Sequence Analysis of 2b Gene of Australian CMV Isolates

Emy Sulistyowati
Indonesian Tobacco and Fiber Crops Research Institute

ABSTRAK

Analisis Sekuen Gen 2b pada Isolat CMV asal Australia. Virus gen 2b dari *Cucumber Mosaic Virus* (CMV) dilaporkan sebagai *suppressor gene silencing*, determinan pergerakan sistemik, dan determinan hipervirylen spesifik. Penelitian dilakukan untuk menganalisis tujuh isolat CMV asal Australia serta untuk sekuensing dan menganalisis gen 2b dari isolat tersebut. Selanjutnya dilakukan pengelompokan (subgrouping) ketujuh isolat tersebut berdasarkan sekuen gen 2b, sekuen protein, dan ukuran dari gen 2b ORF. Hasil penelitian menunjukkan bahwa ketujuh isolat CMV tersebut dapat dikelompokkan menjadi tiga subgrup, yaitu subgrup IA (CMV 207, CMV 237, CMV 242, dan Twa CMV), subgrup IB (CMV 243), dan subgrup II (CMV 241 dan CMV 245). Selanjutnya, gen 2b dalam subgrup Ia and II ukuran dan sekuannya bersifat *conserved* dan mengandung suatu *nuclear localisation signal* (NLS).

Kata kunci: *Cucumber Mosaic Virus* (CMV), *isolate subgrouping*, RT-PCR.

ABSTRACT

Sequence Analysis of 2b Gene of Australian CMV Isolates. Gene 2b of the Cucumber Mosaic Virus (CMV) was reported to function as a suppressor of the gene silencing, a systemic movement determinant, and a host-specific hypervirulence determinant. A study was done to analyze 7 Australian CMV isolates to sequence and analyze their 2b genes. Subgrouping of the 7 CMV isolates has been deduced by analysing the 2b gene sequences, the protein sequences, and the size of the 2b ORF. The CMV isolates were divided into three subgroups, i.e., subgroup IA (CMV 207, CMV 237, CMV 242, and Twa CMV), subgroup IB (CMV 243), and subgroup II (CMV 241 and CMV 245). Moreover, the 2b genes of the Australian CMV isolates within subgroups IA and II were conserved in their sizes and sequences, and contained a nuclear localisation signal (NLS).

Key words: Cucumber Mosaic Virus (CMV), isolate subgrouping, RT-PCR.

INTRODUCTION

Cucumber mosaic virus (CMV) has been reported to infect approximately 1000 species and causing a wide range of symptoms (Palukaitis *et al.* 1992). Seven Australian CMV isolates have been studied, and both biological and molecular characterizations of those CMV isolates confirmed that CMV subgroups IA, IB, and II are present in Australia.

Virulence of plant viruses indirectly reflects their agro-economic properties. It depends upon their ability to move locally and systemically, to evade the host's defensive mechanisms, and to cause disease symptoms by interference with the host metabolism. However, the exact mechanism and molecular interactions by which plant viruses cause disease is poorly understood. The CMV 2b gene has been reported to function as a suppressor of gene silencing (Brignetti *et al.* 1998), as a systemic movement determinant (Ding *et al.* 1995, Ji and Ding 2001), and as a host-specific hypervirulence determinant (Ding *et al.* 1996).

Therefore, the variety of disease symptoms and experimental host range, may be influenced by differences in their 2b genes.

