all isolates. Most size differences occurred in one of two hypervariable regions and most resulted from differences in the number of KE repeats. Amino acid substitutions in this region were mainly conservative. The second region, of approximately six amino acids, was considerably more variable in sequence although many substitutions were still conservative. The remaining N-terminal sequences were quite conserved. In the N terminus of the two Indian isolates, INW and INP-BR, which differed most from other isolates, the regions adjacent to the DAG motif and to the CP core were the same as in the other isolates. The remaining parts of the sequence were extremely variable, particularly in the INP-BR isolate which lacked much of this region.

All sequences were checked using the SISCAN method for the possibility of recombination or selection that would result in conflicting patterns of relatedness. First the 5′ portion of the CP-coding region sequences that encode the repetitive and variable N termini of the CP were removed, and the analysis confined to the conserved portion that encodes, at its 5′ terminus, the amino acid sequence 'GERDVNVG … 7′. Only two related cucurbit isolates, THW-07 and THW-08, gave clear evidence of recombination. The 5′-terminal 100 nucleotides of their CP-coding region was most closely related to the same region of other Thai PRSV-W isolates, especially THW-05, whereas the central part of their CP-coding region was significantly related ($Z > 4{\cdot}0$) to sequences from PRSV isolates from other parts of the world, especially India. This anomalous relationship was only found using the synonymous differences between the sequences; non-synonymous comparisons showed significant relatedness to other Thai isolates, but only of the 5′-terminal part of the sequence.

Sequence diversity in PRSV worldwide

When we looked at pairwise divergence between PRSV sequences from different geographical locations, we found considerable sequence variability within both P and W populations (Fig. 2). The most divergent isolates were the Indian PRSV-W isolate, INW, and the Sri Lankan PRSV-P isolate, SRP, which were 15·2% different from each other. The mean nucleotide sequence divergence between P isolates was 9·2% with a maximum of 14·9% (INP-BR/SRP), while between W isolates the mean divergence was 9·1% with a maximum of 13·6% (INW/THW06). At the amino acid level, the mean divergence was only 5·3% and 5·7% between P and W isolates, respectively, although there was as great as 11% divergence between some isolates (INW/THW05; THW03/PRP).

The nucleotide distances for all 93 isolates were used to generate phylogenetic trees and included two isolates of the related potyvirus, MWMV, as the outgroup. The number of sequences in the Australian, Thai and Vietnamese groups were subsequently trimmed to representative groups, making the trees easier to analyse without changing the general top-

Fig. 1. Alignment of the N-terminal amino acids of the CP of PRSV isolates compared to VNW43C. Sequences were aligned manually using the conserved motifs KQKE, KQK and ASDGN (indicated by a bar) as points of reference and then sorting based on similarities in the hypervariable boxes. The first 'G' residue in the alignment is from the DAG motif. Amino acids that are the same are indicated with a '–'. Deletions are indicated by '*'. The 'D' residue that is conserved in Thai P isolates is boxed.

