Bot. Bull. Acad. Sin. (2001) **42:** 243-250

Complete genome sequence and genetic organization of a Taiwan isolate of *Zucchini yellow mosaic virus*

Shih-Shun Lin[1], Roger F. Hou[1], and Shyi-Dong Yeh[2,*]

[1]*Graduate Institute of Agricultural Biotechnology and* [2]*Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan 402, Republic of China*

(Received March 6, 2001; Accepted May 28, 2001)

Abstract. The complete nucleotide sequence of the RNA genome of a Taiwan isolate of *Zucchini yellow mosaic virus* (ZYMV TW-TN3) was determined from five overlapping cDNA clones 9591 nucleotides in length excluding the poly (A) tail. Computer analysis of the sequence revealed a large open reading frame (ORF) that encodes a polyprotein of 3080 amino acids. Comparison of the gene products of TW-TN3 with those of the reported California (CA), Reunion Island (RU), and Singapore (S) isolates of ZYMV revealed that P1 protein is most variable, with amino acid identities of 59.0-93.2%. The 5′ untranslated region (UTR) of TW-TN3 shares 61.6-83.3% nucleotide identities, and the 3′ UTR shares 90.4-95.7% nucleotide identities, with those of the other isolates. A phylogenetic tree derived from the sequences of P1 proteins of TW-TN3 and the other six reported ZYMV isolates revealed four major genotypes. TW-TN3 was classified in genotype I, and US isolates were in genotype II. The Reunion Island and Singapore isolates were separated into genotypes III and IV, respectively. The distance relationships of P1 protein of genotype I were closer to genotype II, indicating that the Taiwan and US isolates may evolve from the same ancestor. Analyses on the cleavage sites of the C-terminal halves of the polyproteins of TW-TN3, CA, RU, and S isolates revealed that NIa protease cleaves at Q-S, and E-S dipeptide sequences, with a consensus sequence of V-x-x-(Q, E)/(S, A, G). The genetic organization of TW-TN3 was concluded as Vpg/5′ leader/P1 (36 kDa)/HC Pro (52 kDa)/P3 (40 kDa)/6K1 (6 kDa)/CIP (71 kDa)/6K2 (6 kDa)/NIa-Vpg (22 kDa)/NIa-Pro (27 kDa)/NIb (60 kDa)/CP (31 kDa)/3′ UTR-poly(A) tract.

Keywords: Genetic organization; Genome sequence; ZYMV isolates.

Introduction

Zucchini yellow mosaic virus (ZYMV), a member of the genus *Potyvirus* in the family *Potyviridae*, was first reported in Italy in 1973 (Lisa et al., 1981) and was subsequently found causing devastating epidemics in commercial cucurbits worldwide (Lisa and Lecoq, 1984). Symptoms include mosaic, yellowing, shoestring, and distortion on leaves; stunting in plant growth; and deformation of fruits. A Taiwan isolate, designated ZYMV TW-TN3, was collected from diseased sponge gourd (*Luffa cylindrica* Roem.) from Tainan, Taiwan, in 1993 (Lin et al., 1998). Similar to other potyviruses, TW-TN3 RNA was presumed coding a single large polyprotein that undergoes proteolytic processing by virus-encoded proteases to yield several mature gene products (Shukla et al., 1994).

The complete nucleotide sequences of the RNA genomes of three ZYMV isolates, ZYMV CA (California isolate) (Wisler et al., 1995), ZYMV RU (Reunion Island isolate) (Wisler et al., 1995) and ZYMV S (Singapore isolate) (Lee and Wong, 1998) have previously been reported. Based on the degrees of coat protein (CP) homology, these three ZYMV isolates were classified in genotypes II, III, and VI, respectively (Lin et al., 2000). However, the TW-TN3 isolate and most of Taiwan ZYMV isolates were placed in genotype I (Lin et al., 2000). Thus, elucidation of the genomic sequence of TW-TN3 would further understanding of its relationships with other genotypes.

In this study, the complete nucleotide sequence of TW-TN3 was determined and compared with those of the CA, RU, and S isolates. Moreover, since sequence information for the P1 proteins of the additional three ZYMV US isolates (SV, MD, and FL/AT) was also available (Wisler et al., 1995), phylogenetic analysis based on P1 proteins was performed to investigate the evolutionary relationships among the TW-TN3 and the other six ZYMV isolates reported.