The CMV genome consists of three genomic RNAs (RNAs 1, 2, and 3) and two subgenomic RNAs (RNAs 4 and 4A) (Palukaitis *et al.* 1992, Ding *et al.* 1994). RNAs 1 and 2 encode the 1a and 2a proteins, respectively, which are involved in replication. RNA 3 codes for the 3a protein, which is functional in cell-to-cell movement, and the coat protein (CP) which is expressed from RNA 4, a subgenomic RNA derived from RNA 3. The CP is involved in RNA encapsidation and cell-to-cell movement. The fifth protein encoded by the CMV genome is the 2b protein, which is expressed *in vivo* from RNA 4A, a subgenomic RNA derived from RNA 2 (Ding *et al.* 1995a). Infectious clones of RNAs 1, 2, and 3 of several CMV strains, including Q, Sny, and Fny, as well as *Tomato aspermy virus* (TAV) have been engineered (Ding *et al.* 1995b, Rizzo and Palukaitis 1990, Choi *et al.* 2002). The work described in this report is based on the use of infectious clones of Q-CMV in the pCass vector, which enables the use of plasmid DNA as infectious inoculum (Ding *et al.* 1995).

The 2b gene is the latest identified gene of CMV, and was first isolated from Q-CMV (Ding *et al.* 1994). The gene is located at the 3' end of RNA 2, and overlaps with the 3' end of the 2a gene, although it is out of frame of the 2a open reading frame (ORF). The amino terminus of 2b protein, which overlaps with the 2a ORF is highly hydrophilic. The size of 2b ORF in different cucumoviruses varies from 95 to 110 amino acids, and the translational product is approximately 11-13 kDa. To date, 2b protein has been reported to function as a suppressor of post-transcriptional gene silencing (PTGS) (Brignetti *et al.* 1998), as a determinant for long-distance movement (Ding *et al.* 1995), as a virulence determinant (Ding *et al.* 1994), and as a determinant of host specificity (Shi *et al.* 2002).

Further analysis of 2b gene sequence and protein function may shed light on its association with symptom expression. Isolation of 2b genes of those CMV isolates followed by sequence comparison may reveal putative links between sequence variation and observed host range and symptomatology. This paper describes works on the isolation and analysis of 2b genes of Australian CMV isolates.

EXPERIMENTAL PROCEDURE

Isolation of 2b Genes of Australian CMV isolates

The 2b gene was amplified from 7 CMV isolates (207, 237, 241, 242, 243, 245, and Twa) by reverse transcriptase-polymerase chain reaction (RT-PCR). Purified CMV genomic RNA provided by Dr. Neena Mitter, were used as template. Primers were designed based on the consensus sequence of RNA 2 of CMV subgroups IA, IB, and II, respectively following alignment of published sequences of CMV RNA 2 using Sequencher 3.1.1 software (Table 1).

The SuperScript™ One-Step RT-PCR kit with Platinum® *Taq* DNA Polymerase (Invitrogen™ Life Technologies) was used in a model 9700 thermal cycler (PE Applied Biosystem, Foster City, USA). The 50 μL RT-PCR reaction contained approximately 500 ng purified RNA, 1x reaction mix, 0.2 mM of each forward (F) and reverse (R) primers, and 1 μL RT/Platinum *Taq* mix. The RT-PCR temperature profile was 72°C for 1 minute, 50°C for 30 minutes, and 94°C for 2 minutes, 40 cycles of 94°C for 30 seconds, 50°C for 1 minute, and

Table 1. Sequences of primers used to amplify CMV 2b genes.

Subgroup	Primer	Sequences (5'-3')	Reference
IA	ES22a (F)	GGG TTG AGC GTG TAA ATT CCA A (nt 2381-2402)	This study
	ES6a (R)	TGG TCT CCT TTT GGA GGC CCC (nt. 3031-3051)	This study
IB	ES22b (F)	AGC AGC GAA AGA AGA AAG ATG GAA T (nt. 2421-2425)	This study
	RP3 (R)	AAG GAG ACC ACT GCA GGG (3' end of RNA 2)	N. Mitter, *unpublished*
II	SD22 (F)	ACC GTT AAG AAG AAG AAG A (nt. 2394-2412)	Ding *et al.* (1994)
	SD6 (R)	TGG TCT CCT TAT GGA GAA CCT GTG G (nt. 3016-3040)	Ding *et al.* (1994)

