THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Plant Pathology)

DEGREE

Plant Pathology | Plant Pathology

FIELD | **DEPARTMENT**

TITLE: Construction of an Infectious Full-length cDNA Clone of a *Papaya ringspot virus* Type W

NAME: Mr. Pipat Chiampiriyakul

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Associate Professor Supat Attathom, Ph.D.)

COMMITTEE MEMBER

(Assistant Professor Wichai Kositratana, Ph.D.)

COMMITTEE MEMBER

(Associate Professor Sutevee Sukprakarn, Ph.D.)

DEPARTMENT HEAD

(Assistant Professor Pissawan Chiemsombat, Dr.Agr.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Vinai Artkongharn, M.A.)

THESIS

CONSTRUCTION OF AN INFECTIOUS FULL-LENGTH CDNA CLONE OF A *PAPAYA RINGSPOT VIRUS* TYPE W

PIPAT CHIAMPIRIYAKUL

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Plant Pathology)
Graduate School, Kasetsart University
2007

Pipat Chiampiriyakul 2007: Construction of an Infectious Full-length cDNA Clone of a *Papaya ringspot virus* Type W. Doctor of Philosophy (Plant Pathology), Major Field: Plant Pathology, Department of Plant Pathology. Thesis Advisor: Associate Professor Supat Attathom, Ph.D. 349 pages.

The purpose of this research was to construct an infectious full-length cDNA of a Thai isolate of *Papaya ringspot virus* type W (PRSV-W) which causes serious problems in cucurbit crops nationwide. The full-length cDNA of PRSV-W was constructed from eight overlapping cDNA clones covering the whole viral genome except the 3'terminal poly(A) nucleotides.

Two full-length PRSV-W cDNA clones under the control of the enhanced CaMV 35S (pDPCwPN7978) or bacteriophage T7 (pDPCwAN0178) promoters were successfully constructed in this study but they were not stable to be propagated in *E. coli* as bacterial host cells. Due to the instability of the full-length cDNA clones, linear-formed constructs corresponding to the two full-length cDNA clones without plasmid vector sequences were developed from overlapping cDNA clones by overlap extension PCR amplification. Linear full-length PRSV-W cDNAs (PCwAH8584) were constructed downstream from the bacteriophage T7 RNA polymerase promoter. The *in vitro* capped transcripts generated from the PCwAH8584 cDNAs were infectious in pumpkin plants. The infectivity of the *in vitro*-transcribed RNA was confirmed by serological detection of the virus coat protein and by observation of virion by electron microscopy.

The nucleotide sequence of the RNA of PRSV-W Thai isolate is 10323 nucleotides in length excluding the 3'terminal poly(A) tail and contains a single open reading frame (ORF) of 10032 nucleotides encoding a large polyprotein of 3343 amino acids and predicted molecular weight of 380.3 kDa with 5'- and 3'- non coding regions (NCRs) consisted of 85 and 206 nucleotides, respectively.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I would like to thank my committee members, all of whom have conduced to my personal development. I would like to gratefully thank my thesis advisor, Assoc. Prof. Dr. Supat Attathom, for his valuable advice, constant belief in my abilities and giving me the wonderful opportunity to work this research in Molecular Virology Laboratory at KU-BIOTEC, partially financial support for this work and manuscript preparation. I would like to thank Asst. Prof. Dr. Wichai Kositratana and Assoc. Prof. Dr. Sutevee Sukprakarn, my committee members and also Assoc. Prof. Dr. Nawarat Udomprasert from Graduate School for their careful proofreading the manuscript and valuable suggestions.

I thank Dr. Wanpen Srithongchai for providing PRSV-W isolate used in this work, Dr. Srimek Chowpongpang for his valuable guidance in the cloning of a full-length cDNA of PRSV-W genomic RNA and his thoughtfulness, Ms. Namthip Phironrit for the gift of PRSV antibodies, Mrs. Ratree Koohapitakthum for the managing of *in vitro* cultured-pumpkin plants. I also thank Ms. Jiraphan Sopee, who helped me in some parts of preparation of this thesis and of a previous manuscript.

This research was financially supported by a grant from the Graduate School, Kasetsart University. The academic study for Ph.D. degree in Plant Pathology Program at Kasetsart University was financially supported by a Thai Government Scholarship (UDC). Many thanks also to Ms. Sasawan Chiampiriyakul and Police Colonel Akkara Chiampiriyakul, my elder sister and brother, respectively, for their encouragement during the academic study.

Finally, this thesis is dedicated to the memory of my parents, who provide me endless for the most and best education possible.

Pipat Chiampiriyakul

October 2007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

CONSTRUCTION OF AN INFECTIOUS FULL-LENGTH CDNA CLONE OF A *PAPAYA RINGSPOT VIRUS* TYPE W

INTRODUCTION

Papaya ringspot virus type W (PRSV-W) formerly known as *Watermelon mosaic virus* 1 (WMV-1) belongs to the genus *Potyvirus* in the family *Potyviridae* (Purcifull *et al*., 1984a; Provvidenti, 1996) which contains the largest member of plant virus species (Mayo and Pringle, 1998). It is naturally transmitted in a non-persistent manner by several species of aphid. The PRSV-W affects all agricultural species of the Cucurbitaceae, and is the major limiting factor for cucurbit production in warm regions worldwide (Lecoq, 2003) because of its destructiveness. The virus causes diverse symptoms that include mosaic, mottling, blister and deformation on leaves and fruits of cultivated cucurbit species. PRSV, a definitive species, occurs as two closely related pathotypes which can be distinguished by host range (Yeh *et al*., 1984) including papaya-infecting (P) and cucurbit-infecting (W) types. However, the both P and W isolates are found to be indistinguishable serologically (Gonsalves and Ishii, 1980). PRSV isolates are either type P or W. PRSV-P isolates naturally infect both papaya and cucurbits, while PRSV-W isolates naturally infect cucurbits but not infect papaya.

PRSV particles are flexuous rods of approximately 780 nm in length and 12 nm in diameter. Their monopartite genome consists of a single-stranded, positive-sense RNA of about 10.3 kilobases long (Yeh and Gonsalves, 1985; Yeh *et al*., 1992) that is 3' polyadenylated (Hari *et al*., 1979) and contains a 5' terminal genome-linked protein (VPg) (Riechmann *et al*., 1989; Murphy *et al*., 1990). The PRSV RNA genome has a large single open reading frame that is encoded into a large polyprotein of approximately 381 kilodaltons and subsequently proteolytically cleaved into ten functional proteins (Yeh *et al*., 1992; Wang and Yeh, 1997). The proposed genetic organization of PRSV is similar to those of other potyviruses except that the first protein (P1) is larger than those of all reported potyviruses.

Analysis of complete genomic RNA sequence of individual potyviruses has encouraged our understanding of potyviral genome organization and relationships between the genomes besides proposed roles for several mature proteins. Complete nucleotide sequences of the genomic RNAs of several (at least eleven) isolates of PRSV representing the two pathotypes (P or W) have been reported. Around the whole genome excluding the poly (A) tail, the different genomes of the available complete PRSV sequences showed that the genome sizes of these PRSV isolates range from 10320 to 10334 nucleotides. One method to differentiate closely related virus species, closely related virus strains or even different isolates is to compare their nucleotide or amino acid sequences, particularly of the 3'-end noncoding region (NCR) and coat protein (CP) coding sequences. Phylogenetic analysis is also useful for defining potyvirus species and grouping different isolates which belong to the same species.

Molecular genetic analyses of the structure and function of RNA virus genomes and plant–virus interactions have been intensively advanced by the availability of full-length cDNA clones, the sources of infectious RNA transcripts which were prepared either *in vivo* or *in vitro*, that replicate efficiently when introduced into host cells. In the study of molecular biology of viruses with RNA genomes, it has been hampered by the inherent limitations imposed by the nature of their genetic material. With the advancement of recombinant DNA technology, viruses containing either single- or double-stranded RNA genomes have been genetically analyzed at the molecular level. The RNA from viral sequences can be manipulated by converting viral RNA genomes into complementary DNAs and replicating them as plasmid inserted in bacterial hosts. In the case of positive-sense RNA viruses, techniques for generating infectious RNA from *in vivo* or *in vitro* transcription of full-length cDNA clones have been developed. These implicate inoculation with either infectious RNA transcripts from the cDNA synthesized *in vitro* or cDNA plasmids containing the viral genomic sequences flanked by transcription promoter and termination sequences, usually derived from *Cauliflower mosaic virus* (CaMV). These infectious RNA transcripts or cDNA clones have been produced from a variety of RNA plant viruses.

Plant viral positive-sense RNA genome can be cloned as double-stranded DNA in *Escherichia coli*, and that RNAs transcribed *in vitro* from such cloned DNA were infectious to plants (Boyer and Haenni, 1994). This advance built on the ability of promoters derived from viruses infecting *E. coli* (such as SP6, T3 orT7) to be manipulated *in vitro*, using bacteriophage polymerases. Alternately, cloned cDNA of plant viral genome containing a strong eukaryotic RNA polymerase promoter derived from CaMV (the 35S promoter) could be constructed and utilized in expressing of the viral genome in host plant to avoid the *in vitro* transcription step. This approach has proven useful for a number of positive strand RNA viruses.

Infectious RNA is obtained from full-length cDNA clones either by *in vivo* transcription from a strong constitutive promoter active in the host cell or by *in vitro* transcription of RNA ordinarily from a bacteriophage promoter (Boyer and Haenni, 1994). *In vivo*- or *in vitro*-transcribed infectious RNA derived from full-length cDNA clones is a powerful tool for the study of virus pathogenicity and of gene expression at the RNA level, particularly for the large number of RNA viruses whose natural expression does not involve DNA intermediates. *In vivo*- or *in vitro* transcription is a logical bridge between genetic manipulations currently possible only at the cDNA level. The ability to use such genetic manipulations offers considerable potential for a wide range of studies in virology. This approach has proven practicable for a number of positive strand RNA viruses including PRSV type P. *In vivo* and *in vitro* infectious clones of PRSV have been reported for a Hawaii isolate (HA) (Chiang and Yeh, 1997) that was placed in pathotype P. They were constructed downstream from a *Cauliflower mosaic virus* (CaMV) 35S and a bacteriophage T3 promoters, respectively.

As a first step in the development of a molecular genetic system for the consideration of RNA viruses, this thesis describes construction of full-length cDNA clones of a Thai isolate of PRSV-W RNA. They will be later used to produce infectious cDNA clone (for *in vivo* transcription) and RNA transcripts *in vitro*.

Two types of infectious cDNA clones of PRSV-W Thai isolate reported in this thesis were: (1) cDNA clones under the control of the 35S promoter of *Cauliflower mosaic virus*, whose plasmid or linear DNA is directly inoculated onto pumpkin cotyledons by particle bombardment, and (2) a cDNA clone under the control of the bacteriophage T7 promoter, to be used as a template for the synthesis of infectious *in vitro* transcripts.

Additionally, complete nucleotide sequence of the PRSV-W genome excluding poly (A) tail have been determined from eight overlapping viral cDNA clones generated in this research. Their nucleotide and deduced amino acid sequences were analyzed to determine phylogenetic relationships with the different eleven PRSV isolates for which the complete sequence is available. Genome organization and conserved sequence motifs of the PRSV-W Thai isolate were also analyzed and proposed accordingly.

OBJECTIVES

The objectives of this research were:

1. To construct an infectious full-length cDNA clone of a Thai isolate PRSV-W genome.

2. To analyze complete nucleotide and deduced amino acid sequences of a Thai isolate PRSV-W in relation to its genome expression and functions.

3. To determine phylogenetic relationship between the PRSV-W Thai isolate and other different PRSV isolates for which the complete sequence is available.

LITERATURE REVIEW

1. Cucurbits and virus diseases

Cucurbits are important vegetable crops in developing countries where they account for more than 50% of the total fresh vegetable production. About 40 different cucurbit species are cultivated, but in term of production, the 3 major crops are watermelon (*Citrullus lanatus*), cucumber (*Cucumis sativus*) and melon (*Cucumis melo*). Another group of economically important cucurbits includes different squash, pumpkin or marrow types (*Cucurbit pepo*, *Cucurbita moschata* and *Cucurbita maxima*). These species are cultivated worldwide as a diversity of cultivars and landraces adapted to the local user and consumers' demands. Other species have a more limited cultivation area. Bitter gourd (*Momordica charantia*), snake gourd (*Trichosanthes cucumerina*) and wax gourd (*Benincasa hispida*) are mostly grown in Asia and fluted pumpkin (*Telfairia occidentalis*) in Africa. Although distributed worldwide chayote (*Sechium edule*) is particularly popular in Central America and in the Caribbean (Robinson and Deckers-Walters, 1997).

Viral diseases cause important economic losses throughout the world. More than 35 viruses have been isolated from cucurbits (Provvidenti, 1996). These viruses constitute complex and dynamically changing problems (Nameth *et al*., 1986). This viral diversity may be a consequence of the genetic and ecological diversity of their hosts. Since, cucurbits are grown in a large variety of climatic, environmental and agricultural conditions, Thus variety of environments may provide more or less favorable condition for specific viruses or for their vectors.

Cucurbitaceous vegetable crops can be destroyed when infected by several virus families such as *Geminiviridae*, *Closteroviridae*, *Bromoviridae*, *Luteoviridae* and *Potyviridae*. Some viruses in several genera in these families have been reported as a group of major viruses infecting cucurbits that present a worldwide distribution and cause important yield losses (Lecoq, 2003). These cucurbit viruses are listed in Table 1.

Table 1 Major cucurbit infecting viruses that present a worldwide distribution

Family	Genus	Virus	Means of spread	Distribution in countries
Geminiviridae	*Begomovirus*	*Squash leaf curl virus* (SLCV)	Whiteflies	Central America, Caribbeans
	Begomovirus	*Watermelon chlorotic stunt virus* (WmCSV)	Whiteflies	Red sea region, Iran
Closteroviridae	*Crinivirus*	*Cucurbit yellow stunt disorder virus* (CYSDV)	Whiteflies	Middle East, Africa, Mexico
Bromoviridae	*Cucumovirus*	*Cucumber mosaic virus* (CMV)	Aphids	Worldwide
Luteoviridae	*Polerovirus*	*Cucurbit aphid-borne yellows virus* (CABYV)	Aphids	Worldwide
Potyviridae	*Potyvirus*	*Moroccan watermelon mosaic virus* (MWMV)	Aphids	Africa
	Potyvirus	*Papaya ringspot virus* (PRSV)	Aphids	Worldwide, tropical and subtropical
	Potyvirus	*Watermelon mosaic virus* (WMV)	Aphids	Worldwide, temperate
	Potyvirus	*Zucchini yellow mosaic virus* (ZYMV)	Aphids, seed	Worldwide

Source: Lecoq (2003)

2. Potyviruses infecting cucurbits

Four major potyvirus infecting cucurbit including MWMV, WMV and ZYMV are in the same genus (*Potyvirus*), and these viruses relate to *Potyvirus* PRSV-W. General description of these potyviruses is summarized in this review in order to exploit as relative information for further understanding.

2.1 *Moroccan watermelon mosaic virus* (MWMV)

Moroccan watermelon mosaic virus (MWMV) was first described as potyviruses causing losses in cucurbit planting in Morocco and as a strain of WMV due to its biological properties comparable to WMV and reacted to an antiserum to WMV from Germany, although the identity of the virus isolate to which the antiserum was produced was not indicated (Fischer and Lockhart, 1974). MWMV and closely serologically related strains are wide spread in different parts of Africa, including Morocco, Algeria, Sudan, Niger, Cameroon, Zimbabwe, and South Africa and in Southern Europe (Lecoq *et al*., 2001). MWMV infects and causes severe symptoms in cucumber, squash, and watermelon including mosaic and deformations in leaves and fruits. In most melon cultivars symptoms are systemic necrotic spots frequently followed by a complete collapse and necrosis of the plants (Lecoq, 2003).

2.2 *Watermelon mosaic virus* (WMV)

Watermelon mosaic virus (WMV) was formerly known as *Watermelon mosaic virus* 2 (WMV2). The virus was first reported by Webb and Scott (1965). It was isolated from cucurbits at several locations in the United States. This virus and PRSV-W were once considered to be strains of the same virus (Milne and Grogan, 1969), but now are identified as distinct individual based on serological, molecular and biological characteristics (Purcifull and Hiebert, 1979). WMV is easily mechanically transmissible and naturally transmitted by several species of aphid such as *Aphis craccivora*, *A. gossypii* and *Myzus persicae* in a non-persistent manner (Castle *et al*., 1992). Symptoms of WMV in natural hosts include vein banding,

mosaic, leaf distortion and plant stunting. Symptom expressions of WMV are not constant according to the cultivar. The virus can also infect several weed species that can be reservoirs of the virus during the winter season. Various levels of resistance to WMV have been identified in cucumber, melon, watermelon squash, bean and pea.

2.3 *Zucchini yellow mosaic virus* (ZYMV)

Zucchini yellow mosaic virus (ZYMV), tentatively named *Muskmelon yellow stunt virus* (MYSV), was first detected in *Cucurbita pepo* in Italy by Lisa *et al*. (1981). The virus was the first of a series of 'emerging' cucurbit virus that have menacing the cucurbit industry since the early 1980s. ZYMV distribute to the major cucurbit producing region in many countries including Australia, France, Germany, Italy, Israel, Japan, Lebanon, Mauritius, Morocco, Spain, UK and USA (Lisa and Lecoq, 1984). The virus causes severe diseases consisting of mosaic, yellowing, blister, shoestring, leaf distortion, enation or filiformism on leaves, stunting and fruit and seed deformations in zucchini, squash, muskmelon, cucumber and watermelon, making them unmarketable. Its early infections frequently conduct to a complete yield loss. McLean *et al*. (1982) found that ZYMV was able to reduce yields of cucurbit crops from 0 to 99%.

2.4 *Papaya ringspot virus* (PRSV)

Papaya Ringspot virus (PRSV) is a plant virus as a species belonging to genus *Potyvirus* in family *Potyviridae* (Shukla *et al*., 1994; Hull, 2002). PRSV isolates are either type P or W. PRSV-P (papaya-infecting) isolates cause serious problems in papaya throughout the tropical and subtropical regions, while PRSV-W (previously called as *Watermelon mosaic virus* 1) isolates are of economic importance to cucurbit crops throughout the world (Purcifull *et al*., 1984a). It is transmissible mechanically, and by multifarious species of aphid in a non-persistent manner, and has a narrow host range (Purcifull *et al*., 1984a). PRSV was first reported by Jensen (1949) in papaya (PRSV-P) and by Webb and Scott (1965) in cucurbits (PRSV-W). PRSV-W was considered for a long time as a distinct potyvirus, WMV1.

Subsequently, Gonsalves and Ishii (1980) immuno-detected and found that the WMV1 and PRSV-P were indistinguishable serologically, but the WMV1 was shown serologically distinct from WMV (WMV2) (Purcifull and Hiebert, 1979) and MWMV (Fischer and Lockhart, 1974).

Presently, WMV1 (PRSV-W) and PRSV-P are considered as two different pathotypes of PRSV (Purcifull *et al*., 1984a). Based on the host range specificity, PRSV is classified into two pathotypes which PRSV-P infects both papaya and able to infect a number of an experimental host range in cucurbits, while PRSV-W naturally infects and causes important diseases in a wide range of economically important cucurbit crops including squash, watermelon, cucumber cantaloupe and pumpkin, however, it is not able to infect papaya (Milne and Grogan, 1969; Purcifull and Hiebert, 1979; Yeh *et al*., 1984). Although PRSV-P can usually be transmitted experimentally to cucurbits, it is not usually detected in cucurbits in the field (Gonsalves, 1998). The outbreak of PRSV-P in papaya is similar to that of other non-persistent, aphid-borne potyviruses and it is not found in intercropped cucurbits, suggesting that the virus circulates directly from papaya to papaya. There is, however, little information on the epidemiology of PRSV-W in cucurbits.

2.5 *Papaya ringspot virus* pathotype W (PRSV-W)

The pathotype W of *Papaya ringspot virus* (PRSV-W) formerly known as *Watermelon mosaic virus* 1 (WMV-1), one of several potyviruses causing severe damage to the production of cucurbit crops in warm regions worldwide (Purcifull *et al*., 1984a). PRSV-W has been reported as the major virus infecting cucurbits in many regions including The United State of America, Brazil, Mexico, Central America and Caribbean countries, Australia, Germany, France, India, Nepal and South America (Webb *et al*., 1965; Milne *et al*., 1969; Makkouk and Lesemann, 1980; Purcifull *et al*., 1984b; Lecoq, 2003). It seems to be of minor importance in other regions such as the Middle East, Eastern Africa and China (Lecoq, 2003). PRSV-W causes severe mosaic, mottling, blisters and deformation on leaves and fruits of melon, cucumber, squash, cantaloupe, watermelon and other cultivated cucurbit

species. Some strains of PRSV-W induced systemic necrosis in some melon cultivars (Quiot-Douine *et al*., 1990). Cucurbitaceous weeds, such as *Melothria pendula* and *Momordica charantia* serve as reservoirs for the virus isolates (Adlerz, 1972a, b). Losses can be high due to the severe reduction in fruit yield and to altered fruit quality. The virus has been reported to reduce fruit yield of *Cucurbita pepo* by more than 90% in India if plants are infected 14 days after emergence, and fruit formed are very small. PRSV-W host range is mostly restricted to the *Cucurbitaceae* therefore virus sources are either cucurbit crops in regions where cucurbits are grown all year round or wild cucurbit species that can serve as reservoirs of the virus.

3. Genome organization of potyviruses

PRSV has flexuous filamentous particle 780 nm in length and 12 nm in diameter and its genome is positive-sense, single stranded (ss) RNA of about 10.3 kb long (Purcifull *et al*., 1984a; Yeh and Gonsalves, 1985; Yeh *et al*., 1992). The PRSV particle is encapsulated by a coat protein consisting of a single type of capsid protein of approximately 36 kDa (Purcifull and Hiebert, 1979; Gonsalves and Ishii, 1980). The genomic structure and organization of PRSV RNA was tentative proposed by Yeh *et al*. (1992). The genome organization of the virus is similar to those of the other potyviruses except that the first protein processed from the N terminus of the polyprotein, the P1 protein, is 18 kDa to 34 kDa larger than those of all reported potyviruses. The genomic RNA is 10326 nucleotides in length, exclusive of the poly (A) tract, and accommodates on large open reading frame (ORF) that starts at nucleotide position 86 and completes at position 10120, encoding a polyprotein of 3344 amino acid, which, has a calculated molecular weight of about 381 K (Yeh *et al*., 1992). The partial genetic maps of *Tobacco etch virus* (TEV) (Allison *et al*., 1986), *Tobacco vein motting virus* (TVMV) (Domier *et al*., 1986), *Plum pox virus* (PPV) (Maiss *et al*., 1989), *Potato virus Y* (PVY) (Robaglia *et al*., 1989) and *Pea seed-borne mosaic virus* (PSbMV) (Johansen *et al*., 1991). The genome structure and organization of PRSV have been proposed based on comparable similarities with other potyviruses, the main characteristics of the PRSV genome RNA, which have a genome-linked protein (VPg) covalent bound to the 5'terminus (Riechmann *et al*.,

1989) and is polyadenylated at its 3'end. Its genome RNA has a 5'non-coding region (5'NCR) of 85 nucleotides prior to the single large ORF rich in A and U residues and a 3'end non-coding region (3'NCR) contains 209 nucleotides terminating in a poly (A) tract (Yeh *et al*., 1992).

The PRSV genome contain a single large ORF of 10032 nucleotides initialing at residue 86-88 (AUG) and terminating at residue 10118-10120, which could code for a polyprotein with 3344 amino acids (Yeh *et al*., 1992). The polyprotein is subsequently cleaved by these different virus coded proteinases, which include P1, HC-Pro and NIa (Dougherty and Semler, 1993) at specific sites into nine products of smaller protein (Dougherty and Carrington, 1988). The order of these products in the polyprotein is: first protein (P1), helper component (HC), third protein (P3), cylindrical inclusion protein (CI), small cleavage product (6K), small nuclear inclusion protein (NIa) which includes VPg at its N-terminus, large nuclear inclusion protein (NIb) and coat protein (CP). Dougherty and Carrington (1988) found that only the two gene products, VPg and CP, have been detected in virus particles while other gene products such as P1, HC, P3, CI, NIa and Nib have been found in infected plant cells.

4. Functions of potyviral RNA encoded proteins

The Potyvirus RNA genome contains a single long open reading frame (ORF) translated into a large population that is post-translationally cleaved to produce cleavage intermediates or the final products. A large polyprotein is processed up to 9 functional proteins (locations in order: P1, HC-Pro, P3, CI, 6K, VPg, NIa-Pro, NIb and CP) by the activity of three viral proteinase, P1, HC-Pro and NIa. The parts of the Potyvirus genome corresponding to mature proteins are often called cistrons (Johansen *et al.*, 2001). This review focuses on the function of the potyvirus proteins during the viral life cycle. The proteins are discussed in the order in which cistron appear on the genome.

4.1 First protein (P1)

The P1 is the least conserved protein among potyviruses, excluding conserved amino acids detected at the C-terminus of all potyviral proteins and which corresponds to the proteinase catalytic domain (Domier *et al.*, 1987; Urcuqui-Inchima *et al.*, 2001). RNA binding activity has been reported for P1 of *Tobacco vein mottling virus* (TVMV) (Brantley and Hunt, 1993), of *Turnip mosaic virus* (TuMV) (Soumounou and Laliberte, 1994) and of *Potato virus A* (PVA) (Merits *et al.*, 1998). RNA binding of these P1 proteins acts in a sequence-unspecific manner. The RNA-binding properties of P1 may suggest that this protein functions in the RNA replication process. It has been suggested that P1 acts as an accessory factor for stimulatory effect on viral genome amplification (Verchot and Carrington, 1995; Moreno *et al.*, 1998; Merits *et al.*, 1999). Other functions of the P1 protein are an involvement in virus cell-to-cell movement (Arbatova *et al.*, 1998) and an essential factor in overcoming mechanisms of virus resistance which play a role in host range determination (Rajamaki *et al.*, 2005).

4.2 Helper component-proteinase (HC-Pro)

The Potyvirus helper component-proteinase (HC-Pro) is a non-structural and multifunctional protein involved in aphid mediated transmission, genome amplification, polyprotein processing, systemic movement, suppression of gene silencing and synergism. HC-Pro has been reported to be associated with coat protein suggesting an involvement in aphid transmission. By feeding insects with different virus/HC-Pro combinations, there are evidences have been shown that diverse aphid species transmit *Potyviruses* to different extents and correlation of the capacity of HC-Pro to retain a given virion in the insect stylets (Wang *et al.*, 1998). HC-Pro was shown to pass from cell-to-cell, to increase the size exclusion limit (SEL) of plasmodesmata, and to facilitate passage of viral RNA from one cell to another.

4.3 Third protein (P3)

The P3 cistron encoded the third part of the polyprotein of potyviruses. The P3 protein, protein positioned between HC-Pro and 6K1/CI protein products, is one of the most highly variable (Shukla *et al.*, 1994) and least comprehended potyviral proteins (Urcuqui-Inchima *et al*., 2001). The C-terminal region of the P3-6K1 complex contains a determinant of *Plum pox virus* (PPV) pathogenicity (Saenz *et al.*, 2000; Dallot *et al.*, 2001). At the gene-for-gene level, the P3 (+6K1) has been identified as a pathogenicity determinant for *Pea seed-borne mosaic virus* (PSbMV) in pea plants containing the '*sbm-2*' resistance gene (Johansen *et al.*, 2001).

4.4 Cylindrical inclusion protein (CI)

The cylindrical inclusion (CI) protein is one of the functional potyvirus proteins which accumulate in the cytoplasm of infected cells as cylindrical inclusions. The most typical inclusions formed by the CI protein are pinwheel, scroll or laminate-shaped. The high levels of the presence of characteristic cylindrical inclusions in the cytoplasm of the infected cells are a unique feature of potyvirus infections. These inclusions have been used diagnostically to identify different potyviruses. The CI protein is an RNA helicase (Lain *et al*., 1990), which contains a nucleotide binding motif (NTBM) with consensus amino sequences GXXXXGKS is conserved among the CI proteins of potyviruses and is responsible for helicase activity (Walker *et al.*, 1982; Lain *et al.*, 1989).

4.5 6-kilodalton peptide (6K)

The 6K peptide, positioned between the CI and NIa protein, was first determined by appearance of a NIa cleavage situation which was distinguished *in vitro* (Carrington *et al*., 1988). While the presence of mature 6K peptide found in infected plants has been described by Restrepo-Hartwig and Carrigton (1994). The 6K peptide possesses no enzymatic function. However, the function of this protein is still unclear. Nevertheless, its intracellular localization has been studied in some

activities. The 6K peptide is believed to be involved in transport of NIa. When bound to NIa, the TEV 6K peptides prevent translocation of NIa to the nucleus in transgenic plants expressing 6K/NIa, and therefore appears to override nuclear translocation of NIa (Restrepo-Hastwig and Carrington, 1992).

4.6 Small nuclear inclusion protein (NIa)

The small nuclear inclusion protein a (NIa) is positioned between the 6K and NIb proteins. The NIa protein had been first determined as a protein forming an inclusion body in the nuclei of the plant cells infected by TEV (Carrington and Dougherty, 1987). The NIa consists of two functional domains, the N-terminal VPg (viral protein genome-linked) domain and the C-terminal proteinase domain (Dougherty and Parks, 1991). In this review, these two domains will be referred to as VPg and NIa-Pro.

4.6.1 Genome-linked protein (VPg)

The 5'-end of the genomic RNA or DNA of many plants and animal viruses has a genome-linked protein (VPg) covalently bound to the 5'-terminal nitrogenous base by a covalent phosphodiester bond (Wimmer, 1982). The varions of potyviruses contain a 6- to 24-kDa protein (VPg) processed from the N-terminal portion of the nuclear inclusion protein a (NIa) (Hari, 1981; Siaw *et al*., 1985; Riechmann *et al.*, 1989). The potyviral VPg is linked to the 5'-NCR via a tyrosine residue (Murphy *et al.*, 1991; Riechmann *et al*., 1992; Anindya *et al*., 2005) that is conserved in the VPg domains of potyviral polyproteins (Murphy *et al.*, 1990, 1996). Data from several research groups, and the finding that the VPg of poliovirus interacts with the viral 3'-end sequences to prime the negative-strand RNA synthesis (Paul *et al.*, 1998), suggested that the VPg of potyviruses is a putative primer for potyvirus replication as proposed for picornaviruses (Toyoda *et al.*, 1987; Giachetti and Semler, 1991).

4.6.2 Nuclear inclusion a proteinase (NIa-Pro)

The NIa-Pro has been contemplated to be structurally and functionally related to the chymotrypsin-like serine proteinases on the basis of the sequence similarity (Ghabrial *et al*., 1990). The NIa-Pro, a third viral encoded proteinase, is known to be the major viral proteinase which is involved in the post-translational cleavage of the potyvirus polyprotein (Carrington and Dougherty, 1987; Garcia *et al*., 1989abc, 1990).

4.7 Large nuclear inclusion protein (NIb)

The large nuclear inclusion protein (NIb) was shown to be an RNA-dependent RNA polymerase (RdRp) of potyviruses (Hong and Hunt, 1996). Generally, the NIb protein forms inclusions in the nucleus of infected plants although it is necessitated in the cytoplasm or in membranes associated with replication complexes during viral RNA synthesis. However, it is probable that the structural integrating of NIb is implicated in controlling the transport of the protein to nucleus. The function of NIb in the nucleus is unknown.

4.8 Coat protein (CP)

The coat protein (CP) of potyvirus is a multifunctional protein. It can be roughly divided in three domains, the highly variable (in length and sequence) N- and C- terminal domains which are characteristic of a given potyvirus that these terminal regions are expressed on the surface on the virus particle and are sensitive to mild trypsin-proteolytic cleavage, and the trypsin resistant central or domain that is well conserved among the virus of this family (highly conserved within species). Different domains of potyvirus CP have been associated with distinct functions in several stages of the viral life cycle. The primary function of the CP is encapsidation of viral genomic RNA (Ng and Perry, 2004). The CP also plays an essential role in aphid mediated transmission, cell-to-cell movement through the plasmodesmata, long distance (systemic) movement through the vascular tissues and finally in regulation of

viral RNA amplification (Blanc *et al*., 1998; Revers *et al.*, 1999; Fedorkin *et al*. 2000; Varrelmann and Maiss, 2000; Ng and Perry, 2004; Dombrovsky *et al*., 2005).

5. Infectious transcripts and cDNA clones of RNA viruses

On the basis of genome nucleic acid type, more than 80 % of all known plant viruses contain RNA genomes (Hull, 2002). Study of the molecular biology of viruses with positive strand RNA genomes has been hampered by the inherent limitations imposed by the nature of their genetic material that these viruses do not encompass a DNA intermediate step in their replication cycle.

The advent of recombinant DNA technology in the 1970s prompted RNA virologists to convert viral RNA genomes into complementary DNA copies and replicate them as plasmid inserts in bacterial hosts for easier genetic manipulations. The genetic modification of RNA virus genomes depends on the availability of full-length cDNA clones from which functional RNA transcripts can be generated. In the case of positive-strand RNA viruses, the naked (recombinant) genome is infectious which functions as the messenger (m) RNA for viral replication, translation and an infection can simply be initiated by inoculation of a susceptible host. The advent of reverse genetics revolutionized the study of positive-stranded RNA virus that was amenable for cloning as cDNA into plasmids of bacteria.

The production of cDNA clones corresponding to the genomes of a number of positive strand RNA viruses, from which infectious RNA can be transcribed *in vitro* or *in vivo*, is an essential step in the development of reverse genetic systems for these viruses. Infectious RNA is obtained from full-length cDNA clones either by *in vitro* transcription of RNA usually from a phage promoter or by *in vivo* transcription from a strong constitutive promoter active in the host cell. The latter approach is attractive because it bypasses the need for *in vitro* transcription and incorporation of cap analogue. This approach has proven useful for a number of positive strand RNA viruses, including bacteriophage, animal viruses and plant viruses.

The availability of these clones has greatly enhanced the potential of investigations. It has significantly facilitated the study of the genetic expression and replication of RNA viruses by the use of mutagenesis, deletions, insertions and complementation experiments. It has also remarkably enhanced the understanding of the molecular mechanism in various aspects of natural and induced RNA recombination and of plant-virus interactions such as symptom development, cell-to-cell and long distance movement, vector-mediated transmission, interaction with resistance genes, replication, proteolysis and encapsidation.

6. Construction of infectious full-length cDNA clones of RNA viruses

Several variations on the general scheme for the production of full-length cDNA clones have been reported. Typical strategy for constructing infectious clones (as *in vitro*-transcribed RNA or cDNA copies) corresponding to the genomes of RNA viruses consists of complementary DNA synthesis, cloning scheme and the design of sequences bordering the viral insert. These fundamental steps can have a strong effect on the infectivity of the resulting transcripts. In general, the presence of the complete viral sequence is required to obtain infectious clones. Because in most cases the full-length nucleotide sequence of the viral RNA accounted has been determined, technical methods based on the use of viral-specific primers for DNA synthesis have been accommodated compared to those based on vector-primed cDNA synthesis. Petty *et al*. (1988) has developed a select combination of the two strategies for one-step cloning of full-length cDNA of *Barley stripe mosaic virus* (BSMV).

In general, the construction of a cDNA consists of reverse transcribing the viral RNA into a complementary-stranded DNA using a DNA primer (reverse) hybridizing specifically to the 3'end of the viral genomic RNA. After removing of the viral RNA template, the single-stranded DNA is converted into the double-stranded conformation by initiating DNA synthesis with a second primer (forward) including the sequence corresponding to the nucleotides at the 5'end of the viral RNA. Typically, the sequence for an RNA polymerase promoter fused to the viral sequence is included in the second primer.

Presently, a full-length viral cDNA in a high-copy-number bacterial plasmid following a promoter for a commercially purchasable DNA dependent RNA polymerase is a prevailing paradigm. Most viruses with comparatively small RNA genomes have been compliant for such a genetic fashion. However, some viral genomes appears to be synthesis *in vitro* of the full-length first complementary DNA strand, presumably because the polymerization step is impeded by sturdy secondary structures on the viral template RNA. Some research groups but not others have proven out in obtaining a full-length first complementary DNA strand of *Tobacco mosaic virus* (TMV) (Holt and Beachy, 1991) and *Turnip yellow mosaic virus* (TuMV) (Skotnicki *et al*., 1992; Boyer *et al*., 1993). These different eventualities could result from the application of different strains of the viruses or could reflect the importance of the sources of reagents used. The various productions of full-length cDNA clones have been reported. These have several modifications on the general strategy and as a consequence, it is difficult to choose an optimum scheme for any virus. These alterations are for representative the application of an improved vector-primed design (Petty *et al*., 1988), of "glass milk" for the isolation of a more than 9-kb-long first complementary DNA strand (Domier *et al*., 1989) and of the polymerase chain reaction (PCR) for full-length (Hayes and Buck, 1990) or partial (Young *et al*., 1991; Demler *et al*., 1993; Viry *et al*., 1993) viral sequence amplification.

Another problem is that some viruses are inherently difficult to cloning in plasmids due to the high instability of full-length cDNA clone of the viruses in bacterial host cells, because the viral sequences are toxic to the bacterium (Maiss *et al*., 1992; Jakab *et al*., 1997; Lai, 2000). Although the mechanisms fundamental such toxicity are not well understood, it is usually postulated that they are associated to the expression of some viral proteins in the bacterial hosts from RNA transcribed from cryptic promoters (Rodriguez-Cerezo and Shaw, 1991; Fakhfakh *et al*., 1996).

A variety of techniques for cloning of some refractory viral genomes has been reported and could overcome the cloning difficulty because the instability. First, *in vitro* ligation of individually cloned cDNA fragments followed by *in vitro* transcription and inoculation was used to surmount the cloning difficulties of *Beet*

necrotic yellow vein virus (BNYVV) (Quillet *et al*., 1989), *Japanese encephalitis virus* (JEV) (Sumiyoshi *et al*., 1992), *dengue virus* type-2 (DENV-2) (Kapoor *et al*., 1995), *Transmissible gastroenteritis virus* (TGEV) (Yount *et al*., 2000), *Mouse hepatitis virus* (MHV) (Yount *et al*., 2002). Nevertheless the disadvantages of this approach are relatively low levels of full-length cDNA and decreased levels of RNA transcripts to that of *in vitro* transcripts from a complete clone in one plasmid. Second, intron-insertion was used for stabilization of the clone in bacteria (Vancanneyt *et al*., 1990). This approach is to reduce the toxicity of viral sequences in high-copy-number plasmids by inserting introns into the viral sequences. cDNA clones of *Pea seed-borne mosaic virus* (PSbMV) (Johansen, 1996), *Plum pox virus* (PPV) (Lopez-Moya and Garcia, 2000), *Japanese encephalitis virus* (JEV) (Yamshchikov *et al*., 2001) and *Transmissible gastroenteritis virus* (TGEV) (Gonzalez *et al*., 2002) were stabilized by this method. However this strategy requires cytoplasmic viruses to have a nuclear phase and also requires that the viral sequences not have endogenous intron sequences, which has been proposed as a possible problem of large cytoplasmic RNA viruses (Lai, 2000). Johansen (1996) reported that introns interrupting the viral protein, functioned to reduce toxicity problems of potyviral full-length clones, since splicing of introns inserted in *Pea seed-borne mosaic virus* (PSbMV) clones, after their inoculation into host plants, allowed the production of accurate genome-size RNA which resulted in infection. Lopez-Moya and Garcia (2000) attempted to improve a full-length cDNA clone of *Plum pox virus* (PPV) to facilitate its growth in bacteria by intron-insertion strategy and found that intron-insertion in the P3 coding region was adequate to stabilize the PPV clone, it can be concluded that sequences downstream of this insertion-point were responsible for the putative cytotoxicity in bacterial hosts. In addition to faster growth of bacteria and higher plasmid yield, followed the modification which interrupted at the P3 region of the viral open reading frame (ORF). The intron-containing clone was infectious when inoculated into plants after sustaining *in vivo* transcription and splicing. In case of construction of a full-length cDNA clone of *Lettuce mosaic virus* (LMV), this does not seem to be the cases for the PSbMV and PPV clones. Yang *et al*. (1998) suggested that the insertion of introns does not always have the same beneficial effects as in the cases of PSbMV and PPV, considering no

differences were found in the performance of the bacteria between intron-containing and intron-less LMV clones, except a delay in the concern of infection in plants inoculated with the intron-containing clone. Third, avoidance of application of high-copy-number plasmids is an approach could solve the problem and was performed to overcome the cloning difficulties. Almazan *et al*. (2000) used a bacterial artificial chromosome (BAC), a low-copy-number plasmid, as a DNA cloning vector to clone the genome of *Tomato golden mosaic virus* (TGEV), and the TGEV cDNA clone could be propagated in *Escherichia coli*. Other approaches avoid the use of bacteria. *Vacinia virus* was used as a vector to clone the coronaviruses, *Avian infectious bronchitis virus* (AIBV) (Casais *et al*., 2001) and *Human coronavirus* (Thiel *et al*., 2001). Finally, frameshift mutation was used for reducing the toxicity to bacterial host. Satyanarayana *et al.* (2003) reported the approach that is a composition between infectivity of the RNA transcripts and toxicity to *E. coli* caused by introducing frameshift mutations into slippery sequences near the viral toxicity sequences in the *Citrus tristeza virus* (CTV) cDNA clones and concluded that this approach can be used as an additional strategy to clone recalcitrant viral genomic sequences in high-copy-number bacterial plasmids for production of infectious full-length cDNA clones of RNA viruses since this approach is to minimize the toxicity of viral sequences to *E. coli*.

7. *In vitro* transcription

7.1 RNA polymerase

In performing with the design and construction of a full-length cDNA clone from which infectious RNAs are expected to be generated *in vitro* or *in vivo*, the determination of the RNA polymerase promoter is highly significant since it straightly influences the product of transcripts and their nucleotide sequence at extreme termini. Various types of promoters have been utilized, such as the *E. coli* Pm (a modified version of Pr) promoter from bacteriophage λ and the promoters of bacteriophages SP6, T3 or T7. Ahlquist and Janda (1984) designed and constructed the first transcription vector arranged to provide precise control of the transcription initiation

site. This transcription vector, pPM1, permitted direct insertion into a unique *Sma*I cloning site of practically any viral cDNA under the control of the bacteriophage λPm. Subsequent *in vitro* transcription with *E. coli* RNA polymerase multiplied transcripts carrying no extraviral nucleotides at their 5'end. The outstanding applicability of this fashion resides in that in contrast to consensus promoter sequences for the SP6 (Melton *et al*., 1984) or T7 (Dunn and studier, 1983) RNA polymerases, that of *E. coli* lies well upstream of the initiation site for transcription and does not extend beyond the transcription initiation site (Hawley and McClure, 1983), permitting therefore practically any λPm-viral cDNA interfusion to be constructed except influencing promoter activity. Nevertheless, compared to bacteriophage RNA polymerase-based fashion, the *E. coli* RNA polymerase-based fashion confers much lower products of transcript, partly at least because it conducts to an extensive proportion of premature termination products (Angenent *et al*., 1989; Heaton *et al*., 1989). The majority of bacteriophage promoter-viral cDNA interfusions implicates the application of the T7 promoter, eventually the SP6 promoter and exceptionally T3 promoter. Selection for the T7 promoter over the SP6 promoter may be related to the more perfectly studies genetics of bacteriophage T7.

Ordinarily, the interfusions of transcription promoters to the beginning of viral sequences are able to perform in three different strategies.

1) Universal transcription vectors have been constructed in which a restriction site is introduced as close as possible to the transcription initiation site of the promoter. All these constructed vectors including pPM1/λPm (Ahlquist and Janda, 1984), pHST70/SP6 (Jobling *et al*., 1988) and pCa35J/35S (Yamaya *et al*., 1988) could hypothetically direct transcription of viral cDNA sequences to RNAs devoid of vector-derived nucleotides.
2) Foreign sequences between the promoter and viral cDNA have been removed by site-directed mutagenesis.
3) An oligonucleotide containing a promoter has been use as a forward primer in synthesis of second strand cDNA, directly combined to the viral sequence at 5'end (Weiland and Dreher, 1989). Simplicity and

convenience of this procedure are the important features, and it is now widely used.

For 'run-off' transcription, the nucleotide sequence of the 3'end of *in vitro*-transcribed full-length viral RNAs is frequently imposed by the situation of the restriction site. An alternative fashion to eliminate a large number of nonviral nucleotides at this extremity was described by Dzianott and Bujarski (1989). This method was performed by ligating a self-processing sequence of the satellite RNA of *Tobacco ringspot virus* (TRSV) to the 3'end of BMV sequences. The close circular cDNA clones could be transcribed *in vitro* by way of a rolling circle-like mechanism and the resulting multimeric transcripts were processed partially into infectious RNAs of genomic length. In addition to plasmid DNA, PCR products can be used as template for transcription reactions by one of the primers needs to include the promoter sequence of a bacteriophage polymerase so that the PCR product contains a bacteriophage promoter.

7.2 Effect of nonviral nucleotide

Several research groups have explored the consequence of nonviral nucleotides at the extremities of viral transcripts. As a general rule, 5'extensions considerably decrease or exactly abolish infectivity, while 3'extensions are more effortlessly tolerated. Infectivity is notably decreased even for short 5'extensions such as 1 or 2 nucleotides. In some cases of transcripts derived from plant viruses, infectivity is invalidated when these transcripts harbor 5'additional sequences of 14 to 33 nucleotides (Dore *et al*., 1990; Rizzo and Palukaitis, 1990; Chiang and Yeh, 1997).

In the genus *Potyvirus*, the constructs of infectious *in vitro* RNA transcripts derived from cloned cDNA of *Tobacco vein mottling virus* (TVMV) (Domier *et al*., 1989), *Plum pox virus* (PPV) (Riechmann *et al*., 1990), *Zucchini yellow mosaic virus* (ZYMV) (Gal-On *et al*., 1991), *Peanut stripe virus* (PStV) (Flasinski *et al*., 1996), which were contained one additionally nonviral nucleotide at the 5'extreamities of the

potyviral transcripts, were different in percentage of infectivity with 5%, 49%, 11% and 100%, respectively.

The affect of 5'nonviral nucleotides usually distinguished on the biological activity of synthetic viral transcripts. It is difficult to deduce a character for the nucleotides in the nonviral sequences because the number of different extensions demonstrated for various viral transcripts is limited in each case. The nature of the nonviral sequence is also important, as suggested by the difference of infectivity observed between transcripts harboring nonviral 5'extensions analogous in length but varying in their sequence (Duechler *et al*., 1989). All the same, several research groups usually postulated that the presence of these nonviral nucleotides could actively obstruct proper initiation of positive strand RNA synthesis from the 3'end of the negative strand. It gives an appearance of difference that these nucleotides interfere with viral gene translation *in vivo* because they have no initiation codons, and the *in vitro* translation products are analogous to those of viral wild-type RNA (Angenent *et al*., 1989; Dore *et al*., 1990; Gal-On *et al*., 1991; Viry *et al*., 1993).

In contrast, the infectivity of the viral RNA transcripts shows to be comparatively tolerate to short 3'extensions of 1 to 7 nucleotides (Ahlquist *et al*., 1987; Dzianott and Bujarski, 1989; Hayes and Buck, 1990), while long extensions such as 2700 nucleotides for *Brome mosaic virus* (BMV) (Ahlquist *et al*., 1984), 945 nucleotides for TMV (Dawson *et al*., 1986) or more than 82 nucleotides for BMV (Dzianott and Bujarski, 1989) invalidate infectivity. For several viruses, synthetic virus RNA transcripts carrying a long (more than 30 nucleotides) additional sequence at the 3'end are infectious (Boyer *et al*., 1993; Sit and Abouhaidar, 1993), as are *in vivo* transcripts presumably polyadenylated by host cell enzymes. With plant viral genomes presenting a poly (A) tail at their 3'end, it appears conceivable that there is a threshold length of adenine (A) residues below which the stability of the corresponding transcripts would be refashioned (Domier *et al*., 1989; Holy and Abouhaidar, 1993; Viry *et al*. 1993). The appearance of lengthy homopolymeric A sequences at the 3'end of poliovirus *in vitro*-generated transcripts heightens infectivity, while lengthy homopolymeric nucleotide sequences have an opposing

effect (Sarnow, 1989). The poly (A) tail of *Cowpea mosaic virus* (CPMV) transcripts was extended on replication in protoplast, this result was hypothesized the possible involvement of a slipping RNA polymerase event or else the addition of adenine residues by a host terminal nucleotidyl transferase (Eggen *et al*., 1989).

7.3 Cap structure (m^7G(5')ppp(5')G)

For optimum infectivity, a 7-methyl guanosin 5'-cap structure (m^7G(5')ppp(5')G) is necessitated at the 5'end of the RNA transcripts, possibly since it improves their RNA transcripts by providing a better resistance to host cell nucleases (Furuichi *et al*., 1977; Shimotohno *et al*., 1977) and/or enhances translation initiation of the RNA transcripts (Shih *et al*., 1976; Contreras *et al*., 1982). Capped *in vitro* transcripts can be synthesized by substituting cap structure for a portion of the GTP in the transcription reactions. Cap structure is a relatively expensive reagent and its inclusion in a transcription reaction often decreases the yield of the reaction.

Several plant RNA viruses have a VPg (viral protein genome-linked) alternatively of a cap structure (m^7G(5')ppp(5')G) at the 5'extream terminus of their RNA genomes. Some research groups considered this significant by performing with the comparable *in vitro* transcripts evidenced in most cases the affirmative effect of capping on their infectivity (Riechmann *et al*., 1990; Young *et al*., 1991; Veidt *et al*., 1992), suggesting that some of the features of the cap structure could at least partially make up the absence of VPg. It is probable that experimental studies on the role of VPg in infectivity will notably interest from an effective technique for attaching this protein to synthetic RNA transcripts (Boyer and Haenni, 1994).

In most case of infectious *in vitro* transcripts of RNA viruses, however, the infectivity of the transcripts is appreciably lower than that of natural virion RNAs, probably as a result of non-viral residues at the termini of the transcripts. Difficulties may also be caused by exposure of the transcripts to degradative agents during inoculation. For this reason the transcripts are usually stabilized by capping their 5'ends, but this is an inefficient, costly and time-consuming process. The

constructions of directly infectious cDNA clones of genomes of several positive strand RNA viruses have been reported. These constructs make use of the cauliflower mosaic virus (CaMV) 35S promoter sequence linked to the 5'ends of the viral cDNAs to generate infectious transcripts in the plant. This technique potentially overcomes some of the problems encountered with the use of transcripts generated *in vitro* (Dessens and Lomonossoff, 1993).

8. *In vivo* transcription

8.1 Infectious cDNA clones

Infectious cDNAs have several advantages over infectious RNA transcripts. Firstly, DNA is more stable than RNA and therefore easier to use as inoculum. Infectivity is less dependent on RNA degradation because it presumably occurs only within cells where the RNAs are synthesized, and the replication process can overcome damaging effects resulting from degradation. Secondly, *in vitro* transcription is not necessary. This is specifically significant for RNA viruses for which the production of a good yield of highly infectious full-length transcripts can be problematic. Furthermore, expensive reagents such as the cap structure (m^7G(5')ppp(5')G) and RNA polymerase, and technical difficulties associated with *in vitro* transcription are not required. Finally, it renders the expression of viral genes largely independent of the viral replication process. This might be very suitable when considering the localization of proteins translated by mutant viral RNAs unable to replicate in host cells. Van Bokhoven *et al*. (1993) reported that these *in vivo*-generated viral transcripts would then act parallel messenger RNAs produced by a host RNA polymerase, still able to express native or mutant proteins without being replicated.

Recently, several plant viruses have been reported for expressing infections RNAs from *in vivo* transcription of viral cDNA-containing vectors. Infectivity of a number of viral 35S-cDNA clones by manual (mechanical) inoculation onto intact plants has been reported including, *Plum pox virus* (PPV) (Maiss *et al*., 1992),

Zucchini yellow mosaic virus (ZYMV) (Gal-On *et al*.,1995), *Papaya ringspot virus* type P (PRSV-P) (Chiang and Yeh, 1997), *Turnip mosaic virus* (TuMV) (Sanchez *et al*., 1998), and *Lettuce mosaic virus* (LMV) (Yang *et al*., 1998). Instead, infection was performed by means of expression of infectious transcripts through transgenic plants (Yamaya *et al*., 1988) or agroinfection (Leiser *et al*., 1992) such as infection of plants with *Agrobacterium tumefaciens* expressing a viral RNA. Particle bombardment is a well known technique to introduce nucleic acids into plants (Sanford, 1988; Christou, 1992; Klein *et al*., 1992). This technique has also been used for infection of plants by cloned DNA viruses (Garzon-Tiznado *et al*., 1993; Hagen *et al*., 1994) and RNA viruses (Gal-On *et al*., 1995; Fakhfakh *et al*., 1996). The particle bombardment is used as an infection manner whereby microprojectiles are directed onto intact plants in a vacuum chamber, driven by condensed helium. This method is very efficient, but yet is bound by the physical limitations of the technology. A set of materials for plant transformation by particle bombardment ordinarily requires consequential equipment, generally with gas cylinders, vacuum chamber and vacuum pump, under sterile conditions. The equipment is expensive, complex to assemble and cumbersome. Gal-On *et al*. (1997) designed and constructed a simple hand-held device called the HandGun for an infection of plants with viral-encoding constructs by particle bombardment. By this method, a vacuum chamber and helium propulsive gas are not required to achieve an infection. With this technique it is possible to bombard soft plants and seedlings that do not survive particle bombardment by other devices.

8.2 CaMV 35S promoter

The well-studied 35S CaMV promoter is a sequence of about 350 base pairs upstream of the 35S transcript (-343 to +8, with Cap site at +1), about 250 base pairs of which overlap with the 3'end of gene VI, the last of the six large open reading frames (Odell *et al*., 1985). On integrated constructs of the promoter, Fang *et al*. (1989) studied and found that this promoter has three functional regions. The regions between -343 to -208 and -208 to -90 are responsible for about 50% and 40% of the promoter activity, respectively. The region between -90 to -46 region has

inconsiderable activity on its own, but contains determinants for activation of the TATA box and enhances the transcriptional activity of the two upstream regions and also control tissue specificity (Benfey and Chua, 1990).

9. Infectious transcripts and cDNA clones of potyviruses

In plant viruses, infectious *in vitro* transcripts were first successfully made up for *Brome mosaic virus* (BMV) (Ahlquist and Janda,1984), and the construction of an infectious cDNA clone with an ability of direct infection (*in vivo* transcripts) was first reported for RNA3 of *Alfalfa mosaic virus* (AlMV) via coinoculation with RNAs 1, 2 and 4 (Dore and Pinck, 1988). Cloned cDNA material can be used to obtain either *in vitro* or *in vivo* infectious transcripts. Both strategies have been applied successfully to members of *Potyvirus* genus belonging to the largest family of plant viruses.

In the genus *Potyvirus*, infectious transcripts have been synthesized *in vitro* from full-length cDNA clones with bacterial phage promoters recognized by *E. coli* or DNA-dependent RNA polymerases for *Tobacco vein mottling virus* (TVMV) (Domier *et al*., 1989; Nicolas *et al*., 1996), *Plum pox virus* (PPV) (Riechmann *et al*.,1990), *Barley yellow dwarf virus* (BYDV) (Young *et al*., 1991), *Zucchini yellow mosaic virus* (ZYMV) (Gal-On *et al*.,1991; Lin *et al*., 2002), *Tobacco etch virus* (TEV) (Dolja *et al*., 1992), *Pea seed-borne mosaic virus* (PSbMV) (Johansen *et al*., 1996), *Peanut stripe virus* (PStV) (Flasinski *et al*., 1996), *Potato virus A* (PVA) (Puurand *et al*., 1996), *Potato virus Y* (PVY) (Jakab *et al*., 1997), *Papaya ringspot virus* type P (PRSV-P) (Chiang and Yeh, 1997), *Turnip mosaic virus* (TuMV) (Sanchez *et al*., 1998) and *Johnsongrass mosaic virus* (JGMV) (Kim *et al*., 2003). When these RNAs transcribed *in vitro* are inoculated into host cells, they led to viral RNA replication.

In vivo infectious transcripts, which are driven by a *Cauliflower mosaic virus* (CaMV) 35S promoter that can be recognized by host RNA polymerase and terminated by NOS terminator, have also been reported for *Plum pox virus* (PPV) (Maiss *et al*., 1992; Lopez-Moya and Garcia, 2000), *Zucchini yellow mosaic virus* (ZYMV) (Gal-On *et al*.,1995), *Potato virus Y* (PVY) (Fakhfakh *et al*., 1996; Jakab *et*

al., 1997), *Pea seed-borne mosaic virus* (PSbMV) (Johansen, 1996), *Papaya ringspot virus* type P (PRSV-P) (Chiang and Yeh, 1997), *Clover yellow vein virus* (ClYVV) (Takahashi *et al*., 1997), *Turnip mosaic virus* (TuMV) (Sanchez *et al*., 1998), and *Lettuce mosaic virus* (LMV) (Yang *et al*., 1998). In these cases, the plasmid can be directly introduced into host plants to induce infection.

MATERIALS AND METHODS

1. Virus and plant sources

The virus used in this study was the watermelon isolate of PRSV (PRSV-W), the type strain of the virus, which was originally isolated from a naturally infected cucumber (*Cucumis sativus* L.) in Nakhon Sawan province, central region of Thailand. The isolated pathotype W of virus-infected cucumber plants were kindly provided by Dr.Wanpen Srithongchai, Virology section, Plant Pathology and Microbiology Division, Department of Agriculture, Bangkok, Thailand.

1.1 Characteristic confirmation of PRSV-W from infected plant source

In order to confirm the pathotype of the virus obtained from the source plants, infectivity assay of the virus was performed by mechanical inoculation of greenhouse-grown pumpkin and papaya test plants at cotyledon stage and at 2-3 true-leaf stage, respectively. Leaves of the source plants were ground in 0.1 M sodium-potassium phosphate buffer, pH 7.2, and the obtained crude extract was rubbed onto the Celite-dusted leaves of the test plants. Ten min after inoculation, the inoculated leaves were rinsed with tap water and all inoculated plant seedlings were kept in an insect-free and temperature-controlled (24-28 ^{o}C) greenhouse. Over a period of 8 weeks after inoculation, symptom development of inoculated test plants was observed. Non-inoculated plants of each species were included as controls. Around 2-3 weeks after inoculation, systemic mosaic symptom showed in pumpkin but did not in papaya test plants over a period of 2 months. This indicated that the virus was able to infect to pumpkin but was not in papaya test plants.

Further verification of the virus, transmission electron microscopy (TEM) was performed with negative staining of crude sap prepared from the inoculated and mosaic symptom-expressing pumpkin plant tissues to detect a typical flexuous, filamentous particle of 780 x 12 nm. The presence of *Potyvirus* particles and the absence of any other virus particles were visualized.

Following preliminary identification of the pathotype of the virus obtained in the source plants using the strategy described above, a reverse transcription-polymerase chain reaction (RT-PCR) method was performed on total RNA extracted from the inoculated test plants. All plants with or without symptoms used as source of total RNA samples. Total RNA was extracted and were used in RT-PCR using degenerate or specific oligonucleotide primers, 2SC5 5'-ATG TTT TYG ARC AAA G-3' (viral sense) and 2SC4 5'-GCA ATA CAG AGA TCA TCA CCA-3' (viral anti-sense), which were kindly provided by Dr. Srimek Chowpongpang and these primer sequences are complementary and annealable to regions in the nuclear inclusion body (NIb) gene of PRSV. The approximately expected 1-kb fragment was amplified from the total RNA preparations which obtained from all the inoculated-pumpkin test plants, but was not in all the papaya test plants. The one-kb PCR product generated was purified following electrophoresis on a 1% agarose gel using the QIAquick® Gel Extraction Kit (QIAGEN) and cloned into a T-tailed pGEM®-T Easy vector (Promega), prior to transformation into *Escherichia coli* strain DH5α . Screening for insert-containing clones was performed using PCR with the universal primers M13 forward (-20) 5'-GTA AAA CGA CGG CCA GT-3' and M13 reverse 5'-AAC AGC TAT GAC CAT G-3' (section 7.2.1) and by restriction enzyme digestion with *Eco*RI (section 7.2.2). The cloned DNA fragment obtained from the RT-PCR was sequenced by performing as described in section 9. Sequence assembly of nucleotide sequence obtained was carried out using the Lasergene Biocomputing software (DNASTAR, Inc). Nucleotide and amino acid sequences were analyzed and compared with NCBI databases using the basic local alignment search tool (BLAST) programs.

Sequence analysis of the insert fragment of 1075 nucleotides of the viral RNA showed that this genome fragment contains the coding sequence for 16 nucleotides of 3'-end of NIa gene and followed by approximately 66% of 5'-end of NIb gene (1059 nucleotides; 353 amino acids) of PRSV type W. The coding sequence not only shared similarity with NIa and Nib genes of PRSV type W but also contained a characteristic NIa protease cleavage site (VFEQ/S). BLAST search analysis of the viral sequences gave highest scores with PRSV type W (GenBank accession number AY010722.1).

These results indicated that the virus obtained from the original source plants was PRSV type W because of its infectivity on pumpkin but not on papaya, its morphology of the virus particles with a typical flexuous, filamentous particle of 780 x 12 nm observed under TEM and the similarity of the 1075-nucleotides sequence sharing with NIa and NIb genes of PRSV type W.

1.2 Mechanical inoculation and Virus source maintenance

For mechanical inoculation of plants, pumpkin (*Cucurbita moschata*) seedlings were used as source plants to maintain viral cultures for inoculation of PRSV-W (Thai isolate). Inoculum was prepared from infected leaves by grinding young leaves of source plants in an ice-cold mortar in 0.1 M sodium-potassium phosphate buffer (pH 7.2). Inoculations of pumpkin plants were conducted mechanically by cotton tip rubbing Cellite-dusted cotyledons (5-7 days post germination) and then the cotyledons were briefly washed with tap water. All inoculated plant seedlings were kept in a greenhouse (temperature ranging between 26 ^{o}C and 30 ^{o}C). The typical mosaic symptom was observed at 10-14 days after inoculation.

The infected cucumber leaves with mosaic symptoms of the source plants served as the original inoculum source of PRSV pathotype W after confirmation of the virus. PRSV-W was maintained and propagated in mechanically inoculated pumpkin (*Cucurbita moschata*) under greenhouse condition and used for preparation of total RNA and molecular cloning.

2. Isolation of total RNA from healthy and PRSV-W-infected pumpkin leaf tissues

Each healthy and PRSV-W-infected pumpkin leaf tissues was collected, frozen in liquid nitrogen and stored at -80 ^{o}C until required. Approximately 1-3 g of frozen leaf tissues were ground with a mortar and a pestle in the presence of liquid nitrogen to a fine powder, transferred to sterile 1.5-ml microcentrifuge tubes each

containing 300 µl of RNA extraction buffer (1 M Tris-HCl, pH 9.0, and 1% SDS) and 300 µl phenol:chloroform:isoamyl alcohol (25:24:1(v/v)). The extract was vigorously mixed for 2 min by vortexing and centrifuged at 3,000 x g for 1 min at 4°C. The clear aqueous layer was transferred to a sterile tube and was phenol:chloroform:isoamyl alcohol extracted an additional one time. The aqueous layer was transferred to a new tube subsequently 3,000 x g centrifugation for 5 min at 4°C and then added 0.1 volume of sodium acetate (pH5.6) and 3 volume of absolute ethanol. The solution was mixed by invert the tube and stored at -80°C for 10 min. The pellet was retrieved by centrifugation at 3,000 x g for 5 min at 4°C. After the pellet was air-dried, resuspended in 100 µl diethyl pyrocarbonate (DEPC) -treated water and centrifuged at 3,000 x g for 10 min at 4°C and then the supernatant was transferred to a new tube. An equal volume of 4 M lithium chloride (LiCl) was added to the supernatant to a final concentration of 2 M LiCl. RNA was allowed to precipitate by keeping in deep freezer for 1 hour at -80°C or overnight at 4°C before it was pelleted at 3,000 x g for 20 min at 4°C. The pellet was rinsed by adding 1 ml of 2 M LiCl and then centrifuged at 3,000 x g for 2 min at 4°C. The obtained pellet was washed with 70% ethanol, air-dried and resuspended in 10 µl of DEPC-treated water.

3. Isolation of total RNA from infected plant tissues using RNeasy™ Plant Mini Kit

Total RNA from fresh infected pumpkin leaf tissues (10 mg) was extracted in a microcentrifuge tube using RNeasy™ Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. All centrifugation steps were carried out in a microcentrifuge at 13000 x g (maximum speed) at room temperature. The RNA was eluted from an RNeasy spin column using 50 µl of DEPC-treated water and used directly as template for cDNA synthesis or stored at -80 °C until used.

Briefly, ten mg of fresh leaf tissues were placed in a sterile 1.5 ml microcentrifuge tube containing 450 µl of buffer RLT plus 1% (v/v) β-mercaptoethanol (β-ME) and ground using polypropylene pestle designed and compatible for 1.5 ml microcentrifuge tube (USA Scientific, Inc.). The extract

obtained was vigorously vertexed and then transferred to a QIAshradder spin column placing in a 2 ml collection tube. The column was centrifuged at maximum speed for 2 min in a microcentrifuge. The flow-through fraction was carefully transferred to a fresh tube without disturbing the cell debris pellet in the collection tube and then 0.5 volumes absolute ethanol (approximately 225 µl) were added and mixed well by pipetting. The mixture (approximately 675 µl) was applied onto an RNeasy mini spin column placing in a 2 ml collection tube. The column was centrifuged at maximum speed for 15 sec in a microcentrifuge. For washing, 700 µl buffer RW1 were added onto the column and centrifuged for 15 sec. The column was added with 500 µl buffer RPE and centrifuged for 2 min to dry membrane. The column with the dried membrane was transferred and placed into a sterile 1.5 ml microcentrifuge tube. To elute RNA from the column membrane, 50 µl of DEPC-treated water were loaded into the center of a membrane within the column. The column placing in a microcentrifuge tube was centrifuged at maximum speed for 1 min. The total RNA (containing plant RNAs and genomic PRSV-W RNAs) isolated from PRSV-W-infected pumpkin leaf tissues was concentration estimated following spectrophotometric estimation at 260 and 280 nm. RNA was quantified assuming that 1.0 A260 equals 40 µg/ml.

The resulting RNA preparation (1:10 (v/v) e.g., 2.5 µl of this RNA extract to each 25 µl of reaction volume) was used immediately in reverse transcription and polymerase chain reaction (RT-PCR) as RNA template by incubation at 70 ^{o}C or 10 min and snap-cooled on wet ice for 5 min.

4. cDNA synthesis

For cloning of the PRSV-W genome, total RNA was extracted from infected pumpkin leaf tissues as described in section 3. Since the sequence of PRSV-W was unknown, a strategy to amplify the genome in eight overlapping fragments was performed. The cDNA fragments corresponding to the genome of PRSV-W were obtained either by reverse transcription of viral RNA with *Moloney murine leukemia*

virus (MMuLV) reverse transcriptase (Superscript II, Gibco) followed by polymerase chain reaction (coupling RT and PCR) or by using OneStep RT-PCR Kit (QIAGEN).

Several specific or degenerate oligonucleotide sequences were designed based on the alignment of two nucleotide sequences published in the GenBank covering the genome of different PRSV Hawaiian (HA) (GenBank accession number X67673) and Taiwanese YK (GenBank accession number X97251) isolates of the strain PRSV-P, consensus regions were identified. These designed oligonucleotide sequences were used as either forward or reverse primers which were defined in Table 2.

Polymerase chain reaction (PCR) was carried out using *Taq* DNA polymerase (Fermentas) and these primers were arranged to allow RT-PCR amplification of PRSV-W as eight overlapping fragments of approximately 1 to 2.2 kbp with a single adenosine (A) nucleotide overhang at the 3'- terminus at both ends of each cDNA fragment. The eight generated overlapping cDNA fragments corresponding to the genome of PRSV-W were schematic illustrated in Figure 1.

After determination of the RT-PCR products by electrophoresis in a 1% agarose gel, the desired DNA fragments resulted from the RT-PCR were excised from the gel and eluted with the QIAquick® Gel Extraction Kit from QIAGEN as described in section 12.3. The purified RT-PCR products were inserted into a pDrive DNA cloning vector for molecular cloning in bacteria *E. coli*.

4.1 Coupling reverse transcription and polymerase chain reaction (RT-PCR)

First strand cDNA was synthesized using Superscript II reverse transcriptase system (Gibco BRL Life Technologies). The RT reaction was performed using a MJ Research PTC-200 Peltier Thermal Cycler within 200-µl nuclease-free thin-wall PCR tubes. To synthesize first-strand cDNA, 2 µl of 10 µM of a reverse specific primer (20 pmol), 5 µL of the total RNA and 4 µl of sterile DEPC-treated water were heated at 65°C for 5 min, chilled on ice, and then added to a reaction mixture of 4 µl of 5X first-strand buffer (250 mM Tris-HCl (pH 8.3), 375

Table 2 Sequences of synthetic oligonucleotide primers used for cDNA cloning by RT-PCR

Fragment	Size (bp)	Primer	Range	Sequence (5'-3')
wA	963	PC1	1-22	AAATAAAACATCTCAACACAAC
		PC8	963-946	AACTCAAGGCTAGCACTC
wB	2,238	PC11	869-890	GAAGCAGTGCACCTGGCACTCG
		PC2	3106-3086	AAA(CT)TCTCC(AG)CCGACAATGTA
wC	1,973	PC3	2649-2669	CTCCTACCAAGAATCACAT(CT)G
		PC10	4621-4601	GGTGGGCTCTATTAACAACAC
wD	1,791	PC13	4484-4502	GCAGCCACTGTGGCTAGTG
		PC12	6274-6254	TGCATCCTTAATCATCAATGA
wE	1,239	PC15	6129-6146	TGCAGGAAAACAATGTTG
		PC4	7367-7347	CATTCGTGGATATCCA(AG)TGGC
wF	1,527	PC9	6633-6650	CTGACATCTCAATGGTGC
		2SC8*	8159-8141	CCTTTAACGAACCGTTCCA
wG	1,534	2SC9*	7744-7761	AGGTATTTGTCCTCTCTT
		SC2*	9277-9258	ATCCACAGCTTCAGTTTTGG
wH	1,178	2SC17*	9155-9178	CGAGAAAGGGGAGATTCACCTGAG
		3UTR2	10332-10303	$(T)_9$CTCTCATTCTAAGAGGATCGA

* Synthetic oligonucleotide primers were provided by Dr. Srimek Chowpongpang, Department of Plant Pathology, Kasetsart University, Bangkok, Thailand.

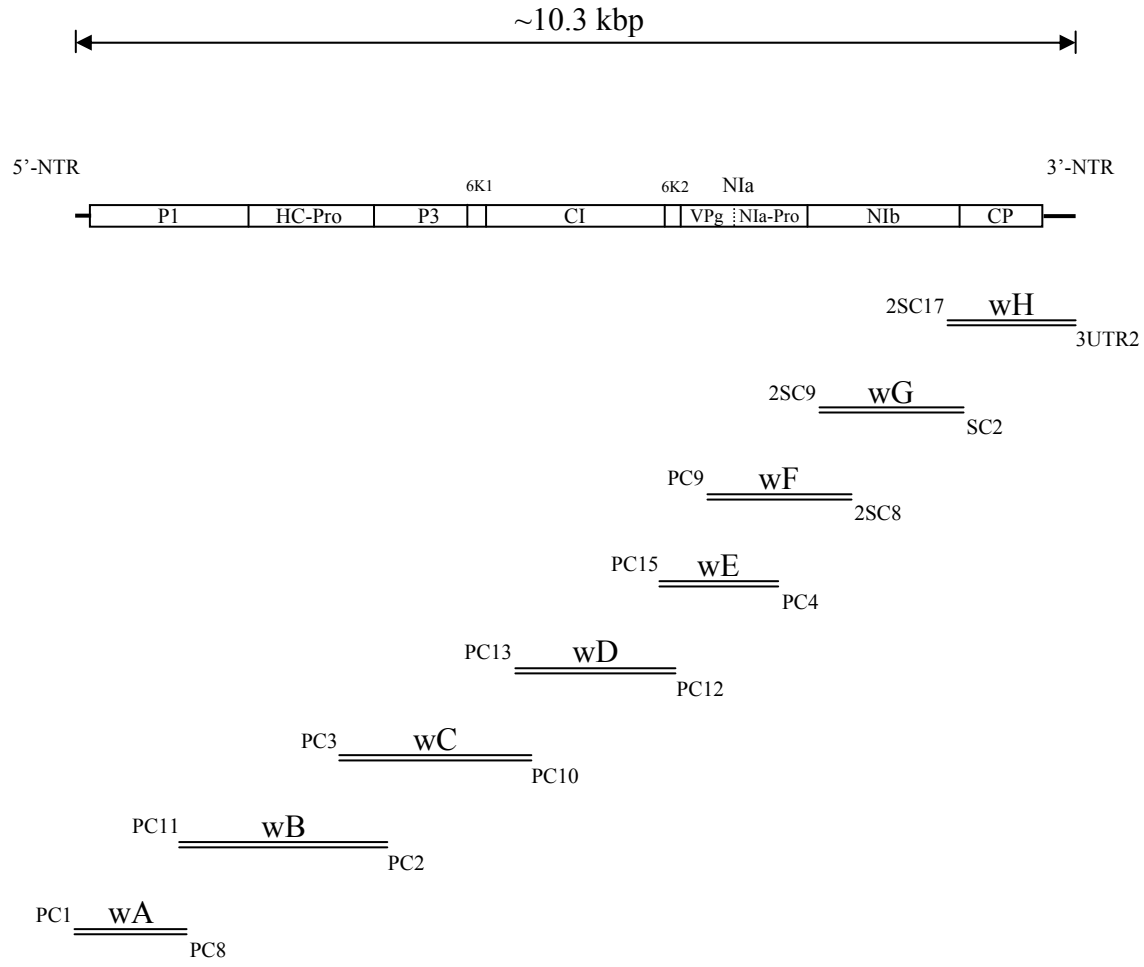


Figure 1 Genome organization of *Papaya ringspot virus* type W (PRSV-W) and the cloning strategy used. The terminal non-coding regions (NCR) are depicted by line, and the open reading frame is represented by open box with the different polyprotein domains (proteins). Cloning and sequencing of the PRSV-W genome was carried out in eight overlapping fragments depicted under the schematic presentation of the viral genome. The oligonucleotide primers used for amplification of the fragments are indicated.

mM KCl, 15 mM $MgCl_2$), 2 µl of 100 mM DTT, 1 µl of 10 mM of dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP) and 1 µl of RNaseOUT recombinant ribonuclease inhibitor (40 units/µl). The combined content was gentle mixed and incubated at 42°C for 2 min. Subsequently addition of 1 µL of SuperScript™ II (units/µl), the reaction mixture was mixed and incubated at 42°C for 50 min. After this period, the reaction was inactivated by heating at 70°C for 15 min. RNA complementary to the synthesized cDNA in the inactivated reaction mixture was removed by adding of 1 µl of (2 units/µl) of *E. coli* RNase H and incubated at 37°C for 20 min. The resulting first-strand cDNAs were used as templates immediately for DNA amplification by polymerase chain reaction or stored at -20°C until used.

For the second strand synthesis and DNA amplification by polymerase chain reaction, 2 µL of the resulting first-strand cDNAs were used as DNA templates in the reaction by adding to a 50 µl final volume of polymerase reaction mixture containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each dATP, dGTP, dCTP and dTTP, 200 nM each of the forward and reverse primers and 1.25 units *Taq* DNA polymerase (Fermentas). The following thermal cycling scheme was used for 35 reaction cycles (MJ Research PTC-200 Peltier Thermal Cycler), an initial template denaturation at 94°C for 2 min., followed by 35 cycles of template denaturation at 94°C for 30 sec, primer annealing at 43-56°C (depending upon the T_m of the primers) for 30 sec and primer extension at 72°C for 2 min. The final extension step was one cycle at 72°C for 10 min.

4.2 One-step RT-PCR

One-step RT-PCRs were performed in a final volume of 25 µl using a one-step RT-PCR reagents kit (QIAGEN), according to the manufacturer's instructions. Briefly, 2.5 µl clarified total RNA was extracted from infected pumpkin leaf tissue as described in section 3 was submitted to an one-step RT-PCR mixture of 5 µl 5x QIAGEN OneStep RT-PCR Buffer (containing Tris·HCl, KCl, $(NH_4)_2SO4$, 12.5 mM $MgCl_2$, DTT, pH 8.7), 1 µl dNTP Mix (contained 10 mM of each dATP, dGTP, dCTP and dTTP), 2 µl each of 10 µM the forward and reverse primers, 11.5 µl of DEPC-

treated water and 1 µl QIAGEN OneStep RT-PCR Enzyme Mix (consisting of an optimized combination of Omniscript reverse transcriptase, Sensiscript reverse transcriptase, and HotStarTaq DNA Polymerase) in a storage buffer (containing 20 mM Tris·HCl, 100 mM KCl, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5% (v/v) Nonidet® P-40, 0.5% (v/v) Tween® 20, 50% glycerol (v/v), pH 9.0). The mixture was mixed thoroughly, and dispensed appropriate volumes into 200-µL nuclease-free thin-wall PCR tubes. The RT-PCR was performed using a MJ Research PTC-200 Peltier Thermal Cycler. The temperature profile of a reverse transcription at 50 °C for 30 min, followed by an initial PCR activation step at 95°C for 15 min, and 35 cycles of amplification (94 °C for 30 sec, 50 °C for 30 sec and 72 °C for 2 min) and the final extension step was one cycle at 72 °C for 10 min.

5. Cloning of RT- PCR and PCR products

5.1 Purification of amplified insert DNA fragments

The resulting *Taq DNA polymerase*-amplified PCR products with a single A-tails were purified by electrophoresis in a 1% agarose gel, the amplified DNA fragments were excised and eluted with either the Qiaquick® Gel Extraction Kit (QIAGEN) or the Highpure® PCR Purification Kit (Roche) as recommended by the manufacturer and recovered in 30-50 µl of DEPC-treated water and then used as insert DNA fragments for cloning of eight overlapping cDNA fragments corresponding to the genome of PRSV-W.

5.2 *In vitro* DNA ligation of RT-PCR fragment into a U-tailed plasmid cloning vector

Taq DNA polymerase-amplified PCR products were cloned into U-tailed vector, pDrive cloning vector from QIAGEN (Appendix A Figure 1), by *in vitro* DNA ligation with T_4 DNA ligase (Fermentas) as described the following. Briefly, the purified PCR products with a single A-tails were combined together at 1-3:1 molar ratio with the pDrive cloning (U-tailed) vector (QIAGEN). The ligation mixture

contained 1X ligation buffer (40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT and 0.5 mM ATP (pH7.8)), 0.1 unit/ µl T_4 DNA ligase and deionized water up to the final volume of 20 µl. The ligation mixture was incubated at 22 ^oC for 1 hour or at 16 ^oC overnight. Following inactivation of the T_4 DNA ligase used in the reaction mixture was carried out by heating the reaction mixture at 65 ^oC for 10 min. After incubation for *in vitro* DNA ligation and heat inactivation of the ligase, the expected ligation products were precipitated by addition of 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of ice-cold absolute ethanol, mixed and then stored on ice for 15-30 min. The precipitate of DNA was allowed to form in the ethanolic solution, the expected DNA ligation products were recovered by centrifugation at maximum speed for 10 min at 4^oC in a microcentrifuge. The resulting pellet of DNA was washed by filling the microcentrifuge tube half way with 70% ethanol and following recentrifugation at maximum speed for 2 min at 4 ^oC in a microcentrifuge. The supernatant was carefully removed, and the remained traces of fluid were evaporated by placing the open microcentrifuge tube on the bench at the room temperature. The air-dried DNA pellet was dissolved in a desired volume of DEPC-treated water. The solution of expected DNA ligation products was used for transformation of competent *E. coli*.

Additionally, PCR products consequently generated from long PCR and overlap extension PCR were also cloned into either an U-tailed vector, pDrive cloning vector from QIAGEN or a T-tailed vector, pTZ57R from Fermentas after attaching the blunt-end PCR products with a single adenosine (A) nucleotide overhang at their 3'- terminus. Therefore, the PCR products with 3' A overhangs are compatible for ligation into the cloning vectors, which contains a single 3'- terminus uracil (U) or thymidine (T) nucleotide overhang at both ends.

All recombinant plasmid preparations were used to transform the recombinant plasmid containing the desired PCR DNA fragment into bacterial host cell of *E. coli*.

6 Transformation of competent *E. coli*

In this research, transformation manners of recombinant DNA plasmids were achieved by using either a calcium chloride (heat shock) or an electroporation method depending on the size of the insert DNA fragment and plasmid vector. The calcium chloride ($CaCl_2$) (heat shock) methods was chosen to perform with a small insert DNA fragment (smaller than 3 kb), whereas electroporation method was carried out with a 3-kb larger insert DNA fragment.

6.1 Bacterial strains used for preparation of transformation competent cells

All bacterial strains used in this research were provided by Plant Genetic Engineering Unit, KU/BIOTEC at Kasetsart University. Three different strains of *E. coli* were used to prepare the competent cells and utilized as a host cell for recombinant DNA.

E. coli strain DH5α (*supE*44 Δ*lac*U169 (ϕ80 *lacZ* ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1).

E. coli strain JM109 (*recA*1 *supE*44 *endA*1 *hsdR*17 *gyrA*96 *relA*1 *thi* Δ(*lac-proAB*), F' [*traD*36 $proAB^{+}$ $lacI^{q}$ *lacZ* ΔM15]).

E. coli strain XL1-Blue (*supE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*, F' [$proAB^{+}$ $lacI^{q}$ *lacZ* ΔM15 Tn*10* (tet^{r})])

These strains carried *lacZ* ΔM15, which permit α-complementation with the amino terminus of β–galactosidase encoded in pUC vectors (allow blue-white screening on X-gal).

6.2 Transformation of competent *E. coli* by electroporation

6.2.1 Preparation of electrocompetent cells

Electrocompetent *E. coli* cells for electroporation were prepared according to a modified procedure of Dower *et al.* (1988). Briefly, a single fresh colony of *E. coli* strain DH5α or JM109 or XL1-Blue was inoculated into a 50-ml flask containing 5 ml of 2xYT broth and grown with vigorous aeration (250 rpm) overnight at 37 ^{o}C. The overnight culture was diluted to 1:100 in 200 ml of pre-warmed 2xYT broth (each 2-liter flask) and grown with vigorous aeration at 37^{o}C until the culture reached an OD_{600} of 0.6-0.8. The 2 flasks of the culture to be harvested were chilled on ice for 30 min. The bacterial cells were pelleted by centrifugation in a cold rotor using Sigma refrigerated centrifuge at 1000 x g for 10 min at 4 ^{o}C. The supernatant was poured out and then pipetted to remove all residual fluid. The pellets were gently resuspended in 200 ml each of sterile ice-cold 10% glycerol. The obtained cell suspension was incubated on wet ice for 30 min and then centrifuged at 1000 x g for 10 min at 4 ^{o}C. The each cell pellet was gently resuspended in 100 ml of sterile ice-cold 10% glycerol. The cell suspension was incubated on ice for 30 min and then centrifuged as in the previous step. The supernatant was poured off as soon as the centrifuge stopped. The cell pellet was resuspended to a final volume of 1-2 ml in ice-cold sterile 10% glycerol. The obtained electrocompetent cells of *E. coli* were dispensed into microcentrifuge tubes chilling on ice. One hundred µl aliquots of competent cells were frozen in liquid nitrogen and used immediately in electroporation or stored at -80^{o}C until used (within 6 months) in a transformation.

6.2.2 Electroporation

One hundred µl aliquot of frozen competent cells were prepared from *E. coli* strain (such as DH5α or JM109 or XL1-Blue) as describe in section 6.2.1. For each transformation reaction by electroporation, one aliquot of the freshly prepared or frozen competent cells was thawed on ice and 5 µl of a ligation-reaction mixture was

added into the thawed component cells. The combination mixture was mixed and transferred to a pre-chilled electroporation cuvette (0.2 cm cuvette, Bio-Rad). A Gene Pulser™ electroporation system (Bio-Rad) was used as a machine for electroporation. The pulse condition was adjusted to 25 μF, capacitor 2.5 kV and the pulse controller unit to 200 Ω, with the time constant of 4 to 5 msec. The mixture of cells and DNA was tapped on the bench several time to eliminate bubbles and brought to the bottom of the cuvette. Cells were transformed by applying one pulse at the above condition settings. A 950 μl aliquot of room temperature SOC medium was added to the cuvette immediately then transferred to a culture tube (sterile 1.5 ml microcentrifuge tube). The tube was incubated in an incubator shaker at 37 °C for 1 hr and 150 rpm. Cells were collected by centrifugation at 4000 xg for 30 sec, and the cell pellet was resuspended in 100 μl of SOC medium. The cell suspension was plated onto 2xYT agar plate containing 100 μg/ml ampicillin, 80 μg/ml X-gal and 0.05 mM IPTG and the plate was incubated at 37°C for 12-14 hrs.

6.3 Transformation of competent *E. coli* using calcium chloride/heat shock

6.3.1 Preparation of competent cells

E. coli cells were prepared according to a modified procedure of Hanahan (1985). Competent cells of strains DH5α or JM109 were prepared and available for a chemical ($CaCl_2$)/heat shock plasmid transformation. Briefly, a single fresh colony of *E. coli* strain DH5α or JM109 was inoculated into a 50-ml flask containing 5 ml of 2xYT broth and grown with vigorous aeration (250 rpm) overnight at 37°C. Two 2-liter flasks containing 200 ml of pre-warmed 2xYT broth were set up and the overnight culture was diluted to 1:100 by adding 2 ml of the overnight culture into the 2-liter flasks containing 2xYT broth and grown with vigorous aeration at 37 °C until the culture reached an OD_{600} of 0.5. The bacterial cells were pelleted by centrifugation in a cold rotor using Sigma refrigerated centrifuge at 1000 x g for 10 min at 4°C after the flasks were chilled on ice for 30 min. The supernatant was poured out and the remained bacteria pellet in each centrifuge bottle was gently resuspended in 25 ml RF1 solution at pH 5.8 (containing 100 mM rubidium chloride

(RbCl), 50 mM manganese chloride ($MnCl_2$), 30 mM potassium acetate ($KC_2H_3O_2$), 10 mM calcium chloride ($CaCl_2$) and 15% (w/v) glycerol). The cell suspension was incubated on wet ice for 1hr and then centrifuged at 1000 x g for 10 min at 4°C. After pouring out the supernatant, each cell pellet was gently resuspended in 5 ml RF2 solution (containing 10 mM 3-(N-morpholino) propanesulfonic acid (MOPS), 10 mM RbCl, 75 mM $CaCl_2$ and 15% (w/v) glycerol). The cell suspension was incubated on wet ice for 15 min. One hundred µl of cell suspension were dispensed into each microcentrifuge tube took out of -80 °C deep freezer. All aliquots of competent cells were frozen in liquid nitrogen and stored at -80°C until used in a transformation.

6.3.2 Chemical ($CaCl_2$)/heat shock plasmid transformation

For each transformation reaction, 100 µl aliquot of fresh or frozen competent cells was prepared from *E. coli* strain (DH5α or JM109) as describe in section 6.3.1. One aliquot of competent cells was combined with 50-500 µg ligated plasmid by adding 5-10 µl of ligation-reaction mixture to competent cells in a microcentrifuge tube. The tube was gently flicked to mix, and placed on ice for 20 min. The bacteria cells were treated for 90 sec in a water bath at exactly 42°C (heat shock). The tube was immediately returned to ice for 10 min. An aliquot of room temperature SOC medium (950 µl) was added to the tube. The bacterial cells were incubated at 37°C for 60 min with shaking (150 rpm). Cells were collected by spinning at 4000xg for 30 sec, and the cell pellet was resuspended in 100 µl of SOC medium. The cell suspension was spread onto 2xYT agar plate containing 100 µg/ml ampicillin, 80 µg/ml X-gal and 0.05 mM IPTG. The plate was incubated at 37°C for 12-14 hrs.

7. Colony screening and analysis of transformants

Successful cloning into these cloning vectors interrupt the coding sequence of β-galactosidase, and recombinant clones can usually be identified by their color in the presence of 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside (X-gal). The site for insertion of foreign DNA fragments into the plasmid occurs right in the middle of this

gene. Insertion of a piece of DNA disrupts the β-galactosidase coding sequence, inactivating the gene. With no fractional β-galactosidase gene, the transformed *E. coli* cells are unable to cleave the dye-containing sugar. The end result is that colonies from transformed *E. coli* cells, containing plasmids that have a disrupted *lacZ* gene, will not turn blue, but show up as white spots on an agar plate with appropriated antibiotic. Therefore, a twofold selection procedure was carried out to determine which of the colonies had an insert of foreign DNA fragment: 1) selection for cells which contain the plasmid by testing for resistance to antibiotic (ampicillin or kanamycin), and 2) selection for plasmid-containing cells (antibiotic resistant) which contain a DNA fragment, by examining their ability to hydrolyze a dye-containing sugar. White colonies which are unable to hydrolyze the dye (white, disrupted *lacZ*) were analyzed by PCR and cultured for plasmid DNA isolation.

7.1 Colony screening

Commercially prepared cloning vectors used, pDrive (QIAGEN), pGEM-T Easy (Promega) and pTZ57R (Fermentas), which come with 3' U or T overhang, were used in this research. The plasmids are the high copy number plasmid that contains T_7 and SP6 RNA polymerase promoters (Appendix A Figure 1, 2 and 3, respectively). And the plasmids exploit the blue/white selection system to facilitate the screening of recombinant clones. The cloning sites frank a bacterial *lacZ* α-peptide expression cassette allowing the discrimination between blue parental clones and white recombinant clones. Insertional inactivation of the enzyme β-galactosidase (*lacZ* product) plasmid-containing cell can grow. The isopropyl-B-D-thiogalactoside (IPTG) is a *lac* inducer, causing transcription of the *lacZ* gene on the plasmid. The X-gal is a colorless substrate that β-galactosidase will cleave into a blue product. Blue colonies developing on the plates are derived from a cell that took up a plasmid that had re-ligated on itself. White colonies developed from plasmids that re-ligated with inserts.

All plasmid vectors used in this research also contained one or more gene for antibiotic resistance. For the all plasmid vectors, the presence of an ampicillin resistance gene was constructed; kanamycin resistance gene was also contained on pDrive. The presence of the gene(s) for antibiotic resistance on the plasmid vectors permitting selection with these antibiotics. Transformants of bacterial cells with these plasmid vectors were selected simultaneously in two ways, blue/white screening and antibiotic selection. Culture medium agar plates (15 ml/ plate), 2x YT or LB media, containing appropriated antibiotic (100 μg ampicillin/ml medium agar) were spread with 12 μl of 100 mg/ml X-gal and 7.5 μl of 100 mM IPTG and the plates were kept in incubator setting at 37 ^{o}C for 30 min to pre-warm. The pre-warm plates were used as selective medium for culturing transformants. After plating suspension of bacterial cells obtained from transformation, the culture plates were incubated at 37 ^{o}C for 14-18 hr. White colonies were analyzed by PCR or plasmid mini-preparation and with appropriate restriction endonuclease to verify presence and direction of inserts in recombination plasmids. White colonies usually contained desired inserts while blue colonies did not.

7.2 Analysis of transformants

For a variety of reasons, however, every cell that takes up a plasmid DNA does not necessarily mean that the plasmid contains the desired PCR product. Some cells may have been transformed with plasmid vectors from failed ligation reactions that either contained no PCR product or small DNA fragments (oligonucleotide primers) instead of the PCR product of interest. The *lacZ* gene encodes the enzyme β-galactosidase which catalyzes the hydrolysis (cleavage) of sugars containing a galactose residue. When this synthetic sugar is cleaved by β-galactosidase, the freed dye compound produces a blue color. Colonies of cells with an active β-galactosidase enzyme will show up as blue spots on the agar plate.

7.2.1 Analysis of transformants by PCR

Colony PCR involves lysing the bacteria and amplifying a portion of the plasmid. Either insert-specific primers or vector-specific primers can be used to screen for recombinant plasmid and orientation. PCR cloning using the A-overhangs left by Taq DNA polymerase and a commercially prepared cloning U-tailed vector, pDrive (QIAGEN), is not a technique that will retain orientation. The orientation can be rapidly assessed with colony PCR using vector-specific primers (M13 forward (-20): 5'-GTA AAA CGA CGG CCA GT-3', and M13 reverse: 5'-AAC AGC TAT GAC CAT G-3') and insert-specific primers as detailed in Table 3.

Each white colony was touched and removed from the agar plate with a sterile toothpick. The toothpick was dipped and swirled in a PCR tube and then streaked it onto a fresh replicate agar plate (2xYT). The replicate agar plate was incubated overnight at 37 ^{o}C, the colonies with expected size of insert fragment were separately cultured at 37^{o}C overnight for plasmid DNA miniprep by alkaline lysis with SDS. For colony screening by PCR, each *E. coli* inoculated PCR tube was added with 20 µl of 1xPCR mixture contained 20 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each dATP, dGTP, dCTP and dTTP, 200 nM each of the forward and reverse primers and 1.25 U *Taq* DNA polymerase (Fermentas, Life Sciences). The mixture were mixed thoroughly, and a thermal cycling scheme was use for 25 reaction cycles (MJ Research PTC-200 Peltier Thermal Cycler), an initial template denatuation at 94^{o}C for 2 min, followed by 25 cycles of template denatuation at 94^{o}C for 30 sec, primer annealing at 43-56^{o}C (depending upon the T_m of the primers) for 30 sec and DNA synthesis at 72^{o}C for 2 min. The final extension step was one cycle at 72^{o}C for 5 min. The amplification (PCR) products were analyzed on a 1% agarose gel in 0.5x TAE.

7.2.2 Analysis of transformants by plasmid DNA miniprep/restriction digestion

Following colony screening and analysis of transformants by PCR, colonies of desired transformants were cultured overnight at 37 ^{o}C with shaking at 250 rpm in 3 ml 2x YT broth containing 100 µg/ml of ampicillin. Plasmid DNA was isolated from the small-scale bacterial (*E. coli*) cell culture by alkaline lysis with SDS method as described in section 8. An aliquot of plasmid DNA was digested with appropriate enzyme(s) to confirm the presence, size and direction of inserts in a vector. When PCR products were cloned in pGEM-T Easy or pDrive, a single digest with *Eco*RI released the insert. Alternatively, and also in pTZ57R, a double digest with *Bam*HI and *Xba*I endonucleases released the insert in the plasmid. Linear, non-recombinant plasmids obtained from digesting their self-ligation with appropriate enzyme were run simultaneously with digested recombinant plasmids to act as a DNA marker and to confirm the presence of a desired insert.

8. Preparation of plasmid DNA by alkaline lysis with SDS: Plasmid DNA miniprep

Plasmid DNA was been able to isolate from small-scale (1-3 ml) bacterial (*E. coli*) cultures by treatment with alkaline and SDS. The resulting plasmid DNA preparation could be further determined by PCR or restriction endonuclease digestion and followed by agarose gel electrophoresis. With further purification by using a QIAquick gel extraction kit (QIAGEN), the preparation could be used as templates in DNA sequencing reactions.

E. coli colonies with expected size of insert fragment from overnight culture on replicate agar plate (2xYT), containing putative recombinant plasmids, were picked and inoculated into 3 ml 2xYT broth containing 100 µg/ml ampicillin and/or 30 µg/ml kanamycin. Culture were grown overnight with vigorous aeration at 250 rpm at 30^{o}C. To harvest the bacterial cells, the overnight cultures were transfer to microcentrifuge tubes and then centrifuged at maximum speed for 30 sec at 4 ^{o}C in a

microcentrifuge. After centrifugation, the medium supernatant was discarded by aspiration and the bacterial pellet was left as dry as possible. The bacterial pellets were resuspended in 100 µl of ice-cold alkaline lysis solution I (50 mM glucose, 25 mM Tris-Cl (pH 8.0) and 10 mM EDTA (pH 8.0)) by vigorous vortexing. Two hundred µl of freshly prepared Alkaline lysis solution II (0.2 N NaOH and 1% (w/v) SDS) was added to each bacterial suspension. The tube caps were closed tightly and the contents were mixed by inverting the tube rapidly five times. The tubes were incubated on ice for 5 min. One hundred and fifty µl of ice-cold alkaline lysis solution III (3 M potassium acetate and 2 M glacial acetic acid) were added. The tubes were closed and alkaline lysis solution III was dispersed through the viscous bacterial lysate by inverting the tube several times. This solution neutralized NaOH in the previous lysis step while precipitating the genomic DNA and SDS in an insoluble while precipitate. The tubes were incubated on ice for 10 min. The bacterial lysate was centrifuged at maximum speed for 5 min at 4^{o}C in a microcentrifuge. The supernatant was transferred to a fresh tube containing 450 µl of chloroform: isoamyl alcohol and the organic and aqueous phases were mixed by vigorous vortexing. The obtained emulsion was centrifuged at maximum speed for 2 min at 4^{o}C in a microcentrifuge and the aqueous upper layer was transferred to a fresh tube containing 900 µl of absolute ethanol at room temperature, and the solution was mixed by inverting the tube and then the mixture was allowed to stand for 2 min. The precipitated DNA plasmids were pelleted by centrifugation at maximum speed for 5 min at 4^{o}C in a microcentrifuge. The supernatant was removed by gentle aspiration and then the tube was stood in an inverted position on a paper towel to allow all of the fluid to drain away. The pellet was rinsed by adding 1 ml of 70% ethanol and the closed tube was inverted several times. The DNA plasmids were recovered by centrifugation at maximum speed for 2 min at 4^{o}C in a microcentrifuge. All of the supernatant was removed again by gentle aspiration and the open tube was stored at room temperature until the ethanol had evaporated and no fluid was visible in the tube. The nucleic acid pellet was dissolved in 50 µl of sterile water containing 20 µg/ml of DNase-free RNase A. The solution was mixed by gentle vortexing for a few second, and then incubated at 37^{o}C for 1 hr. The DNA solution, resulting plasmid DNA preparation, was be further determined by agarose gel electrophoresis and

restriction endonuclease digestion, respectively. The recombinant plasmid DNA, containing insert of the expected size, was further purified and the purified plasmid DNA was used as template in nucleotide sequencing reaction.

9. Nucleotide sequencing

The nucleotide sequence of the complete PRSV-W genome was determined by the dideoxy chain termination as modified method of Sanger *et al.* (1977) using the ABI PRISM® Bigdye™ Terminator Cycle Sequencing Ready Reaction Kits version 2.0 and the reactions obtained were run on either the model 310 or model 377 automatic DNA sequencer (Applied Biosystems). Primers complementary to adjacent vector sequences and to the virus sequences were used for the sequencing reactions of all inserted cDNA fragments in pDrive (QIAGEN) and pGEM-T Easy (Promega) cloning vectors.