Materials and Methods

Virus Purification and RNA Extraction

A typical isolate of ZYMV TW-TN3 was isolated from naturally infected sponge gourd grown in Tainan, Taiwan, (Lin et al., 1998), and propagated in zucchini squash (*Cucurbita pepo* L. var. Zucchini) under greenhouse conditions. Viral particles were purified by PEG precipita-

[3]The GenBank accession number for the sequence data reported in this paper is AF343979.

*Corresponding author. Tel: 886-4-22877021; Fax: 886-4-22877585; E-mail: sdyeh@nchu.edu.tw

tion and Cs_2SO_4 centrifugation (Gonsalves and Ishii, 1980). Genomic RNA was extracted by SDS-proteinase K treatment followed by sucrose density gradient centrifugation as described by Yeh and Gonsalves (1985).

cDNA Synthesis and Cloning

The clone pZCP-36 (Figure 1), covering the CP reading frame and the 3′ untranslated region (UTR) of TW-TN3 RNA, had been previously generated from the cDNA fragments amplified by reverse-transcription polymerase chain reaction (RT-PCR) (Lin et al., 2000). The cDNA segments covering other genomic regions were synthesized and screened from a cDNA library that was constructed according to the protocols of ZAP-cDNA synthesis and cloning kits (Stratagene, La Jolla, CA). The methylated first strand cDNA was synthesized by Moloney murine leukemia virus reverse transcriptase after the addition of ZCP-2 primer (5′-CAACAGCTTCGCGGGC-3′) designed from the known region of the CP gene (Lin et al., 1998). The second strand cDNA was generated according to the method of Gubler and Hoffman (1983) by RNase H digestion and DNA polymerase I extension. The dsDNA was filled in with *Pfu* DNA polymerase (Stratagene) and ligated with *Eco*RI adaptors at both ends. The dsDNA was inserted into the lambda ZAP-II *Eco*RI/CIAP vector (Stratagene) and packaged with capsid proteins (Gold Packaging Extract, Stratagene). The cDNA probe used for plaque hybridization was prepared according to Maniatis et al. (1989), using [α^{32}P]-labeled DNA corresponding to pZCP-36 or the selected clones. In this investigation, the cDNA fragments corresponding to 5′-terminal, P1, HC-Pro, and P3 regions were also synthesized by RT-PCR using purified RNA of TW-TN3 as a template as shown in Figure 1. The cDNA fragment reflecting the 5′ terminal region of TW-TN3 RNA was amplified by the downstream primer MZ1791 (5′-GTGCCATTTGGAAAGTTACGC-3′) complementary to nucleotide positions 1791 - 1811 of ZYMV RU

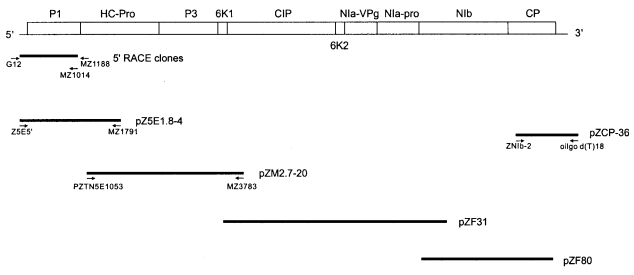

Figure 1. The strategy for determination of the complete nucleotide sequence of the RNA genome of a Taiwan isolate ZYMV TW-TN3. Complenentary cDNA clone pZCP36 that covers the 3′-terminal region of the RNA was obtained from a previous study (Lin et al., 2000). Other cDNA clones were selected from the cDNA library (pZF31 and pZF80) or generated from fragments amplified by reverse-transcription polymerase chain reaction (RT-PCR) (pZ5E1.8, pZM2.7-20, and 5′ RACE clones). Selected cDNA clones were sequenced from both orientations to elucidate the complete nucleotide sequence of ZYMV TW-TN3 RNA.

(Wisler et al., 1995) and the upstream Z5E5′ primer designed from the 5′ terminal consensus sequence of CA and RU isolates (5′-AAAATTAAAACAAATCACAAAGACTAC-3′, positions 1-27) (Wisler et al., 1995). The cDNA reflecting the HC-Pro and P3 regions of nucleotides 1053-3792 was amplified by the upstream primer PZTN5E1053 (5′-GCGAAGTTGACCACTATTCGTCGC-3′) and the downstream primer MZ3783 (5′-GTTCTATAATGTG GTACTGTGCGG-3′), both designed from the consensus sequences of CA and RU isolates (Wisler et al., 1995).