72^{o}C for 1 minute, and additional final extension at 72^{o}C for 10 minutes. The expected size of RT-PCR products was ~ 650 bp**.**

The RT-PCR products were subjected to agarose gel electrophoresis. The DNA bands containing the 2b gene were extracted using a Gel Extraction Kit (Qiagen) and cloned into the pCR 2.1 TOPO TA Cloning kit (Invitrogen) following the manufacturer's instructions. The ligation mix was used to transform Oneshot® chemically-competent *E. coli* cells (Invitrogen) by 30 seconds heat shocking at 42^{o}C. Blue-white colony screening was done on solid Luria Bertani (LB) plates containing 50 mg/l ampicillin and 40 μl of 2% X-gal. White colonies were subjected to PCR colony screening using appropriate primers. Plasmid DNA was isolated from overnight (o/n) cultures of PCR-positive colonies, and analysed by *Hind*III digestion to release the inserts. DNA of three positive colonies were then subjected to automatic nucleic acid sequencing using M13 (F) and M13 (R) primers. Chromatograms and electronic sequence files were checked, aligned and edited manually using Sequencher 3.1.1 software to obtain a consensus sequence for the 2b gene of each isolate.

Analysis of 2b Gene Sequences

Consensus sequences of 2b genes and deduced proteins using the 'Translate' program were aligned to determine variation in nucleotide and amino acid sequences among the 7 CMV isolates. Alignment of nucleotide and amino acid sequences of 2b genes were done using ClustalW. The resulting multiple sequence files (*.msf) was used to determine the level of sequence identity using 'Homology' software. Sequence analysis was done using software available on the ANGIS network.

Grouping of CMV Isolates Based on the 2b Sequences

Published RNA 2 sequences of CMV isolates which known subgroups were downloaded from the genebank from which the 2b sequences were then extracted. The sequences were aligned and were then used to deduce the subgrouping of seven Australian CMV isolates by using the 'rooting' approached.

RESULT AND DISCUSSION

Analysis of 2b Gene Sequences of Australian CMV Isolates

Based on phylogenetic analyses of RNA 3 sequences (Sulistyowati *et al*. 2004), CMV isolates 207, 237, 242, and Twa are members of subgroup IA, whereas isolates 241 and 245 are members of subgroup II. CMV 243 is a subgroup IB isolate. Alignment of the amino acid

(Figure 1) and nucleotide (Figure 2) sequences of 2b ORF from the seven Australian CMV isolates showed a high level of sequence identity within 2b from the same CMV subgroup (Figure 1 and 3).

Large stretches of 2b protein sequence of all isolates from either CMV subgroup IA or II were identical and unique to the 2b protein of each subgroup. The 2b protein of subgroup IB isolate 243 shared more identical sequence with subgroup IA than with subgroup II isolates. Three amino acid boxes were highly conserved across most 2b sequences (Figure 1). All deduced 2b protein sequences contain a nuclear localization signal (NLS) (Figure 1).

The three CMV subgroups were evident not only from sequence homology (Figure 3) and phylogenetic analysis (data not shown), but also from the size of the 2b ORF. 2b ORF of all subgroup IA isolates encompassed 333 nucleotides, that of subgroup IB was 336 nucleotides, and that of subgroup II were 303 nucleotides long. The 2b ORF of subgroup IA isolates had a one C-terminal amino acid deletion, compared to the other isolates, while all subgroup II isolates had an 11 amino acids deletion in the core of the 2b protein. The host range and symptomatology of the CMV isolates used in this experiment have been