In this research, eight generated overlapping recombinant clones containing cDNA fragments of approximately 1 to 2.2 kbp corresponding to the genomic RNA of PRSV-W were generated by RT-PCR and named pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH in the order in which inserted fragments correspond on the virus genome (Figure 1). Each plasmid DNA clone containing the inserted fragments from these recombinant clones was purified using the high pure plasmid isolation kit (Roche) and used as template for automatic sequencing. Each overlapped cDNA clone containing the viral genome was sequenced with universal pUC/M13 forward and reverse primers (vector primers), and internal sequencing primers designed based on the partial nucleotide sequence obtained by primer walking strategy. Both strands of these cDNA clones were sequenced in both directions to avoid sequence ambiguities. The first primers used were the universal primers M13F and M13R and internal sequences of inserted cDNA fragments in the vectors were obtained using specific primers designed from the partial sequences already determined. Overlapping sequences were obtained using universal primers and the specific primers. All primers used to sequence in this work, including 2 universal

primers and 24 specific oligonucleotide primers corresponding to the viral genome designed and *in vitro* synthesized to internal sequences, are listed in Table 3.

Additionally, the 35S and T-Nos separately inserted in pDrive (QIAGEN) cloning vectors were also sequenced with universal pUC/M13 forward and reverse primers (vector primers). Two recombinant plasmids consisted p2324 and p2311 containing 35S CaMV promoter (532 bp) and Nos terminator (237 bp) respectively were kindly provided by Dr. Srimek Chowpongpang and were used as DNA templates in PCRs to amplified with specific primers and cloned into the pDrive cloning vectors.

9.1 Sequencing reaction

Complete nucleotide sequence of PRSV-W was determined from a total of 39 overlapping nucleotide sequences as illustrated in Figure 2. These nucleotide sequences of cloned inserts were obtained by sequencing double-stranded templates of recombinant plasmids. Supercoiled plasmids with inserted DNA were prepared and further purified for sequencing from overnight cultures using the High Pure Plasmid Isolation Kit (Roche). The purified plasmids with the desired DNA insert were used as template in sequencing reactions. Priming for sequencing was either from vector or insert sequences (Table 3). Cycle sequencing reactions were carried out in either a MJ Research PTC-200 Peltier Thermal Cycler or a ThermoHybaid PxE Thermal cycler. Sequencing was performed using dideoxy sequence analysis with the ABI PRISM® Bigdye™ Terminator Cycle Sequencing Ready Reaction Kits version 2.0 (Applied Biosystems Industries; ABI). Sequencing reactions were half the volumes recommended by ABI, consisting of 400-600 ng of plasmid DNA (in maximal volume 5 µl) added to 4 µl of Bigdye™ Terminator Ready Reaction Mix, 3.2 pmol of sequencing primer, made to a final volume of 10 µl with deionized water in a 0.2-ml thin-wall PCR tube. After mixing and brief spinning down, the reactions were incubated for thermal cycling by placing in a thermal cycler. Thermal cycling was set to run on the following program: at 96 ^{o}C for 30 sec followed by 25 cycles of

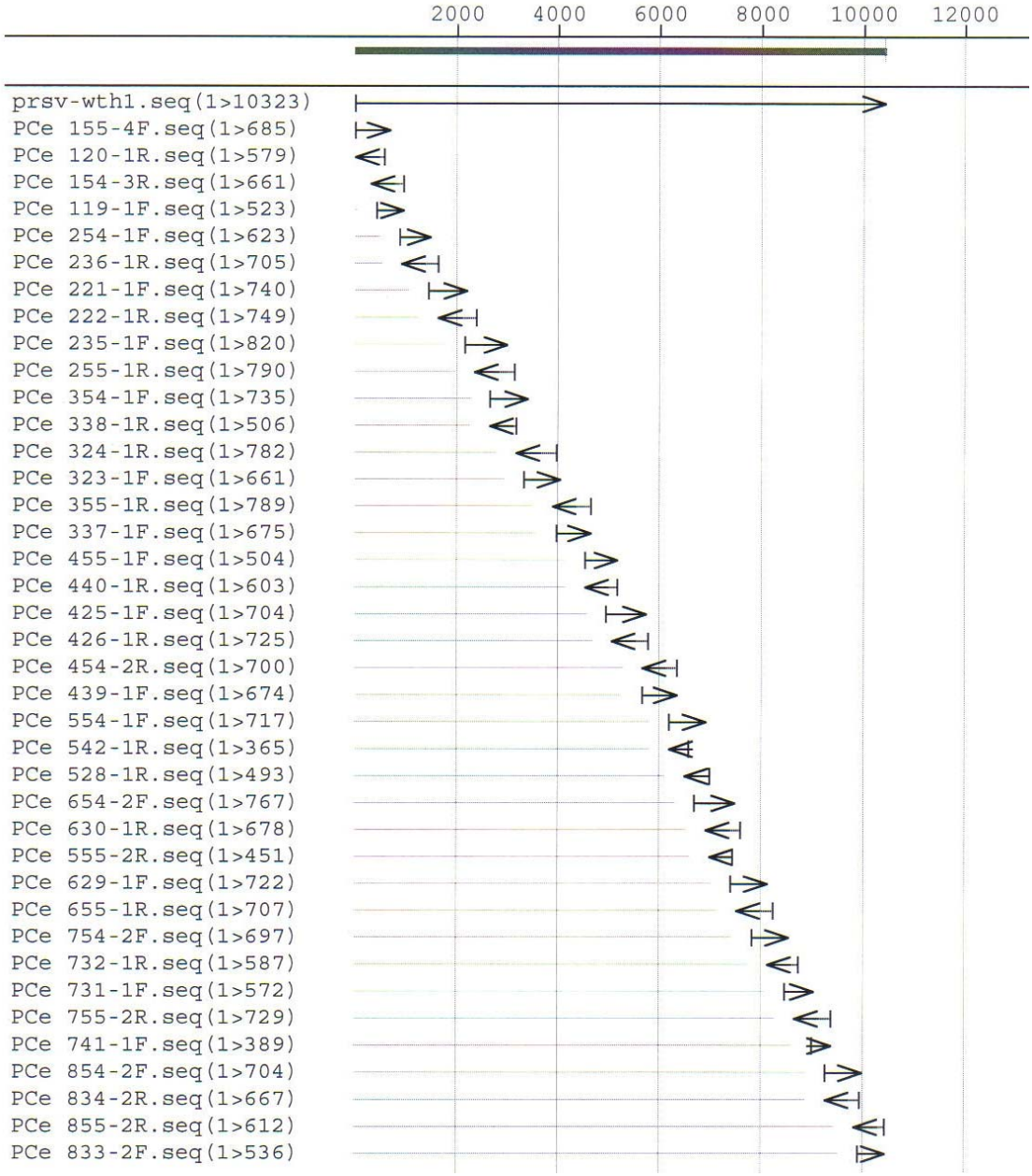


Figure 2 Schematic representation of the nucleotide sequencing of the PRSV-W RNA. Eight cDNA clones (pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH) containing cDNA fragments which correspond to the viral genome (Figure 1) were used separately as template for sequencing. Each cDNA clone was sequenced with universal pUC/M13 forward and reverse primers (vector primers), and internal sequencing primers designed based on the partial nucleotide sequence obtained by primer walking strategy (Table 3). Arrows show the location, direction and length of nucleotide sequences. Nucleotide residues of PRSV-W RNA are numbered in bases beginning at the 5' terminus.

Table 3 Oligonucleotide primers used for nucleotide sequencing

Primer	Primer sequence (5'-3')	Plasmid
Universal primer		
M13F	GTA AAA CGA CGG CCA GT	plasmid vector
M13R	AAC AGC TAT GAC CAT G	plasmid vector
Specific primer		
PC19	GGT TAT GGT GCT GTT ACT CTT G	A 378-399F
PC20	CAC AAA CTT AGG TTC ATC AGG TA	A 603-625R
PC21	GGG ACG GGA AAA ACT ATG TG	B 1422-1441F
PC22	ACC CTT ACT GTG ATC AAT CTT CTC	B 2396-2419R
PC35	GAC GGT CAT CAA CAA TAC GC	B 2125-2144F
PC36	CAA AAT ATC GCC CTG TTC G	B 1613-1631R
PC23	GAT CAA AAG GGA CCA AGA TGT T	C 3277-3298F
PC24	CGC CTT TGC TTG CTT TAG T	C 3943-3961R
PC37	GCC ATG GAG CAT ATA CAA CTA AA	C 3926-3948F
PC38	GCG GAG TTC ATG GGG TTT	C 3152-3169R
PC25	GAA CGC GAT TGA GAA GAA AGT	D 4864-4884F
PC26	TGA TCA AGA ATG GTA TCC GAA C	D 5753-5774R
PC39	GGA ACA GTG CAC CCC AAG	D 5591-5608F
PC40	CCG CCC ATC GAT TTT AGT A	D 5113-5131R
PC27	GAA CTC CGC CAA AAC TGC	E 6751-6768F
PC28	TGT AAT TTC TCA TAC CTT GGC A	E 6944-6965R
PC42	CGT GCA TCC TTC CTT TCT TG	E 6496-6515R
PC29	CAT ATC GCT AGT AAG GTG TCT GAG	F 7286-7309F
PC30	CAT TGT AGC TCC AGT GTT TGC	F 7524-7544R
PC31	CTG ATA GTT GGG TTT ACT GC	G 8358-8377F
PC32	TTC CTG GGC CTC CGT GT	G 8652-8668R
PC41	TCC TAA ACT TGA ACC CGA GAG	G 8863-8883F
PC33	CCT TTG ATT GAG CAC GCA	H 9755-9772F
PC34	GCG GCA TGT ACC TCT CAG T	H 9845-9863R

denaturation at 96 ^{o}C for 20 sec, annealing at 50 ^{o}C for 10 sec and extension at 96 ^{o}C for 4 min.

9.2 Purification of extension products for sequencing

After programmed thermal cycle was complete, the extension products resulted in sequencing reaction mixture were cleaned up by ethanal/sodium acetate precipitation. The sequencing reaction mixture completely incubated was cooled by chilling on ice, briefly centrifuged and then transferred into a 1.5-ml centrifuge tube containing 2.5 vol of 95% ice-cold ethanol (25 µl) and 0.1 vol of 3 M sodium acetate, pH 4.6 (1 µl). After the tube was vertexed briefly, the mixture was cooled by placing in refrigerator at -20 ^{o}C for 20 min. The extension products were precipitated by centrifugation the tube in a microcentrifuge for 20 min at maximum speed. The obtained pellet was rinsed with 125 µl of 70% ethanol. The tube was placed and spun in a microcentrifuge for 5 min at maximum speed. The supernatant was carefully aspirated with a pipette tip and discarded. The pellet was dried either under vacuum condition for 15 min or under heat condition by placing the tube with lid open in a heat block at 90 ^{o}C for 1 min. For sequencing, the pellet of purified extension products resulted from sequencing reaction was sent to the DNA Technology Laboratory, KU/BIOTEC at Kasetsart University, Kamphaengsaen campus, Nakhon Pathom or the DNA Sequencing Service Unit at Kasetsart University, Bangkhen campus, Bangkok using a model 377 and a model 310 automatic DNA sequencer (Applied Biosystems), respectively.

9.3 Computer sequence data analysis

In this research, personal computer-based program DNAStar biocomputing software (DNAStar, Inc.) was used to manage and analyze the obtained sequence data. The software package has several program including EditSeq, SeqMan II, PrimerSelect, MapDraw, MegAlign and Protean. Another of computing software, Vector NTI (InforMax) was also used for constructing and designing of circular- and linear- cDNA molecule and figure drawing of schematic map.

Nucleotide sequence data corresponding to cDNA inserts were obtained following electrophoresis on the automated sequencer. From eight independent cDNA clones containing overlapped fragments which were corresponding to PRSV-W genome, these resulting raw data of each nucleotide sequence were edited, assembled and trimmed to remove primer annealing regions and vector sequence using the program EditSeq and SeqMan II. Alignment of multiple nucleotide and deduced amino acid sequences, representing the full-length cDNA clones of PRSV-W, were constructed and analyzed using the MegAlign program. Via the World Wide Web internet, the complete nucleotide and the deduced amino acid sequences obtained from sequencing and assembling were first compared with all sequences deposited in Gen Bank database. Computer sequence comparisons were performed using the default settings of the basic local alignment search tool (BLAST) available program from http://www.ncbi.nlm.nih.gov/BLAST on the National Center for Biotechnology Information (NCBI) Network server at http://www.ncbi.nlm.nih.gov/Genbank/index.html .

10. *In vitro* amplification of DNA by polymerase chain reaction

In this research, PCRs were performed using either a MJ Research PTC-200 Peltier Thermal Cycler (MJ Reserch, Inc.) or a PxE Thermal Cycler (Thermo Electron Corporation) and the PCR products obtained were catalyzed by a variety of thermostable DNA polymerase in different objectives and operating conditions. The thermostable DNA polymerases used in PCRs were *Taq* (Fermentas), *Tth* (Promega) and *DeepVent* (New England Biolabs Inc) DNA polymerases (Table 4).

10.1 Routine PCR

Routine PCRs were performed with *Taq* DNA polymerase to operate in several applications in this research such as amplification of second strand cDNA or double stranded (ds) DNA fragments, verification of insertion and determination of orientation of cloned fragments in vectors and colony screening of bacterial transformants. Mixture volumes of PCR ranged from 10-50 µl and consisted of 1X

Table 4 Thermostable DNA polymerase used for *in vitro* amplification of DNA by PCR in this study

Enzyme	Optimum Temperature (°C)	Exonu-clease activity	Fidelity	Stability (min. at specified temperature)	Manufacturer
Taq (recombinant)	75-80	5'→ 3'	low	40 min at 95 °C	Fermentas
Tth strain HB-8	75-80	5'→ 3'	low	20 min at 95 °C	Promega
DeepVent (recombinant) strain GB-D	70-80	3'→ 5'	high	480 min at 100 °C	New England Biolabs Inc

PCR buffer (10 mM Tris-HCl, pH8.8, 50 mM KCl and 0.08% (v/v) Nonidet P40), 1.5-2.0 mM $MgCl_2$, 0.2 mM of each dNTP (dATP, dGTP, dCTP and dTTP), 0.1-1.0 µM of each forward and reverse primer, 10-200 ng DNA template and 0.25-1.25 units of *Taq* DNA polymerase (Fermentas). The mixture reactions were prepared in 0.2-ml thin wall PCR tube and carried out in a thermal cycler. A typical PCR thermal cycling profile consisted of a single initial denaturation at 94 °C for 1-4 min, followed by 25-35 cycles of denaturation at 94 °C for 30 sec, primer annealing at temperatures depending on the Tm of the primers used in each reaction (40-60 °C) for 30 sec and primer extension at 72 °C for 1-3 min (approximately 1 kb per min). A final incubation step at 72 °C for 5-7 min was included to ensure complete extension.

10.2 High fidelity PCR

High fidelity PCRs were performed with *DeepVent* DNA polymerase (New England Biolabs Inc) to amplify PCR products with blunt ends. Typically, the

PCR reaction consisted of 20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% (v/v) Triton X-100, 0.2 mM of each dNTP (dATP, dGTP, dCTP and dTTP), 0.4 µM of each forward and reverse primer, 10-200 ng DNA template and 1 unit of *DeepVent* DNA polymerase (New England Biolabs Inc) in a final volume of 50 µl. The mixture reactions were prepared in 0.2-ml thin wall PCR tube and carried out in a thermal cycler. A typical PCR thermal cycling profile consisted of a single initial denaturation at 94 ^{o}C for 3 min, followed by 35 cycles of denaturation at 94 ^{o}C for 30 sec, primer annealing at temperatures depending on the Tm of the primers used in each reaction (40-60 ^{o}C) for 30 sec and primer extension at 68 ^{o}C for 1-3 min (approximately 1 kb per min). A final incubation step at 68 ^{o}C for 7-10 min was included to ensure complete extension.

10.3 Long PCR

The combination of *Taq* (Fermentas) or *Tth* (Promega) DNA polymerase and *DeepVent* DNA polymerase (New England Biolabs Inc) was prepared at a ratio of 20:1 (unit: unit) in a final volume of 90 µl and stored at -20 ^{o}C until use. The amplification reaction consisted of 60 mM Tris-SO_4, pH 9.1, 18 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.2 mM of each dNTP (dATP, dGTP, dCTP and dTTP), 0.4 µM of each forward and reverse primer, 50-100 pg DNA template and 1 µl of the DNA polymerase mix in a final volume of 50 µl. The mixture reactions were prepared in 0.2-ml thin wall PCR tube and carried out in a thermal cycler. A typical PCR thermal cycling profile for long PCR consisted of a single initial denaturation at 94 ^{o}C for 5 min, followed by 25-30 cycles of denaturation at 94 ^{o}C for 1 min, primer annealing at temperatures depending on the Tm of the primers used in each reaction (55-65 ^{o}C) for 1 min and primer extension at 68 ^{o}C for 7-12 min (approximately 1 min for every kilobase pair amplified target). A final incubation step at 68 ^{o}C for 10 min was included to ensure complete extension. After amplification, 2 µl of the reaction mixture was determined on a 0.8% TAE agarose gel. Amplified DNA fragments were visualized by UV light after staining the gel with ethidium bromide. Fragment size was estimated by inclusion of 1kb DNA ladder markers.

10.4 Overlap extension PCR

Primarily, two or more blunted-end DNA fragments, which each has overlapping ends (shared common sequences in their terminal regions) were amplified with *DeepVent* DNA polymerase in high fidelity PCR method (products of smaller than 3 kb) or the DNA polymerase mix in long PCR method (products of larger than 3 kb). The DNA fragments amplified were isolated on 0.8-1.0% TAE agarose gel and purified using a QIAquick® Gel Extraction Kit (QIAGEN) as recommended by the manufacturer. The purified DNA fragments were uses as templates in the next steps in overlap extension PCR.

Overlap extension PCR procedure features in two steps consisting the reaction step A, approximately 50 ng of each DNA fragment (and 10 ng of central DNA fragment in the case of a simultaneous assembly (fusion) of three DNA fragments) was added to the standard amplification reaction mixture (final volume of 50 µl) without both primers (forward and reverse) in the high fidelity PCR method (products of smaller than 3 kb) or in the long PCR method (products of larger than 3 kb). Cycling parameters consisted of initial denaturation at 94 ^{o}C for 3 min, subsequent steps at 94 ^{o}C for 1 min, annealing at 50-60 ^{o}C for 1 min, extension at 72 ^{o}C for 1-3 min, 10 cycles total and then the reaction mixture was hold at 4 ^{o}C. The reaction step B, two µl of unpurified PCR product from step A was added to the standard amplification reaction mixture (final volume of 50 µl) with 0.4 µM of each forward and reverse primer in the high fidelity PCR method (products of smaller than 3 kb) or in the long PCR method (products of larger than 3 kb). PCR thermal cycling profile consisted of a single initial denaturation at 94 ^{o}C for 5 min, followed by 25-30 cycles of denaturation at 94 ^{o}C for 1 min, primer annealing at temperatures depending on the Tm of the primers used in each reaction (55-62 ^{o}C) for 1 min and primer extension at 68 ^{o}C for 7-12 min (approximately 1 min for every kilobase pair amplified target). A final incubation step at 68 ^{o}C for 10 min was included to ensure complete extension. Both mixture reactions in step A and step B were prepared in 0.2-ml thin wall PCR tube and carried out in a thermal cycler.

10.5 Single adenosine (A) nucleotide-tailing reaction

To perform the A-tailing reaction for blunt-ended DNA or PCR fragments generated by proofreading DNA polymerase, the DNA was ethanol precipitated and resuspended in 20 µl DEPC-treated water. The reaction mixture consisted of 10 µl purified PCR product, 1X PCR buffer (10 mM Tris-HCl, pH8.8, 50 mM KCl and 0.08% (v/v) Nonidet P40), 2 mM $MgCl_2$, 0.2 mM of dATP, and 1 unit of *Taq* DNA polymerase (Fermentas). The mixture reactions were prepared in 0.2-ml thin wall PCR tube and incubated in a thermal cycler at 72 ^{o}C for 30 min.

11. Agarose gel electrophoresis

To prepare 100 ml of a 1% agarose solution, one gram agarose powder was put into a flask and then added 100 ml electrophoresis buffer (0.5x TAE). The agarose powder was mixed and then heated in a microwave oven until completely melted. After cooling the solution to about 60^{o}C, it was poured into a casting tray containing a sample comb and allowed to solidify at room temperature. Excess agarose could be stood at room temperature and remelted in a microwave oven. After the gel had solidified, the comb was removed, using care not to rip the bottom of the wells. The gel, still in its plastic tray, was inserted horizontally into the electrophoresis chamber and just covered with electrophoresis buffer. To prepared samples for electrophoresis, 1 µl of 6x gel loading dye was added for every 5 µl of DNA solution and then the mixture was mixed and used as electrophoresis (DNA) sample. Samples containing DNA mixed with loading buffer were then pipetted into the samples wells, the lid and the current was applied by electrophoresis in 0.5x TAE at 50 or 100 volts for 30-60 min or until dye markers had migrated an approximate distance, depending on the size of DNA to be visualized. In this research, DNA size standard markers used included a 1 kb DNA Ladder (New England BioLabs Inc.), GeneRuler^{TM} 1 kb and 100 bp DNA Ladder (Fermentas) and a High Range RNA Ladder (Fermentas).

The gel was stained after electrophoresis by soaking in 0.5 μg/ml ethidium bromide until the DNA had take up the dye (5 min). The gel was removed to a tray of water and allowed it to destain for 5-10 min. DNA (band) fragments were visualized by staining with ethidium bromide. This fluorescent dye intercalate between bases of DNA and RNA. To visualized DNA or RNA, the gel was placed on a ultraviolet transiluminator, DNA bands were visualized under short wave UV light and the DNA band patterns were captured and recorded as figures in different file formats by using GeneSnap (Syngene), if the DNA would not be used further, or with a long-wave UV light if the DNA was to be cut out and purified.

12. DNA purification

DNA fragments were purified from solutions of several reaction mixtures or from bacterial cell cultures. And DNA fragments separated in an agarose gel were also isolated and purified into an aqueous solution. Protocols performed were described as following.

12.1 Phenol-chloroform extraction

To remove proteins from DNA solution, extracting aqueous solutions of DNAs were performed with phenol: chloroform and chloroform organic solvents.

DNA was purified from a solution by extracting with an equal volume of TE saturated phenol: chloroform: isoamyl alcohol (25:24:1) in a 1.5 ml microcentrifuge tube. The mixture was vigorously vortexed for 2 min or until an emulsion formed and then centrifuge at maximum speed for 5 min at room temperature in a microcentrifuge. The aqueous (upper) phase was transferred to a fresh tube, the interface and organic phase were discarded. The aqueous phase obtained was extracted again with an equal volume of TE saturated chloroform: isoamyl alcohol (24: 1). The mixture was vigorously vortexed for 1 min and then centrifuge at maximum speed for 5 min. The aqueous (upper) phase was transferred to a fresh tube. The DNA was then precipitated as described in the next section.

12.2 DNA precipitation

DNA was precipitated with either salt/ethanol solution or with ice-cold isopropanol. Salt/ethanol solutions were made of 0.1 volume of 3 M sodium acetate (pH 5.2) with 2.5 volumes of absolute ethanol. DNA solution was mixed with sodium acetate/ethanol solution or one volume of ice-cold isopropanol, followed by incubation at -20 ^{o}C for 1 hr, and centrifugation at maximum speed for 30 min at 4 ^{o}C in a microcentrifuge. The supernatant was carefully removed with an automatic micropipettor and the DNA pellet was washed with 1 volume of 70% ice-cold ethanol by centrifugation at maximum speed for 5 min, the supernatant was carefully removed. The pellet was dissolved in a desired volume of TE buffer, pH 8 or sterile distilled water after air-drying.

12.3 Isolation of DNA fragment from agarose gel

DNA fragments were isolated from agarose gels using either silica particles in sodium iodide solution in method 1 or using QIAquick® Gel Extraction Kit (QIAGEN) in method 2. These methods were described in this section.

Primarily, the same protocol was performed to prepare the piece of agarose gel in both methods. After electrophoresis and UV visualization of ethidium bromide-stained gel, the desired DNA band was excised from the gel (piece of agarose gel containing the desired DNA fragments) using a razor blade. The piece of gel was placed in a 1.5 ml microcentrifuge tube, and gel weighed only.

Method 1: Three volumes of 6 M sodium iodide (NaI) were added into the tube containing the gel piece (3: 1 (v/w) e.g., 300 µl of 6M NaI to each 100 mg of gel piece). The gel piece in solution was incubated at 50 ^{o}C for 10 min or until the gel piece had dissolved completely. Ten µl of silica particles submerged in 6M NaI were added into the tube and then the tube was vortexed for 5 sec to suspend and followed by slight shaking for 2 min. The silica particles were spun down by centrifugation at maximum speed for 5 sec, and then the supernant was carefully removed with an

automatic micropipettor and the silica particles were washed with 1ml of washing buffer containing 10 mM Tris-HCl buffer, pH 7.5, 1 mM EDTA, 100 mM NaCl and 50% ethanal) by vortexing for 10 sec and followed by centrifugation at maximum speed for 30 sec, the supernatant was carefully removed. The DNA-bound silica particles were dried by incubation at 37 ^{o}C for 10 min. A desired volume of TE buffer, pH 8 or sterile distilled water (30-50 µl) was added into the tube containing the dried particles, and the tube was slightly vortexed followed by centrifugation at maximum speed for 1 min. The supernatant (containing DNA) was transferred to a fresh tube.

Method 2: Three volumes of buffer QG were added into the tube containing the gel piece (one volume). The piece of gel in buffer QG was incubated at 50 ^{o}C for 10 min or until the gel piece was dissolved completely. To help dissolve gel, the tube was vortexed every 3 min during the incubation. After the gel piece was dissolved completely, the solution was applied to a QIAquick column placed in a 2 ml collection tube and then the column was centrifuged at maximum speed for 1 min in a microcentrifuge. The DNA was bound to the column and the flow-through was discarded. The column was placed back in the same collection tube. For washing, 750 µl of buffer PE were added into the column and the column was centrifuged at maximum speed for 1 min. The flow-through was discarded and centrifugation was repeated for an additional 1 min. The column was placed into a sterile 1.5 ml microcentrifuge tube. To elute DNA from column, 30-50 µl of TE buffer or sterile distilled water were loaded into the center of a membrane within the column. After one min of the tube standing, the column placing in a microcentrifuge tube was centrifuged at maximum speed for 1 min.

13. Quantitation of nucleic acids

Purity and concentration of nucleic acids were determined by spectrophotometric analysis. The ratio of absorbance at (A260-A320/A280-A320) should be 1.8 to 2.0 for DNA and higher as 2.0 for RNA. In presence of protein contamination the ratio is less. For the measurement, a model spctronic1201

spectrophotometer from Milton Roy Company was used with a quartz cuvette. The DNA or RNA was diluted 1:100 in distilled water and transferred to a quartz cuvette.

The absorption was at wavelengths of 260 nm, 280 nm and 320 nm. An optical density (OD) of 260 nm of 1.0 is equivalent to 50 µg/ml double stranded DNA, 38 µg/ml single stranded DNA or RNA or 20 µg/ml oligonucleotides. The spectrophotometer machine used was provided with the concentration and the ratio automatically. However, the formula used to calculate the concentration (C) and molarity (M) were the following:

$$C = OD_{260nm} \times \text{dilution factor} \times \text{equivalent} = x\ \mu g/ml$$

$$M = \text{Concentration } (\mu g/ml) \times 1{,}000/330 = x\ \text{pmol}/\mu l$$

14. Restriction endonuclease digestion of DNA

In this study, six different restriction endonucleases were used as tool for manipulation of recombinant DNA and for analysis of plasmid DNA minipreps which were defined in Table 5. Restriction endonuclease digestions were performed by incubating ds DNA molecule with an appropriate amount of restriction enzyme, in its respective buffer as recommended by the supplier, and at the optimal temperature for that specific enzyme. The enzyme was inactivated by heating incubation at the appropriate temperature for 10-15 min. An aliquot of the digestion was assayed by agarose gel electrophoresis reverse non-digested DNA and a size marker.

A portion of the plasmid was screened by restriction endonuclease digestion. In this study, three different cloning vectors, pGEM®-TEasy (Promega), pDrive (QIAGEN) and pTZ57R (Fermentas), were used to clone and construct recombinant DNA. All of the commercially prepared cloning vectors are high-copy plasmids and physical maps of the vectors, pGEM®-TEasy (Promega), pDrive (QIAGEN) and pTZ57R (Fermentas), were illustrated in Appendix A Figure 2, 1, and 3, respectively.

Table 5 Restriction endonucleases used in this research

Enzymes	Recognition sequences	Incubation temperature (°C)	Temperature (°C) for heat inactivation	Manufacturer
*Bam*HI	5′…G↓G A T C C… 3′ 3′…C C T A G↑G… 5′	37°C	80°C for 20 min	Fermentas
*Eco*RI	5′…G↓A A T T C… 3′ 3′…C T T A A↑G… 5′	37°C	65°C for 20 min	Fermentas
*Nco*I	5′…C↓C A T G G… 3′ 3′…G G T A C↑C… 5′	37°C	65°C for 20 min	Fermentas Promega
*Ngo*MIV	5′…G↓C C G G C… 3′ 3′…C G G C C↑G… 5′	37°C	65°C for 20 min	Promega
*Sal*I	5′…G↓T C G A C… 3′ 3′…C A G C T↑G… 5′	37°C	65°C for 20 min	Fermentas
*Xba*I	5′…T↓C T A G A… 3′ 3′…A G A T C↑T… 5′	37°C	65°C for 20 min	Fermentas

The specific cleavage sites are indicated with ↓ and ↑ in each DNA sequence recognized with each restriction endonucleases.

15. Inoculation of an *in vitro*-synthesized cDNA construct by particle bombardment

In this research, particle bombardment was performed to introduction of a constructed cDNA clone of PRSV-W into test plants for infectivity assay with a Particle Inflow Gun (PIG) at the Plant Genetic Engineering Unit (BIOTEC), Kasetsart University, Kamphaengsaen campus, Nakhon Pathom.

15.1 Microcarrier preparation

Tungsten particles (~1.1 μm in diameter) were used as the microparticles (microcarrier) and were prepared as based on the method of Sanford *et al*. (1993). Tungsten particles were washed by adding 30 mg of tungsten particles into a 1.5 ml microcentrifuge tube containing 1 ml of freshly prepared 70% (v/v) ethanol. After the tube was vertexed vigorously for 5 min, the particles were drenched in 70% ethanol for 15 min. The microparticles were precipitated by spinning the tube in a microcentrifuge for 5 sec and then the supernatant was removed and discarded. The microparticles were washed repeatedly three times as the following procedure. One ml of sterile water was added into the tube and then the tube was vertexed vigorously for 1 min. The microparticles were precipitated by briefly spinning the tube in a microcentrifuge for 5 sec and then the liquid was removed and discarded. After the third wash, the microparticle concentration was brought to 60 mg/ml adding 500 μl of sterile 50% glycerol. The tungsten suspension (60 mg/ml) was prepared into 50 μl aliquots and stored at -20 ^{o}C to prevent oxidation.

15.2 Coating DNA onto microcarrier

A suspension of microcarriers coating with DNA was prepared as the following procedure which is sufficient for 6-8 bombardments. An aliquot of the microcarrier suspension (60 mg/ml) prepared in 50% glycerol was vertexed for 5 min to resuspend and disrupt agglomerated particles. Fifty μl of microcarrier suspension (3 mg) were removed to a fresh tube. The aliquot was continuously agitated using

mixer with a platform for holding microcentrifuge tubes during which 5 µg of DNA, 50 µl of 2.5 M $CaCl_2$ and 20 µl of 0.1 M spermidine were added and then the mixture was continuously agitated for an additional 3 min. The microcarriers were left to settle for 1 min and were then precipitated by spinning for 2 sec in a microcentrifuge. After the supernatant was discarded, the pellet in a minimum volume necessary for deposition of the particles on the microcarrier grid was chilled on ice until bombardment. For particle bombardment, 10 µl of the prepared shooting suspension was pipetted during vortexing, and loaded into the center of a 13-mm diameter swinney (filter unit) from which microcarriers were propelled into plant tissues within the bombardment chamber of the apparatus.

15.3 Bombardment conditions for inoculation of pumpkin cotyledons

For infectivity assay, the pumpkin (*Cucurbita moschata*) cotyledons were used as a plant tissue target introduced a constructed cDNA clone of PRSV-W in bombardment experiment. The cotyledons of pumpkin obtained from 6 day-old *in vitro* cultured seedling on water agar under aseptic condition. Before inoculation by particle bombardment, the plant tissue targets were prepared for shooting under aseptic condition as the following step. Each cotyledon with its 4 cm long petiole was cut from the cultured seedling and placed on the center of a Petri dish containing 15 ml of water agar (water agar plate) and the cotyledon was covered with stainless fine netting with 120 µm mesh to reduce shock to cells on bombardment. Pumpkin cotyledons were bombarded with DNA coated microcarrier using 500 kPa helium pressure at a target distance of 10 cm from the filter unit containing the particles, when the vacuum condition in the bombardment chamber of the apparatus reached 50 cm of mercury (Hg). Each cartridge was shot twice (two consecutive discharges) onto each cotyledon target. After bombardment, inoculated cotyledons were then transferred and cultured *in vitro* on water agar under the condition of a 16 hr illumination at 25 ^{o}C for 14 days, for observation of symptom development. After this period, ELISA was performed to detect systemic infectivity of *in vitro*-synthesized cDNA construct of PRSV-W.

16. *In vitro* transcription and infectivity assay of *in vitro* Transcripts

16.1 *In vitro* transcription

In vitro transcription is a process by which an RNA polymerase, in a reaction mixture containing prerequisite components, imitates *in vivo* transcription and catalyses the generation of an RNA transcript from a DNA template. RNA transcripts synthesized *in vitro* are widely applied in various research studies including constructing infectious viral RNA for *in vivo* infection.

In this research, *in vitro* transcription was reacted to synthesize infectious RNA transcripts of PRSV-W using T7 RNA polymerase (Promega). Full-length cDNAs of PRSV-W were directly combined the T7 RNA polymerase promoter at 5'end of the viral sequence by PCR using the 'PC85'primer containing the sequence of the T7 RNA polymerase promoter as a forward primer. The purified PCR products were used as template DNA in the reaction of *in vitro* transcription. *In vitro* transcripts synthesized were capped at the 5'ends with structure 5'-5' triphosphate-linked guanine modified with a 5' 7-methyl group (m^7G(5')ppp(5')G) (7mG RNA Cap Structure Analog, BioLabs).

Capped *in vitro* transcripts were synthesized from the *in vitro* amplified PRSV-W cDNA template mentioned above using runoff transcription method. *In vitro* transcription reaction was carried out in a 150 µl reaction mixture contained 15 µg of template DNA, 30 µl of 5X transcription buffer, 15 µl of 100 mM dithiothreitol (DTT), 3.75 µl of RNase inhibitor (Fermentas, 40 units/µl), 6 µl of rNTP mix containing 25 mM rATP, 25 mM rUTP, 25 mM rCTP and 2.5 mM rGTP, 30 µl of 2.5 mM RNA cap structure analog (BioLabs) and 6 µl of T7 RNA polymerase (Promega, 40 units/µl). The reaction was incubated at 37 ^{o}C for 2 hr. After this period, 10 units of RNase-free DNase I were added into the reaction mixture and then incubated at 37^{o}C for 30 min. The reaction was stopped by adding 5 µl of 500 mM EDTA. The capped *in vitro* transcripts synthesized were determined on 0.8% agarose gel electrophoresis and RNA concentration was determined using spectrophotometry.

The *in vitro* transcription mixture was used immediately as inoculum in infectivity assays without further treatment.

16.2 Infectivity assay of *in vitro* Transcripts

Six day-old seedlings of systemic host pumpkin (*Cucurbita moschata*) at cotyledon stage were used for infectivity assay of an *in vitro* RNA transcripts corresponding to PRSV-W. Approximately 10 µg each of the *in vitro* RNA transcripts from the transcription mixture of 40 µl mixing with carborundum were mechanically inoculated onto both cotyledons (20 µl each) of each pumpkin seedling using a sterilized glass spatula. Healthy seedlings were mock-inoculated with 40 µl transcription buffer and with 10 µg of total RNA extracted and purified from healthy pumpkin leaves, as negative controls. Positive controls in this assay were performed by mechanical inoculation of healthy pumpkin seedlings with 10 µg of total RNA extracted and purified from PRSV-W-infected pumpkin leaves obtained from experimental inoculation. All inoculated plant seedlings were kept in a temperature-controlled (24-28 ^{o}C) greenhouse for observation of symptom development. Further verifications of infectivity of the *in vitro* transcripts were performed using ELISA and observed the presence of morphological *Potyvirus* particles and the absence any other virus particles under TEM.

17. Enzyme-linked immunosorbent assay (ELISA)

The indirect ELISA using plate-trapped antigen technique was performed as described by Clark and Adams (1977) with modifications to detect systemic infection of PRSV or to detect the presence of PRSV-W in plant inoculated with its *in vitro* synthesized construct (infectivity assay). A rabbit polyclonal antibody against capsid protein (CP) of PRSV was kindly provided by the Plant Genetic Engineering Unit, BIOTEC at at Kasetsart University, Kamphaengsaen campus, Nakhon Pathom. The antibody was produced by immunizing rabbit with a gel-purified *E. coli* expression product consisting amino acid sequence of CP of PRSV-P (papaya-infecting) and used as primary antibody in ELISA.

One hundred mg leaf tissues of inoculated and tip leaves from test and control plants were ground in 1 ml of coating buffer (50 mM sodium carbonate buffer, pH9.6). The resultant sap for each sample, 100 µl per well, was added to well on a sterile polystyrene flat bottom 96-well microtiter plate (Polysorp, Nunc). The plate was incubated overnight at 4 ^{o}C for trapping the plate with the target antigen. After the plate was washed three times with phosphate-buffered saline with Tween 20 (PBST containing 10 mM sodium-potassium phosphate, 500 mM sodium chloride at pH 7.4 and 5 ml/l of Tween-20), the rabbit polyclonal antibody against PRSV CP in PBST plus 0.5 % (w/v) bovine serum albumin (BSA) at a dilution of 1: 1000, 100 µl per well, was added and incubated for 2 hr at 37 ^{o}C. After the plate was washed three times with PBST, alkaline phosphatase conjugated goat anti-rabbit immunoglobulin (IgG) in PBST plus 0.5% (w/v) bovine serum albumin (BSA) at a dilution of 1: 10000, 100 µl per well, was added and incubated for 1 hr at 37 ^{o}C. After the plate was washed three times with PBST, alkaline phosphatase substrate solution (0.5 mg/ml ρ-nitrophenyl phosphate (PNPP) in 10% (v/v) diethanolamine, pH 9.8) was added (100 µl per well), and the plate was incubated in the dark at room temperature. Absorbance ($A_{405\ nm}$) values were measured in a Multiskan EX multiscan microplate reader (LABSYSTEMS).

18. Electron microscopy

Electron microscopy (EM) was performed to confirm of the identification of the virus source used in this research and of systemic infection of *in vitro* synthesized-construct of PRSV-W. With morphology of potyvirus particles, typical flexuous, filamentous particles of 780 x 12 nm, presence of *Potyvirus* virions and the absence of any other virus particles were observed using crude sap prepared from the inoculated-plants as sample. Under transmission EM, negative staining technique was performed using uranyl acetate solution. Preparations were examined using a JEM-1230 transmission electron microscope (JEOL) accelerating voltages from 40-120 kV at the Central Biochemistry Laboratory of the Central Laboratory and Greenhouse Complex (CLGC), Kasetsart University, Kamphaengsaen campus, Nakhon Pathom.

EM was performed on 200-mesh formvar-coated copper electron microscope grids. Crude sap preparations were prepared by grinding leaf tissue of 100 mg of inoculated and young leaves from test plants in 200 µl of PBS. The grids were incubated for 30 min at room temperature on 20 µl droplets of crude sap prepared from young leaf tissue of inoculated test plants. After this period, the grids were blotted onto sterile filter paper (Whatman, 3MM) to remove excess solution and then washed three times 1 min on PBS droplets. The grids were then negatively stained with 2% uranyl acetate for 20 min at room temperature. After this period, excess staining solution was removed by blotting with sterile filter paper and then the grids were washed three times with PBS. The grids were blotted using sterile filter paper and air-dried for 20 min at room temperature. The grids were visualized immediately using a JEM-1230 transmission electron microscope (JEOL).

RESULTS AND DISCUSSION

1. Experimental verification of some characteristics of the PRSV-W isolate studying in this research

The isolate of PRSV-W used as virus source in this research was originally isolated from a naturally infected cucumber (*Cucumis sativus* L.) plant growing in Nakhon Sawan province, central region of Thailand. It was identified and transferred by aphid transmission to cucumber seedlings. The virus isolate cultured in cucumber plants under greenhouse condition was collected and five systemically infected cucumber (*C. sativus* L) leaves were kindly donated by Dr. Wanpen Srithongchai, Department of Agriculture (DOA), Bangkok, Thailand. The leaves of infected cucumber (*C. sativus* L) plant with systemic mosaic symptoms were used as an original source of virus isolate in this study.

Initially, the virus was propagated and cultured in greenhouse-grown pumpkin (*Cucurbita moschata*) plants by mechanical inoculation. Pumpkin (*C. moschata*), which was used for virus propagation, showed systemic mosaic symptoms at young leaves 10-14 days after mechanical inoculation. Younger leaves from systemically infected pumpkin plant were used for virus inoculum preparation. Mechanical inoculation of pumpkin seedlings was conducted at cotyledons 5-7 days after germination and inoculated plant seedlings were kept in a greenhouse at Department of Plant Pathology, Kasetsart University, Kamphaengsaen campus, Nakhon Pathom. The mosaic symptom was observed as early as 10 days after mechanical inoculation. The cultured virus isolate was used as virus source for verification of some characteristics to confirm the taxonomic status of the virus in the following preliminary experiments. At about 15-20 days following inoculation, younger leave tissues from systemically infected pumpkin plants were used for inoculum preparation used in infectivity assay (section 1.1), for total RNA extraction from infected plant used as viral template source in RT-PCR and molecular cloning (section 1.2) and for visualization of the virus particles under transmission electron microscope (section 1.3).

1.1 Identification of virus pathotype

As a first step to identify the PRSV isolate used in this research, pathotype of the virus isolate was determined some characterizations of PRSV, a potyvirus, with different host specificity compared with the previously characterized virus isolates of this species.

Under insect-proof and greenhouse conditions, papaya (*Carica papaya*) and two cucurbit plants including pumpkin (*C. moschata*) and zucchini squash (*Cucurbita pepo*) were used as test plants to classify the PRSV isolate into pathotype W or P by mechanical inoculation. Mechanically inoculated pumpkin (Figure 3B) and zucchini squash (Figure 4B) test plants showed systemic mosaic symptoms at younger leaves 10-14 days after mechanical inoculation. Thirty days after inoculation, the infected pumpkin (Figure 3C) and zucchini squash (Figure 4C) test plants were also deformed and stunted. While the inoculated papaya test plants did not develop any symptom development when were observed over a period of 8 weeks after inoculation (data not shown) under the same conditions performing with the two cucurbit species test plants. Four weeks after inoculation, ELISA was performed to detect the presence of PRSV in all test plants and the results showed positive reaction for PRSV with only the inoculated cucurbit (pumpkin and zucchini) test plants but not from the inoculated papaya test plants (data not shown). These ELISA results for PRSV confirmed symptom expressions of virus infectivity in all test plants inoculated. Because of these results, the virus isolate used in this study which was positive for PRSV (ELISA) and was able to systematically infect pumpkin and zucchini squash test plants but not papaya, was classified as pathotype W of PRSV.

1.2 Confirmation of virus species by sequence analysis

In order to confirm the species of the virus isolate used as source in all of this research, viral sequence analysis was performed by searching and comparing with all nucleotide and deduced amino acid sequences deposited in GenBank database

Figure 3 Pumpkin (*Cucurbita moschata*) test plants grown under insect-proof and greenhouse conditions at 30 days after mechanical inoculation. Healthy leaf (A), infected leaf showing mosaic symptoms (B), comparison between healthy (left) and infected (right) plants in side view (C).

Figure 4 Zucchini squash (*Cucurbita pepo*) test plants grown under insect-proof and greenhouse conditions at 30 days after mechanical inoculation. Healthy plant (A), infected plant showing mosaic symptoms (B), comparison between healthy (left) and infected (right) plants in top view (C).

using BLAST available program from http://www.ncbi.nlm.nih.gov/BLAST on the NCBI Network server at http://www.ncbi.nlm.nih.gov/Genbank/index.html.

RT-PCR products of approximately expected 1-kb DNA fragment were amplified *in vitro* with a pair of primers (2SC5 and 2SC4) which their sequences are annealable to the PRSV 5'-NIb gene region by performing on total RNA extracted from systemically infected pumpkin leaf tissues. After DNA purification, the approximately expected 1-kb DNA fragments (RT-PCR products) were cloned into T-tailed pGEM®-T Easy vectors (Promega) (Appendix A Figure 2) before transformation into *E. coli* strain DH5α. After colony screening by PCR using with the universal primers M13 forward (-20) and M13 reverse, the selected clones, namely pPC5NIb, which can be generated an approximately expected 1-kb DNA fragment with the pair of universal M13 primers were further analyzed by digestion with restriction endonucleases *EcoR*I, *Nco*I and *Sal*I (Figure 6).

The RT-PCR insert fragment was sequenced and assembled. The nucleotide and deduced amino acid sequences were analyzed and compared with sequences deposited in NCBI database using BLAST program. Sequence analysis has shown that the insert fragment consists of 1081 nucleotides including two sequences of annealing primers of 16- and 21-nucleotides at the 5'- and 3'-end of the insert sequence, respectively (Figure 5B). The sequence of 1044 nucleotides excluding the two primer sequences at 5'- and 3'-end was predicted to potentially encode a polypeptide of 348 amino acids. BLAST analysis showed that the 1044-nucleotides sequence correspond to the two third region of NIb gene of PRSV isolates. The deduced amino acid sequence not only shared similarity with the two third N-terminal regions of PRSV NIb proteins but also contained a consensus amino acid sequence, VFEQ/S, of NIa-Pro cleavage at site of NIa-Pro/NIb. By BLAST search analysis, the sequence matched most closely with the two third at 5'terminal regions of NIb gene sequences of PRSV species and also gave highest scores of top ten identities with the corresponding region of PRSV isolates. The deduced 348-amino acids sequence corresponding to the two third region of PRSV NIb protein has higher than 94% amino acid sequence identity when compared with the corresponding region of other

PRSV isolates, indicating that the virus isolate might be a different strain of the species PRSV.

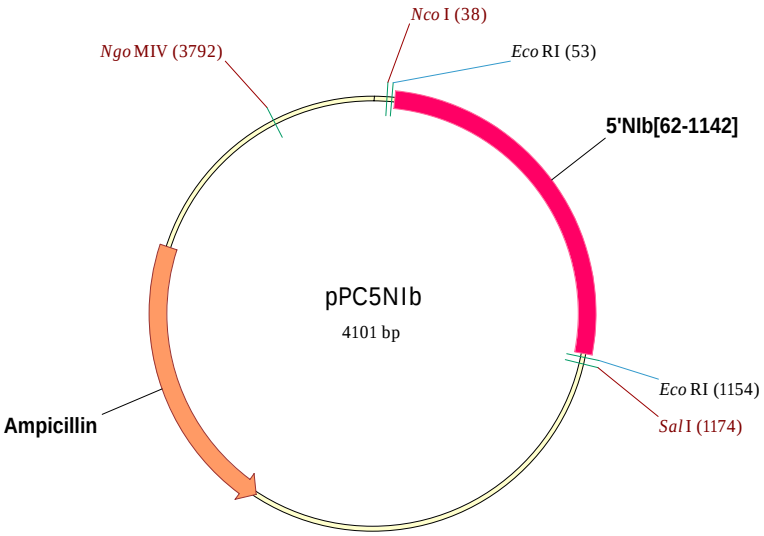
The PCR has been demonstrated to be a very useful tool for virus detection and diagnosis. It is more sensitive than direct probing or serological techniques. This method can be used straightly to generate a DNA product of predicted size that can be verified by visualization using gel electrophoresis under UV exposing. In cases of RNA viruses, RT-PCR using specific primers enables small numbers of viral RNA molecules to be detected in nucleic acid mixtures. In this study, the pair of primers (2SC5 and 2SC4) used in RT-PCR which their sequences are specific to NIb gene region of PRSV were able to recognize and prime for amplification *in vitro* of an approximately expected 1-kb fragment. The nucleotide and deduced amino acid sequences of the DNA fragment generated in this study were shown to be appropriate for molecular genetic information in virus identification in species level apart from CP and 3'-NCR sequences of potyviruses described previously by Shukla *et al*. (1994). Furthermore, the use of RT-PCR with a set of degenerate primers for identification have been reported which it is able to specifically amplify DNA fragments from all species of the genus *Potyvirus* (Langeveld *et al*., 1991; Gibbs and Mackenzie, 1997)

1.3 Visualization of the virus particles using transmission electron microscopy (TEM)

Crude sap prepared from younger leaf tissues of the mechanically inoculated and systemic mosaic symptom-expressing pumpkin plant was negatively stained with 2% uranyl acetate and examined in a JEOL-1230 transmission electron microscope.

Electron microscopy of negatively stained sap preparations from these leaves showed that they contained small numbers of flexuous, filamentous particles measuring of approximately 750 to 800 nm in length and the negatively stained virus

A



B

ATGTATTTGA GCAGAGTGGA AGTCGATGGC TCTTTGACAA GCTGCATGGC AACCTAAAGG GAGTAAGCTC
AGCTCCCAGC AATTTAGTCA CGAAGCACGT TGTGAAAGGA ATCTGTCCTC TCTTTAGGAG TTACCTTGAG
TGTGATGAAG AAGCCAAGGC TTTCTTTAGT CCGCTGATGG GACACTACAT GAAGAGCGTT CTAAGCAAGG
AGGCATACGT GAAGGATTTA TTGAAATATT CGAGTGACAT AATTGTTGGC GAAGTTGACC ACGATGTGTT
TGAGGATAGC GTCGCACAAG TTATCGAGTT GCTAAATGAT CACGAGTGTC CTGAGCTCGA ATACGTCACA
GACAGCGAGG TCATCATACA AGCTCTAAAC ATGGATGCAG CGGTCGGAGC TTTGTACACA GGAAAGAAGA
GAAAATACTT CGAAGGATCA ACAGTTGAGC ACAGACAGGC CCTAGTTCGG AAAAGTTGCG AGCGTCTTTA
TGAAGGGCGG ATGGGTGTTT GGAATGGATC GTTGAAGGCC GAGCTGCGAC CTGCAGAGAA AGTTCTAGCC
AAAAAGACAA GATCTTTCAC AGCAGCGCCT CTCGACACTT TGTTGGGAGC TAAAGTTTGC GTTGATGATT
TCAACAATTG GTTCTACAGC AAAAATATGG AGTGTCCCTG GACTGTTGGC ATGACGAAGT TTTACAAAGG
CTGGGACGAA TTTTTGAGGA AATTCCCTGA TAGTTGGGTT TACTGCGATG CAGATGGATC TCAGTTCGAT
AGTTCTTTGA CACCATATTT GCTTAATGCT GTGTTGTCTA TCCGACTGTG GGCTATGGAG GATTGGGACA
TTGGAGCTCA GATGCTTAAG AACTTGTATG GAGAAATCAC TTACACACCT ATTTTGACAC CTGATGGGAC
GATTGTCAAG AAGTTTAAAG GAAACAACAG TGGCCAACCC TCAACAGTTG TCGATAATAC ATTGATGGTT
TTAATCACAA TGTATTACGC ACTACGCAAG GCTGGTTACG ACACGGAGGC CCAGGAAGAT ATGTGCGTAT
TCTATATCAA TGGTGATGAT CTCTGTATCG C

Figure 5 Map of pPC5NIb (A) and nucleotide sequence of 5'-NIb fragment (B). The pPC5NIb plasmid of 4101 base pairs in closed circular form was obtained from inserting the DNA fragment of about 1 kb amplified by RT-PCR with the pair of primers 2SC5 and 2SC4 into T-tailed pGEM®-T Easy vectors (Promega). The insert fragment '5'NIb' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses. Double strand DNA of the vector is depicted by double line, and the antibiotic (ampicillin) coding region on the cloning vector is represented by a shaded arrowhead. Primer annealing regions are underlined.

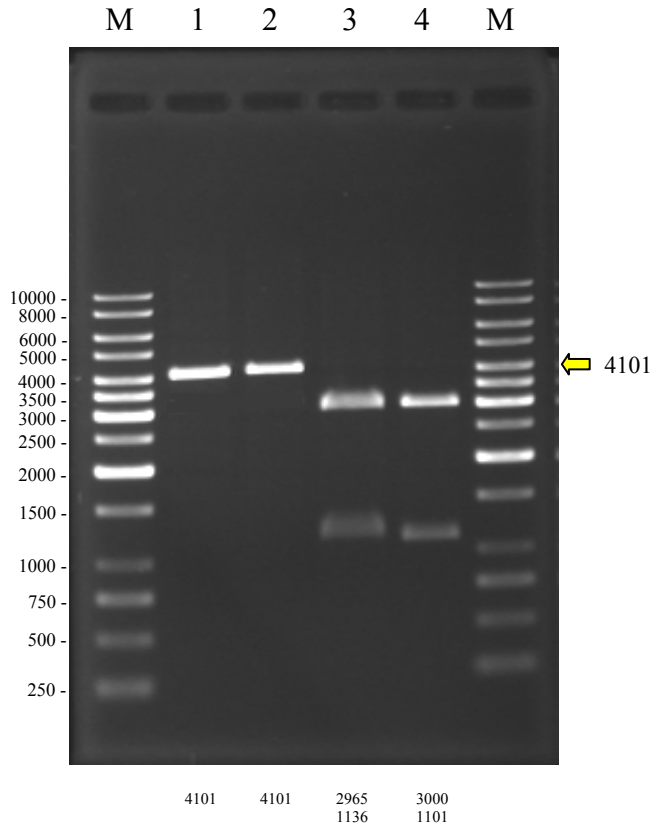


Figure 6 Restriction analyses of the pPC5NIb plasmid. The pPC5NIb plasmid of 4101 base pairs in closed circular form obtained from inserting the DNA fragment of about 1 kb amplified by RT-PCR with the pair of primers 2SC5 and 2SC4 into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPC5NIb plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPC5NIb digested with *Nco*I
Lane 2: pPC5NIb digested with *Sal*I
Lane 3: pPC5NIb digested with *Nco*I and *Sal*I
Lane 4: pPC5NIb digested with *Eco*RI

particles were photographed at a magnification of 25000 (Figure 7). This analysis revealed the morphology of virus particles typical for *Potyvirus* and confirmed the absence of any other virus particles in the virus source cultured and used in this research.

As expected the morphology of the virus isolate is comparable to other potyviruses which have typical flexuous, filamentous particles of approximately 680-900 nm in length (Shukla *et al*., 1994; Hull, 2002) and to other isolate of PRSV particles of 780 nm in length published for type P (Gonsalves and Ishii, 1980).

2. Viral culture

Biological features as host range specificity and symptom expressions on test plants, sequence similarity and electron microscopic analysis all indicated that the virus is an isolate of PRSV pathotype Watermelon (PRSV-W).

After verification of the taxonomic status, the classified PRSV-W isolate was propagated and maintained under insect-proof and greenhouse conditions (temperature ranging between 24 ^{o}C and 28 ^{o}C) on pumpkin (*C. moschata*) plants, and was transmitted experimentally by mechanical inoculation with crude sap every 8-10 weeks. The virus isolate was used as source for all experiments in this research.

3. Primary structure of PRSV-W genome

3.1 Cloning and sequencing

The complete nucleotide sequence of the genome of a Thai isolate of *Papaya ringspot virus* type W (PRSV-W) was determined from eight overlapping viral cDNA clones generated by RT-PCR using degenerate and/or specific primers as shown in Table 2. Eight overlapping fragments representing the complete genomic RNA of the PRSV-W excluding a poly(A) tail were amplified *in vitro* by RT-PCR and the RT-PCR products generated from different reactions were cloned into U-

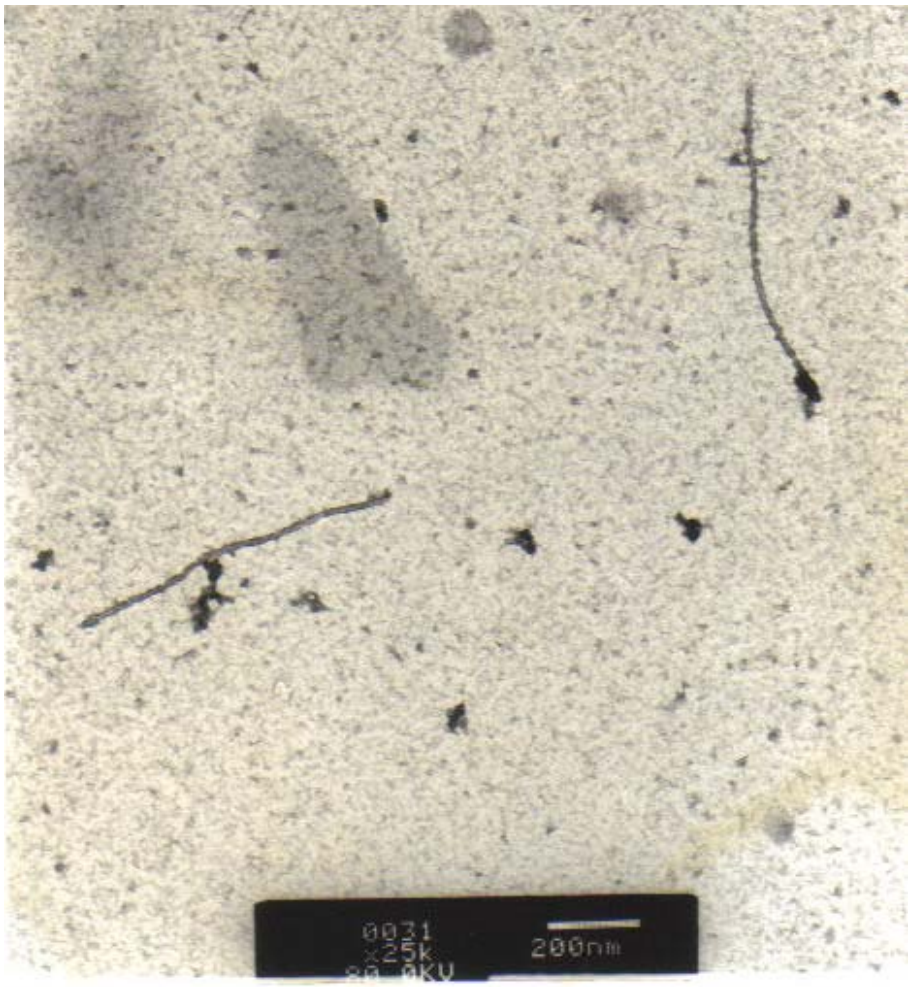


Figure 7 Negatively stained filamentous particles approximately 750 to 800 nm in length observed in the crude sap of systemically infected leaf of greenhouse-grown pumpkin plant mechanically inoculated with the PRSV-W isolate.

tailed vector (pDrive cloning vector from QIAGEN) as shown in Appendix A Figure 1. The eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments of approximately 1 to 2.2 kbp corresponding to the genomic RNA of the virus isolate (Figure 8) were named pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH in the order in which inserted fragments correspond on the virus genome as shown in Figure 1.

The recombinant plasmid DNA from these clones was used as template for automatic sequencing. Both strands of the overlapping viral fragment cDNA inserting in each different clone were sequenced in both directions to elude sequence ambiguities with universal pUC/M13 forward and reverse primers (vector primers), and internal sequencing primers designed based on the partial nucleotide sequence obtained by primer walking strategy. Sequence and position on the viral genome of all primers used in DNA sequencing are listed in Table 3. DNA sequencing for the virus genome was performed by the modified dideoxy chain termination method using the ABI PRISM® Bigdye™ Terminator Cycle Sequencing Ready Reaction Kits version 2.0 and the reactions obtained were run on either the model 310 or model 377 automatic DNA sequencers (Applied Biosystems).

Sequence analyses of the eight cDNA fragments including wA, wB, wC, wD. wE, wF, wG and wH which were cloned separately in eight different recombinant plasmids illustrated that the eight insert cDNA fragments including primer annealing regions are 963, 2238, 1973, 1791, 1238, 1527, 1540 and 1176 base pairs in length, respectively. DNA sequences of these insert fragments are shown separately in Figure 9B for wA fragment of 963 nucleotides, in Figure 10B for wB fragment of 2238 nucleotides, in Figure 11B for wA fragment of 1973 nucleotides, in Figure 12B for wD fragment of 1791 nucleotides, in Figure 13B for wE fragment of 1238 nucleotides, in Figure 14B for wF fragment of 1527 nucleotides, in Figure 15B for wG fragment of 1540 nucleotides and in Figure 16B for wH fragment of 1176 nucleotides.

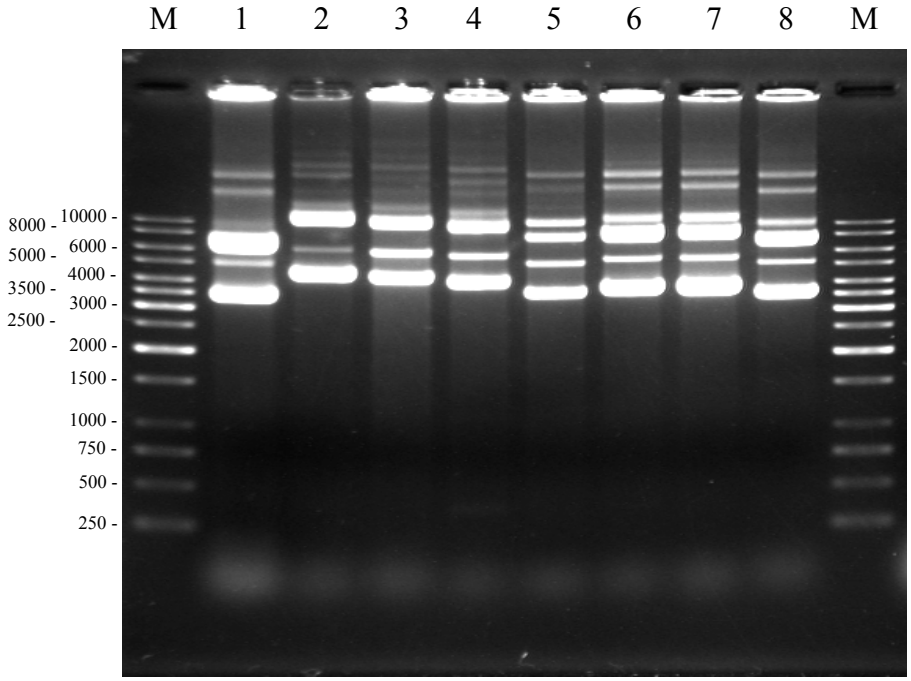
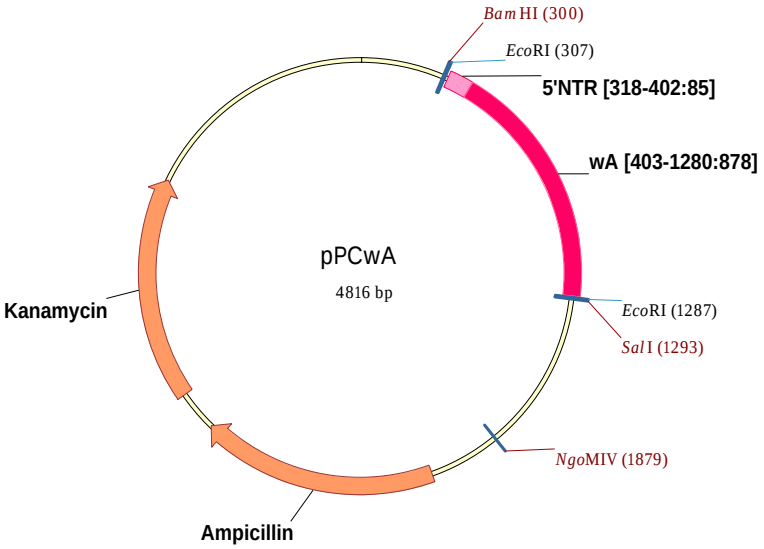


Figure 8 Eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate. These different plasmids (closed circular form) were cloned separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN). The preparations of plasmids were determined by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure.

Lane 1: pPCwA
Lane 2: pPCwB
Lane 3: pPCwC
Lane 4: pPCwD
Lane 5: pPCwE
Lane 6: pPCwF
Lane 7: pPCwG
Lane 8: pPCwH
Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)

A

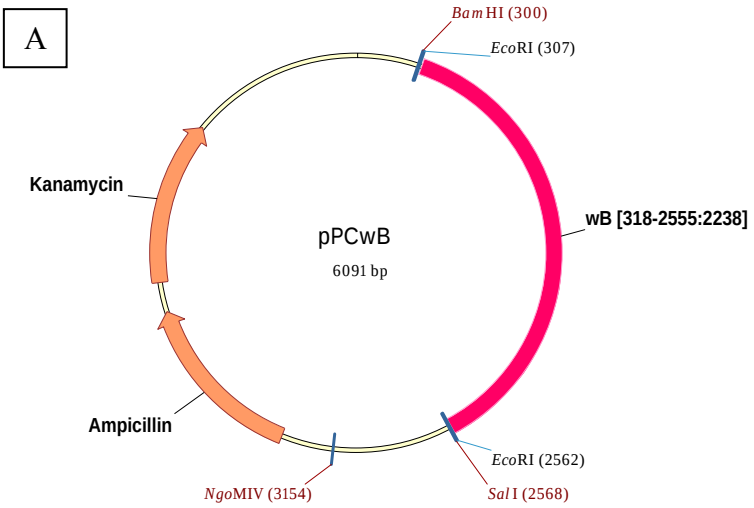


B

```
AAATAAAACA TCTCAACACA ACACAATTCA AAGCATTCAA ACACACTCAA GCAAATTTTA TTCTCAAATT
TCACAATCTG CAAGCATGTC TTCGTTGTAC CAATTGCAAC CAATTGCACT GAAGGACCGT CTCTTGTCTC
ACAAGAGAGG AAAAGGTTGG ATTGAGCACA AGCTCGAAAG GAAAGGAGAG AGAGGAAACA CTCGTCATGT
TGGCGAGTTT GTGATCAGTG AGGGAGCAAA AATTTTGCAA CTCGTGCAAA TTGGAAACGC TGAAATTGGC
AGGACCTTCC TGGAAGGTGA TAGGAGAACT CAAGCTGACA TTTTTGAAAT TATCAAAAAG ACGATGGTCG
GCCACCTGGG GTACGACTTC GAATGTGGGT TATGGTGCTG TTACTCTTGT GACAACACAT CTGACAAGTA
CTTCAAGAAG TGTGATTGCG GTGAGAAGTA CTATTATTCA GAGAAGAATT TGATTAAGTC AATGCATGAC
TTGATGTACC AGTTTGACAT GACAGCCGCT GAAATTGACC AAGTAGGTCT TGACTACTTG GCTGAAGCTG
TAGACTATGC TGAACAGTCC GTCAAGAAAT CAAAGGTACT GGTACCTGAT GAACCTAAGT TTGTGGAGAT
CCTCGCGGCG AGTGAGGAGA GTCTTCTGGT GGTTCCTGAG CCAGAAGTTG CTTCTGTAAC TACAAGAGCT
GAAGAAGCAT GGACGATACA AATTGGGGAA ATTCCTGTCC CACTTGTCGT CATAAGGGAG ACGCCAGTAA
TCAGTGGTGT GAATGGAATG CTGAGCTCAA CTGGTTTTTC ACTTGAAGCG GAAATTGCAA GACCGGCTAA
AAGTACGATT CCTCAAGATG AAGTGGAGGA AGCAGTGCAC CTGGCACTCG AGGTCGGCAA TGAAATCGCC
GAAAAGAAGC CTGAACTCAA ATTGGCACCA TACTGGAGTG CTAGCCTTGA GTT
```

Figure 9 Map of pPCwA (A) and nucleotide sequence of wA cDNA fragment (B). The pPCwA plasmid of 4816 base pairs in closed circular form was obtained from inserting the DNA fragment 'wA' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A



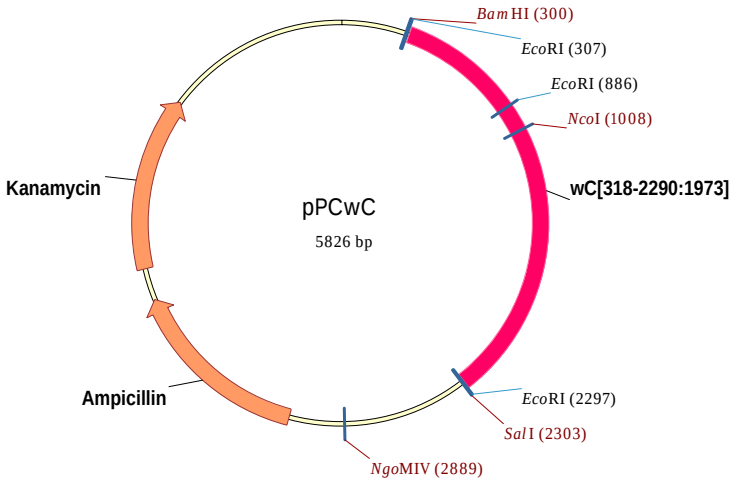
B

```
GAAGCAGTGC ACCTGGCACT
CGAGGTCGGC AATGAAATCG
CCGAAAAGAA GCCTGAACTC
AAATTGGCAC CATACTGGAG
TGCTAGCCTT GAGTTGCACA
AGAGAGTCCG CAAACACAAG
GAACACGCTA AATCTGAAGC
ACTTCGAGTT CGAAAAGAGA
AGGAAAGAGA TCATAGGATT
TTTGCTGCTT TAGAAGCAAA
GCTGAATTTG AAAGCAAGGA
GGCAAGGCCA AGTTGTTGTT
TGCGACAAGA GAGGTACTTT
AAAATGGAAA AAGCATCAAC
AGCGGAAGAG GAACAAAATG
GTTACGCAAT TAAGTGACTC
```

```
TGTCGTCACG AAAATTCACG CCAATTTTGA GTGCAGGACT CCTAATTTTG ATGTTGAGAC TCCTGGCATT
AAGTGTGCAA CCTCGAAGGT GACGCGGAAG AAACAAACGC AACCTAAGAT ATTTGGCAGC AACAAGGTCA
ATTATGTCAT GAAAAATCTG TGTGACATAA TCATCGACAG GAACATTCCT GTTGAGTTAA TCACGAAGCG
CTGCAAAAGG AGAATTTTCC GTAGGGACGG GAAAAACTAT GTGCACTTAA GACACATGGA TGGCAACAAT
GCACCACGGG ATGTAAGCAG CTCGTCAGAC ATGGAGAAGT TGTTTACGCG CTTTTGCAAA TTTTTAATCA
GGAAGCAGTC GATCAACGCA GCGAATCTAA CACATGGTTC GAGTGGTCTC ATCTTTAAAC CGAAGTTCGC
TGACCGAACA GGGCGATATT TTGGTGAATA TTTTATCACG CGTGGACGAT GTGAAGGAAA GCTATTTGAT
GGGAGATCAA AACTAGCTAA GTCGGTTCGC ATGCGCATGG AACAATACAA TGATGTCGCT GAGAAGTTCT
GGCTCGGTTT TAACAGGGCT TTCTTACGGC ACAGGAAGCC GACGGATCAT GTGTGCACGT CTGATATGGA
TGTTACCATG TGTGGTGAAG TGGCGGCTCT TGCAACCATA ATCCTGTTTC CATGTCATAA GATAACTTGT
AACGCTTGCA TGAACAAGGT AAAGGGAAGA GTCATTGACG AAGTCGGTGA GGATCTGAAT TGTGAGCTCG
AACGTTTGCG CGAAACCCTC TCGTCGTATG GAGGCTCATT TGGTCATGTT TCAACATTGC TTGATCAATT
AAACAGAGTT TTAAATGCGC GTAACATGAA TGATGGAGCT TTTAAGGAGA TCGCGAAGAA AATTGATGAG
AAGAAAGAAA GTCCTTGGAT TCATATGACG GTCATCAACA ATACGCTTAT TAAAGGTTCA TTAGCAACTG
GCTATGAATT TGAGAGAGCG TCTGATAGTT TACGAGAAAT TGTAAGATGG CATCTCAAGA GAACAGAGTC
AATCAAAGCT GGCAGTGTTG AGAGCTTCAG AAACAAACGT TCTGGAAAAG CTCATTTCAA CCCGGCCCTC
ACATGTGACA ATCAATTGGA CAGAAATGGC AATTTCTTAT GGGGTGAGAG ACAATATCAT GCCAAAAGAT
TCTTTGCTAA TTACTTTGAG AAGATTGATC ACAGTAAGGG TTATGAGTAC TACAGTCAAC GCCAAAACCC
GAATGGAATT AGAAAGATTG CCATTAGTAA TCTAATATTC TCCACAAATT TGGAGAGGTT TCGGCAGCAA
ATGGTTGAAC ATCACATTGA CCAGGGACCA ATCACTCGTG AGTGTATCGC ACTGCGCAAT AACAATTATG
TTCATGTATG TAGCTGCGTG ACCTTAGATG ATGGAACTCC AGCGACGAGC GAATTGAAAA CTCCCACCAA
GAATCACATT GTTCTTGGTA ATTCTGGTGA TCCTAAGTAT GTCGATTTGC CGACTCTTGA GTCTGACTCA
ATGTATATAG CTAAGAGAGG CTATTGCTAC ATGAACATCT TCCTGGCGAT GCTCATAAAC ATACCTGAGA
ATGAAGCGAA GGACTTTACA AAGAGAGTTC GTGATCTTGT AGGTTCAAAA CTTGGTGAGT GGCCAACGAT
GTTAGATGTC GCAACATGCG CCAACCAACT GGTTATCTTT CATCCTGATG CAGCCAATGC GGAATTGCCA
CGAATCTTAG TGGATCATCG ACAGAAAACA ATGCATGTTA TTGATTCGTT TGGACCTGTA GACTCTGGAT
ATCACATATT GAAGGCGAAC ACAGTGAATC AGTTGATTCA GTTTGCCAGA GAGCCACTTG ATAGTGAGAT
GAAGCACTAC ATTGTTGGTG GAGAGTTT
```

Figure 10 Map of pPCwB (A) and nucleotide sequence of wB cDNA fragment (B). The pPCwB plasmid of 6091 base pairs in closed circular form was obtained from inserting the DNA fragment 'wB' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A

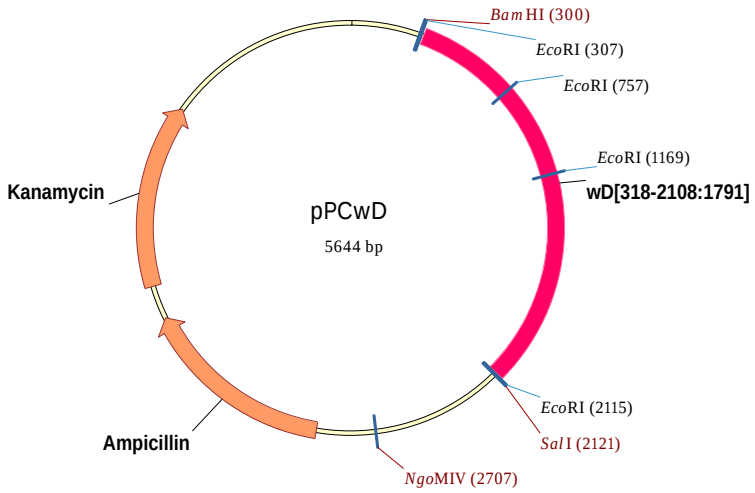


B

```
CTCCCACCAA GAATCACATT GTTCTTGGTA ATTCTGGTGA TCCTAAGTAT GTCGATTTGC CGACTCTTGA
GTCTGACTCA ATGTATATAG CTAAGAGAGG CTATTGCTAC ATGAACATCT TCCTGGCGAT GCTCATAAAC
ATACCTGAGA ATGAAGCGAA GGACTTTACA AAGAGAGTTC GTGATCTTGT AGGTTCAAAA CTTGGTGAGT
GGCCAACGAT GTTAGATGTC GCAACATGCG CCAACCAACT GGTTATCTTT CATCCTGATG CAGCCAATGC
GGAATTGCCA CGAATCTTAG TGGATCATCG ACAGAAAACA ATGCATGTTA TTGATTCGTT TGGACCTGTA
GACTCTGGAT ATCACATATT GAAGGCGAAC ACAGTGAATC AGTTGATTCA GTTTGCCAGA GAGCCACTTG
ATAGTGAGAT GAAGCACTAC ATTGTTGGTG GAGAGTTTGA TCCAACTACT AGTTGCTTGC ATCAGTTGAT
ACGCGTCATC TATAAACCCC ATGAACTCCG CAATCTGCTC AGAAATGAAC CATATCTAAT AGTGATTGCA
TTGATGTCGC CAAGTGTACT TTTGACTTTG TTTAATAGTG GCGCAATTGA GCATGCGTTG AATTATTGGA
TCAAAAGGGA CCAAGATGTT GTTGAGGTTA TTGTCTTGGT GGAGCAATTG TGTAGGAAAG TGACGCTTGC
TAGAACAATT TTGGAGCAAT TCAATGAGAT TCGCCAAAAT GCACGAGATA TACACGAGCT AATGGATCGA
AATAATAAAC CTTGGATTTC ATACGATCGA TCATTGGAGC TACTAAGTGT GTATGCGGAT TCGCAATTAA
CGGATGAGGG CCTGCTCAAG CAAGGTTTTT CAACACTGGA CCCTAGACTG CGCGAAGCTG TTGAAAAAAC
CTACGCCGCT CTTTTGCAGG AAGAGTGGCG TGCTTTAAGT TTGTTTCAAA AGTTGCACTT AAGGTACTTT
GCATTCAAAT CACAACCATC TTTTTCCGAG TGTTTAAAGC CGAAAGGGCG CGCAGATTTG AAAATTGTGA
ACGACTTCTC ACCGAAGTAC TGTGTACACG AGGTCGGAAA GGCGTTGCTC CAACCAATTA AAGCTGGAGC
TGAAATCACA TCACGTCTCC TAAACGGTTG TGGAACGTTT ATCCGGAAAA GTGCCGCAAG GGGATGTGCA
TACATTTTTA AGGACCTCTT CCAGTTTGTT CATGTGGTGT TGGTGTTAAG CATTTTGTTA CAGATATTCA
GGAGCGCACA AGGGATTGCC ATGGAGCATA TACAACTAAA GCAAGCAAAG GCGGAAGCGG TAAAGCAGAA
AGATTTTGAT CGGTTGGAAG CTTTATACGC GGAACTGTGC GTTAAGAATG GTGATCAGCC GACGACTGAA
GAATTCCTTG ATTTTGTGAT GGAGCGTGAA CCAAGGCTCA AGGATCAAGC TTATAGCTTG ATTCACATAC
CAGCCATTCA TCAAGCAAAA TCAGACAATG AAAAGAAACT TGGACAGGTC ATTGCATTCA TAACATTAAT
TTTAATGATG ATCGACGTGG ATAAGAGCGA CTGTGTCTAT CGAATCTTGA ACAAATTTAA AGGCGTGATA
AACTCTTGCA ACACAAATGT TTATCACCAG TCTTTAGATG ATATTCGAGA TTTCTATGAA GACAAGCAAT
TAACCATCGA TTTTGACATC ACTGGTGAGA ATCAGATCAA CAGAGGGCCT ATAGATGTCA CGTTCGAGAA
ATGGTGGGAC AATCAATTGT CCAACAACAA CACAATTGGC CATTACCGAA TTGGGGGAAC GTTCGTTGAA
TTTTCACGGA GCAACGCAGC CACTGTGGCT AGTGAAATAG CCCACAGTCC TGAACGCGAA TTTCTAGTTC
GCGGAGCTGT TGGTAGTGGT AAATCAACAA ACCTTCCTTT CTTACTTAGT AAGCATGGTA GTGTGTTGTT
AATAGAGCCC ACC
```

Figure 11 Map of pPCwC (A) and nucleotide sequence of wC cDNA fragment (B). The pPCwA plasmid of 5826 base pairs in closed circular form was obtained from inserting the DNA fragment 'wC' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A

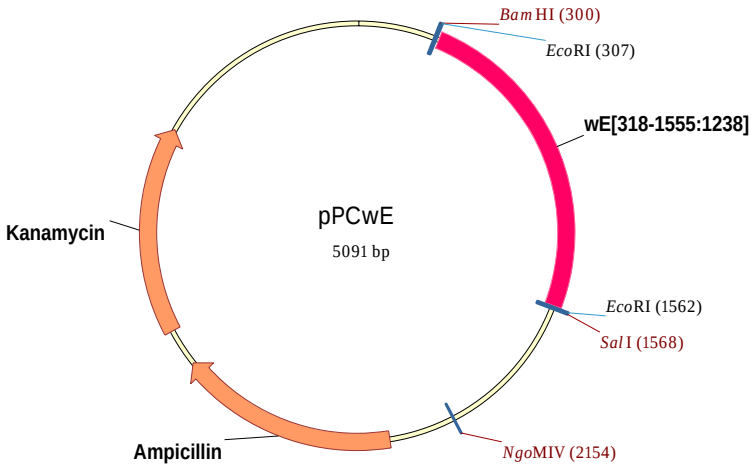


B

```
GCAGCCACTG TGGCTAGTGA AATAGCCCAC AGTCCTGAAC GCGAATTTCT AGTTCGCGGA GCTGTTGGTA
GTGGTAAATC AACAAACCTT CCTTTCTTAC TTAGTAAGCA TGGTAGTGTG TTGTTAATAG AGCCCACCAG
ACCCCTTTGC GAAAATGTTT GCAAGCAATT GCGCGGTGAG CCGTTTCATT GTAATCAAAC CATTCGCATG
CGTGGGTTAA CGGCCTTTGG TTCGACCAAT ATCACTATCA TGACGAGTGG ATTTGCTTTA CACTATTATG
CTCATAACCT TCAGCAATTA AGGCTCTTTG ATTTCATAAT TTTCGATGAA TGTCATGTTA TAGATAGTCA
AGCAATGGCT TTCTACTGCC TGATGGAAGG GAACGCGATT GAGAAGAAAG TTCTTAAAGT TTCTGCCACG
CCACCTGGGC GTGAGGTTGA ATTCTCAACG CAATTCCCAA CAAAAATTGT CACCGAGCGA TCCATAAGTT
TCAAACAGCT GGTTGATAAT TTCGGCACTG GCGCGAATAG CGATGTGACT GCCTTCGCAG ACAACATACT
GGTTTATGTA GCAAGCTATA ATGAAGTGGA TCAACTAAGT AAGCTTTTAT CCGATAAAGG CTACTTAGTT
ACTAAAATCG ATGGGCGGAC AATGAAGGCT GGAAAAACTG AAATTTCAAC TAGTGGCACA AAATCTAAGA
AGCATTTCAT AGTTGCCACT AACATTATTG AAAATGGCGT CACACTTGAC ATAGAAGCTG TGATTGATTT
TGGAATGAAA GTCGTGCCTG AAATGGACTC GGATAATCGC ATGATCAGAT ACTCGAAGCA AGCGATCAGC
TTTGGAGAGA GAATTCAACG GCTTGGCCGC GTTGGAAGAC ATAAAGAGGG AATTGCACTT AGAATCGGTC
ACACTGAGAA AGGCATTCAA GAAATCCCAG AAATGGCGGC TACTGAGGCG GCTTTCCTGA GCTTCACGTA
TGGCTTGCCT GTCATGACTC ATAATGTGGG ACTGAGTTTA CTCAAGAACT GCACCGTGAG GCAAGCACGT
ACAATGCAGC AGTATGAGTT AAGCCCATTT TTCACACAAA ATTTGGTTAA CTTCGATGGA ACAGTGCACC
CCAAGATCGA TGTGTTGTTG CGTCCTTACA AATTGAGGGA TTGTGAAGTC AGATTAAGCG AAGCAGCGAT
ACCGCATGGA GTACAATCTA TTTGGATGTC CGCCAGGGAG TATGAAGCTG TTGGTGGACG CCTTTGTCTG
GAAGGTGATG TTCGGATACC ATTCTTGATC AAGGATGTGC CTGAGCGATT GTACAAGGAA CTGTGGGACA
TTGTCCAAAC TTACAAGCGT GACTTCACAT TTGGGCGAAT TAATTCTGTA TCCGCTGGAA AAATTGCATA
CACATTAAGA ACTGATGTGT ATTCAATTCC CAGAACTCTC ATAACGATTG ATAAACTGAT CGAGAGTGAA
AACATGGAGC ACGCTCACTT TAAAGCCATG ACAAGTTGTA CTGGTCTTAA CTCTAGTTTC TCTCTCCTTG
GTGTCATAAA CACTATTCAG GGTAGATACT TGGTTGACCA TTCGGTTGAA AATATCAGAA AGCTTCAGCT
GGCGAAGGCC CAAATTCAAC AACTTGAAGC TCACGTGCAG GAAAACAATG TTGAAAATCT GATTCAATCT
CTTGGTGCAG TGAGAGCCGT TTATCATCAG AGTGTGGATG GAGTCAAACA CATAAAGCGA GAGTTGGGTT
TAAAAGGAGT CTGGGATGGC TCATTGATGA TTAAGGATGC A
```

Figure 12 Map of pPCwD (A) and nucleotide sequence of wD cDNA fragment (B). The pPCwD plasmid of 5644 base pairs in closed circular form was obtained from inserting the DNA fragment 'wD' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A

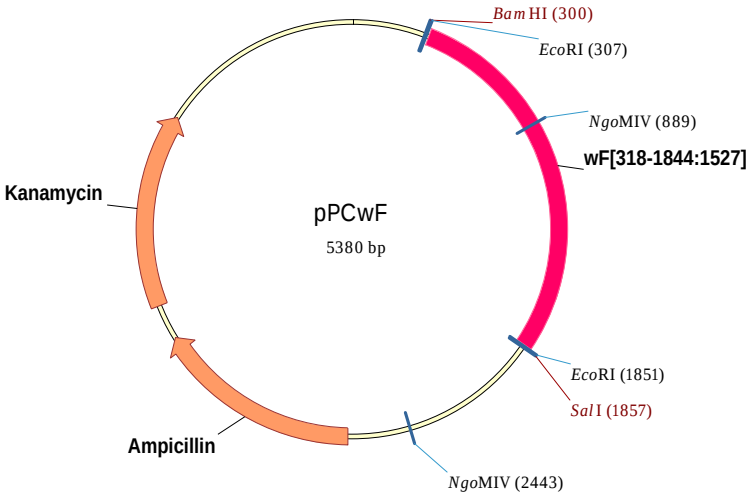


B

```
TGCAGGAAAA CAATGTTGAA AATCTGATTC AATCTCTTGG TGCAGTGAGA GCCGTTTATC ATCAGAGTGT
GGATGGAGTC AAACACATAA AGCGAGAGTT GGGTTTAAAA GGAGTCTGGG ATGGCTCATT GATGATTAAG
GATGCAATTG TATGCGGCTT TACAATGGCT GGTGGTGCGA TGCTATTGTA CCAACATTTT CGTGATAAGC
TCACCAATAT TCATGTGTTT CATCAAGGTT TCTCTGCGCG ACAGCGACAA AAGTTAAGAT TCAAGTCTGC
AGCAAATGCT AAACTTGGTC GAGAAGTTTA CGGGGATGAT GGAACAATTG AGCATTACTT CGGAGAGGCG
TACACGAAGA AGGGGAACAA GAAAGGAAGG ATGCACGGCA TGGGCGTTAA GACAAGGAAA TTCGTTGCTA
CATACGGATT TAAACCAGAG GATTATTCAT ATGTGCGTTA TCTGGACCCC TTAACAGGTG AGACATTAGA
TGAGAGTCCT CAGACTGACA TCTCAATGGT GCAAGAACAT TTTGGTGACA TTAGAAGTAA ATACATGGAC
TCGGACAGCT TCGATCGGCA GGCATTGATA GCAAATAACG TGATTAAAGC CTATTACGTC CGGAACTCCG
CCAAAACTGC ATTAGAAGTT GATTTAACAC CACATAACCC ACTCAAGGTT TGTGATACTA AGTTGACGAT
CGCAGGGTTT CCTGATAGAG AAGCTGAATT GAGGCAAACA GGCCCACCGA GAACAATCTT AGCTGACCAA
GTTCCACCAC CAACAAAGTC TGTTCATCAC GAAGGAAAGA GTCTCTGCCA AGGTATGAGA AATTACAATG
GCATAGCTTC TGTGGTTTGC CATTTAAAGA GCACATCAGG AGATGGTCGA AGCTTATTTG GGGTCGGTTA
CAACTCTTTC ATCATTACAA ACCGACATTT GTTCAAAGAG AATAATGGTG AGCTTATTGT GAAATCCCAA
CATGGAAAGT TTGTTGTTAA GAACACTTCA ACACTTCGGA TAGCTCCGAT TGGAAAGACT GATCTTCTTA
TCATTAGGAT GCCAAAGGGT TTTCCACCAT TCCATAGCAG GGCAAGGTTC AGAGCCATGA AAGCTGGTGA
CAAGGTTTGT ATGATTGGTG TAGATTACCA AGAAAACCAT ATCGCTAGTA AGGTGTCTGA GACCTCAATC
ATTAGTGAGG GCACGGGGGA ATTTGGTTGC CACTGGATAT CCACGAAT
```

Figure 13 Map of pPCwE (A) and nucleotide sequence of wE cDNA fragment (B). The pPCwE plasmid of 5091 base pairs in closed circular form was obtained from inserting the DNA fragment 'wE' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A

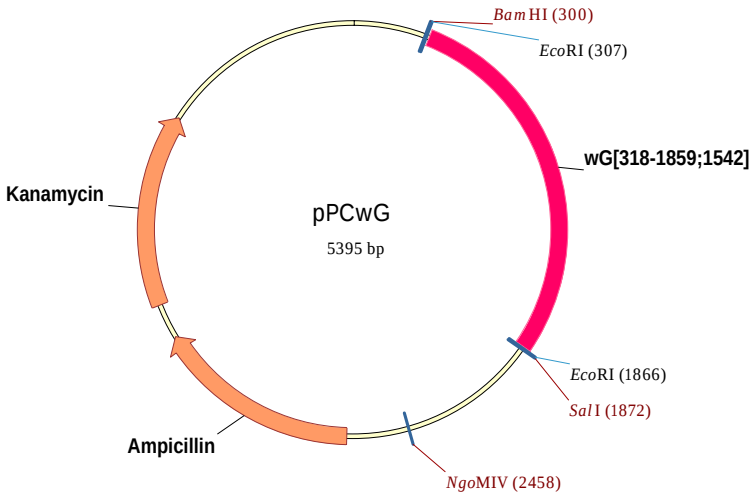


B

```
CTGACATCTC AATGGTGCAA GAACATTTTG GTGACATTAG AAGTAAATAC ATGGACTCGG ACAGCTTCGA
TCGGCAGGCA TTGATAGCAA ATAACGTGAT TAAAGCCTAT TACGTCCGGA ACTCCGCCAA AACTGCATTA
GAAGTTGATT TAACACCACA TAACCCACTC AAGGTTTGTG ATACTAAGTT GACGATCGCA GGGTTTCCTG
ATAGAGAAGC TGAATTGAGG CAAACAGGCC CACCGAGAAC AATCTTAGCT GACCAAGTTC CACCACCAAC
AAAGTCTGTT CATCACGAAG GAAAGAGTCT CTGCCAAGGT ATGAGAAATT ACAATGGCAT AGCTTCTGTG
GTTTGCCATT TAAAGAGCAC ATCAGGAGAT GGTCGAAGCT TATTTGGGGT CGGTTACAAC TCTTTCATCA
TTACAAACCG ACATTTGTTC AAAGAGAATA ATGGTGAGCT TATTGTGAAA TCCCAACATG GAAAGTTTGT
TGTTAAGAAC ACTTCAACAC TTCGGATAGC TCCGATTGGA AAGACTGATC TTCTTATCAT TAGGATGCCA
AAGGGTTTTC CACCATTCCA TAGCAGGGCA AGGTTCAGAG CCATGAAAGC TGGTGACAAG GTTTGTATGA
TTGGTGTAGA TTACCAAGAA AACCATATCG CTAGTAAGGT GTCTGAGACC TCAATCATTA GTGAGGGCAC
GGGGGAATTT GGTTGCCACT GGATATCCAC GAATGATGGT GACTGTGGTA ATCCACTGGT CAGCGTCTCC
GATGGCTTCA TAGTGGGACT GCACAGTCTA TCGACCTCGA CTGGTGATCA AAATTTCTTT GCAAAAATAC
CAGCATTCTT TGAGGAGAAG GTTCTCAGGC GGATTGATGA CCTGACATGG AGCAAACACT GGAGCTACAA
TGTTAATGAA TTGAGCTGGG GGGCTTTGAA AGTGTGGGAG AGCCGGCCAG AGGCTATTTT TAACGCGCAG
AAGGAAATCG ATCAATTAAA TGTATTTGAG CAGAGTGGAA GTCGATGGCT CTTTGACAAG CTGCATGGCA
ACCTAAAGGG AGTAAGCTCA GCTCCCAGCA ATTTAGTCAC GAAGCACGTT GTGAAAGGAA TCTGTCCTCT
CTTTAGGAGT TACCTTGAGT GTGATGAAGA AGCCAAGGCT TTCTTTAGTC CGCTGATGGG ACACTACATG
AAGAGCGTTC TAAGCAAGGA GGCATACGTG AAGGATTTAT TGAAATATTC GAGTGACATA ATTGTTGGCG
AAGTTGACCA CGATGTGTTT GAGGATAGCG TCGCACAAGT TATCGAGTTG CTAAATGATC ACGAGTGTCC
TGAGCTCGAA TACGTCACAG ACAGCGAGGT CATCATACAA GCTCTAAACA TGGATGCAGC GGTCGGAGCT
TTGTACACAG GAAAGAAGAG AAAATACTTC GAAGGATCAA CAGTTGAGCA CAGACAGGCC CTAGTTCGGA
AAAGTTGCGA GCGTCTTTAT GAAGGGCGGA TGGGTGTTTG GAATGGATCG TTGAAGG
```

Figure 14 Map of pPCwF (A) and nucleotide sequence of wF cDNA fragment (B). The pPCwF plasmid of 5380 base pairs in closed circular form was obtained from inserting the DNA fragment 'wF' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A

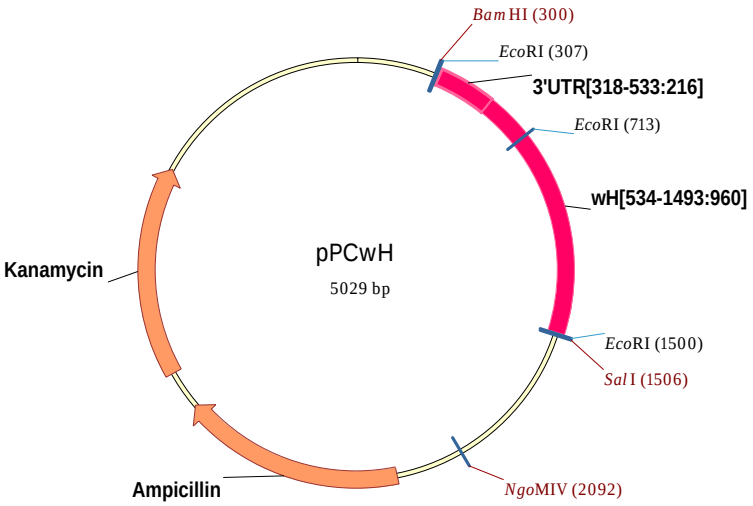


B

```
AGGAATCTGT CCTCTCTTTA GGAGTTACCT TGAGTGTGAT GAAGAAGCCA AGGCTTTCTT TAGTCCGCTG
ATGGGACACT ACATGAAGAG CGTTCTAAGC AAGGAGGCAT ACGTGAAGGA TTTATTGAAA TATTCGAGTG
ACATAATTGT TGGCGAAGTT GACCACGATG TGTTTGAGGA TAGCGTCGCA CAAGTTATCG AGTTGCTAAA
TGATCACGAG TGTCCTGAGC TCGAATACGT CACAGACAGC GAGGTCATCA TACAAGCTCT AAACATGGAT
GCAGCGGTCG GAGCTTTGTA CACAGGAAAG AAGAGAAAAT ACTTCGAAGG ATCAACAGTT GAGCACAGAC
AGGCCCTAGT TCGGAAAAGT TGCGAGCGTC TTTATGAAGG GCGGATGGGT GTTTGGAATG GATCGTTGAA
GGCCGAGCTG CGACCTGCAG AGAAAGTTCT AGCCAAAAAG ACAAGATCTT TCACAGCAGC GCCTCTCGAC
ACTTTGTTGG GAGCTAAAGT TTGCGTTGAT GATTTCAACA ATTGGTTCTA CAGCAAAAAT ATGGAGTGTC
CCTGGACTGT TGGCATGACG AAGTTTTACA AAGGCTGGGA CGAATTTTTG AGGAAATTCC CTGATAGTTG
GGTTTACTGC GATGCAGATG GATCTCAGTT CGATAGTTCT TTGACACCAT ATTTGCTTAA TGCTGTGTTG
TCTATCCGAC TGTGGGCTAT GGAGGATTGG GACATTGGAG CTCAGATGCT TAAGAACTTG TATGGAGAAA
TCACTTACAC ACCTATTTTG ACACCTGATG GGACGATTGT CAAGAAGTTT AAAGGAAACA ACAGTGGCCA
ACCCTCAACA GTTGTCGATA ATACATTGAT GGTTTTAATC ACAATGTATT ACGCACTACG CAAGGCTGGT
TACGACACGG AGGCCCAGGA AGATATGTGC GTATTCTATA TCAATGGTGA TGATCTCTGT ATCGCCATCC
ACCCTGATCA CGAGCATGTT CTTGACTCAT TCTCTAGTTC ATTTGCTGAG CTTGGACTCA AAGATGATTT
CACTCAAAGA CACCGGAATA AACAAGATCT GTGGTTCATG TCGCATCGAG GTATTCTGAT CGATGACATT
TACATTCCTA AACTTGAACC CGAGAGAATT GTGGCCATCC TTGAGTGGGA CAAATCTAAG CTTCCGGAAC
ATCGGTTGGA AGCTATTACA GCAGCAATGA TAGAGTCATG GGGATACGGG GAGTTAACGC ACCAGATCCG
CAGATTTTAT CAATGGGTAC TCGAGCAGGC TCCATTCGAT GAGTTAGCGG AGCAGGGCAG GGCTCCATAT
GTGTCTGAAG TTGGATTGAG GCGCTTGTAC ACTAGTGAGC GTGGGTCAAT GGACGAATTG GAGGCCTACA
TAGATAAATA TTTTGAGCGA GAGAGAGGAG ATTCGCCTGA ACTATTGGTG TATCATGAAT CCAGAAGTAC
TGATGATTAC CAGCTTATTT GCGGTGAGAG CACACACGTG TTCCACCAGT CCAAAAATGA AGCTGTGGAT
```

Figure 15 Map of pPCwG (A) and nucleotide sequence of wG cDNA fragment (B). The pPCwG plasmid of 5395 base pairs in closed circular form was obtained from inserting the DNA fragment 'wG' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

A



B

```
CGAGAGAGAG GAGATTCGCC TGAACTATTG GTGTATCATG AATCCAGAAG TACTGATGAT TACCAGCTTA
TTTGCGGTGA GAGCACACAC GTGTTCCACC AGTCCAAAAA TGAAGCTGTG GATGCTGGTC TGAATGAAAA
GTTCAAAGAG AAAGAGAAAC AAAAAGAAGA AAAAGATAAA CAAAAAGATA AGAATAATGA TGGAGCTAGT
GACGGAAACG ATGTGTCAAC TAGCACAAAA ACTGGAGAGA GAGATAGAGA TGTCAATGCT GGAACTAGTG
GAACTTTCAC TGTTCCAAGA ATGAAGTCAT TTACTAATAA GATGATTTTA CCAAGAATTA AGGGAAAAAC
TGTCCTTAAT TTGAATCATC TTCTTCAGTA CAACCCGCAA CAGATTGACA TCTCAAACAC TCGTGCCACT
CAGTCTCAAT TTGAGAAGTG GTATGAGGGA GTGAGGAATG ATTATGGTCT TAATGGCAAT GAGATGCAGG
TAATGTTGAA TGGTTTGATG GTTTGGTGTA TCGAAAATGG TACATCTCCA GACATATCTG GTGTCTGGGT
TATGGTGGAT GGGGAAACCC AAGTCGATTA TCCTATCAGA CCTTTGATTG AGCACGCAAC TCCTTCATTC
AGGCAAATCA TGGCTCACTT CAGCAACGCG GCAGAGGCAT ACATTGCGAA AAGGAATGCA ACTGAGAGGT
ACATGCCGCG GTATGGAATC AAGAGGAATT TGACTGACAT TAGCCTCGCT AGATATGCTT TCGATTTCTA
TGAGGTGAAT TCAAAAACAC CTGATAGGGC TCGCGAAGCT CATATGCAGA TGAAAGCTGC AGCACTACGC
AATACTGGTC GTAGAATGTT TGGAATGGAC GGCAGTGTCA GTAACAAGGA GGAAAACACG GAGAGACACA
CAGTGGAAGA TGTCAACAAA GACATGCACT CTCTCCTGGG TATGCGCAAT TGAATACTCG CGCTAGTGTG
TTCGTCGGGC CTGGCTCGAC CCTGTTTCAC CTTATAATAC TATATAAGCA TTAGAATACA GTGTGGCTGC
GCCACCGCTT CTATTTTACA GTGAGGGTAG CCCTCCGTGC TTTTAGTGTT ATTCGAGTTC TCTGAGTCTC
CATACAGTGT GGGTGGCCCA CGTGCTATTC GAGCCTCTTA GAATGAGAGA AAAAAA
```

Figure 16 Map of pPCwH (A) and nucleotide sequence of wH cDNA fragment (B). The pPCwH plasmid of 5029 base pairs in closed circular form was obtained from inserting the DNA fragment 'wH' into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wA' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

3.2 Restriction analysis of eight recombinant plasmids

Restriction analyses of the eight different recombinant plasmids containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate illustrated separately DNA fragment (s) patterns with expected size of fragments that can be visualized on gel electrophoresis under UV exposing after digestion with a set of appropriated restriction endonucleases.