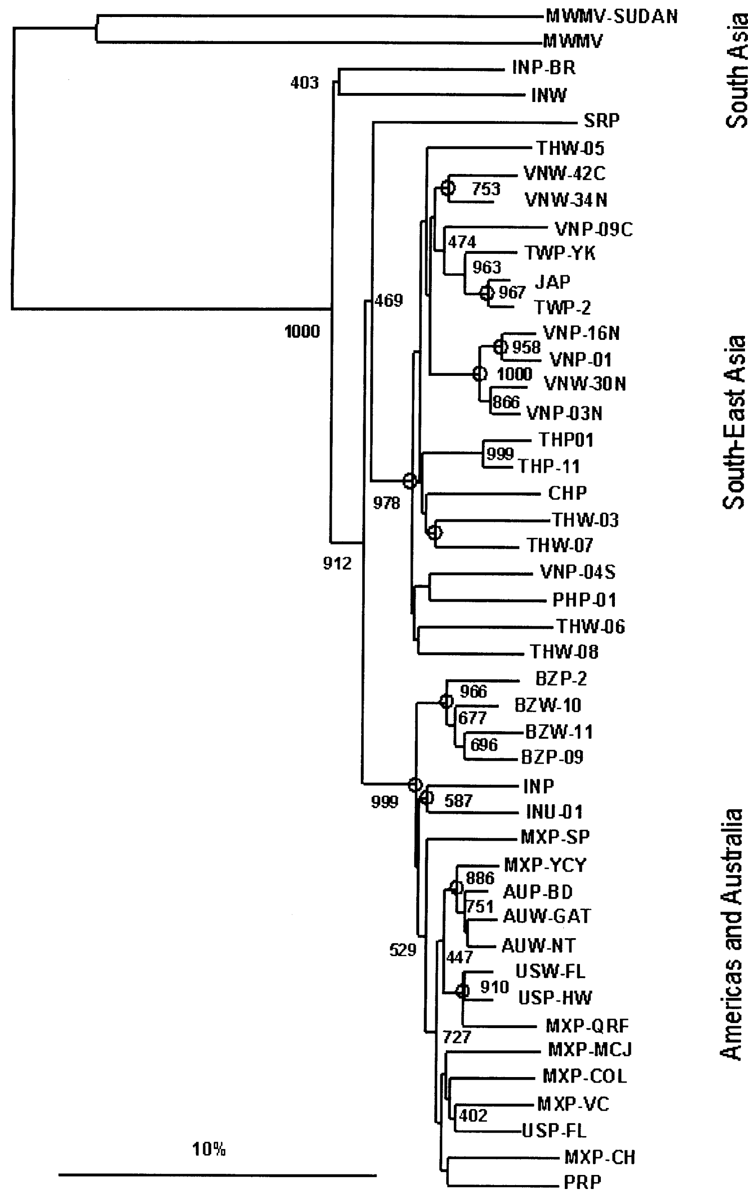
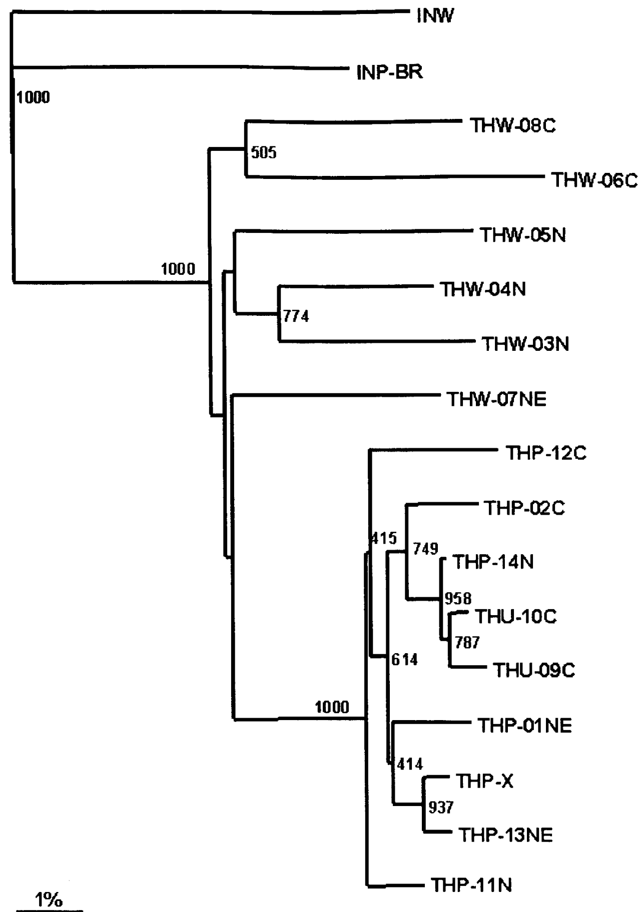