Determination of the Actual Sequence of the 5′ Extreme

The 5′ terminal sequence of TW-TN3 RNA was determined by the 5′ RACE system (GibcoBRL, Gaithersburg, MD). First strand cDNA synthesis was carried out using 0.5 μg of TW-TN3 RNA and 25 ng of primer MZ1188 (5′-GCTAACTCACCACATTCCTCGTTG-3′, nts 1165-1188 designed from the known sequence). After removal of template by RNase Mix (GibcoBRL) digestion, the first strand cDNA was tailed with d(CTP) by terminal deoxynucleotidyl transferase (GibcoBRL). PCR amplification of the dC-tailed cDNA was performed with *EX Taq* DNA polymerase (TaKaRa, Shiga 520-21, Japan), using the upstream primer G12 (5′-CGACTAGTGGGGGGGGGGG-3′) containing a *Spe*I site and the second downstream primer MZ1014 (5′-TCGTCGTCTCTACCTCGAATCACC-3′, nts 991-1014 designed from the known sequence). The amplified cDNA fragments were directly inserted into the pCR2.1-TOPO vector (Invitrogen, San Diego, CA) to generate recombinants that contained the amplified cDNA fragment reflecting the actual 5′-terminal region of ZYMV-TN3 RNA (Figure 1).

DNA Sequencing and Assembly

The clones selected by plaque hybridization or generated by RT-PCR amplification, were digested with *Eco*RI or *Hind*III and subcloned for sequencing. DNA sequencing was performed by the dideoxynucleotide chain-termination method (Sanger et al., 1977), using ssDNA templates generated by phage VCSM13 (Stratagene). DNA sequencing was performed using T7 DNA polymerase following the instructions of the Sequenase Version 2.0 kit (United States Biochemical, Cleveland, OH 44122). Sequence data were assembled, and amino acid sequences were predicted using PC/GENE 6.85 software (IntelliGenetics, Inc., University of Geneva, Switzerland).

Sequence Analysis

Multiple sequences were aligned by the PILEUP program of the GCG package (Version 10.0, 1999, Genetics Computer Group, Madison, WI., USA). Distance matrices for complete P1 sequences were calculated from alignments based on the Dayhoff PAM matrix (Dayhoff et al., 1983), using the program PROTDIST in PHYLIP software package Version 3.5c (Kimura, 1980). Phylogenetic relationships were established from these distances using the Neighbor-Joining routines. The repeatability of the

branching orders obtained was estimated using the SEQBOOT program (PHYLIP software) for bootstrap resampling (200 bootstrap reiterations) the multiple sequence alignment (Felsenstein, 1985). Bootstrap consensus trees were then built using the CONSENSE program and the NEIGHBOR unrooted tree was drawn using the DRAWTREE program of the PHYLIP software.

The GAP program in the GCG package was used to determine the degrees of amino acid and nucleotide homology of individual gene products of TW-TN3 with those of other ZYMV isolates. Sequence for alignments were obtained from the GenBank database under the following accession numbers including CA (L31350), RU (L29569), S (AF014811) MD (L35588), SV (L35589), and FL/AT (L35590) isolates.

Results and Discussion

Complete Nucleotide Sequence of ZYMV TW-TN3

Two cDNA clones, pZF80 and pZF31, were selected from the lambda ZAP library by plaque hybridization; and two cDNA clones, pZM2.7-20 and pZ5E1.8-4, were obtained from RT-PCR. They contained inserts of 2.3, 3.9, 2.7, and 1.8 kb, respectively (Figure 1). These four clones combined with the previously obtained clone pZCP-36 represented 100% of TW-TN3 genome and were used to determine the complete nucleotide sequence of TW-TN3 RNA. In addition, the 5′ RACE clones reflecting the 5′-terminal region were used to reveal the actual nucleotide sequence of the 5′ extreme of the genomic RNA (Figure 1).

The complete RNA genome of TW-TN3 was determined to be 9591 nucleotides in length, excluding the poly (A) tail (Figure 2). The overall base composition of the sequence was found to be 31.05% A, 19.48% C, 23.75% G, and 25.72% U. Computer analysis of the sequence revealed a large open reading frame (ORF) that commences at the first AUG at nucleotide position 139 and terminates at position 9381 with a UAA codon, leaving a 3′ noncoding region of 212 nucleotides. The ORF was predicted encoding a polyprotein of 3079 amino acids with a calculated molecular weight of 351,139 (351 kDa).