The results of the eight different recombinant plasmids digested separately with restriction endonuclease *Eco*RI were analyzed and compared by visualization using gel electrophoresis under UV exposing. Five recombinant plasmids including pPCwA, pPCwB, pPCwE, pPCwF and pPCwG have no *Eco*RI site on their insert fragments. There is one restriction site to *Eco*RI in insert fragments 'wC' and 'wH' presenting on pPCwC and pPCwH, respectively. While 'wD' in pPCwD contains two sites of *Eco*RI (Figure 17)

For *Bam*HI restriction analyses, all eight recombinant plasmids contain only one site (on vector sequence) for digestion and were linearized with this enzyme showing a band (in each) of complete recombinant plasmid, but there is no the *Bam*HI site on all of insert fragments (Figure 18). Similarly, results obtained from *Sal*I digestion on the eight recombinant plasmids indicate that there is no the *Sal*I site on all of insert fragments (Figure 19), one site recognizing to *Sal*I has a location on vector.

Based on restriction map of pDrive cloning vector from QIAGEN as shown in Appendix A Figure 1 and DNA band patterns of *Bam*HI and *Sal*I restriction analyses as shown in Figure 20, the eight different recombinant plasmids were digested separately with restriction endonuclease *Bam*HI and *Sal*I. Vector fragments of 3823 base pairs were observed in all different recombinant plasmids digested with *Bam*HI and *Sal*I. Another fragment in each reaction contains its insert fragment, and upstream and downstream sequences of 17- and 13-base pairs which they are parts of the vector. These fragments of 993-, 2268-, 2003-, 1821-, 1268-, 1557-, 1572- and

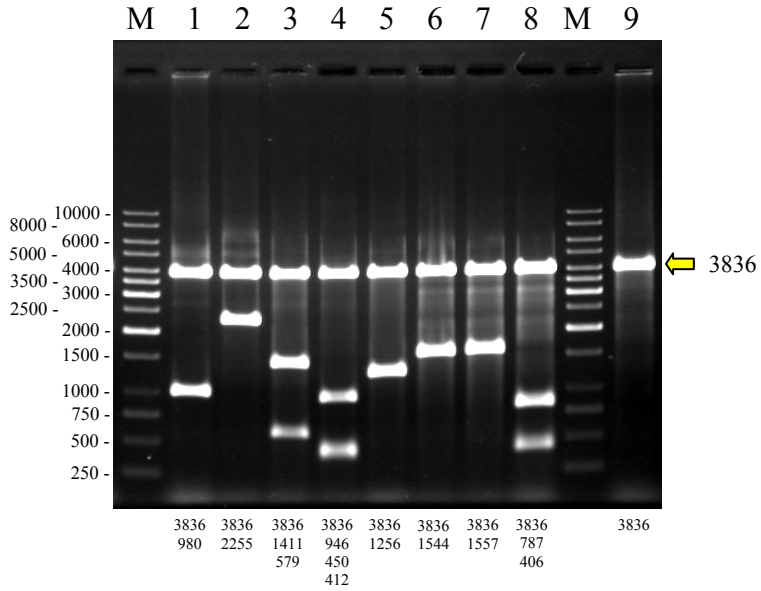


Figure 17 Restriction analyses of the eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate by digestion with *Eco*RI restriction endonuclesae. These different plasmids (closed circular form) obtained separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN) were digested separately under an appropriate condition for digestion of this enzyme. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The linearized vector fragments are depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Eco*RI
Lane 2: pPCwB digested with *Eco*RI
Lane 3: pPCwC digested with *Eco*RI
Lane 4: pPCwD digested with *Eco*RI
Lane 5: pPCwE digested with *Eco*RI
Lane 6: pPCwF digested with *Eco*RI
Lane 7: pPCwG digested with *Eco*RI
Lane 8: pPCwH digested with *Eco*RI
Lane 9: pDrive cloning vector self-ligating digested with *Eco*RI

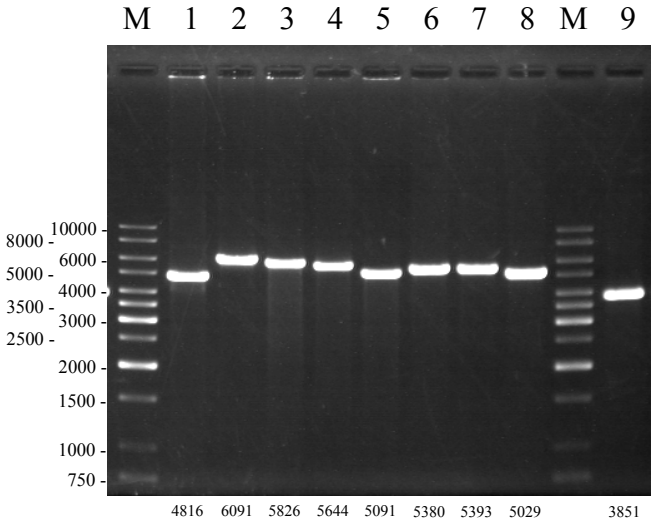


Figure 18 Restriction analyses of the eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate by digestion with *Bam*HI restriction endonuclesae. These different plasmids (closed circular form) obtained separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN) were digested separately under an appropriate condition for digestion of this enzyme. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Bam*HI
Lane 2: pPCwB digested with *Bam*HI
Lane 3: pPCwC digested with *Bam*HI
Lane 4: pPCwD digested with *Bam*HI
Lane 5: pPCwE digested with *Bam*HI
Lane 6: pPCwF digested with *Bam*HI
Lane 7: pPCwG digested with *Bam*HI
Lane 8: pPCwH digested with *Bam*HI
Lane 9: pDrive cloning vector self-ligating digested with *Bam*HI

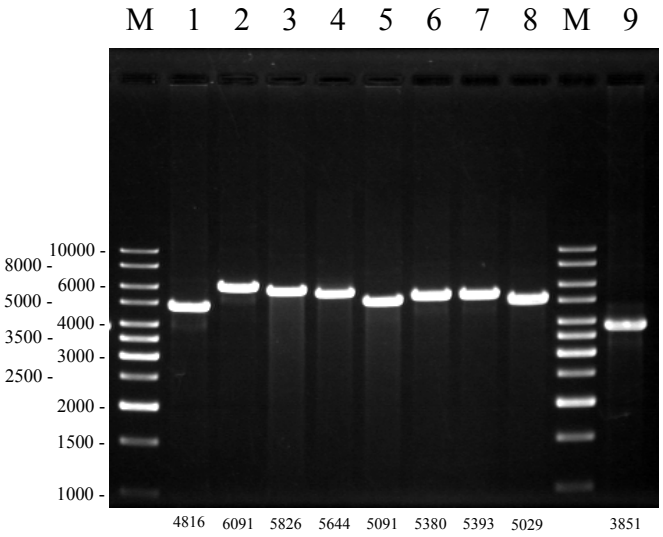


Figure 19 Restriction analyses of the eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate by digestion with *Sal*I restriction endonuclesae. These different plasmids (closed circular form) obtained separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN) were digested separately under an appropriate condition for digestion of this enzyme. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Sal*I
Lane 2: pPCwB digested with *Sal*I
Lane 3: pPCwC digested with *Sal*I
Lane 4: pPCwD digested with *Sal*I
Lane 5: pPCwE digested with *Sal*I
Lane 6: pPCwF digested with *Sal*I
Lane 7: pPCwG digested with *Sal*I
Lane 8: pPCwH digested with *Sal*I
Lane 9: pDrive cloning vector self-ligating digested with *Sal*I

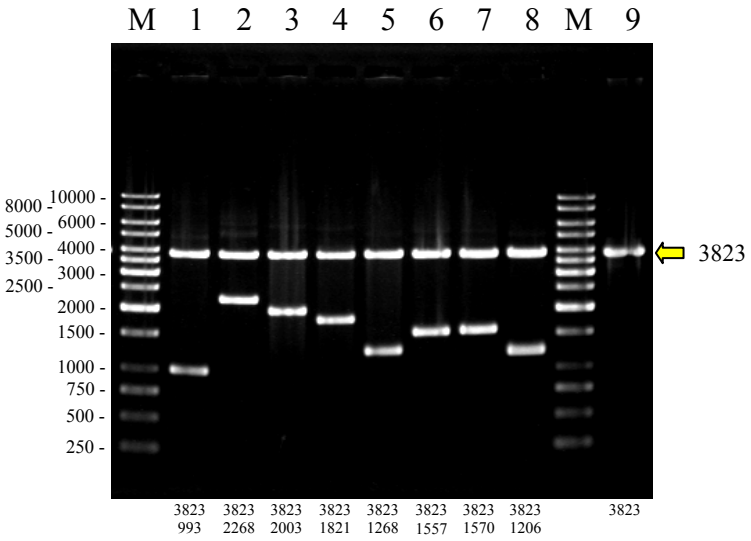


Figure 20 Restriction analyses of the eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate by digestion with *Bam*HI and *Sal*I restriction endonuclesae. These different plasmids (closed circular form) obtained separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN) were digested separately under an appropriate condition for digestion of this enzyme. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The linearized vector fragments are depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Bam*HI and *Sal*I
Lane 2: pPCwB digested with *Bam*HI and *Sal*I
Lane 3: pPCwC digested with *Bam*HI and *Sal*I
Lane 4: pPCwD digested with *Bam*HI and *Sal*I
Lane 5: pPCwE digested with *Bam*HI and *Sal*I
Lane 6: pPCwF digested with *Bam*HI and *Sal*I
Lane 7: pPCwG digested with *Bam*HI and *Sal*I
Lane 8: pPCwH digested with *Bam*HI and *Sal*I
Lane 9: pDrive cloning vector self-ligating digested with *Bam*HI and *Sal*I

1206 base pairs were shown for digesting with *Bam*HI and *Sal*I of pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH, respectively.

There is one site of sequence recognizing to endonuclease *Nco*I. The site is found on insert fragment wC in pPCwC plasmid and can be verified by visualization using gel electrophoresis under UV exposing after digestion the plasmid with *Bam*HI, *Sal*I and *Nco*I (Figure 21). No this site in other plasmids was observed.

By comparisons with self-ligating pDrive cloning vector digested with *Eco*RI (3836 base pairs) and size standard markers (1kb DNA Ladder 'GeneRuler^{TM}', Fermentas), DNA band patterns obtained from restriction analyses with restriction endonucleases *Bam*HI, *Eco*RI, *Nco*I and *Sal*I of the eight different recombinant plasmids are shown separately in Figure 22 for pPCwA, Figure 23 for pPCwB, Figure 24 for pPCwC, Figure 25 for pPCwD, Figure 26 for pPCwE, Figure 27 for pPCwF, Figure 28 for pPCwG, and Figure 29 for pPCwH.

Based on the sequence and restriction analyses of the eight different recombinant plasmids, physical maps of these recombinant plasmids were designed and drawn using Vector NTI computing software (InforMax). Figure 9A, Figure 10A, Figure 11A, Figure 12A, Figure 13A, Figure 14A, Figure 15A, and Figure 16A show predicted maps of pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH, respectively.

3.3 Complete nucleotide sequence of the PRSV-W RNA

The nucleotide sequence of the complete PRSV-W genomic RNA was assembled and analyzed using the personal computer-based program DNAStar biocomputing software (DNAStar, Inc.) The results of the assembled sequences revealed that the full-length genome of PRSV-W consists of 10323 nucleotides excluding the 3' terminal poly(A) tail (Figure 30). The length of the genomic RNA is similar to other fully-sequenced PRSV isolates, which range from 10320 nucleotides (Mexican isolate PRSV-P-MEX; GenBank accession number AY231130) to 10334

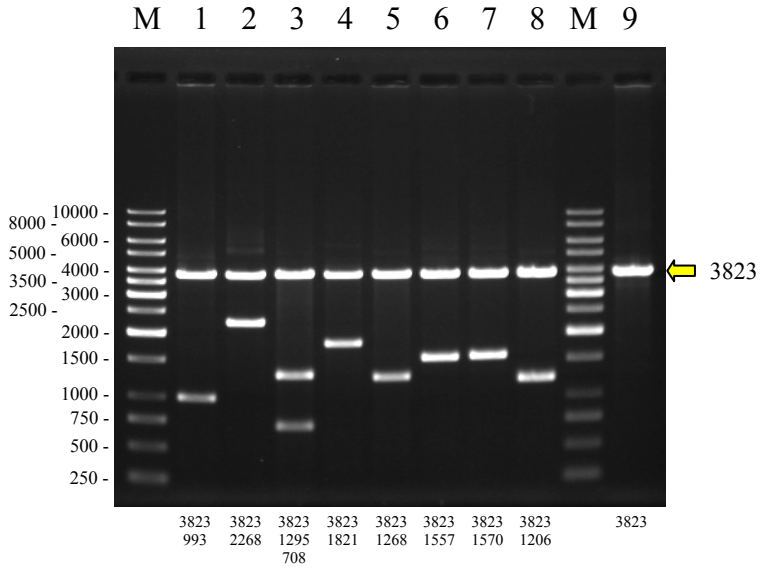


Figure 21 Restriction analyses of the eight recombinant plasmids obtained from different clones containing different overlapping cDNA fragments corresponding to the genomic RNA of the PRSV-W isolate by digestion with *BamH*I, *Nco*I and *Sal*I restriction endonuclesae. These different plasmids (closed circular form) obtained separately from inserting the different DNA fragments into U-tailed pDrive cloning vector (QIAGEN) were digested separately under an appropriate condition for digestion of this enzyme. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The linearized vector fragments are depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Bam*HI, *Nco*I and *Sal*I
Lane 2: pPCwB digested with *Bam*HI, *Nco*I and *Sal*I
Lane 3: pPCwC digested with *Bam*HI, *Nco*I and *Sal*I
Lane 4: pPCwD digested with *Bam*HI, *Nco*I and *Sal*I
Lane 5: pPCwE digested with *Bam*HI, *Nco*I and *Sal*I
Lane 6: pPCwF digested with *Bam*HI, *Nco*I and *Sal*I
Lane 7: pPCwG digested with *Bam*HI, *Nco*I and *Sal*I
Lane 8: pPCwH digested with *Bam*HI, *Nco*I and *Sal*I
Lane 9: pDrive cloning vector self-ligating digested with *Bam*HI, *Nco*I and *Sal*I

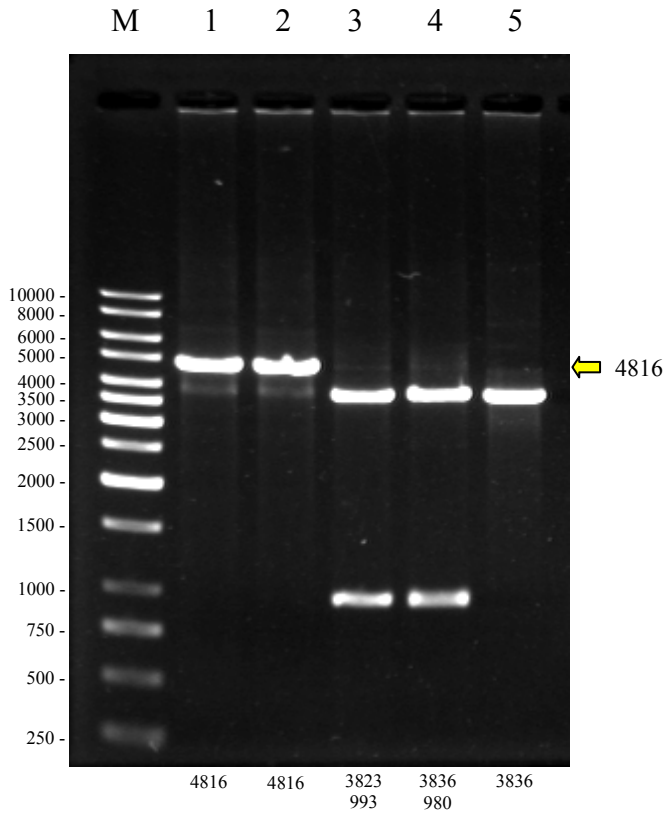


Figure 22 Restriction analyses of the pPCwA plasmid. The pPCwA plasmid of 4816 base pairs in closed circular form obtained from inserting the DNA fragment 'wA' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwA plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Bam*HI
Lane 2: pPCwA digested with *Sal*I
Lane 3: pPCwA digested with *Bam*HI and *Sal*I
Lane 4: pPCwA digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

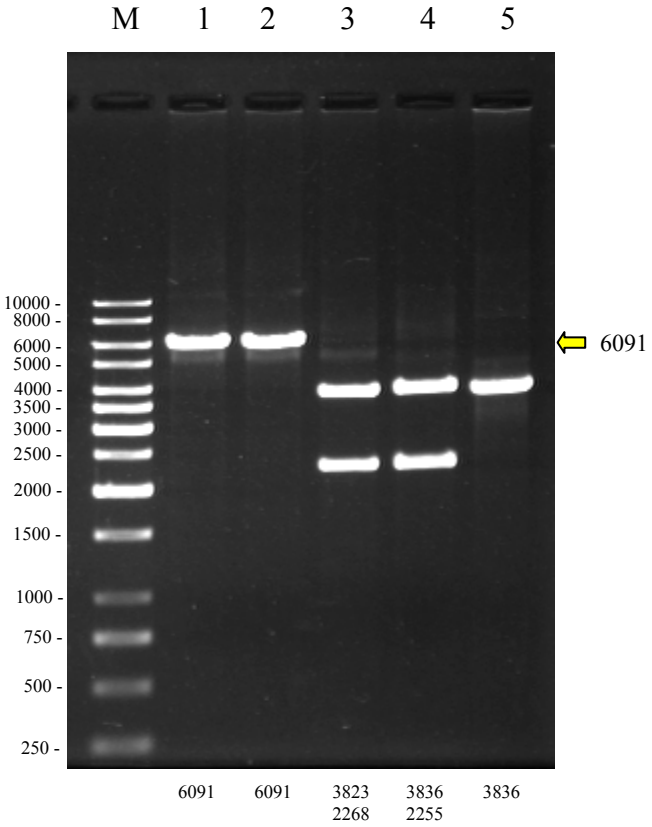


Figure 23 Restriction analyses of the pPCwB plasmid. The pPCwB plasmid of 6091 base pairs in closed circular form obtained from inserting the DNA fragment 'wB' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwB plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwB digested with *Bam*HI
Lane 2: pPCwB digested with *Sal*I
Lane 3: pPCwB digested with *Bam*HI and *Sal*I
Lane 4: pPCwB digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

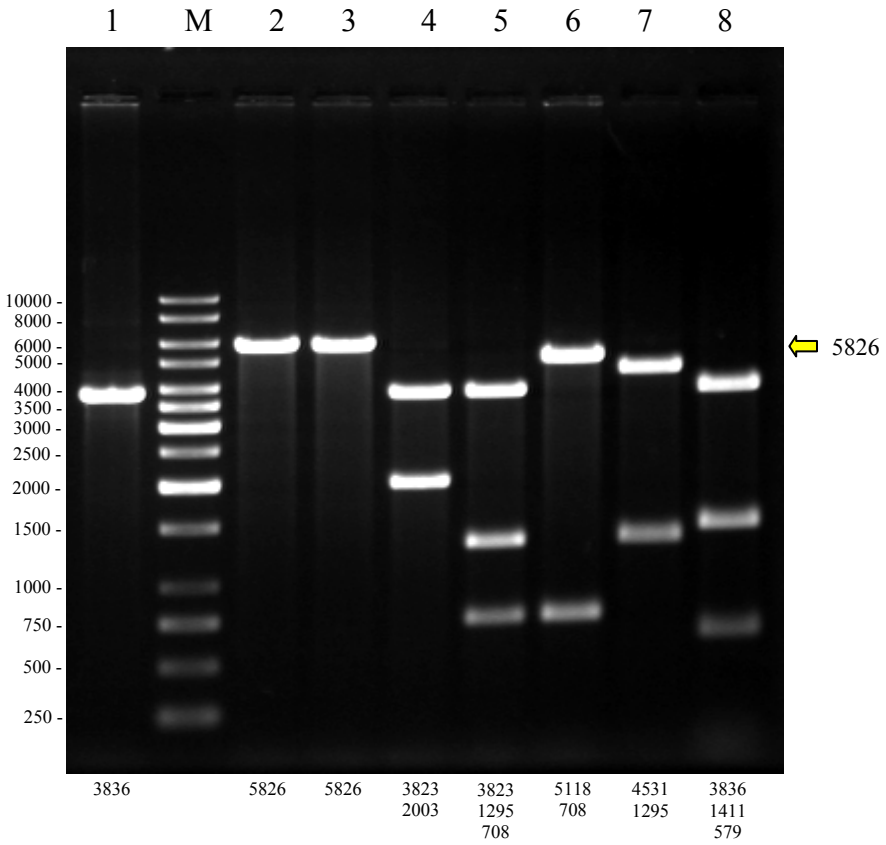


Figure 24 Restriction analyses of the pPCwC plasmid. The pPCwC plasmid of 5826 base pairs in closed circular form obtained from inserting the DNA fragment 'wC' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwC plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane 1: pDrive cloning vector self-ligating digested with *Eco*RI
Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 2: pPCwC digested with *Bam*HI
Lane 3: pPCwC digested with *Sal*I
Lane 4: pPCwC digested with *Bam*HI and *Sal*I
Lane 5: pPCwC digested with *Bam*HI, *Nco*I and *Sal*I
Lane 6: pPCwC digested with *Bam*HI and *Nco*I
Lane 7: pPCwC digested with *Nco*I and *Sal*I
Lane 8: pPCwC digested with *Eco*RI

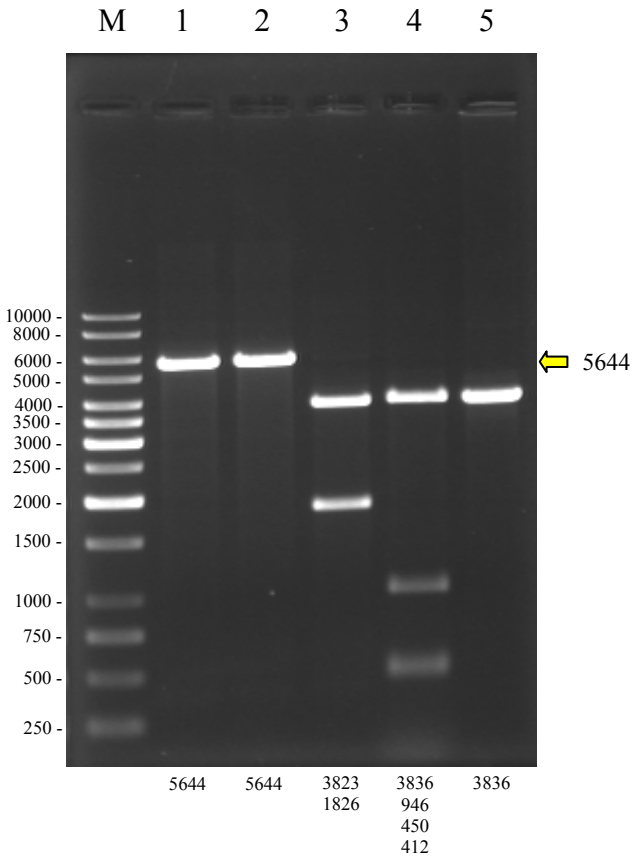


Figure 25 Restriction analyses of the pPCwD plasmid. The pPCwD plasmid of 5644 base pairs in closed circular form obtained from inserting the DNA fragment 'wD' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwD plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwA digested with *Bam*HI
Lane 2: pPCwD digested with *Sal*I
Lane 3: pPCwD digested with *Bam*HI and *Sal*I
Lane 4: pPCwD digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

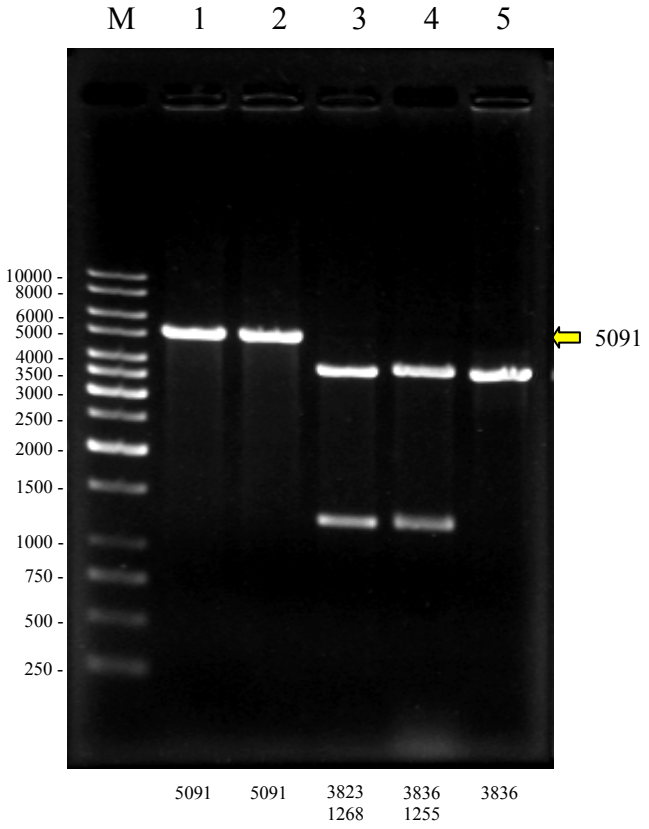


Figure 26 Restriction analyses of the pPCwE plasmid. The pPCwE plasmid of 5091 base pairs in closed circular form obtained from inserting the DNA fragment 'wE' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwE plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwE digested with *Bam*HI
Lane 2: pPCwE digested with *Sal*I
Lane 3: pPCwE digested with *Bam*HI and *Sal*I
Lane 4: pPCwE digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

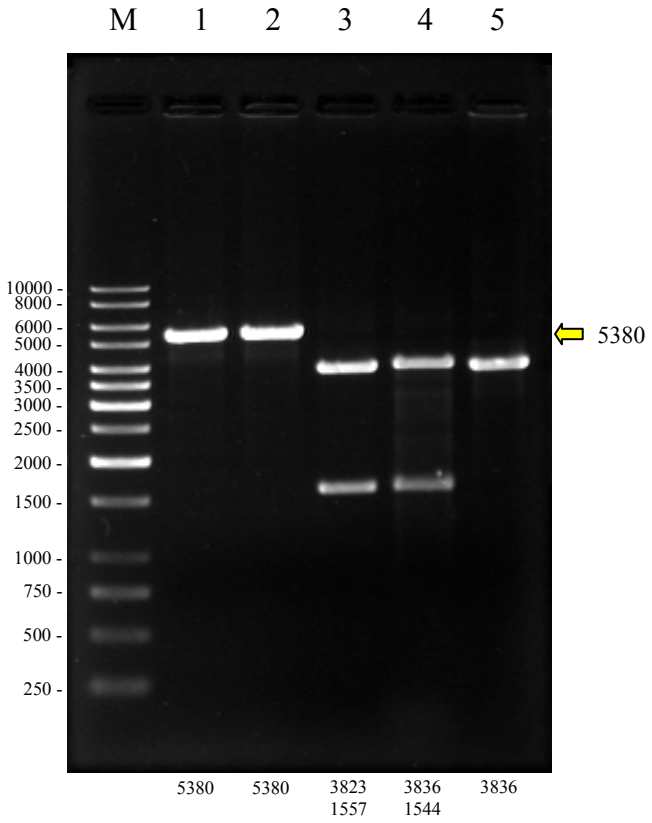


Figure 27 Restriction analyses of the pPCwF plasmid. The pPCwF plasmid of 5380 base pairs in closed circular form obtained from inserting the DNA fragment 'wF' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwF plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwF digested with *Bam*HI
Lane 2: pPCwF digested with *Sal*I
Lane 3: pPCwF digested with *Bam*HI and *Sal*I
Lane 4: pPCwF digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

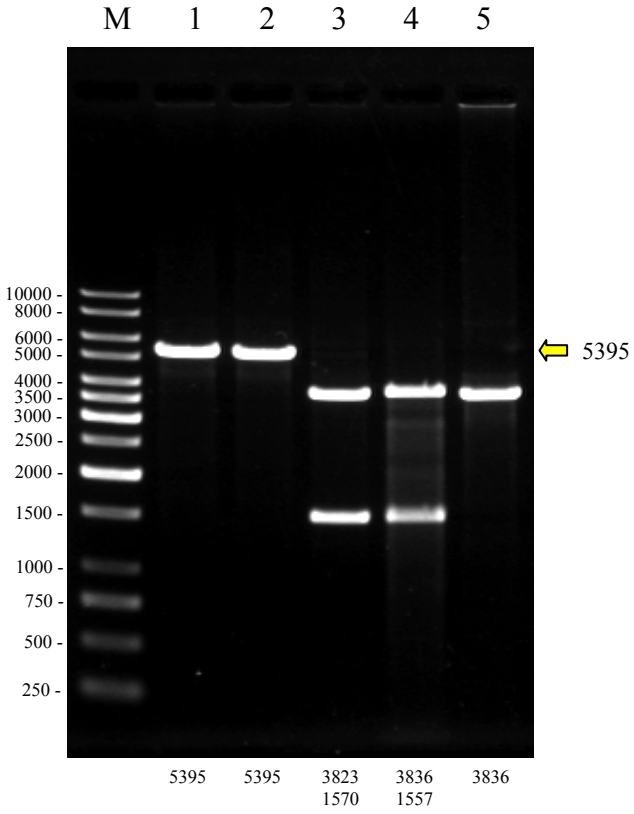


Figure 28 Restriction analyses of the pPCwG plasmid. The pPCwG plasmid of 5395 base pairs in closed circular form obtained from inserting the DNA fragment 'wG' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwG plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwG digested with *Bam*HI
Lane 2: pPCwG digested with *Sal*I
Lane 3: pPCwG digested with *Bam*HI and *Sal*I
Lane 4: pPCwG digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

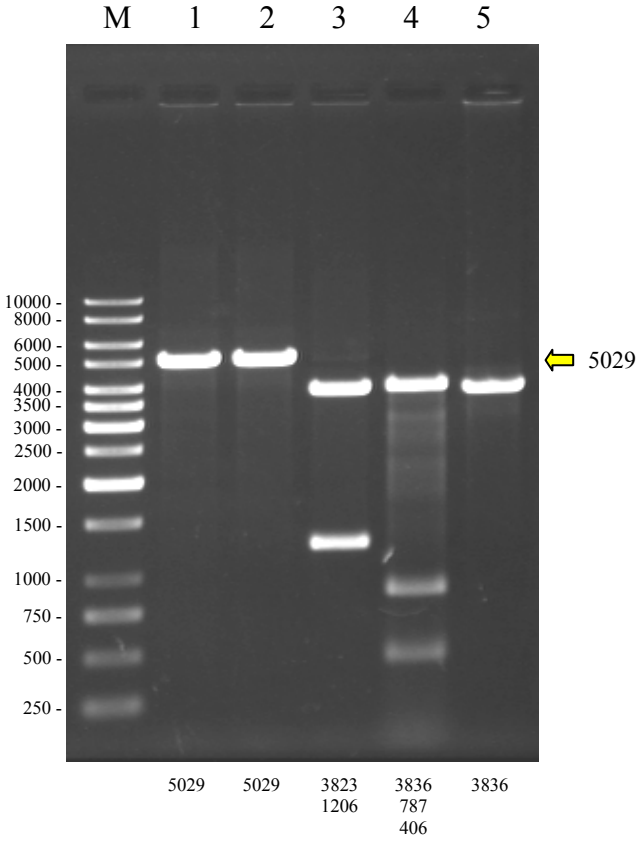


Figure 29 Restriction analyses of the pPCwH plasmid. The pPCwH plasmid of 5029 base pairs in closed circular form obtained from inserting the DNA fragment 'wH' into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pPCwH plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pPCwH digested with *Bam*HI
Lane 2: pPCwH digested with *Sal*I
Lane 3: pPCwH digested with *Bam*HI and *Sal*I
Lane 4: pPCwH digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

```
   1 AAAUAAAACA UCUCAACACA ACACAAUUCA AAGCAUUCAA ACACACUCAA GCAAAUUUUA
                           5′-NCR/P1
                                     M  S   S  L  Y   Q  L  Q   P  I  A  L 12
  61 UUCUCAAAUU UCACAAUCUG CAAGCAUGUC UUCGUUGUAC CAAUUGCAAC CAAUUGCACU

       K  D  R   L  L  S   H  K  R  G   K  G  W   I  E  H   K  L  E  R 32
 121 GAAGGACCGU CUCUUGUCUC ACAAGAGAGG AAAAGGUUGG AUUGAGCACA AGCUCGAAAG

       K  G  E   R  G  N   T  R  H  V   G  E  F   V  I  S   E  G  A  K 52
 181 GAAAGGAGAG AGAGGAAACA CUCGUCAUGU UGGCGAGUUU GUGAUCAGUG AGGGAGCAAA

       I  L  Q   L  V  Q   I  G  N  A   E  I  G   R  T  F   L  E  G  D 72
 241 AAUUUUGCAA CUCGUGCAAA UUGGAAACGC UGAAAUUGGC AGGACCUUCC UGGAAGGUGA

       R  R  T   Q  A  D   I  F  E  I   I  K  K   T  M  V   G  H  L  G 92
 301 UAGGAGAACU CAAGCUGACA UUUUUGAAAU UAUCAAAAAG ACGAUGGUCG GCCACCUGGG

       Y  D  F   E  C  G   L  W  C  C   Y  S  C   D  N  T   S  D  K  Y 112
 361 GUACGACUUC GAAUGUGGGU UAUGGUGCUG UUACUCUUGU GACAACACAU CUGACAAGUA

       F  K  K   C  D  C   G  E  K  Y   Y  Y  S   E  K  N   L  I  K  S 132
 421 CUUCAAGAAG UGUGAUUGCG GUGAGAAGUA CUAUUAUUCA GAGAAGAAUU UGAUUAAGUC

       M  H  D   L  M  Y   Q  F  D  M   T  A  A   E  I  D   Q  V  G  L 152
 481 AAUGCAUGAC UUGAUGUACC AGUUUGACAU GACAGCCGCU GAAAUUGACC AAGUAGGUCU

       D  Y  L   A  E  A   V  D  Y  A   E  Q  S   V  K  K   S  K  V  L 172
 541 UGACUACUUG GCUGAAGCUG UAGACUAUGC UGAACAGUCC GUCAAGAAAU CAAAGGUACU

       V  P  D   E  P  K   F  V  E  I   L  A  A   S  E  E   S  L  L  V 192
 601 GGUACCUGAU GAACCUAAGU UUGUGGAGAU CCUCGCGGCG AGUGAGGAGA GUCUUCUGGU

       V  P  E   P  E  V   A  S  V  T   T  R  A   E  E  A   W  T  I  Q 212
 661 GGUUCCUGAG CCAGAAGUUG CUUCUGUAAC UACAAGAGCU GAAGAAGCAU GGACGAUACA

       I  G  E   I  P  V   P  L  V  V   I  R  E   T  P  V   I  S  G  V 232
 721 AAUUGGGGAA AUUCCUGUCC CACUUGUCGU CAUAAGGGAG ACGCCAGUAA UCAGUGGUGU

       N  G  M   L  S  S   T  G  F  S   L  E  A   E  I  A   R  P  A  K 252
 781 GAAUGGAAUG CUGAGCUCAA CUGGUUUUUC ACUUGAAGCG GAAAUUGCAA GACCGGCUAA

       S  T  I   P  Q  D   E  V  E  E   A  V  H   L  A  L   E  V  G  N 272
 841 AAGUACGAUU CCUCAAGAUG AAGUGGAGGA AGCAGUGCAC CUGGCACUCG AGGUCGGCAA

       E  I  A   E  K  K   P  E  L  K   L  A  P   Y  W  S   A  S  L  E 292
 901 UGAAAUCGCC GAAAAGAAGC CUGAACUCAA AUUGGCACCA UACUGGAGUG CUAGCCUUGA

       L  H  K   R  V  R   K  H  K  E   H  A  K   S  E  A   L  R  V  R 312
 961 GUUGCACAAG AGAGUCCGCA AACACAAGGA ACACGCUAAA UCUGAAGCAC UUCGAGUUCG

       K  E  K   E  R  D   H  R  I  F   A  A  L   E  A  K   L  N  L  K 332
1021 AAAAGAGAAG GAAAGAGAUC AUAGGAUUUU UGCUGCUUUA GAAGCAAAGC UGAAUUUGAA
```

Figure 30 The complete nucleotide sequence of PRSV-W and deduced amino acid sequence of the polyprotein. The predicted cleavage sites (pair of amino acid residues in each site) are indicated in shade boxes, and those of mature protein junction signatures described in the text are shown above their sites. The nucleotide and amino acid positions are indicated at the start and the end, respectively, of each line.

```
      A  R  R   Q  G  Q   V  V  V  C   D  K  R   G  T  L   K  W  K  K 352
1081 AGCAAGGAGG CAAGGCCAAG UUGUUGUUUG CGACAAGAGA GGUACUUUAA AAUGGAAAAA

      H  Q  Q   R  K  R   N  K  M  V   T  Q  L   S  D  S   V  V  T  K 372
1141 GCAUCAACAG CGGAAGAGGA ACAAAAUGGU UACGCAAUUA AGUGACUCUG UCGUCACGAA

      I  H  A   N  F  E   C  R  T  P   N  F  D   V  E  T   P  G  I  K 392
1201 AAUUCACGCC AAUUUUGAGU GCAGGACUCC UAAUUUUGAU GUUGAGACUC CUGGCAUUAA

      C  A  T   S  K  V   T  R  K  K   Q  T  Q   P  K  I   F  G  S  N 412
1261 GUGUGCAACC UCGAAGGUGA CGCGGAAGAA ACAAACGCAA CCUAAGAUAU UUGGCAGCAA

      K  V  N   Y  V  M   K  N  L  C   D  I  I   I  D  R   N  I  P  V 432
1321 CAAGGUCAAU UAUGUCAUGA AAAAUCUGUG UGACAUAAUC AUCGACAGGA ACAUUCCUGU

      E  L  I   T  K  R   C  K  R  R   I  F  R   R  D  G   K  N  Y  V 452
1381 UGAGUUAAUC ACGAAGCGCU GCAAAAGGAG AAUUUUCCGU AGGGACGGGA AAAACUAUGU

      H  L  R   H  M  D   G  N  N  A   P  R  D   V  S  S   S  S  D  M 472
1441 GCACUUAAGA CACAUGGAUG GCAACAAUGC ACCACGGGAU GUAAGCAGCU CGUCAGACAU

      E  K  L   F  T  R   F  C  K  F   L  I  R   K  Q  S   I  N  A  A 492
1501 GGAGAAGUUG UUUACGCGCU UUUGCAAAUU UUUAAUCAGG AAGCAGUCGA UCAACGCAGC

      N  L  T   H  G  S   S  G  L  I   F  K  P   K  F  A   D  R  T  G 512
1561 GAAUCUAACA CAUGGUUCGA GUGGUCUCAU CUUUAAACCG AAGUUCGCUG ACCGAACAGG

      R  Y  F   G  E  Y   F  I  T  R   G  R  C   E  G  K   L  F  D  G 532
1621 GCGAUAUUUU GGUGAAUAUU UUAUCACGCG UGGACGAUGU GAAGGAAAGC UAUUUGAUGG
                                                  P1/HC-Pro
      R  S  K   L  A  K   S  V  R  M   R  M  E   Q  Y/ N   D  V  A  E 552
1681 GAGAUCAAAA CUAGCUAAGU CGGUUCGCAU GCGCAUGGAA CAAUACAAUG AUGUCGCUGA

      K  F  W   L  G  F   N  R  A  F   L  R  H   R  K  P   T  D  H  V 572
1741 GAAGUUCUGG CUCGGUUUUA ACAGGGCUUU CUUACGGCAC AGGAAGCCGA CGGAUCAUGU

      C  T  S   D  M  D   V  T  M  C   G  E  V   A  A  L   A  T  I  I 592
1801 GUGCACGUCU GAUAUGGAUG UUACCAUGUG UGGUGAAGUG GCGGCUCUUG CAACCAUAAU

      L  F  P   C  H  K   I  T  C  N   A  C  M   N  K  V   K  G  R  V 612
1861 CCUGUUUCCA UGUCAUAAGA UAACUUGUAA CGCUUGCAUG AACAAGGUAA AGGGAAGAGU

      I  D  E   V  G  E   D  L  N  C   E  L  E   R  L  R   E  T  L  S 632
1921 CAUUGACGAA GUCGGUGAGG AUCUGAAUUG UGAGCUCGAA CGUUUGCGCG AAACCCUCUC

      S  Y  G   G  S  F   G  H  V  S   T  L  L   D  Q  L   N  R  V  L 652
1981 GUCGUAUGGA GGCUCAUUUG GUCAUGUUUC AACAUUGCUU GAUCAAUUAA ACAGAGUUUU

      N  A  R   N  M  N   D  G  A  F   K  E  I   A  K  K   I  D  E  K 672
2041 AAAUGCGCGU AACAUGAAUG AUGGAGCUUU UAAGGAGAUC GCGAAGAAAA UUGAUGAGAA

      K  E  S   P  W  I   H  M  T  V   I  N  N   T  L  I   K  G  S  L 692
2101 GAAAGAAAGU CCUUGGAUUC AUAUGACGGU CAUCAACAAU ACGCUUAUUA AAGGUUCAUU

      A  T  G   Y  E  F   E  R  A  S   D  S  L   R  E  I   V  R  W  H 712
2161 AGCAACUGGC UAUGAAUUUG AGAGAGCGUC UGAUAGUUUA CGAGAAAUUG UAAGAUGGCA

      L  K  R   T  E  S   I  K  A  G   S  V  E   S  F  R   N  K  R  S 732
2221 UCUCAAGAGA ACAGAGUCAA UCAAAGCUGG CAGUGUUGAG AGCUUCAGAA ACAAACGUUC
```

Figure 30 (continued)

```
      G  K  A  H  F  N  P  A  L  T  C  D  N  Q  L  D  R  N  G  N 752
2281 UGGAAAAGCU CAUUUCAACC CGGCCCUCAC AUGUGACAAU CAAUUGGACA GAAAUGGCAA

      F  L  W  G  E  R  Q  Y  H  A  K  R  F  F  A  N  Y  F  E  K 772
2341 UUUCUUAUGG GGUGAGAGAC AAUAUCAUGC CAAAAGAUUC UUUGCUAAUU ACUUUGAGAA

      I  D  H  S  K  G  Y  E  Y  Y  S  Q  R  Q  N  P  N  G  I  R 792
2401 GAUUGAUCAC AGUAAGGGUU AUGAGUACUA CAGUCAACGC CAAAACCCGA AUGGAAUUAG

      K  I  A  I  S  N  L  I  F  S  T  N  L  E  R  F  R  Q  Q  M 812
2461 AAAGAUUGCC AUUAGUAAUC UAAUAUUCUC CACAAAUUUG GAGAGGUUUC GGCAGCAAAU

      V  E  H  H  I  D  Q  G  P  I  T  R  E  C  I  A  L  R  N  N 832
2521 GGUUGAACAU CACAUUGACC AGGGACCAAU CACUCGUGAG UGUAUCGCAC UGCGCAAUAA

      N  Y  V  H  V  C  S  C  V  T  L  D  D  G  T  P  A  T  S  E 852
2581 CAAUUAUGUU CAUGUAUGUA GCUGCGUGAC CUUAGAUGAU GGAACUCCAG CGACGAGCGA

      L  K  T  P  T  K  N  H  I  V  L  G  N  S  G  D  P  K  Y  V 872
2641 AUUGAAAACU CCCACCAAGA AUCACAUUGU UCUUGGUAAU UCUGGUGAUC CUAAGUAUGU

      D  L  P  T  L  E  S  D  S  M  Y  I  A  K  R  G  Y  C  Y  M 892
2701 CGAUUUGCCG ACUCUUGAGU CUGACUCAAU GUAUAUAGCU AAGAGAGGCU AUUGCUACAU

      N  I  F  L  A  M  L  I  N  I  P  E  N  E  A  K  D  F  T  K 912
2761 GAACAUCUUC CUGGCGAUGC UCAUAAACAU ACCUGAGAAU GAAGCGAAGG ACUUUACAAA

      R  V  R  D  L  V  G  S  K  L  G  E  W  P  T  M  L  D  V  A 932
2821 GAGAGUUCGU GAUCUUGUAG GUUCAAAACU UGGUGAGUGG CCAACGAUGU UAGAUGUCGC

      T  C  A  N  Q  L  V  I  F  H  P  D  A  A  N  A  E  L  P  R 952
2881 AACAUGCGCC AACCAACUGG UUAUCUUUCA UCCUGAUGCA GCCAAUGCGG AAUUGCCACG

      I  L  V  D  H  R  Q  K  T  M  H  V  I  D  S  F  G  P  V  D 972
2941 AAUCUUAGUG GAUCAUCGAC AGAAAACAAU GCAUGUUAUU GAUUCGUUUG GACCUGUAGA

      S  G  Y  H  I  L  K  A  N  T  V  N  Q  L  I  Q  F  A  R  E 992
3001 CUCUGGAUAU CACAUAUUGA AGGCGAACAC AGUGAAUCAG UUGAUUCAGU UUGCCAGAGA
                                    HC-Pro/P3
      P  L  D  S  E  M  K  H  Y  I  V  G/ G  E  F  D  P  T  T  S 1012
3061 GCCACUUGAU AGUGAGAUGA AGCACUACAU UGUUGGUGGA GAGUUUGAUC CAACUACUAG

      C  L  H  Q  L  I  R  V  I  Y  K  P  H  E  L  R  N  L  L  R 1032
3121 UUGCUUGCAU CAGUUGAUAC GCGUCAUCUA UAAACCCCAU GAACUCCGCA AUCUGCUCAG

      N  E  P  Y  L  I  V  I  A  L  M  S  P  S  V  L  L  T  L  F 1052
3181 AAAUGAACCA UAUCUAAUAG UGAUUGCAUU GAUGUCGCCA AGUGUACUUU UGACUUUGUU

      N  S  G  A  I  E  H  A  L  N  Y  W  I  K  R  D  Q  D  V  V 1072
3241 UAAUAGUGGC GCAAUUGAGC AUGCGUUGAA UUAUUGGAUC AAAAGGGACC AAGAUGUUGU

      E  V  I  V  L  V  E  Q  L  C  R  K  V  T  L  A  R  T  I  L 1092
3301 UGAGGUUAUU GUCUUGGUGG AGCAAUUGUG UAGGAAAGUG ACGCUUGCUA GAACAAUUUU

      E  Q  F  N  E  I  R  Q  N  A  R  D  I  H  E  L  M  D  R  N 1112
3361 GGAGCAAUUC AAUGAGAUUC GCCAAAAUGC ACGAGAUAUA CACGAGCUAA UGGAUCGAAA

      N  K  P  W  I  S  Y  D  R  S  L  E  L  L  S  V  Y  A  D  S 1132
3421 UAAUAAACCU UGGAUUUCAU ACGAUCGAUC AUUGGAGCUA CUAAGUGUGU AUGCGGAUUC
```

Figure 30 (continued)

```
      Q  L  T   D  E  G   L  L  K  Q   G  F  S   T  L  D   P  R  L  R 1152
3481 GCAAUUAACG GAUGAGGGCC UGCUCAAGCA AGGUUUUUCA ACACUGGACC CUAGACUGCG

      E  A  V   E  K  T   Y  A  A  L   L  Q  E   E  W  R   A  L  S  L 1172
3541 CGAAGCUGUU GAAAAAACCU ACGCCGCUCU UUUGCAGGAA GAGUGGCGUG CUUUAAGUUU

      F  Q  K   L  H  L   R  Y  F  A   F  K  S   Q  P  S   F  S  E  C 1192
3601 GUUUCAAAAG UUGCACUUAA GGUACUUUGC AUUCAAAUCA CAACCAUCUU UUUCCGAGUG

      L  K  P   K  G  R   A  D  L  K   I  V  N   D  F  S   P  K  Y  C 1212
3661 UUUAAAGCCG AAAGGGCGCG CAGAUUUGAA AAUUGUGAAC GACUUCUCAC CGAAGUACUG

      V  H  E   V  G  K   A  L  L  Q   P  I  K   A  G  A   E  I  T  S 1232
3721 UGUACACGAG GUCGGAAAGG CGUUGCUCCA ACCAAUUAAA GCUGGAGCUG AAAUCACAUC

      R  L  L   N  G  C   G  T  F  I   R  K  S   A  A  R   G  C  A  Y 1252
3781 ACGUCUCCUA AACGGUUGUG GAACGUUUAU CCGGAAAAGU GCCGCAAGGG GAUGUGCAUA

      I  F  K   D  L  F   Q  F  V  H   V  V  L   V  L  S   I  L  L  Q 1272
3841 CAUUUUUAAG GACCUCUUCC AGUUUGUUCA UGUGGUGUUG GUGUUAAGCA UUUUGUUACA

      I  F  R   S  A  Q   G  I  A  M   E  H  I   Q  L  K   Q  A  K  A 1292
3901 GAUAUUCAGG AGCGCACAAG GGAUUGCCAU GGAGCAUAUA CAACUAAAGC AAGCAAAGGC

      E  A  V   K  Q  K   D  F  D  R   L  E  A   L  Y  A   E  L  C  V 1312
3961 GGAAGCGGUA AAGCAGAAAG AUUUUGAUCG GUUGGAAGCU UUAUACGCGG AACUGUGCGU

      K  N  G   D  Q  P   T  T  E  E   F  L  D   F  V  M   E  R  E  P 1332
4021 UAAGAAUGGU GAUCAGCCGA CGACUGAAGA AUUCCUUGAU UUUGUGAUGG AGCGUGAACC
                                                          P3/6K1
      R  L  K   D  Q  A   Y  S  L  I   H  I  P   A  I  H   Q/ A  K  S 1352
4081 AAGGCUCAAG GAUCAAGCUU AUAGCUUGAU UCACAUACCA GCCAUUCAUC AAGCAAAAUC

      D  N  E   K  K  L   G  Q  V  I   A  F  I   T  L  I   L  M  M  I 1372
4141 AGACAAUGAA AAGAAACUUG GACAGGUCAU UGCAUUCAUA ACAUUAAUUU UAAUGAUGAU

      D  V  D   K  S  D   C  V  Y  R   I  L  N   K  F  K   G  V  I  N 1392
4201 CGACGUGGAU AAGAGCGACU GUGUCUAUCG AAUCUUGAAC AAAUUUAAAG GCGUGAUAAA
                                  6K1/CI
      S  C  N   T  N  V   Y  H  Q/ S   L  D  D   I  R  D   F  Y  E  D 1412
4261 CUCUUGCAAC ACAAAUGUUU AUCACCAGUC UUUAGAUGAU AUUCGAGAUU UCUAUGAAGA

      K  Q  L   T  I  D   F  D  I  T   G  E  N   Q  I  N   R  G  P  I 1432
4321 CAAGCAAUUA ACCAUCGAUU UUGACAUCAC UGGUGAGAAU CAGAUCAACA GAGGGCCUAU

      D  V  T   F  E  K   W  W  D  N   Q  L  S   N  N  N   T  I  G  H 1452
4381 AGAUGUCACG UUCGAGAAAU GGUGGGACAA UCAAUUGUCC AACAACAACA CAAUUGGCCA

      Y  R  I   G  G  T   F  V  E  F   S  R  S   N  A  A   T  V  A  S 1472
4441 UUACCGAAUU GGGGGAACGU UCGUUGAAUU UUCACGGAGC AACGCAGCCA CUGUGGCUAG

      E  I  A   H  S  P   E  R  E  F   L  V  R   G  A  V   G  S  G  K 1492
4501 UGAAAUAGCC CACAGUCCUG AACGCGAAUU UCUAGUUCGC GGAGCUGUUG GUAGUGGUAA

      S  T  N   L  P  F   L  L  S  K   H  G  S   V  L  L   I  E  P  T 1512
4561 AUCAACAAAC CUUCCUUUCU UACUUAGUAA GCAUGGUAGU GUGUUGUUAA UAGAGCCCAC

      R  P  L   C  E  N   V  C  K  Q   L  R  G   E  P  F   H  C  N  Q 1532
4621 CAGACCCCUU UGCGAAAAUG UUUGCAAGCA AUUGCGCGGU GAGCCGUUUC AUUGUAAUCA
```

Figure 30 (continued)

```
      T  I  R  M  R  G  L  T  A  F  G  S  T  N  I  T  I  M  T  S 1552
4681 AACCAUUCGC AUGCGUGGGU UAACGGCCUU UGGUUCGACC AAUAUCACUA UCAUGACGAG

      G  F  A  L  H  Y  Y  A  H  N  L  Q  Q  L  R  L  F  D  F  I 1572
4741 UGGAUUUGCU UUACACUAUU AUGCUCAUAA CCUUCAGCAA UUAAGGCUCU UUGAUUUCAU

      I  F  D  E  C  H  V  I  D  S  Q  A  M  A  F  Y  C  L  M  E 1592
4801 AAUUUUCGAU GAAUGUCAUG UUAUAGAUAG UCAAGCAAUG GCUUUCUACU GCCUGAUGGA

      G  N  A  I  E  K  K  V  L  K  V  S  A  T  P  P  G  R  E  V 1612
4861 AGGGAACGCG AUUGAGAAGA AAGUUCUUAA AGUUUCUGCC ACGCCACCUG GGCGUGAGGU

      E  F  S  T  Q  F  P  T  K  I  V  T  E  R  S  I  S  F  K  Q 1632
4921 UGAAUUCUCA ACGCAAUUCC CAACAAAAAU UGUCACCGAG CGAUCCAUAA GUUUCAAACA

      L  V  D  N  F  G  T  G  A  N  S  D  V  T  A  F  A  D  N  I 1652
4981 GCUGGUUGAU AAUUUCGGCA CUGGCGCGAA UAGCGAUGUG ACUGCCUUCG CAGACAACAU

      L  V  Y  V  A  S  Y  N  E  V  D  Q  L  S  K  L  L  S  D  K 1672
5041 ACUGGUUUAU GUAGCAAGCU AUAAUGAAGU GGAUCAACUA AGUAAGCUUU UAUCCGAUAA

      G  Y  L  V  T  K  I  D  G  R  T  M  K  A  G  K  T  E  I  S 1692
5101 AGGCUACUUA GUUACUAAAA UCGAUGGGCG GACAAUGAAG GCUGGAAAAA CUGAAAUUUC

      T  S  G  T  K  S  K  K  H  F  I  V  A  T  N  I  I  E  N  G 1712
5161 AACUAGUGGC ACAAAAUCUA AGAAGCAUUU CAUAGUUGCC ACUAACAUUA UUGAAAAUGG

      V  T  L  D  I  E  A  V  I  D  F  G  M  K  V  V  P  E  M  D 1732
5221 CGUCACACUU GACAUAGAAG CUGUGAUUGA UUUUGGAAUG AAAGUCGUGC CUGAAAUGGA

      S  D  N  R  M  I  R  Y  S  K  Q  A  I  S  F  G  E  R  I  Q 1752
5281 CUCGGAUAAU CGCAUGAUCA GAUACUCGAA GCAAGCGAUC AGCUUUGGAG AGAGAAUUCA

      R  L  G  R  V  G  R  H  K  E  G  I  A  L  R  I  G  H  T  E 1772
5341 ACGGCUUGGC CGCGUUGGAA GACAUAAAGA GGGAAUUGCA CUUAGAAUCG GUCACACUGA

      K  G  I  Q  E  I  P  E  M  A  A  T  E  A  A  F  L  S  F  T 1792
5401 GAAAGGCAUU CAAGAAAUCC CAGAAAUGGC GGCUACUGAG GCGGCUUUCC UGAGCUUCAC

      Y  G  L  P  V  M  T  H  N  V  G  L  S  L  L  K  N  C  T  V 1812
5461 GUAUGGCUUG CCUGUCAUGA CUCAUAAUGU GGGACUGAGU UUACUCAAGA ACUGCACCGU

      R  Q  A  R  T  M  Q  Q  Y  E  L  S  P  F  F  T  Q  N  L  V 1832
5521 GAGGCAAGCA CGUACAAUGC AGCAGUAUGA GUUAAGCCCA UUUUUCACAC AAAAUUUGGU

      N  F  D  G  T  V  H  P  K  I  D  V  L  L  R  P  Y  K  L  R 1852
5581 UAACUUCGAU GGAACAGUGC ACCCCAAGAU CGAUGUGUUG UUGCGUCCUU ACAAAUUGAG

      D  C  E  V  R  L  S  E  A  A  I  P  H  G  V  Q  S  I  W  M 1872
5641 GGAUUGUGAA GUCAGAUUAA GCGAAGCAGC GAUACCGCAU GGAGUACAAU CUAUUUGGAU

      S  A  R  E  Y  E  A  V  G  G  R  L  C  L  E  G  D  V  R  I 1892
5701 GUCCGCCAGG GAGUAUGAAG CUGUUGGUGG ACGCCUUUGU CUGGAAGGUG AUGUUCGGAU

      P  F  L  I  K  D  V  P  E  R  L  Y  K  E  L  W  D  I  V  Q 1912
5761 ACCAUUCUUG AUCAAGGAUG UGCCUGAGCG AUUGUACAAG GAACUGUGGG ACAUUGUCCA

      T  Y  K  R  D  F  T  F  G  R  I  N  S  V  S  A  G  K  I  A 1932
5821 AACUUACAAG CGUGACUUCA CAUUUGGGCG AAUUAAUUCU GUAUCCGCUG GAAAAAUUGC
```

Figure 30 (continued)

```
      Y  T  L   R  T  D   V  Y  S  I   P  R  T   L  I  T   I  D  K  L 1952
5881 AUACACAUUA AGAACUGAUG UGUAUUCAAU UCCCAGAACU CUCAUAACGA UUGAUAAACU

      I  E  S   E  N  M   E  H  A  H   F  K  A   M  T  S   C  T  G  L 1972
5941 GAUCGAGAGU GAAAACAUGG AGCACGCUCA CUUUAAAGCC AUGACAAGUU GUACUGGUCU

      N  S  S   F  S  L   L  G  V  I   N  T  I   Q  G  R   Y  L  V  D 1992
6001 UAACUCUAGU UUCUCUCUCC UUGGUGUCAU AAACACUAUU CAGGGUAGAU ACUUGGUUGA

      H  S  V   E  N  I   R  K  L  Q   L  A  K   A  Q  I   Q  Q  L  E 2012
6061 CCAUUCGGUU GAAAAUAUCA GAAAGCUUCA GCUGGCGAAG GCCCAAAUUC AACAACUUGA

      A  H  V   Q  E  N   N  V  E  N   L  I  Q   S  L  G   A  V  R  A 2032
6121 AGCUCACGUG CAGGAAAACA AUGUUGAAAA UCUGAUUCAA UCUCUUGGUG CAGUGAGAGC
                CI/6K2
      V  Y  H   Q/ S  V   D  G  V  K   H  I  K   R  E  L   G  L  K  G 2052
6181 CGUUUAUCAU CAGAGUGUGG AUGGAGUCAA ACACAUAAAG CGAGAGUUGG GUUUAAAAGG

      V  W  D   G  S  L   M  I  K  D   A  I  V   C  G  F   T  M  A  G 2072
6241 AGUCUGGGAU GGCUCAUUGA UGAUUAAGGA UGCAAUUGUA UGCGGCUUUA CAAUGGCUGG

      G  A  M   L  L  Y   Q  H  F  R   D  K  L   T  N  I   H  V  F  H 2092
6301 UGGUGCGAUG CUAUUGUACC AACAUUUUCG UGAUAAGCUC ACCAAUAUUC AUGUGUUUCA
     6K2/VPg
      Q/ G  F   S  A  R   Q  R  Q  K   L  R  F   K  S  A   A  N  A  K 2112
6361 UCAAGGUUUC UCUGCGCGAC AGCGACAAAA GUUAAGAUUC AAGUCUGCAG CAAAUGCUAA

      L  G  R   E  V  Y   G  D  D  G   T  I  E   H  Y  F   G  E  A  Y 2132
6421 ACUUGGUCGA GAAGUUUACG GGGAUGAUGG AACAAUUGAG CAUUACUUCG GAGAGGCGUA

      T  K  K   G  N  K   K  G  R  M   H  G  M   G  V  K   T  R  K  F 2152
6481 CACGAAGAAG GGGAACAAGA AAGGAAGGAU GCACGGCAUG GGCGUUAAGA CAAGGAAAUU

      V  A  T   Y  G  F   K  P  E  D   Y  S  Y   V  R  Y   L  D  P  L 2172
6541 CGUUGCUACA UACGGAUUUA AACCAGAGGA UUAUUCAUAU GUGCGUUAUC UGGACCCCUU

      T  G  E   T  L  D   E  S  P  Q   T  D  I   S  M  V   Q  E  H  F 2192
6601 AACAGGUGAG ACAUUAGAUG AGAGUCCUCA GACUGACAUC UCAAUGGUGC AAGAACAUUU

      G  D  I   R  S  K   Y  M  D  S   D  S  F   D  R  Q   A  L  I  A 2212
6661 UGGUGACAUU AGAAGUAAAU ACAUGGACUC GGACAGCUUC GAUCGGCAGG CAUUGAUAGC

      N  N  V   I  K  A   Y  Y  V  R   N  S  A   K  T  A   L  E  V  D 2232
6721 AAAUAACGUG AUUAAAGCCU AUUACGUCCG GAACUCCGCC AAAACUGCAU UAGAAGUUGA

      L  T  P   H  N  P   L  K  V  C   D  T  K   L  T  I   A  G  F  P 2252
6781 UUUAACACCA CAUAACCCAC UCAAGGUUUG UGAUACUAAG UUGACGAUCG CAGGGUUUCC

      D  R  E   A  E  L   R  Q  T  G   P  P  R   T  I  L   A  D  Q  V 2272
6841 UGAUAGAGAA GCUGAAUUGA GGCAAACAGG CCCACCGAGA ACAAUCUUAG CUGACCAAGU
                                      VPg/NIa-Pro
      P  P  P   T  K  S   V  H  H  E/  G  K  S   L  C  Q   G  M  R  N 2292
6901 UCCACCACCA ACAAAGUCUG UUCAUCACGA AGGAAAGAGU CUCUGCCAAG GUAUGAGAAA

      Y  N  G   I  A  S   V  V  C  H   L  K  S   T  S  G   D  G  R  S 2312
6961 UUACAAUGGC AUAGCUUCUG UGGUUUGCCA UUUAAAGAGC ACAUCAGGAG AUGGUCGAAG

      L  F  G   V  G  Y   N  S  F  I   I  T  N   R  H  L   F  K  E  N 2332
7021 CUUAUUUGGG GUCGGUUACA ACUCUUUCAU CAUUACAAAC CGACAUUUGU UCAAAGAGAA
```

Figure 30 (continued)

```
      N  G  E  L  I  V  K  S  Q  H  G  K  F  V  V  K  N  T  S  T 2352
7081 UAAUGGUGAG CUUAUUGUGA AAUCCCAACA UGGAAAGUUU GUUGUUAAGA ACACUUCAAC

      L  R  I  A  P  I  G  K  T  D  L  L  I  I  R  M  P  K  G  F 2372
7141 ACUUCGGAUA GCUCCGAUUG GAAAGACUGA UCUUCUUAUC AUUAGGAUGC CAAAGGGUUU

      P  P  F  H  S  R  A  R  F  R  A  M  K  A  G  D  K  V  C  M 2392
7201 UCCACCAUUC CAUAGCAGGG CAAGGUUCAG AGCCAUGAAA GCUGGUGACA AGGUUUGUAU

      I  G  V  D  Y  Q  E  N  H  I  A  S  K  V  S  E  T  S  I  I 2412
7261 GAUUGGUGUA GAUUACCAAG AAAACCAUAU CGCUAGUAAG GUGUCUGAGA CCUCAAUCAU

      S  E  G  T  G  E  F  G  C  H  W  I  S  T  N  D  G  D  C  G 2432
7321 UAGUGAGGGC ACGGGGGAAU UUGGUUGCCA CUGGAUAUCC ACGAAUGAUG GUGACUGUGG

      N  P  L  V  S  V  S  D  G  F  I  V  G  L  H  S  L  S  T  S 2452
7381 UAAUCCACUG GUCAGCGUCU CCGAUGGCUU CAUAGUGGGA CUGCACAGUC UAUCGACCUC

      T  G  D  Q  N  F  F  A  K  I  P  A  F  F  E  E  K  V  L  R 2472
7441 GACUGGUGAU CAAAAUUUCU UUGCAAAAAU ACCAGCAUUC UUUGAGGAGA AGGUUCUCAG

      R  I  D  D  L  T  W  S  K  H  W  S  Y  N  V  N  E  L  S  W 2492
7501 GCGGAUUGAU GACCUGACAU GGAGCAAACA CUGGAGCUAC AAUGUUAAUG AAUUGAGCUG

      G  A  L  K  V  W  E  S  R  P  E  A  I  F  N  A  Q  K  E  I 2512
7561 GGGGGCUUUG AAAGUGUGGG AGAGCCGGCC AGAGGCUAUU UUUAACGCGC AGAAGGAAAU
                        NIa-Pro/NIb
      D  Q  L  N  V  F  E  Q/ S  G  S  R  W  L  F  D  K  L  H  G 2532
7621 CGAUCAAUUA AAUGUAUUUG AGCAGAGUGG AAGUCGAUGG CUCUUUGACA AGCUGCAUGG

      N  L  K  G  V  S  S  A  P  S  N  L  V  T  K  H  V  V  K  G 2552
7681 CAACCUAAAG GGAGUAAGCU CAGCUCCCAG CAAUUUAGUC ACGAAGCACG UUGUGAAAGG

      I  C  P  L  F  R  S  Y  L  E  C  D  E  E  A  K  A  F  F  S 2572
7741 AAUCUGUCCU CUCUUUAGGA GUUACCUUGA GUGUGAUGAA GAAGCCAAGG CUUUCUUUAG

      P  L  M  G  H  Y  M  K  S  V  L  S  K  E  A  Y  V  K  D  L 2592
7801 UCCGCUGAUG GGACACUACA UGAAGAGCGU UCUAAGCAAG GAGGCAUACG UGAAGGAUUU

      L  K  Y  S  S  D  I  I  V  G  E  V  D  H  D  V  F  E  D  S 2612
7861 AUUGAAAUAU UCGAGUGACA UAAUUGUUGG CGAAGUUGAC CACGAUGUGU UUGAGGAUAG

      V  A  Q  V  I  E  L  L  N  D  H  E  C  P  E  L  E  Y  V  T 2632
7921 CGUCGCACAA GUUAUCGAGU UGCUAAAUGA UCACGAGUGU CCUGAGCUCG AAUACGUCAC

      D  S  E  V  I  I  Q  A  L  N  M  D  A  A  V  G  A  L  Y  T 2652
7981 AGACAGCGAG GUCAUCAUAC AAGCUCUAAA CAUGGAUGCA GCGGUCGGAG CUUUGUACAC

      G  K  K  R  K  Y  F  E  G  S  T  V  E  H  R  Q  A  L  V  R 2672
8041 AGGAAAGAAG AGAAAAUACU UCGAAGGAUC AACAGUUGAG CACAGACAGG CCCUAGUUCG

      K  S  C  E  R  L  Y  E  G  R  M  G  V  W  N  G  S  L  K  A 2692
8101 GAAAAGUUGC GAGCGUCUUU AUGAAGGGCG GAUGGGUGUU UGGAAUGGAU CGUUGAAGGC

      E  L  R  P  A  E  K  V  L  A  K  K  T  R  S  F  T  A  A  P 2712
8161 CGAGCUGCGA CCUGCAGAGA AAGUUCUAGC CAAAAAGACA AGAUCUUUCA CAGCAGCGCC

      L  D  T  L  L  G  A  K  V  C  V  D  D  F  N  N  W  F  Y  S 2732
8221 UCUCGACACU UUGUUGGGAG CUAAAGUUUG CGUUGAUGAU UUCAACAAUU GGUUCUACAG
```

Figure 30 (continued)

```
      K  N  M  E  C  P  W  T  V  G  M  T  K  F  Y  K  G  W  D  E 2752
8281 CAAAAAUAUG GAGUGUCCCU GGACUGUUGG CAUGACGAAG UUUUACAAAG GCUGGGACGA

      F  L  R  K  F  P  D  S  W  V  Y  C  D  A  D  G  S  Q  F  D 2772
8341 AUUUUUGAGG AAAUUCCCUG AUAGUUGGGU UUACUGCGAU GCAGAUGGAU CUCAGUUCGA

      S  S  L  T  P  Y  L  L  N  A  V  L  S  I  R  L  W  A  M  E 2792
8401 UAGUUCUUUG ACACCAUAUU UGCUUAAUGC UGUGUUGUCU AUCCGACUGU GGGCUAUGGA

      D  W  D  I  G  A  Q  M  L  K  N  L  Y  G  E  I  T  Y  T  P 2812
8461 GGAUUGGGAC AUUGGAGCUC AGAUGCUUAA GAACUUGUAU GGAGAAAUCA CUUACACACC

      I  L  T  P  D  G  T  I  V  K  K  F  K  G  N  N  S  G  Q  P 2832
8521 UAUUUUGACA CCUGAUGGGA CGAUUGUCAA GAAGUUUAAA GGAAACAACA GUGGCCAACC

      S  T  V  V  D  N  T  L  M  V  L  I  T  M  Y  Y  A  L  R  K 2852
8581 CUCAACAGUU GUCGAUAAUA CAUUGAUGGU UUUAAUCACA AUGUAUUACG CACUACGCAA

      A  G  Y  D  T  E  A  Q  E  D  M  C  V  F  Y  I  N  G  D  D 2872
8641 GGCUGGUUAC GACACGGAGG CCCAGGAAGA UAUGUGCGUA UUCUAUAUCA AUGGUGAUGA

      L  C  I  A  I  H  P  D  H  E  H  V  L  D  S  F  S  S  S  F 2892
8701 UCUCUGUAUC GCCAUCCACC CUGAUCACGA GCAUGUUCUU GACUCAUUCU CUAGUUCAUU

      A  E  L  G  L  K  D  D  F  T  Q  R  H  R  N  K  Q  D  L  W 2912
8761 UGCUGAGCUU GGACUCAAAG AUGAUUUCAC UCAAAGACAC CGGAAUAAAC AAGAUCUGUG

      F  M  S  H  R  G  I  L  I  D  D  I  Y  I  P  K  L  E  P  E 2932
8821 GUUCAUGUCG CAUCGAGGUA UUCUGAUCGA UGACAUUUAC AUUCCUAAAC UUGAACCCGA

      R  I  V  A  I  L  E  W  D  K  S  K  L  P  E  H  R  L  E  A 2952
8881 GAGAAUUGUG GCCAUCCUUG AGUGGGACAA AUCUAAGCUU CCGGAACAUC GGUUGGAAGC

      I  T  A  A  M  I  E  S  W  G  Y  G  E  L  T  H  Q  I  R  R 2972
8941 UAUUACAGCA GCAAUGAUAG AGUCAUGGGG AUACGGGGAG UUAACGCACC AGAUCCGCAG

      F  Y  Q  W  V  L  E  Q  A  P  F  D  E  L  A  E  Q  G  R  A 2992
9001 AUUUUAUCAA UGGGUACUCG AGCAGGCUCC AUUCGAUGAG UUAGCGGAGC AGGGCAGGGC

      P  Y  V  S  E  V  G  L  R  R  L  Y  T  S  E  R  G  S  M  D 3012
9061 UCCAUAUGUG UCUGAAGUUG GAUUGAGGCG CUUGUACACU AGUGAGCGUG GGUCAAUGGA

      E  L  E  A  Y  I  D  K  Y  F  E  R  E  R  G  D  S  P  E  L 3032
9121 CGAAUUGGAG GCCUACAUAG AUAAAUAUUU UGAGCGAGAG AGAGGAGAUU CGCCUGAACU

      L  V  Y  H  E  S  R  S  T  D  D  Y  Q  L  I  C  G  E  S  T 3052
9181 AUUGGUGUAU CAUGAAUCCA GAAGUACUGA UGAUUACCAG CUUAUUUGCG GUGAGAGCAC
                NIb/CP
      H  V  F  H  Q/ S  K  N  E  A  V  D  A  G  L  N  E  K  F  K 3072
9241 ACACGUGUUC CACCAGUCCA AAAAUGAAGC UGUGGAUGCU GGUCUGAAUG AAAAGUUCAA

      E  K  E  K  Q  K  E  E  K  D  K  Q  K  D  K  N  N  D  G  A 3092
9301 AGAGAAAGAG AAACAAAAAG AAGAAAAAGA UAAACAAAAA GAUAAGAAUA AUGAUGGAGC

      S  D  G  N  D  V  S  T  S  T  K  T  G  E  R  D  R  D  V  N 3112
9361 UAGUGACGGA AACGAUGUGU CAACUAGCAC AAAAACUGGA GAGAGAGAUA GAGAUGUCAA

      A  G  T  S  G  T  F  T  V  P  R  M  K  S  F  T  N  K  M  I 3132
9421 UGCUGGAACU AGUGGAACUU UCACUGUUCC AAGAAUGAAG UCAUUUACUA AUAAGAUGAU
```

Figure 30 (continued)

```
       L  P  R  I  K  G  K  T  V  L  N  L  N  H  L  L  Q  Y  N  P 3152
9481 UUUACCAAGA AUUAAGGGAA AAACUGUCCU UAAUUUGAAU CAUCUUCUUC AGUACAACCC

       Q  Q  I  D  I  S  N  T  R  A  T  Q  S  Q  F  E  K  W  Y  E 3172
9541 GCAACAGAUU GACAUCUCAA ACACUCGUGC CACUCAGUCU CAAUUUGAGA AGUGGUAUGA

       G  V  R  N  D  Y  G  L  N  G  N  E  M  Q  V  M  L  N  G  L 3192
9601 GGGAGUGAGG AAUGAUUAUG GUCUUAAUGG CAAUGAGAUG CAGGUAAUGU UGAAUGGUUU

       M  V  W  C  I  E  N  G  T  S  P  D  I  S  G  V  W  V  M  V 3212
9661 GAUGGUUUGG UGUAUCGAAA AUGGUACAUC UCCAGACAUA UCUGGUGUCU GGGUUAUGGU

       D  G  E  T  Q  V  D  Y  P  I  R  P  L  I  E  H  A  T  P  S 3232
9721 GGAUGGGGAA ACCCAAGUCG AUUAUCCUAU CAGACCUUUG AUUGAGCACG CAACUCCUUC

       F  R  Q  I  M  A  H  F  S  N  A  A  E  A  Y  I  A  K  R  N 3252
9781 AUUCAGGCAA AUCAUGGCUC ACUUCAGCAA CGCGGCAGAG GCAUACAUUG CGAAAAGGAA

       A  T  E  R  Y  M  P  R  Y  G  I  K  R  N  L  T  D  I  S  L 3272
9841 UGCAACUGAG AGGUACAUGC CGCGGUAUGG AAUCAAGAGG AAUUUGACUG ACAUUAGCCU

       A  R  Y  A  F  D  F  Y  E  V  N  S  K  T  P  D  R  A  R  E 3292
9901 CGCUAGAUAU GCUUUCGAUU UCUAUGAGGU GAAUUCAAAA ACACCUGAUA GGGCUCGCGA

       A  H  M  Q  M  K  A  A  A  L  R  N  T  G  R  R  M  F  G  M 3312
9961 AGCUCAUAUG CAGAUGAAAG CUGCAGCACU ACGCAAUACU GGUCGUAGAA UGUUUGGAAU

       D  G  S  V  S  N  K  E  E  N  T  E  R  H  T  V  E  D  V  N 3332
10021 GGACGGCAGU GUCAGUAACA AGGAGGAAAA CACGGAGAGA CACACAGUGG AAGAUGUCAA
                                      CP/3'-NCR
       K  D  M  H  S  L  L  G  M  R  N                          3343
10081 CAAAGACAUG CACUCUCUCC UGGGUAUGCG CAAUUGAAUA CUCGCGCUAG UGUGUUCGUC

10141 GGGCCUGGCU CGACCCUGUU UCACCUUAUA AUACUAUAUA AGCAUUAGAA UACAGUGUGG

10201 CUGCGCCACC GCUUCUAUUU UACAGUGAGG GUAGCCCUCC GUGCUUUUAG UGUUAUUCGA

10261 GUUCUCUGAG UCUCCAUACA GUGUGGGUGG CCCACGUGCU AUUCGAGCCU CUUAGAAUGA

10321 GAG
```

Figure 30 (continued)

nucleotides (Taiwanese isolate PRSV-W-TW; GenBank accession number AY027810). The variation in length of the twelve different PRSV genomes fully-sequenced is reflected by the length of coat protein (CP) coding region, which range from 855 nucleotides (PRSV-P-MEX) to 867 nucleotides (PRSV-W-TW and Brazilian isolate PRSV-W-1; GenBank accession number DQ374153). Among the twelve PRSV isolates, PRSV-W-TW 5' non coding region (5'-NCR) consisting 87 nucleotides is 2 nucleotides larger than those of all the eleven fully-sequenced isolates and this reflects for the largest genome of the virus. The RNA genome of the PRSV-W isolate is identical in size to the two PRSV isolates originated from Thailand (cucurbit-infecting PRSV-W-ThRb; GenBank accession number AY010722 and papaya-infecting PRSV-P-TH; GenBank accession number AY010722) (Attasart *et al*., 2002; Charoensilp *et al*., 2003), when compared with eleven PRSV isolates for which the complete genomic sequence is available. It presents all the characteristic features of potyviruses, with a 5' non-coding region (5' NCR), a single open reading frame (ORF) and a 3' non-coding region (3' NCR). The overall nucleotide composition of the viral RNA is 30.33% adenine, 18.96% cytosine, 24.55% guanine and 26.16% uracil, which is similar to those of the other PRSV isolates and also similar to those of the potyviruses. The nucleotide composition of each putative RNA gene encoding different viral proteins and the two NCRs are shown in Table 6.

The length of the 5' non-coding region (5' NCR) of this PRSV-W RNA, 85 nucleotides long, is the smallest one among those of potyviruses (85-205 nucleotides) known so far. The 5' NCR has a significantly different nucleotide composition (45.8% A, 25.8% C, 4.7% G, 23.5% U) from that of the complete sequence. This is comparable to the leader sequences of several potyviruses which contain only 5.9 to 13.1% guanine (Johansen *et al*., 1991; Wang and Yeh, 1997). The low G content of the leader (4.7% for the PRSV-W) is characteristic of the 5' NCRs of potyviruses and other plant viruses (Gallie *et al*., 1987). The high A content (45.8%) is found in the 5' NCR of the PRSV-W. A comparison of the 5' NCR sequence of the genomic RNAs of several species of *Potyviridae* has discovered a high content of adenine and a similar motif based on different numbers of adenine and uracil (Gotz and Maiss, 2002). The high AU (69.4%) or low GC (30.6) content is found in the 5' NCR, it has

Table 6 Nucleotide composition of each putative gene including NCRs of complete genomic RNA excluding the 3' terminal poly(A) tail of the PRSV-W isolate

	Number of bases						Total number of bases	Range of sequence
	A (%)	G (%)	U (%)	C (%)	A+U (%)	C+G (%)		
5'NCR	39 (45.88)	4 (4.71)	20 (23.53)	22 (25.88)	59 (69.41)	26 (30.59)	85 (0.82)	1-85
P1	518 (31.57)	428 (26.08)	392 (23.89)	303 (18.46)	910 (55.45)	731 (44.55)	1641 (15.90)	86-1726
HC-Pro	417 (30.42)	325 (23.71)	374 (27.28)	255 (18.60)	791 (57.70)	580 (42.30)	1371 (13.28)	1727-3097
P3	302 (29.18)	252 (24.35)	286 (27.63)	195 (18.84)	588 (56.81)	447 (43.19)	1035 (10.03)	3098-4132
6K1	59 (37.82)	29 (18.59)	42 (26.92)	26 (16.67)	101 (64.74)	55 (35.26)	156 (1.51)	4133-4288
Cl	566 (29.71)	442 (23.20)	518 (27.19)	379 (19.90)	1084 (56.90)	821 (43.10)	1905 (18.45)	4289-6193
6K2	48 (28.07)	46 (26.90)	52 (30.41)	25 (14.26)	100 (58.48)	71 (41.52)	171 (1.66)	6194-6364
VPg	186 (32.80)	137 (24.16)	129 (22.75)	115 (20.28)	315 (55.56)	252 (44.44)	567 (5.49)	6365-6931
NIa-Pro	207 (28.99)	185 (25.91)	193 (27.03)	129 (18.07)	400 (56.02)	314 (43.98)	714 (6.92)	6932-7645
NIb	454 (28.18)	422 (26.19)	425 (26.38)	310 (19.24)	879 (54.56)	732 (45.44)	1611 (15.61)	7646-9256
CP	294 (34.27)	212 (24.71)	204 (23.78)	148 (17.25)	498 (58.04)	360 (41.96)	858 (8.31)	9257-10114
3'NCR	41 (19.62)	52 (24.88)	66 (31.58)	50 (23.92)	107 (51.20)	102 (48.80)	209 (2.02)	10115-10323
Full-length	3131 (30.33)	2534 (24.55)	2701 (26.16)	1957 (18.96)	5832 (56.50)	4491 (43.50)	10323 (100.00)	1-10323

been proposed that this high AU or low GC content facilitates the melting of secondary structures of RNA for efficient translation (Jobling and Gehrke, 1987; Altmann *et al*., 1990).

The 3' NCR of the PRSV-W is 209 nucleotides in length excluding the poly(A) tail which is the same length as those of other different PRSV isolates. In TVMV, Rodriguez-Cerezo *et al.* (1991) reported that a large insertion in the 3' NCR was able to dramatically attenuate symptoms on tobacco. However, the virus 3' NCR length is identical. In TEV, Haldeman-Cahill *et al*. (1998) suggested that secondary structures formed by the 3' NCR and the 3'-portion of the CP-encoded RNA are involved in viral genome amplification.

4. Putative polyprotein of the PRSV-W isolate

For translation, computer analysis of the RNA revealed a single large ORF of 10032 nucleotides. A 5' NCR of 85 nucleotides preceding the ORF was found to be the smallest one among those of potyviruses known so far. In-frame with the ORF, the first AUG at nucleotide positions 86-88 was considered to be the initiation codon and the encoded polyprotein is terminated with a UGA (termination codon) at positions 10115-10117, followed by a 3' NCR of 206 nucleotides. According to the initiation codon in-frame with the ORF, the single ORF of 10032 nucleotides was predicted to potentially encode a putative polyprotein of 3343 amino acids with a calculated molecular weight of 380295.21 Da (380.3 kDa) (Figure 30). The length of the putative complete polyprotein amino acid sequence is similar to those of other fully-sequenced PRSV isolates, which range from 3342 amino acids (PRSV-P-MEX) to 3346 amino acids (PRSV-W-TW and PRSV-W-1). The variation in length of the twelve different PRSV polyproteins fully-sequenced is reflected by the length of coat protein (CP), which range from 285 amino acids (PRSV-P-MEX) to 289 amino acids (PRSV-W-TW and PRSV-W-1). The polyprotein of the PRSV-W isolate is identical in size to the two Thai isolates of PRSV (PRSV-W-ThRb and PRSV-P-TH) when compared with those of the eleven PRSV isolates for which the complete genomic sequence is available. Based on previously predicted potyviral cleavage sites,

putative cleavage sites in the PRSV-W polyprotein were identified by amino acid sequence alignment and searching for conserved peptide sequences corresponding to the cleavage site consensus of the three virus-encoded proteinases. Based on the likely location of the cleavage sites, and based on multiple alignments of amino acid sequences for each functional protein, the polyprotein of the PRSV-W isolate is probably processed into ten putative proteins (Figure 30). Like other fully-sequenced PRSV isolates, the polyprotein of the PRSV-W isolate contains nine putative cleavage sites in polyprotein processing. Ten putative mature proteins generated from polyprotein processing of the PRSV-W isolate (Table 7) are similar in size and in the genome organization order to different putative mature proteins from other fully-sequenced PRSV isolates. The amino acid composition of each putative mature cleavage product (functional protein) proteolytic processing from PRSV-W polyprotein and its overall polyprotein are shown in Table 8.

5. Putative proteolytic cleavage sites in the putative polyprotein of the PRSV-W isolate

From the nucleotide sequence and analogy with the other potyviruses, putative proteolytic cleavage sites in the PRSV-W polyprotein were identified by searching for peptide sequences corresponding to the cleavage site consensus of the potyvirus proteinases P1-Pro, HC-Pro and NIa-Pro (Dougherty and Carrington, 1988; Riechmann *et al*., 1992). These proteinases are involved in the complete processing of the potyviral protein. The P1- and HC-proteinases autoproteolytically cleave at their carboxy-termini *in cis* between themselves and the HC-Pro and P3, respectively (Carrington *et al*., 1989a, b; Verchot *et al*., 1991; Merits *et al*., 2002; Ruiz-Ferrer *et al*., 2004) while the NIa proteinase catalyzes cleavage of the remainder of the proteins out from the polyprotein (Riechmann *et al*., 1992; Shukla *et al*., 1994; Merits *et al*., 2002).

Potyviral precursor polyproteins are cleaved at specific sites into smaller mature proteins by virus-encoded proteinases, NIa, HC-Pro and P1. By sequence alignment, putative cleavage sites in the PRSV-W polyprotein were identified (Figure

Table 7 Putative mature proteins expected after proteolytic processing from predicted polyprotein of the PRSV-W isolate

Putative protein	Range of amino acid sequence	Size in amino acid residues	Range of nucleotide sequence	Size in nucleotides
5’ NCR	-	-	1-85	85
P1	1-547	547	86-1726	1641
HC-Pro	548-1004	457	1727-3097	1371
P3	1005-1349	345	3098-4132	1035
6K1	1350-1401	52	4133-4288	156
CI	1402-2036	635	4289-6193	1905
6K2	2037-2093	57	6194-6364	171
VPg	2094-2282	189	6365-6931	567
NIa-Pro	2283-2520	238	6932-7645	714
NIb	2521-3057	537	7646-9256	1611
CP	3058-3343	286	9257-10114	858
3’ NCR	-	-	10118-10323	206

Table 8 Amino acid composition of each putative cleavage product (mature proteins) proteolytic processing from predicted polyprotein of the PRSV-W isolate

Amino acid	Functional proteins proteolytically processed from PRSV-W polyprotein										Total amino acid
	P1	HC-Pro	P3	6K1	Cl	6K2	VPg	NIa-Pro	NIb	CP	
Ala (A)	35	29	27	2	39	3	14	11	3	19	212
Arg (R)	41	29	22	1	37	2	13	11	26	20	202
Asn (N)	19	33	11	5	32	1	6	15	15	25	162
Asp (D)	26	27	17	4	31	4	14	11	40	19	193
Cys (C)	13	12	7	2	9	1	1	5	10	1	61
Gln (Q)	19	12	20	2	28	2	7	7	15	11	123
Glu (E)	46	29	26	1	42	1	11	13	44	21	234
Gly (G)	32	26	12	2	42	7	14	22	32	19	208
His (H)	15	18	10	1	18	4	6	8	16	6	102
Ile (I)	34	30	23	6	50	4	7	18	26	12	210
Leu (L)	43	38	48	4	55	6	12	16	54	15	291
Lys (K)	55	29	22	6	33	5	16	16	33	22	237
Met (M)	12	12	4	2	14	3	4	4	14	14	83
Phe (F)	21	20	17	2	32	3	8	15	23	9	150
Pro (P)	18	16	13	0	19	0	11	7	19	9	112
Ser (S)	34	26	19	3	43	2	10	23	39	17	216
Thr (T)	25	25	12	2	43	2	14	11	24	19	177
Trp (W)	5	5	3	0	4	1	0	5	12	3	38
Tyr (Y)	16	14	11	2	19	1	10	4	27	9	113
Val (V)	38	27	21	5	45	5	11	16	35	16	219
Number of amino acids	547	457	345	52	635	57	189	238	537	286	3343
Protein position	1-547	548-1004	1005-1349	1350-1401	1402-2036	2037-2093	2094-2282	2283-2520	2521-3057	3058-3343	1-3343
MW (kDa)	62.6	52.1	39.7	5.9	71.5	6.4	21.4	26.5	61.7	32.6	380.3

30) based on known and previously predicted potyviral cleavage sites, the consistency in genomic organization, and the size and uniformity of potyviral proteins. This was supported by the presence of conserved motifs in the deduced amino acid sequence of PRSV-W polyprotein which are required for the proteolytic process of potyviral polyprotein (Dougherty and Semler, 1993), such as the serine-type proteinase domain H_{456}-(8X)-D_{465}-(31X)-G_{497}-(X)-S_{499}-G_{500} and the proteolytic domain FITRGR in P1 protein at amino acid positions 519-524, C_{890}-(72X)-H_{963} in HC-Pro, and H_{2327}-(34X)-D_{2362}-(68X)-C_{2431} in NIa-Pro. The polyprotein contains nine putative cleavage sites, yielding ten putative functional proteins that have motifs conserved in homologous proteins of other potyviruses.

Based on the likely location of cleavage sites, and based on alignments of amino acid sequences for each protein, the polyprotein of PRSV-W is probably proteolytically processed into ten putative mature proteins with nine cleavage sites by the three viral encoded proteinases. These indicated that polyprotein of PRSV-W is cleaved in a manner similar to other potyviruses.

Amino acid residues in the P1-Pro cleavage site between P1 and HC-Pro of the PRSV-W are located at positions 547/548 with tyrosine (Y) and asparagine (N) which are identical to those found in 11 other PRSV isolates reported previously in GenBank. In TEV, the N-terminal protein was reported to have proteinase activity and to cleave itself from the polyprotein (Verchot *et al*., 1991) and a consensus of (Y or F)/S cleavage sites has been observed in several potyviruses (Mavankal and Rhoads, 1991).

The cleavage site catalyzing with HC-Pro between HC-Pro and P3 is located at positions between 1004 and 1005 with glycine (G) and glycine (G), conserved in all known potyviruses (Shukla *et al*. 1994) and is consistent with the proposed consensus, Y(X)VG/G (Carrington and Herndon, 1992). And the conserved amino acid sequence (MKHYIVG/G) in this cleavage site is also identical with all PRSV isolates reported previously.

Potyviral NIa proteinase is responsible for several of the polyprotein processing events (Shukla *et al*., 1994). NIa proteinase cleavage sites are defined by sequences of 7 amino acids, which have been characterized by a variety of experimental approaches (Riechmann *et al*., 1992). The NIa-Pro cleavage sites of the PRSV-W were deduced and found that there are 7 cleavage sites in the order in which cistron appear on the genome are located at positions 1349/1350, 1401/1402, 2036/2037, 2093/2094, 2282/2283, 2520/2521 and 3057/3058. These NIa-Pro cleavage positions are corresponding to the cleavage sites of P3/6K1, 6K1/CI, CI/6K2, 6K2/VPg, VPg/NIa-Pro, NIa-Pro/NIb and NIb/CP which contain amino acid residues at these sites glutamine (Q)/alanine (A), glutamine (Q)/serine (S), glutamine (Q)/serine (S), glutamine (Q)/glycine (G), glutamic acid (E)/glycine (G), glutamine (Q)/serine (S) and glutamine (Q)/serine (S), respectively. All pair of amino acid residues in each site is identical to those all of other PRSV isolates reported previously. In PRSV-W, the deduced consensus amino acid sequence of NIa-Pro cleavage sites is (V or A) (F or Y or I or H) (H or E) (Q or E) / (S or G or A) (amino acids in parentheses represent alternatives at that position) (Table 9). A consensus amino acid sequence is sufficient to define the cleavage sites recognized by NIa-Pro which usually cleaves at V-(2X)-Q/(A or E or G or S or T or V) sequences (Shukla *et al*., 1994). Generally, a consensual amino acid, valine (V), is found at position P4 in amino acid sequence of NIa-Pro cleavage sites. In PRSV-W, amino acid sequence (AIHQ/A) of NIa-Pro cleavage site at P3/6K1 is shown to differ at the position P4 from the proposed consensus (an alanine (A) replaces a consensual valine (V)). Both amino acids alanine (A) and valine (V) are classified as having nonpolar side chains, which have a hydrophobic characteristic (Voet and Voet, 1990). In *Sweetpotato mild mottle virus* (SPMMV), Colinet *et al*. (1998) found that virus-specific variations may include hydrophobic residues other than valine at the position P4. Exchange of valine for alanine at position P4 of the cleavage site is unusual. Only PRSV-W isolate (in this study) has the amino acid variation at this site, and differs from those of other 11 PRSV isolates as shown in Table 10. All sites containing a glutamine (Q) are those predicated to be cleave by the NIa proteinase. Cleavage at site of VPg/NIa-Pro is a late event separating the VPg (21K) from the proteinase (29K) (Dougherty and Parks, 1991). The VPg/NIa-Pro cleavage site of the PRSV-W contains a glutamic acid (E),

Table 9 The position of amino acid residues in the nine predicted cleavage sites of PRSV-W polyprotein and the amino acid sequence adjacent to the proposed sites.