Fig. 3. Neighbour-joining tree showing the relationships of three representative PRSV sequences. The number of bootstrap trees (greater than 400/1000) in which particular nodes were found is shown, and the nodes also found by the Tree-Puzzle method are circled; the INP-BR, INW and SRP sequences were linked by a single trichotomous node by Tree-Puzzle.

ography of the tree (Fig. 3). In the general tree, as mentioned above, the three sequences from India and Sri Lanka formed a sister cluster to all the other sequences; in the neighbour-joining tree the Sri Lankan isolate was sister to the two Indian isolates (Fig. 3), but in the maximum-likelihood tree they formed a trichotomy (not shown). The other isolates formed two major lineages. Those from the Americas (Brazil, Mexico, Puerto Rico and the USA) and Australia formed one, in which a cluster of Brazilian isolates and two from India formed two lineages, and the third was a cluster of Australian, Mexican and US isolates. Within the Australian/American lineage the Mexican isolates were the most diverse, suggesting that this was the population from which the Australian and USA isolates diverged. The grouping of Australian sequences within the North American clade suggests they were probably imported from there. The low level of variation in Australia (Fig. 2) suggests that this occurred relatively recently.

The other major lineage included all the isolates from South-East Asia and the Western Pacific, including Thailand, Vietnam, China, Japan and the Philippines. However, the subclustering of isolates did not correlate well with their geographical origins, and they appeared to be a single mixed population with some well-defined subpopulations. Vietnamese and Thai isolates of both P and W biotypes intermingled with other Asian isolates. All Vietnam isolates, with the exception of P isolates from south Vietnam (represented by VNP04S), diverged from a common branch that also included P isolates from Japan and Taiwan. Isolates from south Vietnam diverged with the Philippines isolate and were closely associated with several Thai W isolates as was the Chinese

A.



B.