The 5′ untranslated region (UTR) of TW-TN3 RNA was determined as 138 nucleotides in length with a base composition of 43.48% A, 21.01% C, 10.87% G, and 24.64% U. The high AU content (68.12%) in the 5′ leader sequence is common in plant viruses (Gallie et al., 1987). This feature was proposed to aid melting of secondary structures of RNA for efficient translation (Jobling and Gehrke, 1987). When six clones obtained from 5′ RACE using primers MZ1188 and MZ1014 were used to analyze the 5′ end of TW-TN3 isolate, three of them contained the complete 5′ end. The actual sequence of the 5′-terminal region was determined as "5′-AAATTAAAACAAA…-3′." The first to third residues of the 5′-terminal sequence were determined as $(A)_3$-start and were different from the designed primer Z5E5′ that contained $(A)_4$-start. However, the other nucleotide residues of 5′-terminal region were the same as those designed for the primer Z5E5′ which was designed from the 5′ terminal consensus sequence of CA and RU isolates (Wisler et al., 1995). Our results indicated that TW-TN3 contains an $(A)_3$-start the same as isolate S (Lee and Wong, 1998), but different from the $(A)_4$-start of isolates CA and RU (Wisler et al., 1995).

Proteolytic Processing of TW-TN3 Polyprotein

The polyprotein of a potyvirus is cleaved by the virus-encoded proteases P1, HC-Pro, and NIa, resulting in at least nine functional, mature proteins (Shukla et al., 1994). By comparison of the deduced amino acid sequence, TW-TN3 also possessed all nine potential protease cleavage sites. The proposed cleavage sites and the deduced products of TW-TN3 are shown in Figures 1 and 2.

The processing of the N-terminal region of the potyviral polyprotein involves the proteases HC-Pro and the P1 (Shukla et al., 1994). The C-terminal part of P1 protein has been identified as a serine-type protease responsible for the autocatalytic cleavage between P1 (36 kDa) and HC-Pro (52 kDa) proteins (Verchot et al., 1991). After comparing the homologous regions in the polyproteins of TW-TN3 and other potyviruses (Yeh et al., 1992; Lee and Wong, 1998), we concluded that the cleavage site for P1 and HC-Pro proteins of TW-TN3 occurs at Y_{310}/S_{311}. The serine protease active site motif, GCSG, was found in all seven ZYMV isolates CA, MD, SV, FL/AT, RU, S, and TW-TN3. The P1 consensus motif LVIRG for all reported ZYMV isolates was also found at amino acid positions 284-288 of TW-TN3.

HC-Pro (helper component-protease) protein is released from the polyprotein by autocatalysis and the HC-Pro/P3 cleavage site has been shown to be at G-G in TEV (Carrington et al., 1989). By comparison, the sequence KHYRVGG at amino acid positions 761-767 was predicted as the cleavage site for the HC-Pro protein of TW-TN3, and this motif was also present in the other three ZYMV isolates CA, RU, and S.

The NIa protein (nuclear inclusion protein a) is the major protease of potyviruses and is involved in processing the C-terminal two-thirds of the polyprotein (Shukla et al., 1994). It belongs to the class of cysteine-type proteases (Bazan and Fletterick, 1988), and the motif GDCG at amino acid positions 2190-2194 of the polyprotein of TW-TN3 was considered the catalytic site responsible for NIa protease activity (Shukla et al., 1994). A histidine residue that may play a role in substrate binding (Dougherty and Semler, 1993) was also found conserved in TW-TN3 at amino acid postion 2208.

Comparison of cleavage sites of the C-terminal halves of the polyproteins of the three reported ZYMV isolates with those predicted from TW-TN3 revealed that the NIa protease cleaves at Q-S, and E-S dipeptides (Table 1). Moreover, Shukla et al. (1994) concluded that the Q-A, Q-G, E-A, and E-G are also NIa protease cleavage sites. We found that Q-A and Q-G are responsible for the junction sites of P3/6K1 and 6K1/CIP of ZYMV isolates, respec-