Cleavage site	Amino acid position (from N-terminus to C-terminus)											Proteinase
	P7	P6	P5	P4	P3	P2	P1	*	P'1	P'2	P'3	
P1 / HC-Pro	R	M	R	M	E	Q	Y	*	N	D	V	P1-Pro
HC-Pro / P3	M	K	H	Y	I	V	G	*	G	E	F	HC-Pro
P3 / 6K1	H	I	P	A	I	H	Q	*	A	K	S	NIa-Pro
6K1 / CI	N	T	N	V	Y	H	Q	*	S	L	D	NIa-Pro
CI / 6K2	V	R	A	V	Y	H	Q	*	S	V	D	NIa-Pro
6K2 / VPg	N	I	H	V	F	H	Q	*	G	F	S	NIa-Pro
VPg / NIa-Pro	T	K	S	V	H	H	E	*	G	K	S	NIa-Pro
NIa-Pro / NIb	Q	L	N	V	F	E	Q	*	S	G	S	NIa-Pro
Nib / CP	S	T	H	V	F	H	Q	*	S	K	N	NIa-Pro

The cleavage site is indicated by asterisk (*)

Table 10 Alignment of putative cleavage sites in the polyprotein of PRSV isolates and adjacent amino acid sequence

PRSV isolate	Amino acid sequence (N- to C-terminus) adjacent to the putative site		
	P1 proteinase cleavge site	HC proteinase cleavage site	NIa proteinase cleavage site
	P1/HC-Pro	HC-Pro/P3	P3/6K1
Consensus	RMNMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-W	RMRMEQY / NDV	MKHYIVG / GEF	HIPAIHQ / AKS
PRSV-W-ThRb	RMRMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-W-TW	RMKMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-W-C	YAKMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-W-1	YAKMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-TH	RMRMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-HA	YAKMDQY / NDV	MKHYIVG / GEF	YIPVIHQ / AKS
PRSV-P-YK	RMNMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-MEX	YARMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-DF	RMNMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-SMN	RMNMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS
PRSV-P-SM	RMNMEQY / NDV	MKHYIVG / GEF	HIPVIHQ / AKS

PRSV isolate	Amino acid sequence (N- to C-terminus) adjacent to the putative site		
	NIa proteinase cleavage site	NIa proteinase cleavage site	NIa proteinase cleavage site
	6K1/CI	CI/6K2	6K2/VPg
Consensus	NTNVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-W	NTNVYHQ / SLD	VRAVYHQ / SVD	NIHVFHQ / GFS
PRSV-W-ThRb	NTNVYHQ / SLD	VGAVYHQ / SVD	NIHVFHQ / GFS
PRSV-W-TW	NTNVYHQ / SLD	VRAVYHQ / SVD	NVYVFHQ / GFS
PRSV-W-C	NTNVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-W-1	NTNVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-TH	DTNVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-HA	NTNVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-YK	NSDVYHQ / SLD	IRAVYHQ / GVD	SVHVFHQ / GFS
PRSV-P-MEX	DANVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-DF	NTDVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-SMN	NTDVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS
PRSV-P-SM	NTDVYHQ / SLD	VRAVYHQ / SVD	NVHVFHQ / GFS

PRSV isolate	Amino acid sequence (N- to C-terminus) adjacent to the putative site		
	NIa proteinase cleavage site	NIa proteinase cleavage site	NIa proteinase cleavage site
	VPg/NIa-Pro	NIa-Pro/NIb	Nib/CP
Consensus	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-W	TKSVHHE / GKS	QLNVFEQ / SGS	STHVFHQ / SKN
PRSV-W-ThRb	TKSVHHE / GKS	QLNVFEQ / SGS	STHVFHQ / SKN
PRSV-W-TW	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-W-C	SKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-W-1	SKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-P-TH	TKSVHHE / GKS	QLNVFEQ / SGS	DTHVFHQ / SKN
PRSV-P-HA	SKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-P-YK	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-P-MEX	SRSVHHE / GKS	QLNAFEQ / SGS	NTAVFHQ / SKN
PRSV-P-DF	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-P-SMN	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN
PRSV-P-SM	TKSVHHE / GKS	QLNVFEQ / SGS	NTHVFHQ / SKN

which is consensus among known potyviruses (Dougherty and Parks, 1991; Riechmann *et al*., 1992). The proposed genomic map for PRSV-W is shown in Figure 31. Apparently, the gene order of PRSV-W genome is identical to those of previously analyzed potyviruses. The sizes of most predicted mature proteins fall within the range established from those of other PRSV isolates.

A multiple alignment of all complete PRSV amino acid sequences available, including the 11 isolates of PRSV and the PRSV-W (in this study) sequences, was performed. The predicted cleavage sites of the polyproteins were aligned in each case, and at these positions possible cleavage sequences were also found in the PRSV-W. The putative autoproteolytic cleavage sites for P1 and HC-Pro and seven NIa proteolytic cleavage sites conserved in other potyvirus and highly conserved among PRSV isolates were identified in the polyprotein of the PRSV-W. The peptide profiles of nine putative proteolytic cleavage sites in the PRSV-W polyprotein are shown in Table 9. These cleavage sites within the putative coding region (polyprotein) of PRSV-W were compared by amino acid sequence alignment with those of 11 other PRSV isolates (Table 10). The comparison reveals similar cleavage sites and adjacent amino acid sequences for proteinase P1, HC-Pro and NIa-Pro.

6. Conserved sequence motifs

6.1 Conserved sequence motifs in non coding regions

6.1.1 5’ non coding region (5’NCR)

In most of the 5’-termini of the genomes of the *Potyviridae* are present, namely; (3-5)A-U-(2-5)A, (3-5)A-U-U-(3-4)A and (4-6)A-U-A-U-(4-5)A (Gotz and Maiss, 2002). In the PRSV-W, 5’-AAA UAA AA-3’ (the first 8 nucleotides) sequence is located at the 5’-termini of the genome. Furthermore, the first 23 nucleotides of the virus RNA within the 5’ NCR are identical among both type W and P isolates of PRSV. This consensus sequence, 5’-AAA UAA AAC AUC UCA ACA CAA CA-3’, was also conserved in other potyviruses (Yeh *et al*., 1992;

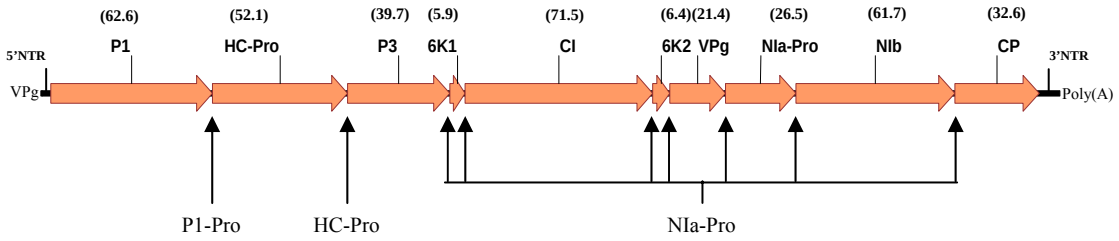


Figure 31 Proposed genome organization of *Papaya ringspot virus* type W (PRSV-W). The terminal non-coding regions (NCR) are depicted by line, and the open reading frame is represented by shaded arrowheads with the different polyprotein domains (proteins). Putative cleavage sites for the mature proteins with proteinases (P1-Pro, HC-Pro and NIa-Pro) responsible for each cleavage site are depicted below the polyprotein. The size of each protein (in kDa unit) is shown in parentheses ( ) above the protein acronym. VPg, genome-linked protein; poly(A) tail.

Shukla *et al*., 1994). The highly conserved sequence may act for binding to VPg or may act as a polymerase recognition signal for RNA replication or may act as the assembly initiation site of potyvirus particle which these activities are the common and important functions for potyviruses.

The 5' NCR contains ten CAA nucleotide repeat sequences for the PRSV-W (Figure 32) which the CAA repeats have been reported earlier and described for the 5' NCR of *Tobacco mosaic virus* (TMV) (Gallie and Walbot, 1992) and *Sugarcane mosaic virus* (SCMV-BJ, a potyvirus, GenBank accession number AY 042184) (Fan *et al*., 2003). This structure of repeats of CAA sequence was similar to the core translation regulation elements in the *Tobacco mosaic virus* (TMV) 5' leader sequence, which are demonstrated to be associated with a role translation enhancement (Gallie and Walbot, 1992).

In addition, two highly conserved regions called potyboxes "a" and "b" (Turpen, 1989; Atreya, 1992) exist within the 5' NCR. The nucleotide sequences UCA ACA CAA CAC (12 nucleotides) for "potybox a" motif and UCA AGC A (7 nucleotides) for "potybox b" motif are also present to reside in the sequence of the 5' NCR of PRSV-W between 13-24 nucleotides and between 47-53 nucleotides, respectively. The functional significance of these sequences has not been established thus far.

6.1.2 3' non coding region (3'NCR)

Like many potyviruses, the 3' NCR of PRSV-W contains a TATA box-like sequence UAUAUAA which is located at nucleotide positions 10175-10181 (Figure 32). Additionally, a directly repeated nucleotide sequence AUACAGUGUGG is eleven nucleotides in length and found in the 3' NCR of PRSV-W (Figure 32). Two locations of this repeated sequence presenting in the 3'NCR are located at nucleotides positions 10190-10200 and 10276-10286 (Figure 32).

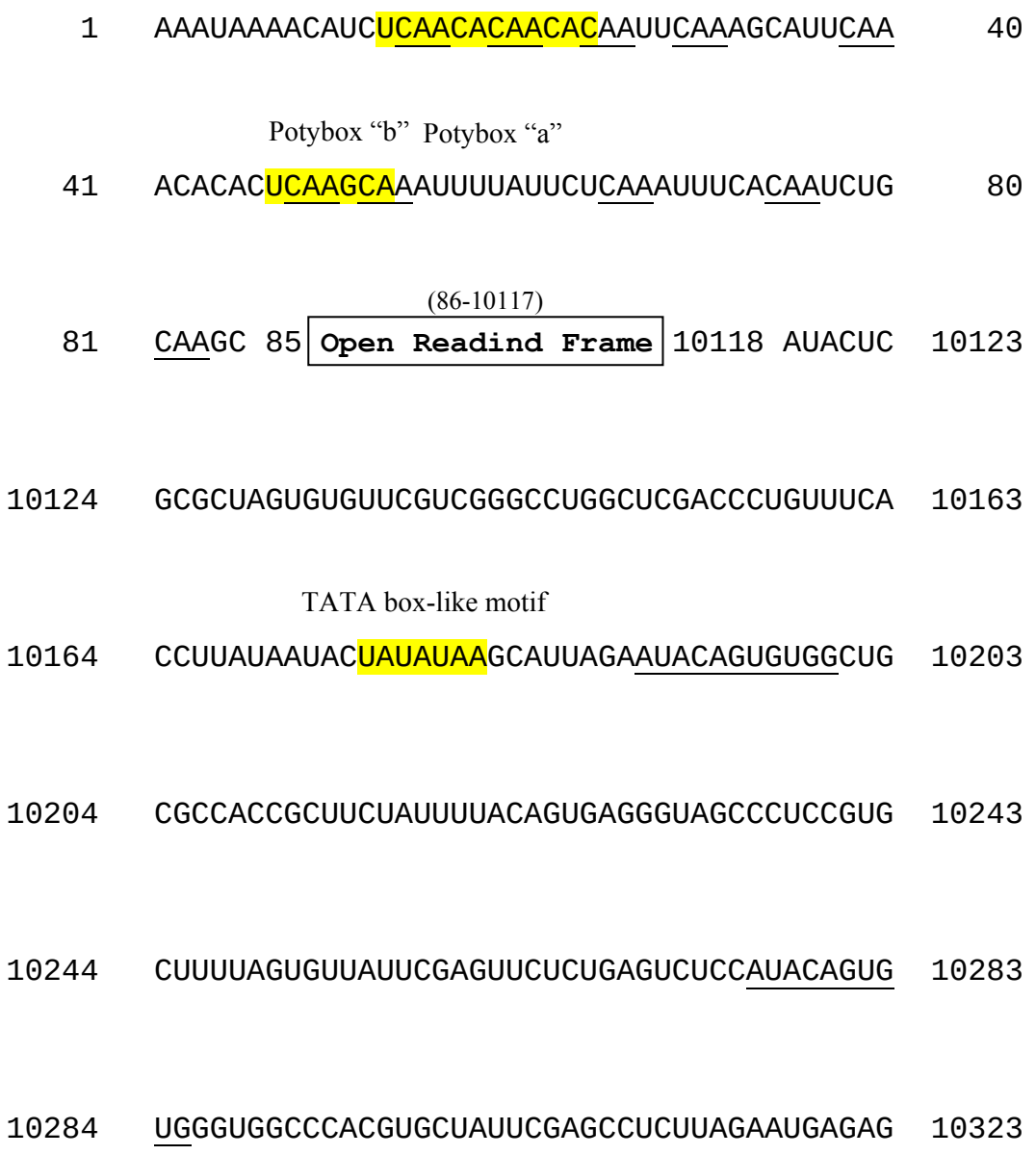


Figure 32 The nucleotide sequence of the 5' and 3' noncoding regions of PRSV-W RNA. The predicted amino acid sequence of the genome open reading frame (boxed) was omitted. Motifs within the 5'-NCR, Potybox "a" and Potybox "b", and within the 3'-NCR, TATA box-like motif, are highlighted and repeated sequences are underlined. The nucleotide position is indicated at the start and the end of each line.

6.2 Conserved sequence motifs in encoded proteins

Based on the proposed cleavage sites of gene products, ten mature proteins are predicted for PRSV-W (Figure 30) and these were identified for the important conserved amino acid sequences with proposed functions for other potyviruses.

Analysis of the deduced amino acid sequences of the P1, HC-Pro, P3, 6K, CI, NIa, NIb and CP proteins of PRSV-W genome revealed several functionally important motifs or sequences associated with virus replication, gene regulation or virus translocation. Most of the conserved amino acid sequences or motifs of potyviruses (Shukla *et al*., 1994) were detected in PRSV-W polyprotein.

6.2.1 First protein (P1)

The first (P1) protein generated from the N-terminus of the potyviral polyprotein differs most from the corresponding proteins of other *Potyviridae*. The first (P1) protein encoded by PRSV-W is composed of 547 amino acids residues (62.6 kDa), and is the second largest P1 protein ever reported. The largest is that of *Sweet potato feathery mottle virus* severe strain (SPFMV-S) composed of 664 amino acids (74 kDa) (Sakai *et al*., 1997).

The number of amino acid residues of P1 protein varied from 237 (for JGMV) to 664 (for SPFMV-S) contrasted with other encoding proteins (Shukla *et al*., 1994). The P1 of TVMV was previously assumed to be related to the movement protein of tobamoviruses based on sequence comparison (Domier *et al*., 1987) and Shukla *et al*. (1994) suggested that potyvirus P1 may be involved in a virus-host interaction. However, it is reported that this protein has a proteinase function for processing between P1 and HC-Pro suggesting the proteinase domain is in the C-terminal half of P1.

In the PRSV-W P1 protein, conserved catalytic triad composed of amino acids histidine (H), aspartic acid (D) and serine (S) in the sequence H-(8X)-D-(31X)-G-(X)-

SG (where (X) designates any unspecified amino acid) representing the active site of the P1 serine proteinase (Verchot *et al*., 1991, 1992; Riechmann *et al*., 1992) is completely conserved and located in the C-terminal P1 protein of PRSV-W at amino acid positions 456, 465 and 499 of the polyprotein, respectively. The corresponding amino acid sequence in PRSV-W P1 protein to the catalytic active site for P1 proteinase is H_{456}-(8X)-D_{465}-(31X)-G_{497}-(X)-$S_{499}G_{500}$. Moreover, the conserved active site sequence G-(X)-SG found in all potyviruses (Shukla *et al*., 1994) is also present in the P1 protein of PRSV-W at amino acid positions 497-500 of the polyprotein. As shown in Table 14, the heterogeneity of P1 is very high (66.7-97.1%).

6.2.2 Helper component-proteinase (HC-Pro)

The helper component protein (HC-Pro) encoded by PRSV-W is composed of 457 amino acid residues (52.1 kDa). This protein has several functions including a helper component for aphid transmission of viruses, a proteinase for processing between HC-Pro and P3, and involvement in long distance movement (Maia *et al*., 1996). The N-terminal domain was shown to be important for virus transmission by aphids in HC-Pro of PVY (Atreya *et al*., 1992).

In HC-Pro, several possible conserved motifs from other potyviruses were found. The conserved motif G_{888}-Y_{889}-C_{890}-Y_{891}-(71X)-H_{963} described for HC-Pro (Oh and Carrington, 1989) is present in the C-terminal region of the corresponding protein of PRSV-W, with 72 amino-acids distance between cysteine (C) and histidine (H), essential for the putative proteinase active site of HC-Pro, at positions 890 and 963, respectively.

The conserved cysteine cluster C-(8X)-C-(13X)-C-(4X)-C-(2X)-C in the N-terminal third of the HC-Pro of several species of the *Potyviridae* is similar to the zinc finger metal-binding motifs of several nucleic acid-binding proteins (Robaglia *et al*., 1989; Shukla *et al*., 1994). This motif is probably involved in aphid transmission of potyviruses and/or cell-to-cell movement (Atreya *et al*., 1992). Like most other

potyviruses, the conserved cluster of 5 cysteine residues, C_{573}-(8X)-C_{582}-(13X)-C_{596}-(4X)-C_{601}-(2X)-C_{604}, is completely present in PRSV-W HC-Pro. This cluster contains KITC box which is one of the boxes associated with stylet binding and indicated to be important for virus transmission by aphids (Thornbury *et al*., 1990; Atreya and Pirone, 1993). Aphid transmission of potyviruses is hypothesized to result from an interaction between the mouthparts of the aphid, HC-Pro and CP. In PRSV-W, this box is found at amino acid positions 598-601 of the polyprotein. Additionally, the conserved sequence motifs FRNK and PTK in HC-Pro are also believed to be essential in the transmission of the virus by aphids (Revers *et al*., 1999). The conserved FRNK and PTK motifs are perfectly conserved in the corresponding protein at amino acid positions 727-730 and 856-858 of the PRSV-W polyprotein, respectively.

Several protein-protein interactions have been identified for proteins from several species of different genera (*Potyvirus, Tritimovirus*) of the family *Potyviridae*. The conserved histidine (H) and two cysteine (C) residues within the cysteine-rich region of HC-Pro were involved in HC-Pro self-interaction (Urcuqui-Inchima *et al*., 1999a, b). This conserved motif, H_{571}-(X)-C_{573}-(27X)-C_{601}, was completely present in the HC-Pro of PRSV-W.

A conserved sequence ‘CCCVT’-box in the central region of HC-Pro was demonstrated to be involved in the long distance movement of TEV (Cronin *et al*., 1995), but although the motif is found in the HC-Pro of all potyviruses, it is not perfectly conserved. The replacement of the second cysteine (C) by serine (S) in the sequence of PRSV-W was also recognized in *Cocksfooot streak virus* (CSV) (Gotz and Maiss, 2002). The ‘CSCVT’-box is found at amino acid positions 838-842 of the PRSV-W polyprotein. For PRSV-W HC-Pro, considerable levels of identity (95.9-98.5%) were found when compared with those of other PRSV isolates (Table 16).

6.2.3 Third protein (P3)

P3 protein of PRSV-W is composed of 345 amino acid residues (39.7 kDa), showing 92.2-97.7% similarity to other PRSV isolates. The protein is not known in detail but the presence of this potyviral sequence in the P3 region suggests that the P3 protein could be involved in the regulation of the proteolytic processing of the potyviral polyprotein, similar to what is seen with comoviruses (Riechmann *et al*., 1992). The immunogold study shows both its location in the cytoplasmic inclusion and the very early detection in protoplasts indicating the involvement in virus replication (Rodriguez-Cerezo *et al*., 1993). Together with 6K1, P3 gene has been shown to influence symptom induction, systemic movement or host specific pathogenicity (Saenz *et al*., 2000; Dallot *et al*., 2001; Hjulsager *et al*., 2002).

The conserved motif EPY-(7X)-SP-(2X)-L-(X)-A-(2X)-N-(X)-G-(2X)-E-(5X)-W is found in the P3 region of PRSV-W at amino acid positions 1034-1064 (a threonine (T) replaces a consensual alanine (A) at position 1050). The presence of this motif suggests a role for PRSV-W P3 in regulation of proteolytic processing of the potyviral protein as was proposed for a similar motif in the 32kDa protein of *Cowpea mosaic virus* (*Comovirus*; CPMV) (Riechmann *et al*., 1992).

6.2.4 6-kilodalton peptide (6K)

The 6K1 protein of PRSV-W has 52 amino acid residues (5.9 kDa) and 92.5-98.1% similarity to other PRSV isolates. The function of this protein is unclear. However, the protein is suggested to be involved in virus replication based on the similar location in the genome to the picornaviral 2B peptide (Riechman *et al*., 1992).

6.2.5 Cylindrical inclusion protein (CI)

The CI protein encoded by PRSV-W is composed of 635 amino acid residues (71.5 kDa). The potyvirus CI protein has been demonstrated to have helicase

activity and ATPase activity and is probably involved in the virus replication (Lain *et al*., 1990, 1991; Riechman *et al*., 1992).

The RNA helicase motif DE-(X)-H is found in the potyviral CI proteins. In PRSV-W, the motif DECH sequence is completely conserved in the CI protein at amino acid positions 1575-1578. In addition, six nucleotide-binding sequence motifs characteristic of helicase proteins which are found in the CI proteins of all sequenced members of the family *Potyviridae* (Hodgeman, 1988; Lain *et al*., 1990; Kadare and Haenni, 1997) reside in the amino acid sequence of the PRSV-W CI protein. These motifs were identified as clusters GAVGSGKST (9 amino acids; at positions 1486-1494), VLLIEPTRPL (10 amino acids; at positions 1506-1515), KVSAT (5 amino acids; at positions 1602-1606), LVYV (4 amino acids; at positions 1653-1656), VATNIIENGVTL (12 amino acids; at positions 1704-1715) and GERIQRLGRVGR (12 amino acids; at positions 1748-1759) which are perfectly conserved in the corresponding protein of the PRSV-W.

6.2.6 Small nuclear inclusion protein (NIa)

The NIa protein of PRSV-W, such as that of all members of the family *Potyviridae* sequenced to date, can be divided into two domains, VPg and proteinase. The VPg and NIa proteinase (NIa-Pro) of PRSV-W has 189 amino acid residues (21.4 kDa) and 238 amino acid residues (26.5 kDa), respectively.

The VPg protein contained a tyrosine residue at the conserved amino acid position 186 which is required for linking VPg to potyviral RNA (Murphy *et al*., 1996). A conserved triad of amino acids histidine (H), aspartic acid (D) and cysteine (C) in the consensus motif H-(38X)-D-(71X)-GDCG represents the active site of the NIa-Pro (Dougherty *et al*., 1989a, b). The amino acid sequence of PRSV-W includes these three amino acids, though the distances between the amino acids histidine and aspartic acid and between the amino acids aspartic acid and cysteine are slightly different and the motif has a sequence presenting in the NIa-Pro of H_{2327}-(34X)-D_{2362}-(66X)-GDC_{2431}G.

6.2.7 Large nuclear inclusion protein (NIb)

The protein NIb, which is the most conserved potyvirus protein, is assumed to be the viral RNA polymerase because it contains a conserved motif present in all positive-stranded RNA virus RNA-dependent RNA polymerase (RdRp) (Kamer and Argos, 1984). An RdRp motif, CDADGS, was perfectly conserved in the PRSV-W NIb protein at amino acids 2764-2769. Furthermore, the conserved motif (S/T)G-(3X)-T-(3X)-N(S/T)-(18-37X)-GDD is also proposed to be the active site of the RNA-dependent RNA polymerase (Domier *et al*., 1987; Koonin, 1991). The conserved motif of $S_{2829}G_{2830}$-(3X)-T_{2834}-(3X)-$N_{2838}T_{2839}$-(30X)-$G_{2870}D_{2871}D_{2872}$ was completely conserved in the sequence of PRSV-W polyprotein.

6.2.8 Coat protein (CP)

Like most other aphid-transmissible potyviruses, a conserved tripeptide DAG (aspartic acid, alanine and glycine) motif presented in the N-terminal region of the PRSV-W CP which is considered to be essential for the interaction with HP-Pro in the transmission of the virus by the aphids (Atreya *et al*., 1991; Pirone and Blanc, 1996). The conserved DAG motif is perfectly conserved in the CP at amino acid positions 3064-3066 of the PRSV-W polyprotein. The conserved motif R-(43X)-D in the CP is presumably involved in long-distance movement of potyviruses (Dolja *et al*., 1994, 1995). The amino acid sequence of PRSV-W CP includes these two amino acids, and the motif has a sequence R_{3234}-(43X)-D_{3278} in the sequence of polyprotein.

7. Comparison of nucleotide and amino acid sequence of the PRSV-W isolate with those of eleven other PRSV isolates

In this research, all alignments and phylogenetic relationships of complete nucleotide and deduced amino acid sequences of eleven fully-sequenced isolates in the PRSV species currently available in the GenBank database and of the PRSV-W sequences were performed using the program MegAlign in the DNAStar

biocomputing software (DNAStar, Inc.). Each phylogenetic analysis was carried out by aligning the sequences using the Jotun Hein algorithm with gap penalty of 11 and gap length penalty of 3 as provided by the MegAlign program. For comparing nucleotide and amino acid residues based on evolutionary substitution patterns, charge, structural and chemical similarity, genetic distances between pairs of nucleotide sequences were automatically calculated using ambiguous weights. The residue weight values were assigned and shown as in weight table (Appendix B Table 1). Percent accepted mutation 100 (PAM 100) matrix representing one mutation per 100 residues was used to automatically calculate the protein distances between pairs of amino acid sequences. The weight values of amino acid residues were assigned and shown as in residue distance table (Appendix B Table 2).

The eleven available complete nucleotide sequences, including 11 isolates (Table 11), of PRSV used for the comparisons were obtained from the GenBank accession numbers: AY010722 (PRSV-W-ThRB; Thailand isolate of PRSV type W), AY027810 (PRSV-W-TW; Taiwan isolate of PRSV type W), DQ374152 (PRSV-W-C; Brazil isolate of PRSV type W, severe strain), DQ371453 (PRSV-W-1; Bazil isolate of PRSV type W, mild strain), AY162218 (PRSV-P-TH; Thailand isolate of PRSV type P), X67673 (PRSV-P-HA; Hawaii isolate of PRSV type P), X97251 (PRSV-P-YK; Taiwan isolate of PRSV type P), AY231130(PRSV-P-MEX; Mexico isolate of PRSV type P), DQ340769 (PRSV-P-DF; Taiwan isolate of PRSV type P, strain of leaf-deformation), DQ340770 (PRSV-P-SMN; Taiwan isolate of PRSV type P, strain of leaf-mottling and necrosis) and DQ340771 (PRSV-P-SM; Taiwan isolate of PRSV type P, strain of leaf-mottling).

7.1 Phylogenetic analysis

Sequence alignments and phylogenetic analyses were carried out separately for each protein-encoding region at the nucleotide and amino acid level, and each part of 5’- and 3’-NCRs nucleotide sequences using the program MegAlign. The complete genomic RNA and amino acid (polyprotein) sequences of PRSV-W

Table 11 Sources and sequences accession numbers of genome of PRSV isolates used in sequence multiple alignments and phylogenetic analyses

PRSV isolate	Source (country)	GenBank accession number	Genomic RNA size (nucleotide)	Polyprotein size (amino acid)
PRSV-W	cucurbits (Thailand)	this study	10323	3343
PRSV-W-ThRb	cucurbits (Thailand)	AY010722	10323	3343
PRSV-W-TW	cucurbits (Taiwan)	AY027810	10334	3346
PRSV-W-C	cucurbits (Brazil)	DQ374152	10326	3344
PRSV-W-1	cucurbits (Brazil)	DQ374153	10332	3346
PRSV-P-TH	papaya (Thailand)	AY162218	10323	3343
PRSV-P-HA	papaya (USA)	X67673	10326	3344
PRSV-P-YK	papaya (Taiwan)	X97251	10326	3344
PRSV-P-MEX	papaya (Mexico)	AY231130	10320	3342
PRSV-P-DF	papaya (Taiwan)	DQ340769	10326	3344
PRSV-P-SMN	papaya (Taiwan)	DQ340770	10326	3344
PRSV-P-SM	papaya (Taiwan)	DQ340771	10326	3344

were also aligned and phylogenetically analyzed separately at the nucleotide and amino acid level using the same parameter in the same program.

7.1.1 Phylogenetic analyses based on the NCRs

Non coding regions (NCRs) on PRSV genome including 5'-NCR and 3'-NCR at 5'- and 3'-terminus, respectively were analyzed separately at the nucleotide level by multiple alignments and phylogenetic analyses. By ORF searching and sequence analysis, 85 nucleotides in length at the 5' terminus of PRSV-W is the 5'NCR. Sizes of 5'NCR nucleotide sequences of all isolates of the available complete PRSV sequences are identical size in length (85 nucleotides) except for the PRSV-W-TW isolate (87 nucleotides). At 3'terminal region of PRSV genome excluding poly(A) tail, all of isolates of the available complete PRSV sequences and PRSV-W isolate are identical in size (206 nucleotides).

a) 5' NCR nucleotide sequences

Multiple alignment of 5'NCR nucleotide sequences of the twelve isolates (Table 12) showed that the identities range from 61.6 to 98.8%, and the PRSV-W 5'NCR has 66.3-98.8% identical nucleotides when compared with the eleven other PRSV isolates. A phylogenetic tree was constructed from the multiple alignment of the 5'-NCR nucleotide sequences of twelve PRSV isolates, including all the fully-sequenced isolates in the species. The phylogenetic tree (Figure 33A) shows a degree of relationship between the PRSV isolates. Phylogenetic analysis of the twelve PRSV isolates reveals a relationship between the geographic locations of the isolates. These isolates of PRSV species aligned could be divided into two clusters. Both geographic locations of the isolates from Taiwan and Thailand including of eight Asian isolates separately formed their own phylogenetic branch. In the Asian isolate cluster, three PRSV isolates from Thailand including PRSV-W, PRSV-W-ThRb and PRSV-P-TH formed a sub-cluster separating from five PRSV isolates from Taiwan. In the sub-cluster of the PRSV isolates from Thailand, the PRSV-W isolate is most closely related to PRSV-W-ThRb (1.2% different nucleotides) and PRSV-P-TH

Table 12 Percent identities (top right) and differences (bottom left) between the 5' non coding region (5'-NCR) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available.

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.8	89.5	66.3	68.6	89.5	69.8	86.0	69.8	86.0	87.2	89.5
PRSV-W-ThRb	1.2		88.4	67.4	69.8	88.4	70.9	84.9	70.9	84.9	87.2	88.4
PRSV-W-TW	11.5	12.8		67.4	65.1	86.0	66.3	96.5	69.8	95.3	94.2	97.7
PRSV-W-C	45.0	42.9	42.8		88.4	67.4	79.1	66.3	82.6	70.9	66.3	66.3
PRSV-W-1	40.8	38.9	47.5	12.7		65.1	81.4	61.6	82.6	68.6	66.3	64.0
PRSV-P-TH	11.4	12.7	15.8	42.9	47.8		66.3	84.9	66.3	83.7	81.4	86.0
PRSV-P-HA	39.2	37.3	46.0	25.7	22.4	45.7		65.1	90.7	67.4	65.1	65.1
PRSV-P-YK	15.8	17.2	3.6	44.8	54.7	17.2	48.1		68.6	91.9	90.7	94.2
PRSV-P-MEX	39.0	37.1	39.0	20.6	20.4	45.4	10.2	40.9		70.9	68.6	68.6
PRSV-P-DF	16.0	17.4	4.9	36.8	40.8	19.0	43.7	8.8	37.0		91.9	93.0
PRSV-P-SMN	14.2	14.1	6.1	45.0	44.8	21.9	47.8	10.0	41.1	8.7		93.0
PRSV-P-SM	11.5	12.8	2.4	45.0	49.9	15.8	48.5	6.1	41.1	7.5	7.4	

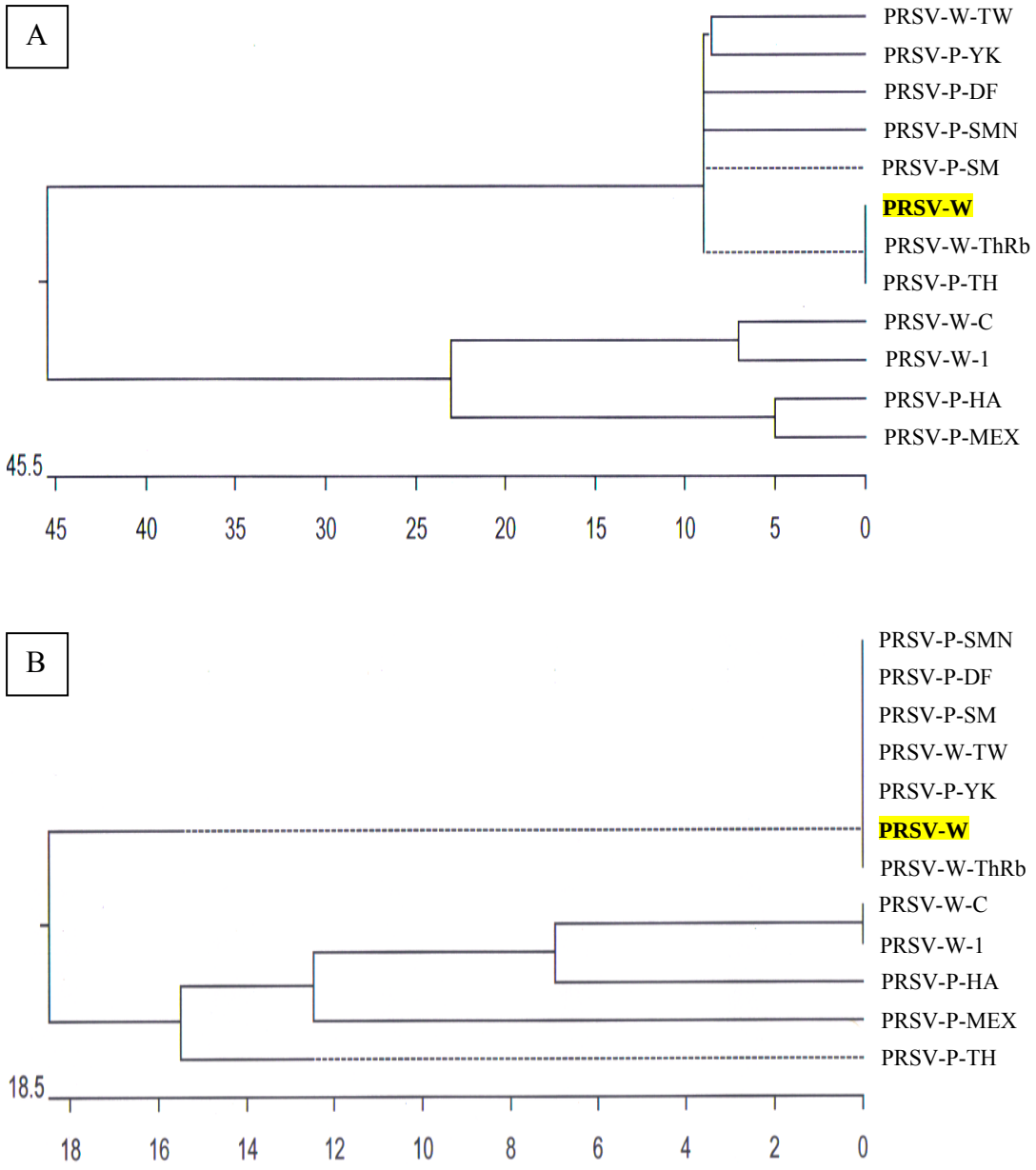


Figure 33 Phylogenetic trees constructed from the multiple alignments of the non coding regions (NCRs) nucleotide sequences based on the sequences of 5'-NCRs (A) and 3'-NCRs (B) of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

(11.4% different nucleotides) isolates. Another cluster separating from the Asian cluster including of four non-Asian isolates consists of two sub-clusters. Within the cluster of non-Asian isolates, both isolates of PRSV from Brazil (PRSV-W-C and PRSV-W-1 isolates) grouped together to form a sub-cluster separating from another sub-cluster of PRSV-P-HA (from USA) and PRSV-P-MEX (from Mexico) isolates which were grouped together. The differences between 5'-NCR nucleotide sequence of the PRSV-W isolate and those of the Asian isolates within the same cluster was 1.2 to 16.0%, while it differed by 39.0 to 45.0% when compared with those of the non-Asian isolates placing within another cluster.

b) 3' NCR nucleotide sequences

In the PRSV 3'-NCR, multiple alignment of nucleotide sequences of the twelve isolates (Table 13) showed that the identities range from 88.4 to 100.0%, and the PRSV-W 3'NCR has 93.2-100.0% identical nucleotides when compared with the eleven other PRSV isolates. No difference of PRSV-W 3'-NCR nucleotide was shown when compared with that of PRSV-W-ThRb (100.0% identity). A phylogenetic tree (Figure 33B) constructed from the multiple alignment of 3'-NCR nucleotide sequences of twelve fully-sequenced PRSV isolates shows a pattern of relationship between the aligned PRSV isolates. These PRSV isolates were divided into two clusters. In the first cluster, the PRSV-W was grouped together with PRSV-W-ThRb, PRSV-P-YK, PRSV-W-TW, PRSV-P-SM, PRSV-P-DF and PRSV-P-SMN which it has 0.0-5.6% different nucleotides when its 3'-NCR compared with those of PRSV isolates placing within the same cluster. In second cluster, five PRSV isolates including PRSV-W-C, PRSV-W-1, PRSV-P-HA, PRSV-P-MEX and PRSV-P-TH were grouped together by phylogenetic branching from the first cluster.

7.1.2 Phylogenetic analyses based on the P1s

Sequence analysis demonstrated that the first protein (P1) coding region of PRSV-W consists of 1641 nucleotides, locating at nucleotide positions 86 to 1726 on the viral genome and representing characteristic PRSV RNA, which encodes

Table 13 Percent identities (top right) and differences (bottom left) between the 3' non coding region (3'-NCR) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		100.0	98.1	95.2	95.2	97.1	93.7	98.1	93.2	94.7	95.2	95.2
PRSV-W-ThRb	0.0		98.1	95.2	95.2	97.1	93.7	98.1	93.2	94.7	95.2	95.2
PRSV-W-TW	2.0	2.0		93.2	93.2	95.2	91.8	98.1	91.3	96.1	96.6	96.6
PRSV-W-C	5.0	5.0	7.2		97.6	94.7	94.2	94.2	91.3	90.8	91.3	91.3
PRSV-W-1	5.1	5.1	7.2	2.5		95.2	94.7	94.2	91.8	90.8	91.3	91.3
PRSV-P-TH	3.0	3.0	5.0	5.5	5.0		93.7	96.1	93.2	91.8	92.3	92.3
PRSV-P-HA	6.6	6.6	6.8	6.1	5.6	6.6		92.8	92.8	88.4	88.9	88.9
PRSV-P-YK	2.0	2.0	2.0	6.1	6.1	4.0	7.7		92.3	94.7	95.2	95.2
PRSV-P-MEX	7.2	7.2	9.5	9.4	8.9	7.2	7.7	8.4		88.9	89.4	89.4
PRSV-P-DF	5.6	5.6	4.0	10.0	10.0	8.9	12.9	5.6	12.4		99.5	99.5
PRSV-P-SMN	5.1	5.1	3.5	9.4	9.5	8.3	12.3	5.1	11.8	0.5		100
PRSV-P-SM	5.1	5.1	3.5	9.4	9.5	8.3	12.3	5.1	11.8	0.5	0.0	

a putative protein of 547 amino acids (Table 7) with a calculated molecular weight of 62.6 kDa (Table 8). The PRSV-W P1 gene (1641 nucleotides) and putative protein (547 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) P1 nucleotide sequences

Multiple alignment of the twelve P1 coding region nucleotide sequences (Table 14) showed that their P1 gene identities and differences varied from 69.5 to 98.4% and 1.6 to 40.8%, respectively. The P1 nucleotide sequence of PRSV-W has the similar ranges of identity (69.5-98.4%) and of difference (1.6-40.8%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the P1 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 34A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates. As the result obtained from multiple alignment of P1 coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.4%) and the lowest distance (1.6%) with PRSV-W-ThRb isolate.

Table 14 Percent identities (top right) and differences (bottom left) between the first protein (P1) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.4	80.8	70.7	69.5	81.9	70.6	81.6	70.3	80.9	81.1	81.4
PRSV-W-ThRb	1.6		81.6	70.9	69.8	82.5	70.7	82.3	70.8	81.8	82.0	82.2
PRSV-W-TW	23.2	21.9		70.6	70.2	82.8	71.1	93.2	70.5	93.4	94.1	93.1
PRSV-W-C	38.3	38.0	38.7		92.1	70.9	87.9	70.5	87.3	70.2	70.1	70.5
PRSV-W-1	40.8	40.2	39.5	8.4		70.4	87.1	70.2	87.0	70.2	70.0	69.9
PRSV-P-TH	21.4	20.5	20.3	38.1	39.0		70.8	83.1	71.3	82.8	83.1	82.6
PRSV-P-HA	38.4	38.3	37.7	13.6	14.5	38.1		70.9	89.9	70.2	70.2	70.0
PRSV-P-YK	22.0	20.9	7.2	39.0	39.7	20.0	38.2		70.4	93.4	94.1	92.8
PRSV-P-MEX	39.1	38.2	39.1	14.3	14.7	37.3	11.1	39.2		69.9	70.1	70.4
PRSV-P-DF	22.9	21.7	7.0	39.6	39.6	20.3	39.5	7.0	40.1		94.8	94.5
PRSV-P-SMN	22.6	21.3	6.2	39.7	40.0	19.8	39.5	6.2	39.7	5.4		94.8
PRSV-P-SM	22.2	21.1	7.3	39.1	40.2	20.6	39.8	7.7	39.2	5.8	5.5	

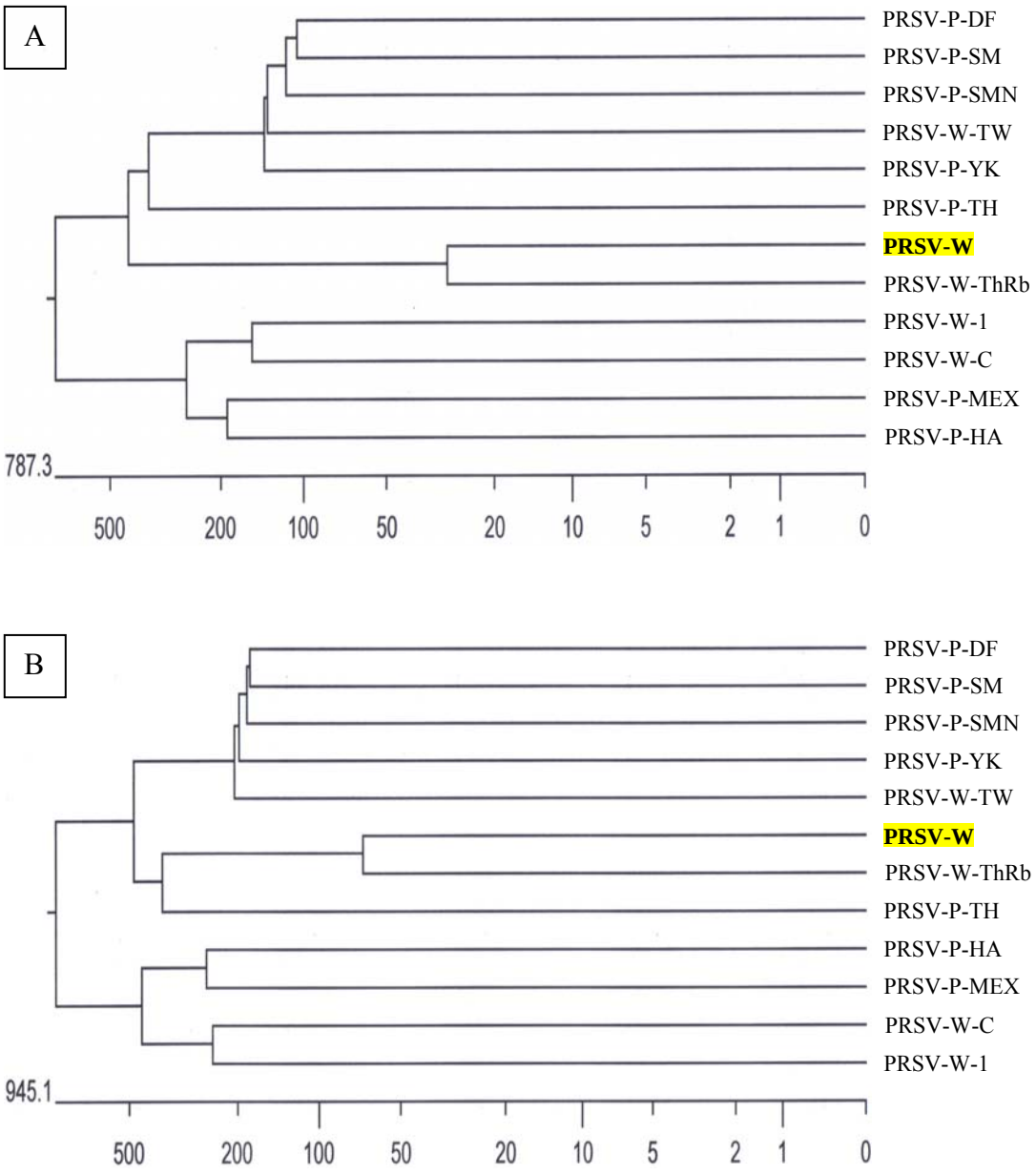


Figure 34 Phylogenetic trees constructed from the multiple alignments of the first protein (P1) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

b) P1 amino acid sequences

Multiple alignment of the twelve putative P1 amino acid sequences (Table 15) showed that their P1 protein identities and differences varied from 65.4 to 97.1% and 3.0 to 46.2%, respectively. The P1 amino acid sequence of PRSV-W has the similar ranges of identity (66.7-97.1%) and of difference (3.0-43.9%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the P1 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 34B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, all five isolates from Taiwan including PRSV-W-TW, PRSV-P-YK, PRSV-P-SMN, PRSV-P-SM and PRSV-P-DF grouped the isolates according to their provenance as a sub-cluster. In another sub-cluster, the three Thai isolates grouped together, with a further branch separating the two isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) from PRSV-P-TH (papaya-infecting). As the result obtained from multiple alignment of P1 coding region amino acid sequences (Table 15) and the branching pattern of the phylogenetic tree (Figure 34B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (97.1%) and the lowest distance (3.0%) with PRSV-W-ThRb isolate. The differences between the putative P1 amino acid sequence of the PRSV-W isolate and those of the five Taiwanese isolates within the same cluster was 22.3 to 24.6%, while it differed by 42.3 to 43.9% when compared with those of the non-Asian isolates placing within another cluster in the same dendrogram.

Table 15 Percent identities (top right) and differences (bottom left) between the putative first protein (P1) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		97.1	80.6	67.2	66.7	82.2	67.0	80.7	67.6	79.1	80.6	79.3
PRSV-W-ThRb	3.0		81.7	68.0	67.6	82.9	67.2	81.5	68.7	79.8	81.3	79.6
PRSV-W-TW	22.6	21.1		66.7	67.2	81.8	66.7	92.0	67.4	90.3	92.2	91.2
PRSV-W-C	43.0	41.7	43.9		89.4	67.2	85.2	66.3	85.9	65.4	65.9	66.1
PRSV-W-1	43.9	42.3	43.0	11.4		68.0	83.2	66.9	84.3	65.4	66.9	66.3
PRSV-P-TH	20.4	19.4	20.9	43.0	41.7		67.6	80.6	68.7	80.6	81.8	79.6
PRSV-P-HA	43.3	43.0	43.9	16.5	19.1	42.3		67.4	87.0	65.6	66.3	65.7
PRSV-P-YK	22.3	21.3	8.5	44.6	43.6	22.6	42.6		67.4	91.1	93.4	90.7
PRSV-P-MEX	42.3	40.4	42.6	15.6	17.7	40.4	14.3	42.6		66.5	67.6	66.9
PRSV-P-DF	24.6	23.6	10.4	46.2	46.2	22.6	45.9	9.5	44.3		92.2	92.7
PRSV-P-SMN	22.6	21.6	8.3	45.2	43.6	20.9	44.6	6.9	42.3	8.3		93.1
PRSV-P-SM	24.3	23.8	9.3	44.9	44.6	23.8	45.6	10.0	43.6	7.7	7.3	

7.1.3 Phylogenetic analyses based on the HC-Pros

Sequence analysis demonstrated that the helper component-proteinase (HC-Pro) coding region of PRSV-W consists of 1371 nucleotides, locating at nucleotide positions 1727 to 3097 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 457 amino acids (Table 7) with a calculated molecular weight of 52.1 kDa (Table 8). The PRSV-W HC-Pro gene (1371 nucleotides) and putative protein (457 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) HC-Pro nucleotide sequences

Multiple alignment of the twelve HC-Pro coding region nucleotide sequences (Table 16) showed that their HC-Pro gene identities and differences varied from 85.2 to 99.0% and 1.0 to 17.2%, respectively. The HC-Pro nucleotide sequence of PRSV-W has the similar ranges of identity (85.2-99.0%) and of difference (1.0-17.2%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the HC-Pro coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 35A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the

Table 16 Percent identities (top right) and differences (bottom left) between the helper component-proteinase (HC-Pro) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		99.0	89.8	86.5	85.6	89.9	85.2	89.3	85.6	89.2	89.0	89.4
PRSV-W-ThRb	1.0		90.1	86.7	85.7	90.6	85.3	89.6	85.8	89.7	89.1	89.9
PRSV-W-TW	11.3	10.9		85.9	86.0	92.1	86.4	97.1	87.0	96.0	96.2	96.3
PRSV-W-C	15.5	15.2	16.3		95.0	86.5	91.8	85.8	91.8	86.4	86.2	86.4
PRSV-W-1	16.8	16.6	16.1	5.3		85.5	91.9	86.2	92.6	86.4	86.3	86.3
PRSV-P-TH	11.1	10.3	8.5	15.4	16.8		86.6	92.1	86.7	91.5	91.3	91.3
PRSV-P-HA	17.2	17.0	15.5	8.8	8.8	15.3		86.5	93.2	86.7	86.4	86.4
PRSV-P-YK	11.9	11.5	3.0	16.4	15.8	8.5	15.4		87.2	96.5	96.4	96.4
PRSV-P-MEX	16.6	16.4	14.8	8.9	8.0	15.2	7.2	14.5		87.9	87.5	87.4
PRSV-P-DF	12.0	11.4	4.2	15.5	15.5	9.2	15.2	3.6	13.6		96.4	97.3
PRSV-P-SMN	12.3	12.1	3.9	15.9	15.7	9.4	15.5	3.8	14.1	3.7		97.0
PRSV-P-SM	11.7	11.2	3.9	15.5	15.7	9.4	15.5	3.8	14.2	2.8	3.1	

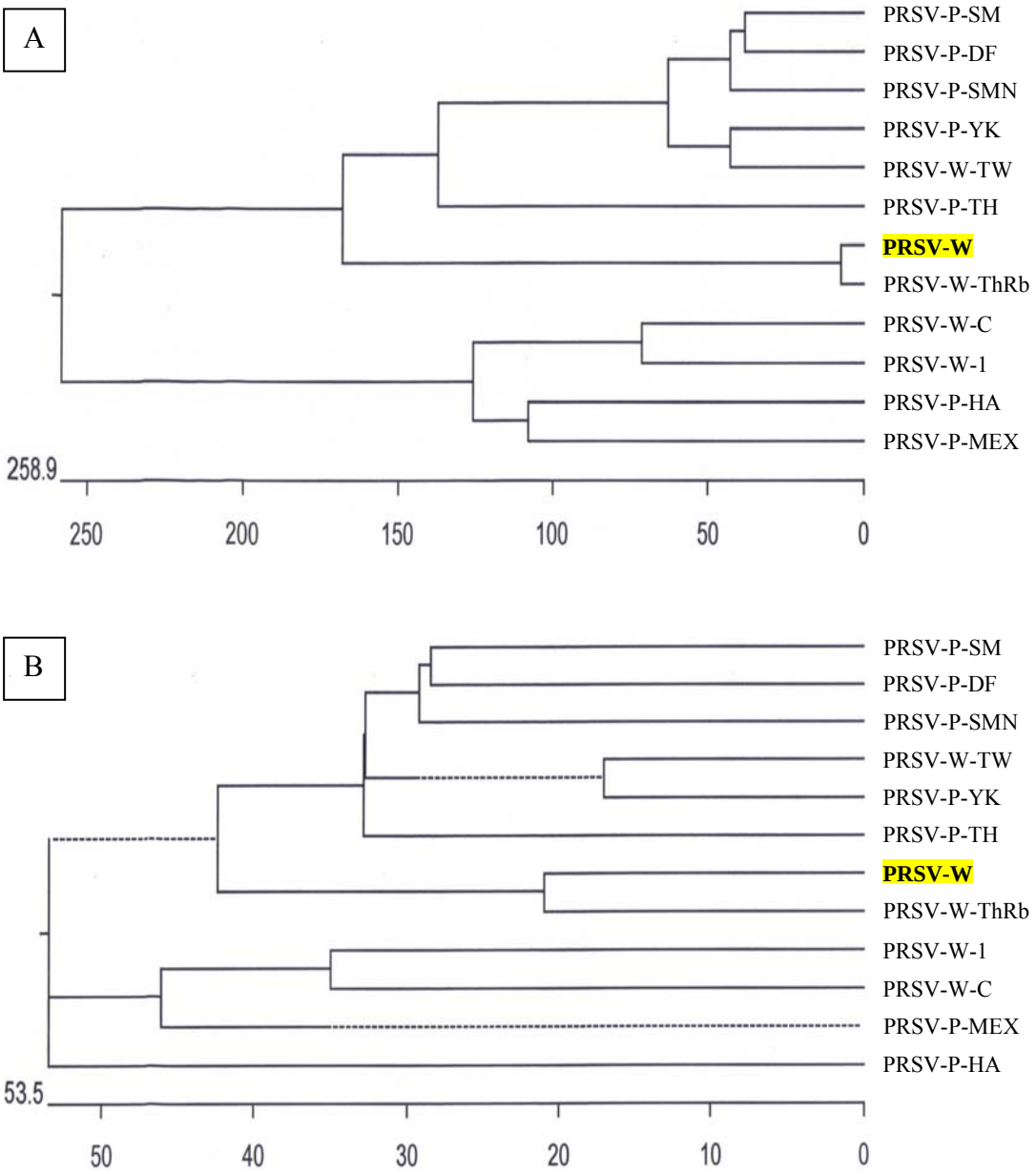


Figure 35 Phylogenetic trees constructed from the multiple alignments of the helper component-proteinase (HC-Pro) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

phylogenetic branch formed a sub-cluster separating from six other Asian isolates. As the result obtained from multiple alignment of HC-Pro coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.4%) and the lowest distance (1.6%) with PRSV-W-ThRb isolate.

b) HC-Pro amino acid sequences

Multiple alignment of the twelve putative HC-Pro amino acid sequences (Table 17) showed that their HC-Pro amino acid identities and differences varied from 94.8to 98.9% and 1.1 to 5.4%, respectively. The HC-Pro amino acid sequence of PRSV-W has the similar ranges of identity (95.9-98.5%) and of difference (1.5-4.3%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the HC-Pro coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 35B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into three clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, three non-Asian isolates including PRSV-W-1, PRSV-W-C and PRSV-P-MEX grouped together to form a cluster. While PRSV-P-HA did not group with the non-Asian isolates, it was separated by branching as a cluster from all isolates. In the cluster of non-Asian PRSV isolates, the two Brazilian isolates grouped together, with a further branch separating the two isolates of cucurbit-infecting PRSV (PRSV-W-1 and PRSV-W-C) from PRSV-P-MEX (papaya-infecting). In the cluster of the Asian isolates, all five isolates from Taiwan and one isolate from Thailand including PRSV-P-YK, PRSV-W-TW, PRSV-P-SMN, PRSV-P-DF, PRSV-P-SM and PRSV-P-TH grouped the isolates as a sub-cluster. In another sub-cluster, the two Thai isolates grouped together, separating the two isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) from the Asian isolates. As the result obtained from multiple alignment of HC-Pro coding

Table 17 Percent identities (top right) and differences (bottom left) between the putative helper component-proteinase (HC-Pro) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.5	96.5	96.9	96.9	97.8	96.5	96.7	95.9	96.3	96.5	95.9
PRSV-W-ThRb	1.5		96.7	97.2	97.2	98.0	96.5	96.9	95.9	96.9	96.3	96.5
PRSV-W-TW	3.6	3.4		95.6	95.6	97.8	95.4	98.9	95.2	98.0	97.8	97.6
PRSV-W-C	3.1	2.9	4.5		98.0	97.4	96.5	95.9	96.5	95.4	95.6	95.0
PRSV-W-1	3.1	2.9	4.5	2.0		96.9	96.7	95.9	96.5	95.4	95.6	95.2
PRSV-P-TH	2.2	2.0	2.2	2.7	3.1		96.3	98.0	96.1	97.6	97.4	97.2
PRSV-P-HA	3.6	3.6	4.7	3.6	3.4	3.8		95.6	96.3	95.2	95.4	94.8
PRSV-P-YK	3.4	3.1	1.1	4.3	4.3	2.0	4.5		95.6	98.5	98.3	98.0
PRSV-P-MEX	4.3	4.3	5.0	3.6	3.6	4.0	3.8	4.5		95.9	95.6	95.0
PRSV-P-DF	3.8	3.1	2.0	4.7	4.7	2.4	5.0	1.5	4.3		98.0	98.7
PRSV-P-SMN	3.6	3.8	2.2	4.5	4.5	2.7	4.7	1.8	4.5	2.0		97.6
PRSV-P-SM	4.3	3.6	2.4	5.2	5.0	2.9	5.4	2.0	5.2	1.3	2.4	

region amino acid sequences (Table 17) and the branching pattern of the phylogenetic tree (Figure 35B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (98.5%) and the lowest distance (1.5%) with PRSV-W-ThRb isolate. The differences between the putative HC-Pro amino acid sequence of the PRSV-W isolate and those of the six Asian isolates placing within the same cluster was 2.2 to 4.3%, while it differed by 3.1 to 4.3% when compared with those of the non-Asian isolates placing within another cluster in the same dendrogram.

7.1.4 Phylogenetic analyses based on the P3s

Sequence analysis demonstrated that the third protein (P3) coding region of PRSV-W consists of 1035 nucleotides, locating at nucleotide positions 3098 to 4132 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 345 amino acids (Table 7) with a calculated molecular weight of 39.7 kDa (Table 8). The PRSV-W HC-Pro gene (1035 nucleotides) and putative protein (345 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) P3 nucleotide sequences

Multiple alignment of the twelve P3 coding region nucleotide sequences (Table 18) showed that their P3 gene identities and differences varied from 85.3 to 98.5% and 1.6 to 16.9%, respectively. The P3 nucleotide sequence of PRSV-W has the similar ranges of identity (85.3-98.5%) and of difference (1.6-16.8%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the P3 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 36A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two

Table 18 Percent identities (top right) and differences (bottom left) between the third protein (P3) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.5	89.0	86.3	85.8	88.2	85.3	89.8	85.9	88.8	88.6	88.5
PRSV-W-ThRb	1.6		89.5	86.7	86.1	88.8	85.9	90.3	86.3	89.3	89.1	88.8
PRSV-W-TW	12.2	11.6		87.8	86.9	90.4	86.2	95.9	86.6	95.1	94.8	95.1
PRSV-W-C	15.5	15.0	13.6		95.3	86.4	91.9	87.8	91.4	86.8	87.2	86.8
PRSV-W-1	16.2	15.9	14.9	4.9		86.0	90.7	87.4	90.8	86.4	86.8	86.2
PRSV-P-TH	13.2	12.4	10.5	15.4	16.0		85.3	91.5	85.4	90.0	89.8	89.5
PRSV-P-HA	16.8	16.1	15.8	8.7	10.1	16.9		86.7	96.2	85.5	85.8	85.5
PRSV-P-YK	11.3	10.6	4.2	13.6	14.3	9.2	15.1		86.6	96.3	95.8	95.8
PRSV-P-MEX	16.0	15.5	15.3	9.3	10.0	16.7	3.9	15.2		85.6	86.1	85.5
PRSV-P-DF	12.4	11.8	5.2	15.0	15.5	11.0	16.6	3.8	16.5		95.8	96.4
PRSV-P-SMN	12.7	12.1	5.5	14.4	14.9	11.3	16.2	4.4	15.8	4.3		95.5
PRSV-P-SM	12.8	12.4	5.1	14.9	15.7	11.6	16.6	4.3	16.6	3.7	4.7	

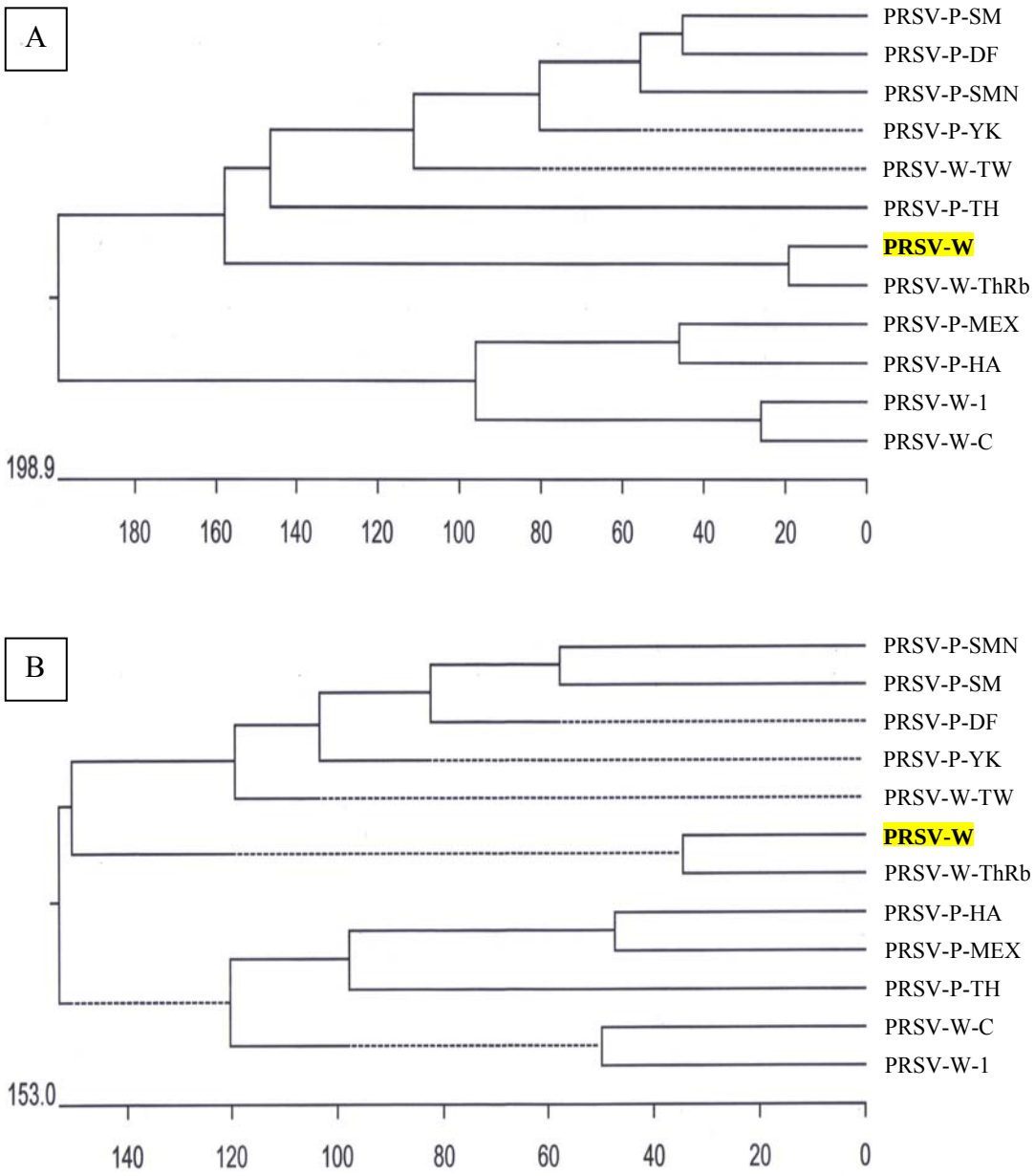


Figure 36 Phylogenetic trees constructed from the multiple alignments of the third protein (P3) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates including PRSV-P-TH, PRSV-W-TW, PRSV-P-YK, PRSV-P-SMN, PRSV-P-DF and PRSV-P-SM. As the result obtained from multiple alignment of P3 coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.5%) and the lowest distance (1.6%) with PRSV-W-ThRb isolate.

b) P3 amino acid sequences

Multiple alignment of the twelve putative P3 amino acid sequences (Table 19) showed that their P3 amino acid identities and differences varied from 92.5 to 98.0% and 2.1 to 7.9%, respectively. The P3 amino acid sequence of PRSV-W has the similar ranges of identity (92.5-97.7%) and of difference (2.4-7.9%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the P3 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 36B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Five Taiwanese isolates and two cucurbit-infecting Thai isolates (PRSV-W and PRSV-W ThRb) grouped together to form their own phylogenetic branch within the same cluster, the two Thai isolates grouped together, with a further branch separating the two Thai isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) as a sub-cluster from five Taiwanese isolates of PRSV. In another cluster,

Table 19 Percent identities (top right) and differences (bottom left) between the putative third protein (P3) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		97.7	95.1	93.9	94.2	93.6	92.2	95.1	92.5	94.5	94.2	93.9
PRSV-W-ThRb	2.4		95.7	94.5	94.8	94.2	92.8	95.7	93.1	95.1	94.8	94.2
PRSV-W-TW	5.1	4.5		95.4	95.1	94.8	93.6	98.0	93.9	96.8	96.5	96.0
PRSV-W-C	6.3	5.7	4.8		96.5	94.8	94.2	95.4	94.2	94.5	94.2	93.9
PRSV-W-1	6.0	5.4	5.1	3.6		95.1	93.1	95.1	93.6	95.1	93.9	93.6
PRSV-P-TH	6.7	6.0	5.4	5.4	5.1		93.1	94.8	94.2	93.6	94.2	93.4
PRSV-P-HA	8.3	7.6	6.7	6.0	7.3	7.3		93.6	95.7	93.1	92.5	93.4
PRSV-P-YK	5.1	4.5	2.1	4.8	5.1	5.4	6.7		93.9	97.1	97.1	96.2
PRSV-P-MEX	7.9	7.3	6.3	6.0	6.7	6.0	4.5	6.3		93.9	92.8	93.1
PRSV-P-DF	5.7	5.1	3.3	5.7	5.1	6.7	7.3	2.9	6.3		96.2	96.2
PRSV-P-SMN	6.0	5.4	3.6	6.0	6.3	6.0	7.9	2.9	7.6	3.9		96.0
PRSV-P-SM	6.3	6.0	4.2	6.3	6.7	7.0	7.0	3.9	7.3	3.9	4.2	

PRSV-P-TH grouped with the Four non-Asian Isolates, with a further branch separating the two Brazilian isolates of cucurbit-infecting PRSV (PRSV-W-C and PRSV-W-1) grouped together as a sub-cluster from the three papaya-infecting isolates including PRSV-P-TH (Thai isolate), PRSV-P-MEX (Mexican isolate) and PRSV-P-HA (American isolate). As the result obtained from multiple alignment of P3 coding region amino acid sequences (Table 19) and the branching pattern of the phylogenetic tree (Figure 36B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (97.7%) and the lowest distance (2.4%) with PRSV-W-ThRb isolate. The differences between the putative P3 amino acid sequence of the PRSV-W isolate and those of the five Taiwanese isolates placing within the same cluster was 5.1 to 6.3%, while it differed by 6.0 to 8.3% when compared with those of the isolates placing within another cluster in the same dendrogram.

7.1.5 Phylogenetic analyses based on the 6K1s

Sequence analysis demonstrated that the 6-kilodalton peptide 1(6K1) coding region of PRSV-W consists of 156 nucleotides, locating at nucleotide positions 4133 to 4288 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 52 amino acids (Table 7) with a calculated molecular weight of 5.9 kDa (Table 8). The PRSV-W 6K1 gene (156 nucleotides) and putative protein (52 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) 6K1 nucleotide sequences

Multiple alignment of the twelve 6K1 coding region nucleotide sequences (Table 20) showed that their 6K1 gene identities and differences varied from 85.4 to 98.7% and 1.3 to 17.1%, respectively. The 6K1 nucleotide sequence of PRSV-W has the similar ranges of identity (86.6-98.7%) and of difference (1.3-15.2%) when compared with the corresponding region of eleven other PRSV isolates.

Table 20 Percent identities (top right) and differences (bottom left) between the 6-kilodalton peptide1 (6K1) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.7	90.4	87.9	87.9	88.5	88.5	89.8	86.6	89.8	91.7	91.1
PRSV-W-ThRb	1.3		91.7	89.2	89.2	89.8	89.8	91.1	86.6	91.1	93.0	92.4
PRSV-W-TW	10.4	8.9		88.5	90.4	90.4	88.5	91.7	86.0	93.6	93.0	94.9
PRSV-W-C	13.6	12.0	12.8		93.6	87.3	91.1	86.6	89.2	88.5	87.9	88.5
PRSV-W-1	13.7	12.0	10.5	6.8		87.3	91.1	87.3	89.2	89.2	88.5	89.2
PRSV-P-TH	12.8	11.2	10.4	14.4	14.5		86.6	89.8	85.4	92.4	91.7	93.6
PRSV-P-HA	12.8	11.2	12.8	9.7	9.8	15.3		89.2	90.4	88.5	87.9	88.5
PRSV-P-YK	11.0	9.5	8.9	15.1	14.3	11.1	11.9		85.4	94.3	94.9	95.5
PRSV-P-MEX	15.2	15.2	16.1	12.0	12.0	17.1	10.4	16.9		85.4	87.3	86.6
PRSV-P-DF	11.1	9.6	6.8	12.8	12.0	8.2	12.8	6.0	17.1		96.8	98.7
PRSV-P-SMN	8.9	7.4	7.5	13.6	12.8	8.9	13.6	5.3	14.5	3.3		98.1
PRSV-P-SM	9.6	8.1	5.3	12.8	12.0	6.7	12.8	4.6	15.3	1.3	1.9	

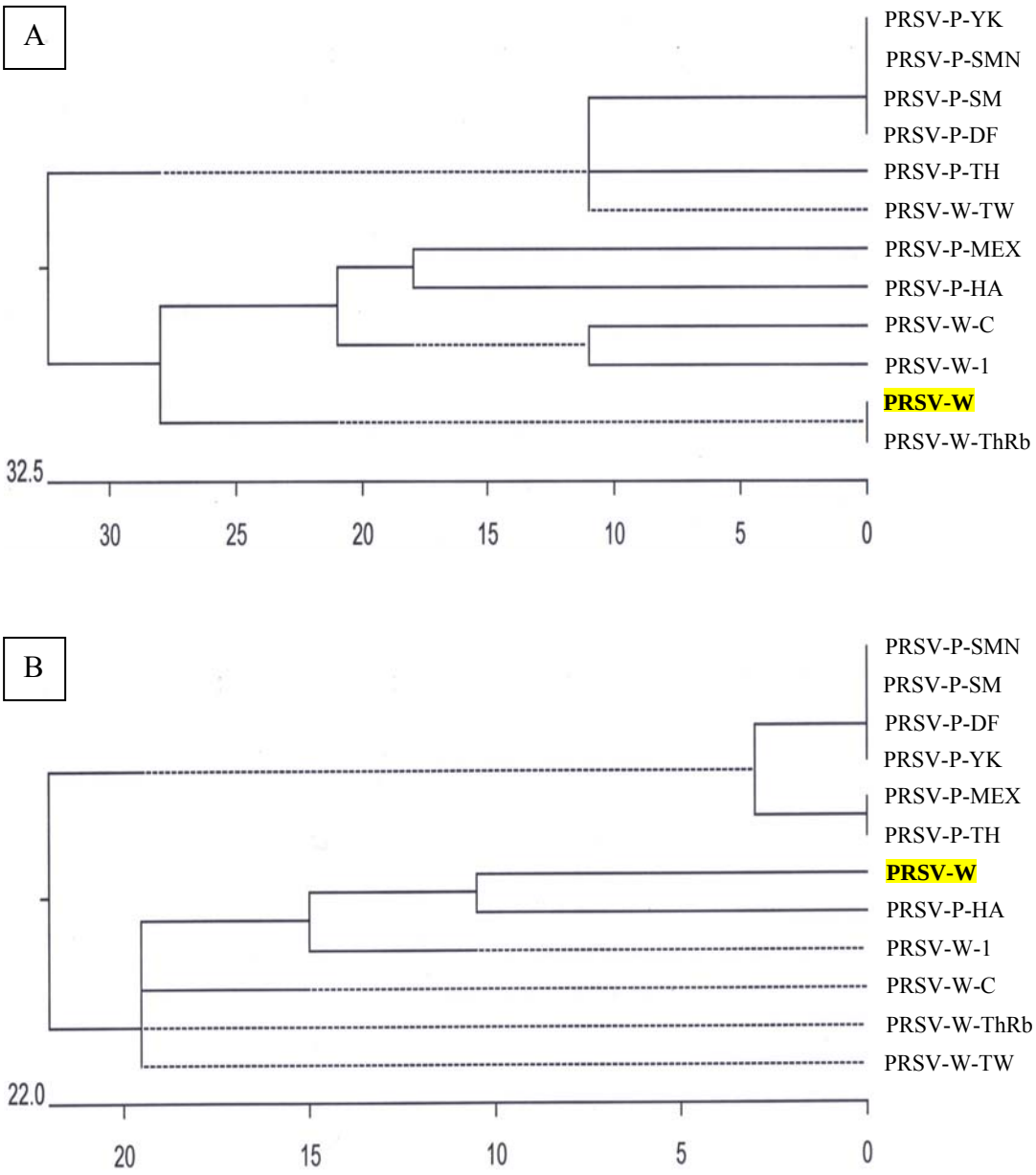


Figure 37 Phylogenetic trees constructed from the multiple alignments of the 6-kilodalton peptide 1(6K1) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the 6K1 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 37A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Six Asian isolates including of all isolates from Taiwan and PRSV-P-TH from Thailand grouped together to form their own phylogenetic branch within the same cluster. In this cluster, all papaya-infecting PRSV (type P) isolates from Taiwan grouped together, with a further branch separating the four PRSV-P isolates as a sub-cluster from PRSV-P-TH (Thai isolate) and PRSV-W-TW (Taiwan isolate, cucurbit-infecting). In another cluster, there are two sub-clusters separating the two Thai isolates (PRSV-W and PRSV-W-ThRb) grouped together separating from another sub-cluster consisting of PRSV-P-MEX (from Mexico), PRSV-P-HA (from USA), PRSV-W-C (from Brazil) and PRSV-W-1 (from Brazil). In this cluster, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from the four non-Asian isolates. As the result obtained from multiple alignment of 6K1 coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.7%) and the lowest distance (1.3%) with PRSV-W-ThRb isolate.

b) 6K1 amino acid sequences

Multiple alignment of the twelve putative 6K1 amino acid sequences (Table 21) showed that their 6K1 amino acid identities and differences varied from 90.6 to 100.0% and 0.0 to 10.1%, respectively. The 6K1 amino acid sequence of PRSV-W has the similar ranges of identity (92.5-98.1%) and of difference (1.9-8.0%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the 6K1 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 37B) shows a branching pattern of relationship between the

Table 21 Percent identities (top right) and differences (bottom left) between the putative 6-kilodalton peptide 1 (6K1) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.1	98.1	94.3	96.2	94.3	96.2	92.5	92.5	94.3	94.3	94.3
PRSV-W-ThRb	1.9		100	96.2	98.1	96.2	98.1	94.3	94.3	96.2	96.2	96.2
PRSV-W-TW	1.9	0.0		96.2	98.1	96.2	98.1	94.3	94.3	96.2	96.2	96.2
PRSV-W-C	5.9	3.9	3.9		94.3	92.5	94.3	90.6	90.6	92.5	92.5	92.5
PRSV-W-1	3.9	1.9	1.9	5.9		94.3	96.2	92.5	92.5	94.3	94.3	94.3
PRSV-P-TH	5.9	3.9	3.9	8.0	5.9		94.3	94.3	98.1	96.2	96.2	96.2
PRSV-P-HA	3.9	1.9	1.9	5.9	3.9	5.9		92.5	92.5	94.3	94.3	94.3
PRSV-P-YK	8.0	5.9	5.9	10.1	8.0	5.9	8.0		94.3	98.1	98.1	98.1
PRSV-P-MEX	8.0	5.9	5.9	10.1	8.0	1.9	8.0	5.9		94.3	94.3	94.3
PRSV-P-DF	5.9	3.9	3.9	8.0	5.9	3.9	5.9	1.9	5.9		100.0	100.0
PRSV-P-SMN	5.9	3.9	3.9	8.0	5.9	3.9	5.9	1.9	5.9	0.0		100.0
PRSV-P-SM	5.9	3.9	3.9	8.0	5.9	3.9	5.9	1.9	5.9	0.0	0.0	

twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. In the first cluster, four Taiwanese isolates of PRSV-P grouped together to form a sub-cluster separating from another sub-cluster grouping of PRSV-P-MEX and PRSV-P-TH. The PRSV-W placed within the second cluster, it grouped together with PRSV-P-HA and PRSV-W-1 as a sub-cluster separating from PRSV-W-C, PRSV-W-ThRb and PRSV-W-TW. From results of multiple alignment of 6K1 coding region amino acid sequences (Table 21), the PRSV-W isolate is most closely related to PRSV-W-ThRb and PRSV-W-TW isolates and it has the highest identity (98.1%) and the lowest distance (1.9%) with the two isolates. The differences between the putative 6K1 amino acid sequence of the PRSV-W isolate and those of the five isolates placing within the same cluster was 1.9 to 5.9%, while it differed by 5.9 to 8.0% when compared with those of the isolates placing within another cluster in the same dendrogram.

7.1.6 Phylogenetic analyses based on the CIs

Sequence analysis demonstrated that the cylindrical inclusion protein (CI) coding region of PRSV-W consists of 1905 nucleotides, locating at nucleotide positions 4289 to 6193 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 635 amino acids (Table 7) with a calculated molecular weight of 71.5 kDa (Table 8). The PRSV-W CI gene (1905 nucleotides) and putative protein (635 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) CI nucleotide sequences

Multiple alignment of the twelve CI coding region nucleotide sequences (Table 22) showed that their CI gene identities and differences varied from 85.2 to 99.1% and 0.9 to 17.1%, respectively. The CI nucleotide sequence of PRSV-W has the similar ranges of identity (85.8-99.1%) and of difference (0.9-16.3%) when

Table 22 Percent identities (top right) and differences (bottom left) between the cylindrical inclusion (CI) protein nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		99.1	89.0	86.4	85.8	88.8	86.6	89.3	86.2	89.0	89.7	89.0
PRSV-W-ThRb	0.9		89.4	86.9	86.4	89.1	87.0	89.9	86.5	89.5	90.1	89.6
PRSV-W-TW	12.3	11.8		87.1	86.0	90.5	87.1	95.6	86.9	96.1	96.1	95.8
PRSV-W-C	15.5	14.8	14.6		95.1	85.9	93.3	87.0	91.9	87.3	86.7	86.6
PRSV-W-1	16.3	15.6	16.0	5.1		85.2	92.6	86.4	91.4	86.4	85.9	85.7
PRSV-P-TH	12.4	12.0	10.5	16.1	17.1		86.4	91.2	86.1	91.3	91.4	91.2
PRSV-P-HA	15.2	14.7	14.6	7.2	7.9	15.5		87.1	95.0	87.1	87.0	86.8
PRSV-P-YK	11.8	11.2	4.6	14.7	15.6	9.5	14.6		87.1	96.2	96.1	96.0
PRSV-P-MEX	15.7	15.3	14.8	8.8	9.3	15.8	5.2	14.6		86.9	86.8	86.5
PRSV-P-DF	12.2	11.7	4.1	14.4	15.5	9.5	14.6	4.0	14.9		98.2	97.7
PRSV-P-SMN	11.4	10.9	4.1	15.1	16.2	9.3	14.8	4.1	14.9	1.8		97.5
PRSV-P-SM	12.2	11.6	4.4	15.3	16.5	9.5	15.0	4.2	15.4	2.4	2.6	

compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree was prepared from an alignment of the nucleotide sequences of the CI coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 38A) shows a degree of relationship between the different isolates of the PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from PRSV-P-TH and the five Taiwanese isolates (PRSV-P-YK, PRSV-P-YK, PRSV-W-TW, PRSV-P-SM, PRSV-P-DF and PRSV-P-SMN. As the result obtained from multiple alignment of CI coding region nucleotide sequences, the PRSV-W isolate has the highest identity (99.1%) and the lowest distance (0.9%) with PRSV-W-ThRb isolate.

b) CI amino acid sequences

Multiple alignment of the twelve putative CI amino acid sequences (Table 23) showed that their CI amino acid identities and differences varied from 96.7 to 99.7% and 0.3 to 3.4%, respectively. The CI amino acid sequence of PRSV-W has the similar ranges of identity (97.3-99.1%) and of difference (0.9-2.7%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the CI coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 38B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern

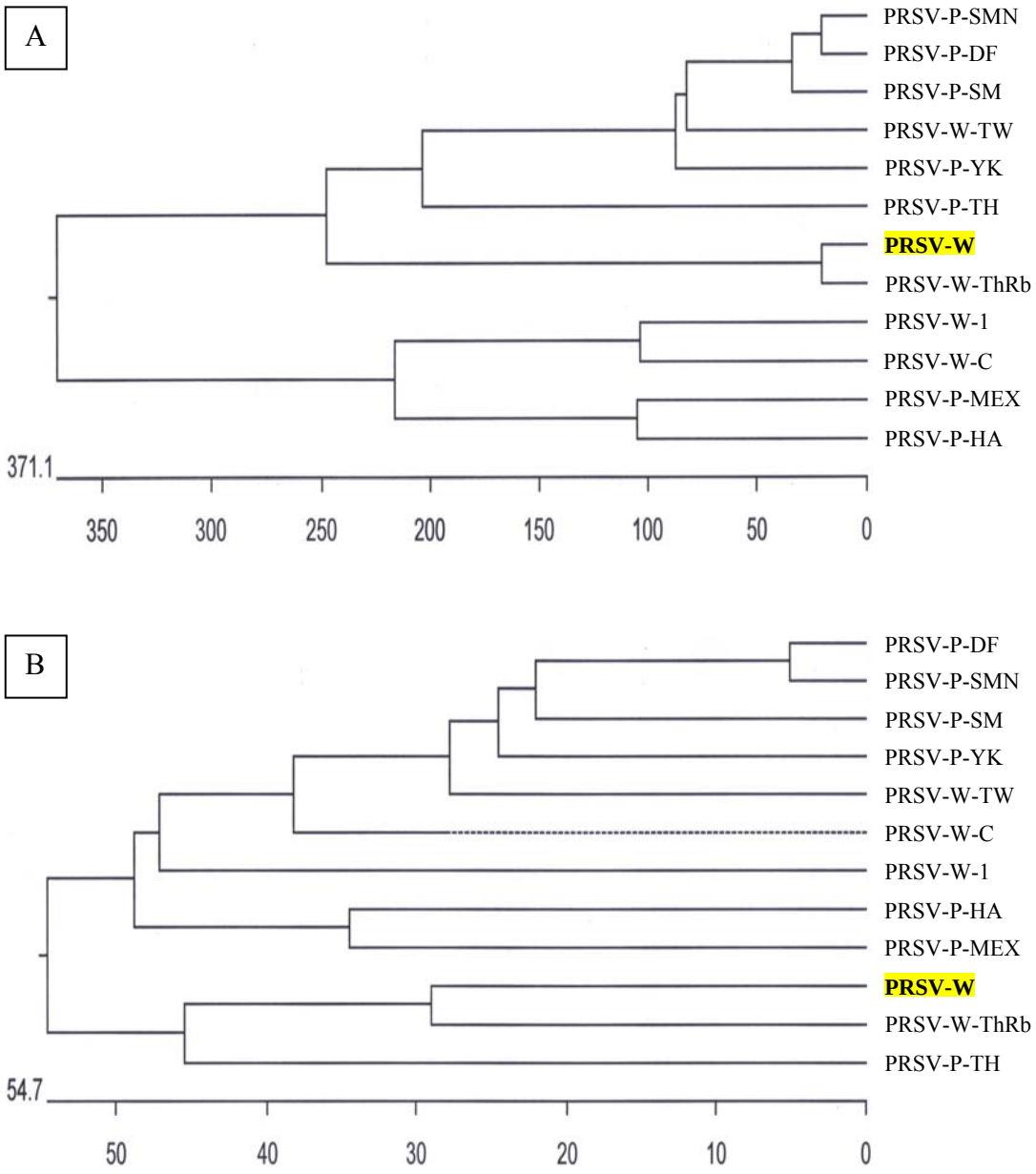


Figure 38 Phylogenetic trees constructed from the multiple alignments of the cylindrical inclusion (CI) protein coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

Table 23 Percent identities (top right) and differences (bottom left) between the putative cylindrical inclusion (CI) protein amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		99.1	97.8	98.1	97.5	97.6	98.1	97.3	97.6	97.5	97.8	97.5
PRSV-W-ThRb	0.9		98.4	98.7	98.1	98.3	98.7	98.0	98.3	98.1	98.4	98.1
PRSV-W-TW	2.2	1.6		99.1	98.1	97.6	98.1	98.6	97.8	98.7	99.1	98.7
PRSV-W-C	1.9	1.3	0.9		98.4	97.6	98.4	98.6	98.1	98.7	99.1	98.7
PRSV-W-1	2.6	1.9	1.9	1.6		96.7	97.8	97.6	97.5	97.8	98.1	97.8
PRSV-P-TH	2.4	1.8	2.4	2.4	3.4		97.6	97.2	97.3	97.3	97.6	97.8
PRSV-P-HA	1.9	1.3	1.9	1.6	2.2	2.4		97.6	98.4	97.8	98.1	97.8
PRSV-P-YK	2.7	2.1	1.4	1.4	2.4	2.9	2.4		97.3	98.6	98.9	98.6
PRSV-P-MEX	2.4	1.8	2.2	1.9	2.6	2.7	1.6	2.7		97.6	97.8	97.8
PRSV-P-DF	2.6	1.9	1.3	1.3	2.2	2.7	2.2	1.4	2.4		99.7	99.1
PRSV-P-SMN	2.2	1.6	0.9	0.9	1.9	2.4	1.9	1.1	2.2	0.3		99.1
PRSV-P-SM	2.6	1.9	1.3	1.3	2.2	2.2	2.2	1.4	2.2	0.9	0.9	

of the tree, the twelve isolates of PRSV species aligned could be divided into three clusters. Three Thai isolates including PRSV-W, PRSV-W-ThRb and PRSV-P-TH grouped together to form a cluster, with a further branch separating the two isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) from PRSV-P-TH (papaya-infecting). In another cluster, five Taiwanese isolates grouped with four non-Asian isolates (from Brazil, Mexico and USA) to form their own phylogenetic branch within the same cluster. As the result obtained from multiple alignment of CI coding region amino acid sequences (Table 23) and the branching pattern of the phylogenetic tree (Figure 38B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (99.1%) and the lowest distance (0.9%) with PRSV-W-ThRb isolate. The differences between the putative CI amino acid sequence of the PRSV-W isolate and those of the two Thai isolates placing within the same cluster was 0.9 to 2.4%, while it differed by 1.9 to 2.7% when compared with those of the isolates placing within another cluster in the same dendrogram.

7.1.7 Phylogenetic analyses based on the 6K2s

Sequence analysis demonstrated that the 6-kilodalton peptide 2(6K2) coding region of PRSV-W consists of 171 nucleotides, locating at nucleotide positions 6194 to 6364 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 57 amino acids (Table 7) with a calculated molecular weight of 6.4 kDa (Table 8). The PRSV-W 6K2 gene (171 nucleotides) and putative protein (57 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) 6K2 nucleotide sequences

Multiple alignment of the twelve 6K2 coding region nucleotide sequences (Table 24) showed that their 6K2 gene identities and differences varied from 83.7 to 100.0% and 0.0 to 19.1%, respectively. The 6K2 nucleotide sequence of PRSV-W has the ranges of identity (88.4-100.0%) and of difference (0.0-13.0%) when compared with the corresponding region of eleven other PRSV isolates.

Table 24 Percent identities (top right) and differences (bottom left) between the 6-kilodalton peptide 2 (6K2) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		100.0	89.0	90.1	88.4	89.5	90.7	88.4	88.4	89.5	88.4	88.4
PRSV-W-ThRb	0.0		89.0	90.1	88.4	89.5	90.7	88.4	88.4	89.5	88.4	88.4
PRSV-W-TW	12.3	12.3		85.5	83.7	89.0	87.8	93.6	85.5	94.2	94.2	94.2
PRSV-W-C	10.8	10.8	16.8		93.6	89.5	91.3	87.2	92.4	84.3	84.9	84.9
PRSV-W-1	12.9	12.9	19.1	6.8		87.8	89.5	87.8	88.4	83.7	85.5	85.5
PRSV-P-TH	11.5	11.5	12.3	11.5	13.7		87.2	91.9	86.6	88.4	87.8	87.8
PRSV-P-HA	10.1	10.1	13.7	9.4	11.5	14.4		88.4	94.2	87.2	86.6	86.6
PRSV-P-YK	13.0	13.0	6.8	14.4	13.6	8.7	13.0		87.2	93.0	93.6	93.6
PRSV-P-MEX	13.0	13.0	16.9	8.1	13.0	15.3	6.1	14.6		84.9	84.3	84.3
PRSV-P-DF	11.4	11.4	6.1	18.2	18.9	13.0	14.4	7.4	17.6		97.1	97.1
PRSV-P-SMN	13.0	13.0	6.2	17.6	16.7	13.7	15.3	6.8	18.6	3.0		100.0
PRSV-P-SM	13.0	13.0	6.2	17.6	16.7	13.7	15.3	6.8	18.6	3.0	0.0	

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the 6K2 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 39A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Six Asian isolates including of all isolates from Taiwan and PRSV-P-TH from Thailand grouped together to form their own phylogenetic branch within the same cluster. In this cluster, five Taiwanese isolates from Taiwan grouped together, with a further branch separating the five isolates as a sub-cluster from PRSV-P-TH (Thai isolate). In another cluster, there are two sub-clusters separating the two Thai isolates (PRSV-W and PRSV-W-ThRb) grouped together separating from another sub-cluster consisting of PRSV-W-1 (from Brazil), PRSV-W-C (from Brazil), PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In this cluster, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from the four non-Asian isolates. As the result obtained from multiple alignment of 6K1 coding region nucleotide sequences, the PRSV-W isolate has the highest identity (100.0%) and the lowest distance (0.0%) with PRSV-W-ThRb isolate.

b) 6K2 amino acid sequences

Multiple alignment of the twelve putative 6K2 amino acid sequences (Table 25) showed that their 6K2 amino acid identities and differences varied from 91.4 to 100.0% and 0.0 to 9.2%, respectively. The 6K2 amino acid sequence of PRSV-W has the ranges of identity (93.1-100.0%) and of difference (0.0-7.2%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the 6K2 coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 39B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into three

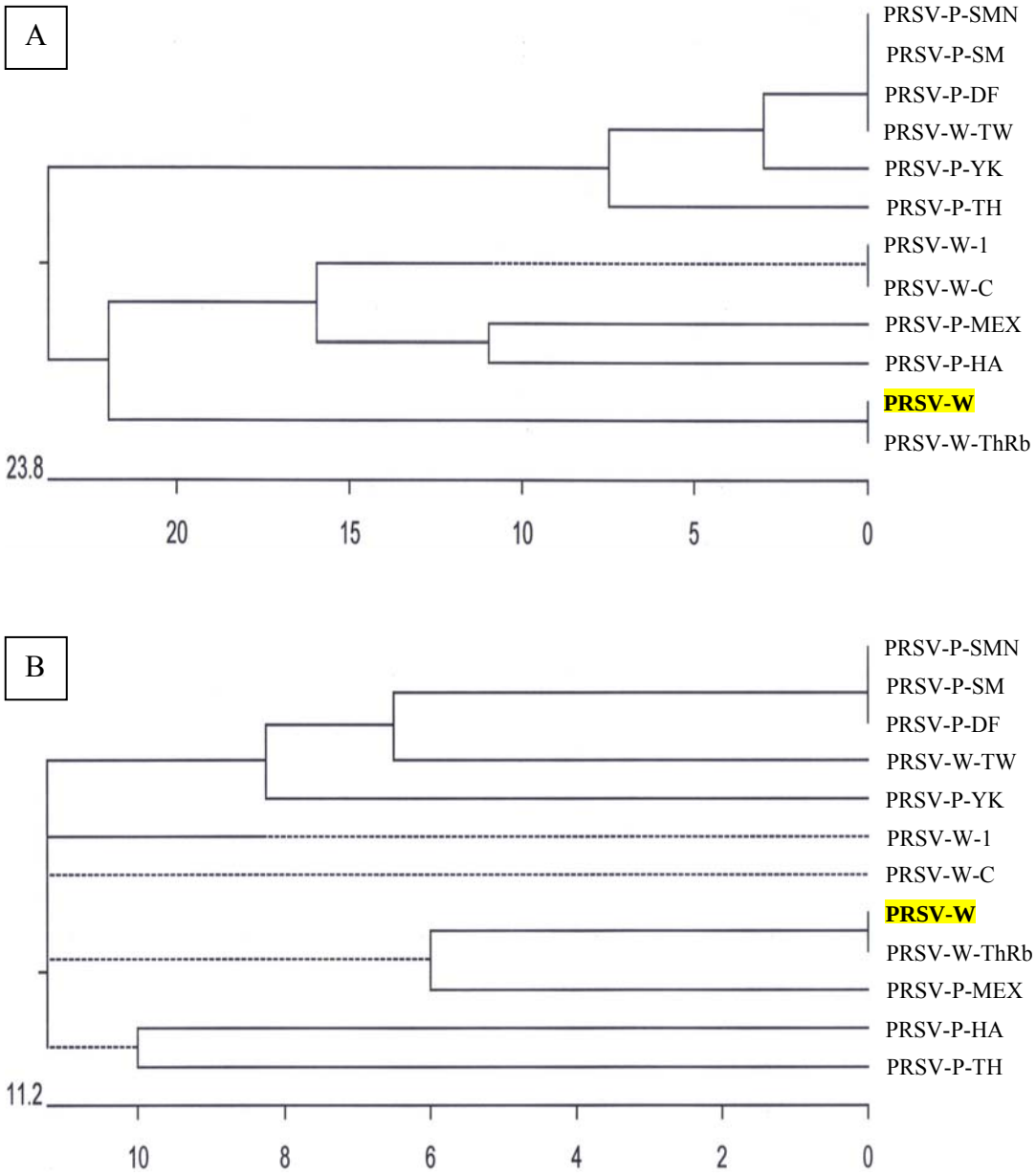


Figure 39 Phylogenetic trees constructed from the multiple alignments of the 6-kilodalton peptide 2 (6K2) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

Table 25 Percent identities (top right) and differences (bottom left) between the putative 6-kilodalton peptide 2 (6K2) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		100.0	93.1	98.3	98.3	96.6	94.8	93.1	94.8	93.1	93.1	93.1
PRSV-W-ThRb	0.0		93.1	98.3	98.3	96.6	94.8	93.1	94.8	93.1	93.1	93.1
PRSV-W-TW	7.2	7.2		94.8	94.8	93.1	91.4	93.1	91.4	96.6	96.6	96.6
PRSV-W-C	1.7	1.7	5.4		100.0	98.3	96.6	94.8	96.6	94.8	94.8	94.8
PRSV-W-1	1.7	1.7	5.4	0.0		98.3	96.6	94.8	96.6	94.8	94.8	94.8
PRSV-P-TH	3.5	3.5	7.2	1.7	1.7		96.6	93.1	94.8	93.1	93.1	93.1
PRSV-P-HA	5.4	5.4	9.2	3.5	3.5	3.5		91.4	93.1	91.4	91.4	91.4
PRSV-P-YK	7.2	7.2	7.2	5.4	5.4	7.2	9.2		91.4	93.1	93.1	93.1
PRSV-P-MEX	5.4	5.4	9.2	3.5	3.5	5.4	7.2	9.2		91.4	91.4	91.4
PRSV-P-DF	7.2	7.2	3.5	5.4	5.4	7.2	9.2	7.2	9.2		100.0	100.0
PRSV-P-SMN	7.2	7.2	3.5	5.4	5.4	7.2	9.2	7.2	9.2	0.0		100.0
PRSV-P-SM	7.2	7.2	3.5	5.4	5.4	7.2	9.2	7.2	9.2	0.0	0.0	

clusters. In the tree, both isolates from Brazil (PRSV-W-1 and PRSV-W-C) did not group with other isolates, they were separated individually by phylogenetic branching from all remaining isolates. First cluster was placed with five PRSV isolates from Taiwan including PRSV-P-YK, PRSV-W-TW, PRSV-P-DF, PRSV-P-SM and PRSV-P-SMN. In second cluster, PRSV-W grouped together with PRSV-W-ThRb as a sub-cluster separating from PRSV-P-MEX in the same cluster. PRSV-P-HA (from USA) grouped with PRSV-P-TH (from Thailand) to form their own phylogenetic branch as the third cluster. As the result obtained from multiple alignment of 6K2 coding region amino acid sequences (Table 25) and the branching pattern of the phylogenetic tree (Figure 39B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (100.0%) and the lowest distance (0.0%) with PRSV-W-ThRb isolate. The differences between the putative 6K2 amino acid sequence of the PRSV-W isolate and those of the Taiwanese isolates placing within the first cluster was 7.2%, while it differed by 3.5 to 5.4% when compared with those of the two isolates placing within the third cluster in the same dendrogram.