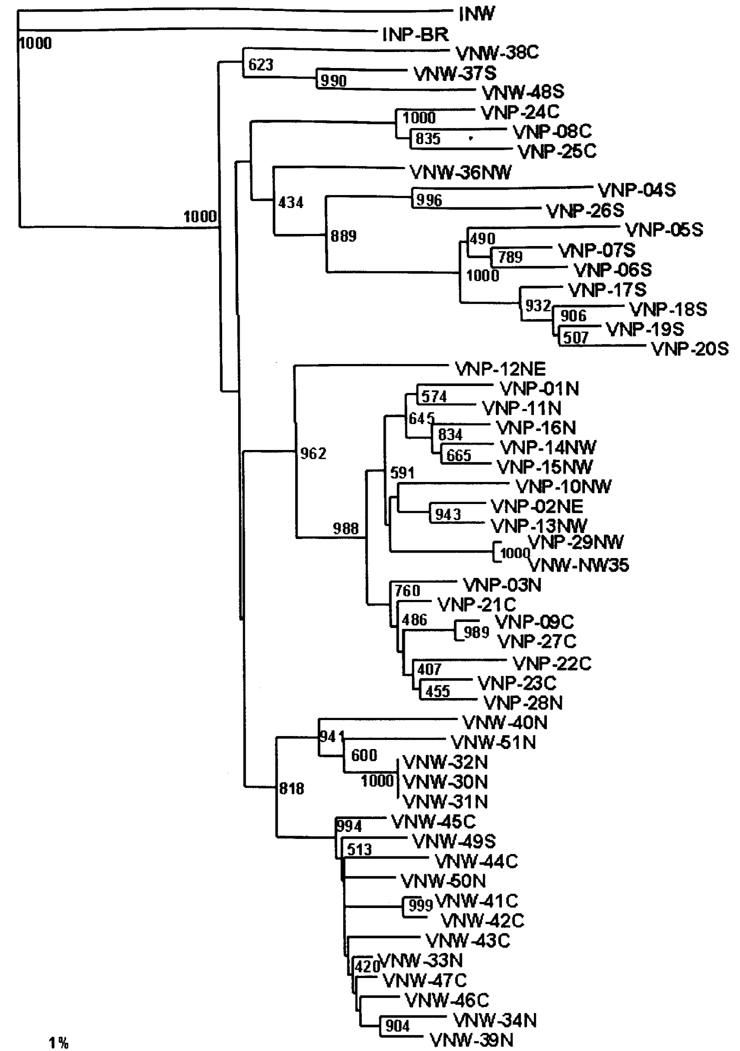


Fig. 4. Neighbour-joining trees showing the relationships of PRSV sequences collected within Thailand (A) and Vietnam (B). Acronyms are followed by region within the country where the isolate was collected: N, north; NE, north-east; NW, north-west; C, central; S, south. The number of bootstrap trees (greater than 400/1000) in which particular nodes were found is shown.

isolate. P isolates from Thailand diverged together, while Thai W isolates were dispersed among other South-East Asian isolates. These relationships suggest that there has been considerable mixing and movement of isolates in South-East Asia, fitting to some degree with the relative proximity of these countries.

Our phylogenetic analyses indicate that the most diverse isolates are from the Indian subcontinent (Indian/Sri Lankan); they form one of the basal groups, sister to all others, and two are present in one of the major sublineages, and hence they probably represent the oldest population of PRSV and indicate that PRSV may have arisen in South Asia. The node linking all PRSV isolates to the outgroup was also consistent when analysed with zucchini yellow mosaic virus (ZYMV) as the outgroup (data not shown).

Sequence diversity within Thailand and Vietnam

We also looked closely at the molecular variability in PRSV within Thailand and Vietnam. We compared sequences from a total of 15 isolates of P and W from Thailand and found up to 10·6% nucleotide and 10·1% amino acid sequence divergence within the PRSV population as a whole; however, the majority of this variability was between the cucurbit isolates (Fig. 2). There was 5·4–9·3% nucleotide (mean 7·6%) divergence between W isolates that were definitively biotyped (THW#) and mean divergence of 8·2% between the cucurbit isolates THU09 and THU10 and other PRSV-W isolates. In contrast, isolates of PRSV-P were much closer with a mean divergence of 2·6% at both nucleotide and amino acid levels. When analysed phylogenetically, we found all isolates of PRSV-P diverged from a common branch within a diverse background of W isolates (Fig. 4A). Interestingly the P cluster also included two cucurbit isolates, THU09 and THU10, collected from central Thailand.

Unfortunately, these isolates were not part of the original collection and the biotype was not confirmed by host range. There did not appear to be any significant grouping of P or W isolates according to region and bootstrap values for several of the nodes linking W isolates were low, suggesting uncertainty in their relationship to each other. There is strong evidence from the tight clustering of P isolates in such a variable background of W isolates that all the PRSV-P isolates in Thailand arose from a single mutation in an individual of a more ancient PRSV-W population and the resulting PRSV-P population spread throughout the country. Similarity in the N terminus, including the presence of a conserved aspartic acid residue seven amino acids downstream of DAG motif in all Thai P isolates (Fig. 1), supports this.