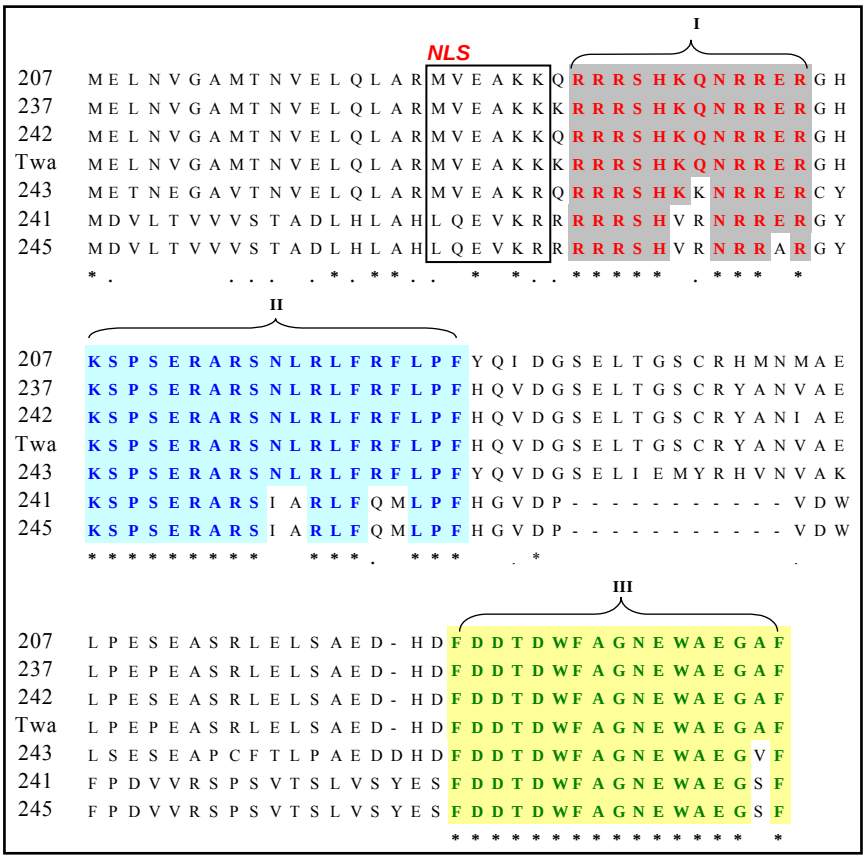

Figure 1. ClustalW multiple amino acid sequence alignment of the 2b ORF of seven Australian CMV isolates. Asterisks and dots indicate identical and similar residues, respectively. Dashes represent gaps introduced for optimal alignment. Conserved boxes are shown in colour. Nuclear localization signals (NLS) are shown in opened-box.