7.1.8 Phylogenetic analyses based on the VPgs

Sequence analysis demonstrated that the genome-linked protein (VPg) coding region of PRSV-W consists of 567 nucleotides, locating at nucleotide positions 6365 to 6931 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 189 amino acids (Table 7) with a calculated molecular weight of 21.4 kDa (Table 8). The PRSV-W VPg gene (567 nucleotides) and putative protein (189 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) VPg nucleotide sequences

Multiple alignment of the twelve VPg coding region nucleotide sequences (Table 26) showed that their VPg gene identities and differences varied from 81.0 to 100.0% and 0.0 to 22.7%, respectively. The VPg nucleotide sequence of

Table 26 Percent identities (top right) and differences (bottom left) between the genome-linked protein (VPg) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.6	87.0	83.5	84.5	86.6	82.6	89.4	84.2	87.9	87.5	87.5
PRSV-W-ThRb	1.4		87.5	83.1	84.2	87.1	82.6	90.0	84.5	88.4	88.0	88.0
PRSV-W-TW	14.8	14.1		81.3	82.6	89.8	81.0	94.5	81.2	96.0	95.2	95.2
PRSV-W-C	19.3	19.8	22.3		96.7	83.8	95.1	82.0	93.1	82.2	81.7	81.7
PRSV-W-1	17.9	18.3	20.5	3.5		84.7	94.0	82.9	92.1	82.7	82.2	82.2
PRSV-P-TH	15.2	14.5	11.2	18.8	17.6		82.7	89.1	83.1	90.1	89.4	89.4
PRSV-P-HA	20.5	20.5	22.7	5.1	6.3	20.1		81.9	94.4	81.3	81.0	81.0
PRSV-P-YK	11.7	11.0	5.7	21.3	20.1	12.1	21.5		82.4	94.4	94.0	94.0
PRSV-P-MEX	18.3	17.8	22.5	7.3	8.5	19.6	5.9	20.8		81.5	81.2	81.2
PRSV-P-DF	13.7	13.1	4.2	21.0	20.3	10.8	22.2	5.9	22.0		98.2	98.2
PRSV-P-SMN	14.1	13.5	5.0	21.8	21.0	11.7	22.7	6.3	22.5	1.8		100.0
PRSV-P-SM	14.1	13.5	5.0	21.8	21.0	11.7	22.7	6.3	22.5	1.8	0.0	

PRSV-W has the ranges of identity (82.6-98.6%) and of difference (1.4-20.5%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the VPg coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 40A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. All five isolates from Taiwan and PRSV-P-TH from Thailand grouped together to form their own phylogenetic branch within the same cluster. In another cluster, there are two sub-clusters separating the two Thai isolates (PRSV-W and PRSV-W-ThRb) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico), PRSV-P-HA (from USA), PRSV-W-C (from Brazil) and PRSV-W-1 (from Brazil). The PRSV-W isolate is most closely related to PRSV-W-ThRb isolate. As the result obtained from multiple alignment of VPg coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.6%) and the lowest distance (1.4%) with PRSV-W-ThRb isolate.

b) VPg amino acid sequences

Multiple alignment of the twelve putative VPg amino acid sequences (Table 27) showed that their VPg amino acid identities and differences varied from 92.1 to 100.0% and 0.0 to 8.4%, respectively. The VPg amino acid sequence of PRSV-W has the ranges of identity (94.7-98.9%) and of difference (1.1-5.5%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the VPg coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 40B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own

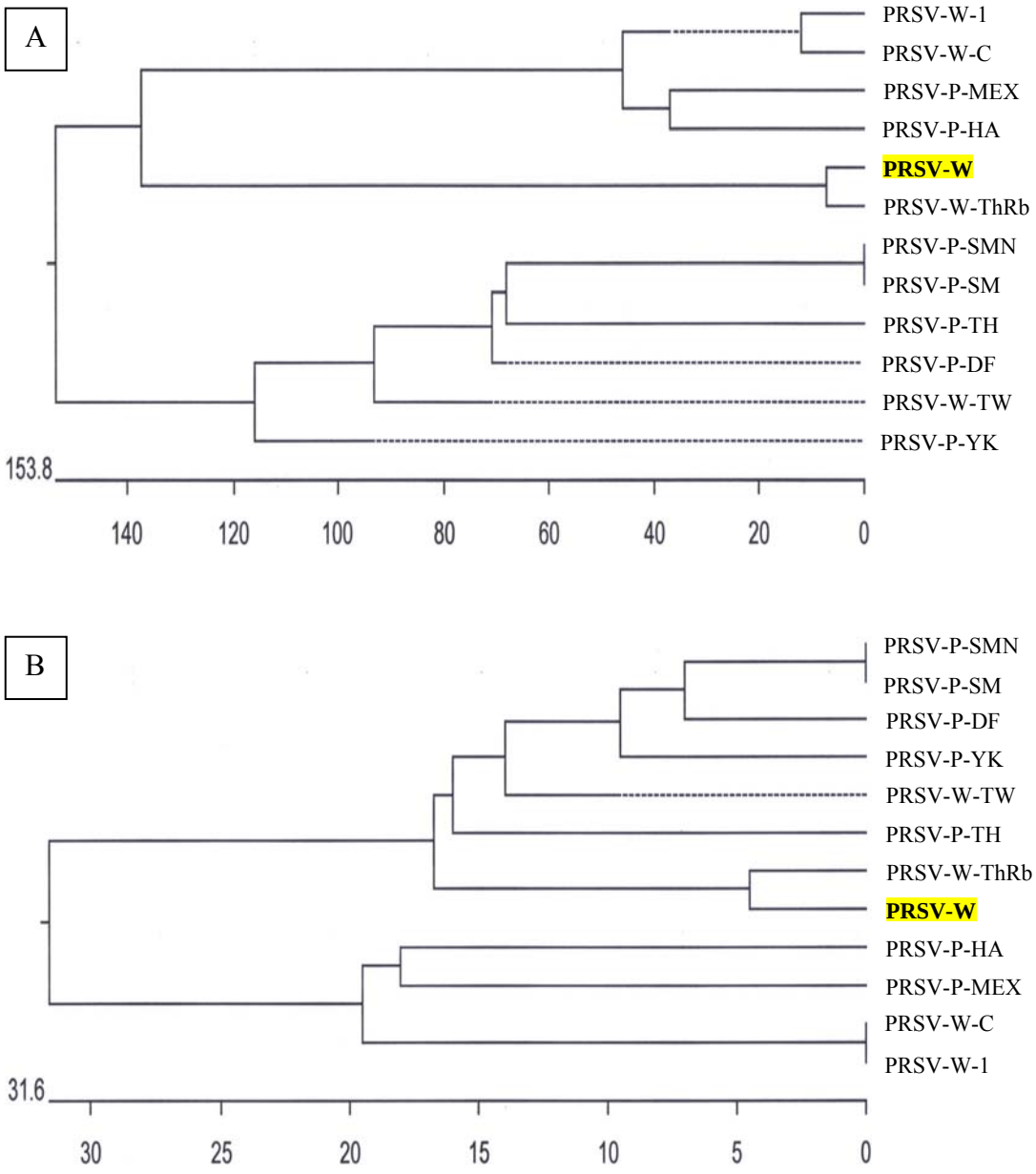


Figure 40 Phylogenetic trees constructed from the multiple alignments of the genome-linked protein (VPg) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

Table 27 Percent identities (top right) and differences (bottom left) between the putative genome-linked protein (VPg) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.9	96.8	96.3	96.3	95.8	94.7	96.3	94.7	96.8	95.8	95.8
PRSV-W-ThRb	1.1		97.4	95.3	95.3	96.3	94.2	96.8	94.7	97.4	96.3	96.3
PRSV-W-TW	3.2	2.7		94.7	94.7	95.8	93.2	98.4	93.7	97.9	96.8	96.8
PRSV-W-C	3.8	4.9	5.5		100.0	93.7	96.3	94.2	96.3	93.7	92.6	92.6
PRSV-W-1	3.8	4.9	5.5	0.0		93.7	96.3	94.2	96.3	93.7	92.6	92.6
PRSV-P-TH	4.3	3.8	4.3	6.6	6.6		92.1	95.3	93.7	96.8	95.8	95.8
PRSV-P-HA	5.5	6.0	7.2	3.8	3.8	8.4		93.7	96.3	93.2	92.1	92.1
PRSV-P-YK	3.8	3.2	1.6	6.0	6.0	4.9	6.6		94.2	98.4	97.4	97.4
PRSV-P-MEX	5.5	5.5	6.6	3.8	3.8	6.6	3.8	6.0		93.7	92.6	92.6
PRSV-P-DF	3.2	2.7	2.1	6.6	6.6	3.2	7.2	1.6	6.6		98.9	98.9
PRSV-P-SMN	4.3	3.8	3.2	7.8	7.8	4.3	8.4	2.7	7.8	1.1		100.0
PRSV-P-SM	4.3	3.8	3.2	7.8	7.8	4.3	8.4	2.7	7.8	1.1	0.0	

phylogenetic branch within the same cluster, while the four non-Asian isolates including PRSV-W-1, PRSV-W-C and PRSV-P-MEX grouped together to form another cluster with two branches separating the two Brazilian isolates from the another branch of two isolates from Mexico and USA. In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate (1.1% different amino acids) and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates. The differences between the VPg amino acid sequence of the PRSV-W isolate and those of the Asian isolates within the same cluster (with the exception of the PRSV-W-ThRb isolate) was 3.2 to 4.3%, while it differed by 3.8 to 5.5% when compared with those of the non-Asian isolates placing within another cluster in the same dendrogram.

7.1.9 Phylogenetic analyses based on the NIa-Pros

Sequence analysis demonstrated that the nuclear inclusion a proteinase (NIa-Pro) coding region of PRSV-W consists of 714 nucleotides, locating at nucleotide positions 6932 to 7645 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 238 amino acids (Table 7) with a calculated molecular weight of 26.5 kDa (Table 8). The PRSV-W NIa gene (714 nucleotides) and putative protein (238 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) NIa-Pro nucleotide sequences

Multiple alignment of the twelve NIa-Pro coding region nucleotide sequences (Table 28) showed that their NIa-Pro gene identities and differences varied from 81.7 to 99.2% and 0.8 to 21.6%, respectively. The NIa-Pro nucleotide sequence of PRSV-W has the similar ranges of identity (81.7-99.2%) and of difference (0.8-21.5%) when compared with the corresponding region of eleven other PRSV isolates.

Table 28 Percent identities (top right) and differences (bottom left) between the nuclear inclusion a proteinase (NIa) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		99.2	88.7	82.5	82.4	87.1	82.7	88.0	81.7	87.0	86.9	86.7
PRSV-W-ThRb	0.8		89.0	82.8	82.7	87.4	82.9	88.3	82.0	87.3	87.1	87.0
PRSV-W-TW	12.6	12.3		82.9	84.2	89.9	82.7	95.1	82.2	94.8	95.0	94.7
PRSV-W-C	20.5	20.1	19.9		95.2	83.4	91.7	82.4	90.8	82.8	83.4	82.5
PRSV-W-1	20.7	20.3	18.2	5.0		83.1	91.7	83.4	91.2	83.2	83.8	83.5
PRSV-P-TH	14.6	14.2	11.1	19.3	19.7		83.4	91.3	82.9	89.5	89.9	89.4
PRSV-P-HA	20.2	19.8	20.2	8.9	8.9	19.2		83.1	95.7	81.8	82.7	82.5
PRSV-P-YK	13.5	13.2	5.1	20.7	19.3	9.4	19.6		82.5	95.2	95.7	95.7
PRSV-P-MEX	21.5	21.1	20.8	10.1	9.5	19.8	4.5	20.4		81.7	82.2	82.5
PRSV-P-DF	14.8	14.4	5.4	20.1	19.5	11.6	21.4	5.0	21.6		98.5	96.5
PRSV-P-SMN	14.9	14.6	5.3	19.3	18.7	11.0	20.2	4.5	20.8	1.6		96.9
PRSV-P-SM	15.1	14.8	5.6	20.5	19.1	11.7	20.4	4.5	20.4	3.6	3.2	

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the NIa coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 41A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-MEX (from Mexico) and PRSV-P-HA (from USA). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates including PRSV-P-TH, PRSV-W-TW, PRSV-P-YK, PRSV-P-SMN, PRSV-P-SM and PRSV-P-DF. As the result obtained from multiple alignment of NIa coding region nucleotide sequences, the PRSV-W isolate has the highest identity (99.2%) and the lowest distance (0.8%) with PRSV-W-ThRb isolate.

b) NIa-Pro amino acid sequences

Multiple alignment of the twelve putative NIa-Pro amino acid sequences (Table 29) showed that their NIa-Pro amino acid identities and differences varied from 92.9 to 99.6% and 0.4 to 7.5%, respectively. The NIa-Pro amino acid sequence of PRSV-W has the ranges of identity (92.9-98.3%) and of difference (1.7-7.5%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the NIa coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 41B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Two cucurbit-infecting Thai isolates (PRSV-W and PRSV-W ThRb) and

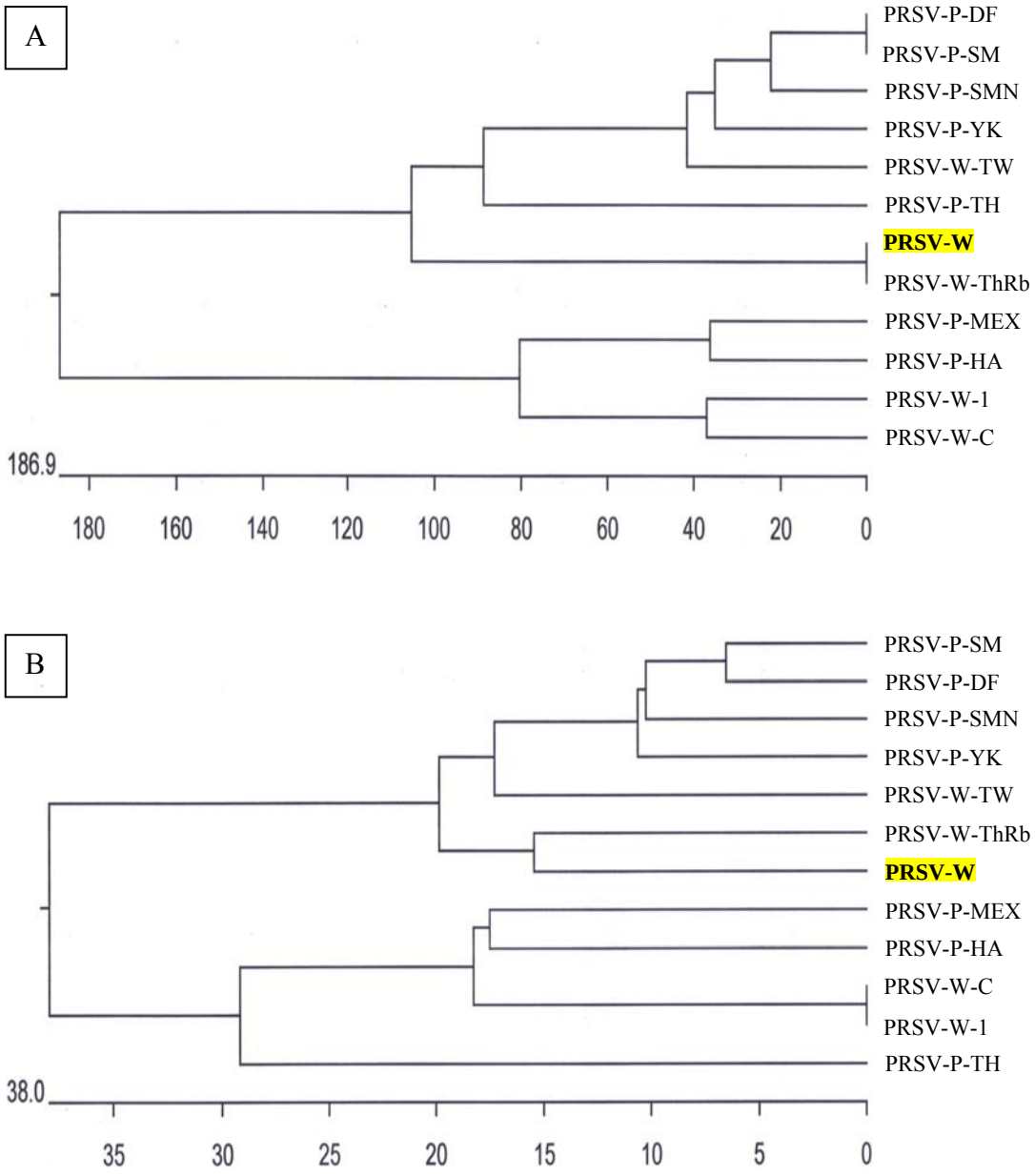


Figure 41 Phylogenetic trees constructed from the multiple alignments of the nuclear inclusion a proteinase (NIa-Pro) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

Table 29 Percent identities (top right) and differences (bottom left) between the putative nuclear inclusion a proteinase (NIa) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.3	96.2	95.4	95.8	95.4	93.7	95.4	92.9	94.6	95.0	94.1
PRSV-W-ThRb	1.7		96.2	95.4	95.8	95.4	93.7	95.4	92.9	94.6	95.0	94.1
PRSV-W-TW	3.9	3.9		95.4	95.8	94.6	93.7	97.1	93.3	96.2	97.1	95.8
PRSV-W-C	4.8	4.8	4.8		99.6	95.8	97.5	95.8	96.7	95.0	95.0	96.2
PRSV-W-1	4.3	4.3	4.3	0.4		96.2	97.9	96.2	97.1	95.4	95.4	95.8
PRSV-P-TH	4.8	4.8	5.7	4.3	3.9		95.0	96.2	94.1	95.4	95.4	95.0
PRSV-P-HA	6.6	6.6	6.6	2.6	2.1	5.2		95.0	97.5	94.1	94.1	94.6
PRSV-P-YK	4.8	4.8	3.0	4.3	3.9	3.9	5.2		94.6	98.3	98.3	97.9
PRSV-P-MEX	7.5	7.5	7.0	3.4	3.0	6.1	2.6	5.7		94.6	94.6	94.1
PRSV-P-DF	5.7	5.7	3.9	5.2	4.8	4.8	6.1	1.7	5.7		98.3	98.7
PRSV-P-SMN	5.2	5.2	3.0	5.2	4.8	4.8	6.	1.7	5.7	1.7		97.1
PRSV-P-SM	6.1	6.1	4.3	3.9	4.3	5.2	5.7	2.1	6.1	1.3	3.0	

five Taiwanese isolates grouped together to form their own phylogenetic branch within the same cluster, the two Thai isolates grouped together, with a further branch separating the two Thai isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) as a sub-cluster from five Taiwanese isolates of PRSV. In another cluster, PRSV-P-TH clustered with the four non-Asian isolates, with a further branch separating the PRSV-P-TH isolates from a sub-cluster of the four non-Asian isolates including PRSV-W-1, PRSV-W-C, PRSV-P-HA and PRSV-P-MEX. As the result obtained from multiple alignment of NIa coding region amino acid sequences (Table 29) and the branching pattern of the phylogenetic tree (Figure 41B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (98.3%) and the lowest distance (1.7%) with PRSV-W-ThRb isolate. The differences between the putative NIa amino acid sequence of the PRSV-W isolate and those of the five Taiwanese isolates placing within the same cluster was 3.9 to 6.1%, while it differed by 4.3 to 7.5% when compared with those of the isolates placing within another cluster in the same dendrogram.

7.1.10 Phylogenetic analyses based on the NIbs

Sequence analysis demonstrated that the large nuclear inclusion protein (NIb) coding region of PRSV-W consists of 1611 nucleotides, locating at nucleotide positions 7646 to 9256 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 537 amino acids (Table 7) with a calculated molecular weight of 61.7 kDa (Table 8). The PRSV-W NIb gene (1611 nucleotides) and putative protein (537 amino acids) were identical in size with all available eleven isolates of PRSV which were fully-sequenced and submitted to GenBank.

a) NIb nucleotide sequences

Multiple alignment of the twelve NIb coding region nucleotide sequences (Table 30) showed that their NIb gene identities and differences varied from 81.3 to 99.1% and 0.9 to 22.3%, respectively. The NIb nucleotide sequence of

PRSV-W has the ranges of identity (81.8-99.1%) and of difference (0.9-21.6%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the NIb coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 42A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Eight Asian isolates both from Taiwan and Thailand formed their own phylogenetic branch within the same cluster, while four non-Asian isolates grouped together to form another cluster. In the cluster of non-Asian PRSV isolates, there are two sub-clusters separating the two Brazilian isolates (PRSV-W-1 and PRSV-W-C) grouped together from another sub-cluster consisting of PRSV-P-HA (from USA) and PRSV-P-MEX (from Mexico). In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates including PRSV-P-TH, PRSV-P-YK, PRSV-W-TW, PRSV-P-SMN, PRSV-P-DF and PRSV-P-SM. As the result obtained from multiple alignment of NIb coding region nucleotide sequences, the PRSV-W isolate has the highest identity (99.1%) and the lowest distance (0.9%) with PRSV-W-ThRb isolate.

b) NIb amino acid sequences

Multiple alignment of the twelve putative NIb amino acid sequences (Table 31) showed that their NIb amino acid identities and differences varied from 93.1 to 98.9% and 1.1 to 7.2%, respectively. The NIb amino acid sequence of PRSV-W has the ranges of identity (94.8-98.5%) and of difference (1.5-5.4%) when compared with the corresponding region of eleven other PRSV isolates.

Table 30 Percent identities (top right) and differences (bottom left) between the large nuclear inclusion protein (NIb) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		99.1	89.4	81.8	82.1	88.3	83.3	89.8	83.1	89.0	88.6	89.8
PRSV-W-ThRb	0.9		89.6	81.9	82.1	88.5	83.5	90.0	83.2	89.3	88.8	90.0
PRSV-W-TW	11.7	11.4		82.9	82.7	89.9	82.8	94.4	82.9	95.8	95.6	96.3
PRSV-W-C	21.6	21.4	20.0		94.7	81.3	91.6	82.5	91.1	82.1	81.4	82.4
PRSV-W-1	21.2	21.2	20.2	5.5		81.5	91.4	82.4	90.6	81.6	81.3	81.9
PRSV-P-TH	13.1	12.9	11.2	22.2	21.9		82.1	88.8	82.2	89.0	89.0	89.3
PRSV-P-HA	19.4	19.2	20.2	9.1	9.3	21.2		82.6	93.1	81.7	81.8	82.1
PRSV-P-YK	11.3	11.0	5.9	20.5	20.7	12.5	20.4		83.1	94.5	94.2	95.2
PRSV-P-MEX	19.8	19.6	19.9	9.7	10.3	20.9	7.3	19.7		82.3	82.1	82.8
PRSV-P-DF	12.2	11.9	4.4	21.0	21.9	12.2	21.7	5.7	20.9		96.4	98.3
PRSV-P-SMN	12.7	12.4	4.6	22.1	22.3	12.2	21.6	6.2	21.0	3.7		97.0
PRSV-P-SM	11.3	11.0	3.9	20.7	21.4	11.8	21.2	5.1	20.2	1.7	3.1	

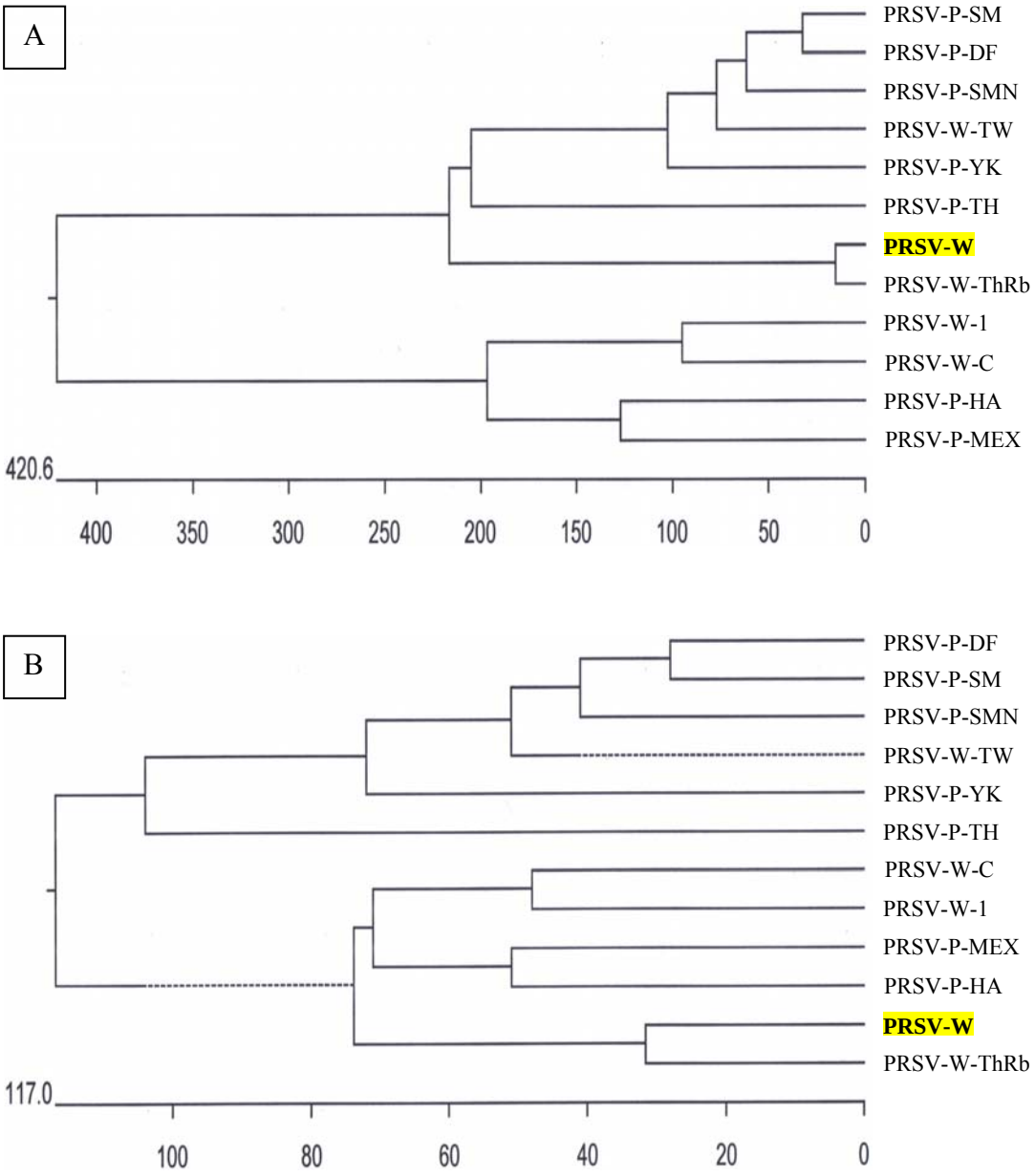


Figure 42 Phylogenetic trees constructed from the multiple alignments of the large nuclear inclusion protein (NIb) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

Table 31 Percent identities (top right) and differences (bottom left) between the putative large nuclear inclusion protein (NIb) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.5	95.5	94.8	95.5	94.8	95.9	95.4	95.2	94.8	95.5	95.5
PRSV-W-ThRb	1.5		96.1	95.4	95.7	95.2	96.7	95.9	95.9	95.4	95.9	96.1
PRSV-W-TW	4.6	4.0		94.1	94.6	95.0	94.8	96.7	94.4	97.4	97.6	98.0
PRSV-W-C	5.4	4.8	6.2		97.0	93.1	95.5	93.5	95.0	93.3	93.5	93.7
PRSV-W-1	4.6	4.4	5.6	3.0		93.5	96.5	93.9	95.9	93.3	93.7	93.9
PRSV-P-TH	5.4	5.0	5.2	7.2	6.8		93.9	94.1	93.5	93.9	94.4	94.4
PRSV-P-HA	4.2	3.4	5.4	4.6	3.6	6.4		94.6	96.8	93.7	94.6	94.4
PRSV-P-YK	4.8	4.2	3.4	6.8	6.4	6.2	5.6		94.2	96.5	97.0	97.2
PRSV-P-MEX	5.0	4.2	5.8	5.2	4.2	6.8	3.2	6.0		93.3	94.1	94.1
PRSV-P-DF	5.4	4.8	2.7	7.0	7.0	6.4	6.6	3.6	7.0		97.6	98.9
PRSV-P-SMN	4.6	4.2	2.5	6.8	6.6	5.8	5.6	3.0	6.2	2.5		98.3
PRSV-P-SM	4.6	4.0	2.1	6.6	6.4	5.8	5.8	2.8	6.2	1.1	1.7	

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the NIb coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 42B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. Two cucurbit-infecting isolates (PRSV-W and PRSV-W ThRb) from Thailand and four non-Asian isolates grouped together to form their own phylogenetic branch within the same cluster, the two Thai isolates grouped together, with a further branch separating the two Thai isolates of cucurbit-infecting PRSV as a sub-cluster from the four non-Asian isolates of PRSV. In another cluster, PRSV-P-TH clustered with all five Taiwanese isolates, with a further branch separating the PRSV-P-TH isolates from a sub-cluster of the Taiwanese isolates including PRSV-P-YK, PRSV-W-TW, PRSV-P-SMN, PRSV-P-SM and PRSV-P-DF. As the result obtained from multiple alignment of NIb coding region amino acid sequences (Table 31) and the branching pattern of the phylogenetic tree (Figure 42B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (98.5%) and the lowest distance (1.5%) with PRSV-W-ThRb isolate. The differences between the putative NIb amino acid sequence of the PRSV-W isolate and those of the four non-Asian isolates placing within the same cluster was 4.2 to 5.4%, while it differed by 4.6 to 5.4% when compared with those of the isolates placing within another cluster in the same dendrogram.

7.1.11 Phylogenetic analyses based on the CPs

Sequence analysis demonstrated that the coat protein (CP) coding region of PRSV-W consists of 858 nucleotides, locating at nucleotide positions 9257 to 10114 on the viral genome and representing characteristic PRSV RNA, which encodes a putative protein of 286 amino acids (Table 7) with a calculated molecular weight of 32.6 kDa (Table 8).

a) CP nucleotide sequences

Over the entire CP coding region, size comparisons between nucleotide sequences of the different CP coding regions of the available complete PRSV sequences showed that the CP sizes of these PRSV isolates range from 855 to 867 nucleotides. The PRSV-W CP (858 nucleotides) is the third largest CP of all completely sequenced PRSV isolates. It is 9 nucleotides smaller than those of PRSV-W-TW (AY027810) and PRSV-W-1 (DQ374153) isolates which their CP gene consist of 867 nucleotides, and 3 nucleotides smaller than those of PRSV-W-C (DQ374152), PRSV-P-HA (X67673), PRSV-P-YK (X97251), PRSV-P-DF (DQ340769), PRSV-P-SMN (DQ340770) and PRSV-P-SM (DQ340771) isolates which their CP gene consist of 861 nucleotides. The entire CP gene of the PRSV-W was identical in size (858 nucleotides) with the two other PRSV isolates from the same country (Thailand) including PRSV-W-ThRb (AY010722) and PRSV-P-TH (AY162218). However, the CP gene of the PRSV-W was considerably larger than that of PRSV-P-MEX (855 nucleotides).

Multiple alignment of the twelve CP coding region nucleotide sequences (Table 32) showed that their CP gene identities and differences varied from 89.3 to 98.8% and 1.2 to 11.9%, respectively. The CP nucleotide sequence of PRSV-W has the similar ranges of identity (89.3-98.8%) and of difference (1.2-1.2%) when compared with the corresponding region of eleven other PRSV isolates.

A phylogenetic tree constructed from the multiple alignment using the nucleotide sequences of the CP coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 43A) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. All four non-Asian isolates (PRSV-W-C, PRSV-W-1, PRSV-P-MEX and PRSV-P-HA) clustered with all three Thai isolates (PRSV-W, PRSV-W-ThRb and PRSV-P-TH), with two further branches separating between sub-clusters of non-Asian isolates and Thai isolates according to their geographic locations. While all

Table 32 Percent identities (top right) and differences (bottom left) between the coat protein (CP) nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.8	93.0	90.2	89.7	92.2	90.9	92.8	89.3	92.7	92.8	93.0
PRSV-W-ThRb	1.2		93.7	90.4	90.2	93.1	91.5	93.6	89.8	93.4	93.6	93.7
PRSV-W-TW	7.5	6.7		92.6	91.6	93.8	92.0	96.9	90.7	97.7	97.3	97.7
PRSV-W-C	10.7	10.4	7.9		95.9	90.4	95.0	91.8	93.7	92.5	92.7	92.7
PRSV-W-1	11.3	10.7	9.0	4.2		90.0	94.4	90.8	93.9	91.9	91.6	91.8
PRSV-P-TH	8.4	7.3	6.6	10.4	11.0		90.7	93.3	89.5	93.8	93.7	93.9
PRSV-P-HA	9.9	9.2	8.6	5.2	5.9	10.1		91.4	95.6	92.2	92.3	92.6
PRSV-P-YK	7.7	6.9	3.2	8.9	10.0	7.1	9.3		89.6	96.3	96.3	96.5
PRSV-P-MEX	11.9	11.1	10.1	6.7	6.4	11.6	4.6	11.4		90.5	90.6	90.7
PRSV-P-DF	7.8	7.0	2.4	8.0	8.7	6.6	8.3	3.8	10.4		98.3	98.8
PRSV-P-SMN	7.6	6.8	2.7	7.7	9.0	6.7	8.2	3.8	10.2	1.8		98.7
PRSV-P-SM	7.5	9.7	2.4	7.8	8.9	6.4	7.9	3.6	10.1	1.2	1.3	

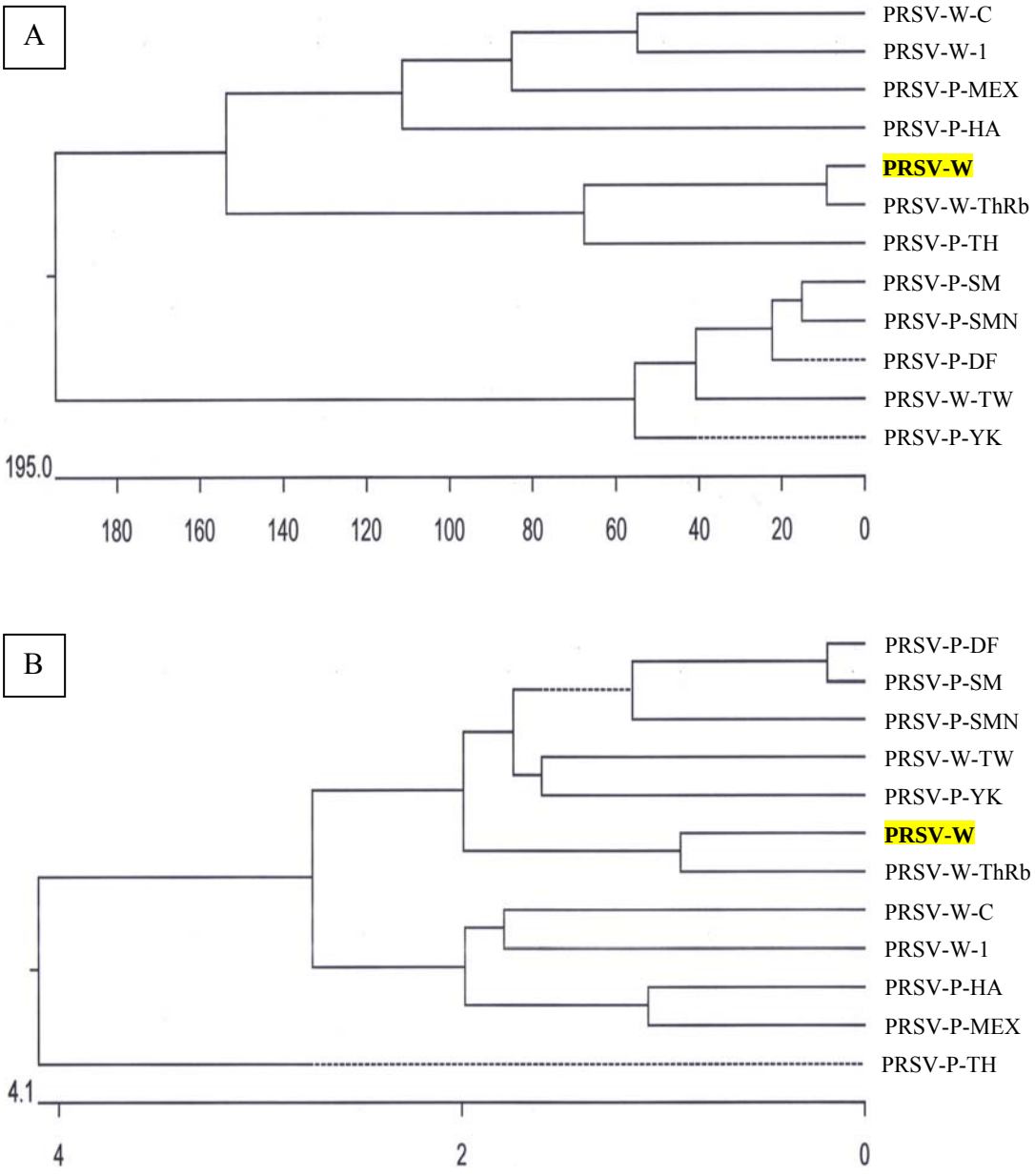


Figure 43 Phylogenetic trees constructed from the multiple alignments of the coat protein (CP) coding region based on the nucleotide (A) and the putative amino acid (B) sequences of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

five Taiwanese isolates (PRSV-P-SM, PRSV-P-SMN, PRSV-P-DF, PRSV-W-TW and PRSV-P-YK) grouped together to form a cluster separating from the first cluster. In the sub-cluster of PRSV isolates from Thailand, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch separated the two cucurbit-infecting isolates from the papaya-infecting isolates (PRSV-P-TH) placing within the same sub-cluster. As the result obtained from multiple alignment of CP coding region nucleotide sequences, the PRSV-W isolate has the highest identity (98.8%) and the lowest distance (1.2%) with PRSV-W-ThRb isolate.

b) CP amino acid sequences

Sizes comparisons between amino acid sequences of the CPs of the available complete PRSV sequences showed that the sizes of different PRSV CPs ranged from 285 to 289 amino acids. The putative CP of the PRSV-W resulted from computer analyzing in this research (286 amino acids) is 3 amino acids smaller than those of PRSV-W-TW (AY027810) and PRSV-W-1 (DQ374153) isolates which their CPs consist of 289 amino acids, and one amino acid smaller than those of PRSV-W-C (DQ374152), PRSV-P-HA (X67673), PRSV-P-YK (X97251), PRSV-P-DF (DQ340769), PRSV-P-SMN (DQ340770) and PRSV-P-SM (DQ340771) isolates which their putative CPs consist of 287 amino acids. The putative CP of the PRSV-W (585 amino acids) was identical in size with the corresponding region of two other PRSV isolates from the same country (Thailand) including PRSV-W-ThRb (AY010722) and PRSV-P-TH (AY162218), while the size of the PRSV-W CP was larger than that of PRSV-P-MEX (285 amino acids).

Multiple alignment of the twelve putative CP amino acid sequences (Table 33) showed that their CP amino acid identities and differences varied from 92.8 to 99.7% and 0.3 to 7.0%, respectively. The CP amino acid sequence of PRSV-W has the ranges of identity (92.8-98.3%) and of difference (1.8-6.6%) when compared with the corresponding region of eleven other PRSV isolates.

Table 33 Percent identities (top right) and differences (bottom left) between the putative coat protein (CP) amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.3	94.9	93.5	92.8	94.2	93.8	94.5	93.8	95.5	95.2	95.5
PRSV-W-ThRb	1.8		96.6	94.9	94.2	95.9	95.5	96.2	95.5	97.3	96.9	97.6
PRSV-W-TW	4.3	2.5		93.5	93.2	94.9	94.9	96.2	94.5	97.3	96.6	97.6
PRSV-W-C	6.6	5.1	5.4		95.9	93.2	95.2	93.2	95.5	94.2	93.5	94.5
PRSV-W-1	6.6	5.1	5.1	3.6		93.2	94.5	92.8	94.9	93.8	93.2	94.2
PRSV-P-TH	6.2	4.3	4.3	7.0	6.2		94.2	94.5	94.2	96.2	95.2	96.6
PRSV-P-HA	5.5	3.6	4.0	3.6	4.3	5.1		94.5	96.6	95.5	95.2	95.9
PRSV-P-YK	5.4	3.6	3.2	6.6	6.2	5.4	5.1		93.8	96.9	95.9	97.3
PRSV-P-MEX	6.2	4.3	4.3	4.0	4.0	5.8	2.1	5.8		94.9	94.2	95.2
PRSV-P-DF	4.3	2.5	2.1	5.4	5.1	3.6	4.0	3.2	4.7		97.6	99.7
PRSV-P-SMN	4.7	2.9	2.8	6.2	5.8	4.7	4.3	4.3	5.5	2.5		97.9
PRSV-P-SM	4.0	2.1	1.8	5.1	4.7	3.2	3.6	2.8	4.3	0.3	2.1	

A phylogenetic tree constructed from the multiple alignment using the amino acid sequences of the CP coding region of the twelve PRSV isolates. The phylogenetic tree (Figure 43B) shows a branching pattern of relationship between the twelve isolates of the fully-sequenced PRSV species. Based on the branching pattern of the tree, the twelve isolates of PRSV species aligned could be divided into two clusters. PRSV-P-TH isolate from Thailand did not group with other isolates and it was separated by phylogenic branching as a cluster from all isolates. Seven Asian isolates of PRSV both from Taiwan and Thailand (with the exception of the PRSV-P-TH isolate) grouped with the four non-Asian isolates to form a phylogentic cluster. In this cluster, there are two sub-clusters separating between group of Asian isolates and group of non-Asian isolates. According to geographic locations in the sub-cluster of Asian isolates, the two Thai isolates grouped together, separating the two isolates of cucurbit-infecting PRSV (PRSV-W and PRSV-W-ThRb) from the five Taiwanese isolates including PRSV-P-YK, PRSV-W-TW, PRSV-P-SMN, PRSV-P-SM and PRSV-P-DF. In the sub-cluster of non-Asian PRSV isolates, there are two groups separating the two Brazilian isolates (PRSV-W-C and PRSV-W-1) grouped together from another group consisting of PRSV-P-HA (from USA) and PRSV-P-MEX (from Mexico). As the result obtained from multiple alignment of CP coding region amino acid sequences (Table 33) and the branching pattern of the phylogenetic tree (Figure 43B), the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and it has the highest identity (98.3%) and the lowest distance (1.8%) with PRSV-W-ThRb isolate. The differences between the putative CP amino acid sequence of the PRSV-W isolate and those of the five Taiwanese isolates placing within the same sub-cluster was 4.0 to 5.4%, while it differed by 5.5 to 6.6% when compared with those of the four non-Asian isolates placing within another sub-cluster (in the same cluster).

7.1.12 Phylogenetic analyses based on the complete genomic nucleotide sequences

Throughout the whole genome excluding the poly(A) tail, size comparisons between nucleotide sequences of the different genomes of the available complete PRSV sequences showed that the genome sizes of these PRSV isolates

range from 10320 to 10334 nucleotides (Table 11). The PRSV-W genome analyzing in this research (10323 nucleotides) is the fourth largest genome of all completely sequenced PRSV isolates. It is 11 nucleotides smaller than that of PRSV-W-TW (AY027810) isolate (10334 nucleotides) and 9 nucleotides smaller than that of PRSV-W-1 (DQ374153) isolate (10332 nucleotides) and 3 nucleotides smaller than those of PRSV-W-C (DQ374152) isolate, PRSV-P-HA (X67673) isolate, PRSV-P-YK (X97251) isolate, PRSV-P-DF (DQ340769) isolate, PRSV-P-SMN (DQ340770) isolate and PRSV-P-SM (DQ340771) isolate which their complete genome consist of 10326 nucleotides. The complete RNA genome of the PRSV-W was identical in size (10323 nucleotides) with the two other PRSV isolates from the same country (Thailand) including PRSV-W-ThRb (AY010722) and PRSV-P-TH (AY162218). However, the genome size of the PRSV-W was considerably larger than that of PRSV-P-MEX (10320 nucleotides). Multiple alignment of these nucleotide sequences (Table 34) showed that their completely sequenced genome identities and differences varied from 82.7 to 98.9 % and 1.1 to 20.3%, respectively. The PRSV-W has the similar ranges of identity (82.7 to 98.9 % identical nucleotides) and of difference (1.1 to 20.3% different nucleotides) when compared with all other fully-sequenced PRSV isolates. A phylogenetic tree was prepared from an alignment of the complete nucleotide sequences of PRSV genomes, including all the fully-sequenced isolates in the species. The phylogenetic tree based on multiple alignment of the complete nucleotide sequences (Figure 44A) shows a degree of relationship between the different isolates of the PRSV species. From the result of the phylogenetic analysis of complete nucleotide sequences, these isolates of PRSV species aligned could be divided into two clusters. The eight Asian isolates both Taiwan and Thailand formed their own phylogenetic branch, while four non-Asian isolates clustered together, with two branches separating the two Brazilian isolates from the another branch of two isolates from Mexico and USA. In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates. As the result obtained from multiple alignment of complete nucleotide sequences, the PRSV-W isolate has the highest identity (98.9%) and the lowest distance (1.1%) with PRSV-W-ThRb isolate.

Table 34 Percent identities (top right) and differences (bottom left) between the complete genomic nucleotide sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.9	88.3	83.1	82.7	88.0	83.2	88.6	83.0	88.0	88.0	88.2
PRSV-W-ThRb	1.1		88.7	83.4	83.0	88.4	83.5	89.0	83.3	88.4	88.5	88.6
PRSV-W-TW	13.1	12.6		83.5	83.2	89.6	83.4	95.2	83.2	95.4	95.4	95.5
PRSV-W-C	19.7	19.3	19.2		94.7	83.1	91.8	83.3	91.0	83.3	83.0	83.2
PRSV-W-1	20.3	19.9	19.6	5.6		82.7	91.3	83.2	90.8	83.0	82.8	82.9
PRSV-P-TH	13.5	12.9	11.5	19.7	20.3		83.1	89.8	83.0	89.4	89.4	89.4
PRSV-P-HA	19.6	19.2	19.3	8.9	9.4	19.7		83.4	93.7	82.9	82.9	82.9
PRSV-P-YK	12.8	12.2	5.0	19.4	19.6	11.2	19.3		83.3	95.2	95.2	95.2
PRSV-P-MEX	19.9	19.5	19.5	9.8	10.1	19.8	6.7	19.4		82.9	83.0	83.0
PRSV-P-DF	13.5	12.9	4.7	19.5	19.8	11.7	20.0	5.0	19.9		96.7	97.3
PRSV-P-SMN	13.4	12.9	4.7	19.9	20.1	11.8	20.0	5.0	19.9	3.4		97.0
PRSV-P-SM	13.2	12.6	4.7	19.5	20.1	11.7	20.0	5.0	19.9	2.8	3.1	

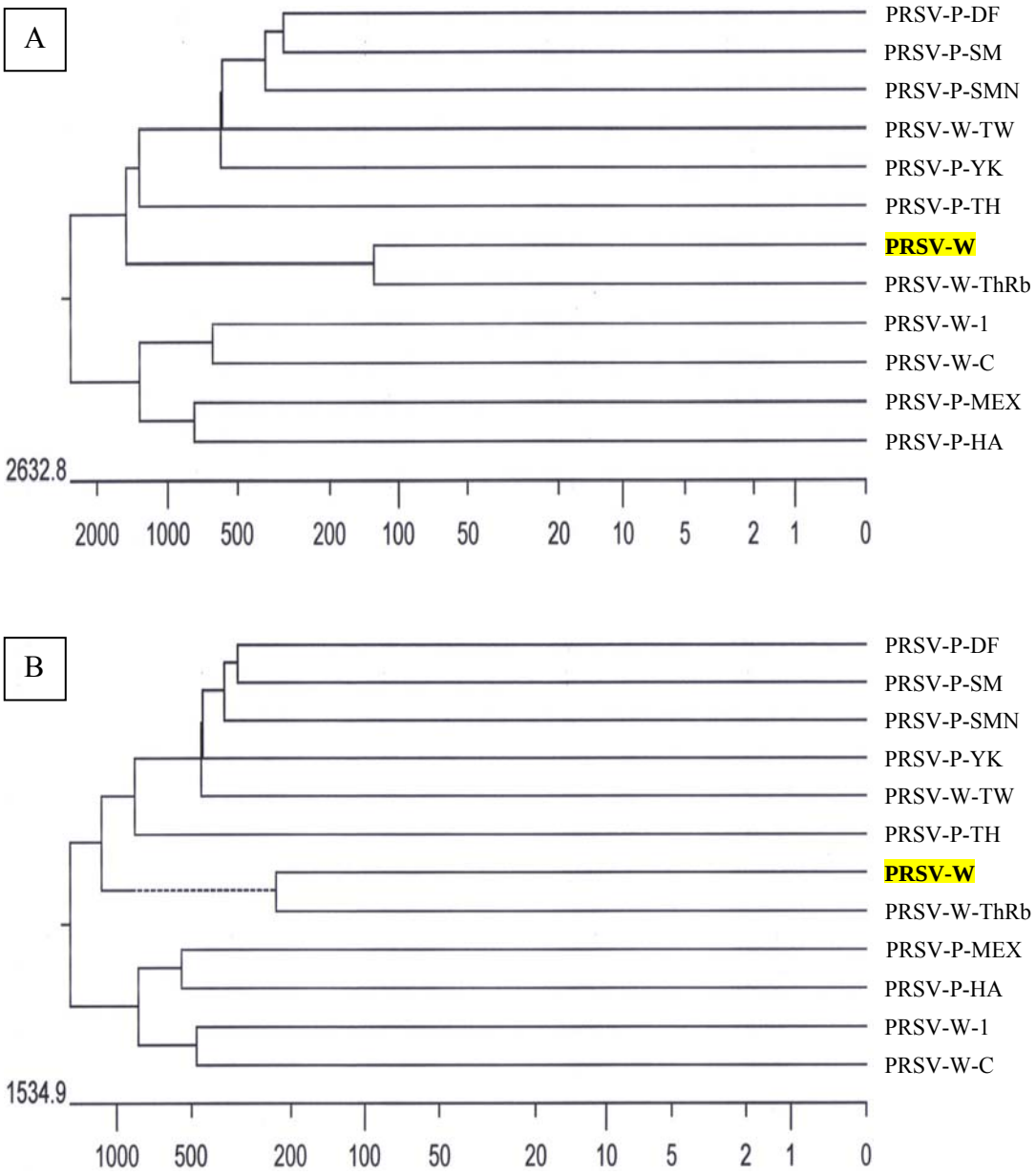


Figure 44 Phylogenetic trees constructed from the multiple alignments of the complete genomic nucleotide sequences (A) and the putative complete polyprotein amino acid sequences (B) of twelve PRSV isolates (including PRSV-W) for which the complete sequence is available. Vertical distances are arbitrary, and horizontal distances reflect the number of nucleotide (for (A)) and amino acid (for (B)) differences between branch nodes, with a scale bar shown at the bottom of each dendrogram. Detailed description of PRSV isolates used in these comparisons (the acronyms present in the dendrograms) is listed in Table 11. PRSV-W isolate analyzing of full-length sequence in this research is indicated in shaded box.

7.1.13 Phylogenetic analyses based on the putative polyprotein amino acid sequences

Like other potyviruses, PRSV genome has a large single open reading frame (ORF) generating an approximately 380-kDa polyprotein. One possible ORF resulted from sequence analysis of PRSV-W isolate was found. The ORF region of the PRSV-W consists of 10032 nucleotides, locating at nucleotide positions 86 to 10117 on the viral genome, which is predicted to encode a putative polyprotein of 3343 amino acids with a calculated molecular weight of 380.3 kDa. Sizes comparisons between amino acid sequences of the putative polyproteins of the available complete PRSV sequences showed that the sizes of different PRSV polyproteins ranged from 3342 to 3346 amino acids (Table 11). The PRSV-W polyprotein resulted from computer analyzing in this research (3343 amino acids) is 3 amino acids smaller than those of PRSV-W-TW (AY027810) and PRSV-W-1 (DQ374153) isolates which their complete polyproteins consist of 3346 amino acids, and one amino acid smaller than those of PRSV-W-C (DQ374152), PRSV-P-HA (X67673), PRSV-P-YK (X97251), PRSV-P-DF (DQ340769), PRSV-P-SMN (DQ340770) and PRSV-P-SM (DQ340771) isolates which their complete polyproteins consist of 3344 amino acids. The complete polyprotein of the PRSV-W (3343 amino acids) was identical in size with the two other PRSV isolates from the same country (Thailand) including PRSV-W-ThRb (AY010722) and PRSV-P-TH (AY162218), while the size of the PRSV-W polyprotein was larger than that of PRSV-P-MEX (3342 amino acids). A multiple alignment of the complete polyprotein amino acid sequences of the twelve isolates (Table 35) showed that their putative polyprotein identities and differences ranged from 90.5 to 98.3% and 1.7 to 10.2%, respectively. The PRSV-W polyprotein has the similar ranges of identity (90.8 to 98.3% identical amino acids) and of difference (1.7 to 9.9% different amino acids) when compared with those of all other fully-sequenced PRSV isolates. Among these PRSV isolates, the overall identity of the PRSV-W polyprotein amino acid sequence is highest with PRSV-W-ThRb (98.3%), followed by PRSV-W-TW (93.8%), PRSV-P-TH (93.7%), PRSV-P-SMN (93.5%), PRSV-P-YK (93.4%), PRSV-P-DF (93.1%), PRSV-P-SM (93.1%), PRSV-W-C (91.4%), PRSV-W-1 (91.3%), PRSV-P-HA

Table 35 Percent identities (top right) and differences (bottom left) between the putative complete polyprotein amino acid sequences in twelve PRSV isolates (including PRSV-W) for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.3	93.8	91.4	91.3	93.7	91.2	93.4	90.8	93.1	93.5	93.1
PRSV-W-ThRb	1.7		94.5	91.9	91.8	94.3	91.7	94.1	91.4	93.8	94.0	93.7
PRSV-W-TW	6.5	5.7		91.3	91.3	93.9	91.0	96.7	90.9	96.5	96.7	96.5
PRSV-W-C	9.2	8.6	9.2		96.4	91.0	94.6	90.9	94.5	90.7	90.7	90.8
PRSV-W-1	9.2	8.7	9.2	3.6		91.1	94.3	90.9	94.3	90.6	90.7	90.7
PRSV-P-TH	6.6	6.0	6.4	9.6	9.5		90.9	93.5	91.0	93.5	93.8	93.4
PRSV-P-HA	9.4	8.8	9.6	5.6	5.9	9.7		90.9	95.3	90.5	90.7	90.6
PRSV-P-YK	6.9	6.2	3.3	9.7	9.8	6.8	9.7		90.8	96.6	96.9	96.4
PRSV-P-MEX	9.9	9.1	9.7	5.7	6.0	9.6	4.9	9.8		90.7	90.7	90.6
PRSV-P-DF	7.2	6.5	3.6	10.0	10.1	6.8	10.2	3.5	10.0		97.2	97.7
PRSV-P-SMN	6.8	6.3	3.3	9.9	9.9	6.5	9.9	3.1	9.9	2.8		97.3
PRSV-P-SM	7.3	6.6	3.6	9.8	10.0	6.9	10.0	3.7	10.0	2.3	2.8	

(91.2%) and PRSV-P-MEX (90.8%). A phylogenetic tree was prepared from the alignment of the complete polyprotein amino acid sequences of PRSV isolates, including all the fully-sequenced isolates in the species. The phylogenetic tree (Figure 44B) shows a degree of relationship between the different isolates of the PRSV species. From the phylogenetic trees (Figure 44), the clustering obtained based solely on the complete polyprotein amino acid sequences of the twelve PRSV isolates is very similar to the branching pattern of relationship in the phylogenetic dendrogram obtained based on the complete genome nucleotide sequences of the same group of PRSV isolates which were aligned and compared in this research. Similarly, these isolates of PRSV species aligned could be divided into two clusters consisting of cluster of Asian isolates and non-Asian isolates. The eight Asian isolates both Taiwan and Thailand formed their own phylogenetic branch, while four non-Asian isolates clustered together, with two branches separating the two Brazilian isolates from the another branch of two isolates from Mexico and USA. In the cluster of the Asian isolates, the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate (1.7% different amino acids) and the phylogenetic branch formed a sub-cluster separating from six other Asian isolates. The differences between the complete polyprotein amino acid sequence of the PRSV-W isolate and those of the Asian isolates within the same cluster (with the exception of the PRSV-W-ThRb isolate) was 6.5 to 7.3%, while it differed by 9.2 to 9.9% when compared with those of the non-Asian isolates placing within the another cluster in the same dendrogram.

Phylogenetic relationships of the PRSV-W isolate to other fully-sequenced isolates of the species PRSV was assessed by comparing the sequences both in nucleotide and amino acid levels. The degrees of relationships between the twelve isolates of the fully-sequenced PRSV are separately illustrated in the 22 phylogenetic trees (section 7.1) generated from multiple alignments based on both nucleotide and amino acid sequences of different genomic regions including two NCRs (5'-NCR and 3'-NCR) and ten putative mature proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP).

Nucleotide sequence comparisons of NCRs indicate that the 3'-NCR, with 88.4-100.0% sequence identity (Table 13) is more conservative than the 5'-NCR, with 61.6-98.8% sequence identity (Table 12) among the twelve fully-sequenced PRSV isolates. Sequence comparisons of predicted mature proteins reveal that the most conserved protein region among the twelve fully-sequenced PRSV isolates is the cylindrical inclusion (CI) protein, with 96.7-99.7% sequence identity (Table 23). The region with the lowest identity among the PRSV isolates studied is the first protein (P1), with identity ranging from 65.4 to 97.1%. All of the phylogenetic trees illustrate that the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate except the tree generated from multiple amino acid sequence alignment of 6K1 peptide (Figure 37B) which illustrated the closest relationship of the PRSV-W isolate with PRSV-P-HA.

In these 22 phylogenetic trees, branching patterns of 15 phylogenetic trees illustrated a place in each tree of the PRSV-W isolate which was clustered with the Asian isolates including PRSV isolates from Thailand and Taiwan. Four trees of these, the Asian isolates were grouped together without PRSV-P-TH. There are two phylogenetic trees constructed separately from 5'-NCR nucleotide sequences (Figure 33A) and from P1 amino acid sequences (Figure 34B) which showed a phylogenetic grouping of all three Thai isolates as a subcluster separating from all five Taiwanese isolates which were placed in the same cluster. The Asian cluster (in each tree) including three Thai isolates and five Taiwanese isolates separated from another cluster of non-Asian isolates. The both trees illustrated a phylogenetic relationship depending on the geographic location of the isolates.

While the PRSV-W isolate was grouped to form a phylogenetic cluster with non Asian isolates in 5 trees including 4 trees constructing based on nucleotide sequences from 6K1 (Figure 37A), 6K2 (Figure 39A), VPg (Figure 40A) and CP (Figure 43A), and a tree constructing based on amino acid sequences from NIb (Figure 42B). In the tree constructed from CP nucleotide sequences, all three isolates from Thailand grouped together with a further phylogenic branch separately from all four non-Asian isolates which were grouped together. Both groups of the Thai and

non-Asian isolates placed together within the same cluster separating from another cluster which all five Taiwanese isolates grouped together.

Two trees constructing based on amino acid sequences from 6K1 (Figure 37B) and 6K2 (Figure 39B) illustrated separately phylogenetic branching patterns which the cluster patterns are not correlated with geographic locations of the virus isolates. While the three Thai isolates were grouped together to form a cluster separating from other isolates in the tree (Figure 38B) constructing based on amino acid sequences from CI.

Phylogenetic branching pattern of more than two clusters was observed and found in two phylogenetic trees such as the trees constructing based on amino acid sequences of separated HC-Pro (Figure 35B) and 6K2 (Figure 39B) which illustrated for three and five clusters, respectively. No phylogenetic tree grouping depend on "type (P or W)" was found in all of these 22 trees, except the tree constructed from 6K1 amino acid sequences (Figure 37B) which it illustrated a phylogenetic relationship of partially separated clusters depending on the types of the PRSV isolates.

Interestingly, the phylogenetic tree constructed from CP amino acid sequences (Figure 43B) illustrated a branch pattern which the PRSV-P-TH isolate has a position separating from other eleven isolates grouping as a cluster. This seemed to be as an outgroup isolate illustrating in the tree, and other eleven isolates grouped together to form a cluster with further phylogenetic branchs depending on their originally geographic locations.

By multiple alignment and phylogenetic analysis, the complete genomic nucleotide and the putative polyprotein amino acid sequences among twelve fully-sequenced isolates of PRSV were also applied to phylogenetically analyze by comparing in the same condition as performing with the others. The percent identities of the complete genomic nucleotide sequences and amino acid sequences of the

putative polyproteins among the twelve PRSV isolates are shown as a homology matrix in Table 36.

Both phylogenetic trees constructed separately from multiple alignments of the complete genomic nucleotide sequences (Figure 44A) and the putative overall polyprotein amino acid sequences (Figure 44B) of twelve fully-sequenced PRSV isolates (including PRSV-W isolate) are similar in branching pattern and the both trees illustrate that the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate.

In the both phylogenetic trees, the PRSV-W isolate grouping together with PRSV-W-ThRb isolate has a position in the cluster of all Asian isolates with a further phylogenetic branch to form a sub-cluster separating from other Asian isolates including five Taiwanese isolates and PRSV-P-TH isolate from Thailand. The Asian isolates (including PRSV-W isolate) grouped together to form phylogenetic cluster separating from four non-Asian isolates clustered together, with two branches separating the two Brazilian isolates from the another branch of two isolates from Mexico and USA. The both trees illustrated a phylogenetic relationship depending on the geographic location of the isolates.

The separate twelve nucleotide sequences and ten predicted amino acid sequences from different genomic regions of the PRSV-W isolate were separately compared with those nucleotide and amino acid sequences, respectively, corresponding regions of the eleven fully-sequenced PRSV isolates. Percentage nucleotide and amino acid identities of different genomic regions between the PRSV-W isolate and the eleven other PRSV isolates are shown in Table 37. The degree of identities both nucleotide and amino acid to the different genomic regions of the PRSV-W isolate varied depending on the originally geographic locations of the isolates and the genomic regions.

Table 36 Homology matrix of the percent identities of the complete genomic nucleotide sequences (bottom left) and deduced amino acids sequences in the polyproteins (top right) of twelve isolates of PRSV species for which the complete sequence is available

	PRSV-W	PRSV-W-ThRb	PRSV-W-TW	PRSV-W-C	PRSV-W-1	PRSV-P-TH	PRSV-P-HA	PRSV-P-YK	PRSV-P-MEX	PRSV-P-DF	PRSV-P-SMN	PRSV-P-SM
PRSV-W		98.3	93.8	91.4	91.3	93.7	91.2	93.4	90.8	93.1	93.5	93.1
PRSV-W-ThRb	98.9		94.5	91.9	91.8	94.3	91.7	94.1	91.4	93.8	94.0	93.7
PRSV-W-TW	88.3	88.7		91.3	91.3	93.9	91.0	96.7	90.9	96.5	96.7	96.5
PRSV-W-C	83.1	83.4	83.5		96.4	91.0	94.6	90.9	94.5	90.7	90.7	90.8
PRSV-W-1	82.7	83.0	83.2	94.7		91.0	94.3	90.9	94.3	90.6	90.7	90.7
PRSV-P-TH	88.0	88.4	89.6	83.1	82.7		90.9	93.5	91.0	93.5	93.8	93.4
PRSV-P-HA	83.2	83.5	83.4	91.8	91.3	83.1		90.9	95.3	90.5	90.7	90.6
PRSV-P-YK	88.6	89.0	95.2	83.3	83.2	89.8	83.4		90.8	96.6	96.9	96.4
PRSV-P-MEX	83.0	83.3	83.2	91.0	90.8	83.0	93.7	83.3		90.7	90.7	90.6
PRSV-P-DF	88.0	88.4	95.4	83.3	83.0	89.4	82.9	95.2	82.9		97.2	97.7
PRSV-P-SMN	88.0	88.5	95.4	83.0	82.8	89.4	82.9	95.2	83.0	96.7		97.3
PRSV-P-SM	88.2	88.6	95.5	83.2	82.9	89.4	82.9	95.2	83.0	97.3	97.0	

Sequence comparisons of the separate twelve nucleotide sequences and ten predicted amino acid sequences from different genomic regions between the PRSV-W isolate and those of corresponding regions of the eleven fully-sequenced PRSV isolates reveal that the PRSV-W isolate has the highest nucleotide and amino acid identities in all of different genomic region sequences with PRSV-W-ThRb isolate, which the percentages of nucleotide and amino acid sequence identities range from 98.4 to 100.0% and from 97.1 to 100.0%, respectively (Table 37).

In non coding regions (NCRs), the highest percentage 5'-NCR and 3'-NCR nucleotide identities of the PRSV-W isolate are 98.8% and 100.0%, respectively, with the corresponding regions of PRSV-W-ThRb isolate. The 5'-NCR and 3'-NCR of the PRSV-W isolate has the nucleotide sequence identities varied from 66.3 to 89.5% and 93.2 to 98.1%, respectively, when compared with the corresponding region of other fully-sequenced PRSV isolates except those of PRSV-W-ThRb. These indicate that the PRSV 5'-NCR is highly variable while the 3'-NCR of PRSV is a conserved region among the PRSV species.

In ten coding regions of predicted mature proteins, both nucleotide and amino acid sequences of the different genomic regions within polyprotein of the PRSV-W isolate were compared separately with those of the corresponding regions of the other fully-sequenced PRSV isolates.

At nucleotide level, sequence comparisons show that the highest percentage identities of P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP coding regions of the PRSV-W isolate are 98.4%, 99.0%, 98.5%, 98.7%, 99.1%, 100.0%, 98.6%, 99.2%, 99.1% and 98.8%, respectively, with the corresponding regions of PRSV-W-ThRb isolate. In all of ten predicted mature proteins, nucleotide sequences of the PRSV-W isolate were compared separately with those of the corresponding regions of the other fully-sequenced PRSV isolates except those of PRSV-W-ThRb. The sequence comparisons with the exception of the PRSV-W-ThRb isolate show that the percentages of nucleotide identities of predicted mature proteins of the PRSV-W isolate range from 69.5% with that of PRSV-W-1 P1 to 93.0% with those of the CPs

Table 37 Percentage nucleotide (upper) and amino acid (lower) identities between different genomic regions of PRSV-W and the corresponding regions of eleven other isolates of PRSV for which the complete sequence is available

PRSV isolate	Percent (%) identity of nucleotide (upper) and amino acid (lower) sequence between PRSV-W and other isolates of PRSV											
	5' NCR	Mature cleavage proteins of PRSV isolates										3' NCR
		P1	HC-Pro	P3	6K1	CI	6K2	VPg	NIa-Pro	NIb	CP	
PRSV-W-ThRb	98.8	98.4	99.0	98.5	98.7	99.1	100.0	98.6	99.2	99.1	98.8	100.0
(AY010722)		97.1	98.5	97.7	98.1	99.1	100.0	98.9	98.3	98.5	98.3	
PRSV-W-TW	89.5	80.8	89.8	89.0	90.4	89.0	89.0	87.0	88.7	89.4	93.0	98.1
(AY027810)		80.6	96.5	95.1	98.1	97.8	93.1	96.8	96.2	95.5	94.9	
PRSV-W-C	66.3	70.7	86.5	86.3	87.9	86.4	90.1	83.5	82.5	81.8	90.2	95.2
(DQ374152)		67.2	96.9	93.9	94.3	98.1	98.3	96.3	95.4	94.8	93.5	
PRSV-W-1	68.6	69.5	85.6	85.8	87.9	85.8	88.4	84.5	82.4	82.1	89.7	95.2
(DQ374153)		66.7	96.9	94.2	96.2	97.5	98.3	96.3	95.8	95.5	92.8	
PRSV-P-TH	89.5	81.9	89.9	88.2	88.5	88.8	89.5	86.6	87.1	88.3	92.2	97.1
(AY162218)		82.2	97.8	93.6	94.3	97.6	96.6	95.8	95.4	94.8	94.2	
PRSV-P-HA	69.8	70.6	85.2	85.3	88.5	86.6	90.7	82.6	82.7	83.3	90.9	93.7
(X67673)		67.0	96.5	92.2	96.2	98.1	94.8	94.7	93.7	95.9	93.8	
PRSV-P-YK	86.0	81.6	89.3	89.8	89.8	89.3	88.4	89.4	88.0	89.8	92.8	98.1
(X97251)		80.7	96.7	95.1	92.5	97.3	93.1	96.3	95.4	95.4	94.5	
PRSV-P-MEX	69.8	70.3	85.6	85.9	86.6	86.2	88.4	84.2	81.7	83.1	89.3	93.2
(AY231130)		67.6	95.9	92.5	92.5	97.6	94.8	94.7	92.9	95.2	93.8	
PRSV-P-DF	86.0	80.9	89.2	88.8	89.8	89.0	89.5	87.9	87.0	89.0	92.7	94.7
(DQ340769)		79.1	96.3	94.5	94.3	97.5	93.1	96.8	94.6	94.8	95.5	
PRSV-P-SMN	87.2	81.1	89.0	88.6	91.7	89.7	88.4	87.5	86.9	88.6	92.8	95.2
(DQ340770)		80.6	96.5	94.2	94.3	97.8	93.1	95.8	95.0	95.5	95.2	
PRSV-P-SM	89.5	81.4	89.4	88.5	91.1	89.0	88.4	87.5	86.7	89.8	93.0	95.2
(DQ340771)		79.3	95.9	93.9	94.3	97.5	93.1	95.8	94.1	95.5	95.9	

of PRSV-W-TW and PRSV-P-SM isolates. The P1 coding region of the PRSV-W isolate has the nucleotide sequence identities varied from 69.5 with that of PRSV-W-1 to 81.9% with that of PRSV-P-TH when compared with the corresponding region of other fully-sequenced PRSV isolates except that of PRSV-W-ThRb. With the exception of the P1 coding region in comparisons, the nucleotide identities of nine predicted genomic coding regions the PRSV-W isolate are higher than 80% identical when compared with the corresponding regions of other fully-sequenced PRSV isolates. These indicate that nucleotide sequence in the P1 coding region is highly variable while the remainders are the conserved regions among the PRSV species.

At amino acid level, sequence comparisons show that the highest percentage identities of P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP coding regions of the PRSV-W isolate are 97.1%, 98.5%, 97.7%, 98.1%, 99.1%, 100.0%, 98.9%, 98.3%, 98.5% and 98.3%, respectively, with the corresponding regions of PRSV-W-ThRb isolate. Ten amino acid sequences of different predicted mature proteins of the PRSV-W isolate were compared separately with those of the corresponding regions of the other fully-sequenced PRSV isolates except those of PRSV-W-ThRb. The sequence comparisons show that the percentages of amino acid identities of predicted mature proteins of the PRSV-W isolate range from 66.7% with that of PRSV-W-1 P1 to 98.3% with that of PRSV-W-C CP. The putative P1 of the PRSV-W isolate has the amino acid sequence identities varied from 66.7 with that of PRSV-W-1 to 82.2% with that of PRSV-P-TH when compared with the corresponding region of other fully-sequenced PRSV isolates except that of PRSV-W-ThRb. With the exception of the P1 coding region in comparisons, the amino acid identities of nine predicted mature proteins of the PRSV-W isolate are higher than 90% identical when compared with the corresponding proteins of other fully-sequenced PRSV isolates. Similar with the sequence comparisons at the nucleotide level, the amino acid sequence comparisons indicate that P1 amino acid sequence is highly variable while the remainders are the conserved regions among the fully-sequenced PRSV isolates.

The genome and putative polyprotein of the PRSV-W isolate have the ranges of 82.7 to 88.6% identical nucleotides and 90.8 to 93.8% identical amino acids,

respectively, when compared with those of other fully-sequenced PRSV isolates (with the exception of PRSV-W-ThRb isolate). This indicates that the majority of nucleotide changes are silent. The PRSV-W isolate has the highest percentage nucleotide (98.9%) and amino acid (98.3%) identities of the complete genome and putative polyprotein, respectively, with those of PRSV-W-ThRb isolate when compared with those of other fully-sequenced PRSV isolates. Both PRSV-W and PRSV-W-ThRb isolates are classified into the cucurbit-infecting (W) type. They were isolated from the same country (Thailand) but they are different in geographic location of origin. Geographic origin of the PRSV-W isolate is located in Nakhon Sawan province, while the PRSV-W-ThRb is an isolate originating in Ratchaburi.

7.2 Comparison of the deduced amino acid sequences between PRSV-W and PRSV-W-ThRb polyproteins.

Amino acid sequence comparison between polyproteins of PRSV-W and PRSV-W-ThRb isolates was performed and the sequences compared are shown in Figure 45. By sequence alignment, the amino acid identity between the two isolates varies from 97.1 to 100.0%. However, the CI and 6K2 proteins are distinct from the other proteins because they have an identity level higher than 99%.

Fifty seven residue changes are found in the comparison of the whole polyprotein (Figure 45). As shown among other fully-sequenced PRSV isolates, the P1 is the most variable region of polyprotein between the two isolates of PRSV-W and PRSV-W-ThRb isolates with 97.1% amino acid sequence identity. The largest number of non-conservative amino acid differences occurs in the P1 protein. These differences include sixteen residue changes which represent 28.1% of the overall changes in the polyprotein, indicating that the P1 protein is more variable than the other proteins of this virus. There is no change of amino acid residue in 6K2 peptide (100.0% identity). The numbers of amino acid differences occur in the remains including HC-Pro, P3, 6K1, CI, VPg, NIa-Pro, NIb and CP, are 7, 8, 1, 6, 2, 4, 8 and 5 residue changes, respectively.

```
MSSLYQLQPIALKDRLLSHKRGKGWIEHKLERKGERGNTRHVGEFVISEG 50

AKILQLVQIGNAEIGRTFLEGDRRTQADIFEIIKKTMVGHLGYDFECGLW 100

CCYSCDNTSDKYFKKCDCGEKYYYSEKNLIKSMHDLMYQFDMTAAEIDQV 150
  H                       R

GLDYLAEAVDYAEQSVKKSKVLVPDEPKFVEILAASEESLLVVPEPEVAS 200
                      P     E   V       H        V

VTTRAEEAWTIQIGEIPVPLVVIRETPVISGVNGMLSSTGFSLEAEIARP 250
                                                  K

AKSTIPQDEVEEAVHLALEVGNEIAEKKPELKLAPYWSASLELHKRVRKH 300
    V     K

KEHAKSEALRVRKEKERDHRIFAALEAKLNLKARRQGQVVVCDKRGTLKW 350
            R                                     I

KKHQQRKRNKMVTQLSDSVVTKIHANFECRTPNFDVETPGIKCATSKVTR 400
  R         A                       L

KKQTQPKIFGSNKVNYVMKNLCDIIIDRNIPVELITKRCKRRIFRRDGKN 450
                                                M

YVHLRHMDGNNAPRDVSSSSDMEKLFTRFCKFLIRKQSINAANLTHGSSG 500

                                                P1/HC-Pro
LIFKPKFADRTGRYFGEYFITRGRCEGKLFDGRSKLAKSVRMRMEQYNDV 550

AEKFWLGFNRAFLRHRKPTDHVCTSDMDVTMCGEVAALATIILFPCHKIT 600
```

Figure 45 Comparison of the deduced amino acid sequences of PRSV-W and PRSV-W-ThRb. The complete sequence of the PRSV-W polyprotein is presented. Amino acids of PRSV-W-ThRb polyprotein that differ from PRSV-W are shown below the PRSV-W sequence. The predicted cleavage sites (pair of amino acid residues in each site) are indicated in shade boxes, and those of mature protein junction signatures described in the text are shown above their sites. The regions around the putative cleavage sites (from P7 to P'3) are underlined. The amino acid positions are indicated at the end of each line.

```
CNACMNKVKGRVIDEVGEDLNCELERLRETLSSYGGSFGHVSTLLDQLNR 650
  T                                        R

VLNARNMNDGAFKEIAKKIDEKKESPWIHMTVINNTLIKGSLATGYEFER 700
                                  A

ASDSLREIVRWHLKRTESIKAGSVESFRNKRSGKAHFNPALTCDNQLDRN 750
       V

GNFLWGERQYHAKRFFANYFEKIDHSKGYEYYSQRQNPNGIRKIAISNLI 800
                                                  G

FSTNLERFRQQMVEHHIDQGPITRECIALRNNNYVHVCSCVTLDDGTPAT 850


SELKTPTKNHIVLGNSGDPKYVDLPTLESDSMYIAKRGYCYMNIFLAMLI 900


NIPENEAKDFTKRVRDLVGSKLGEWPTMLDVATCANQLVIFHPDAANAEL 950


PRILVDHRQKTMHVIDSFGPVDSGYHILKANTVNQLIQFAREPLDSEMKH 1000
                    S        N
HC-Pro/P3
YIVGGEFDPTTSCLHQLIRVIYKPHELRNLLRNEPYLIVIALMSPSVLLT 1050
                                        V

LFNSGAIEHALNYWIKRDQDVVEVIVLVEQLCRKVTLARTILEQFNEIRQ 1100
                                        I

NARDIHELMDRNNKPWISYDRSLELLSVYADSQLTDEGLLKQGFSTLDPR 1150
                                  N

LREAVEKTYAALLQEEWRALSLFQKLHLRYFAFKSQPSFSECLKPKGRAD 1200
                                          Y

LKIVNDFSPKYCVHEVGKALLQPIKAGAEITSRLLNGCGTFIRKSAARGC 1250
    Y

AYIFKDLFQFVHVVLVLSILLQIFRSAQGIAMEHIQLKQAKAEAVKQKDF 1300
                                          E    Y
                                             P3/6K1
DRLEALYAELCVKNGDQPTTEEFLDFVMEREPRLKDQAYSLIHIPAIHQA 1350
                                            V
```

Figure 45 (continued)

```
KSDNEKKLGQVIAFITLILMMIDVDKSDCVYRILNKFKGVINSCNTNVYH 1400
        E
6K1/CI
QSLDDIRDFYEDKQLTIDFDITGENQINRGPIDVTFEKWWDNQLSNNNTI 1450

GHYRIGGTFVEFSRSNAATVASEIAHSPEREFLVRGAVGSGKSTNLPFLL 1500

SKHGSVLLIEPTRPLCENVCKQLRGEPFHCNQTIRMRGLTAFGSTNITIM 1550
                                    P
TSGFALHYYAHNLQQLRLFDFIIFDECHVIDSQAMAFYCLMEGNAIEKKV 1600

LKVSATPPGREVEFSTQFPTKIVTERSISFKQLVDNFGTGANSDVTAFAD 1650
                            Q
NILVYVASYNEVDQLSKLLSDKGYLVTKIDGRTMKAGKTEISTSGTKSKK 1700
                                      V
HFIVATNIIENGVTLDIEAVIDFGMKVVPEMDSDNRMIRYSKQAISFGER 1750

IQRLGRVGRHKEGIALRIGHTEKGIQEIPEMAATEAAFLSFTYGLPVMTH 1800

NVGLSLLKNCTVRQARTMQQYELSPFFTQNLVNFDGTVHPKIDVLLRPYK 1850

LRDCEVRLSEAAIPHGVQSIWMSAREYEAVGGRLCLEGDVRIPFLIKDVP 1900

ERLYKELWDIVQTYKRDFTFGRINSVSAGKIAYTLRTDVYSIPRTLITID 1950

KLIESENMEHAHFKAMTSCTGLNSSFSLLGVINTIQGRYLVDHSVENIRK 2000
        K                               S
                                    CI/6K2
LQLAKAQIQQLEAHVQENNVENLIQSLGAVRAVYHQSVDGVKHIKRELGL 2050
                                  G
                                                6K2/VPg
KGVWDGSLMIKDAIVCGFTMAGGAMLLYQHFRDKLTNIHVFHQGFSARQR 2100
```

Figure 45 (continued)

```
QKLRFKSAANAKLGREVYGDDGTIEHYFGEAYTKKGNKKGRMHGMGVKTR 2150

KFVATYGFKPEDYSYVRYLDPLTGETLDESPQTDISMVQEHFGDIRSKYM 2200
                                                  N

DSDSFDRQALIANNVIKAYYVRNSAKTALEVDLTPHNPLKVCDTKLTIAG 2250
G
                                    VPg/NIa-Pro
FPDREAELRQTGPPRTILADQVPPPTKSVHHEGKSLCQGMRNYNGIASVV 2300

CHLKSTSGDGRSLFGVGYNSFIITNRHLFKENNGELIVKSQHGKFVVKNT 2350
    N

STLRIAPIGKTDLLIIRMPKGFPPFHSRARFRAMKAGDKVCMIGVDYQEN 2400
                     D   S       T

HIASKVSETSIISEGTGEFGCHWISTNDGDCGNPLVSVSDGFIVGLHSLS 2450

TSTGDQNFFAKIPAFFEEKVLRRIDDLTWSKHWSYNVNELSWGALKVWES 2500

              NIa-Pro/NIb
RPEAIFNAQKEIDQLNVFEQSGSRWLFDKLHGNLKGVSSAPSNLVTKHVV 2550

KGICPLFRSYLECDEEAKAFFSPLMGHYMKSVLSKEAYVKDLLKYSSDII 2600
  V

VGEVDHDVFEDSVAQVIELLNDHECPELEYVTDSEVIIQALNMDAAVGAL 2650

YTGKKRKYFEGSTVEHRQALVRKSCERLYEGRMGVWNGSLKAELRPAEKV 2700

LAKKTRSFTAAPLDTLLGAKVCVDDFNNWFYSKNMECPWTVGMTKFYKGW 2750

DEFLRKFPDSWVYCDADGSQFDSSLTPYLLNAVLSIRLWAMEDWDIGAQM 2800
         G    G

LKNLYGEITYTPILTPDGTIVKKFKGNNSGQPSTVVDNTLMVLITMYYAL 2850
```

Figure 45 (continued)

```
RKAGYDTEAQEDMCVFYINGDDLCIAIHPDHEHVLDSFSSSFAELGLKDD 2900
       K                                          Y

FTQRHRNKQDLWFMSHRGILIDDIYIPKLEPERIVAILEWDKSKLPEHRL 2950

EAITAAMIESWGYGELTHQIRRFYQWVLEQAPFDELAEQGRAPYVSEVGL 3000
                                    N   K

RRLYTSERGSMDELEAYIDKYFERERGDSPELLVYHESRSTDDYQLICGE 3050
                                                   V
    NIb/CP
STHVFHQSKNEAVDAGLNEKFKEKEKQKEEKDKQKDKNNDGASDGNDVST 3100

STKTGERDRDVNAGTSGTFTVPRMKSFTNKMILPRIKGKTVLNLNHLLQY 3150
                        I    D

NPQQIDISNTRATQSQFEKWYEGVRNDYGLNGNEMQVMLNGLMVWCIENG 3200
                                   D

TSPDISGVWVMVDGETQVDYPIRPLIEHATPSFRQIMAHFSNAAEAYIAK 3250
          M          K

RNATERYMPRYGIKRNLTDISLARYAFDFYEVNSKTPDRAREAHMQMKAA 3300

ALRNTGRRMFGMDGSVSNKEENTERHTVEDVNKDMHSLLGMRN        3343
```

Figure 45 (continued)

8. Construction of full-length cDNA clones of PRSV-W

As shown in Figure 30, the complete nucleotide sequence of the Thai isolate of *Papaya ringspot virus* type W (PRSV-W) which was used as source for all experiments in this research contains 10323 nucleotides in length excluding the poly(A) tail. The complete nucleotide sequence of the virus was determined from eight overlapping viral cDNA clones which were generated from RT-PCR with degenerate and/or specific primers (Table 2). Initially, all oligonucleotide primers used in these RT-PCRs were designed separately in their DNA sequence based on the alignment of two nucleotide sequences published in the GenBank representing the genome of two different PRSV Hawaiian (HA) (GenBank accession number X67673) and Taiwanese YK (GenBank accession number X97251) isolates of the strain PRSV-P, consensus regions were identified. Then the different RT-PCR products amplified were cloned separately in U-tailed pDrive cloning vector (QIAGEN) prior to their insert sequences were analyzed as described in section 3.1. Based on the complete nucleotide sequence of the PRSV-W and the eight nucleotide sequences of different inserts of the eight overlapping viral cDNA clones, a set of virus-sequence specific internal oligonucleotide primers was designed and synthesized to use for amplify the desired viral DNA fragments which contain overlapping DNA regions at their 5'- and 3'-terminal regions with the adjacent cDNA fragments. Table 38 lists all oligonucleotide primers used in this study.

Eight overlapping viral DNA fragments were amplified separately from the eight overlapping viral cDNA clones constructed in this research as described in section 3.1 and 3.2 (named pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH in the order in which inserted fragments correspond on the virus genome as shown in Figure 1) using a pair of virus-sequence specific internal primers complementary to each clones (Table 38). The eight cDNA fragments with blunt ends amplified separately from the eight overlapping viral cDNA clones as DNA templates in high fidelity PCRs were used as viral genomic cDNA source for constructing full-length clones of the PRSV-W.

Table 38 Primers used for constructing full-length cDNA clones of PRSV-W

Primer	Sequence (5'-3')[a]	Plasmid: Position (F/R)[b]
Virus-sequence specific internal primers		
PC60	AACTCAAGGCTAGCACTC	pPCwA: 946-963 (R)
PC61	TCGCCGAAAAGAAGCCTG	pPCwB: 906-923 (F)
PC62	TTTTCTGTCGATGATCCACTAAG	pPCwB: 2943-2966 (R)
PC63	CCAACTGGTTATCTTTCATCCT	pPCwC: 2893-2904 (F)
PC64	TAGAAATTCGCGTTCAGGAC	pPCwC: 4515-4534 (R)
PC65	AGTGAAATAGCCCACAGTCCT	pPCwD: 4499-4519 (F)
PC66	CAGACTCCTTTTAAACCCAACT	pPCwD: 6225-6246 (R)
PC67	AGTCAAACACATAAAGCGAGAGT	pPCwE: 6205-6227 (F)
PC68	CCTTACTAGCGATATGGTTTTCTT	pPCwE: 7278-7301 (R)
PC69	AGGTTTGTATGATTGGTGTAGATT	pPCwF: 7251-7274 (F)
PC70	TAACTTGTGCGACGCTATCC	pPCwF: 7915-7934 (R)
PC71	GGCGAAGTTGACCACGAT	pPCwG: 7889-7906 (F)
PC74	GAAAC$(T)_{24}$CTCTCATTCT	pPCwH: 10313-10352 (R)
PC75	CAGCTTATTTGCGGTGAGAG	pPCwH: 9218-9237 (F)
PC76	CATTTTTGGACTGGTGGAACA	pPCwG: 9246-9266 (R)
PC77	CATTTGGAGAGGAAATAAAACATCTCAACAC	pPCwA: 1-19 (F)
PC78	*AAGCTT*CCCGATCTAGTAACATAGATGACACC	pDNosTer: NOS (R)
PC79	*AAGCTT*CATGGAGTCAAAGATTCAAATAGAGG	pD35SPro: CaMV 35S (F)
Vector-sequence specific primers		
PC80	GTAAAACGACGGCCAGT	pDrive and pTZ57R vectors
PC81	AACAGCTATGACCATG	pDrive and pTZ57R vectors

[a] Overhang nucleotide sequences designed for generating overlapping DNA regions with the adjacent DNA fragment are underlined. Restriction sites shown in *italic* are not present in the native CaMV 35S promoter (PC79) and NOS terminator (PC78) nucleotide sequences.

[b] Number represents nucleotide position in the complete nucleotide sequence of the PRSV-W which was used as source for all experiments in this research. (F) = forward primer, (R) = reverse primer.

For *in vitro* amplification of DNA in this study, three different applications of PCRs including 'high fidelity PCR', 'long PCR' and 'overlap extension PCR' were performed to amplify desired cDNA fragments with a limit size of 3 – 4 kilobase pairs, with a size more than 4 kilobase pairs and cDNA fragments which were generated from assembling of two or more linear cDNA fragments, respectively. All three PCR methods are based on *DeepVent* DNA polymerase purchased from New England Biolabs Inc., which has a 3'→5' proofreading activity and the obtained PCR products from these reaction have both blunted ends. Plasmids pDrive (QIAGEN) and pTZ57R (Fermentas) were used as vectors. PCR products generated with *Deep Vent* proofreading DNA polymerase (or DNA polymerase mix containing *Deep Vent*) do not contain a single adenosine (A) nucleotide overhang at the 3'- terminus (blunt ends) and therefore are not compatible for direct ligation into either pDrive (QIAGEN) or pTZ57R (Fermentas) cloning vector, which contains a single 3'-terminus uracil (U) or thymidine (T) nucleotide overhang at both ends, respectively. Prior to cloning blunted end fragment in either U- or T-tailed plasmid cloning vectors, a single A-tailing reaction was performed to add adenosine nucleotides to the PCR fragments at their 3'- termini, which can hybridize to ligate the fragments into the available plasmid cloning vectors.

Cloning procedure was carried out essentially as mentioned in materials and methods. Three different strains of *Escherichia coli* including DH5α, JM109 and XL1-Blue were prepared the competent cells and used as host cells for propagation of plasmids and cDNA clones. Transformation manners of recombinant DNA plasmids were performed by using either an electroporation or a calcium chloride (heat shock) method depending on the expected size of the recombinant plasmid vector. The calcium chloride (heat shock) method was carried out with a small insert DNA fragment (smaller than 3 kilobase pairs), while the electroporation method was usually performed with a 3-kilobase pairs larger insert DNA fragment. Transformants of bacterial cells which contain these recombinant plasmids were selected simultaneously in two ways, blue/white colony screening and antibiotic resistance selection on an appropriated culture medium agar plate (2xYT or LB media) containing one or two appropriated antibiotic(s), X-gal and IPTG. Colony selection

was performed after incubating the culture plates spreaded with bacterial cell suspension at 37^{o}C for 14-18 hr. After colony selection, analysis of recombinant plasmids was performed to verify the presence and orientation of the insert fragment. The selected transformants were analyzed to confirm the presence of the insert fragment by PCR and plasmid DNA miniprep/restriction digestion. Restriction enzymes *Bam*HI, *Eco*RI, *Nco*I *Ngo*MIV *Sal*I and *Xba*I used in this study were purchased and used as specified by the manufacturers (Table 5). 'Routine PCR' was performed to verify an orientation of an insert fragment in an obtained recombinant plasmid using vector-specific primers and insert specific primers. Recombinant plasmid DNAs were prepared and isolated from small-scale (1-3 ml) bacterial cultures by treatment with alkaline and SDS.

In this study, two full-length cDNA clones of PRSV-W, named pDPCwPN7978 and pDPCwAN0178, were constructed separately by cloning in a commercially available high copy number plasmid U-tailed pDrive (QIAGEN). All prerequisite clones for constructing the full-length cDNA clones of PRSV-W were separately cloned in either U-tailed pDrive (QIAGEN) or T-tailed pTZ57R (Fermentas) cloning vectors. The recombinant plasmid pDPCwPN7978 designed and constructed in pDrive cloning vector (QIAGEN) was made from assembly of the full-length cDNA of PRSV-W with *Cauliflower mosaic virus* (CaMV) 35S promoter and nopaline synthase (NOS) terminator and was used directly for infectivity assay as described in section 8.4. Another recombinant plasmid, pDPCwAN0178, is identical to the plasmid pDPCwPN7978 except that it has no CaMV 35S promoter and its entire fragment which contains the full-length cDNA of PRSV-W and NOS terminator has been inserted in the opposite orientation when compared with the pDPCwPN7978 recombinant plasmid. The recombinant plasmid pDPCwAN0178 (without CaMV 35S promoter) was designed and constructed for using as viral cDNA template in an *in vitro* transcription reaction for synthesizing infectious RNA transcripts of PRSV-W used in infectivity assay as described in section 8.5. Both two full-length cDNA clones of PRSV-W, pDPCwPN7978 and pDPCwAN0178, containing with and without CaMV 35S promoter, respectively, were constructed as follows.

8.1 Construction of pDPCwPD7940 and pDPCwAD0140 plasmids

A recombinant plasmid, named pDPCwPD7940, was constructed by inserting a 'wPD7940' cDNA fragment of 5663 base pairs into U-tailed pDrive cloning vector (QIAGEN). The 'wPD7940' cDNA fragment was developed from overlap extension PCR of two overlapping cDNA fragments consisting 'wPB8062' of 3628 base pair and 'wCD6366' of 3354 base pairs which were generated separately from pDPCwPB7962 and pDPCwCD6366 recombinant plasmids, respectively. In the pDPCwPD7940 plasmid, insert cDNA fragment 'wPD7940' (5663 base pairs) contains CaMV 35S promoter and 5131-base pairs cDNA fragment (wAD0140) corresponding to 5'-NCR, P1, HC-Pro, P3 and 5'-terminal region of CI genes of the PRSV-W isolate studied in this research. Plasmid pDPCwAD0140 was constructed by cloning of cDNA fragment 'wAD0140' (5131 base pairs) into U-tailed pDrive cloning vector (QIAGEN). The 'wAD0140' cDNA fragment was amplified from the pDPCwPD7940 plasmid using primers PC1 and PC40 in a long extension PCR. The linear cDNA construct 'wAD0140' is identical to the 'wPD7940' cDNA fragment except that it has no cDNA region of CaMV 35S promoter but it contains only 5131-base pairs cDNA fragment (wAD0140) corresponding to 5'-NCR, P1, HC-Pro, P3 and 5'-terminal half of CI genes of the PRSV-W isolate studied in this research. The two constructs, pDPCwPD7940 and pDPCwAD0140 plasmids, were used for further constructions.

Cauliflower mosaic virus (CaMV) 35S promoter used in this research was amplified from p2324 plasmid and cloned into U-tailed pDrive cloning vector (QIAGEN). The source of the CaMV 35S promoter, p2324 plasmid, was kindly donated by Dr. Srimek Chowpongpang, Department of Plant Pathology, Kasetsart University. The CaMV 35S promoter cDNA fragment was amplified from plasmid p2324 using primers 35S-1H (5'-AAG CTT CAT GGA GTC AAA GAT TCA-3') and 5U-35S-2 (5'-AAT GTT TTA TTT CCT CTC CAA ATG AAA TG-3'). The obtained PCR product of approximately 0.5 kilobase pairs in size was cloned into U-tailed pDrive cloning vector (QIAGEN) after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). Sequencing of the CaMV

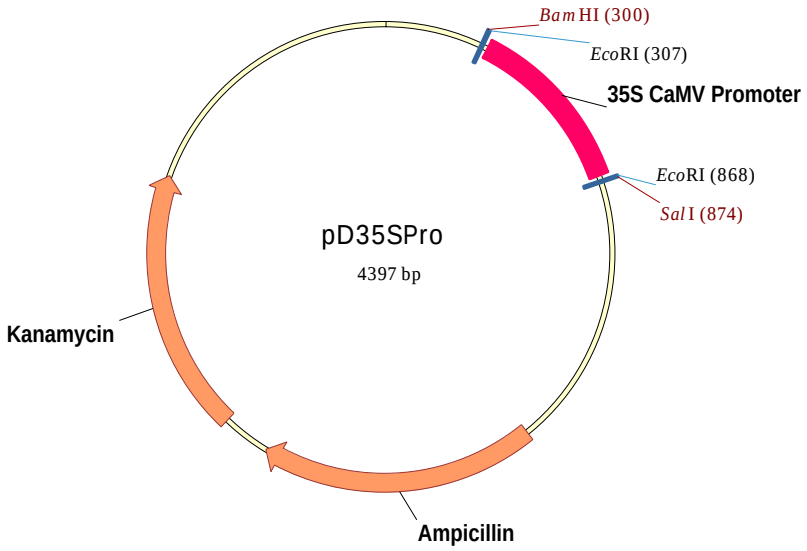
35S promoter cDNA in the obtained recombinant plasmid, named pD35Pro (Figure 46A), was performed using the ABI PRISM® Bigdye™ Terminator Cycle Sequencing Ready Reaction Kits version 2.0 and the reactions obtained were run on either the model 377 automatic DNA sequencer (Applied Biosystems). DNA sequence of the insert fragment is shown in Figure 46B for CaMV 35S promoter of 526 base pairs with additional 6-base pair sequence specific to *Hin*dIII endonuclease at the 5'-terminus. The pD35Pro plasmid used as source of CaMV 35S promoter in this study was used for amplifying fragment '35S CaMV promoter' to fuse to the full-length genomic cDNA construct of the PRSV-W isolate studying in this research.

8.1.1 Construction of pDPCwPB7962 plasmid

Three different recombinant plasmids constructed in this research, pD35SPro (Figure 46), pPCwA (Figure 9) and pPCwB (Figure 10), were used as templates for generating three overlapping cDNA fragments consisting '35S CaMV promoter', 'wA7760' and 'wB6102' fragments, respectively, by high fidelity PCR. All three blunt-ended PCR products were gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as templates in overlap extension PCR for assembly of proposed linear DNA construct.

Initial step for constructing pDPCwPB7962 plasmid, an overlap extension PCR for assembly of two linear DNA fragments between '35S CaMV promoter' and 'wA7760' fragments was performed to generate a linear DNA construct which contains CaMV 35S promoter and cDNA fragment of 5'-terminal region of PRSV-W together with no unwanted sequences at the junction. In construction of the fused-linear DNA 'wPA7960', two primers 5U-35S-2 (reverse) and PC77 (forward) were designed and synthesized to use separately with primer 35S-1H and PC60, respectively, for generating an 24-base pairs overlapping sequences between CaMV 35S promoter and 'wA7760' fragments at 3'- and 5'-terminal regions, respectively. CaMV 35S promoter cDNA fragment of 544 base pairs was amplified from plasmid pD35SPro using primers 35S-1H and 5U-35S-2. Fragment 'wA7760' of 975 base pairs was amplified from plasmid pPCwA using primers PC77 and PC60. Both

A



B

AAGCTTCATG GAGTCAAAGA TTCAAATAGA GGACCTAACA GAACTCGCCG TAAAGACTGG CGAACAGTTC
ATACAGAGTC TCTTACGACT CAATGACAAG AAGAAAATCT TCGTCAACAT GGTGGAGCAC GACACACTTG
TCTACTCCAA AAATATCAAA GATACAGTCT CAGAAGACCA AAGGGCAATT GAGACTTTTC AACAAAGGGT
AATATCCGGA AACCTCCTCG GATTCCATTG CCCAGCTATC TGTCACTTTA TTGTGAAGAT AGTGGAAAAG
GAAGGTGGCT CCTACAAATG CCATCATTGC GATAAAGGAA AGGCCATCGT TGAAGATGCC TCTGCCGACA
GTGGTCCCAA AGATGGACCC CCACCCACGA GGAGCATCGT GGAAAAAGAA GACGTTCCAA CCACGTCTTC
AAAGCAAGTG GATTGATGTG ATATCTCCAC TGACGTAAGG GATGACGCAC AATCCCACTA TCCTTCGCAA
GACCCTTCCT CTATATAAGG AAGTTCATTT CATTTGGAGA GG*AAATAAAA CATC*

Figure 46 Map of pD35SPro (A) and nucleotide sequence of 35S CaMV Promoter cDNA fragment of 526 base pairs with *Hind*III specific sequence of 6 base pairs (underlined) and partial 5'NCR sequence of 12 base pairs (in italic) of the PRSV-W isolate at 5'- and 3'-terminal regions, respectively, of the insert fragment (B). The pD35SPro plasmid of 4397 base pairs in closed circular form was obtained from inserting the 35S CaMV Promoter cDNA fragment of 544 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment '35S CaMV Promoter' is depicted by a shaded box. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and two antibiotic-resistant genes (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

fragments of ‘35S CaMV promoter’ and ‘wA7760’ separately amplified by high fidelity PCR were used together as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). For assembly these fragments together, overlap extension PCR was performed with the two fragments which share the 24-base pairs overlapping cDNA region (‘35S CaMV promoter’ and ‘wA7760’) as templates using primers PC79 and PC60 (Table 38). Amplified fragment obtained from assembly of ‘35S CaMV promoter’ and ‘wA7760’ in overlap extension PCR, named ‘wPA7960’, is 1495 base pairs in size. The blunt-ended fragment ‘wPA7960’ was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailed reaction. Recombinant plasmid containing the ‘wPA7960’ fragment in the pDrive vector, named pDPCwPA7960, is circular DNA of 5348 base pairs in size. Outline of construction of pDPCwPA7960 is shown in Figure 47. Orientation of the insert fragment ‘wPA7960’ in the recombinant plasmid pDPCwPA7960 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including 35S-1H and PC60 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwPA7960 plasmid as substrate by digesting with three restriction endonucleases including *Bam*HI, *Eco*RI and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 48. Proposed circular DNA map of pDPCwPA7960 plasmid constructing as described above is shown in Figure 49. The pDPCwPA7960 plasmid was used as template in a high fidelity PCR for further constructions.