In Vietnam we found a different scenario. Initial information using heteroduplex analysis to look at variability in the CP suggested there was significant sequence diversity in both PRSV-P and W isolates (results not shown). This was confirmed through sequence comparison of the CP-coding region from 29 isolates of PRSV-P and 22 isolates of PRSV-W representing south, central, north, north-east and north-west Vietnam (Fig. 2). Between P isolates the mean divergence was 6·4% (max. 10·9%) and 4·1% (max. 7·6%) at nucleotide and amino acid levels, respectively, while between W isolates the mean divergence was only 4·1% at both nucleotide and amino acid levels. When analysed phylogenetically (Fig. 4B), we found that all isolates of PRSV-P from south Vietnam formed a single clade with strong support from bootstrapping (889/1000). As well, all P isolates from north Vietnam grouped together (962/1000) in a cluster which also included five of the eight central Vietnamese P isolates. The remaining central P isolates formed a third cluster although their relationship to other isolates was tenuous. Most isolates of PRSV-W from north and central Vietnam, with the exception of two north-west isolates (VNW-36 and VNW-35) and a central isolate (VNW-38), formed a single clade of closely related sequences. These isolates were collected from a range of different hosts including pumpkin, cucumber and loofah. This group also included one southern W isolate (VNW-49). The wide geographical occurrence of this sequence and the low variation suggests a recent spread of this 'strain' throughout country. The remaining two southern W isolates grouped together with a central W isolate (VNW-38). Overall, there was greater variability in the southern isolates of both P and W compared to those from the north. While the situation in Vietnam is more complex that seen in other countries, the relationship between several W isolates (VNW-35, VN-W36) and the P isolates, supports the theory that P arises by mutation from PRSV-W on occasion.

There is also evidence, as seen in Fig. 3, to suggest that in addition to mutation and local movement of PRSV within Vietnam, there have been introductions from other regions. This is evident from the grouping of PRSV-P sequences from south Vietnam and the Philippines, as well as the relationship between isolates of PRSV-P from Taiwan and Japan with several isolates of PRSV-P from central Vietnam.

Discussion

Much sequence variation was found within PRSV populations, but all were clearly isolates of PRSV (Shukla *et al*., 1994; Ward *et al*., 1995) and their differences were still less than has been reported for some other potyviruses, such as *Yam mosaic virus* (YMV) (Bousalem *et al*., 2000), where nucleotide sequence diversity as great as 28% has been reported. The greatest variability as well as all differences in length of the CP occurred in the N terminus of the CP; however, there was still considerable conservation within this region. This is particularly true of the regions adjacent to the DAG motif, which is associated with aphid transmission (Shukla *et al*., 1994), and the core. The fact that so much of the N terminus of the PRSV CP was conserved certainly suggests that this region has a functional or structural role, as previously suggested (Silva-

Rosales *et al*., 2000). There is no evidence implicating the N terminus, or any other part of the CP, in the ability to infect papaya.

What is shaping PRSV populations?

In general, variation in PRSV is still primarily related to geographical location rather than biotype (host range) as was previously reported (Bateson *et al*., 1994). The clustering of P and W isolates within regions including India, Australia, Brazil, USA, Thailand and Vietnam supports the hypothesis that PRSV-P does arise from PRSV-W by mutation, but that this occurs rarely. It is not known what factors contribute to this; however, evidence from comparisons of full-length genomes of PRSV-P and -W in Australia suggests that this mutation involves changes to one or more single amino acids (N. Jayathilake, J. Henderson, J. L. Dale & M. J. Bateson, unpublished). It may be that these changes arise frequently in cucurbit PRSV populations as a result of random, natural mutation but the virus is only established in papaya when there is a critical host mass such as the establishment of commercial plantations. Thus, it appears that movement of the virus around the world in cucurbits and then mutation to infect papaya is still a major factor in the molecular epidemiology of PRSV. However, at the same time, there is also considerable evidence to suggest movement of the virus in papaya. Unfortunately, the lack of sequence data for PRSV-W from many countries makes it difficult to confirm this. The relative contribution of PRSV-P to variation in cucurbit PRSV populations is not certain, as there is no evidence that PRSV-P subsequently reinfects cucurbits in the field. This is seen in the inability to detect PRSV-P in cucurbits in the field, despite extensive surveys (Gonsalves, 1998). It is possible that following mutation, as the virus population gains fitness through selection in papaya, it may gradually lose the ability to infect cucurbits. This is supported by the apparent difficulty in mechanically inoculating some PRSV-P isolates to some or all cucurbit species. Interestingly, field transmission of PRSV-P from papaya to cucurbits has only been demonstrated in a single experimental trial where cucurbits were grown near a field of PRSV-P infected papaya in Australia (Persley, 1998) and was attributed to the isolate being a new mutation.