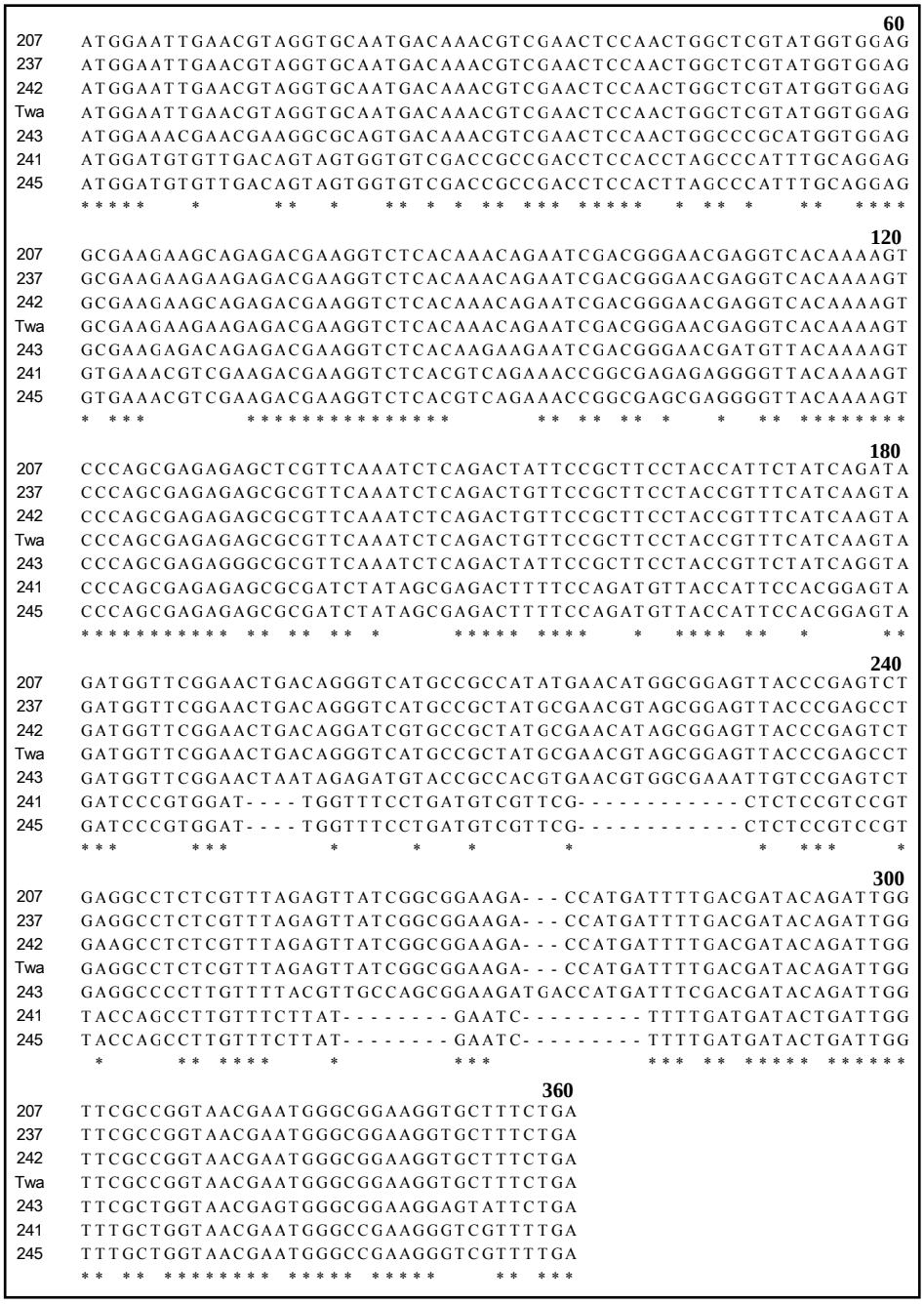

Figure 2. ClustalW multiple nucleotide sequence alignment of the 2b ORF of seven Australian CMV isolates. Asterisks indicate identical nucleotides, and dashes represent gaps introduced for optimal alignment.

differentiated by its unique differentiating disease symptoms and host differentials, such as Nicotiana species, sweet corn, capsicum, and tomato.

207	237	242	Twa	243	241	245	
	95.5	97.3	95.5	85.6	56.8	55.9	207
		98.2	100	81.1	58.6	57.7	237
			98.2	82.9	57.7	56.8	242
				81.1	58.6	57.7	Twa
					54.0	53.2	243
						99	241

Figure 3. Amino acid sequence identity (%) between the 2b ORF of seven Australian CMV isolates. Values are color-coded in pink for level of identity >90%, green for values between 60% and 90%, and black for values <60%. Subgroups IA, IB, and II isolates are shaded in pink, green, and yellow, respectively.

Figure 4 represents the subgrouping of Australian CMV isolates deduced from the 2b sequences. Result of this analysis agrees with the above subgrouping of those isolates based on the protein sequences and the size of the 2b ORF. By involving 2b sequences of CMV isolates of subgroup, it can be summarised that Twa CMV, CMV 207, CMV 237, and CMV 242 are members of CMV subgroup Ia, whereas CMV 243 is a member of CMV subgroup Ib. In addition, CMV 241 and CMV 245 are in the same subgroup of Q-CMV, a subgroup II CMV. These also agrees with that based on the RNA 3 sequences (Sulistyowati *et al.* 2004).