Plasmid pDPCwPB7962 was constructed by inserting a linear cDNA fragment generated from assembly of two DNA fragments ‘wPA8160’ and ‘wB6102’ into U-tailed pDrive cloning vector (QIAGEN). For amplification of ‘wPA8160’ fragment of 1604 base pairs, high fidelity PCR was performed with pDPCwPA7960 plasmid as template using primers PC81 and PC60 for DNA polymerization priming. Fragment ‘wB6102’ of 2201 base pairs was amplified from pPCwB plasmid (Figure 10) using primers PC61 and PC2. The two fragments, ‘wPA8160’ and ‘wB6102’, which share the 58-base pairs overlapping cDNA region at 3’- and 5’-terminal regions,

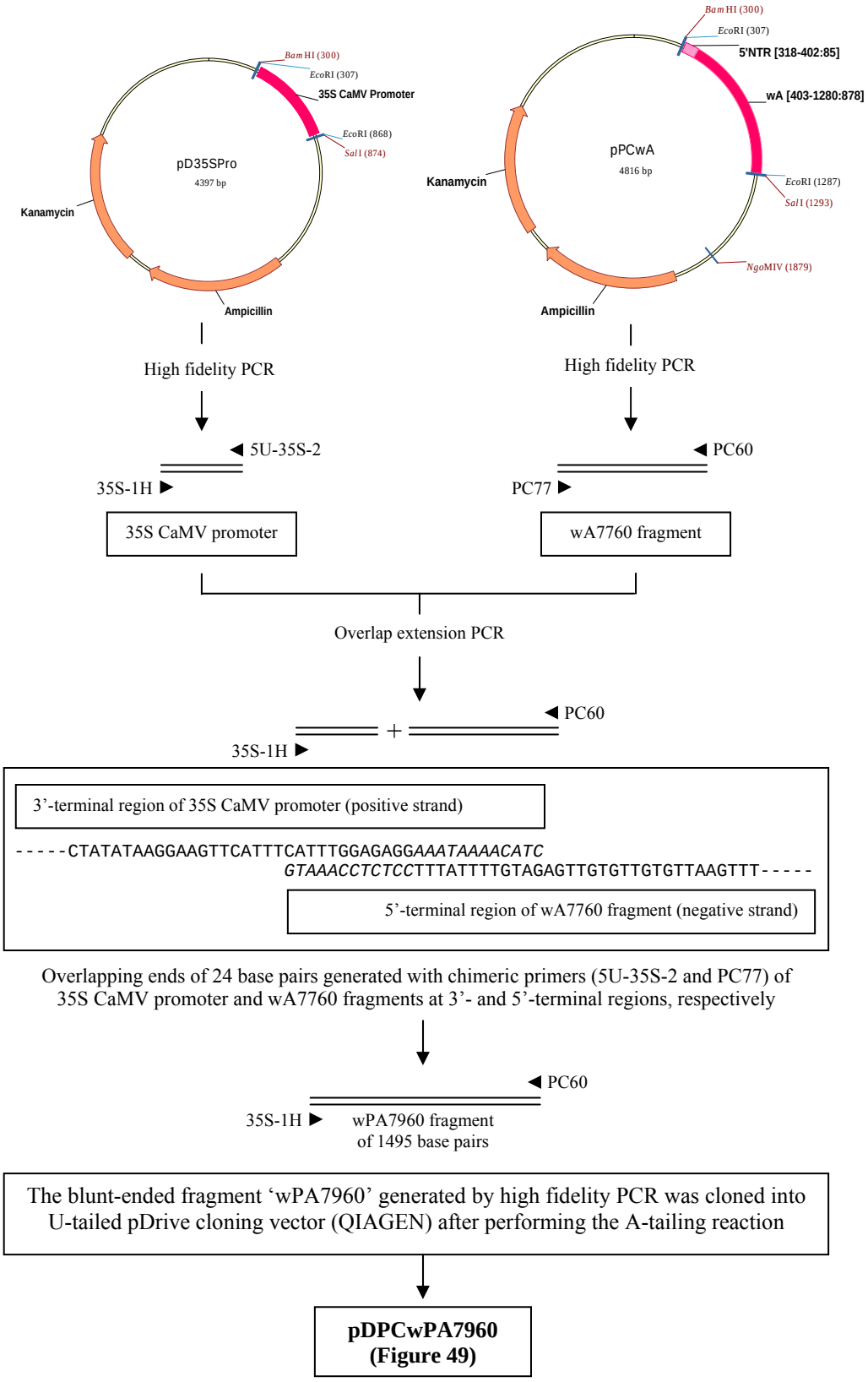


Figure 47 Outline of construction of pDPCwPA7960

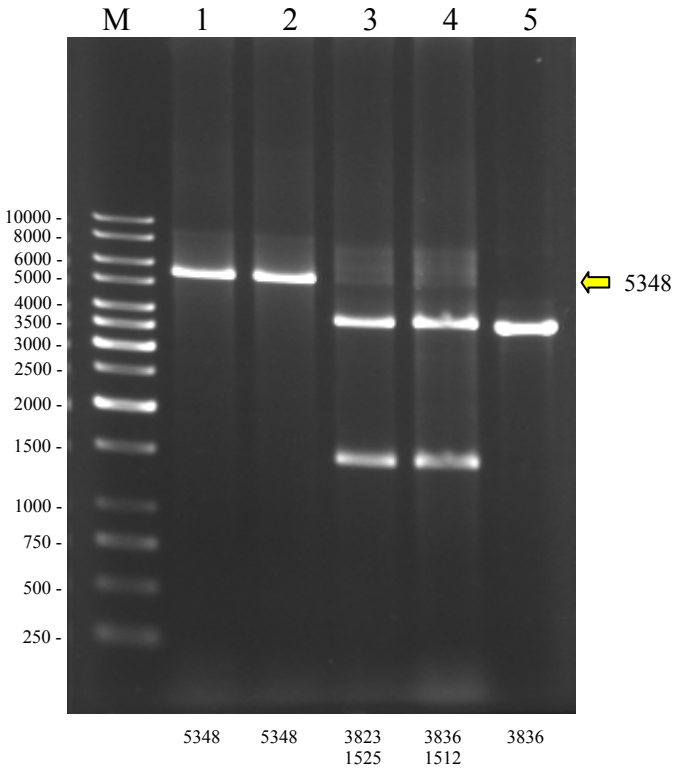


Figure 48 Restriction analyses of the pDPCwPA7960 plasmid. The pDPCwPA7960 plasmid of 5348 base pairs in closed circular form obtained from inserting the DNA fragment 'wPA' of 1495 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwPA plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDPCwPA digested with *Bam*HI
Lane 2: pDPCwPA digested with *Sal*I
Lane 3: pDPCwPA digested with *Bam*HI and *Sal*I
Lane 4: pDPCwPA digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

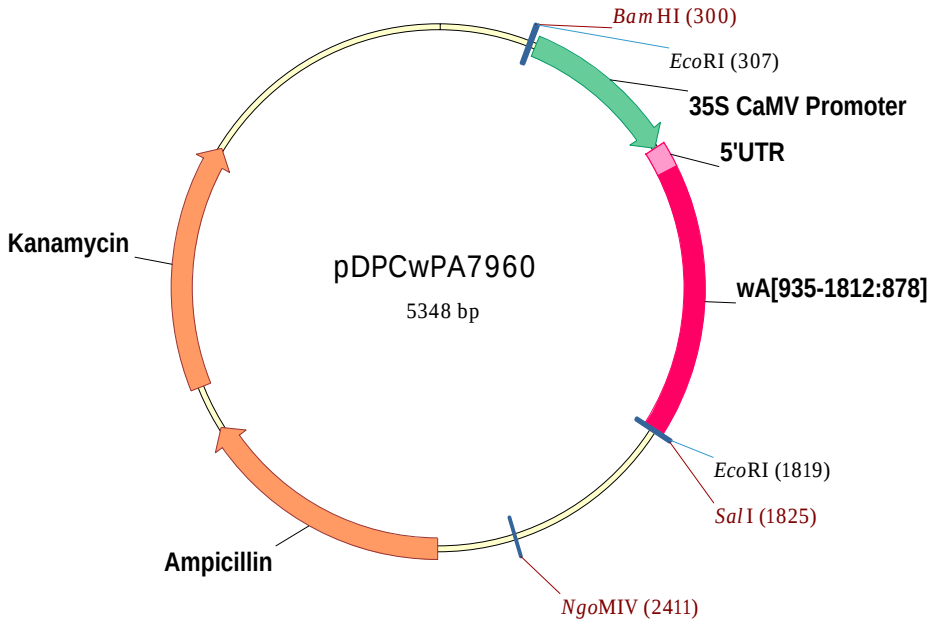


Figure 49 Map of pDPCwPA7960. The pDPCwPA7960 plasmid of 5348 base pairs in closed circular form was obtained from inserting the DNA fragment 'wPA7960' of 1495 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wPA7960' is depicted by a shaded arrowhead for 35S CaMV Promoter region, and a shaded box for fragment 'wA'region. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). To assemble the two linear DNA fragments, overlap extension PCR was performed with 'wPA8160' and 'wB6102' fragments as overlapping DNA templates using primers PC81 and PC2. Approximately 3.7-kilobase pairs DNA fragment, named 'wPB8102', amplified from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wPB7962' fragment. The 'wPB7962' fragment was amplified from 'wPB8102' fragment as template using primers PC79 and PC62 in high fidelity PCR. The blunt-ended 'wPB7962' fragment of 3498 base pairs in size obtained from the high fidelity PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction. Recombinant plasmid containing the 'wPB7962' fragment in the pDrive vector, named pDPCwPB7962, is circular DNA of 7351 base pairs in size. Outline of constructing the recombinant plasmid pDPCwPB7962 is shown in Figure 50. Orientation of the insert fragment 'wPB7962' of 3498 base pairs in the recombinant plasmid pDPCwPB7962 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC60 and PC61 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwPB7962 plasmid as substrate by digesting with three restriction endonucleases including *Bam*HI, *Eco*RI and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 51. Proposed circular DNA map of pDPCwPB7962 plasmid constructing as described above is shown in Figure 52. The pDPCwPB7960 plasmid was used as template in a high fidelity PCR for further constructions.

8.1.2 Construction of pDPCwCD6366 plasmid

Plasmid pDPCwCD6366 was constructed from inserting a linear DNA construct, named 'wCD6366', generated from assembly of two overlapping DNA fragments 'wC0364' and 'wD6512' into U-tailed pDrive cloning vector (QIAGEN). Two different recombinant plasmids constructed in this research, pPCwC

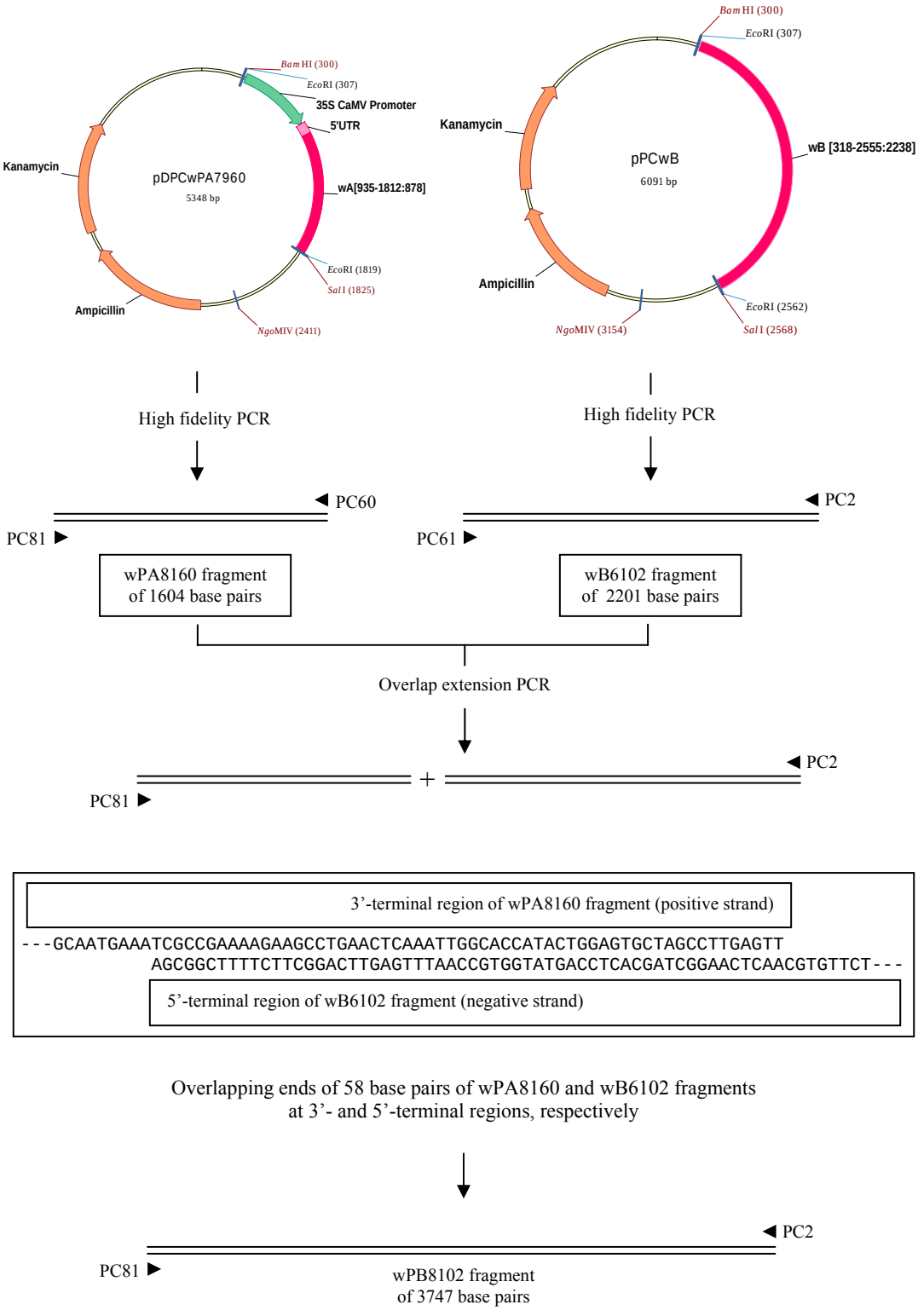
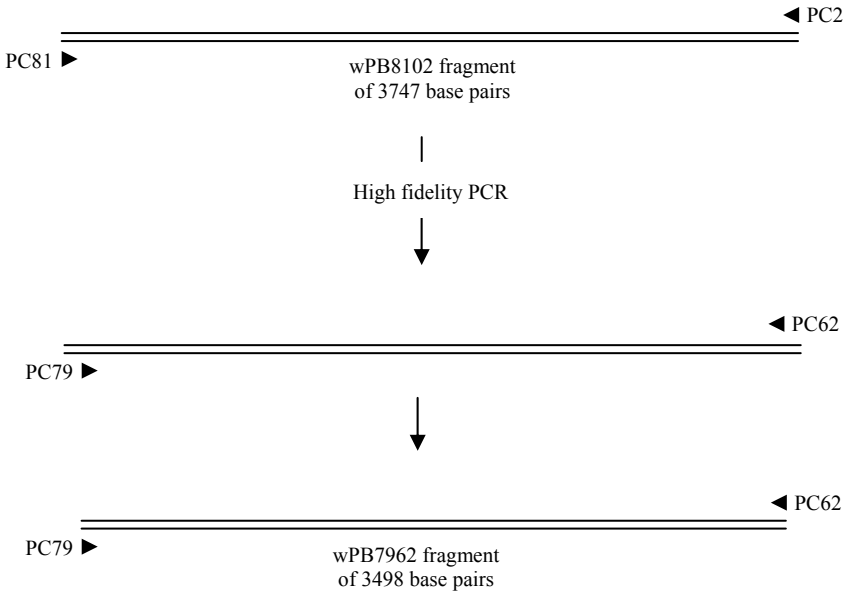


Figure 50 Outline of construction of pDPCwPB7962

The fragment ‘wPB8102’ generated by overlap extension PCR was used as template for amplification of ‘wPB7962’ DNA fragment in high fidelity PCR



The blunt-ended fragment ‘wPB7962’ generated by high fidelity PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction

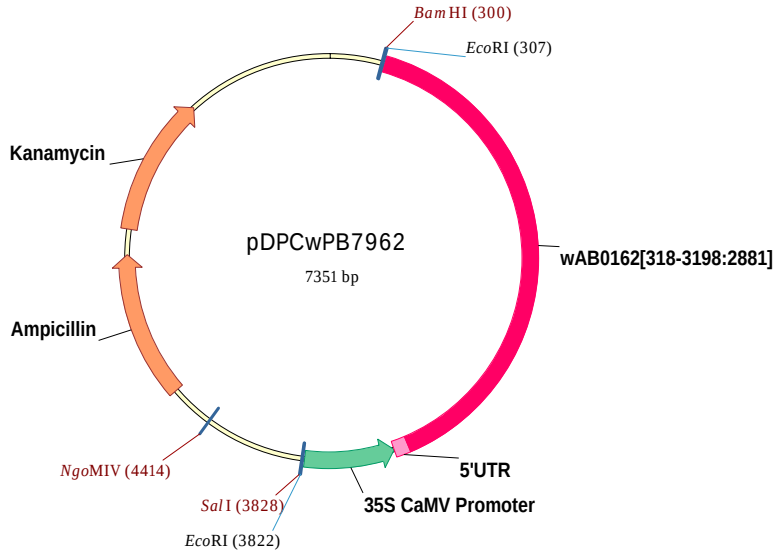


Figure 50 (continued)

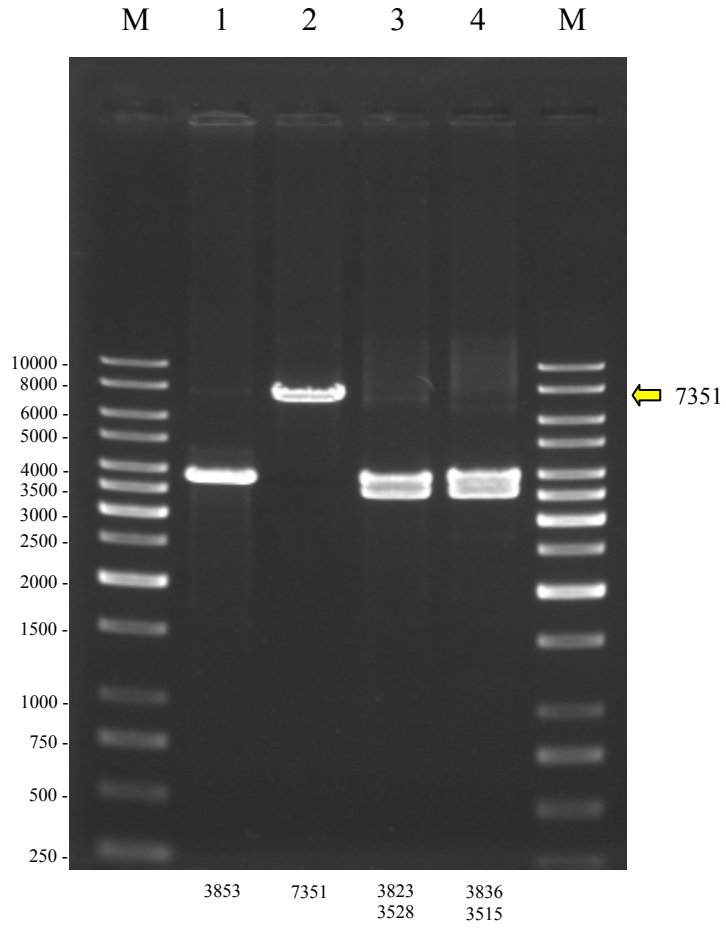


Figure 51 Restriction analyses of the pDPCwPB7962 plasmid. The pDPCwPB7962 plasmid of 7351 base pairs in closed circular form obtained from inserting the DNA fragment 'wPB7962' of 3498 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwPB7962 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive cloning vector self-ligating digested with *Bam*HI
Lane 2: pPCwPB7962 digested with *Bam*HI
Lane 3: pPCwPB7962 digested with *Bam*HI and *Sal*I
Lane 4: pPCwPB7962 digested with *Eco*RI

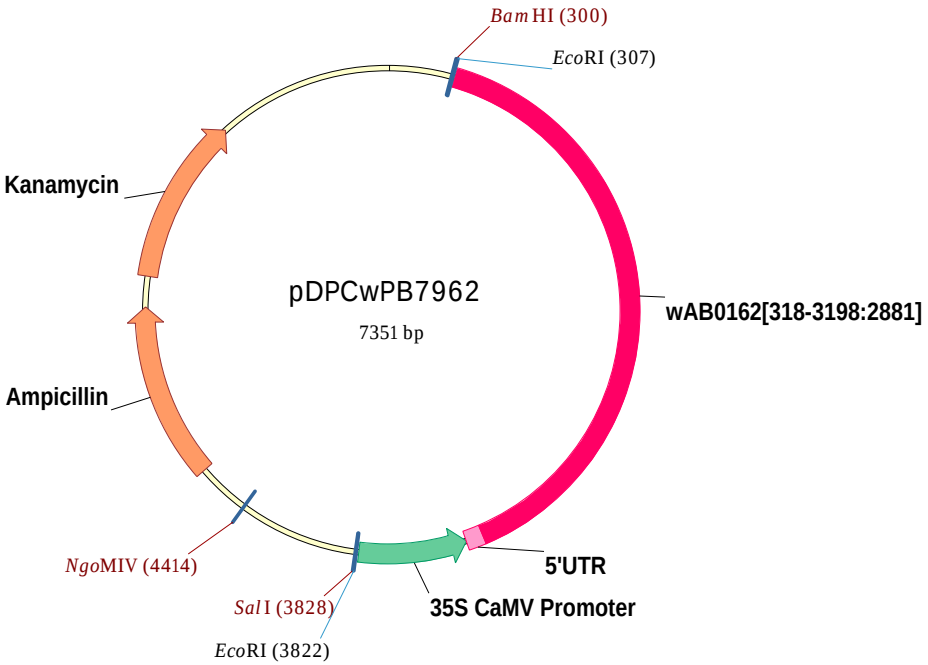


Figure 52 Map of pDPCwPB7962. The pDPCwPB7962 plasmid of 7351 base pairs in closed circular form was obtained from inserting the DNA fragment 'wPA7960' of 3498 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wPB7962' is depicted by a shaded arrowhead for 35S CaMV Promoter region, and a shaded box for fragment 'wAB'region. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

(Figure 11) and pPCwD (Figure 12), were used as templates for generating two overlapping cDNA fragments consisting 'wC0364' and 'wD6512' fragments, respectively, by high fidelity PCR. High fidelity PCR was performed with pPCwC plasmid as template using primers PC3 and PC64 for DNA amplification of 'wC0364' fragment of 1886 base pairs. Fragment 'wD6512' of 1776 base pairs was amplified from pPCwD plasmid using primers PC65 and PC12 in high fidelity PCR. The two fragments, 'wC0364' and 'wD6512', which share a 36-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). For assembly of the two linear DNA fragments 'wC0364' and 'wD6512', overlap extension PCR was performed with primers PC3 and PC12. Overlap extension PCR product of approximately 3.6 kilobase pairs, named 'wCD0312', was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wCD6366' fragment. The linear DNA construct 'wCD6366' of 3354 base pairs was amplified from 'wCD0312' DNA fragment (3629 base pairs in size) as template using primers PC63 and PC66 in high fidelity PCR. The blunt-ended 'wCD6366' fragment obtained from the high fidelity PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction. The selected recombinant plasmid containing the 'wCD6366' fragment in the pDrive vector, named pDPCwCD6366, is circular DNA of 7207 base pairs in size. Outline of constructing the recombinant plasmid pDPCwCD6366 is shown in Figure 53. Orientation of the insert fragment 'wCD6366' of 3354 base pairs in the recombinant plasmid pDPCwCD6366 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC64 and PC65 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwCD6366 plasmid as substrate by digesting with three restriction endonucleases including *Bam*HI, *Nco*I and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 54. Proposed circular DNA map of pDPCwCD6366 plasmid constructing as described above is shown in Figure 55. The pDPCwCD6366 plasmid was used as template in a high fidelity PCR for further constructions.

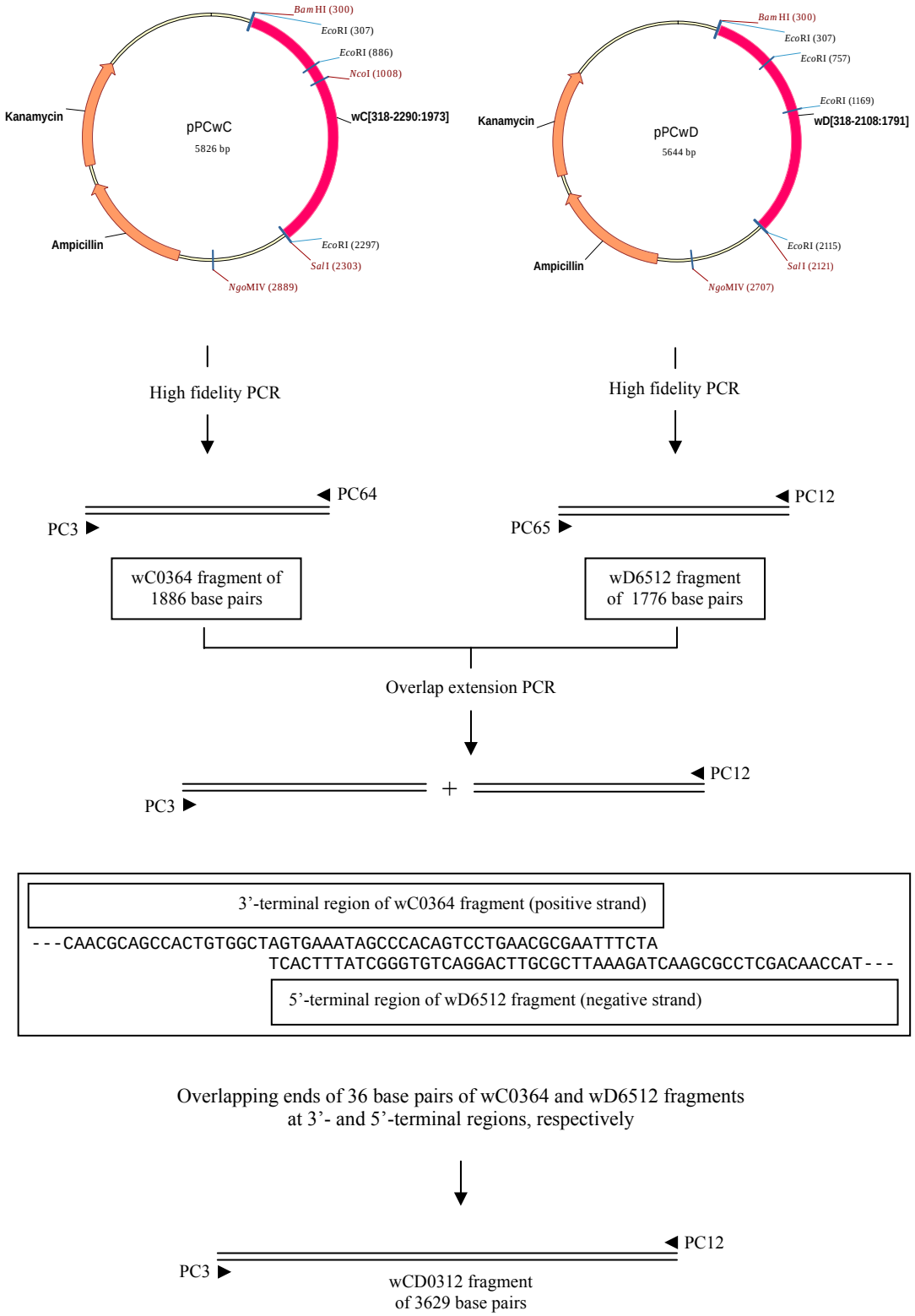
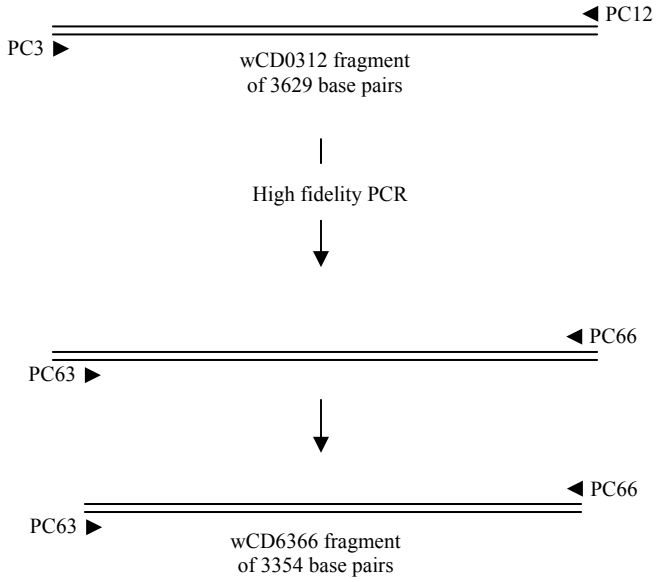


Figure 53 Outline of construction of pDPCwCD6366

The fragment ‘wCD0312’ generated by overlap extension PCR was used as template for amplification of ‘wCD6366’ DNA fragment in high fidelity PCR



The blunt-ended fragment ‘wCD6366’ generated by high fidelity PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction

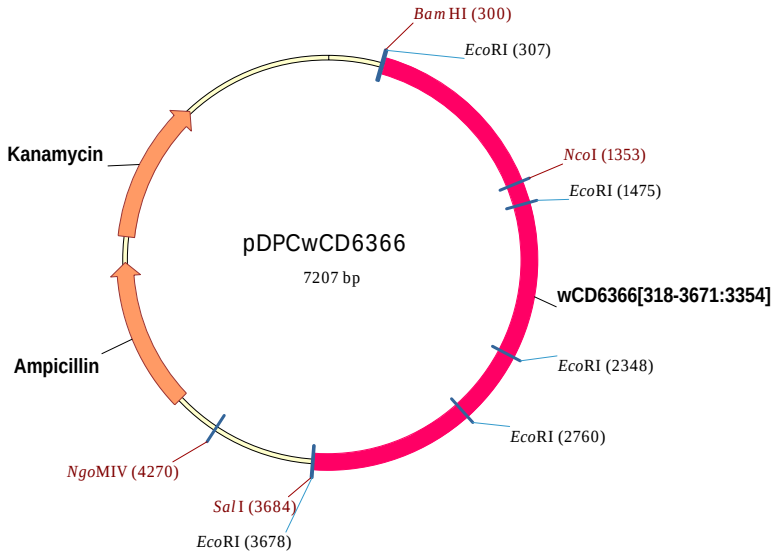


Figure 53 (continued)

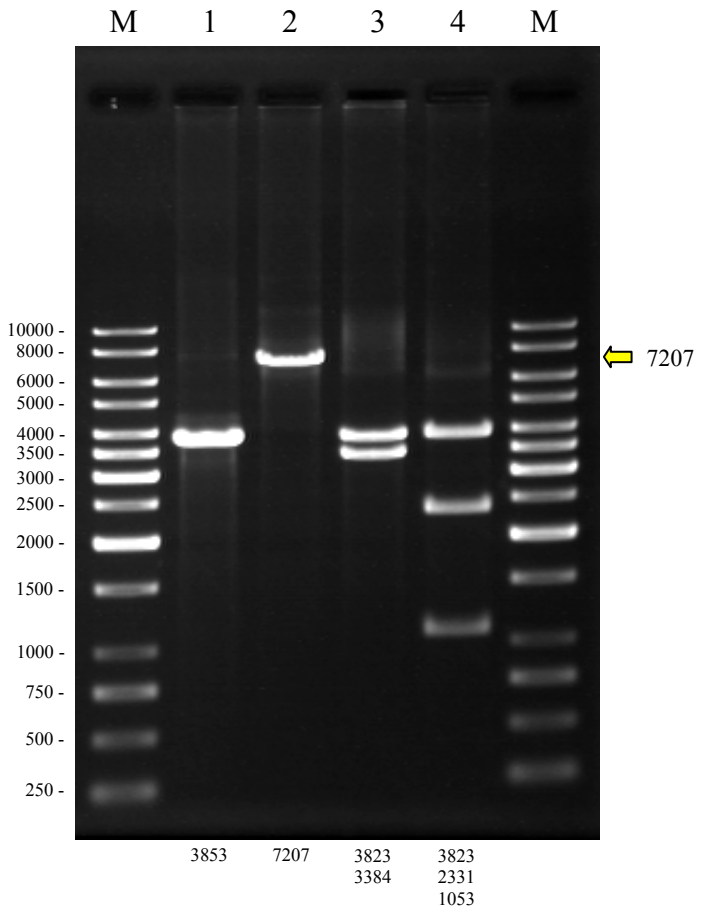


Figure 54 Restriction analyses of the pDPCwCD6366 plasmid. The pDPCwCD6366 plasmid of 7207 base pairs in closed circular form obtained from inserting the DNA fragment 'wCD6366' of 3354 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwCD6366 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive cloning vector self-ligating digested with *Bam*HI
Lane 2: pPCwCD6366 digested with *Bam*HI
Lane 3: pPCwCD6366 digested with *Bam*HI and *Sal*I
Lane 4: pPCwCD6366 digested with *Bam*HI, *Nco*I and *Sal*I

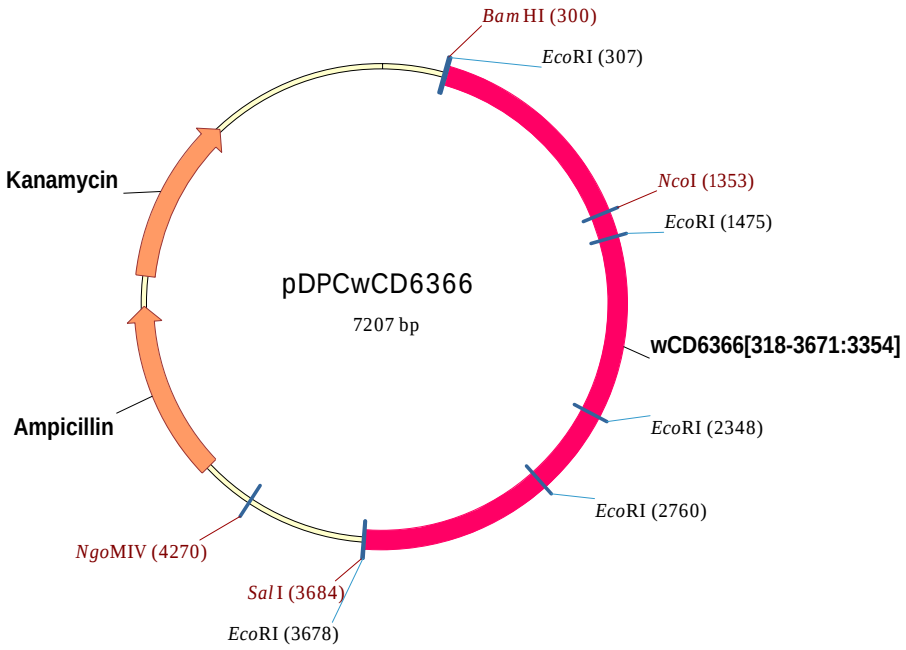


Figure 55 Map of pDPCwCD6366. The pDPCwCD6366 plasmid of 7207 base pairs in closed circular form was obtained from inserting the DNA fragment 'wCD6366' of 3354 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wCD6366' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

8.1.3 Construction of pDPCwPD7940 plasmid

Plasmid pDPCwPD7940 was constructed from inserting a linear DNA construct 'wPD7940' generated from assembly of two overlapping DNA fragments 'wPB8062' and 'wCD6366' into U-tailed pDrive cloning vector (QIAGEN). Two different recombinant plasmids constructed in this research, pDPCwPB7962 (Figure 52) and pDPCwCD6366 (Figure 55), as described in section 8.1.1 and 8.1.2, respectively, were used as templates for generating two overlapping cDNA fragments consisting 'wPB8062' and 'wCD6366' fragments, respectively, by high fidelity PCR. For amplification of 'wPB8062' fragment of 3628 base pairs, high fidelity PCR was performed with pDPCwPB7962 plasmid as template using primers PC80 and PC62 for DNA polymerization priming. Fragment 'wCD6366' of 3354 base pairs was amplified from pDPCwCD6366 plasmid using primers PC63 and PC66. The two fragments, 'wPB8062' and 'wCD6366', which share the 74-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). To assemble the two linear DNA fragments, overlap extension PCR was performed with 'wPB8062' and 'wCD6366' fragments as overlapping DNA templates using primers PC80 and PC66. Approximately 6.9-kilobase pairs DNA fragment, named 'wPD8066', amplified from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wPD7940' fragment. The 'wPD7940' fragment was amplified from 'wPD8066' fragment as template using primers PC79 and PC40 in long PCR. The blunt-ended 'wPD7940' fragment of 5663 base pairs in size obtained from the long PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction. Recombinant plasmid containing the 'wPD7940' fragment in the pDrive vector, named pDPCwPD7940, is circular DNA of 9516 base pairs in size. Outline of constructing the recombinant plasmid pDPCwPD7940 is shown in Figure 56. Orientation of the insert fragment 'wPD7940' of 5663 base pairs in the recombinant plasmid pDPCwPD7940 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific

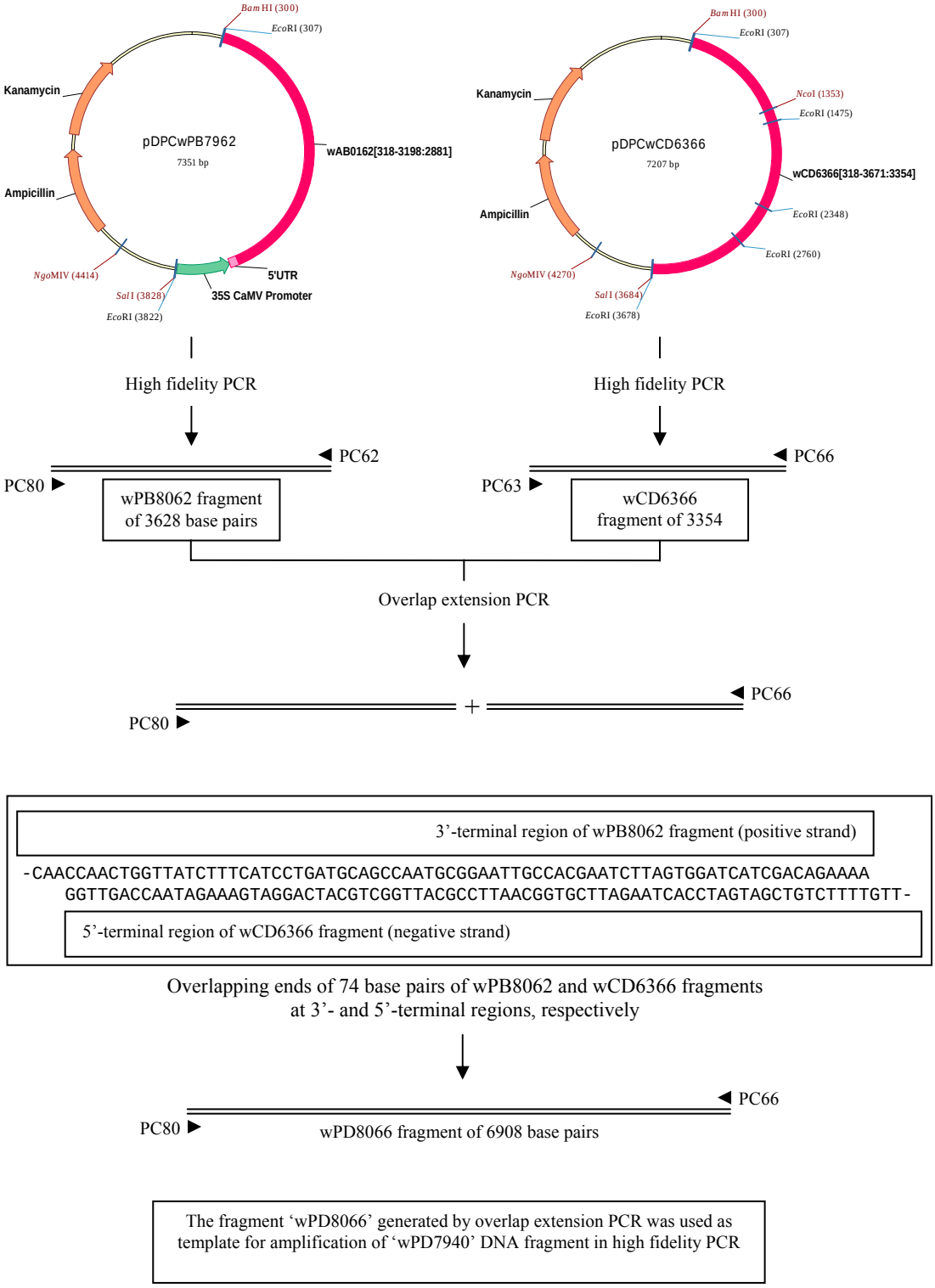


Figure 56 Outline of construction of pDPCwPD7940

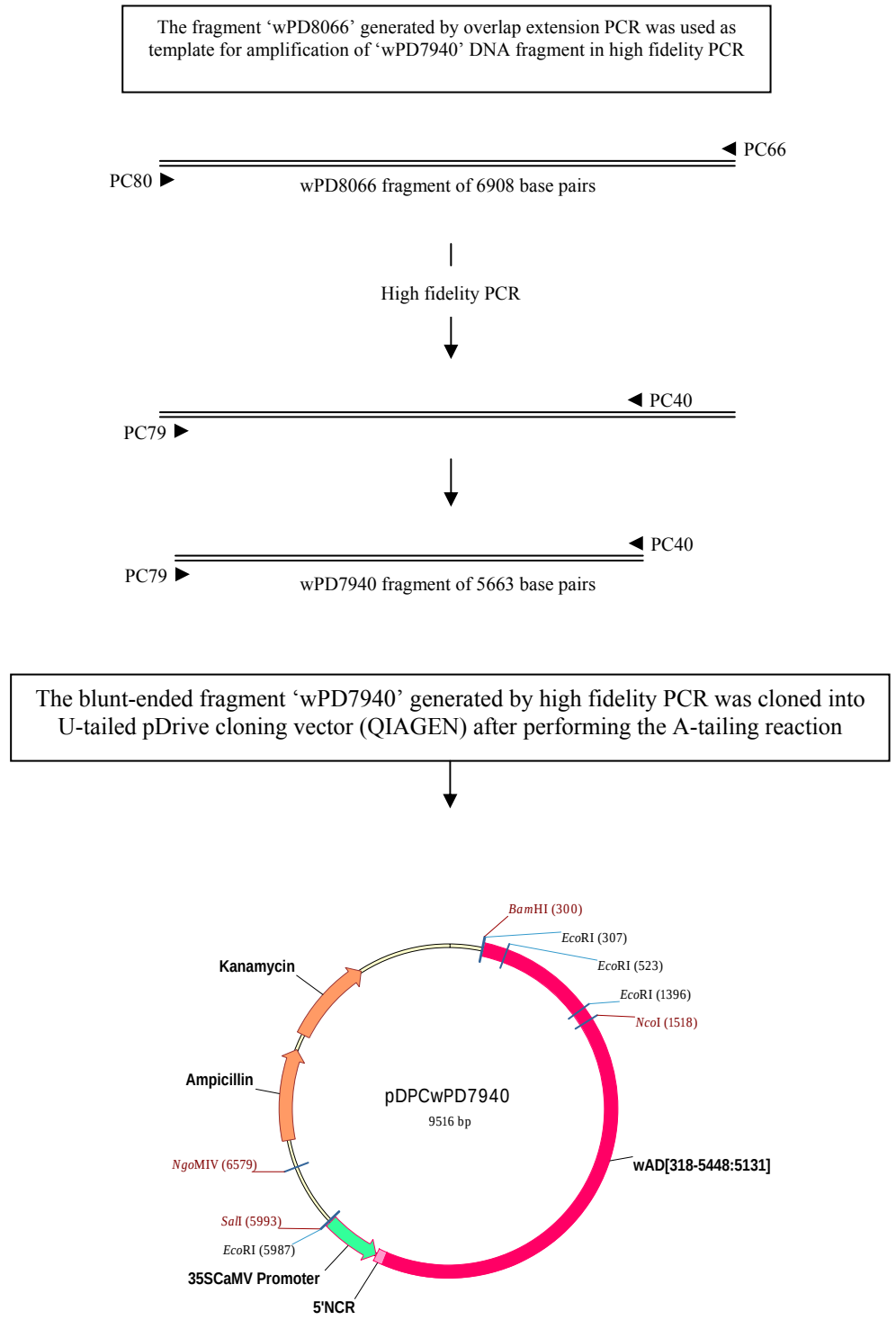


Figure 56 (continued)

primers including PC60 and PC37 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwPD7940 plasmid as substrate by digesting with four restriction endonucleases including *Bam*HI, *Nco*I, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 57. Proposed circular DNA map of pDPCwPD7940 plasmid constructing as described above is shown in Figure 58. The pDPCwPD7940 plasmid was used for further constructions.

8.1.4 Construction of pDPCwAD0140 plasmid

Plasmid pDPCwAD0140 was constructed from inserting a linear DNA construct 'wAD0140' of 5131 base pairs into U-tailed pDrive cloning vector (QIAGEN). The pDPCwPD7940 plasmid was used as template in a long PCR for amplification of 'wAD0140' fragment. The blunt-ended 'wAD0140' fragment obtained from the long PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction. Recombinant plasmid containing the 'wAD0140' fragment in the pDrive vector, named pDPCwAD0140, is circular DNA of 8984 base pairs in size. Outline of constructing the recombinant plasmid pDPCwAD0140 is shown in Figure 59. Orientation of the insert fragment 'wAD0140' of 5131 base pairs in the recombinant plasmid pDPCwAD0140 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC36 and PC37 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwAD0140 plasmid as substrate by digesting with four restriction endonucleases including *Bam*HI, *Nco*I, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 60. Proposed circular DNA map of pDPCwAD0140 plasmid constructing as described above is shown in Figure 61. The pDPCwAD0140 plasmid was used for further constructions.

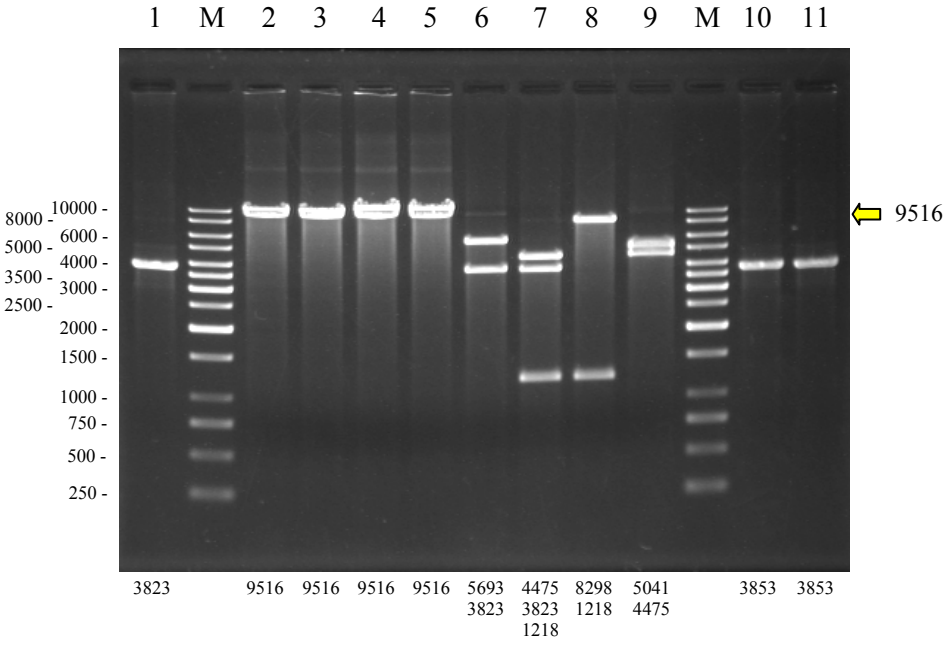


Figure 57 Restriction analyses of the pDPCwPD7940 plasmid. The pDPCwPD7940 plasmid of 9516 base pairs in closed circular form obtained from inserting the DNA fragment 'wPD7940' of 5663 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwPD7940 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive vector self-ligating digested with *Bam*HI and *Sal*I
Lane 2: pDPCwPD7940 digested with *Bam*HI
Lane 3: pDPCwPD7940 digested with *Sal*I
Lane 4: pDPCwPD7940 digested with *Xba*I
Lane 5: pDPCwPD7940 digested with *Nco*I
Lane 6: pDPCwPD7940 digested with *Bam*HI and *Sal*I
Lane 7: pDPCwPD7940 digested with *Bam*HI, *Nco*I and *Sal*I
Lane 8: pDPCwPD7940 digested with *Bam*HI and *Nco*I
Lane 9: pDPCwPD7940 digested with *Nco*I and *Sal*I
Lane 10: pDrive vector self-ligating digested with *Bam*HI and *Nco*I
Lane 11: pDrive vector self-ligating digested with *Nco*I and *Sal*I

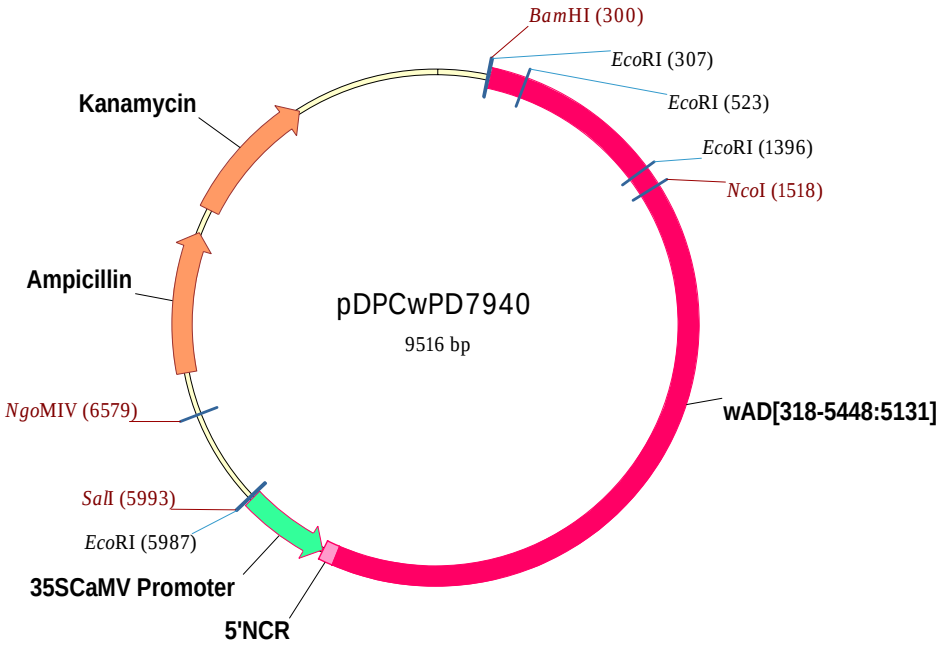


Figure 58 Map of pDPCwPD7940. The pDPCwPD7940 plasmid of 9516 base pairs in closed circular form was obtained from inserting the DNA fragment 'wPD7940' of 5663 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wPD7940' is depicted by a shaded arrowhead for 35S CaMV Promoter region, and a shaded box for fragment 'wAD'region. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

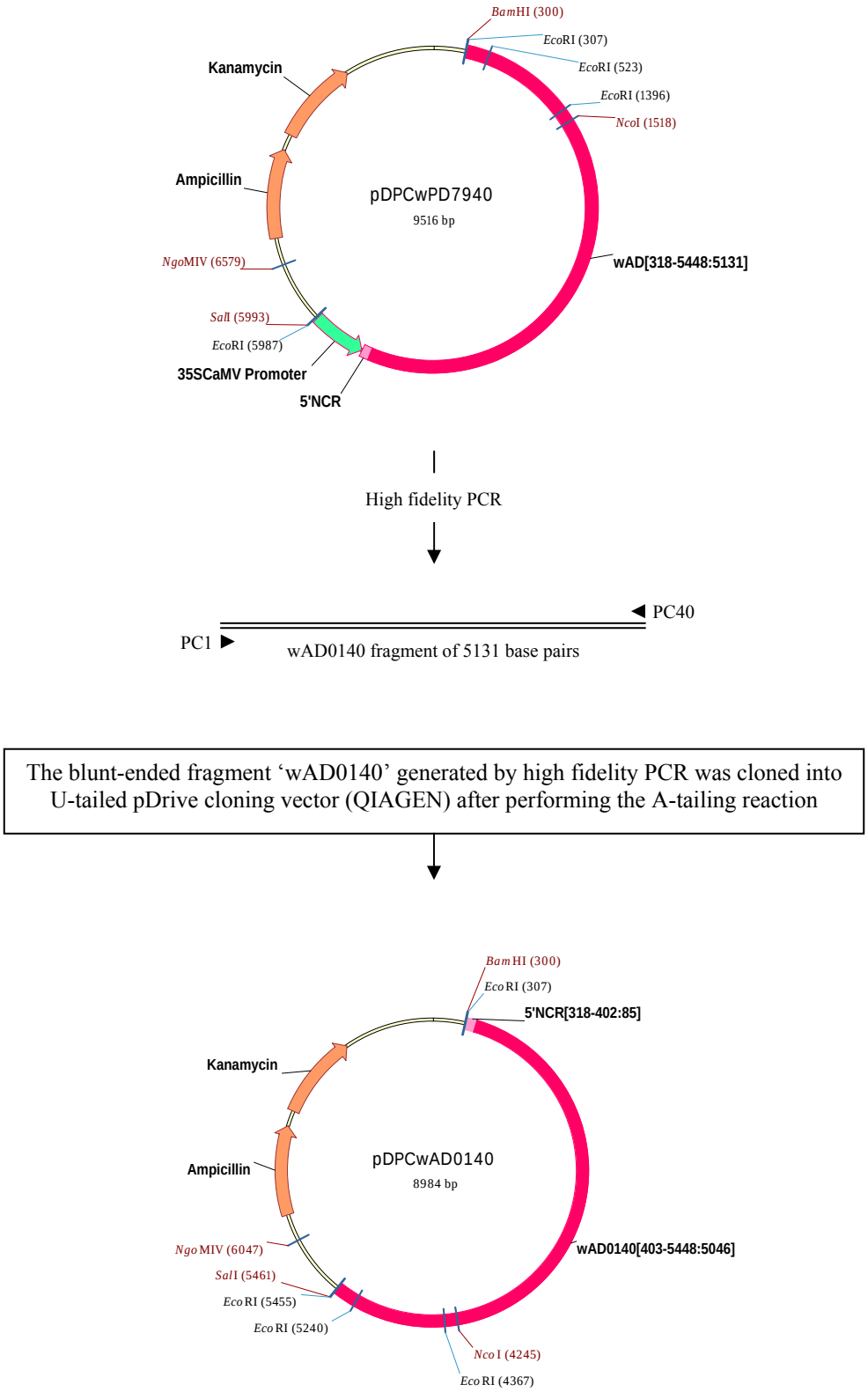


Figure 59 Outline of construction of pDPCwAD0140

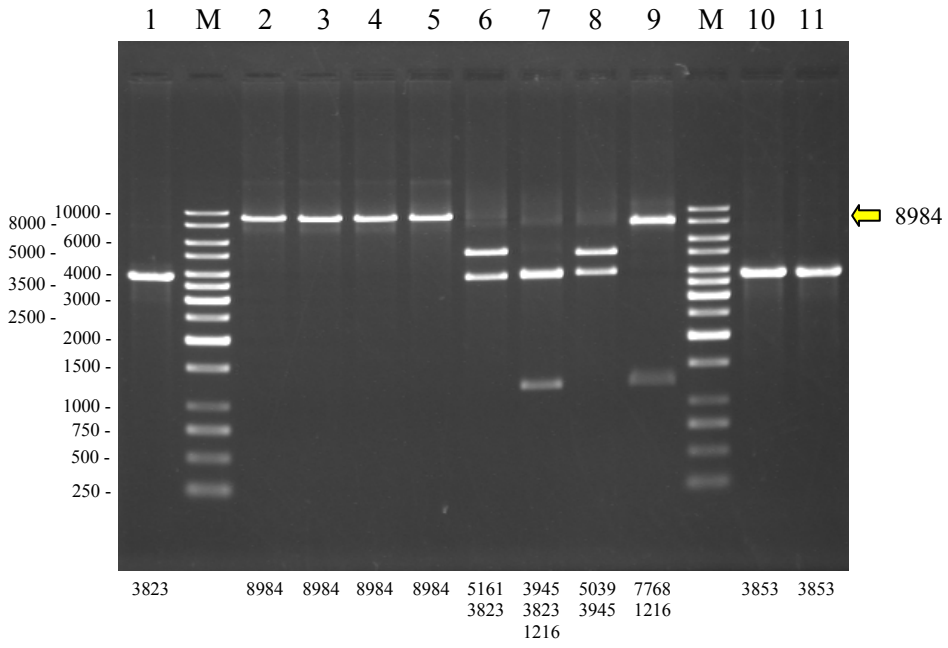


Figure 60 Restriction analyses of the pDPCwAD0140 plasmid. The pDPCwAD0140 plasmid of 8984 base pairs in closed circular form obtained from inserting the DNA fragment 'wAD0140' of 5131 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwAD0140 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive vector self-ligating digested with *Bam*HI and *Sal*I
Lane 2: pDPCwAD0140 digested with *Bam*HI
Lane 3: pDPCwAD0140 digested with *Sal*I
Lane 4: pDPCwAD0140 digested with *Xba*I
Lane 5: pDPCwAD0140 digested with *Nco*I
Lane 6: pDPCwAD0140 digested with *Bam*HI and *Sal*I
Lane 7: pDPCwAD0140 digested with *Bam*HI, *Nco*I and *Sal*I
Lane 8: pDPCwAD0140 digested with *Bam*HI and *Nco*I
Lane 9: pDPCwAD0140 digested with *Nco*I and *Sal*I
Lane 10: pDrive vector self-ligating digested with *Bam*HI and *Nco*I
Lane 11: pDrive vector self-ligating digested with *Nco*I and *Sal*I

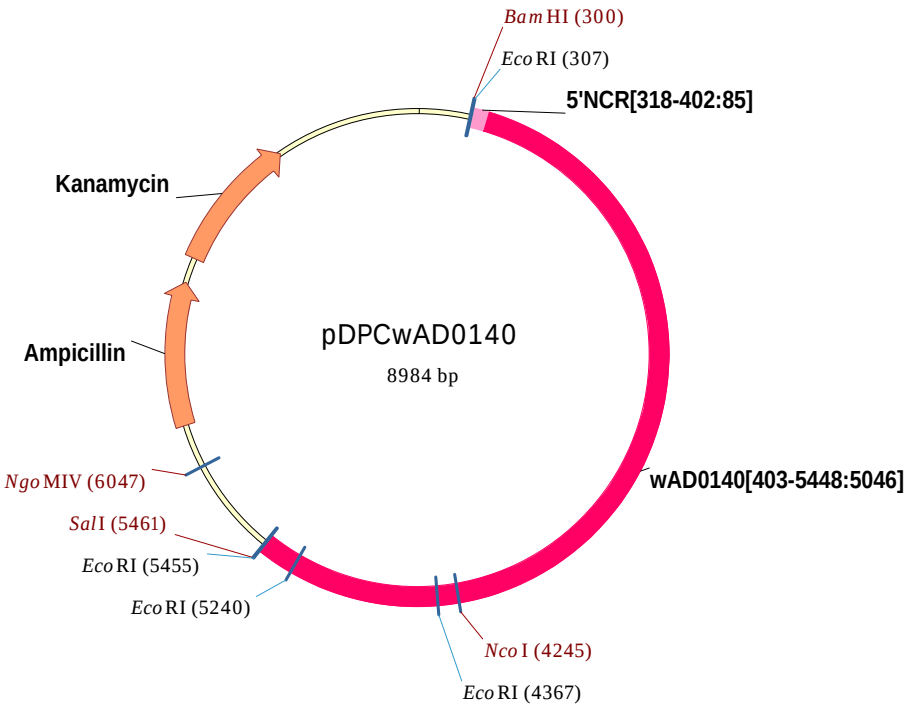


Figure 61 Map of pDPCwAD0140. The pDPCwAD0140 plasmid of 8984 base pairs in closed circular form was obtained from inserting the DNA fragment 'wAD0140' of 5136 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wAD0140' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

8.2 Construction of pTPCwCF6370 plasmid

Plasmid pTPCwCF6370 was constructed from inserting a linear DNA construct ‘wCF6370’ amplified from an overlap extension PCR product of two overlapping DNA fragments ‘wCD8026’ and ‘wDF392SC8’ into T-tailed pTZ57R cloning vector (Fermentas). By high fidelity PCR using primers PC80 and PC26, the ‘wCD8026’ DNA fragment of 3012 base pairs was amplified from the plasmid pDPCwCD6366 (Figure 55) constructed as described in section 8.1.2. While the ‘wDF392SC8’ was amplified from pDPCwDF6570 plasmid constructed from three overlapping DNA fragments amplified separately from three different recombinant plasmids constructed in this research, pPCwD (Figure 12), pPCwE (Figure 13) and pPCwF (Figure 14), as described in section 3.1.

8.2.1 Construction of pDPCwDF6570 plasmid

Initial step for constructing pDPCwDF6570 plasmid, an overlap extension PCR for assembly of two linear DNA fragments between ‘wE1568’ and ‘wF692SC8’ fragments was performed to generate a linear DNA construct which contains DNA sequences assembling from the two DNA fragments. For amplification of ‘wE1568’ fragment of 1173 base pairs, high fidelity PCR was performed with pPCwE plasmid as template using primers PC15 and PC68 for DNA polymerization priming. Fragment ‘wF692SC8’ of 909 base pairs was amplified from pPCwF plasmid using primers PC69 and 2SC8. The two fragments, ‘wE1568’ and ‘wF692SC8’, which share the 51-base pairs overlapping cDNA region at 3’- and 5’-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). To assemble the two linear DNA fragments, overlap extension PCR was performed with ‘wE1568’ and ‘wF692SC8’ fragments as overlapping DNA templates using primers PC15 and 2SC8. Approximately 2-kilobase pairs DNA fragment, named ‘wEF152SC8’, generated from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of ‘wEF672SC8’ fragment. The ‘wEF672SC8’

fragment was amplified from ‘wEF152SC8’ fragment as template using primers PC67 and 2SC8 in high fidelity PCR. The blunt-ended ‘wEF672SC8’ fragment of 1955 base pairs in size obtained from the high fidelity PCR was used as an overlapping DNA template with another ‘wD1366’ overlapping DNA template for further construction of a linear DNA fragment of ‘wDF132SC8’

To construct the linear DNA ‘wDF132SC8’ fragment, the fragment ‘wD1366’ of 1763 base pairs was amplified by high fidelity PCR from pPCwD plasmid as template using primers PC13 and PC66 and was then used as an overlapping DNA template together with the ‘wEF672SC8’ fragment generated as described above. The two fragments, ‘wD1366’ of 1763 bases pairs and ‘wEF672SC8’ of 1955 base pairs, which share the 42-base pairs overlapping cDNA region at 3’- and 5’-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). By overlap extension PCR using primers PC13 and 2SC8, a linear DNA ‘wDF132SC8’ fragment of 3629 base pairs was generated from assembly of the two overlapping DNA fragments (‘wD1366’ and ‘wEF672SC8’) and used as DNA template for amplification of ‘wDF6570’ DNA fragment after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN).

For construction of pDPCwDF6570 plasmid, the ‘wDF6570’ fragment was amplified from the ‘wDF132SC8’ fragment as template using primers PC65 and PC70 in high fidelity PCR. The blunt-ended ‘wDF6570’ fragment of 3436 base pairs in size obtained from the high fidelity PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction. Recombinant plasmid containing the ‘wDF6570’ fragment in the pDrive vector, named pDPCwDF6570, is circular DNA of 7289 base pairs in size. Outline of constructing the recombinant plasmid pDPCwDF6570 is shown in Figure 62. Orientation of the insert fragment ‘wDF6570’ of 3436 base pairs in the recombinant plasmid pDPCwDF6570 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC66 and PC67 (data not shown). Restriction analyses were performed with the close circular construct of

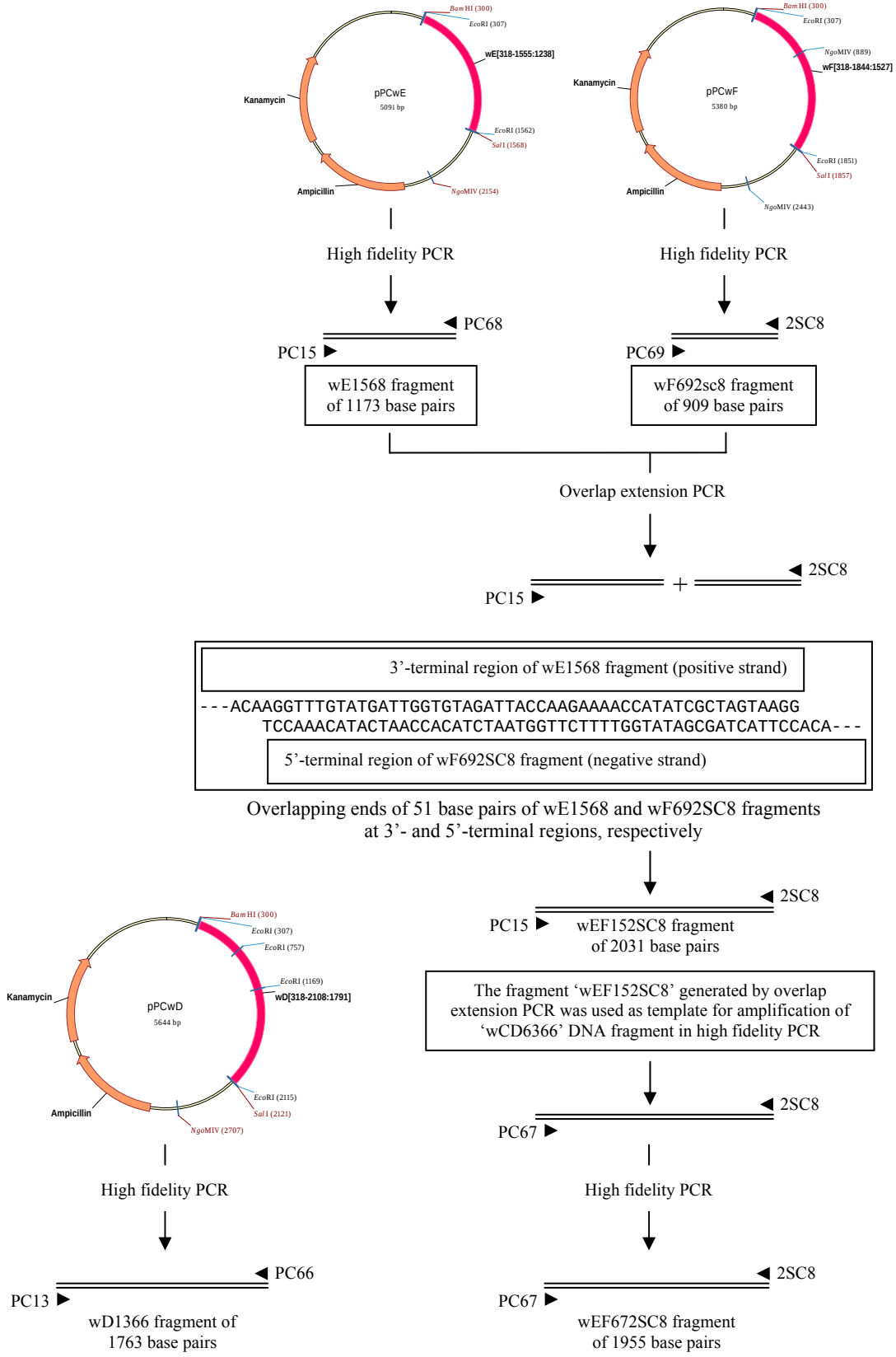


Figure 62 Outline of construction of pDPCwDF6570

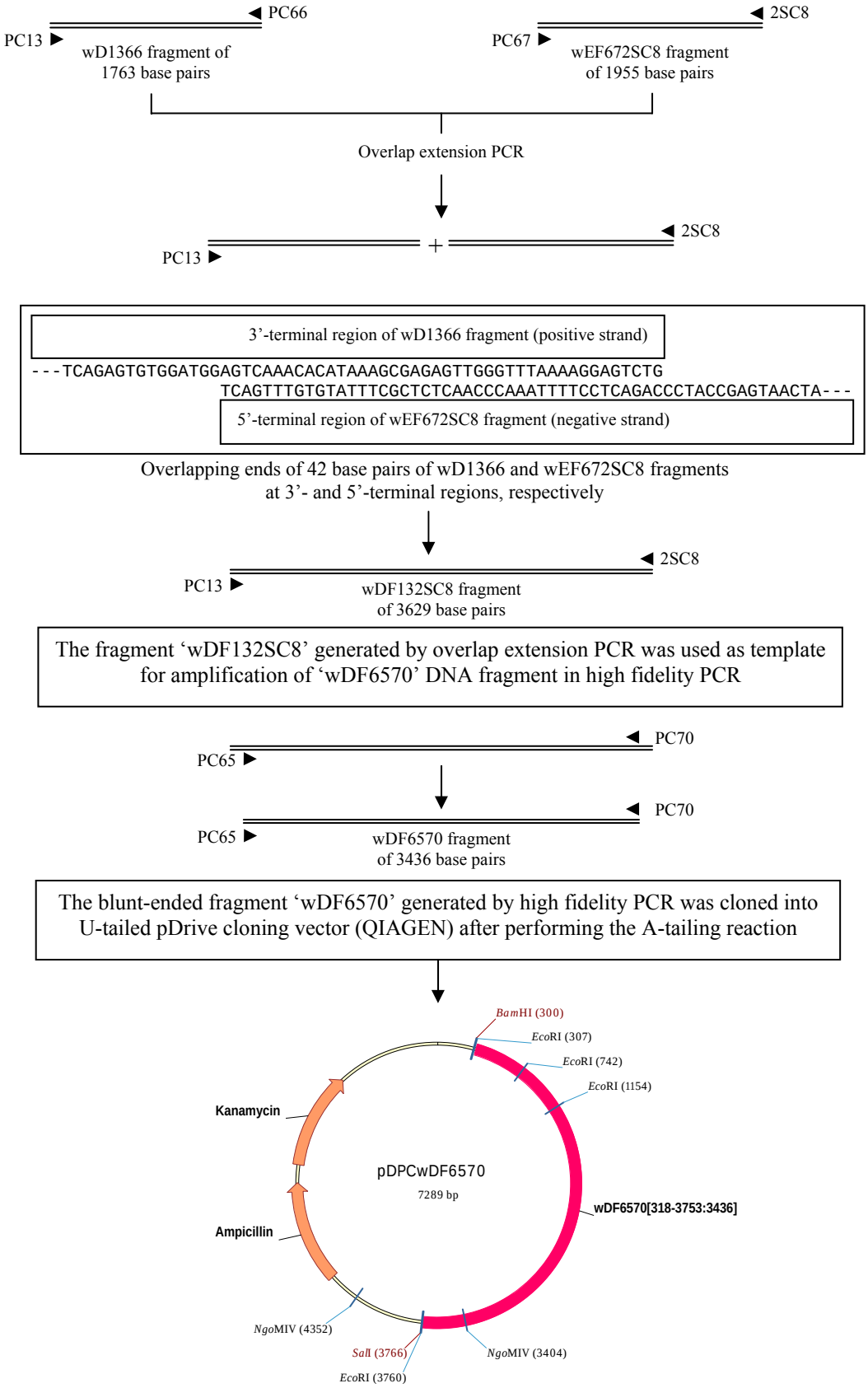


Figure 62 (continued)

pDPCwDF6570 plasmid as substrate by digesting with three restriction endonucleases including *Bam*HI, *Eco*RI and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 63. Proposed circular DNA map of pDPCwDF6570 plasmid constructing as described above is shown in Figure 64. The pDPCwDF6570 plasmid was used for further constructions.

8.2.2 Construction of pTPCwCF6370 plasmid

By separate high fidelity PCRs, the two recombinant plasmids pDPCwCD6366 (Figure 55) constructed as described in section 8.1.2 and pDPCwDF6570 (Figure 62) constructed as described in section 8.2.1 were used separately as templates for generating two overlapping cDNA fragments consisting 'wCD8026' and 'wDF392SC8' fragments, respectively. The 'wCD8026' fragment was amplified from pDPCwCD6366 plasmid of 7207 base pairs as template using primers PC80 and PC26, while the 'wDF392SC8' fragment was amplified from pDPCwDF6570 plasmid of 7289 base pairs as template using primers PC39 and 2SC8. Both 'wCD8026' of 3012 base pairs and 'wDF392SC8' of 2569 base pairs fragments amplified separately in high fidelity PCRs were used simultaneously as overlapping DNA templates for assembling an approximately 5.4 kilobase pairs recombinant product from the two fragments. The approximately 5.4-kilobase pairs DNA fragment, named 'wCF802SC8', amplified from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wCF6370' fragment. The 'wCF6370' fragment was amplified from the 'wCF802SC8' fragment as template using primers PC63 and PC70 in a long PCR. The blunt-ended 'wCF6370' fragment of 5042 base pairs in size obtained from the long PCR was cloned into T-tailed pTZ57R cloning vector (Fermentas) after performing the A-tailing reaction.

Recombinant plasmid containing the 'wCF6370' fragment in the pTZ57R vector, named pTPCwCF6370, is circular DNA of 7930 base pairs in size. Outline of constructing the recombinant plasmid pTPCwCF6370 is shown in Figure 65. Orientation of the insert fragment 'wCF6370' of 5042 base pairs in the recombinant

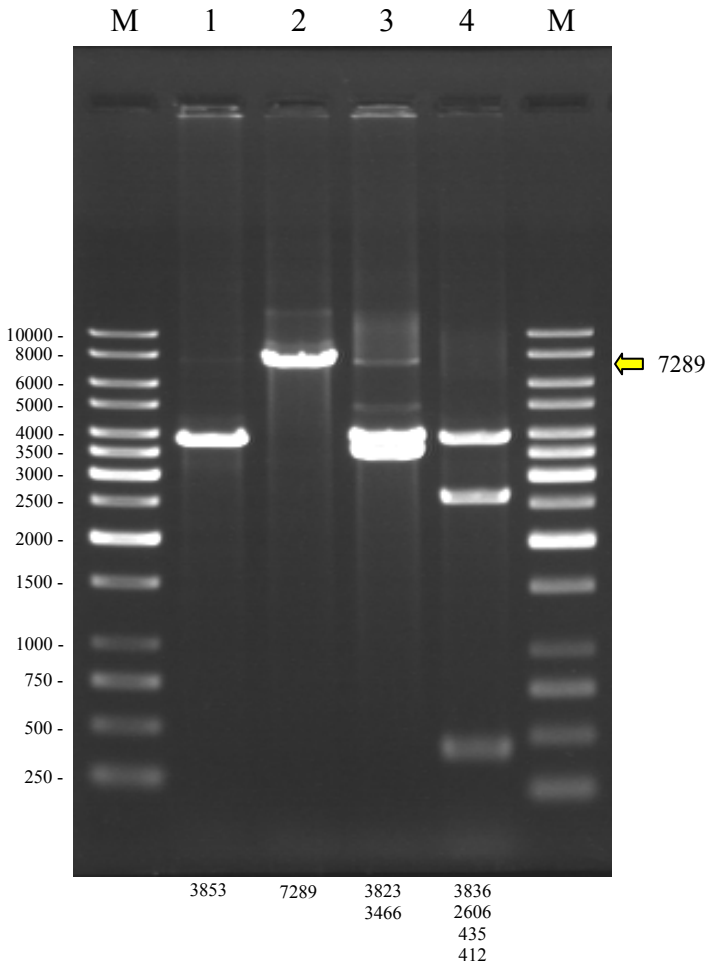


Figure 63 Restriction analyses of the pDPCwDF6570 plasmid. The pDPCwDF6570 plasmid of 7289 base pairs in closed circular form obtained from inserting the DNA fragment 'wDF6570' of 3436 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwDF6570 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive cloning vector self-ligating digested with *Bam*HI
Lane 2: pPCwDF6570 digested with *Bam*HI
Lane 3: pPCwDF6570 digested with *Bam*HI and *Sal*I
Lane 4: pPCwDF6570 digested with *Eco*RI

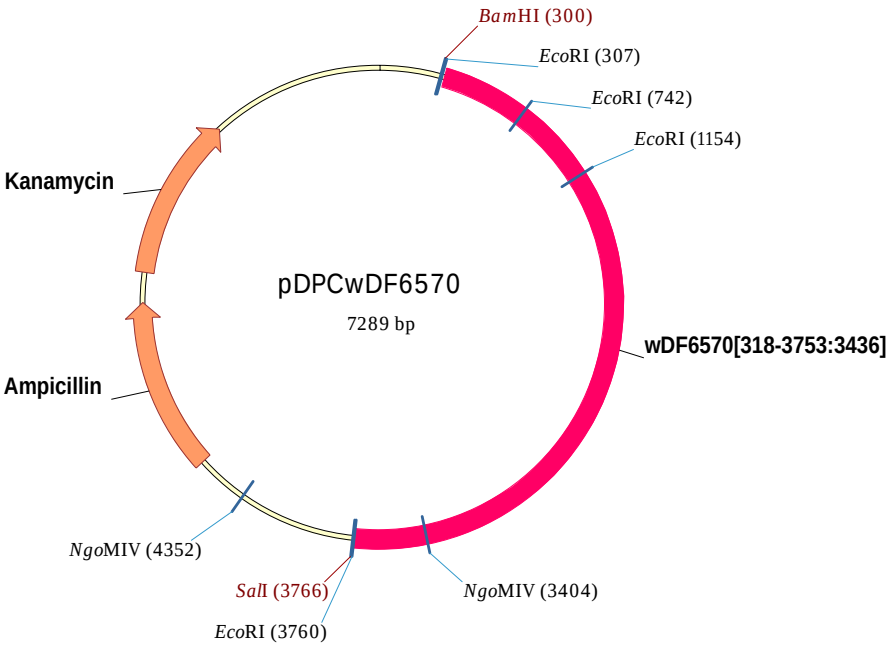


Figure 64 Map of pDPCwDF6570. The pDPCwDF6570 plasmid of 7289 base pairs in closed circular form was obtained from inserting the DNA fragment 'wDF6570' of 3436 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wDF6570' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

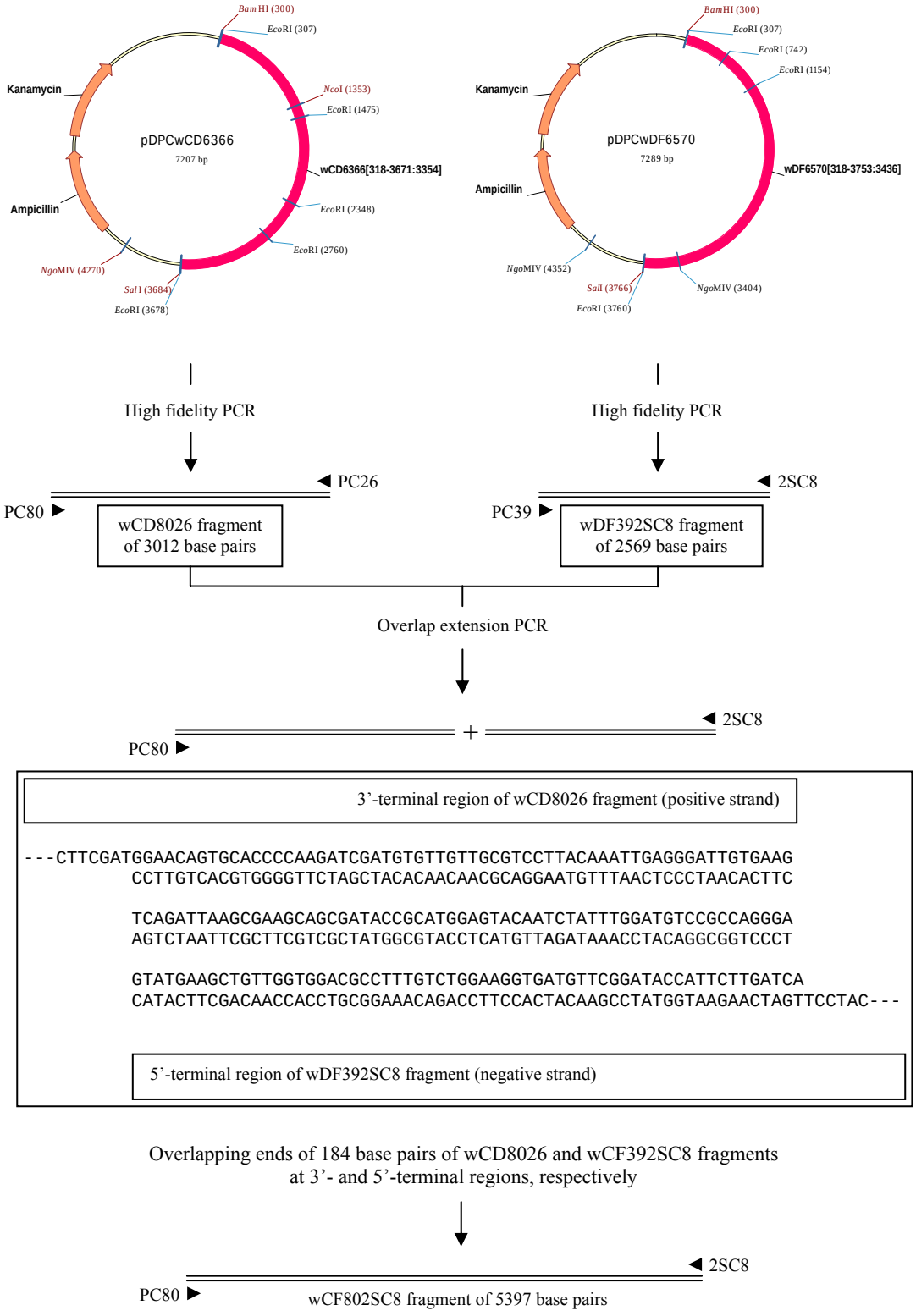


Figure 65 Outline of construction of pTPCwCF6370

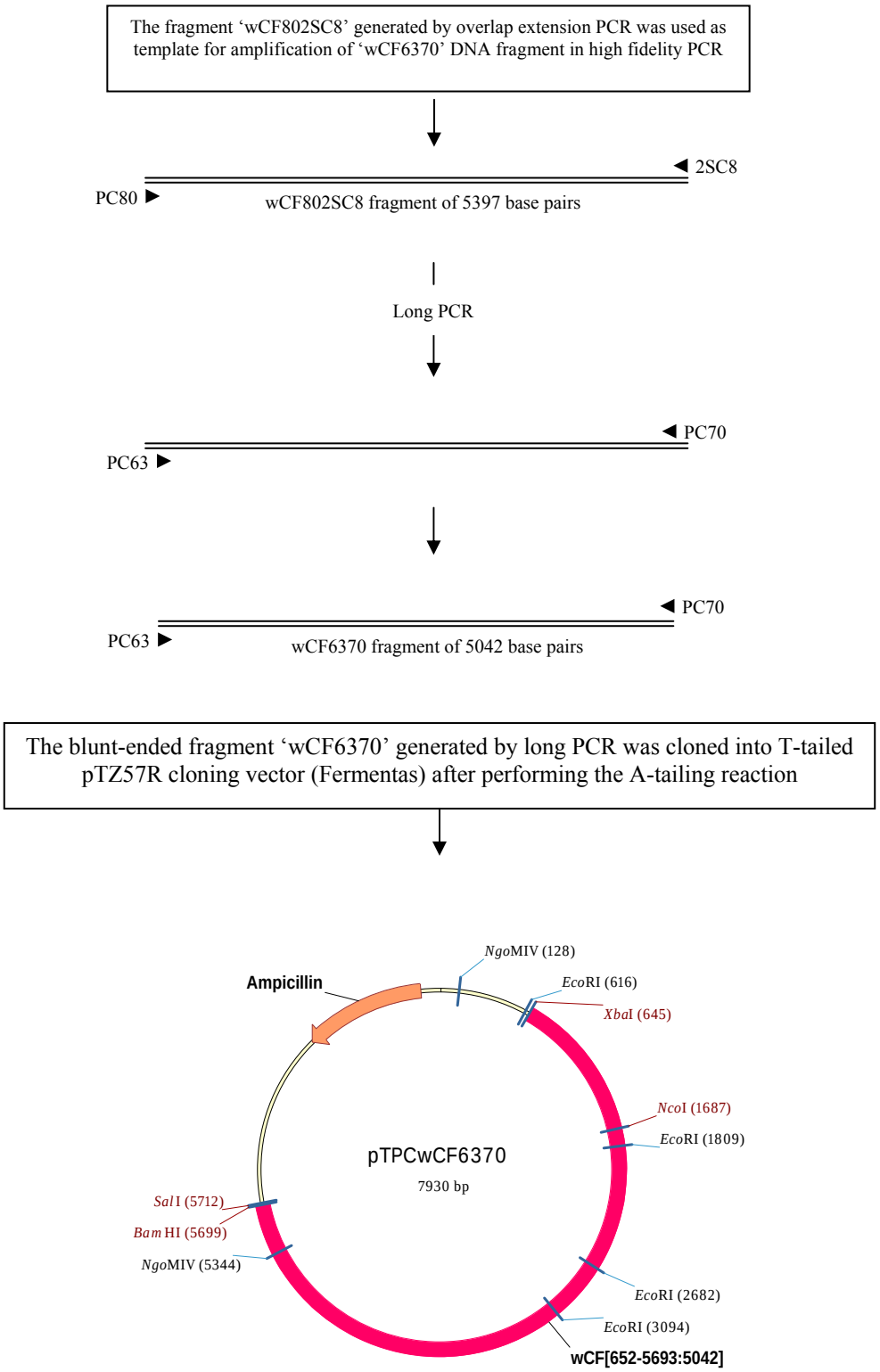


Figure 65 (continued)

plasmid pTPCwCF6370 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC64 and PC67 (data not shown). Restriction analyses were performed with the close circular construct of pTPCwCF6370 plasmid as substrate by digesting with five restriction endonucleases including *Bam*HI, *Nco*I, *Ngo*MIV, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 66. Proposed circular DNA map of pTPCwCF6370 plasmid constructing as described above is shown in Figure 67. The pTPCwCF6370 plasmid was used for further constructions.

3.8.3 Construction of pTPCwDN6578 plasmid

A recombinant plasmid, named pTPCwDN6578, was constructed by inserting a 'wDN6578' cDNA fragment of 6086 base pairs into T-tailed pTZ57R cloning vector (Fermentas). The insert cDNA fragment contains DNA sequences corresponding to the 3'-terminal half of the PRSV-W genome region at nucleotide positions 4499 to 10347 including the additional 24-mer of A-tail and 237-base pair NOS terminator sequence with additional 6-base pair sequence specific to *Hin*dIII endonuclease at the 3'-terminus. The 5825-base pair 5'-terminal region of the 'wDN6578' fragment corresponding to the 3'-terminal nine tenth of the CI gene, 6K1, VPg, NIa-Pro, NIb, CP genes and 3'-NCR of the PRSV-W isolate studied in this research. The 'wDN6578' linear DNA construct of 6086 base pairs was developed from two recombinant plasmids constructed in this research, pTPCwCF6370 (Figure 67) as described in section 8.2 and pDPCwFN6978 described below.

Nopaline synthase (NOS) terminator used in this research was amplified from p2311 plasmid and cloned into U-tailed pDrive cloning vector (QIAGEN). The source of the NOS terminator, p2311 plasmid, was kindly donated by Dr. Srimek Chowpongpang, Department of Plant Pathology, Kasetsart University. The NOS terminator cDNA fragment was amplified from plasmid p2311 using primers ANOS-1 (5'-$(A)_{12}$ GTT TCT TAA GAT TGA ATC-3') and NOS-4H (5'-AAG CTT CCC GAT CTA GTA ACA TAG-3'). The obtained PCR product of approximately 0.23

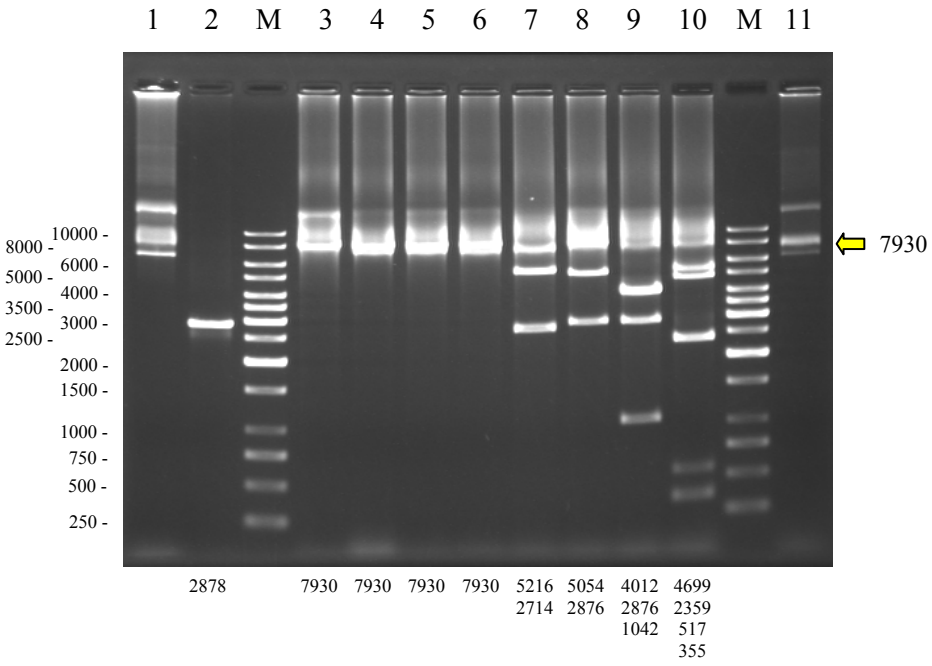


Figure 66 Restriction analyses of the pTPCwCF6370 plasmid. The pTPCwCF6370 plasmid of 7930 base pairs in closed circular form obtained from inserting the DNA fragment 'wCF6370' of 5042 base pairs into T-tailed pTZ57R cloning vector (Fermentas) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pTPCwCF6370 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pTPCwCF6370 (circular form, 1 µl)
Lane 2: pTZ57R vector self-ligating digested with *Bam*HI and *Xba*I
Lane 3: pTPCwCF6370 digested with *Bam*HI
Lane 4: pTPCwCF6370 digested with *Sal*I
Lane 5: pTPCwCF6370 digested with *Xba*I
Lane 6: pTPCwCF6370 digested with *Nco*I
Lane 7: pTPCwCF6370 digested with *Ngo*MIV
Lane 8: pTPCwCF6370 digested with *Bam*HI and *Xba*I
Lane 9: pTPCwCF6370 digested with *Bam*HI, *Nco*I and *Xba*I
Lane 10: pTPCwCF6370 digested with *Bam*HI, *Ngo*MIV and *Xba*I
Lane 11: pTPCwCF6370 (circular form, 0.5 µl)

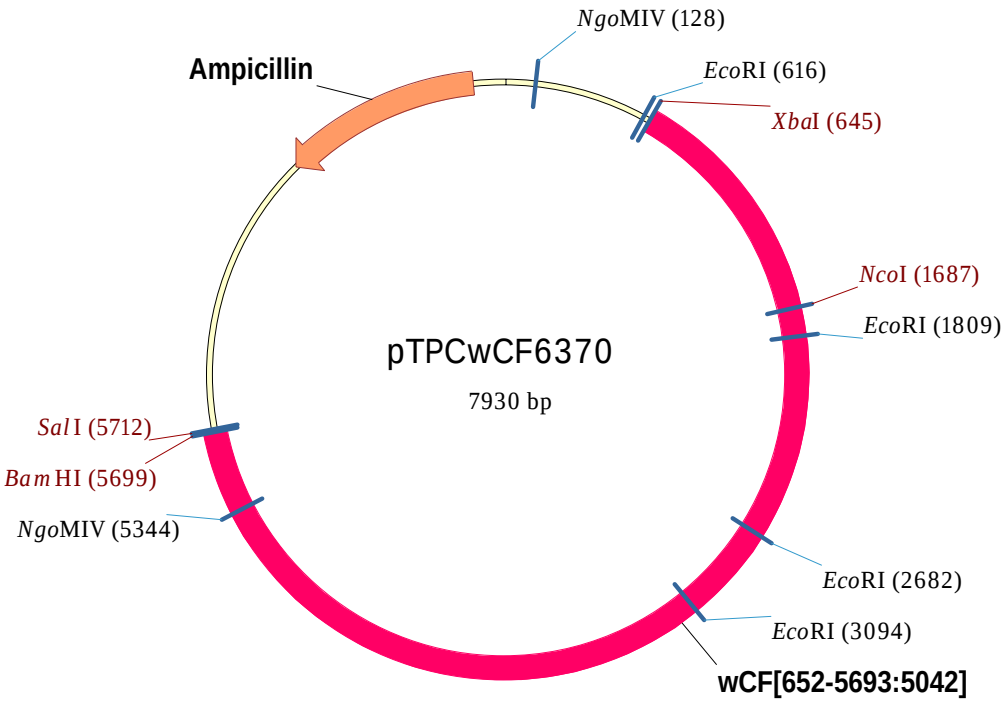


Figure 67 Map of pTPCwCF6370. The pTPCwCF6370 plasmid of 7930 base pairs in closed circular form was obtained from inserting the DNA fragment 'wCF6370' of 5042 base pairs into T-tailed pTZ57R cloning vector (Fermentas). The insert fragment 'wCF6370' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV, *Sal*I and *Xba*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the ampicillin coding region on the cloning vector is represented by a shaded arrowhead.

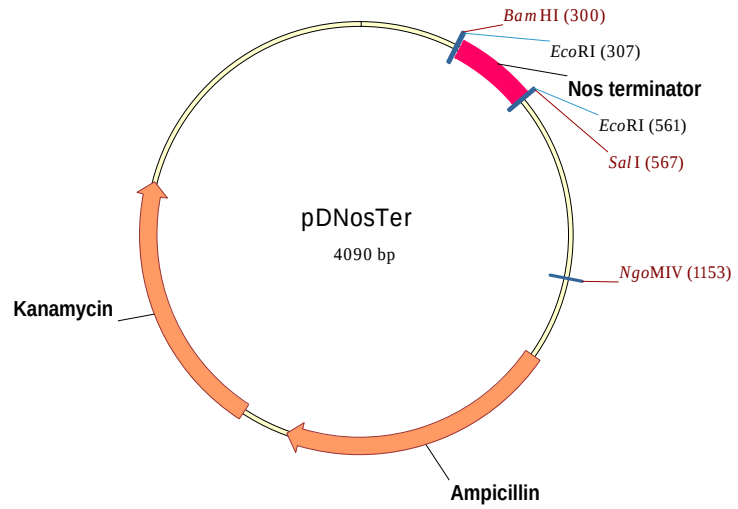
kilobase pairs in size was cloned into U-tailed pDrive cloning vector (QIAGEN) after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). Sequencing of the NOS terminator cDNA insert fragment in the obtained recombinant plasmid, named pDNosTer (Figure 68A), was performed using the ABI PRISM® Bigdye™ Terminator Cycle Sequencing Ready Reaction Kits version 2.0 and the reactions obtained were run on either the model 377 automatic DNA sequencer (Applied Biosystems). DNA sequence of the insert fragment is shown in Figure 68B for NOS terminator of 231 base pairs with additional 6-base pair sequence (5'-AAGCTT-3') specific to *Hin*dIII endonuclease at the 3'-terminus. The pDNosTer plasmid used as source of NOS terminator in this study was used for amplifying fragment 'Nos terminator' to fuse to the full-length genomic cDNA construct of the PRSV-W isolate studying in this research.

8.3.1 Construction of pDPCwFN6978

Four different recombinant plasmids constructed in this research, pDPCwF (Figure 14), pPCwG (Figure 15), pPCwH (Figure 16) and pDNosTer (Figure 68), were used as templates for amplification of four overlapping cDNA fragments consisting 'wF0970', 'wG7176' 'wH7574' and 'Nos terminator' fragments, respectively, by high fidelity PCR. All four blunt-ended PCR products were gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as templates in overlap extension PCR for assembly of proposed linear DNA construct.

Initial step for constructing pDPCwFN6978 plasmid, an overlap extension PCR for assembly of two linear DNA fragments between 'wH7574' and 'Nos terminator' fragments was performed to generate a linear DNA construct which contains cDNA fragment of 3'-terminal region ('wH7574') of PRSV-W and 'NOS terminator' fragment together with no unwanted sequences at the junction. In construction of the fused- linear DNA 'wHN7578', two primers PC74 (reverse) and A Nos-1 (forward) were designed and synthesized to use separately with primer PC75 and Nos-4H, respectively, for generating an 17-base pairs overlapping sequences

A



B

```
GTTTCTTAAG ATTGAATCCT GTTGCCGGTC TTGCGATGAT TATCATATAA TTTCTGTTGA ATTACGTTAA
GCATGTAATA ATTAACATGT AATGCATGAC GTTATTTATG AGATGGGTTT TTATGATTAG AGTCCCGCAA
TTATACATTT AATACGCGAT AGAAAACAAA ATATAGCGCG CAAACTAGGA TAAATTATCG CGCGCGGTGT
CATCTATGTT ACTAGATCGG GAAGCTT
```

Figure 68 Map of pDNosTer (A) and nucleotide sequence of Nos terminator cDNA fragment (B). The pDNosTer plasmid of 4090 base pairs in closed circular form was obtained from inserting the Nos terminator cDNA fragment into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'Nos terminator' is depicted by a shaded box. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

between ‘wH7574’ and ‘Nos terminator’ fragments at 3’- and 5’-terminal regions, respectively. Fragment ‘wH7574’ of 1135 base pairs was amplified from plasmid pPCwH using primers PC75 and PC74. NOS terminator cDNA fragment of 249 base pairs was amplified from plasmid pDNosTer using primers A Nos-1 and Nos-4H. Both fragments of ‘wH7574’ and ‘Nos terminator’ separately amplified by high fidelity PCR were used together as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). For assembly these fragments together, overlap extension PCR was performed with the two fragments which share the 17-base pairs overlapping cDNA region (‘wH7574’ and ‘Nos terminator’) as templates using primers PC75 and Nos-4H. Amplified fragment obtained from assembly of ‘wH7574’ and ‘Nos terminator’ in overlap extension PCR, named ‘wHN7578’, is 1367 base pairs in size. The blunt-ended fragment ‘wHN7578’ was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailed reaction. Recombinant plasmid containing the ‘wHN7578’ fragment in the pDrive vector, named pDPCwHN7578, is circular DNA of 4090 base pairs in size. Outline of construction of pDPCwHN7578 is shown in Figure 69. Orientation of the insert fragment ‘wHN7578’ in the recombinant plasmid pDPCwHN7578 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC75 and Nos-4H (data not shown). Restriction analyses were performed with the close circular construct of pDPCwHN7578 plasmid as substrate by digesting with three restriction endonucleases including *Bam*HI, *Eco*RI and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 70. Proposed circular DNA map of pDPCwHN7578 plasmid constructing as described above is shown in Figure 71. The pDPCwHN7578 plasmid was used as template in a high fidelity PCR for further constructions.