In addition to natural mutation, recombination has been reported to contribute significantly to the generation of sequence diversity in a number of plant virus families (reviewed in Garcia-Arenal *et al*., 2001). When we looked at the possible influence of recombination in PRSV populations, we found only a single, definite crossover event in all of the sequences studied. This is similar to that reported for a natural population of *Plum pox virus* (Cervera *et al*., 1993). Few data are available for recombination in natural potyvirus populations, although recently multiple recombination events were identified in YMV populations (Bousalem *et al*., 2000). Based on analysis of the CP, it would seem that recombination is much less significant than movement and mutation in the molecular evolution of PRSV. However, recombination may be more frequent in other regions of the genome.

The origin of PRSV

From this study, there is evidence that PRSV may have arisen in Asia and, based on analysis of available sequences, this appears to be in the region of the Indian subcontinent (India/Sri Lanka). As discussed above, there is also significant, albeit somewhat circumstantial, evidence from numerous sources that PRSV is primarily a pathogen of cucurbits. This is also supported by the diversity of cucurbit-infecting potyviruses and virus isolates that are serologically related to PRSV (Quiot-Douine *et al*., 1990), a situation not seen in papaya. It is interesting to note that PRSV has not been reported from southern Africa while MWMV, which is phylogenetically closer to PRSV than to other cucurbit potyviruses (Lecoq *et al*., 2001), has been found predominantly only in Africa (North and South) and Europe (Quiot-Douine *et al*., 1990). This relationship supports the hypothesis that PRSV originated in cucurbits somewhere in the region that extends between North Africa and India. This is also in agreement with the origin of many cucurbits in Africa and Asia [*Cucumis sativus* (cucumber), *Benincasa hispida* (wax or white gourd), *Luffa* and possibly *Cucumis melo*] (IBPGR, 1983). It is possible that PRSV may have then gone into papaya in this region when papaya arrived in the Indo/China region in the 16th–17th century (Purseglove, 1968).

Mexico also appears to have a pivotal role in the epidemiology of PRSV. The grouping of different Mexican isolates with those from Australia, USA and Puerto Rico and the level of divergence seen between the Mexican P isolates certainly implicates Mexico in the spread of PRSV to these countries. It may be significant that three of the four genera of the *Caricaceae*, including *Carica*, are native to tropical America (Purseglove, 1968). However, the relative role of papaya and cucurbits in the spread of PRSV from this region is unknown, although in Australia it appears to have been through cucurbits, as PRSV-W was present for at least 20 years before PRSV-P was reported (Greber, 1978; Thomas & Dodman, 1993). Unfortunately, no sequence data for PRSV-W from Mexico are available.

Ultimately, the origin and epidemiology of PRSV can only be further defined with the generation of more sequence data for isolates of PRSV-P and -W within different countries, particularly South America, North Africa, the Middle East and India.

Variation within countries

Previously, there have been very few studies looking at sequence variability in PRSV from different countries/regions and most reports have suggested relatively low levels of variation within these regions (Silva-Rosales *et al*., 2000; Davis & Ying, 1999; Bateson *et al*., 1994). As well, much of this data