The size of 2b gene sequences of the Australian CMV isolates and their amino acid sequence homology indicated their subgroup memberships, which mirrored those deduced by the phylogenetic analysis based on the RNA 3 (Sulistyowati *et al.* 2004). Szilassy *et al.* (1999) and Roossinck (2002) also reported the presence of three CMV subgroups based on analysis of the 2b gene and complete genome sequences.

Three blocks of conserved amino acids were identified from the alignment of 2b sequences of seven Australian CMV isolates. Block I consisted of 12 residues, which were arginine rich. Such arginine rich sequences are typical of NLS (Alberts *et al.* 1989), which direct the 2b protein to the nucleus (Mayers *et al.* 2000). The location of NLS were the same for all 2b sequences, but the NLS differed slightly between subgroups. NLS of CMV 241 and 245, the subgroup II isolates, were identical to that of Q-CMV, ^{22}KRRRRR^{27} in sequence and position (Lucy *et al.* 2000). The CMV 243 NLS, ^{22}KRQRRR^{27}, was identical to that of K-CMV (Lucy *et al.* 2000). CMV isolates 207 and 242 have identical NLS' as Fny-CMV, ^{22}KKQRRR^{27}, whereas CMV 237 and Twa NLS were identical to that of WAII-CMV (Lucy *et al.* 2000). It is unknown whether this difference will affect the efficiency of 2b protein of different CMV subgroups to act as a suppressor of gene silencing. The conserved stretches of 2b sequence, which appear unique for the subgroups may be responsible for the host-specific 2b effect on virulence. This sould be investigated by site-directed mutagenesis in future experiments.

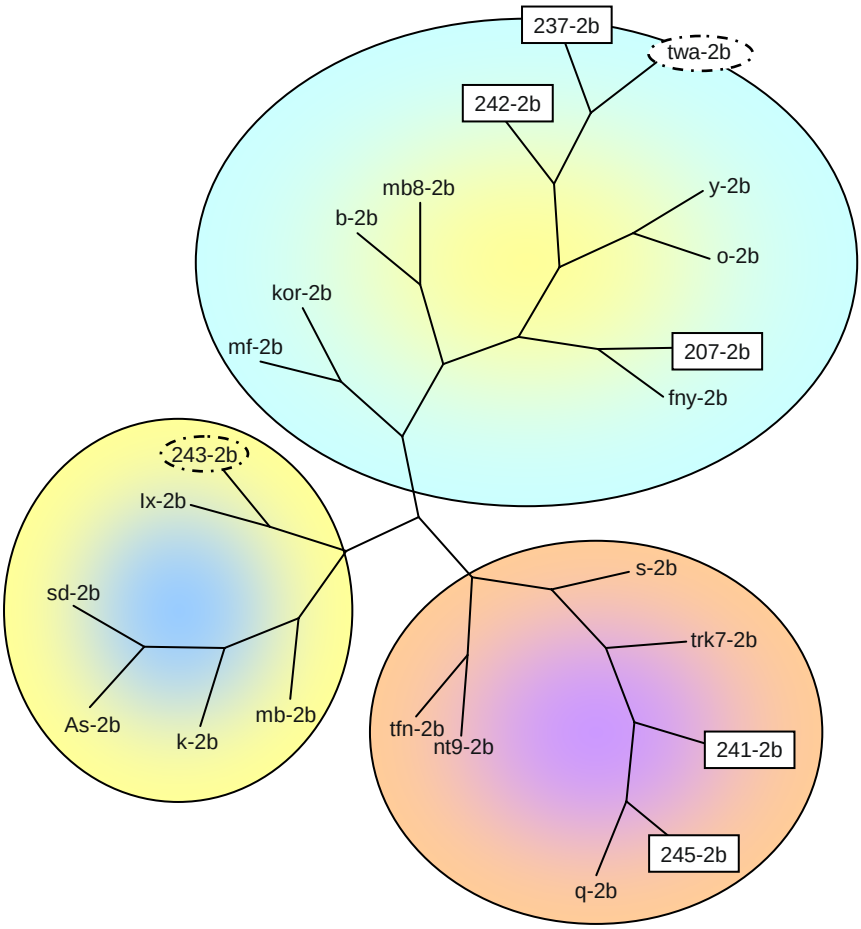

Figure 4. Root grouping of CMV isolates based on the 2b sequences.