Plasmid pDPCwFN6978 was constructed from inserting a linear DNA construct, named ‘wFN6978’, into U-tailed pDrive cloning vector (QIAGEN). The ‘wFN6978’ fragment was amplified from an approximately 4 kilobase pairs recombinant product, named ‘wFN0980’, assembled from three overlapping fragments. Three different recombinant plasmids constructed in this research, pPCwF

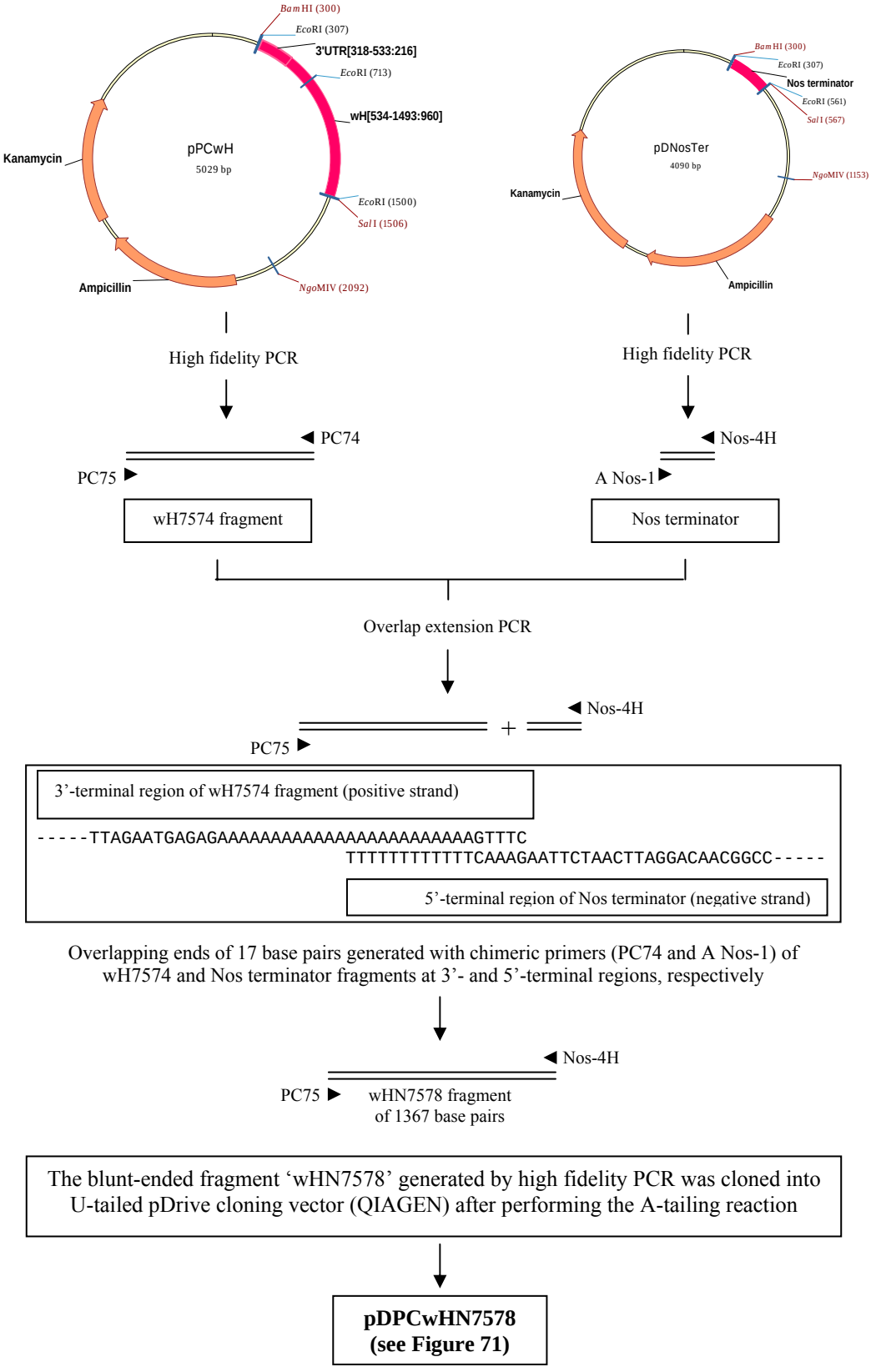


Figure 69 Outline of construction of pDPCwHN7578

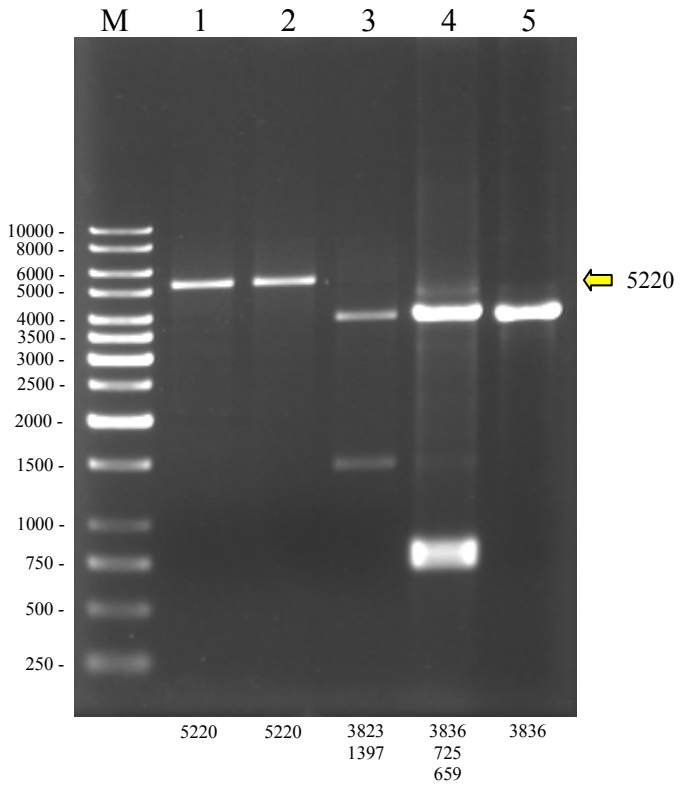


Figure 70 Restriction analyses of the pDPCwHN7578 plasmid. The pDPCw HN7578 plasmid of 5220 base pairs in closed circular form obtained from inserting the DNA fragment 'wHN' of 1367 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwHN plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDPCwHN digested with *Bam*HI
Lane 2: pDPCwHN digested with *Sal*I
Lane 3: pDPCwHN digested with *Bam*HI and *Sal*I
Lane 4: pDPCwHN digested with *Eco*RI
Lane 5: pDrive cloning vector self-ligating digested with *Eco*RI

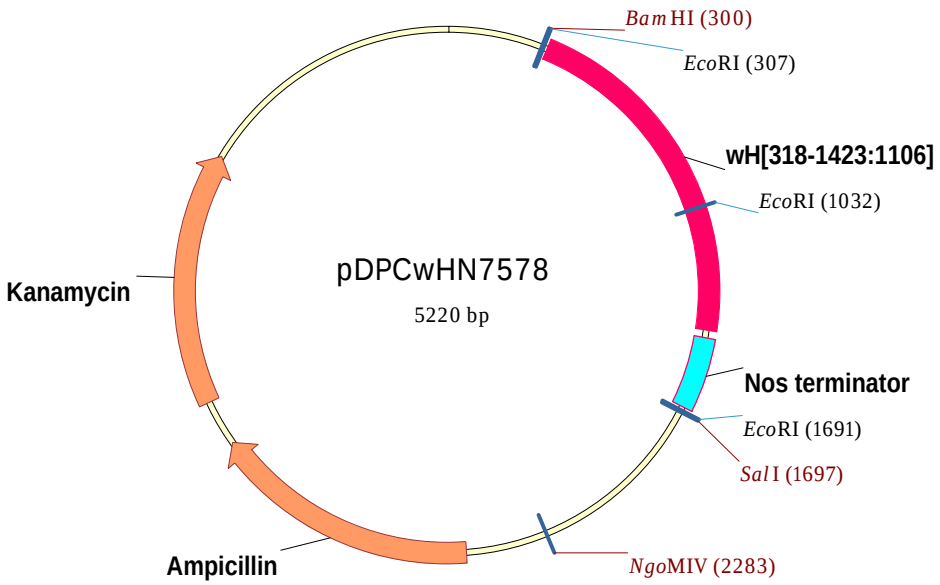


Figure 71 Map of pDPCwHN7578. The pDPCw HN7578 plasmid of 5220 base pairs in closed circular form was obtained from inserting the DNA fragment 'w HN7578' of 1367 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wHN7578' is depicted separately by shaded boxes for fragment 'wH' and Nos terminator regions. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

(Figure 14), pPCwG (Figure 15) and pPCwHN7578 (Figure 70), were used separately as templates for amplification of three overlapping cDNA fragments consisting 'wF0970', 'wG7176' and 'wHN7580' fragments, respectively. By high fidelity PCRs, the 'wF0970' fragment of 1302 base pairs was amplified from pPCwF plasmid as DNA template using primers PC9 and PC70, the 'wG7176' fragment of 1378 base pairs was amplified from pPCwG plasmid as DNA template using primers PC13 and PC76, and the 'wHN7580' fragment of 1497 base pairs was amplified from pPCwHN7578 plasmid as DNA template using primers PC75 and PC80. All three blunt-ended PCR products were gel-purified separately with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as templates in overlap extension PCR for assembly of proposed linear DNA construct. The three fragments, 'wF0970', 'wG7176' and 'wHN7580', which share the 46- and 49-base pairs overlapping cDNA regions between 'wF0970' and 'wG7176' fragments and between 'wG7176' and 'wHN7580' fragments at 3'- and 5'-terminal regions of each pair, respectively, were used simultaneously as templates in an overlap extension PCR performing with two primers PC9 and PC80. Approximately 4-kilobase pairs DNA fragment, named 'wFN0980', generated from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wFN6978' fragment. The 'wFN6978' fragment was amplified from the 'wFN0980' fragment as template using primers PC69 and PC78 in long PCR. The blunt-ended 'wFN6978' fragment of 3334 base pairs in size obtained from the long PCR was cloned into U-tailed pDrive cloning vector (QIAGEN) after performing the A-tailing reaction.

Recombinant plasmid containing the 'wFN6978' fragment in the pDrive vector, named pDPCwFN6978, is circular DNA of 7187 base pairs in size. Outline of constructing the recombinant plasmid pDPCwFN6978 is shown in Figure 72. Orientation of the insert fragment 'wFN6978' of 3334 base pairs in the recombinant plasmid pDPCwFN6978 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC70 and PC75 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwFN6978 plasmid as substrate by

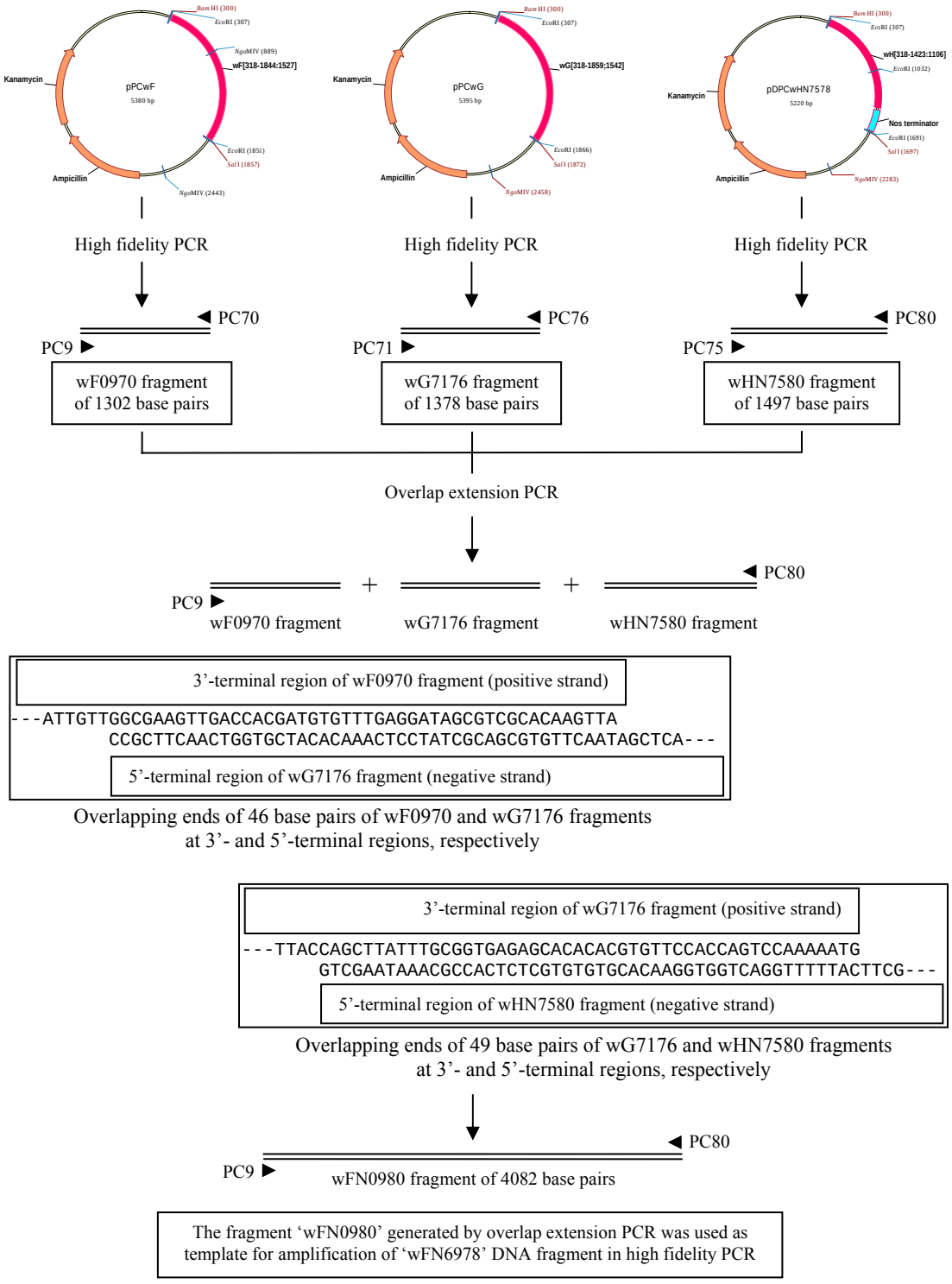


Figure 72 Outline of construction of pDPCwFN6978

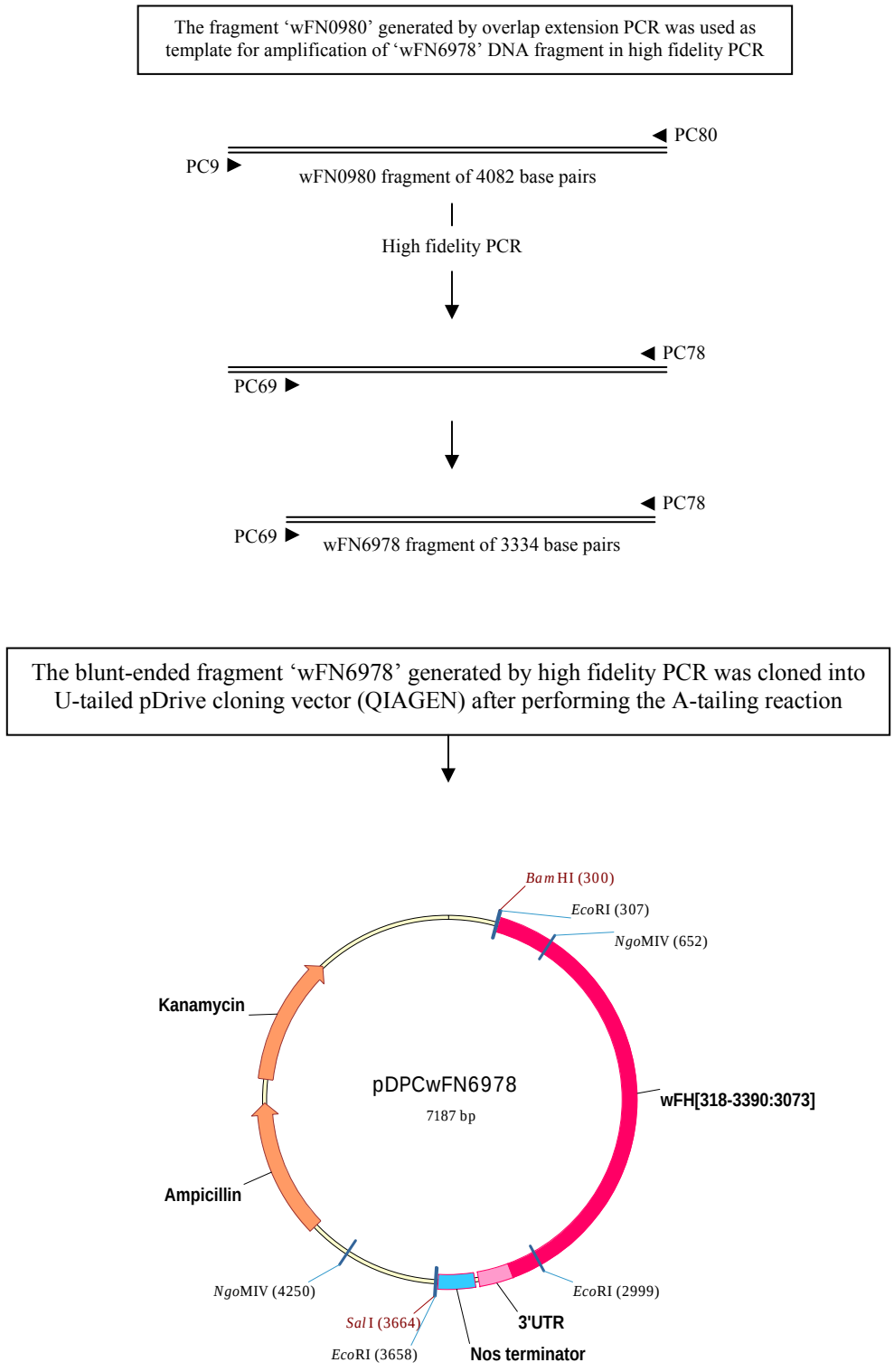


Figure 72 (continued)

digesting with three restriction endonucleases including *Bam*HI, *Eco*RI and *Sal*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 73. Proposed circular DNA map of pDPCwFN6978 plasmid constructing as described above is shown in Figure 74. The pDPCwFN6978 plasmid was used for further constructions.

8.3.2 Construction of pTPCwDN6578 plasmid

In order to construct the linear recombinant 'wDN6578' DNA fragment, 'wCN6380' DNA fragment used as template for amplifying the 'wDN6578' DNA fragment was developed from two overlapping DNA fragments 'wCF6370' and 'wFN6980'. The two overlapping DNA fragments share a 684-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, which were amplified separately from pDPCwCF6370 plasmid (Figure 67) using primers PC63 and PC70, and from pDPCwFN6978 plasmid (Figure 74) using primers PC69 and PC80, respectively, in two separate high fidelity PCRs. In the overlapping cDNA region, both DNA fragments contain a 6-base pair sequence (GCCGGC) specific to *Ngo*MIV endonuclease at the 3'- and 5'-terminal regions, repectively. The two amplified fragments, 'wCF6370' of 5042 base pairs and 'wFN6980' of 3464 base pairs, were digested separately with *Ngo*MIV endonuclease. Approximately 4.7- and 3.1-kilobase pairs DNA fragments from the separate *Ngo*MIV digestion reaction of 'wCF6370' and 'wFN6980', respectively, were gel-purified separately with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as substrates for an *in vitro* sticky end ligation to generate the 'wCN6380' recombinant DNA fragment from the two fragments. *In vitro* ligation of DNA was performed with T4 DNA ligase (Fermentas) using 4.7- and 3.1-kilobase pairs DNA fragments of 'wCF6370' and 'wFN6980', respectively, which were digested with *Ngo*MIV endonuclease as substrates, and the reaction was incubated at 10 ^{o}C for 16 hr. After incubation period of the *in vitro* ligation reaction, agarose gel electrophoresis with ethidium bromide staining was performed to separate the DNA products obtained from the reaction. A DNA fragment with the expect size of approximately 7.8 kilobase pairs was harvested and gel-purified with extraction from agarose using

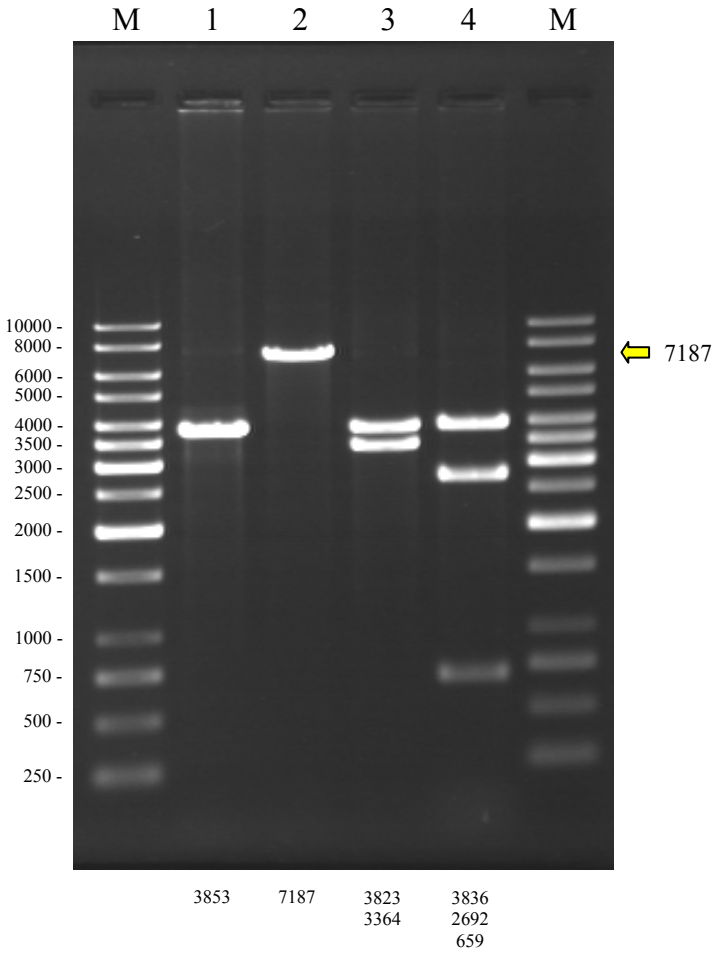


Figure 73 Restriction analyses of the pDPCwFN6978 plasmid. The pDPCwFN6978 plasmid of 7187 base pairs in closed circular form obtained from inserting the DNA fragment 'wFN6978' of 3334 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwFN6978 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDrive cloning vector self-ligating digested with *Bam*HI
Lane 2: pPCwFN6978 digested with *Bam*HI
Lane 3: pPCwFN6978 digested with *Bam*HI and *Sal*I
Lane 4: pPCwFN6978 digested with *Eco*RI

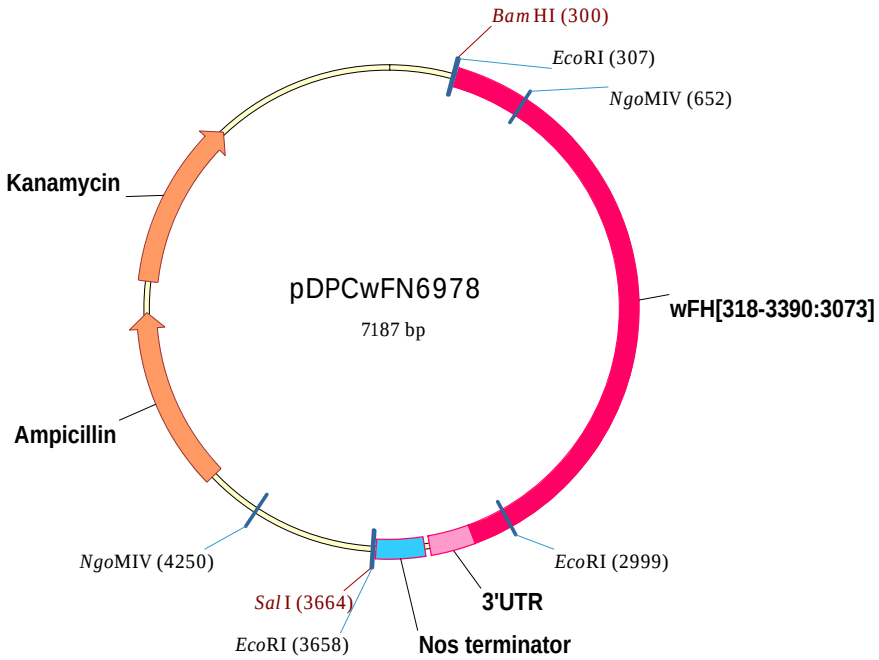


Figure 74 Map of pDPCwFN6978. The pDPCwFN6978 plasmid of 7187 base pairs in closed circular form was obtained from inserting the DNA fragment 'wFN6978' of 3334 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wFN6978' is depicted by shaded boxes. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

QIAquick® Gel Extraction kit (QIAGEN). The approximately 7.8-kilobase pairs DNA fragment obtained from the ligation reaction, named 'wCN6380', was used as DNA template for amplifying the 'wDN6578' fragment. A long PCR was performed with the 'wCN6380' fragment as DNA template using primers PC65 and PC78 to amplify the 'wDN6578' DNA fragment of 6086 base pairs. The blunt-ended 'wDN6578' fragment of 6086 base pairs in size obtained from the long PCR was cloned into T-tailed pTZ57R cloning vector (Fermentas) after performing the A-tailing reaction. Recombinant plasmid containing the 'wDN6578' fragment in the pTZ57R vector, named pTPCwDN6578, is circular DNA of 8974 base pairs in size. Outline of constructing the recombinant plasmid pTPCwDN6578 is shown in Figure 75. Orientation of the insert fragment 'wDN6578' of 6086 base pairs in the recombinant plasmid pTPCwDN6578 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC64 and PC75 (data not shown). Restriction analyses were performed with the close circular construct of pTPCwDN6578 plasmid as substrate by digesting with four restriction endonucleases including *Bam*HI, *Ngo*MIV, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 76. Proposed circular DNA map of pTPCwDN6578 plasmid constructing as described above is shown in Figure 77. The pTPCwDN6578 plasmid was used for further constructions.

8.4 Construction of a full-length cDNA clone of a Thai isolate of PRSV-W for expressing viral RNA through *in vivo* transcription

A recombinant plasmid, named pDPCwPN7978, was constructed by inserting a full-length cDNA with nucleotide sequence representing the genomic RNA of a Thai isolate of PRSV-W following a poly(A) 24 tract assembled with a CaMV 35S promoter and a NOS terminator fragment of 11116 base pairs, named 'wPN7978', into U-tailed pDrive cloning vector (QIAGEN). In the insert fragment, the full-length cDNA corresponding to the PRSV-W genome was assembled with a CaMV 35S promoter and a NOS terminator cDNA sequences at 5'- and 3'-termini, respectively. The recombinant plasmid pDPCwPN7978 was developed from three

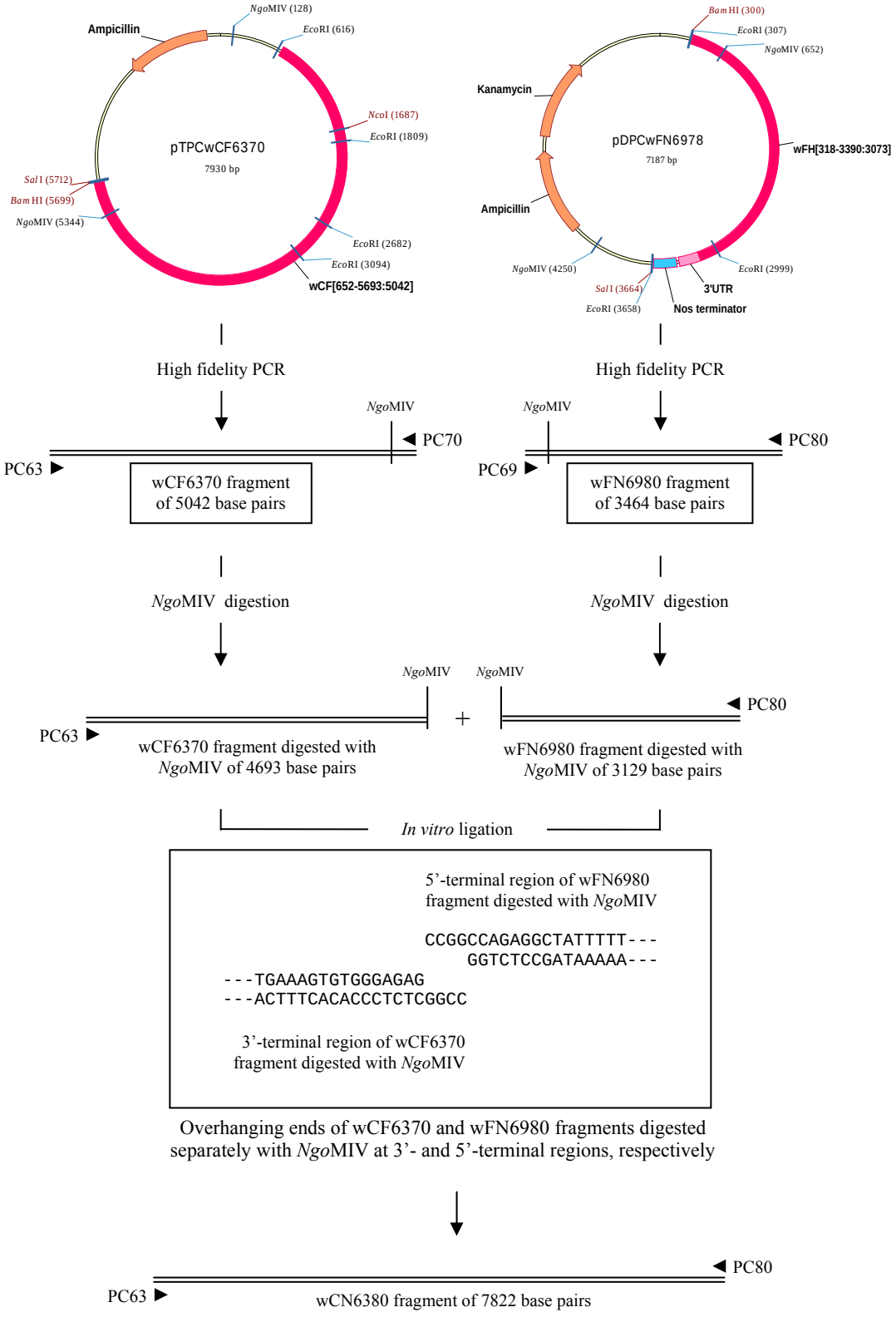


Figure 75 Outline of construction of pTPCwDN6578

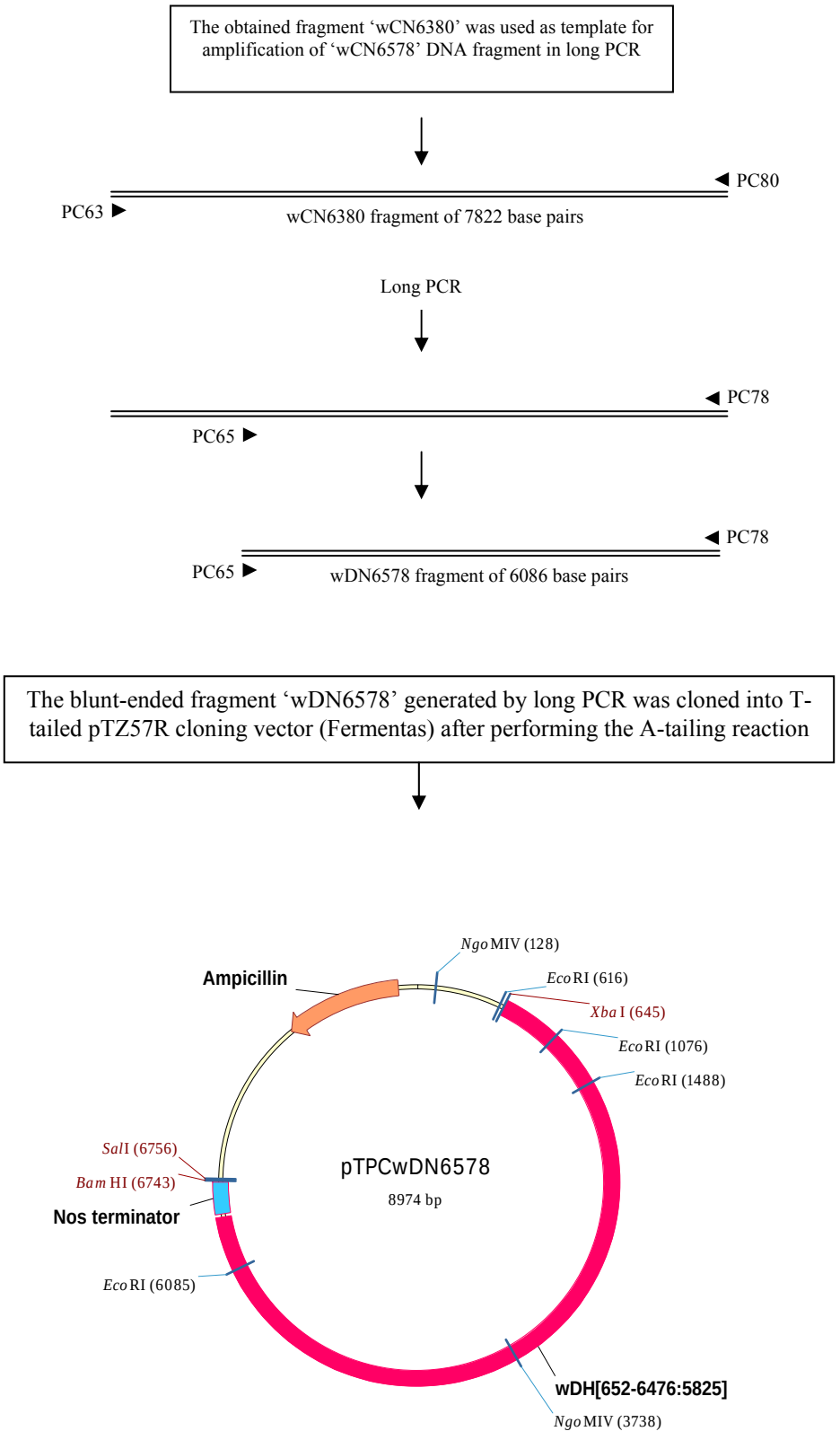


Figure 75 (continued)

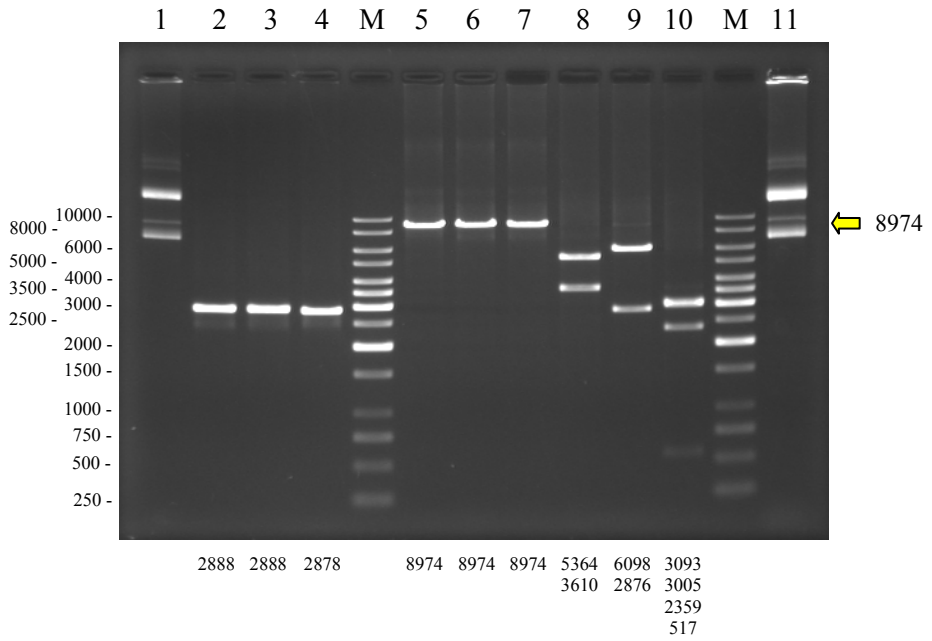


Figure 76 Restriction analyses of the pTPCwDN6578 plasmid. The pTPCwDN6578 plasmid of 8974 base pairs in closed circular form obtained from inserting the DNA fragment 'wDN6578' of 6086 base pairs into T-tailed pTZ57R cloning vector (Fermentas) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pTPCwDN6578 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pTPCwDN6578 (circular form, 0.5 µl)
Lane 2: pTZ57R vector self-ligating digested with *Xba*I
Lane 3: pTZ57R vector self-ligating digested with *Ngo*MIV
Lane 4: pTZ57R vector self-ligating digested with *Bam*HI and *Xba*I
Lane 5: pTPCwDN6578 digested with *Bam*HI
Lane 6: pTPCwDN6578 digested with *Sal*I
Lane 7: pTPCwDN6578 digested with *Xba*I
Lane 8: pTPCwDN6578 digested with *Ngo*MIV
Lane 9: pTPCwDN6578 digested with *Bam*HI and *Xba*I
Lane 10: pTPCwDN6578 digested with *Bam*HI, *Ngo*MIV and *Xba*I
Lane 11: pTPCwDN6578 (circular form, 1 µl)

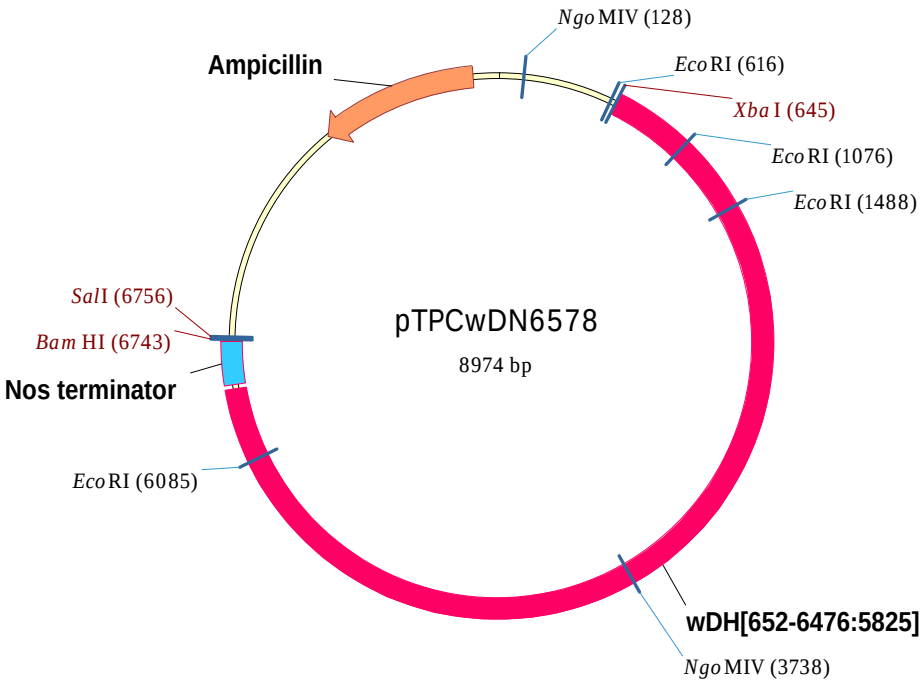


Figure 77 Map of pTPCwDN6578. The pTPCwDN6578 plasmid of 8974 base pairs in closed circular form was obtained from inserting the DNA fragment 'wDN6578' of 6086 base pairs into T-tailed pTZ57R cloning vector (Fermentas). The insert fragment 'wDN6578' is depicted by shaded boxes. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the ampicillin coding region on the cloning vector is represented by a shaded arrowhead.

recombinant plasmids constructed in this research consisting pDPCwPD7940 (Figure 58), pTPCwCF6370 (Figure 67) and pTPCwDN6578 (Figure 77). The plasmid pDPCwPN7978 containing the full-length cDNA clone of the PRSV-W with the CaMV 35S promoter and a NOS terminator was constructed as follows.

Initial step for constructing pDPCwPN7978 plasmid, an overlap extension PCR for assembly of two linear DNA fragments between 'wCF6328' and 'wFN2781' fragments was performed to generate a linear DNA construct, named 'wCN6381', which contains DNA sequences assembling from the two DNA fragments. For amplification of 'wCF6328' fragment of 4073 base pairs, long PCR was performed with pTPCwCF6370 plasmid as template using primers PC63 and PC28 for DNA polymerization priming. Another fragment 'wFN2781' of 3931 base pairs was amplified from pTPCwDN6578 plasmid using primers PC27 and PC81 in a long PCR. The two fragments, 'wCF6328' and 'wFN2781', which share a 215-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). To assemble the two linear DNA fragments, overlap extension PCR was performed with 'wCF6328' and 'wFN2781' fragments as overlapping DNA templates using primers PC63 and PC81. Approximately 7.8-kilobase pairs DNA fragment, named 'wCN6381', generated from the overlap extension PCR was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as template for amplification of 'wCN2381' fragment. The 'wCN2381' fragment was amplified from 'wCF6381' fragment as template using primers PC23 and PC81 in long PCR.

The blunt-ended 'wCN2381' fragment of 7405 base pairs in size obtained from the long PCR was digested simultaneously with two restriction endonucleases *Bam*HI and *Nco*I. Approximately 6.7-kilobase pairs DNA fragments from the *Bam*HI and *Nco*I digestion reaction of 'wCN2381' fragment, was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as insert fragment for an *in vitro* sticky end ligation with an approximately 8.3 kilobase pairs DNA fragment obtained from double-digestion of pDPCwPD7940 with the same

enzymes (*Bam*HI and *Nco*I) using with the 'wCN2381' fragment. To construct the pDPCwPN7978 recombinant plasmid from the two fragments, *in vitro* ligation of DNA was performed with T4 DNA ligase (Fermentas) using the 6.7- and 8.3-kilobase pairs DNA fragments of the *Bam*HI and *Nco*I digested 'wCN2381' and the *Bam*HI and *Nco*I digested pDPCwPD7940 plasmid, respectively as substrates, and the reaction was incubated at 10 ^{o}C for 16 hr. After incubation period of the *in vitro* ligation reaction, transformation of competent *E. coli* by electroporation was performed with the mixture of the ligation reaction using *E. coli* strain XL1-Blue as bacterial host cells to clone the recombinant plasmid pDPCwPN7978. The strategy for the construction of the recombinant plasmid pDPCwPN7978 which contains the insert fragment of the full-length cDNA clone of a Thai isolate of PRSV-W with the CaMV 35S promoter and a NOS terminator is outlined in Figure 78. Orientation of the insert fragment 'wPN7978' of 11116 base pairs in the recombinant plasmid pDPCwPN7978 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC64 and PC75 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwPN7978 plasmid as substrate by digesting with five restriction endonucleases including *Bam*HI, *Nco*I, *Ngo*MIV, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 79. Proposed circular DNA map of pDPCwPN7978 plasmid constructing as described above is shown in Figure 80. The pDPCwPN7978 plasmid was used directly for infectivity assay as described in section 9.1.

8.5 Construction of a full-length cDNA clone of a Thai isolate of PRSV-W for expressing viral RNA through *in vitro* transcription

Another recombinant plasmid, named pDPCwAN0178, was constructed by inserting a full-length cDNA with nucleotide sequence representing the genomic RNA of a Thai isolate of PRSV-W following a poly(A) 24 tract assembled with a NOS terminator fragment of 10584 base pairs, named 'wAN0178', into U-tailed pDrive cloning vector (QIAGEN). In the insert fragment, the full-length cDNA corresponding to the PRSV-W genome was assembled with a NOS terminator cDNA

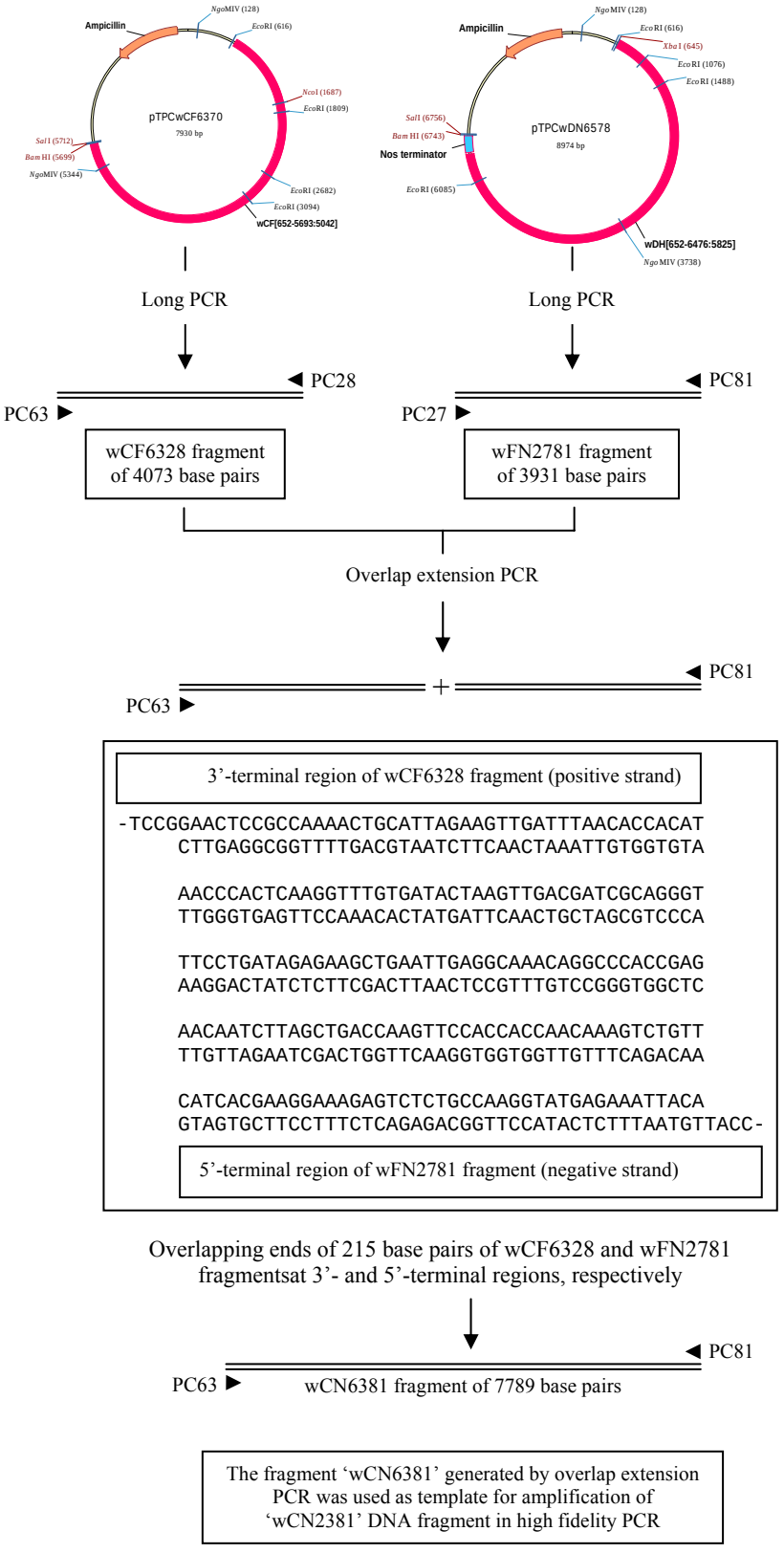


Figure 78 Outline of construction of pDPCwPN7978

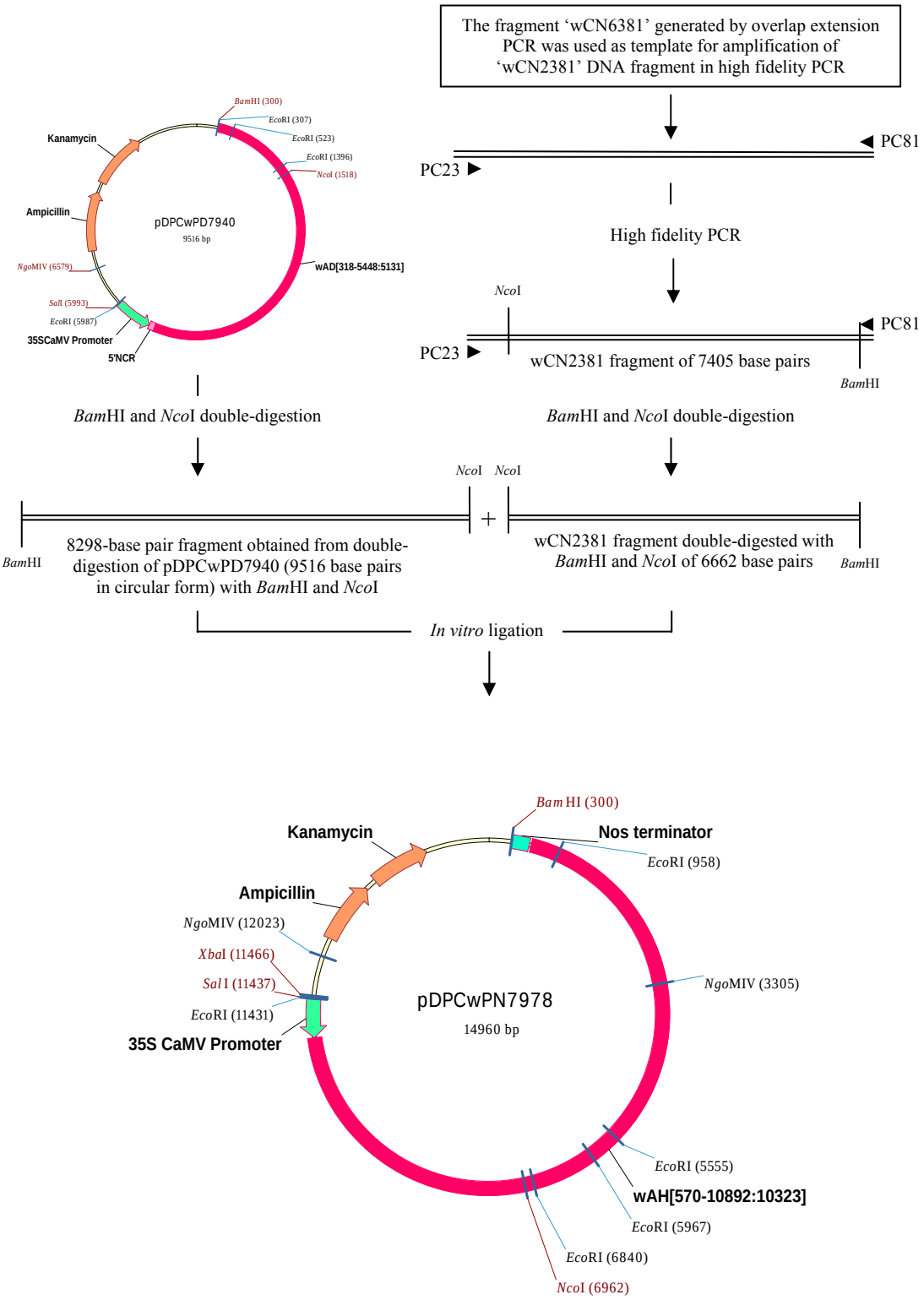


Figure 78 (continued)

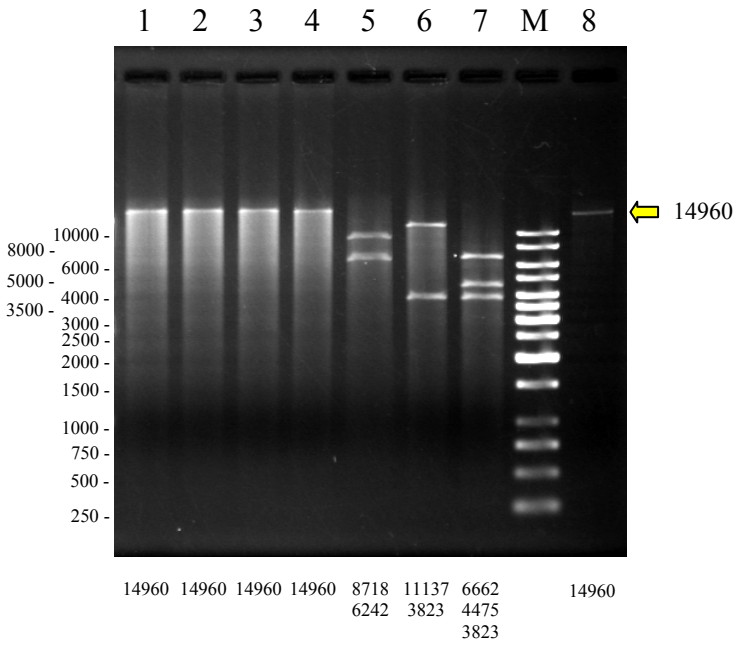


Figure 79 Restriction analyses of the pDPCwPN7978 plasmid. The pDPCwPN7978 plasmid of 14960 base pairs in closed circular form obtained from inserting the DNA fragment 'wPN7978' of 11116 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwPN7978 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane 1: pDPCwPN7978 digested with *Bam*HI
Lane 2: pDPCwPN7978 digested with *Sal*I
Lane 3: pDPCwPN7978 digested with *Xba*I
Lane 4: pDPCwPN7978 digested with *Nco*I
Lane 5: pDPCwPN7978 digested with *Ngo*MIV
Lane 6: pDPCwPN7978 digested with *Bam*HI and *Sal*I
Lane 7: pDPCwPN7978 digested with *Bam*HI, *Nco*I and *Sal*I
Lane 8: pDPCwPN7978 (circular form, 0.25 µl)

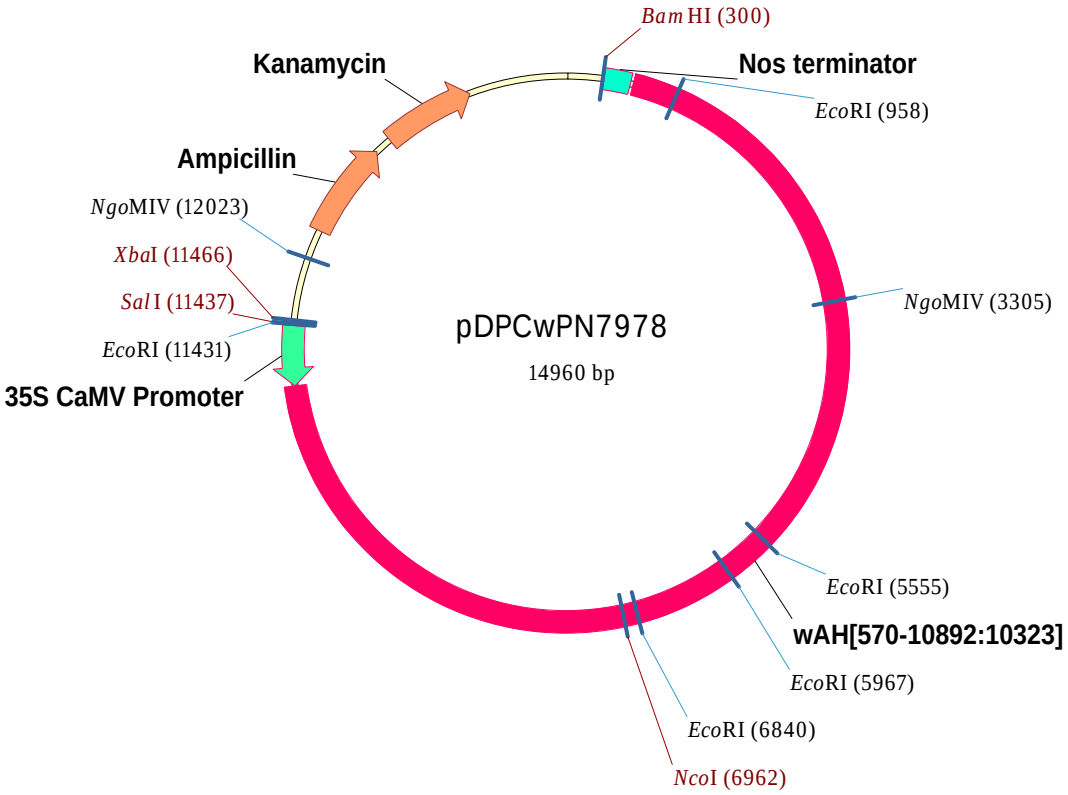


Figure 80 Map of pDPCwPN7978. The pDPCwPN7978 plasmid of 14960 base pairs in closed circular form was obtained from inserting the DNA fragment 'wPN7978' of 11116 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wPN7978' is depicted by a shaded arrowhead for 35S CaMV Promoter region, and shaded boxes for fragment 'wAH' and Nos terminator regions. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

sequences 3'-termini. The recombinant plasmid pDPCwAN0178 was developed from three recombinant plasmids constructed in this research consisting pDPCwAD0140 (Figure 61), pTPCwCF6370 (Figure 67) and pTPCwDN6578 (Figure 77). The plasmid pDPCwAN0178 containing the full-length cDNA clone of the PRSV-W with a NOS terminator was constructed as follows.

Like the initial step for construction of pDPCwPN7978 plasmid, the approximately 7.8-kilobase pairs DNA fragment 'wCN2381' was amplified from 'wCF6328' fragment which was developed from two recombinant plasmids pTPCwCF6370 and pTPCwDN6578 as described in section 8.4. After the 'wCN2381' linear recombinant fragment amplified from the 'wCN6381' which was generated from the assembling of the 'wCF6328' and 'wFN2781'fragments was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) and then the blunt-ended 'wCN2381' fragment of 7405 base pairs in size amplified from the long PCR was digested simultaneously with two restriction endonucleases *Nco*I and *Sal*I.

Approximately 6.7-kilobase pairs DNA fragments from the *Nco*I and *Sal*I digestion reaction of 'wCN2381' fragment, was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN) before using as insert fragment for an *in vitro* sticky end ligation with an approximately 8.3 kilobase pairs DNA fragment obtained from double-digestion of pDPCwAD0140 with the same enzymes (*Nco*I and *Sal*I) using with the 'wCN2381' fragment. To construct the pDPCwAN0178 recombinant plasmid from the two fragments, *in vitro* ligation of DNA was performed with T4 DNA ligase (Fermentas) using the 6.7- and 8.3-kilobase pairs DNA fragments of the *Bam*HI and *Nco*I digested 'wCN2381' and the *Bam*HI and *Nco*I digested pDPCwPD7940 plasmid, respectively as substrates, and the reaction was incubated at 10 ^{o}C for 16 hr. After incubation period of the *in vitro* ligation reaction, transformation of competent *E. coli* by electroporation was performed with the mixture of the ligation reaction using *E. coli* strain XL1-Blue as bacterial host cells to clone the recombinant plasmid pDPCwAN0178. The strategy for the construction of the recombinant plasmid pDPCwAN0178 which contains the

insert fragment of the full-length cDNA clone of a Thai isolate of PRSV-W with a NOS terminator is outlined in Figure 81. Orientation of the insert fragment 'wAN0178' of 10584 base pairs in the recombinant plasmid pDPCwAN0178 was determined in routine PCR using vector-specific primers, PC80 (M13 forward (-20)) and PC81 (M13 reverse), and insert specific primers including PC36 and PC75 (data not shown). Restriction analyses were performed with the close circular construct of pDPCwAN0178 plasmid as substrate by digesting with five restriction endonucleases including *Bam*HI, *Nco*I, *Ngo*MIV, *Sal*I and *Xba*I. Results of these analyses are shown by agarose gel electrophoresis with ethidium bromide staining in Figure 82. Proposed circular DNA map of pDPCwAN0178 plasmid constructing as described above is shown in Figure 83. The pDPCwAN0178 plasmid was used as DNA template in a transcription *in vitro* reaction for synthesizing RNA transcripts of PRSV-W used in infectivity assay as described in section 9.2.

Two full-length cDNA clones of PRSV-W were constructed separately in a commercially available plasmid U-tailed pDrive (QIAGEN). These were assembled together with or without CaMV 35S promoter at the 5'-end of the PRSV-W cDNA sequence, whereas both clones gather together for a NOS terminator at their 3'-ends which are constructed downstream from a poly$(A)_{24}$ tract. The full- length cDNA constructs were developed from the eight overlapping cDNA clones of a Thai isolate of PRSV-W which was used as source for all experiments in this research. The sources of sequences for full-length cDNA constructions described above are pPCwA, pPCwB, pPCwC, pPCwD, pPCwE, pPCwF, pPCwG and pPCwH in the order in which inserted fragments correspond on the virus genome as shown in Figure 1.

The fusions of the CaMV 35S promoter and NOS terminator to the 5'- and 3'-ends of PRSV-W cDNA , respectively, were made separately through overlap extension PCR amplifications of two pairs of different cDNA inserts which were cloned in separate plasmids pD35SPro and pPCwA, and pPCwH and pDNosTer. Two pairs of chimeric oligonucleotide primers, pair of 5U-35S-2 (reverse priming DNA polymerization on CaMV 35S promoter sequence at the 3'-terminal region which cloned in plasmid pD35SPro) and PC77 (forward priming DNA polymerization

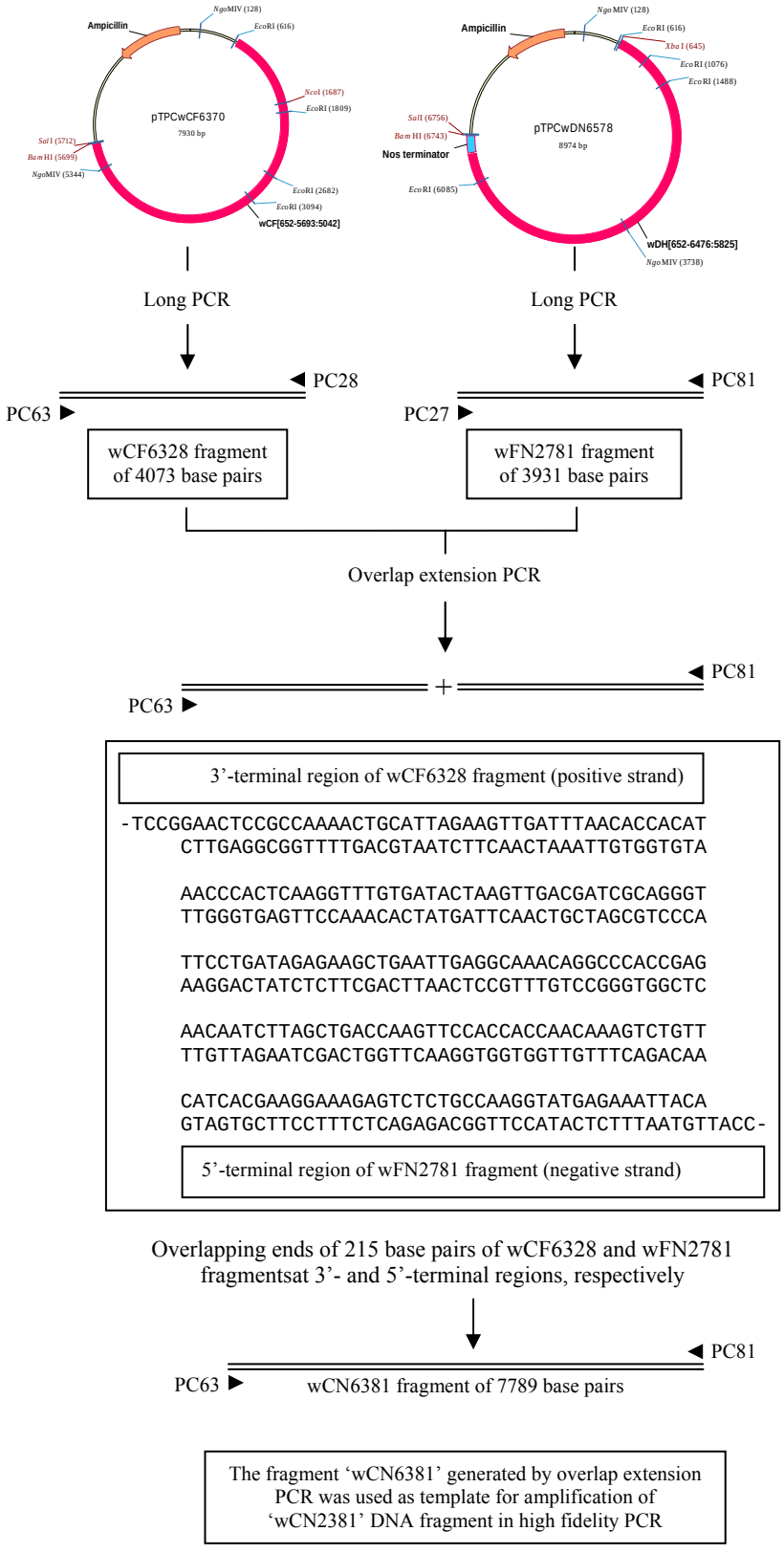
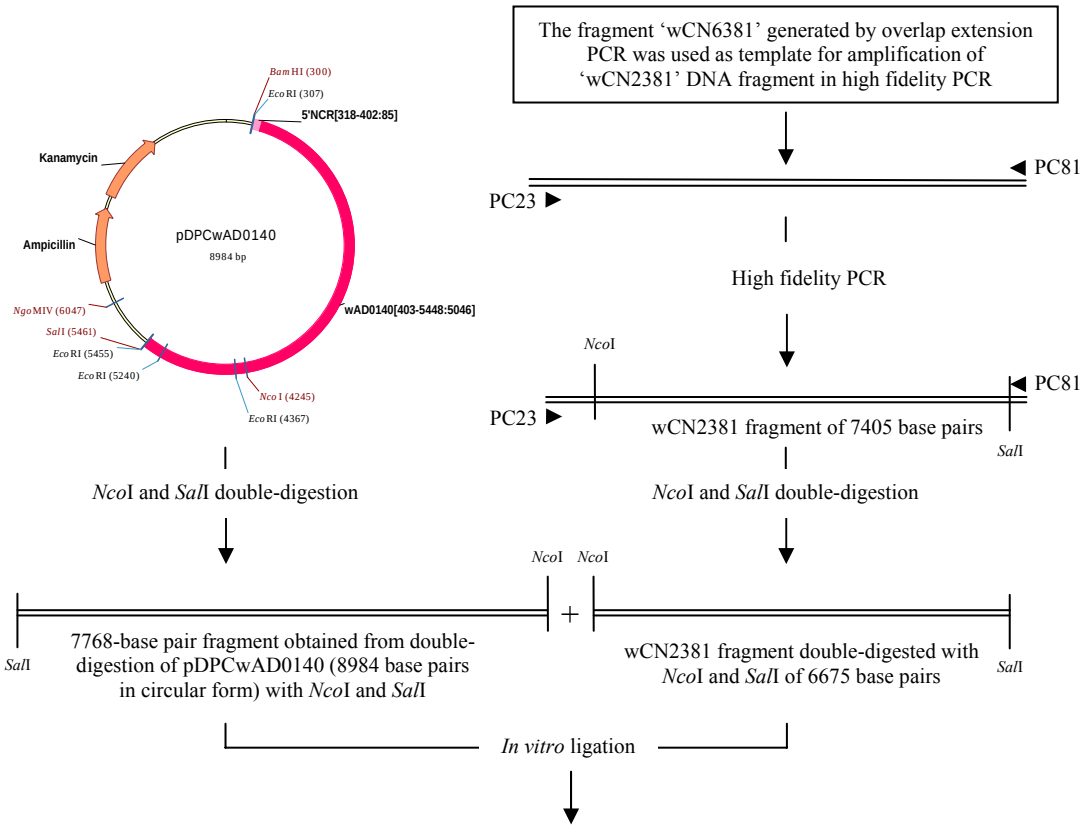


Figure 81 Outline of construction of pDPCwAN0178



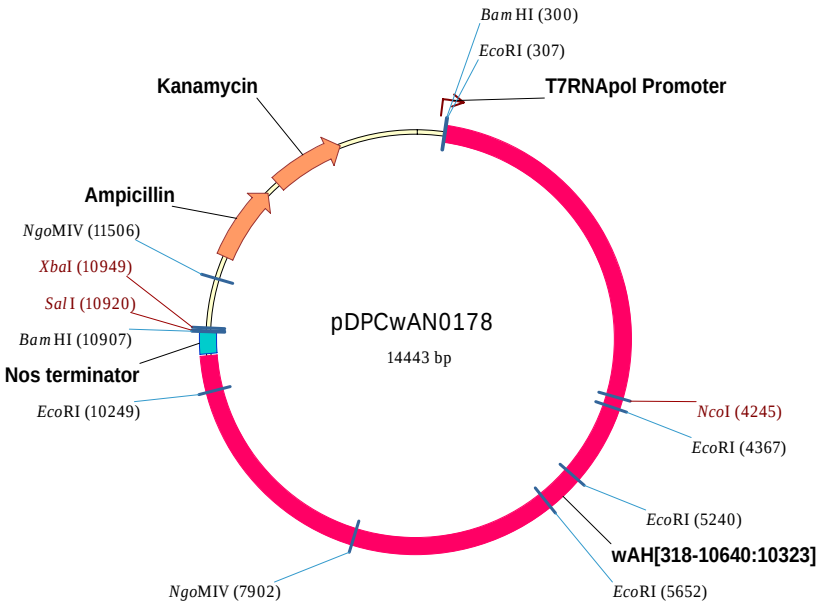


Figure 81 (continued)

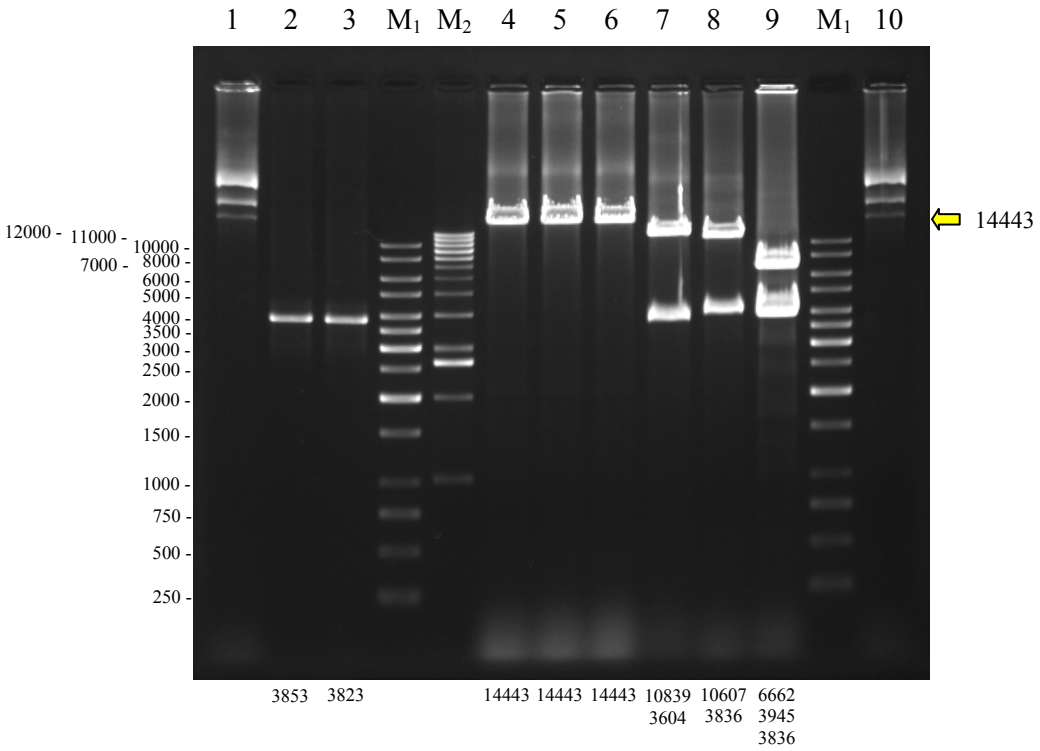


Figure 82 Restriction analyses of the pDPCwAN0178 plasmid. The pDPCwAN0178 plasmid of 14443 base pairs in closed circular form obtained from inserting the DNA fragment 'wAN0178' of 10584 base pairs into U-tailed pDrive cloning vector (QIAGEN) was digested separately in each condition. The results of reactions were determined and analyzed by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the left of figure. Set of DNA fragment sizes (in base pairs) after digestion for each reaction is shown below each lane. The full-length DNA fragment of the linearized pDPCwAN0178 plasmid is depicted by an arrow with size (in base pairs) at the right of the figure.

Lane M_1: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane M_2: 1kb DNA Ladder
Lane 1: pDPCwAN0178 (circular form, 0.5 µl)
Lane 2: pDrive vector self-ligating digested with *Bam*HI
Lane 3: pDrive vector self-ligating digested with *Bam*HI and *Sal*I
Lane 4: pDPCwAN0178 digested with *Sal*I
Lane 5: pDPCwAN0178 digested with *Xba*I
Lane 6: pDPCwAN0178 digested with *Nco*I
Lane 7: pDPCwAN0178 digested with *Ngo*MIV
Lane 8: pDPCwAN0178 digested with *Bam*HI
Lane 9: pDPCwAN0178 digested with *Bam*HI and *Nco*I
Lane 10: pDPCwAN0178 (circular form, 0.25 µl)

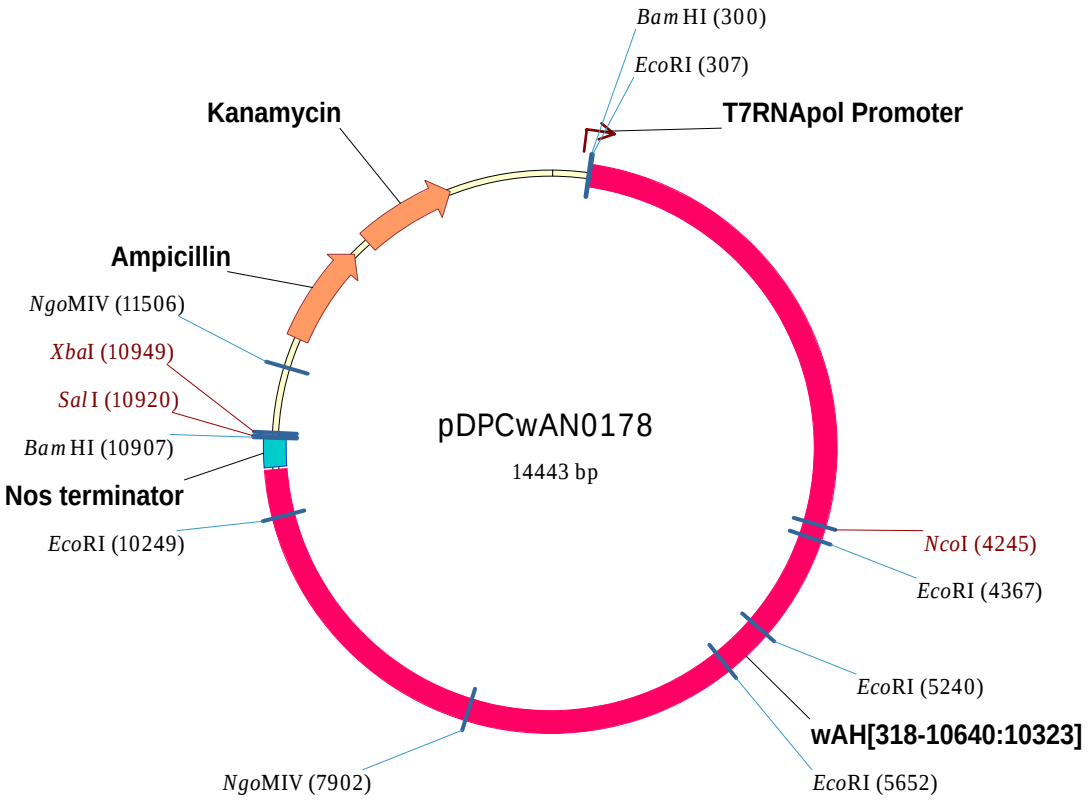


Figure 83 Map of pDPCwAN0178. The pDPCwAN0178 plasmid of 14443 base pairs in closed circular form was obtained from inserting the DNA fragment 'wAN0178' of 10584 base pairs into U-tailed pDrive cloning vector (QIAGEN). The insert fragment 'wAN0178' is depicted by shaded boxes for fragment 'wAH' and Nos terminator regions. Position range on map and size in base pairs of the insert fragment are shown in parentheses [ ]. Restriction sites for digestions with restriction endonucleases (*Bam*HI, *Eco*RI, *Nco*I, *Ngo*MIV and *Sal*I) using for analyses of the plasmid are indicated on the map circle and each cleavage sites is shown in parentheses ( ). Double strand DNA of the pDrive cloning vector is depicted by double line, and the two antibiotic (ampicillin and kanamycin) coding regions on the cloning vector are represented by shaded arrowheads.

on ‘wA’ DNA inserted fragment at the 5’-terminal region which cloned in plasmid pPCwA) and another pair of PC74 (reverse priming DNA polymerization on ‘wH’ DNA inserted fragment at the 3’-terminal region which cloned in plasmid pPCwH) and A Nos-1 (forward priming DNA polymerization on NOS terminator sequence at the 5’-terminal region which cloned in plasmid pDNosTer) were designed and synthesized to use separately for amplifying overlapping DNA fragments. These four chimeric primers contained long 5’ additional sequences, which resulted in introduction of sequence homologous to the end of a prospective adjacent DNA fragment. The primer 5U-35S-2 (reverse) and PC77 (forward) were used separately as one of a pair of primers with primer 35S-1H (forward) and PC60 (reverse), respectively, for generating 24-base pairs overlapping sequences between CaMV 35S promoter and ‘wA7760’ fragments at 3’- and 5’-terminal regions, respectively. The primer PC74 (reverse) and A Nos-1 (forward), another of primer pair, were used separately as one of a pair of primers with primer PC75 (forward) and Nos-4H (reverse), respectively, for generating 17-base pairs overlapping sequences between ‘wH7574’ and ‘Nos terminator’ fragments at 3’- and 5’-terminal regions, respectively. Plasmids pDPCwPA7960 and pDPCwHN7578 resulting from separate cloning of overlap extension products, ‘wPA7960’ and ‘wHN7578’ fragments, in pDrive cloning vectors (QIAGEN) were used separately as templates for sequencing of the fusion points between CaMV 35S promoter and ‘wA7760’ DNA fragment, and between ‘wH7574’ DNA fragment and ‘Nos terminator’, respectively. Sequence analyses of the two fusion points between CaMV 35S promoter and PRSV-W genomic cDNA region at 3’- and 5’-ends, respectively, and between PRSV-W genomic cDNA region and NOS terminator at 3’- and 5’-ends, respectively, were verified and sequences of these regions clearly showed the presence of precise and complete overlapping regions which contain sequences identical to the sequences of the chimeric primers as designed and described above.

A 24-nucleotides poly (A) tail was generated to the complete cDNA of PRSV-W at 3’ end using high fidelity PCR amplification with one of primer pair, chimeric primer PC74 containing 24 nucleotides of poly (A) sequence, which was used as a

reverse primer for amplifying 3' terminal region of PRSV-W cDNA such as 'wH7574' DNA fragment of 1135 base pairs.

Generally, it is known that amplification of DNA fragments larger than one kilobase pairs with *Taq* DNA polymerase is problematic. Absence of proofreading capacity lets misincorporations, which decreases the efficiency of amplification. *In vitro* amplification of DNA fragments using PCR with a proofreading DNA polymerase takes misincorporated nucleotides away (Cheng *et al*., 1994). However, the yield of amplification product typically reduces with expanding DNA length. The combination of *Taq* DNA polymerase for high processivity (5'→3' polymerase activity) and a small amount of some different DNA polymerases that evidence a characteristic 3'→5' exonuclease activity (usually *Pfu* or *Pwo* proofreading DNA polymerase) has been demonstrated to successfully amplify DNA fragments (PCR product), which are much larger than those achieved with conventional *Taq* DNA polymerase alone.

In this study, long PCR was performed for *in vitro* amplification of DNA fragments with longer than 3 kilobase pairs using a DNA polymerase mix which was prepared by combining 20 units of *Taq* (Fermentas) or *Tth* (Promega) DNA polymerase and one unit of *DeepVent* DNA polymerase (New England Biolabs INC). The DNA polymerase mix resulted in optimal PCR amplification results. Using the long PCR protocol, DNA fragments (PCR products) up to 7.8 kilobase pairs could be amplified from linear recombinant DNA template which was generated from an overlap extension PCR. Amplification of the 7.8 kilobase pairs DNA fragment with no apparent nonspecific amplification products (data not shown) suggests that the procedure was specific for the target DNA.

High fidelity PCR amplification was performed as the first step in preparing all linear DNA fragments (used as DNA templates in overlap extension PCR) with shorter than 3 kilobase pairs. This high fidelity PCR method performed in this research was based on *DeepVent* DNA polymerase, which is a high-fidelity

thermophillic DNA polymerase and optimized to produce high fidelity PCR products with blunt ends.

The problem with *Taq* DNA polymerase is that it leaves single A-overhangs at 3' ends, which disrupt priming of the overlapping regions during the overlap extension reaction (Sambrook and Russel, 2001). Therefore, the initial DNA fragments amplified from PCR have to be generated with a DNA polymerase enzyme that has 3'→5' proofreading exonuclease activity, and the obtained PCR products have both blunted ends. An alternative to DNA fragment amplifying with *Taq* DNA polymerase is to flush the ends of PCR products using T4 polymerase treatment (Warrens *et al*., 1997).

Overlap extension PCR can be applied to fuse necessary DNA fragments exactly and rapidly. This method is due largely to constraints resulting from the absence of adjacent restriction sites at joints. The method for splicing two short DNA fragments, overlap extension PCR, was described by Yolov and Shabarova (1990) can be employed to fuse necessary fragments precisely and quickly (Yon and Fried, 1989). This method, however, is limited to regular PCR which its amplified products of approximately 3 to 4 kilobase pairs in length and to assembly of no more than two fragments of DNA at a time (Pont-Kingdon, 1997; Kawayama *et al*., 2002).

The assembly of linear recombinant cDNA constructs was performed by using an overlap extension PCR method to fuse linear DNA fragments. A procedure of overlap extension PCR performed to assemble linear DNA fragments in this study was modified and developed by combination of long PCR and overlap extension PCR. The modified method is based on *DeepVent* polymerase mix, which has a 3'→5' proofreading exonuclease activity. The usefulness of a mixture of either *Taq* or *Tth* non-proofreading DNA polymerase and the proofreading *DeepVent* DNA polymerase was shown and performed to simultaneously splice up to four fragments, rather than two, with high yield. Additionally, it can be used to splice and/or amplify long PCR products with longer than 4 kilobase pairs.

In this study, ten different linear DNA recombinants ranging from 1.4 to 7.8 kilobase pairs were successfully assembled using the modified overlap extension PCR method. All were satisfyingly cloned into plasmid DNA cloning vectors. Cloning system performed in this study was used with plasmid vectors either pDrive (QIAGEN) or pTZ57R (Fermentas) as cloning vectors which both have a characteristic of high-copy-number replication in compatible host cells (*E. coli*). Three different *E. coli* strains including DH5α, JM109 and XL1-Blue were prepared separately the competent cells and used as host cells for propagation of plasmids and cDNA clones. All recombinant DNA plasmids obtained in this study were transformed separately into *E. coli* host cells performing in two manners either electroporation or a calcium chloride (heat shock) which depended on the expected size of the recombinant plasmid vector. Simultaneously, blue/white colony screening and antibiotic resistance selection were two methods which were used for selecting transformants of bacterial cells which contain these recombinant plasmids. Both routine PCR and restriction analysis were performed for verification and confirmation of the presence and orientation of the insert fragments in each transformants.

The construction of full-length PRSV-W cDNA clones under the control of the enhanced CaMV 35S or T7 promoters is reported in this study. The recombinant plasmid pDPCwPN7978 contained a construct in the arrangement of the CaMV 35S promoter, the complete PRSV-W cDNA, a 24 poly(A) tract and a NOS polyadenylation signal as a terminator (Figure 80). The clone pDPCwAN0178 contained the full-length cDNA sequence of PRSV-W (without CaMV 35S promoter at the 5' end of the viral cDNA sequence) and a poly(A) tail of 24 nucleotides at the 3' end followed by a NOS terminator (Figure 83). Both full-length PRSV-W cDNA clones, pDPCwPN7978 and pDPCwAN0178, were constructed separately in pDrive cloning vector (Fermentas) as a backbone vector. The pDPCwPN7978 and pDPCwAN0178 of 14960- and 14443-nucleotides in sizes, respectively, were designed and constructed for used as viral DNA templates for *in vivo* and *in vitro* transcription reactions, respectively. Clones pDPCwPN7978 and pDPCwAN0178 were unstable and toxic in *E. coli* resulting in amplification of plasmids with deletions. The both constructs of full-length PRSV-W cDNA clones, pDPCwPN7978

and pDPCwAN0178, which separately cloned in a high copy number plasmid (pDrive, QIAGEN) cloning vector could not be propagated in *E. coli* generally used as bacterial host cells. Whereas the other partial recombinant cDNA clones of ranging from 5.2 to 9.5 kilobase pairs generated in this construction process were obtained in the bacterial host cells (*E. coli*). Instability of full-length cDNA clones in bacteria have been reported for several animal and plant viruses, possibly due to the potential toxicity and instability of viral sequences in bacteria (Boyer and Haenni, 1994). In some cases of potyviruses, the construction of full-length clones of PStV and TuMV in *E. coli* strain DH5α have been reported (Flasinski *et al*., 1996; Sanchez *et al*., 1998). Instability of a full-length clone of PVY in *E. coli* has also been unsuccessful, and the reasons for the instability have been attributed to potential expression of toxic product (Fakhfakh *et al*., 1996).

Full-length cDNA of several potyviruses have been obtained and cloned in *E. coli* under the control of either a CaMV 35S promoter or a bacteriophage promoter (Boyer and Haenni, 1994; Gal-On *et al*., 1995; Flasinski *et al*., 1996; Johansen, 1996; Takahashi *et al*., 1997). Nevertheless, molecular cloning of full-length viral cDNAs is usually problematic due to the unexpected toxicity of these constructions to the bacterial cells (Boyer and Haenni, 1994; Fakhfakh *et al*., 1996; Johansen, 1996). Plasmid toxicity is not altogether understood but may cause from vector-specific besides insert-specific occasions (Boyer and Haenni, 1994; Lai, 2000; Yamshchikov *et al*., 2001). Accordingly, a key parameter in full-length cDNA clone construction is the plasmid backbone. Additionally, it is assumed that there are mechanisms which are linked to the expression of viral proteins in the bacterial hosts from RNA transcribed from cryptic promoters (Rodriguez-Cerezo and Shaw, 1991; Fakhfakh *et al*., 1996). These concepts are supported by the phenomenon that insertion of introns into the genomic cDNA of PStMV and PPV strongly attenuates the instability problems (Johansen, 1996; Lopez-Moya and Garcia, 2000).

Currently, the mechanisms controlling the stability of such a clone are not completely understood. Success in some cases might be due to the application of divergent combinations of the cloning vector, bacterial host and incubation

temperature. However the effects of these strategies of the stabilization of full-length virus cDNA plasmids during their propagation in *E. coli* have never been studied so far. Attempts to propagate the two full-length PRSV-W cDNA clones, pDPCwPN7978 and pDPCwAN0178, in *E. coli* were performed by separate incubating the two different bacterial transformants at reduced temperature (23-25 ^{o}C) and selecting with low concentration of ampicillin (25 µg/ml) without kanamycin. But these recombinant plasmids could also not be propagated in *E. coli*. This strategy was unsuccessful to multiply both full-length PRSV-W cDNA plasmids in this study. Whereas this strategy was used successfully to propagate a full-length *Dengue 2 virus* recombinant plasmid constructed using high copy number plasmid in *E. coli* (Sriburi *et al*., 2001). In case of TuMV, the instability detected in full-length clones has been overcome by growing culture containing 100 µg/ml of ampicillin at 30 ºC (Sanchez *et al*., 1998).

The procedures utilized for construction of clones containing a full-length PRSV-W cDNA under control of either the CaMV 35S promoter (pDPCwPN7978) or the T7 polymerase promoter (pDPCwAN0178) are described in detail in section 8 and the maps of these constructs are illustrated in Figures 80 and 83, respectively. Although both full-length constructs (pDPCwPN7978 and pDPCwAN0178) were able to successfully clone into U-tailed pDrive cloning vector (QIAGEN) which has a characteristic of high copy number plasmid, but these obtained recombinant plasmids of the full-length constructs could not be propagated in *E. coli*. Clones pDPCwPN7978 and pDPCwAN0178 were unstable and toxic in *E. coli* resulting in amplification of plasmids with deletions.

Due to the instability of the two different full-length PRSV-W cDNA plasmids (pDPCwPN7978 and pDPCwAN0178 which were designed for viral infectious clones) during propagation in *E. coli*, overlap extension PCR was used as an alternative procedure for constructing the two linear full-length PRSV-W cDNA constructs which were controlled separately with the CaMV 35S promoter or the T7 polymerase promoter. This strategy was performed to construct linear full-length cDNAs of PRSV-W in order to keep away from the instability problem. Alternately,

the overlap extension PCR products, named PCwPN7978 and PCwAG8584, could be obtained and might be used separately as DNA templates for *in vivo* and *in vitro* transcription reactions, respectively, in infectivity assays. Construction of the linear recombinant constructs, PCwPN7978 and PCwAG8584, which contains the full-length cDNA clone of a Thai isolate of PRSV-W with the CaMV 35S and T7 polymerase promoters are described in section 8.6 and section 8.7, respectively.

8.6 Construction of a linear full-length cDNA of PRSV-W containing of the CaMV 35S promoter and NOS terminator (PCwPN7978)

To generate the linear cDNA construct PCwPN7978 placing a copy cDNA of the full-length of the PRSV-W genome with a poly$(A)_{24}$ tract (at the 3'end) between a CaMV 35S promoter and a NOS terminator, two overlapping cDNA fragments (named PCwPD8040 and PCwDN2581) were amplified separately using pDPCwPD7940 and pTPCwDN6578 plasmids as DNA templates, respectively, in separate long PCRs. Following is a protocol for assembling the 11.1-kilobase pairs 'PCwPN7978' cDNA fragment from two fragments (PCwPD8040 and PCwDN2581). Two different recombinant plasmids constructed in this research, pDPCwPD7940 (Figure 58) and pTPCwDN6578 (Figure 77), were used separately as templates for generating two overlapping cDNA fragments consisting 'PCwPD8040' and 'PCwDN2581' fragments, respectively, by long PCR. Long PCR was performed with pPCwPD7940 plasmid as template using primers PC80 and PC40 for DNA amplification of 'PCwPD8040' fragment of 5793 base pairs (Figure 84, Lane 3). Fragment 'PCwDN2581' of 5859 base pairs (Figure 84, Lane 6) was amplified from pPCwDN6578 plasmid using primers PC25 and PC81 in long PCR. The two fragments, 'PCwPD8040' and 'PCwDN2581', which share a 268-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). For assembly of the two linear DNA fragments 'PCwPD8040' and 'PCwDN2581', overlap extension PCR was performed with primers PC79 and PC78. Overlap extension PCR product of approximately 11.1 kilobase pairs, named 'PCwPN7978' (Figure 84, Lane 8), was

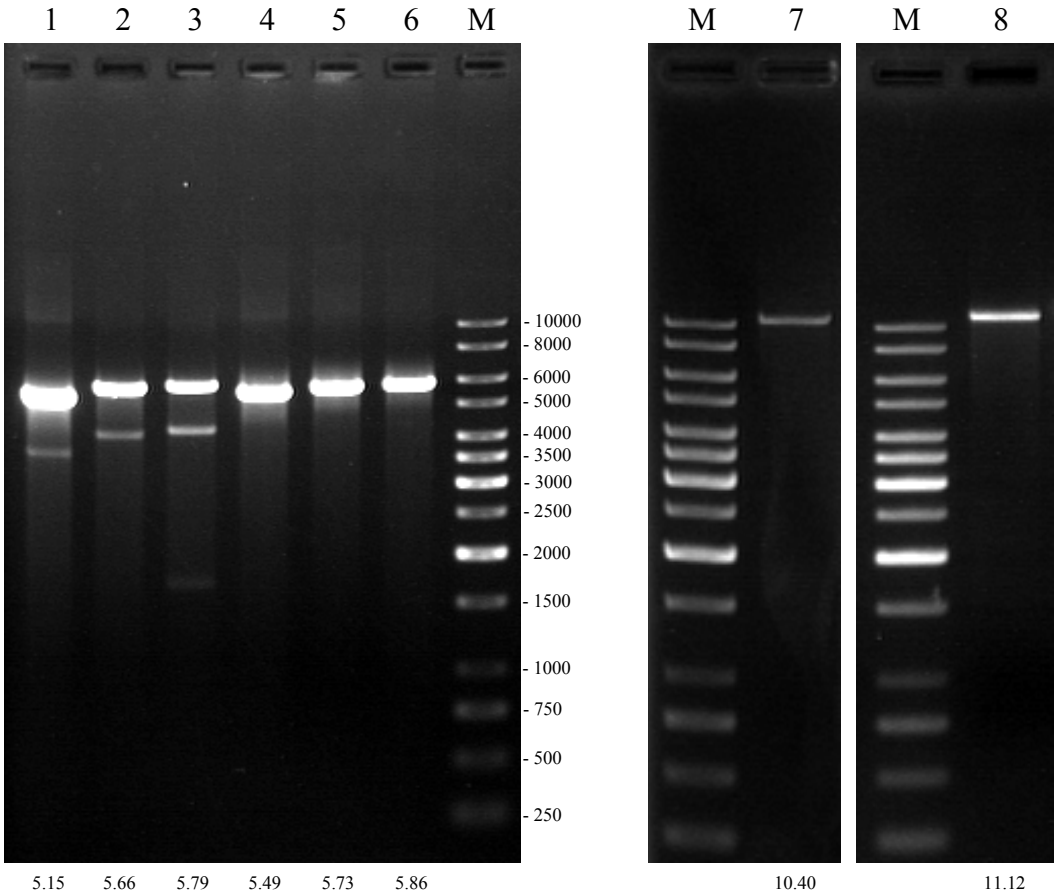


Figure 84 PCR amplifications of three pairs overlapping cDNA fragments used as templates for construction of the linear full-length cDNAs of PRSV-W. PCR products PCwAD8340 were amplified from pDPCwAD0140, while two PCR products PCwPD7940 and PCwPD8040 were amplified from pDPCwPD7940. And three PCR products PCwDG2584, PCwDN2578 and PCwDN2581 were amplified from pTPCwDN6578. Purified PCwAG and PCwPN7978 PCR products were amplified separately by overlap extension PCRs. The products of PCRs were determined by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA fragment of standard marker (in base pairs) is shown in the right of figure. Proposed DNA fragment size (in kilobase pairs) for each PCR product is shown below each lane.

Lane 1: PCwAD8340 PCR products
Lane 2: PCwPD7940 PCR products
Lane 3: PCwPD8040 PCR products
Lane 4: PCwDG2584 PCR products
Lane 5: PCwDN2578 PCR products
Lane 6: PCwDN2581 PCR products
Lane 7: PCwAG8584 PCR products
Lane 8: PCwPN7978 PCR products
Lane M: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)

gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). The linear cDNA construct PCwPN7978 was used as inoculum for inoculation of pumpkin cotyledons by particle bombardment.

8.7 Construction of a linear full-length cDNA of PRSV-W containing of the bacteriophage T7 promoter (PCwAG8584)

There are two reasons to generate the linear cDNA construct PCwAG8584 placing a copy of the full-length of the PRSV-W genome with a poly$(A)_{24}$ tract (at the 3'end) downstream from a bacteriophage T7 promoter by overlap extension PCR from two overlapping cDNA fragments (named PCwAD8540 and PCwDG2584) which were amplified separately from the two overlapping plasmids (pDPCwAD0140 and pTPCwDN6578) used as DNA templates, respectively, in separate long PCRs. Firstly, the full-length PRSV-W cDNA plasmid pDPCwAN0178 was unstable and toxic in *E. coli* resulting in amplification of plasmids with deletions. Secondly, for introducing a sequence of the T7 polymerase promoter into the 5'-end cDNA fragment of the PRSV-W genome (PCwAD8540) at the 5'-end (upstream the full-length cDNA of PRSV-W) with no unwanted sequences (non viral sequence) at the junction.

Following is a protocol for generating the linear cDNA fragment 'PCwAG8584' of approximately 10.4 kilobase pairs which was designed and constructed for used as DNA template in a transcription *in vitro* reaction for synthesizing RNA transcripts of PRSV-W. The linear cDNA construct PCwAG8584 containing a copy cDNA of the full-length of the PRSV-W genome with a poly$(A)_{24}$ tract (at the 3'end) positioned downstream from a T7 polymerase promoter was developed from two overlapping plasmids constructed in this research (pDPCwAD0140 and pTPCwDN6578).