has been for P isolates. The only comparison of W isolates from a single country was from Australia (Bateson *et al*., 1994) where there was less than 2% variation. In contrast to this, Jain *et al*. (1998) reported very high sequence variation between a single P and W isolate from India; however, there was no information on the variability within Indian P or W populations. It is now clear that there are high levels of sequence variation in PRSV within many countries. Here we have demonstrated, for the first time, much higher levels of variation in W populations within a single geographical region. In Thailand, particularly, variation in W isolates is not a single highly divergent sequence but a highly variable population of sequences, suggesting that the virus has been present and evolving for some time in cucurbits in that country. This also appears to be true for W isolates in Vietnam, although the level of variation was lower. In contrast, variation in PRSV-P differed markedly in these countries. In Thailand variation in P was low and could be attributed to a relatively recent mutation, which fits with the first report of PRSV-P in Thailand (Srisomchai, 1975); however, in Vietnam, variation was much greater, apparently the result of introduction from other countries. It appears that the profile of variation will be different in each country. This can be attributed to the relative contribution of natural variation, mutation to infect papaya and movement of isolates, all of which will be influenced by proximity to and interaction with other countries, and agronomic practices which may enhance the opportunity to change host. In a practical light, the complex scenarios and high levels of divergence in PRSV in different countries will pose a significant challenge to control of the virus. The success of many current control strategies being used against PRSV, particularly genetically engineered resistance (reviewed in Gonsalves, 1998) and mild strain cross-protection (Yeh & Gonsalves, 1994; Rezende & Pacheco, 1998), relies on low sequence variation within the countries/regions being targeted (Tennant *et al*., 1994) and the continued exclusion of isolates with greater variability. This may be achieved in countries such as Australia, where low variation already exists in both P and W populations and where effective quarantine and restriction on movement of infected material can be adequately ensured (Thomas & Dodds, 1993). However, in other countries, the success of particular control methods may differ depending on their individual PRSV profiles. Even where divergence in PRSV-P is low, highly divergent cucurbit populations may act as potential reservoirs of new P sequences, while exclusion of isolates may also be more difficult where countries are in close proximity.

The work reported here was funded by the Australian Centre for International Agricultural Research (ACIAR). W. Chaleeprom was supported by a scholarship from the Government of Thailand and C. V. Ha by the Hanoi Agricultural University. Professor V. T. Mann is thanked for his help with sample collection and for use of facilities in Vietnam. Authors Lines, Chaleeprom and Revill contributed equally to the work reported in this paper.

References

Bateson, M. F., Henderson, J., Chaleeprom, W., Gibbs, A. J. & Dale, J. L. (1994). Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P in Australia. *Journal of General Virology* **75**, 3547–3553.

Bousalem, M., Douzery, E. J. P. & Fargette, D. (2000). High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution. *Journal of General Virology* **81**, 243–255.

Cervera, M. T., Riechmann, J. L., Martin, M. T. & Garcia, J. A. (1993). 3′-Terminal sequence of the plum pox virus PS and ŏ6 isolates: evidence for RNA recombination within the potyvirus group. *Journal of General Virology* **74**, 329–334.

Chaleeprom, W. (1997). *Genome analysis of Papaya ringspot potyvirus and a related virus*. PhD thesis. Queensland University of Technology, Australia.

Davis, M. J. & Ying, Z. (1999). Genetic diversity of the *Papaya ringspot virus* in Florida. *Proceedings of the Florida State Horticultural Society* **112**, 194–196.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Feng, D. F. & Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* **25**, 351–360.

Garcia-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology* **39**, 157–186.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573–582.

Gonsalves, D. (1998). Control of papaya ringspot virus in papaya: a case study. *Annual Review of Phytopathology* **36**, 415–437.

Greber, R. S. (1978). *Watermelon mosaic virus* 1 and 2 in Queensland cucurbit crops. *Australian Journal of Agricultural Research* **29**, 1235–1245.

IBPGR (1983). *Genetic Resources of Cucurbitaceae*. Rome, Italy: Crop Genetic Resources Centre, Plant Production and Protection Division, Food and Agriculture Organization of the United Nations.

Jain, R. K., Pappu, H. R., Pappu, S. S., Varma, A. & Ram, R. D. (1998). Molecular characterisation of *Papaya ringspot potyvirus* isolates from India. *Annals of Applied Biology* **132**, 413–425.

Jensen, D. D. (1949). Papaya virus diseases with special reference to papaya ringspot. *Phytopathology* **39**, 191–211.