CONCLUSIONS

The 2b genes of Australian CMV isolates within subgroups IA and II are conserved in size and sequence and contain a nuclear localisation signal (NLS). The subgrouping of 7 Australian CMV isolates can be deduced by analysing the 2b gene sequences.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. 1989. Molecular Biology of the Cell. Garland Publishing, Inc. New York.

Brigneti, G., O. Voinnet, W.X. Li, L.H. Ji, S.W. Ding, and D.C. Baulcombe. 1998. Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. EMBO J. 17:6739-6746.

Choi, S.K., J.K. Choi, K.H. Ryu, and W.M. Park. 2002. Generation of infectious cDNA clones of a Korean strain of *Tomato aspermy virus*. Mol. Cells. 13:52-60.

Ding, B., Q. Li, L. Nguyen, P. Palukaitis, and W.J. Lucas. 1995. *Cucumber mosaic virus* 3a movement protein potentiates cell-to-cell trafficking of CMV RNA in tobacco plants. Virology 207:345-353.

Ding, S-W., B.J. Anderson, H.R. Haase, and R.H. Symons. 1994. New overlapping gene encoded by *Cucumber mosaic virus* gene. Virology 198:593-601.

Ding, S-W., W-X. Li, and R.H. Symons. 1995a. A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long-distance virus movement. EMBO J. 14:5762-5772.

Ding, S-W., J.P. Rathjen, W-X. Li, R. Swanson, H. Healy, and R.H. Symons. 1995b. Efficient infection from cDNA clones of *Cucumber mosaic cucumovirus* RNAs in a new plasmid vector. J. Gen. Virol. 76:459-464.

Ji, L-H. and S-W. Ding. 2001. The suppressor of transgene RNA silencing encoded by *Cucumber mosaic virus* Interferes with salicylic acid-mediated virus resistance. MPMI 14(6):715-724.

Lucy, A.P., H-S. Guo, W-X. Li, and S-W. Ding. 2000. Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. EMBO J. 19:1672-1680.

Mayers, C.N., P. Palukaitis, and J.P. Carr. 2000. Subcellular distribution analysis of the *Cucumber mosaic virus* 2b protein. J. Gen. Virol. 81:219-226.

Palukaitis, P., M.J. Roossinck, R.G. Dietzgen, and R.I.B. Francki. 1992. *Cucumber mosaic virus*. Adv. Virus Res. 41:281-348.

Rizzo, T.M. and P. Palukaitis. 1990. Construction of full-length *Cucumber mosaic virus* RNAs 1, 2, and 3: generation of infectious RNA transcripts. Mol. Gen. Genet. 222:249-256.

Roossinck, M.J. 2002. Evolutionary history of *Cucumber mosaic virus* deduced by phylogenetic analyses. J. Virol. 76(7):3382-3387.

Shi, B-J., P. Palukaitis, and R.H. Symons. 2002. Differential virulence by strains of *Cucumber mosaic virus* is mediated by the 2b gene. Mol. Plant-Microbe Interact. 15:947-955.

Sulistyowati, E., N. Mitter, S. Bastiaan-Net, M.J. Rossinck, and R.G. Dietzgen. 2004. Host range, symptom expression, and RNA 3 sequaence analysis of six Australian strains of *Cucumber mosaic virus.* Australasian Plant Pathology 33:505-512.

Szilassy, D., K. Salanki, and E. Balazs. 1999. Molecular evidence for the existence of two distinct subgroups in *Cucumber mosaic cucumovirus*. Virus Genes 18:221-227.