The two different recombinant plasmids constructed, pDPCwAD0140 (Figure 61) and pTPCwDN6578 (Figure 77), were used separately as templates for generating two overlapping cDNA fragments consisting 'PCwAD8340' and 'PCwDG2584'

fragments, respectively, by long PCR. The bacteriophage T7 RNA polymerase promoter was fused to the first nucleotide of the full-length cDNA of the PRSV-W genome with a poly$(A)_{24}$ tract (at the 3'end) by long PCR amplification of PCwAD0140 cDNA insert (pDPCwAD0140 plasmid) using primers PC83 (5'-*TAATACGACTCACTATAG*AAATAAAACATCTCAACACAAC-3') and PC40. Primer PC83 (40 mer) contains the sequence corresponding to T7 promoter and the additional G in order to provide a good initiating residue (Bailey *et al*., 1983) (italics), and 22 nucleotides corresponding to the 5'nucleotide sequence of the PRSV-W genome. Long PCR was performed with pDPCwAD0140 plasmid as template using primers PC83 and PC40 for DNA amplification of 'PCwAD8340' fragment of 5149 base pairs (Figure 84, Lane 1). Fragment 'PCwDG2584' of 5493 base pairs (Figure 84, Lane 4) was amplified from pTPCwDN6578 plasmid using primers PC25 and PC84 (5'-TTAGGATCCTTTTTTTTTTTTTTTTTTTTTTTCTCTCATT-3') in long PCR. Primer PC84 (41 mer) contains the following sequences: a sequence of six nucleotides specific to *Bam*HI endonuclease (underlined), the sequence of 24 nucleotides (T) complementary to the poly$(A)_{24}$ tract and 8 nucleotides cmplementary to the 3' sequence of the PRSV-W cDNA. The two fragments, 'PCwAD8340' and 'PCwDG2584', which share a 268-base pairs overlapping cDNA region at 3'- and 5'-terminal regions, respectively, were used as templates in overlap extension PCR after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). For assembly of the two linear DNA fragments 'PCwAD8340' and 'PCwDG2584', overlap extension PCR was performed with primers PC85 and PC84. Primer PC85 (5'-AACAGCTATGACC ATGTCGAC*TAATACGACTCACTATAG*-3') contains a sequence of six nucleotides specific to *Sal*I endonuclease (underlined) and the sequence corresponding to T7 promoter and the additional G (italics). Overlap extension PCR product of approximately 10.4 kilobase pairs, named 'PCwAG8584' (Figure 84, Lane 7), was gel-purified with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN). The linear cDNA construct PCwAG8584 was used as DNA template in a transcription *in vitro* reaction for synthesizing RNA transcripts of PRSV-W. The mixture of synthesized RNA transcripts were used as inoculum for mechanical inoculation of greenhouse grown pumpkin plants in infectivity assay.

9. Infectivity assays

9.1 Infectivity assay of PRSV-W full-length cDNA constructs containing of the CaMV 35S promoter and NOS terminator

As described in section 8, a full-length cDNA clone of PRSV-W was developed by overlap extension PCR and *in vitro* ligation from three overlapping cDNA clones which were constructed in this research including pDPCwPD7940 (Figure 58), pTPCwCF6370 (Figure 67) and pTPCwDN6578 (Figure 77). The full-length cDNA clone, pDPCwPN7978 (Figure 80), was modified with CaMV 35S promoter and NOS terminator so that its biological activity of infectivity could be assessed by particle bombardment for inoculation of plasmid DNA onto pumpkin cotyledons. Systemic infectivity of pDPCwPN7978 was determined two weeks after inoculation by assaying plant seedlings developing from cultured cotyledons for the presence of viral coat protein by ELISA using a PRSV CP specific polyclonal antiserum.

The plasmid pDPCwPN7978 contained the full-length PRSV-W cDNA (10323 base pairs) and a poly(A) tail cDNA of 24 base pairs at the 3'end downstream from the CaMV 35S promoter and upstream the NOS terminator which was prepared from *E. coli* strain XL1-Blue by alkaline lysis with SDS with a final concentration of approximately 1 µg/µl. Fifty µl (approximately 5 µg total DNA) of the purified plasmid pDPCwPN7978 were used as target DNA for coating onto tungsten (60 mg/ml suspension) used as microcarrier in the particle bombardment.

The amplified linear cDNA fragment of approximately 11.1 kilobase pairs, PCwPN7978 (section 8.6), containing a copy cDNA of the full-length of the PRSV-W genome with a poly$(A)_{24}$ tract (at the 3'end) between a CaMV 35S promoter and a NOS terminator at the 5'-and 3'-end of the full-length cDNA fragment of the virus genome with the poly$(A)_{24}$ tract, respectively, was also used as an inoculum for a comparative experiment in infectivity assay by particle bombardment.

Two main differences between pDPCwPN7978 and PCwPN7978 constructs are their form and size of cDNAs. The pDPCwPN7978 plasmid of 14960 base pairs (closed circular form) was developed from insertion of the PCwPN7978 cDNA fragment into the pDrive cloning vector (QIAGEN). Whereas the PCwPN7978 construct was generated from two overlapping cDNA fragments, PCwPD8040 (5'-fragment) in pDPCwPD7940 plasmid and PCwDN2581 (3'-fragment) in pTPCwDN6578 plasmid which were used together as DNA templates in an overlap extension PCR for assembling and amplifying the linear cDNA fragment of 11116 base pairs without nucleotide sequence from the cloning vectors (pDrive and pTZ57R cloning vectors).

Similarly in the two constructs, the full-length cDNA sequence to PRSV-W RNA (10323 nucleotides) with a poly(A) tail of 24 nucleotides at the 3' end was placed downstream from the CaMV 35S promoter (526 base pairs) and upstream the NOS terminator (231 base pairs) without any extra nucleotides from the two junctions. There are two 6-nucleotide sequences placing at the 5'- and 3' termini of the PCwPN7978 which were fused with *Hind*III-specific nucleotide sequences (AAGCTT). Therefore, proposed size of the cDNA construct PCwPN7978 is 11116 base pairs in length.

For infectivity assay of the two PRSV-W full-length cDNA constructs (circular-formed pDPCwPN7978 plasmid and the linear-formed PCwPN7978 cDNA fragment) by particle bombardment, systemic host pumpkin (*Cucurbita moschata*) cotyledons harvested from 6 day-old seedling which were *in vitro* cultured on water agar media under aseptic condition were used as a plant tissue target for inoculating the PRSV-W full-length cDNA constructs. One pumpkin cotyledon with its petiole was placed on the center of a petri dish containing water agar. Each bombardment cartridge was shot twice with DNA coated microcarrier onto each cotyledon target. In this experiment, mock inoculation by particle bombardment with spontaneously self-ligated pDrive cloning vector which has a circular form was employed as control (5 pumpkin cotyledons, group 7). Two groups of 10 pumpkin cotyledons were used separately for infectivity assay with the circular-formed pDPCwPN7978 (group 8)

and the linear-formed PCwPN7978 (group 9) constructs. After bombardment, all of the inoculated cotyledons were transferred and cultured *in vitro* on water agar.

Under sterile conditions during a period of 14 days after the bombardment, all pumpkin seedlings developing from the pDPCwPN7978 (circular-formed)-inoculated or PCwPN7978 (linear-formed)-inoculated cotyledons did not show systemic infection of the true leaves as confirmation from negative ELISA reactions (Figure 86, group 9 and 10) and the lack of symptom development, similar to those found in the pumpkin seedlings developing from the self-ligated pDrive-inoculated cotyledons. These results show that the two full-length cDNA constructs of the PRSV-W, circular-formed pDPCwPN7978 plasmid and the linear-formed PCwPN7978 cDNAs, are not infectious and confirmed by negative reactions of ELISA.

9.2 Infectivity assay of *in vitro* transcripts of PRSV-W full-length cDNA clone in pumpkin

Construction of the recombinant plasmid pDPCwAN0178 which contains the insert fragment of the full-length cDNA clone of a Thai isolate of PRSV-W with a NOS terminator is described in section 8.5 and its proposed map is shown in Figure 83. Due to the instability of the full-length cDNA clone (pDPCwAN0178) in *E. coli*, overlap extension PCR was performed as an alternative procedure to produce a linear cDNA construct of the full-length cDNA copy of the PRSV-W which is placed downstream from a bacteriophage T7 polymerase promoter (section 8.7). The bacteriophage T7 polymerase promoter was fused to the first nucleotide of the full-length cDNA of PRSV-W by PCR amplification of PCwAG8384 cDNAs (overlap extension PCR products) which were developed from pDPCwAD0140 and pTPCwDN6578 plasmids (section 8.7). The full-length cDNA fragments of approximately 10.4 kilobase pairs fused downstream from a bacteriophage T7 polymerase promoter, PCwAG8584 (Figure 84, Lane 7), amplified from purified PCwAG8384 cDNA fragments (overlap extension PCR products) were directly used as template DNAs for *in vitro* transcription reaction after gel-purification with extraction from agarose using QIAquick® Gel Extraction kit (QIAGEN).

In vitro transcription of the purified PCwAG8584 cDNA (25 µg in 250 µl final volume) was performed in the presence of 7mG(ppp)G RNA cap structure analog (BioLaps Inc.) using T7 RNA polymerase (Promega). Quality and quantity of the capped *in vitro* transcripts of the full-length cDNA (PCwAG8584) were assessed on 0.8% agarose gel (Figure 85) after treatment with DNase I, and RNA concentration was also determined using spectrophotometry. Approximately 60 µg of *in vitro* transcripts were produced in 200 µl reaction mixture from 15 µg of DNA template. The synthesized *in vitro* transcripts were added with carborundum and directly inoculated onto both cotyledons of each pumpkin seedling using a sterilized glass spatula. Total RNA preparations extracted separately from healthy- or PRSV-W-infected pumpkin leaves which were used for negative or positive control were also determined on 0.8% agarose gel (Figure 85).

Each six-day old seedling (group of five seedlings, group 7) of systemic host pumpkin at cotyledon stage was mechanically inoculated with 20 µl transcription mixtures with the presence of suspended carborundum (final quantity of approximately 5 µg RNA) per cotyledon. Four groups of 5 six-day old healthy pumpkin seedlings were mock-inoculated separately in each group with 40 µl 0.1 M phosphate buffer, pH7.2 per seedling (group 1), with 40 µl transcription buffer per seedling (group 2), with 100 µl crude sap (in 0.1 M phosphate buffer, pH7.2) extracted from healthy pumpkin leaves per seedling (group 3), with 100 µl total RNA (final quantity of approximately 10 µg RNA in DEPC-treated water) extracted from healthy pumpkin leaves per seedling (group 4), as negative controls. And two groups of 10 six-day old healthy pumpkin seedlings were mock-inoculated separately in each group with 100 µl crude sap extracted from PRSV-W-infected pumpkin leaves per seedling (group 5), with 100 µl total RNA (final quantity of approximately 10 µg RNA in DEPC-treated water) extracted from PRSV-W-infected pumpkin leaves per seedling (group 6), as positive controls. All forty-five inoculated pumpkin seedlings were kept in a temperature- controlled (24-28 ^{o}C) greenhouse for observation of virus symptom development. Appearance and development of typical mosaic symptom on the RNA transcripts-inoculated pumpkin seedlings were recorded and compared daily

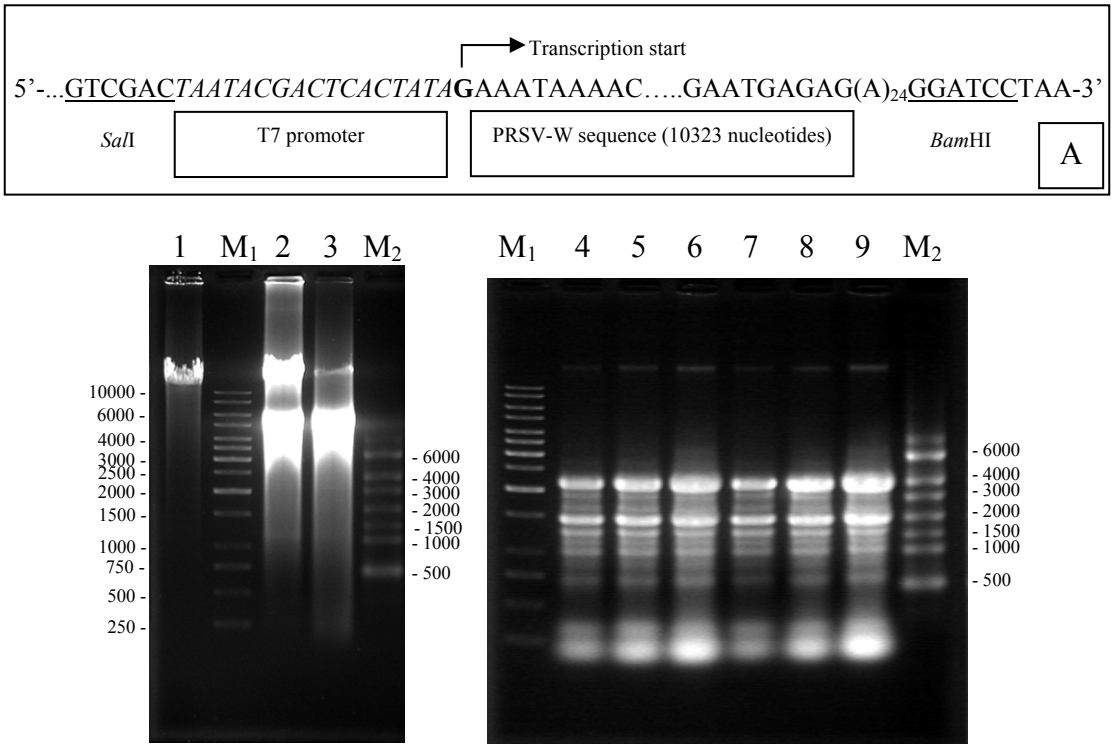


Figure 85 *In vitro* RNA transcripts of PRSV-W and total RNA preparations extracted from healthy- or PRSV-W-infected pumpkin leaves. Schematic representation of the junction between the T7 promoter and the 5'end of PRSV-W sequence (10323 nucleotides) contained an extra guanosine residue, and the 3'end of the viral sequence was followed by a poly(A)24 tract with a *Bam*HI site (A). *In vitro* RNA transcripts of PRSV-W (B) and total RNA preparations (C) extracted from healthy- or PRSV-W-infected pumpkin leaves were determined by 0.8% agarose gel electrophoresis with ethidium bromide staining. Size of each DNA (M_1, in base pairs) and RNA (M_2, in base) fragment of standard markers are separately shown in the left and right of figure, respectively.

Lane 1: 1 µl of purified 'PCwAG8584' PCR products used template
Lane 2: 1 µl of *in vitro* transcription reaction mixture
Lane 3: 1 µl of *in vitro* transcription reaction mixture after treatment with DNase I
Lane 4-6: 1 µl, 2 µl and 3 µl per lane, respectively, of total RNA preparations extracted from healthy pumpkin leaf tissues
Lane 7-9: 1 µl, 2 µl and 3 µl per lane, respectively, of total RNA preparations extracted from PRSV-W-infected pumpkin leaf tissues
Lane M_1: 1kb DNA Ladder 'GeneRuler^{TM}' (Fermentas)
Lane M_2: RNA Ladder, High Range (Fermentas)

to those of both crude sap and total RNA which were extracted separately from PRSV-W-infected pumpkin leaves (group 5 and 6) over a period of two weeks.

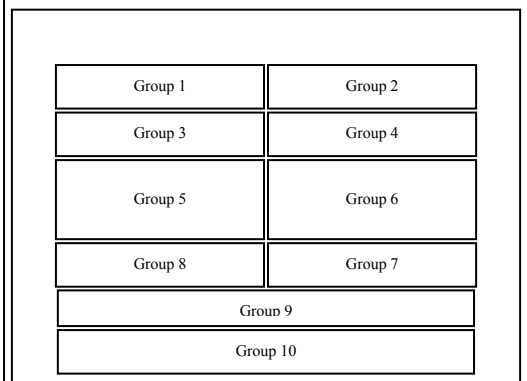
In vitro capped transcripts generated from cDNA template PCwAG8384 infected pumpkin seedlings (4 out of 5 inoculated plants of group 7) after they were mechanically inoculated. The *in vitro* transcripts of PCwAG8384 cDNA were able to induce typical mosaic symptom on pumpkin seedlings 12 to 14 days after inoculation, similar to those induced by crude sap or by total RNA which were extracted separately from PRSV-W-infected pumpkin leaves (group 5 and 6). Number of infected plants which expressed the typical mosaic symptom on all inoculated test plants 1 month after inoculation was shown in Table 39. Similarly, inoculation of pumpkin cotyledons with separate mixture of crude sap (10 out of 10 inoculated plants) or total RNA (8 out of 10 inoculated plants) which were extracted from PRSV-W-infected pumpkin leaves resulted in development of systemic mosaic symptom expression in uninoculated upper leaves at the same rate as shown in the transcript-infected pumpkin plants. Negative control plants in the experiment were shown to be always free from symptom expression. These results indicated that the capped *in vitro* transcripts from PCwAG8384 full-length cDNA fragment are infectious and result in induction of systemic mosaic symptom on pumpkin leaves.

To confirm systemic infection, all inoculated plant seedlings were analyzed by indirect-ELISA using PRSV CP antiserum to detect the presence of PRSV CP at 14 days after inoculation (Figure 86). Uninoculated upper leaf tissues of all test plants were extracted and the obtained crude sap preparations were used as samples in ELISA. A perfect correlation was found between symptom appearance and positive ELISA tests. The quantity of the CP in plants infected with the *in vitro* transcripts of PCwAG8384 cDNA was similar to those infected with both of crude sap and total RNA which were extracted separately from PRSV-W-infected pumpkin leaves (group 5 and 6). These results were also recorded by measurement of the absorbancy at 405 nm by ELISA reader (Figure 86C).

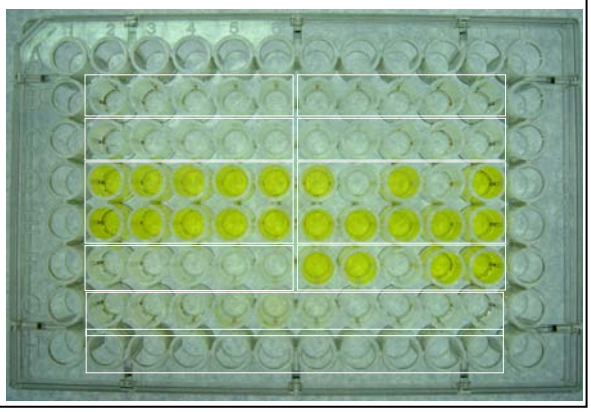
Table 39 Infectivity assay of *in vitro* capped transcripts of PRSV-W generated from PCwAG8584 cDNA which were mechanically inoculated on greenhouse grown pumpkin plants.

Inoculum	Infection rate (%)	Symptom appearance (days after inoculation)	Symptom expression	Mean ELISA absorbance value (A_{405})
Phosphate buffer, pH7.2 (group 1)	0/5 (0%)	-	No symptom	0.0986 ± 0.00297
Transcription buffer (group 2)	0/5 (0%)	-	No symptom	0.0982 ± 0.00327
Crude sap extracted from healthy pumpkin leaf tissues (group 3)	0/5 (0%)	-	No symptom	0.1168 ± 0.05887
Total RNA extracted from healthy pumpkin leaf tissues (group 4)	0/5 (0%)	-	No symptom	0.0922 ± 0.00476
Crude sap extracted from PRSV-W-infected pumpkin leaf tissues (group 5)	10/10 (100%)	12-14	Systemic mosaic	2.4677 ± 0.37615
Total RNA extracted from PRSV-W-infected pumpkin leaf tissues (group 6)	8/10 (80%)	12-14	Systemic mosaic	2.7834 ± 0.32967
In vitro capped transcripts of PRSV-W (group 7)	4/5 (80%)	12-14	Systemic mosaic	2.9965 ± 0.23204

A



B



C

Group 1	0.103	0,099	0.095	0.097	0.099	0.097	0.099	0.094	0.103	0.098	Group 2
Group 3	0.222	0.089	0.090	0.095	0.088	0.088	0.091	0.088	0.099	0.095	Group 4
Group 5	2.064 2.481	1.785 2.893	2.077 2.731	2.654 2.685	2.441 2.866	2.131 2.668	0.095 2.741	2.628 2.866	0.098 2.969	3.056 3.208	Group 6
Group 8	0.098	0.099	0.095	0.101	0.100	2.980	2.687	0.102	3.081	3.238	Group 7
Group 9	0.107	0.111	0.111	0.121	0.183	0.104	0.082	0.086	0.096	0.084	
Group 10	0.094	0.097	0.089	0.095	0.103	0.097	0.090	0.098	0.089	0.090	

Figure 86 Indirect-ELISA using PRSV CP antiserum to detect the presence of PRSV CP from experimental inoculations of pumpkin plants at 14 days after inoculation. Uninoculated upper leaf tissues of all test plants were extracted separately and the obtained crude sap preparations were used as samples in ELISA. Schematic representation of sample group positions on an ELISA microtiter plate (A) and the ELISA results showing visible color (yellow) or colorless signals (B) after color devlopment in the dark at room temperature for 15 min. The absorbance values (A_{405}) for each sample of each group (C) are shown and the positive reactions of the ELISA are indicated in shade boxes (yellow). See in section 9.1 and section 9.2 for details of ELISA sample groups.

To confirm that the development of symptoms in pumpkin seedlings was caused by the *in vitro* RNA transcripts of PCwAG8384 and the transcript-derived virus can be assembled into mature virions, crude sap from leaf tissues of infected pumpkin seedlings expressing mosaic symptom and positive reactions of ELISA was observed the presence or not of morphological *Potyvirus* particles under TEM. By electron microscopy, filamentous virus particles of 760-800 nm in length typical of PRSV-W were observed in the samples of upper leaf tissue extracted from RNA transcript-infected test plants (Figure 87). Virus yields from systemically infected pumpkin leaves after RNA transcript inoculation were similar to those observed in pumpkin leaves mechanically inoculated with crude sap of pumpkin leaves systemically infected with the original, greenhouse-cultured virus. The virions produced from inoculation with the *in vitro* RNA transcripts of PCwAG8384 were indistinguishable from those of both crude sap and total RNA which were extracted separately from PRSV-W-infected pumpkin leaves.

These results show that the *in vitro* RNA transcripts of PCwAG8584 are infectious which confirmed by positive reactions of ELISA and the presence of typical *Potyvirus* particle in shape and size. Therefore, the progeny virus also reacted positively with homologous antiserum when the virus CP was detected using ELISA.

Infectivity of the progeny virus derived from transcripts was tested by mechanical inoculation of systemic host plants with crude sap from transcript-infected pumpkin leave tissues. Crude sap from pumpkin plants showing systemic mosaic symptom of viral infection after inoculation with capped *in vitro* transcripts of PCwAG8384, was obtained by grinding leaves in 0.1 M sodium-potassium phosphate buffer, pH 7.2 and used to inoculate new sets of plants with cotton sticks. At Maejo University in Chiang Mai province, pumpkin and zucchini squash plants inoculated with crude sap which was extracted from leaf tissues of the initial transcript-infected pumpkin plants expressing mosaic symptom were considered infected and expressed the typical mosaic symptom on all inoculated test plants 12 to 14 days after inoculation, similar to those induced by crude sap which was extracted from PRSV-W-infected pumpkin leaves. The symptom, systemic mosaic on uninoculated upper

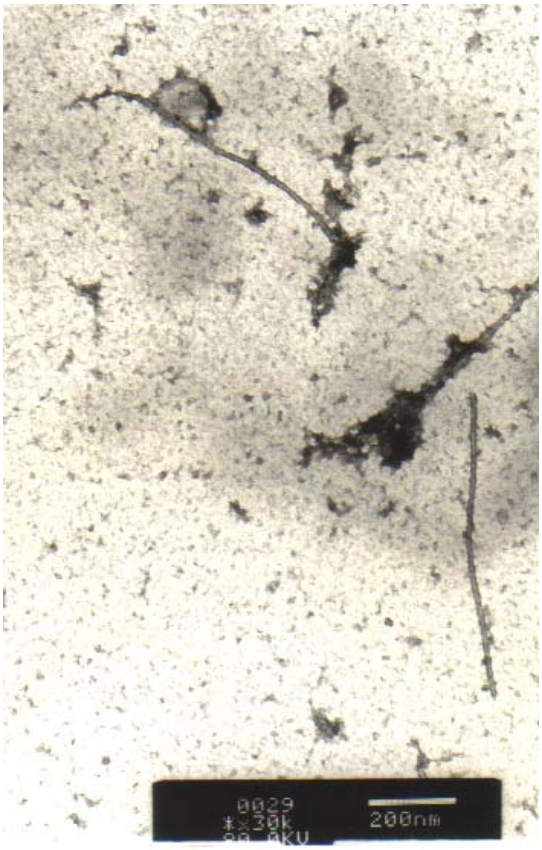


Figure 87 Electron micrograph showing PRSV-W particles in a crude sap prepared from pumpkin plants inoculated with *in vitro* capped transcripts of PRSV-W. The crude sap was prepared from an upper, uninoculated pumpkin leaf from plants mechanically inoculated with PRSV-W in vitro-transcribed RNA on the cotyledons.

leaves, was indistinguishable from those caused by the native virus. The time of symptom appearance was the same for plants inoculated with viral progenies of PCwAG8584 or by the native virus. It is concluded that the progeny virus derived from *in vitro* transcripts was also infectious.

These results clearly showed that progeny virus derived from infectious *in vitro* transcripts-infected plants was efficiently transmitted by sap inoculation on its host plants, and typical mosaic symptom expression was the same with parental wild type virus. There is no difference of symptom expression between the progeny virus from transcripts RNA and parental virus when the progeny virus was passaged from initial transcript-infected pumpkin to other healthy host plants. In conclusion, these results show that the PRSV-W *in vitro* transcript system is fully functional.

In this research, full-length cDNA clones of PRSV-W Thai isolate were successfully constructed downstream a CaMV 35S promoter or a bacteriophage T7 RNA polymerase promoter. However, toxicity-associated problem in these PRSV-W full-length cDNA clones was noticed. The recombinant plasmids, pDPCwPN7978 and pDPCwAN0178, could not be propagated in *E. coli* used as bacterial host cells. Cloning of the full-length cDNA of the PRSV-W into two subclones and overlap PCR amplification to fuse the two overlapping cDNA fragments which were inserted separately in each clone, before used as inoculum (for *in vivo* transcription) or used as template DNA (for *in vitro* transcription) were performed as a strategy to avoid these problems.

Both circular-(pDPCwPN7978) and linear-(PCwPN7978) forms of the PRSV-W full-length cDNAs which contained CaMV 35S promoter and NOS terminator failed to establish infection by particle bombardment inoculation on the ten pumpkin cotyledons for each constructs. These show that the two constructs were not infectious by confirmation of the results with the lack of symptom development and negative reactions of ELISA. The unsuccessful infection of these constructs may be due to the inoculation method (particle bombardment) which was not an appropriate condition for the delivery of the constructs into plant nuclei. This may be responsible

for the lack of infection on *in vitro* cultured pumpkin seedlings. This is dissimilar to the results of infection of PRSV-P and ZYMV full-length cDNA clones which were inoculated onto papaya and zucchini squash plants, respectively, by particle bombardment (Chiang and Yeh, 1997; Lin *et al*., 2002). Infectious cDNA clones have been illustrated for members of several families of animal and plant viruses (Boyer and Haenni, 1994). The infectivity of a number of potyviral cDNA clones under control of CaMV 35S promoter by mechanical inoculation onto plants has been reported, including PPV (Maiss *et al*., 1992), ZYMV (Gal-On *et al*., 1995; Lin *et al*., 2002) and PRSV-P (Chiang and Yeh, 1997). Low or no infectivity after the mechanical inoculation of test plants using these clones was referred (Gal-On *et al*., 1995). Differences in infectivity efficiencies of infectious cDNA clones have been reported as dependent on the amount of DNA applied per inoculated host, method of inoculation (manually, particle bombardment, agroinfection), presence of polyadenylation signal, plasmid linearization and/or excision of vector sequences.

The reason for pDPCwPN7978 not being infectious might be due to the amount of the plasmid used as inoculum was too low and not enough to infect the pumpkin plants. While the linear form of PCwPN7978 cDNA was not available in used as inoculum. Another possibility might be that the *in vitro* cultured pumpkin cotyledons used for inoculation of plant were not appropriate for bombardment. Finally, the failure of pDPCwPN7978 (circular formed) or PCwPN7978 (linear formed) to be infectious may be due to mistake in the viral nucleotide sequence in these cDNA constructs. However, this appears questionable because PRSV-W sequences used for the construction of pDPCwPN7978 (circular formed) or PCwPN7978 (linear formed) and pDPCwAN0178 (circular formed) or PCwAG8584 (linear formed) had essentially the same origin.

In vitro capped transcripts synthesized from PCwAG8584 cDNA template was able to infect up to 80% of pumpkin plants. This full-length cDNA template of the PRSV-W which was placed downstream from a bacteriophage T7 polymerase promoter was developed from two overlapping plasmids, pDPCwAD0140 and pTPCwDN6578. The two overlapping subclones were constructed for the full-length

cDNA of PRSV-W and the viral cDNA sequences were assembled before used as DNA template in *in vitro* transcription reaction, in order to avoid toxicity-associated problems when the full-length cDNA of the virus was cloned in a high copy number plasmid. This strategy has been used to avoid undesired expression of cloned full-length material in bacteria (Jakab *et al*., 1997). This is a main goal which was proposed at the beginning of this research to optimize an appropriate condition in order to produce infectious cDNA.

In initial experiments, *in vitro* uncapped transcripts of PCwAG8584 cDNA fragments inoculated on pumpkin plants, failed to produce symptoms despite repeated attempts. The lack of infectivity of the transcript probably due to the absence of 7mG(ppp)G RNA cap structure analog at the 5'end of the full-length transcript RNAs. Infectivities of *in vitro* capped transcripts of potyviral clones were initially described in the 5 to 49% range (Domier *et al*., 1989; Riechmann *et al*., 1990), and recently reached 100% (Flasinski *et al*., 1996; Nicolas *et al*., 1996), probably due to optimization of capping and transcription efficiencies with new commercially available kits. The presence of a cap (7mG(ppp)G) is required at the 5'end for optimal infectivity of potyviral transcripts by mechanical inoculation (Domier *et al*., 1989; Riechmann *et al*., 1990; Gal-On *et al*., 1991; Dolja *et al*., 1992; Gal-On *et al*., 1995; Flasinski *et al*., 1996; Nicolas *et al*., 1996; Chiang and Yeh, 1997; Lin *et al*., 2002). In this study, only the capped transcript of PRSV-W was infectious, possibly because it protected the transcript from exonucleolytic degradation in host cells (Green *et a*l., 1983) and increased translation initiation of viral proteins (Contreras *et al*., 1982). Accordingly, a 7mG(ppp)G RNA cap structure analog can reinstate the function of the VPg. This is in understanding with the applications made for a number of other viruses possessing a 5'VPg (Riechmann *et al*., 1990).

As a general rule, nonviral nucleotides at the 5'ends considerably reduce or even invalidate infectivity of an *in vitro* transcript (Boyer and Haenni, 1994). The infectious transcript of PRSV-W with the T7 promoter was synthesized *in vitro* and the RNA transcripts were introduced one extra guanosine (G) residue at its 5'end by PCR using primer PC83 which contains the sequence corresponding to T7 promoter

and the additional G. This is similar to most of the infectious constructs of potyviruses in which the nonviral nucleotide introduced at the 5'end is a G residue (Domier *et al*., 1989; Riechmann *et al*., 1990; Gal-On *et al*., 1991; Dolja *et al*., 1992; Chiang and Yeh, 1997; Lin *et al*., 2002). Both T3 and T7 RNA polymerases initiate transcription primarily at G residues (Bailey *et al*., 1983; Domier *et al*., 1989; Chiang and Yeh, 1997; Lin *et al*., 2002) in order to provide a higher efficiency for capping the *in vitro* synthesized RNA transcripts.

The presence of a long A sequence in poly(A) tail at the 3'end is proposed to be essential in the stabilization of the RNA transcript and in enhancement of an efficient translation (Jackson and Standart, 1990). In the synthesized poly(A) tails at the 3'end of RNA transcripts, their sizes in length may be responsible for the infectivity of the viral full-length RNA transcripts. In TYMV, RNA transcripts with 12 nucleotides poly(A) tails were not infectious, while the longer (37 or 96 nucleotides) poly(A) tails which were exchanged resulted in the production of infectious transcripts (Domier *et al*., 1989). The poly(A) tail of the PRSV-W full-length cDNAs which were constructed in this study constituted only 24 A residues and is shorter than other infectious transcripts of potyviruses such as TVMV with 37 A residues (Domier *et al*., 1989), ZYMV with 66 or 95 A residues (Gal-On *et al*., 1995; Lin *et al*., 2002) and PRSV-W with 36 A residues (Chiang and Yeh, 1997).

Additionally, the presence of 9 additional residues (5'-GGAUCCUAA-3') after the poly(A) track of the PRSV-W transcript from PCwAG8584 cDNA did not interfere with its infectivity. In potyviruses, the length of additional nonviral nucleotides at the 3'end of viral sequence (after the poly(A) track) ranging from 1 to 12 residue were reported from several infectious *in vitro* transcripts (Domier *et al*., 1989; Riechmann *et al*., 1990; Gal-On *et al*., 1991; Dolja *et al*., 1992; Chiang and Yeh, 1997; Lin *et al*., 2002). While the longer (64 residues) nonviral sequence of PRSV-P transcript from pT3-Xb interfered with its infectivity. Nonviral nucleotides at the 3'end of viral sequence are believed to be influential in the formation of the secondary structure which may be an improper structure for initiation of RNA synthesis (Boyer and Haenni, 1994).

Systemic mosaic symptom of pumpkin plants inoculated with infectious *in vitro* transcripts of PRSV-W was not considerably delayed when compared to those induced by the control crude sap and total RNA which were extracted separately from PRSV-W-infected pumpkin leaves. This is similar to the case of papaya plants induced by the infectious PRSV-P transcripts which they were constructed by Chiang and Yeh (1997), but different from those induced by the other infectious potyviral transcripts which symptom appearances were slightly delayed. The *in vitro* capped transcripts of PRSV-W produced in this study also had a high infection rate of 80% when compared with the infectious transcripts of other potyviruses. The lack of delay in symptom appearance and the high level of infection rate may be due to the cDNA clones derived from an isolate of highly infectious viral RNA used in the preparation, and the presence of a cap (7mG(ppp)G) at the 5'end of the transcripts of PRSV-W.

CONCLUSION

The virus used in this study is an isolate of PRSV-W which belongs to genus *Potyvirus* in family *Potyviridae*. The taxonomic status was supported by host range specificity and typical mosaic symptom on systemic hosts for identifying virus pathotype, ELISA and viral sequence analysis for confirmation of virus species, and visualization of the presence of typical *Potyvirus* particle in shape and size.

Analysis of the assembled sequence reveals that the PRSV-W RNA contains 10323 nucleotides excluding the poly(A) sequence of unknown length at the 3'end. The overall nucleotide composition of the viral RNA is 30.33% A, 18.96% C, 24.55% G and 26.16% U. The RNA genome of PRSV-W contains a single long ORF of 10032 nucleotides, which accounts for 97.2% of the genomic RNA sequence, coding for a polyprotein. The long ORF begins at nucleotide positions 86-88 (AUG) and terminates at a UGA stop codon at nucleotide positions 10115-10117. The ORF was predicted to potentially encode a putative polyprotein of 3343 amino acids with a calculated molecular weight of 380.3 kDa. The 5'- and 3'- NCRs consisted of 85 and 206 residues, respectively. The nucleotide composition of the 5'NCR, 45.8% A, 25.8% C, 4.7% G and 23.5% U, is significantly different from that of the complete nucleotide sequence. The PRSV-W 3'NCR of 206 nucleotides excluding the poly(A) tail is the same length as those of other different PRSV isolates.

Like other fully-sequenced PRSV isolates, polyprotein of the PRSV-W isolate is probably proteolytically processed into ten putative mature proteins consisting P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP, with nine cleavage sites by the three viral encoded proteinases. All pairs of amino acid residues in each site are identical to those of other PRSV isolates reported previously. This was supported by the presence of conserved motifs which are required for the proteolytic process of potyviral polyprotein in these sequences, such as the serine-type proteinase domain H_{456}-(8X)-D_{465}-(31X)-G_{497}-(X)-S_{499}-G_{500} and the proteolytic domain FITRGR in P1 protein at amino acid positions 519-524, C_{890}-(72X)-H_{963} in HC-Pro, and H_{2327}-(34X)-D_{2362}-(68X)-C_{2431} in NIa-Pro.

The first 23-nucleotide sequence of the virus RNA at the 5'-terminus of the genome within the 5'NCR is identical among PRSV species. Ten CAA repeat sequences which have been reported earlier and described for the 5' NCR of TMV and SCMV are also found in the region of the PRSV-W 5'NCR. Potyboxes "a" and "b" were also found within the PRSV-W 5'NCR which are located at nucleotde positions 13-24 and 47-53, respectively. The TATA box-like sequence was found within the PRSV-W 3'NCR. Additionally, two locations of repeated sequence, AUACAGUGUGG, were found and they are located at nucleotides positions 10190-10200 and 10276-10286 in the 3' NCR of PRSV-W. Most of the functionally important motifs or sequences associated with virus replication, gene regulation or virus translocation of potyviruses were detected in the PRSV-W polyprotein when the conserved amino acid sequence motifs were identified in the ten putative mature proteins of the PRSV-W.

Nucleotide sequence comparisons of NCRs indicate that the 3'-NCR, with 88.4-100.0% sequence identity is more conservative than the 5'-NCR, with 61.6-98.8% sequence identity among the twelve fully-sequenced PRSV isolates. Sequence comparisons of predicted mature proteins reveal that the most conserved protein region among the twelve fully-sequenced PRSV isolates is the cylindrical inclusion (CI) protein, with 96.7-99.7% sequence identity. The region with the lowest identity among the PRSV isolates in this study is the first protein (P1), with identity ranging from 65.4 to 97.1%. All of the phylogenetic trees illustrate that the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate except the tree generated from multiple amino acid sequence alignment of 6K1 peptide which shown the closest relationship of the PRSV-W isolate with PRSV-P-HA.

Phylogenetic branching patterns of 15 trees illustrated a place in each tree of the PRSV-W isolate which was clustered with the Asian isolates including other PRSV isolates from Thailand and Taiwan. While the three Thai isolates were grouped together to form a cluster separating from other isolates in the tree constructing based on amino acid sequences from CI. However, no phylogenetic tree grouping depend on "type (P or W)" was found in all of these 22 trees, except the tree

constructed from 6K1 amino acid sequences which it illustrated a phylogenetic relationship of partially separated clusters depending on the types of the PRSV isolates.

Both phylogenetic trees constructed separately from multiple alignments of the complete genomic nucleotide sequences and the putative overall polyprotein amino acid sequences of twelve fully-sequenced PRSV isolates (including PRSV-W isolate) are similar in branching pattern and the both trees illustrate that the PRSV-W isolate is most closely related to PRSV-W-ThRb isolate. These phylogenetic relationships are depended on the geographic location of the isolates.

The degree of identities both nucleotide and amino acid to the different genomic regions of the PRSV-W isolate varied depending on the originally geographic locations of the isolates and the genomic regions. These analyses reveal that the PRSV-W isolate has the highest nucleotide (range from 98.4 to 100.0%) and amino acid (range from 97.1 to 100.0%) identities in all of different genomic region sequences with PRSV-W-ThRb isolate.

In NCRs, the highest percentage 5'-NCR and 3'-NCR nucleotide identities of the PRSV-W isolate are 98.8% and 100.0%, respectively, with the corresponding regions of PRSV-W-ThRb isolate. The 5'-NCR and 3'-NCR of the PRSV-W isolate has the nucleotide sequence identities varied from 66.3 to 89.5% and 93.2 to 98.1%, respectively, when compared with the corresponding region of other fully-sequenced PRSV isolates except those of PRSV-W-ThRb.

The P1 coding region of the PRSV-W isolate has the nucleotide sequence identities varied from 69.5 with that of PRSV-W-1 to 81.9% with that of PRSV-P-TH and the amino acid sequence identities varied from 66.7 with that of PRSV-W-1 to 82.2% with that of PRSV-P-TH, when compared with the corresponding region of other fully-sequenced PRSV isolates except that of PRSV-W-ThRb. These indicate that the P1 coding region is highly variable while the remainders are the conserved regions among the PRSV species.

The genomic RNA and putative polyprotein of the PRSV-W isolate have the ranges of 82.7 to 88.6% identical nucleotides and 90.8 to 93.8% identical amino acids, respectively, when compared with those of other fully-sequenced PRSV isolates (with the exception of PRSV-W-ThRb isolate). This indicates that the majority of nucleotide changes are silent.

The PRSV-W isolate has the highest percentage nucleotide (98.9%) and amino acid (98.3%) identities of the complete genome and putative polyprotein, respectively, with those of PRSV-W-ThRb isolate when compared with those of other fully-sequenced PRSV isolates. Both PRSV-W and PRSV-W-ThRb isolates are classified into the cucurbit-infecting (W) type. They were isolated from the same country (Thailand) but they are different in geographic location of origin. Geographic origin of the PRSV-W isolate is located in Nakhon Sawan province, while the PRSV-W-ThRb is an isolate originating in Ratchaburi.

Two full-length PRSV-W cDNA clones under the control of the enhanced CaMV 35S (named pDPCwPN7978) or bacteriophage T7 (pDPCwAN0178) promoters were successfully constructed in this study. Two main differences between the two full-length cDNA clones are the presence or not of the CaMV 35S promoter and their opposite orientations of the inserts in cloning vector (pDrive, QIAGEN). The both full-length PRSV-W cDNA clones, pDPCwPN7978 with a size of 14960 nucleotides and pDPCwAN0178 with a size of 14443 nucleotides, could not be propagated in *E. coli* used as bacterial host cells whereas the other partial cDNA clones of ranging from 5.2 to 9.5 kilobase pairs were stable and they could be propagated in the bacterial host cells.

Full-length PRSV-W cDNAs under the control of the CaMV 35S promoter both pDPCwPN7978 plasmid (circular-formed in pDrive cloning vector (QIAGEN) and linear-formed construct of PCwPN7978 cDNA fragment were not infectious when inoculated separately onto pumpkin cotyledons with these purified cDNA constructs by particle bombardment, according to no expression of induced symptoms and confirmed by negative reactions of ELISA.

In vitro capped RNA transcripts of PCwAG8584 are highly infectious on pumpkin plants according to systemically induced symptoms which confirmed by positive reactions of ELISA and the presence of typical *Potyvirus* particle in shape and size. The presence of the virions in crude sap prepared from leaf tissues of infected pumpkin seedlings indicated that the transcript-derived virus can be assembled into mature virions. The virions multiplied from inoculation with infectious RNA transcripts were indistinguishable from those of the parental virus. Therefore, the progeny virus also reacted positively with homologous antiserum when the virus CP was detected using ELISA. Full-length cDNA fragments, PCwAG8584, constructed under the control of the bacteriophage T7 RNA promoter could be transcribed to produce *in vitro* transcript RNAs. The transcript RNAs produced from the uncloned PCR products was apparently full-length and approximately 10.3 kilobases long, the same size as the wild-type RNA genome of the virus, when compared with RNA size standard markers. *In vitro* capped transcripts from template PCwAG8584 infected 80% of the pumpkin plants (4 of 5 inoculated) when mechanically inoculated, whereas *in vitro* uncapped transcripts of the PCwAG8584 were not infectious in pumpkin plants when initially tested. The appearance of symptoms in the pumpkin plants inoculated with the infectious transcript was observed at 12-14 days after inoculation which was similar to those induced by crude sap or by total RNA which were extracted separately from PRSV-W-infected pumpkin leaves. Presence of PRSV-W CP in inoculated and upper leaves of the tested plants was verified ELISA. The amounts of viral CP detected were similar level for plants inoculated with both of crude sap and total RNA which were extracted separately from PRSV-W-infected pumpkin leaves. From these results it can be concluded that *in vitro* capped RNA transcripts of PCwAG8584 are highly infectious on pumpkin plants and an RNA cap structure analog, 7mG(ppp)G, is needed for infectivity. Furthermore, progeny virus from infectious *in vitro* transcripts-infected plants could also be passaged to healthy plant and induced the same symptoms. The transcript-derived virus was efficiently transmitted by sap inoculation on its systemic host plants (pumpkin and zucchini squash). No delay in symptom appearance was observed when compared with the appearance of the symptom of control plants which were inoculated with crude sap extracted from PRSV-W-infected pumpkin leaves.

LITERATURE CITED

Adlerz, W.C. 1972a. *Melothria pendula* plants infected with watermelon mosaic virus 1 as a source of inoculum for cucurbits in Collier Country, Florida. **J. Econ. Entomol.** 65: 1303-1306.

_____. 1972b. *Momordica charantia* as source of watermelon mosaic virus 1 for cucurbit crops in Palm beach Country, Florida. **Plant Dis. Rept.** 56: 563-564.

Ahlquist, P., R. French and J.J. Bujarski. 1987. Molecular studies of brome mosaic virus using infectious transcripts from cloned cDNA. **Adv. Virus Res.** 32: 215-242.

_____, _____, M. Janda and L.S. Loesch-Fries. 1984. Multicomponent RNA plant virus infection derived from cloned viral cDNA. **Proc. Natl. Acad. Sci. USA** 81: 7066-7070.

_____ and M. Janda. 1984. cDNA cloning and *in vitro* transcription of the complete brome mosaic virus genome. **Mol. and Cellular Biol.** 12: 2876-2882.

Allison, R.F., R.E. Johnston and W.G. Dougherty. 1986. The nucleotide sequence of the coding region of tobacco, etch virus genomic RNA: evidence for the synthesis of a single polyprotein. **Virol.** 154: 9-20.

Almazan, F., J.M. Gonzalez, Z. Penzes, A. Izeta, E. Calvo, J. Plana-Duran and J. Enjuanes. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. **Proc. Natl. Acad. Sci. USA** 97: 5516-5521.

Altmann, M., S. Blum, T.M.A. Wilson and H. Trachsel. 1990. The 5'-leadder sequence of tobacco mosaic virus RNA mediates initiation-factor-4E-independent, but still initiation-factor-4A-dependent translation in yeast extracts. **Gene** 91: 127-129.

Angenent, G.C., E. Posthumus and J.F. Bol. 1989. Biological activity of transcripts synthesized *in vitro* from full-length and mutated DNA copies of tobacco rattle virus RNA 2. **Virol.** 173: 68-76.

Anindya, R., S. Chittori and H.S. Savithri. 2005. Tyrosine 66 of Pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. **Virol.** 336: 154-162.

Arbatova, J., K. Lehto, E. Pehu and T. Pehu. 1998. Localization of the P1 protein of potato Y potyvirus in association with cytoplasmic inclusion bodies and in the cytoplasm of infected cells. **J. Gen. Virol.** 79: 2319-2323.

Atreya, C.D. 1992. Application of genome sequence information in potyvirus taxonomy: an overview. **Arch. Virol.** 5: 17-23.

_____ and T.P. Pirone. 1993. Mutational analysis of the helper component-proteinase gene of a potyvirus: effects of amino acid substitutions, deletions, and gene replacement on virulence and aphid transmissibility. **Proc. Natl. Acad. Sci. USA** 90: 11919-11923.

Atreya, P.L., C.D. Atreya and T.P. Pirone. 1991. Amino acid substutions in the coat protein result in loss of insect transmissibility of a plant virus. **Proc. Natl. Acad. Sci. USA** 88: 7887-7891.

Attasart, P., G. Charoensilp, S. Kertbundit, S. Panyim and M. Juricek. 2002. Nucleotide sequence of a Thai isolate of papaya ringspot virus type W. **Acta Virologica** 46: 241-246.

Bailey, J.N., J.F. Klement and W.T. McAllister. 1983. Relationship between promoter structure and template specificities exhibited by the bacteriophage T3 and T7 RNA polymerases. **Proc. Natl. Acad. Sci. USA** 80: 2814-2818.

Benfey, P.N. and N.H. Chua. 1990. The cauliflower mosaic virus 35S promoter: Combinatorial regulation of transcription in plants. **Science.** 254: 959-966.

Blanc, S., E.D. Ammar, S. Garcia-Lampason, V.V. Dolja, C. Llave, J. Baker and T.P. Pirone. 1998. Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. **J. Gen. Virol.** 79: 3119-3122.

Boyer, J.C., G. Drugeon, K. Seron, M.D. Morch-Devignes, F. Agnes and A.L. Haenni. 1993. *In vitro* transcripts of turnip yellow mosaic virus encompassing a long 3' extension or produced from a full-length cDNA clone harboring a 2-kb-long PCR amplified segment are infectious. **Res. Virol.** 144: 339-348.

_____ and A.L. Haenni. 1994. Infectious transcripts and cDNA clones of RNA viruses. **Virol.** 198: 415-426.

Brantley, J.D. and A.G. Hunt. 1993. The N-terminal protein of the polyprotein encoded by the potyvirus tobacco vein mottle virus is an RNA-binding protein. **J. Gen. Virol.** 74: 1157-1162.

Carrington, J.C., S.M. Cary and G.D. Dougherty. 1988. Mutational analysis of tobacco etch virus polyprotein processing *cis* and *trans* proteolytic activities of polyproteins containing the 49-kilodalton proteinase. **J. Virol.** 62: 2313-2320.

_____, _____, T.D. Parks and G.D. Dougherty. 1989a. A second proteinase encoded by a plant potyvirus genome. **EMBO J.** 80: 365-370.

_____ and G.D. Dougherty. 1987. Small nuclear inclusion protein encoded by a plant potyvirus genome is a proteinase. **J. Virol.** 61: 2540-2548.

_____, D.D. Freed and T.C. Sanders. 1989b. Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and *in vitro*. **J. Virol.** 63: 4459-4463.

_____ and K.L. Herndon. 1992. Characterization of the potyviral HC-Pro autoproteolytic cleavage site. **Virol.** 187: 308-315.

Casais, R., V. Thiel, S.G. Siddell, D. Cavanagh and P. Britton. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. **J. Virol.** 75: 12359-12369.

Castle, S.J., T.M. Perring, C.A. Farrar and A.N. Kishaba. 1992. Field and laboratory transmission of watermelon mosaic virus 2 and zucchini yellow mosaic virus by various aphid species. **Phytopathol.** 82: 235-240.

Charoensilp, G., P. Attasart, M. Juricek, S. Panyim and S. Kertbundit. 2003. Sequencing and characterization of Thai *Papaya ringspot virus* isolate type P (PRSVthP). **ScienceAsia** 29: 89-94.

Cheng, S., C. Fockler, W.M. Barnes and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. **Proc. Natl. Acad. Sci. USA** 91: 5695–5699.

Chiang, C.H. and S.D. Yeh. 1997. Infectivity assays of in vitro and in vivo transcripts of papaya ringspot potyvirus. **Bot. Bull. Acad. Sin.** 38: 153-163.

Christou, P. 1992. Genetic transformation of crop plants using microprojectile bombardment. **Plant J.** 2: 275-281.

Clark, M.F. and A.N. Adams. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses**.** **J. Gen. Virol.** 34: 475-483.

Colinet, D., J. Kummert and P. Lepoivre. 1998. The nucleotide sequence and genome organization of the whitefly transmitted sweetpotato mild mottle virus: a close relationship with members of the family *Potyviridae*. **Virus Res.** 53: 187-196.

Contreras, R., H. Cheroutre, W. Degrave and W. Fiers. 1982. Simple, efficient *in vitro* synthesis of capped RNA useful for direct expression of cloned eukaryotic genes. **Nucleic Acids Res.** 10: 6353-6362.

Cronin, S., J. Verchot, R. Haldeman-Cahill, M.C. Schaad and J.C. Carrington. 1995. Long-distance movement factor: a transport function of the potyvirus helper component proteinase. **Plant Cell** 7: 549-559.

Dallot, S., L. Quiot-Douine, P. Saenz, M.T. Cervera, J.A. Gracia and J.B. Quiot. 2001. Identification of *Plum pox virus* determinants implicated in specific interaction with different *Prunus* spp. **Phytopathol.** 91: 159-164.

Dawson, W.O., D.L. Beck, D.A. Knorr and G.L. Grantham. 1986. cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. **Proc. Natl. Acad. Sci. USA** 83: 1832-1836.

Demler, S.A., D.G. Rucker and G.A. De Zoeten. 1993. The chimeric nature of the genome of pea enation mosaic virus: The independent replication of RNA2. **J. Gen. Virol.** 74: 1-14.

Dessens, J.T. and G.P. Lomonossoff. 1993. Cauliflower mosaic virus 35S promoter-controlled DNA copies of cowpea mosaic virus RNAs are infectious on plants. **J. Gen. Virol.** 74: 889-892.

Dolja, V.V., R. Haldeman-Cahill, A.E. Montgomery, K.A. Vandenbosch and J.C. Carrington. 1995. Capsid protein determinants involved in the cell-to-cell

and long distance movement of tobacco etch potyvirus. **Virol.** 206: 1007-1016.

_____, _____, N.L. Robertson, W.G. Dougherty and J.C. Carrington. 1994. Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants. **EMBO J.** 13: 1482-1491.

_____, H.J. McBride and J.C. Carrington. 1992. Tagging of plant potyvirus replication and movement by insertion of beta-glucuronidase into the viral polyprotein. **Proc. Natl. Acad. Sci. USA** 89: 10208-10212.

Dombrosky, A., H. Huet, N. Chejanovsky and B. Raccah. 2005. Aphid transmission of a potyvirus depends on suitability of the helper component and the N terminus of the coat protein. **Arch. Virol.** 150: 287-298.

Domier, L.L., K.M. Franklin, A.G. Hunt, R.E. Rhoads and J.G. Shaw. 1989. Infectious *in vitro* transcripts from cloned cDNA of a potyvirus, tobacco vein mottling virus. **Proc. Natl. Acad. Sci. USA** 86: 3509-3513.

_____, _____, M. Shahabuddin, G.M. Hellmann, J.H. Overmayer, S.T. Hiremath, M.F.E. Siaw, G.P. Lomonossoff, J.G. Shaw and R.E. Rhoads. 1986. The nucleotide sequence of tobacco vein mottling virus RNA. **Nucleic Acids Res.** 14: 5417-5430.

_____, J.G. Shaw and R.E. Rhoads. 1987. Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. **Virol.** 158: 20-27.

Dore, J.M., C. Erny and L. Pinck. 1990. Biologically active transcripts of alfalfa mosaic virus RNA3. **FEBS Lett.** 264: 183-186.

_____ and L. Pinck. 1988. Plasmid DNA containing a copy of RNA 3 can substitute for RNA 3 in alfalfa mosaic virus RNA inocula. **J. Gen Virol.** 69: 1331-1338.

Dougherty, W.G. and J.C. Carrington. 1988. Expression and function of potyvirus gene products. **Ann. Rev. Phytopathol.** 26: 123-143.

_____, S.M. Cary and T.D. Parks. 1989a. Molecular genetics analysis of a plant virus protein cleavage site: A model. **Virol.** 171: 356-364.

_____ and T.D. Parks. 1991. Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and protease domains. **Virol.** 183: 449-456.

_____, _____, S.M. Cary, J.F. Bazan and R.J. Fletterick. 1989b. Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. **Virol.** 172: 302-310.

_____ and B.L. Semler. 1993. Expression of virus-encoded proteinase: functional and structural similarities with cellular enzymes. **Microbiol. Rev.** 57: 781-822.

Dower, W.J. 1988. Transformation of *E. coli* to extremely high efficiency by electroporation. **Plant Mol. Biol. Rept.** 6: 3-4.

Duechler, M., T. Skern, D. Blaas, B. Berger, W. Sommergruber and E. Kuechler. 1989. Human rhinovirus serotype 2: *In vitro* synthesis of an infectious RNA. **Virol.** 168: 159-161.

Dunn, J.J. and F.W. Studier. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the location of T7 genetic elements. **J. Mol. Biol.** 166: 477-485.

Dzianott, A.M. and J.J. Bujarski. 1989. Derivation of an infectious viral RNA by autocatalytic cleavage of *in vitro* transcribed viral cDNAs. **Proc. Natl. Acad. Sci. USA** 86: 4823-4827.

Eggen, R., J. Verver, J. Wellink, A. De Jong, R. Goldbach and A. Van Kammen. 1989. Improvements of the infectivity of *in vitro* transcripts from cloned cowpea mosaic virus cDNA: Impact of terminal nucleotide sequences. **Virol.** 173: 447-455.

Fakhfakh, H., F. Vilaine, M. Makni and C. Robaglia. 1996. Cell-free cloning and biolistic inoculation of an infectious cDNA of *Potato virus Y*. **J. Gen. Virol.** 77: 519-523.

Fan, Z.F., H.Y. Chen, X.M. Liang and H.F. Li. 2003. Complete sequence of the genomic RNA of the prevalent strain of a potyvirus infecting maize in China. **Arch. Virol.** 148: 773-782.

Fang, R.X., S. Sivasubramaniam and N.H. Chua. 1989. Multiple cis-regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. **Plant Cell** 1: 141-150.

Fedorkin, O.N., A. Merits, J. Lucchese, A.G. Solovyev, M. Saarma, S.Y. Morozv and K. Makinen. 2000. Complementation of the movement-deficient mutations in potato virus X: potyvirus coat protein mediates cell-to-cell trafficking of C-terminal truncation but not detection mutant of potexvirus coat protein. **Virol.** 270: 31-42.

Fischer, H.U. and B.E.L. Lockhart. 1974. Serious losses in cucurbits caused by watermelon mosaic virus in Morocco. **Plant Dis. Rept.** 58: 143-146.

Flasinski, S., U.B. Gunasinghe, R.A. Gonzales and B.G. Cassidy. 1996. The cDNA sequence and infectious transcripts of *Peanut stripe virus*. **Gene** 171: 299-300.

Furuichi, Y., A. Lafiandra and A.J. Shatkin. 1977. 5'-Terminal structure and mRNA stability. **Nature** 266: 235-239.

Gal-On, A., Y. Antignus, A. Rosner and B. Raccah. 1991. Infectious in vitro RNA transcripts derived from cloned cDNA of the cucurbit potyvirus *Zucchini yellow mosaic virus*. **J. Gen. Virol.** 72: 2639-2643.

Gal-On, A., E. Meiri, H. Huet, W.J. Hua, B. Raccah and V. Gaba. 1995. Particle bombardment drastically increases the infectivity of cloned DNA of zucchini yellow mosaic potyvirus. **J. Gen. Virol.** 76: 3223-3227.

Gallie, D.R., D.E. Sleat, J.W. Watts, P.C. Turner and T.M.A. Wilson. 1987. A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. **Nucleic Acids Res.** 15: 8693-8711.

_____ and V. Walbot. 1992. Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. **Nucleic Acids Res.** 20: 4631-4638.

Garcia, J.A., S. Lain, M.T. Cervera, J.L. Riechmann and M.T. Martin. 1990. Mutational analysis of the plum pox potyvirus polyprotein processing by the NIa protease in *Escherichia coli*. **J. Gen. Virol.** 71: 2773-2779.

_____, J.L. Richmann and S. Lain. 1989a. Proteolytic activity of the plum pox potyvirus NIa-like protein in *Escherichia coli*. **Virol.** 170: 362-369.

_____, _____ and _____. 1989b. Artificial cleavage site recoqnized by plum pox potyvirus protease in *Escherichia coli*. **Virol.** 63: 2457-2460.

_____, _____ and _____. 1989c. Proteolytic activity of the plum pox potyvirus NIa-protein on excess of natural and artificial substrates in *Escherichia coli*. **FEBS Lett.** 257: 269-273.

Garzon-Tiznado, J.A., I. Torres-Pacheco, J.T. Ascencio-Ibanez, L. Herrera-Estrella and R.F. Rivera-Bustamante. 1993. Inoculation of peppers with infectious clones of a new geminivirus by a biolistic procedure. **Phytopathol.** 83: 514-521.

Ghabrial, S.A., H.A. Smith, T.D. Parks and W.G. Dougherty. 1990. Molecular genetic analyses of the soybean mosaic virus NIa proteinase. **J. Gen. Virol.** 71: 1921-1927.

Giachetti, C. and B.L. Semler. 1991. Role of a viral membrane polypeptide in strand-specific initiation of polyvirus RNA synthesis. **J. Virol.** 65: 2647-2654.

Gibbs, A. and A. Mackenzie. 1997. A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. **J. Virol. Methods** 63: 9-16.

Gonsalves, D. 1998. Control of papaya ringspot virus in papaya: a case study. **Ann. Rev. Phytopathol.** 36: 415-437.

_____ and M. Ishii. 1980. Purification and serology of papaya ringspot virus. **Phytopathol.** 70: 1028-1032.

Gonzalez, J.M., P. Penzes, F.Almazan, E. Calvo and L. Enjuanes. 2002. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. **J. Virol.** 76: 4655-4661.

Gotz, R. and E. Maiss. 2002. The complete sequence of the genome of Cocksfoot streak virus (CSV), a grass infecting *Potyvirus*. **Arch. Virol.** 147: 1573-1583.

Green, M.R., T. Maniatis and D.A. Melton. 1983. Human *β*–globin pre-mRNA synthesized in vitro is accurately spiced in *Xenopus oocyte* nuclei. **Cell** 32: 681-694.

Hagen, L.S., H. Lot, C. Godon, M. Tepfer and M. Jacquemond. 1994. Infection of *Theobroma cacao* using cloned DNA of cacao swollen shoot virus and particle bombardment. **Phytopathol.** 84: 1239-1243.

Haldeman-Cahill, R., J.A. Daros and J.C. Carrington. 1998. Secondary structures in the capsid protein coding sequence and 3′-non translated region involved in amplification of the tobacco etch virus genome. **J. Virol.** 72: 4072–4079.

Hanahan, D. 1985. Techniques for transformation of *E. coli*, pp. 109-135. *In* D.M. Glover, ed. **DNA Cloning: A Practical Approach, Volume 1.** IRL Press, Oxford, United Kingdom.

Hari, V., A. Siegel, C. Rozek and W.E. Tinberlake. 1979. The RNA of tobacco etch virus contain Poly(A). **Virol.** 92: 568-571.

Hawley, D. and W. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. **Nucleic Acids Res.** 11: 2237-2255.

Hayes, R.J. and K.W. Buck. 1990. Infectious cucumber mosaic virus RNA transcribed *in vitro* from clones obtained from cDNA amplified using the polymerase chain reaction. **J. Gen. Virol.** 71: 2503-2508.

Heaton, L.A., J.C. Carrington and T.J. Morris. 1989. Turnip crinkle virus infection from RNA synthesized *in vitro*. **Virol.** 170: 214-218.

Hjulsager, C.K., O.S. Lund and I.E. Johansen. 2002. A new pathotype of *Pea seedborne mosaic virus* explained by properties of the P3-6K1- and viral genome-linked protein (VPg)-coding regions. **Mol. Plant Microbe Interaction** 15: 169-171.

Hodgeman, T.C. 1988. A conserved NTP-motif in putative helicases. **Nature** 333: 22-23.

Holt, C.A. and R.N. Beachy. 1991. *In vivo* complementation of infectious transcripts from mutant tobacco mosaic virus cDNAs in transgenic plants. **Virol.** 181: 109-117.

Holy, S. and M.G. Abouhaidar. 1993. Production of infectious *in vitro* transcripts from a full-length clover yellow mosaic virus cDNA clone. **J. Gen. Virol.** 74: 781-784.

Hong, Y. and A.G. Hunt. 1996. RNA polymerase activity catalysed by a potyvirus-encoded RNA-dependent RNA polymerase. **Virol.** 226: 146-151.

Hull, R. 2002. **Matthew's Plant Virology**. 4th ed. Academic Press, Bath, Great Britain.

Jackson, R.J. and N. Standart. 1990. Do the poly(A) tail and 3' untranslated region control mRNA translation? **Cell** 62: 15-24.

Jakab, G., E. Droz, G. Brigneti, D. Baulcombe and P. Malnoe. 1997. Infectious in vivo and in vitro transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. **J. Gen. Virol.** 78: 3141-3145.

Jensen, D.D. 1949. Papaya virus diseases with special reference to papaya ringspot. **Phytopathol.** 39: 191-211.

Jobling, S.A., C.M. Cuthbert, S.G. Rogers, R.T.Fraley and L. Gehrke. 1988. *In vitro* transcription and translational efficiency of chimeric SP6 messenger RNAs devoid of 5' vector nucleotides. **Nucleic Acids Res.** 10: 4483-4498.

_____ and L. Gehrke. 1987. Enhanced translation of chimeric messenger RNAs containing a plant viral untranslated leader sequence. **Nature** 325: 622-625.

Johansen, E. 1996. Intron insertion facilitates amplification of cloned virus cDNA in *Escherichia coli* while biological activity is reestablished after transcription *in vivo*. **Proc. Natl. Acad. Sci. USA** 93:12400-12405.

_____, O.F. Rasmussen, M. Heide and B. Borkhardt. 1991. The complete nucleotide sequence of pea seed-borne mosaic virus RNA. **J. Gen. Virol.** 72: 2625-2632.

_____, O.S. Lund, C.K. Hjulsager and J. Laursen. 2001. Recessive resistance in *Pisum sativum* and potyvirus pathotype resolved in a gene for cistron correspondence between host and virus. **J. Virol.** 75: 6609-6614.

Kadare, G. and A.L. Haenni. 1997. Virus-encoded RNA helicases. **J. Virol.** 71: 2583-2590.

Kamer, G. and P. Argos. 1984. Primary structural comparison of RNA-dependent polymerase from plant, animal and bacterial viruses. **Nucleic Acids Res.** 12: 7269-7282.

Kapoor, M., L. Zhang, P.M. Mohan and R. Padmanabhan. 1995. Synthesis and characterization of an infectious dengue virus type-2 RNA genome (New Guinea C strain). **Gene** 162: 175-180.

Kawayama, H., S. Obara, T. Morio, M. Katoh, H. Urushihara and Y. Tanaka. 2002. PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors. **Nucleic Acids Res.** 30: e2.

Kim, K.S., H.Y. Oh, S. Suranto, E. Nurhayati, K.H. Gough, D.D. Shukla and C.K. Pallaghy. 2003. Infectivity of in vitro transcript of Johnsongrass mosaic potyvirus full-length cDNA clones in maize and sorghum. **Arch. Virol.** 148: 563-574.

Klein, T.M., R. Arentzen, P.A. Lewis and S. Fitzpatrick-McElligott. 1992. Transformation of microbes, plants and animals by particle bombardment. **Bio/Technol.** 10: 286-291.

Koonin, E.V. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. **J. Gen. Virol.** 72: 2197-2206.

Lai, M.M.C. 2000. The making of infectious viral RNA: no size limit in sight. **Proc. Natl. Acad. Sci. USA** 97: 5025-5027.

Lain, S., M.T. Martin, J.L. Riechmann and J.A. Garcia. 1991. Novel catalytic activity associated with positive-strand RNA virus infection: nucleic acid-stimulated ATPase activity of the plum pox potyvirus helicaselike protein. **J. Virol**. 65: 1-6.

_____, J.L. Riechmann and J.A. Garcia. 1990. RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. **Nucleic Acids Res.** 18: 7003-7006.

_____, _____, M.T. Martin and J.A. Garcia. 1989. Homologous potyvirus and flavivirus proteins belong to a superfamily of helicase-like proteins. **Gene** 82: 357-362.

Langeveld, S.A., J.M. Dore, J. Memelink, A.F.L.M. Derk, C.I.M. van der Vlugt, C.J. Asjes and J.F. Bol. 1991. Identification of potyviruses using the polymerase chain reaction with degenerate primers. **J. Gen. Virol.** 72: 1531-1541.

Lecoq, H. 2003. Cucurbits, pp. 665-687. *In* G. Loebenstein and G. Thottappilly, eds. **Virus and Virus like Diseases of Major Crops in Developing Countries**. Kluwer Academic Publishers, Dardrecht, Netherlands.

_____, G. Dafalla, C. Desbiez, C. Wipf-Scheibel, B. Delecolle, T. Lanina, Z. Ullah and R. Grumet. 2001. Biological and molecular characterization of *Moroccan watermelon mosaic virus* and a potyvirus isolate from Eastern Sudan. **Plant Dis.** 85: 547-552.

Leiser, R.M., V. Ziegler-Graff, A. Reutenauer, E. Herrbach, O. Lemaire, H. Guilley, K. Richards and G. Jonard. 1992. Agroinfection as an alternative to insects for infecting plants with beet western yellows luteovirus. **Proc. Natl. Acad. Sci. USA** 89: 9136-9140.

Lin, S.S., R.F. Hou and S.D. Yeh. 2002. Construction of in vitro and in vivo infectious transcripts of a Taiwan strain of *Zucchini yellow mosaic virus*. **Bot. Bull. Acad. Sin.** 43: 261-269.

Lisa, V., G. Boccardo, G. D'Agostino, G. Dellavalle and M. D'Aguilo. 1981. Characterization of a potyvirus that causes zucchini yellow mosaic virus. **Phytopathol.** 71: 667-672.

_____ and H. Lecoq. 1984. Zucchini yellow mosaic virus. **CMI/AAB Descriptions of Plant Viruses No. 282**. Commonwealth Mycol Inst/Association of Applied Biologists, Kew, Surrey, UK.

Lopez-Moya, J.J. and J.A. Garcia. 2000. Construction of a stable and highly infectious intron-containing cDNA clone of plum pox potyvirus and its use to infect plants by particle bombardment. **Virus Res.** 68: 99-107.

Maia, I.G. and F. Bernardi. 1996. Nucleic acid-binding properties of a bacterially expressed potato virus Y helper component-proteinase. **J. Gen. Virol.** 77: 869-877.

Maiss, E., U. Timpe, A. Brisske-Rode, W. Jelkmann, R. Casper, G. Hinnler, D. Mattanovich and H.W.D. Katinger. 1989. The complete nucleotide sequence of plum pox virus RNA. **J. Gen. Virol.** 70: 513-524.

_____, _____, _____, D.E. Lesemann and R. Casper. 1992. Infectious *in vivo* transcripts of a Plum pox potyvirus full-length cDNA clone containing the *Cauliflower mosaic virus* 35S promoter. **J. Gen. Virol**. 73: 709-713.

Makkouk, K. and D.E. Lesemann. 1980. A severe mosaic of cucumber in Lebanon caused by watermelon mosaic virus-1. **Plant Dis.** 64: 799-801.

Mavankal, G. and R.E. Rhoads. 1991. *In vitro* cleavage at or near the N-terminus of the helper component protein in tobacco vein mottling virus polyprotein. **Virol.** 185: 721-731.

Mayo, M.A. and C.R. Pringle. 1998. Virus taxonomy-1997. **J. Gen. Virol.** 79: 649-657.

McLean, G.D., J.R. Burt, D.W. Thomas and A.N. Sproul. 1982. The use of reflective mulch to reduce the incidence of watermelon mosaic virus in Western Australia. **Crop Protection** 1: 491-496.

Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA

hybridization probes from plasmids containing a bacteriophage SP6 promoter. **Nucleic Acids Res.** 12: 7035-7056.

Merits, A., D. Guo, L. Jarvekulg and M. Sarrma. 1999. Biochemical and genetic evidence for interactions between potato A potyvirus-encoded protein P1 and P3 and proteins of the putative replication complex. **Virol.** 263: 15-22.

_____, _____and M. Sarrma. 1998. VPg coat protein and five non-structural proteins of potato A potyvirus bind RNA in a sequence-unspecific manner. **J. Gen. Virol.** 79: 3123-3127.

_____, M.L. Rajamaki, P. Lindholm, P. Runeberg-Roos, T. Kekarainen, P. Puustinen, K. Makelainen, J.P.T. Valkonen and M. Saarma. 2002. Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. **J. Gen. Virol.** 83: 1211-1221.

Milne, K.S. and R.G. Grogan. 1969. Characterization of watermelon mosaic virus strains by serology and other properties. **Phytopathol.** 59: 809-818.

Moreno, M., J.J. Bernal, I. Jimenez and E. Rodriguez-Cerezo. 1998. Resistance in plants transformed with the P1 or P3 gene of tobacco vein mottling potyvirus. **J. Gen. Virol.** 79: 2819-2827.

Murphy, J.F., P.G. Klein, A.G. Hunt and J.G. Shaw. 1996. Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. **Virol.** 220: 535-538.

_____, R.A. Rhoads, A.G. Hunt and J.G. Shaw. 1990. The VPg of tobacco etch virus RNA is the 49 kDa proteinase or the N-terminal 24 kDa part of the proteinase. **Virol.** 178: 285-288.

_____, W. Rychlik, R.A. Rhoads, A.G. Hunt and J.G. Shaw. 1991. A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. **J. Virol.** 65: 511-513.

Nameth, S.T., J.A. Dodds, A.O. Paulus and F.F. Laemmlen. 1986. Cucurbit viruses of California: an ever-changing problem. **Plant Dis.** 70: 8-11.

Ng, J.C.K. and K.L. Perry. 2004. Transmission of plant viruses by aphid vectors. **Mol. Plant Pathol.** 5: 505-511.

Nicolas, O., T.P. Pirone and G.M. Hellmann. 1996. Construction and analysis of infectious transcripts from a resistance-breaking strain of tobacco vein mottling potyvirus. **Arch. Virol.** 141: 1535-1552.

Odell, J.T., F. Nagy and N.H. Chua. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. **Nature** 313: 810-812.

Oh, C.S. and J.C. Carrington. 1989. Identification of essential residues in potyvirus poteinase HC-Pro by site directed mutagenesis. **Virol.** 173: 692-699.

Paul, A.V., J.H. van Boom, D. Filippov and E. Wimmer. 1998. protein-primed RNA systhesis by purified poliovirus RNA polymerase. **Nature** 393: 280-284.

Petty, I.T.D., B.G. Hunter and A.O. Jackson. 1988. A novel strategy for one-step cloning of full-length cDNA and its application to the genome of barley stripe mosaic virus. **Gene** 74: 423-432.

Pirone, T.P. and S. Blanc. 1996. Helper-dependent vector transmission of plant viruses. **Ann. Rev. Phytopathol.** 34: 227-247.

Pont-Kingdon, G. 1997. Creation of chimeric junctions, deletions, and insertions by PCR. **Methods Mol. Biol.** 67: 167-172.

Provvidenti, R. 1996. Diseases caused by viruses, pp. 37-45. *In* T.A. Zitter, D.L. Hophins and C.E. Thomas, eds. **Compendium of Cucurbit Diseases**. APS Press, St. Paul, Minn.

Purcifull, D.E., J.R. Edwardson, E. Hiebert and D. Gonsalves. 1984a. Papaya ring spot virus. **CMI/AAB Descriptions of Plant Viruses No. 292 (No. 84 revised)**. Commonwealth Mycol Inst/Association of Applied Biologists, Kew, Surrey, UK.

_____and E. Hiebert. 1979. Serological distinction of watermelon mosaic virus isolates. **Phytopathol.** 69: 112-116.

_____, _____and J. Edwardson. 1984b. Watermelon mosaic virus-2. **CMI/AAB Descriptions of Plant Viruses No. 293 (No. 63 revised)**. Commonwealth Mycol Inst/Association of Applied Biologists, Kew, Surrey, UK.

Puurand, U., J.P.T. Valkonen, K. Makinen, F. Rabenstein and M. Saarma. 1996. Infectious *in vitro* transcripts from cloned cDNA of the potato A potyvirus. **Virus Res.** 40: 135-140.

Quillet, L., H. Guilley, G. Jonard and K. Richards. 1989. In vitro synthesis of biologically active beet necrotic yellow vein virus RNA. **Virol.** 172: 293-301.

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

Rajamaki, M.L., J. Kelloniemi, A. Alminaite, T. Kekarainen, F. Rabenstein and J.P.T. Valkonen. 2005. A novel insertion site inside the potyvirus P1 cistron allows

expression of heterologous proteins and suggests some P1 functions. **Virol.** 342: 88-101.

Restrepo-Hartwig, M.A. and J.C. Carrington. 1992. Regulation of nuclear transport of a plant potyvirus protein by autoproteolysis. **J. Virol.** 66: 5662-5666.

_____ and _____. 1994. The tobacco etch potyvirus 6 kilodalton protein is membrane associated and involved in viral replication. **J. Virol.** 68: 2388-2397.

Revers, F., O. Le Gall, T. Candresse and A.J. Maule. 1999. New advances in understanding the molecular biology of plant/potyvirus interactions. **Mol. Plant-Microbe Interactions** 12: 367-376.

Riechmann, J.L., S. Lain and J.A. Garcia. 1989. The genome-linked protein and 5'end RNA sequence of plum pox potyvirus. **J. Gen. Virol.** 70: 935-947.

_____, _____and _____. 1990. Infectious in vitro transcripts from a plum pox potyvirus cDNA clone. **Virol.** 177: 710-716.

_____, _____ and _____. 1992. Highlights and prospects of potyvirus molecular biology. **J. Gen. Virol.** 73: 1-16.

Rizzo, T.M. and P. Palukaitis. 1990. Construction of full-length cDNA clones of cucumber mosaic virus RNAs 1, 2 and 3: Generation of infectious transcripts. **Mol. Gen. Genet.** 222: 249-256.

Robaglia, C., M. Durand-Tardif, M. Tronchet, G. Boudazin, S. Astier-Manifacier and F. Casse-Debart. 1989. Nucleotide sequence of potato virus Y (N strain) genomic RNA. **J. Gen. Virol.** 70: 935-947.

Robinson, R.W. and D.S. Deckers-Walters. 1997. **Cucurbits**. CAB International, Wallingford, UK.

Rodriguez-Cerezo, E., E.D. Ammar, T.P. Pirone and J.G. Shaw. 1993. Association of the non-structural P3 viral protein with cylindrical inclusions in potyvirus-infected cells. **J. Gen. Virol.** 74: 1945-1949.

_____, P. Gamble Klein and J.G. Shaw. 1991. A determinant of disease symptom severity is located in the 3'-terminal noncoding region of the RNA of a plant virus. **Proc. Natl. Acad. Sci. USA** 88: 9863-9867.

_____ and J.G. Shaw. 1991. Two newly detected nonstructural viral proteins in potyvirus-infected cells. **Virol.** 185: 572-579.

Ruiz-Ferrer, V., E. Goytia, B. Martinez-Garcia, D. Lopez-Abella and J. Lopez-Moya. 2004. Expression of functionally active helper component protein of *Tobacco etch potyvirus* in *Pichia pastoris*. **J. Gen. Virol.** 85: 241–249.

Saenz, P., M.T. Cervera, S. Dallot, L. Quiot, J.B. Quiot, J.L. Riechmann and J.A. Garcia. 2000. Identification of a pathogenicity determinant of plum pox virus in the sequence encoding the C-terminal region of protein P3+6K1. **J. Gen. Virol.** 81: 557-566.

Sakai, J., M. Mori, T. Morishita, M. Tanaka, K. Hanada, T. Usugi and M. Nishiguchi. 1997. Complete nucleotide sequence and genome organization of sweet potato feathery mottle virus (S strain) genomic RNA: the large coding region of the P1 gene. **Arch. Virol.** 142: 1553-1562.

Sambrook, J. and D.W. Russel. 2001. **Molecular Cloning: A Laboratory Manual**. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanchez, F., D. Martinez-Herrera, I. Aguilar and F. Ponz. 1998. Infectivity of turnip mosaic potyvirus cDNA clones and transcripts on the systemic host *Arabidopsis thaliana* and local lesion hosts. **Virus Res.** 55: 207-219.

Sanford, J.C. 1988. The biolistic process. **Trends Biotechnol.** 6: 299-302.

_____, F.D. Smith and J.A. Russell. 1993. Optimizing the biolistic process for different biological applications. **Methods in Enzymol.** 217: 483-509.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. **Proc. Natl. Acad. Sci. USA** 74: 5463-5467.

Sarnow, P. 1989. Role of 3'-end sequences in infectivity of polio virus transcripts made *in vitro***. J. Virol.** 63: 467-470.

Satyanarayana, T., S. Gowda, M.A. Ayllon and W.O. Dawson. 2003. Frameshift mutations in infectious cDNA clones of *Citrus tristeza virus*: a strategy to minimize the toxicity of viral sequences to *Escherichia coli*. **Virol.** 313: 481-491.

Shih, D.S., R. Dasgupta and P. Kaesberg. 1976. 7-Methyl-guanosine and efficiency of RNA translation. **J. Virol.** 19: 637-642.

Shimotohno, K., Y. Kodama, J. Hashimoto and K.I. Miura. 1977. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. **Proc. Natl. Acad. Sci. USA** 74: 2734-2738.

Shukla, D.D., C.W. Ward and A.A. Brunt. 1994. **The Potyviridae**. CAB International, Wallingford, UK.

Siaw, M.F.E., M. Shahabuddin, S. Ballard, J.G. Shaw and R.E. Rhoads. 1985. Identification of a protein covalently linked to the 5'-terminus of tobacco vein mottling virus RNA. **Virol.** 142: 134-143.

Sit, T.L. and M.G. Abouhaidar. 1993. Infectious RNA transcripts derived from cloned cDNA of papaya mosaic virus: effect of mutations to the capsid and polymerase proteins. **J. Gen. Virol.** 74: 1133-1140.

Skotnicki, M.L., A.M. MacKenzie and A.J. Gibbs. 1992. Turnip yellow mosaic virus variants produced from DNA clones encloding their genomes. **Arch. Virol.** 127: 25-35.

Soumounou, Y. and J.-F. Laliberte. 1994. Nucleic-acid binding properties of the P1 protein of turnip mosaic potyvirus produced in *Escherichia coli*. **J. Gen. Virol.** 75: 2567-2573.

Sriburi, R., P Keelapang, T. Duangchinda, S. Pruksakorn, N. Maneekarn, P. Malasit and N. Sittisombat. 2001. Construction of infectious dengue 2 virus cDNA clones using high copy number plasmid. **J. Virological Methods** 92: 71-82.

Sumiyoshi, H., C.H. Hoke and D.W. Trent. 1992. Infectious Japanese encephalitis virus RNA can be synthesized from *in vitro*-ligated cDNA templates. **J. Virol.** 66: 5425-5431.

Takahashi, Y. T. Takahashi and I. Uyeda. 1997. A cDNA clone to clover yellow vein potyvirus genome is highly infectious. **Virus Genes** 14: 235-243.

Thiel, V., J. Herold, B. Schelle and S.G. Siddell. 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. **J. Gen. Virol.** 82: 1273-1281.

Thornbury, D.W., C.A. Patterson, J.T. Dessens and T.P. Pirone. 1990. Comparative sequence of the helper component (HC) region of potato virus Y and a HC-defective strain, potato virus C. **Virol.** 178: 573-578.

Toyoda, H., C.F. Yang, N. Takeda, A. Nomoto and E. Wimmer. 1987. Analysis of RNA synthesis of type I poliovirus by using an in vitro molecular genetic approach. **J. Virol.** 61: 2816-2822.

Turpen, T. 1989. Molecular cloning of potato virus Y genome: nucleotide sequence homology in non-coding regions of potyviruses. **J. Gen. Virol.** 70: 1951-1960.

Urcuqui-Inchima, S., A.-L. Haenni and F. Bernardi. 2001. Potyvirus proteins: a wealth of functions. **Virus Res.** 74: 157-175.

_____, I.G. Maia, P., G. Drugeon, A.L. Haenni and F. Bernardi. 1999a. Effect of mutations within the Cys-rich region of potyvirus helper component-proteinase on self-interaction. **J. Gen. Virol.** 80: 2809-2812.

_____, J. Walter, G. Drugeon, S. German-Retana, A.L. Haenni, T. Candresse, F. Bernardi and O. LeGall. 1999b. Potyvirus helper component-proteinase self interaction in the yeast two-hybrid system and delineation of the interaction domain involved. **Virol.** 258: 95-99.

Van Bokhoven, H., J. Verver, J. Wellink and A. Van Kammen. 1993. Protoplasts transiently expressing the 200K coding sequence of cowpea mosaic virus B-RNA support replication of M-RNA. **J. Gen. Virol.** 74: 2233-2241.

Vancanneyt, G., R. Schmidt, A. O'Connor-Sanchez, L. Willmiter and M. Rocha-Sosa. 1990. Construction of an intron-containing marker gene: splicing of the intron in transgenic plants and its use in monitoring early events in Agrobacterium-mediated plant transformation. **Mol. Gen. Genet.** 220: 245-250.

Varrelmann, M. and R. Maiss. 2000. Mutations in the coat protein gene of plum pox virus suppress particle assembly, heterollogous encapsidation and complementation in transgenic plants of *Nicotiana benthamiana*. **J. Gen. Virol.** 81: 567-576.

Veidt, I., S.E. Bouzoubaa, R.M. Leiser, V. Ziegler-Graff, H. Guilley, K. Richards and G. Jonard. 1992. Synthesis of full-length transcripts of beet western yellows virus RNA: Messenger properties and biological activity in protoplasts. **Virol.** 186: 192-200.

Verchot, J. and J.C. Carrington. 1995. Evidence that the potyvirus P1 proteinase functions in *trans* as an accessory factor for genome amplification. **J. Virol.** 69: 3668-3674.

_____, K.L. Herndon and J.C. Carrington. 1992. Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. **Virol.** 190: 298-306.

_____, E.V. Koonin and J. Carrington. 1991. The 35-kDa protein from N-terminus of potyviral polyprotein functions as a third virus-encoded proteinase. **Virol.** 185: 527-533.

Viry, M., M.A. Serghini, F. Hans, C. Ritzenthaler, M. Pinck and L. Pinck. 1993. Biologically active transcripts from cloned cDNA of genomic grapevine fanleaf nepovirus RNAs. **J. Gen. Virol.** 74: 169-174.

Voet, D. and J.G. Voet. 1990. **Biochemistry.** John Wiley & Sons, Inc., New York, USA.

Walker, J.E., M. Saraste, M.J. Runswick and N.J. Gay. 1982. Distantly related sequences in the α- and β-subunits of ATP synthetase, myosin, kinases and

other ATP-requiring enzymes and a common nucleotide binding fold. **EMBO J.** 1: 945-951.

Wang, C.H. and S.D. Yeh. 1997. Divergence and conservation of the genomic RNAs of Taiwan and Hawaii strains of papaya ringspot potyvirus. **Arch. Virol.** 142: 271-285.

_____, G. Powell, J. Hardie and T.P. Pirone. 1998. Role of the helper component in vector-specific transmission of potyviruses. **J. Gen. Virol.** 79: 1519-1524.

Warrens, A.N., M.D. Jones and R.I. Lechler. 1997. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. **Gene** 186: 29-35.

Webb, R.E., G.W. Bohn and H.A. Scott. 1965. Watermelon mosaic virus 1 and 2 in southern and western cucurbit production areas. **Plant Dis. Rept.** 49: 532-535.

_____ and H.A. Scott. 1965. Isolation and identification of watermelon mosaic virus 1 and 2. **Phytopathol.** 55: 895-900.

Weiland, J.J. and T.W. Dreher. 1989. Infectious TYMV RNA from cloned cDNA: Effects *in vitro* and *in vivo* of point substitutions in the initiation codons of two extensively overlapping ORFs. **Nucleic Acids Res.** 17: 4675-4687.

Wimmer, E. 1982. Genome-linked proteins of viruses. **Cell** 28: 199-201.

Yamaya, Y., M. Yoshioka, T. Meshi, Y. Okada and T. Ohno. 1988. Expression of tobacco mosaic virus RNA in transgenic plants. **Mol. Gen. Genet.** 211: 520-525.

Yamshchikov, V., V. Mishin and F. Cominelli. 2001. A new strategy in design of (+) RNA virus infectious clones enabling their stable propagation in *E.coli*. **Virol.** 281: 272-280.

Yang, S.J., F. Revers, S. Souche, H. Lot, O.L. Gall, T. Candresse and J. Dunez. 1998. Construction of full-length cDNA clones of lettuce mosaic virus (LMV) and the effects on their viability in *Escherichia coli* and on their infectivity to plants. **Arch. Virol.** 143: 2443-2451.

Yeh, S.D. and D. Gonsalves. 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. **Virol.** 143: 260-270.

_____, _____and R. Provvidenti. 1984. Comparative studies on host range and serology of papaya ringspot virus and watermelon mosaic virus 1. **Phytopathol.** 74: 1081-1085.

_____, F.J. Jam, C.H. Chiang, T.J. Doong, M.C. Chen, P.H. Chung and H.J. Bau. 1992. Complete nucleotide sequence and genetic organization of papaya ring spot virus RNA. **J. Gen. Virol.** 73: 2531-2541.

Yolov, A.A. and Z.A. Shabarova. 1990. Constructing DNA by polymerase recombination. **Nucleic Acids Res.** 18: 3983-3986.

Yon, J. and M. Fried. 1989. Precise gene fusion by PCR. **Nucleic Acids Res.** 17: 4895.

Young, M.J., L. Kelly, P.J. Larkin, P.M. Waterhouse and W.L. Gerlach. 1991. Infectious *in vitro* transcripts from a cloned cDNA of barley yellow dwarf virus. **Virol.** 180: 372-379.

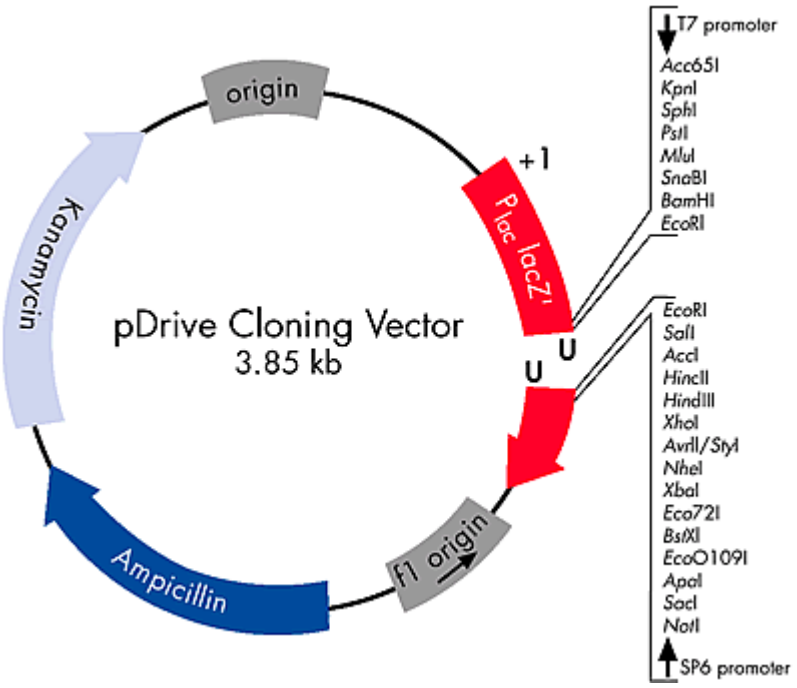
Yount, B., K.M. Curtis and R.S. Baric. 2000. Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. **J. Virol.** 74: 10600-10611.

_____, M.R. Denison, S.R. Weiss and R.S. Baric. 2002. Systematic assembly of a full-length infectious cDNA clone of mouse hepatitis virus strain A59. **J. Virol.** 76: 11065-11078.

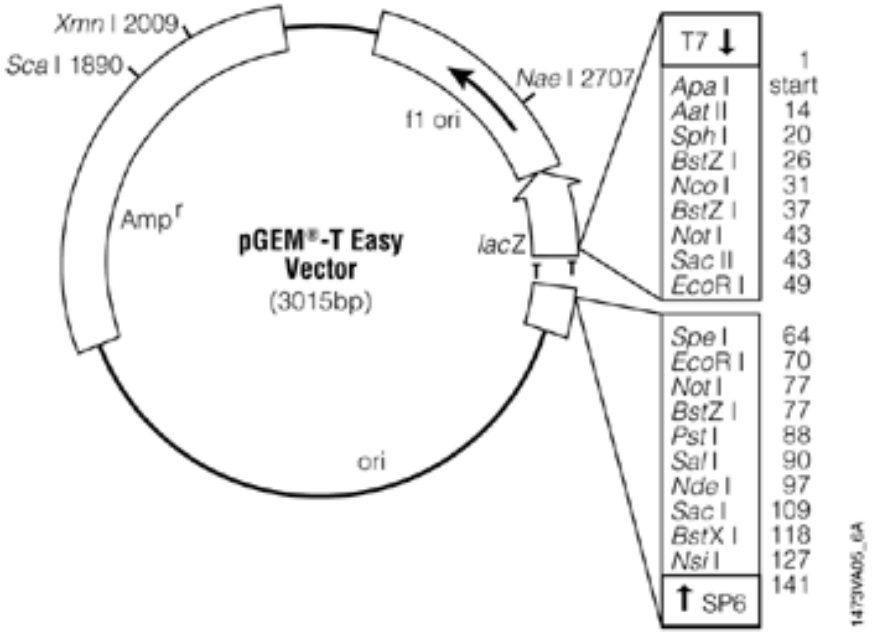
APPENDICES

Appendix A

Physical maps of plasmid DNA cloning vectors

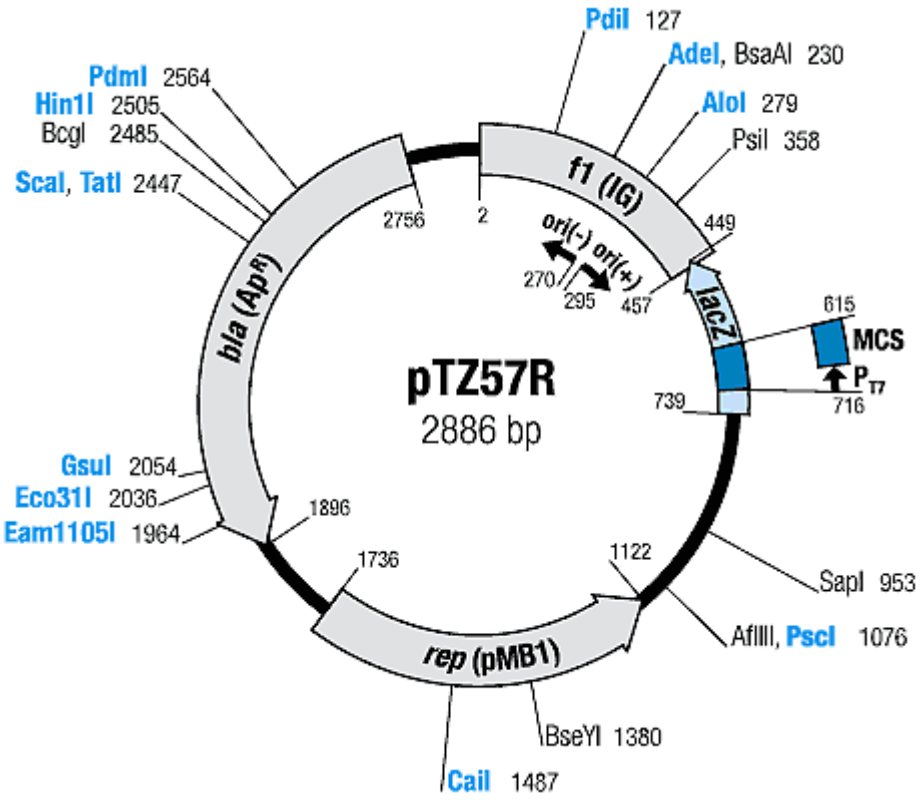


Appendix Figure 1 Map of pDrive cloning vector from QIAGEN

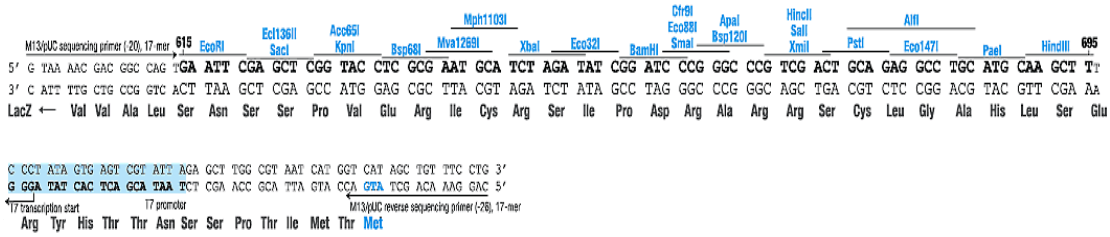


Appendix Figure 2 Map of pGEM®-T Easy cloning vector from Promega

A



B



Appendix Figure 3 Map of pTZ57R cloning vector from Fermentas

Appendix B

Weight values of nucleotide or amino acid residues
used for comparing genetic distances

Appendix Table 1 Weight values of nucleotide residues used for comparing genetic distances between pairs of nucleotide sequences using the program MegAlign in the DNAStar biocomputing software (DNAStar, Inc.)

	A	**C**	**G**	**T**	**R**	**Y**	**W**	**S**	**M**	**K**	**B**	**D**	**H**	**V**	**X**	**-**	
A	4																**A**
C	0	4															**C**
G	0	0	4														**G**
T	0	0	0	4													**T**
R	0	0	0	0	2												**R**
Y	0	0	0	0	0	2											**Y**
W	0	0	0	0	0	0	2										**W**
S	0	0	0	0	0	0	0	2									**S**
M	0	0	0	0	0	0	0	0	2								**M**
K	0	0	0	0	0	0	0	0	0	2							**K**
B	0	0	0	0	0	0	0	0	0	0	1						**B**
D	0	0	0	0	0	0	0	0	0	0	0	1					**D**
H	0	0	0	0	0	0	0	0	0	0	0	0	1				**H**
V	0	0	0	0	0	0	0	0	0	0	0	0	0	1			**V**
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		**X**
-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**-**
	A	**C**	**G**	**T**	**R**	**Y**	**W**	**S**	**M**	**K**	**B**	**D**	**H**	**V**	**X**	**-**	

Appendix Table 2 Weight values of amino acid residues used for comparing genetic distances between pairs of amino acid sequences using the program MegAlign in the DNAStar biocomputing software (DNAStar, Inc.)

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	14																				C
S	-1	6																			S
T	-5	2	7																		T
P	-6	1	-1	10																	P
A	-5	2	2	1	6																A
G	-8	1	-3	-3	1	8															G
N	-8	2	0	-3	-1	-1	7														N
D	-11	-1	-2	-4	-1	-1	4	8													D
E	-11	-2	-3	-3	0	-2	1	5	8												E
Q	-11	-3	-3	-1	-2	-5	-1	1	4	9											Q
H	-6	-4	-5	-1	-5	-7	2	-1	-2	4	11										H
R	-6	-1	-4	-2	-5	-8	-3	-6	-5	1	1	10									R
K	-11	-2	-1	-4	-4	-5	1	-2	-2	-1	-3	3	8								K
M	-11	-4	-2	-6	-3	-8	-5	-8	-6	-2	-7	-2	1	13							M
I	-5	-4	-1	-6	-3	-7	-4	-6	-5	-5	-7	-4	-4	2	9						I
L	-12	-7	-5	-5	-5	-8	-6	-9	-7	-3	-5	-7	-6	4	2	9					L
V	-4	-4	-1	-4	0	-4	-5	-6	-5	-5	-6	-6	-6	1	5	1	8				V
F	-10	-5	-6	-9	-7	-8	-6	-11	-11	-10	-4	-7	-11	-2	0	0	-5	12			F
Y	-2	-6	-6	-11	-6	-11	-3	-9	-7	-9	-1	-10	-10	-8	-4	-5	-6	6	13		Y
W	-13	-4	-10	-11	-11	-13	-8	-13	-14	-11	-7	1	-9	-11	-12	-7	-14	-2	-2	19	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	