Lecoq, H., Dafalla, G., Desbiez, C., Wipf-Scheibel, C., Delécolle, B., Lanina, T., Ullah, Z. & Grumet, R. (2000). Biological and molecular characterization of *Moroccan watermelon mosaic virus* and a potyvirus isolate from Eastern Sudan. *Plant Disease* **85**, 547–552.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Persley, D. M. (1998). *Identification, epidemiology and control of papaya ringspot virus, recently recorded in papaya (Carica papaya) in Australia*. Masters thesis, Queensland University of Technology, Australia.

Purcifull, D. E., Edwardson, J. R., Hiebert, E. & Gonsalves, D. (1984). Papaya ringspot virus. *CMI/AAB Descriptions of Plant Viruses*, no. 292. Wallingford, UK: CAB International.

Purseglove, J. W. (1968). *Caricaceae. Tropical Crops: Dicotyledons*. London: Longman.

Quemada, H., Hostis, B. L., Gonsalves, D., Reardon, I. M., Heinrikson,

R., Hiebert, E. L., Sieu, L. C. & Slightom, J. L. (1990). The nucleotide sequence of the 3′ terminal regions of papaya ringspot virus strains W and P. *Journal of General Virology* **71**, 203–210.

Quiot-Douine, L., Lecoq, H., Quiot, J. B., Pitrat, M. & Labonne, G. (1990). Serological and biological variability of virus isolates related to strains of papaya ringspot virus. *Phytopathology* **80**, 256–263.

Rezende, J. A. M. & Pacheco, D. A. (1998). Control of papaya ringspot virus type-W in zucchini squash by cross protection in Brazil. *Plant Disease* **82**, 171–175.

Shukla, D. D., Ward, W. W. & Brunt, A. A. (1994). *The Potyviridae*. Wallingford, UK: CAB International.

Silva-Rosales, L., Becerra-Leor, N., Ruiz-Castro, S., Téliz-Ortiz, D. & Noa-Carrazana, J. C. (2000). Coat protein sequence comparisons of three Mexican isolates of *Papaya ringspot virus* with other geographical isolates reveal a close relationship to American and Australian isolates. *Archives of Virology* **145**, 835–843.

Srisomchai, T. (1975). Studies on papaya ringspot virus disease. Office of Northeast Agriculture and Co-operative Annual Report, Tha Phra, Khon Kaen, Thailand. (In Thai.)

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.

Strimmer, K. & von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences, USA* **94**, 6815–6819.

Tennant, P. F., Gonsalves, C., Ling, K. S., Fitch, M., Manshardt, R., Slightom, J. L. & Gonslaves, D. (1994). Differential protection against papaya ringspot virus isolates in coat protein gene transgenic papaya and classically cross-protected papaya. *Phytopathology* **84**, 1359–1366.

Thomas, J. E. & Dodman, R. L. (1993). The first record of papaya ringspot virus-type P in Australia. *Australian Plant Pathology* **22**, 2–7.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876–4882.

Tomlinson, J. A. (1987). Epidemiology and control of virus diseases and vegetables. *Annals of Applied Biology* **110**, 661–681.

Wang, C. H., Bau, H. J. & Yeh, S. D. (1994). Comparison of the nuclear inclusion B protein and coat protein genes of five papaya ringspot virus strains distinct in geographic origin and pathogenicity. *Phytopathology* **84**, 1205–1210.

Ward, C. W., Weiller, G. F., Shukla, D. D. & Gibbs, A. J. (1995). Molecular systematics of the *Potyviridae*, the largest plant virus family. In *Molecular Basis of Virus Evolution*, pp. 477–497. Edited by A. Gibbs, C. H. Calisher & F. Garcia-Arenal. Cambridge: Cambridge University Press.

Yeh, S. D. & Gonsalves, D. (1994). Practice and perspectives of control of papaya ringspot virus by cross protection. *Advances in Disease Vector Research* **10**, 237–257.

Received 26 February 2002; Accepted 10 June 2002