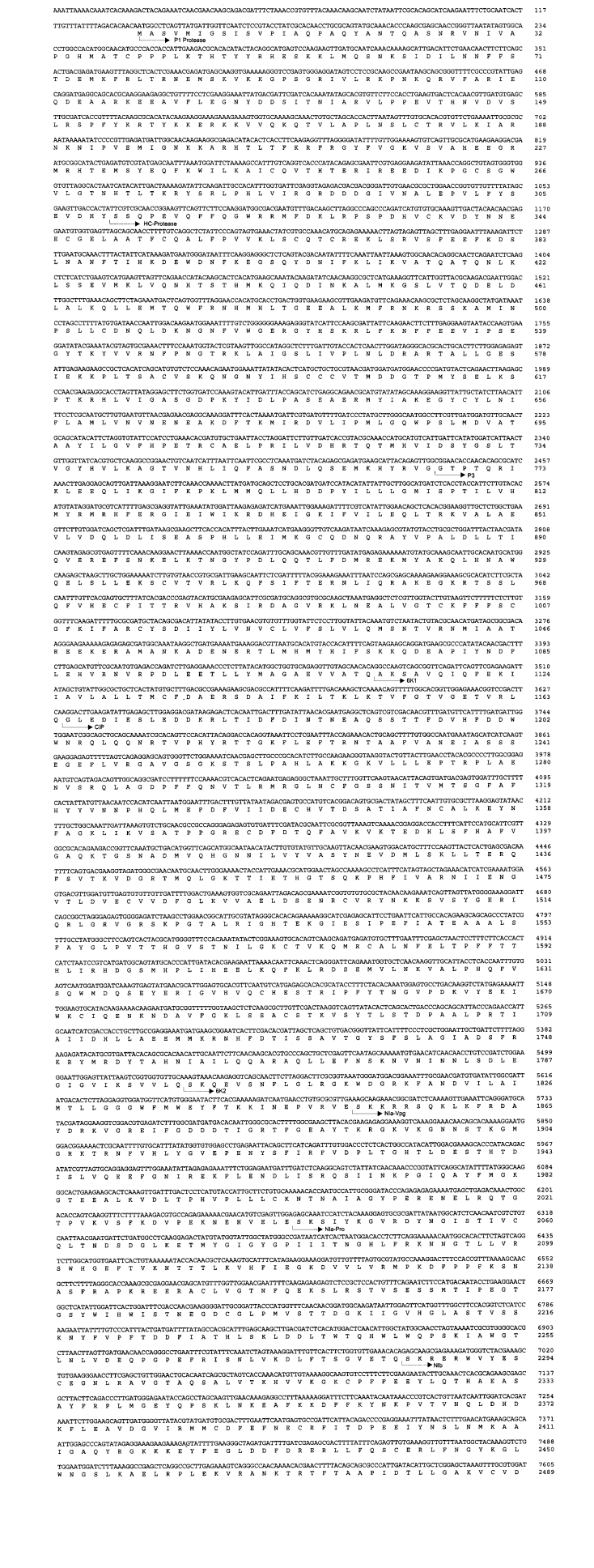

Figure 2. The complete nucleotide sequence of ZYMV TW-TN3 RNA (in the cDNA form) and deduced amino acids of the polyprotein. Arrows indicate the cleavage sites of putative viral proteins.

tively (Table 1). In all cleavage sites, valine was found present 100% at the upstream P4 position. The comparison of individual cleavage sites of NIa protease ZYMV isolates are listed in Table 1, and the consensus sequence for NIa cleavage sites was found to be V-x-x-(Q, E)/(S, A, G). The results also revealed that glutamine presents at the P6 position as 57%, leucine at P2 position as 46%, and lysine at P2´ position down stream from the cleavage site as 57%. These residues may affect processing efficiency and regulate production of corresponding products (Shukla et al., 1994). In addition, there is an alanine at the position P3 of the P3/6K1 cleavage site of TW-TN3, which is not the case with the serine of isolate CA or the threonine of isolates RU and S (Table 1). The different amino acid might result from mis-incorporation by one of the DNA polymerase used to create the cDNAs. However, Figure 1 shows that the region of the P3/6K1 cleavage site of TW-TN3 is covered by two cDNA clones, pZM2.7-20 and pZF31. The cDNA in pMZ2.7-20 was synthesized by RT-PCR, and the cDNA in pZF31 was selected from the cDNA library. The data from both clones showed that both sequences covered the P3/6K1 cleavage site of TW-TN3 and were thus identical. Therefore, the possibility that the different alanine in the position P3 of the P3/6K1 cleavage site was caused by synthesis or sequence error is minimal. Whether this unique amino acid affects the biological functions of the virus remains to be investigated.

The NIa protease recognized the predicted cleavage sites in *cis* and *trans* manners to generate products of P3 protein (40 kDa), 6K1 protein (6 kDa), CIP (cytoplasmic inclusion protein, 71 kDa), 6K2 protein (6 kDa), NIa-Vpg protein (22 kDa), NIa-Pro protein (27 kDa), NIb protein (nuclear inclusion protein b, 60 kDa), and CP (31 kDa) (Figure 2). The proteolytic processing of these proteins by the NIa protease is the common characteristic of potyviral gene expression, and the consensus sequence of the NIa cleavage site can be used for cleavage of the fusion protein in protein engineering.

Several other functional motifs of potyviral polyprotein were also found in the polyprotein of TW-TN3. The RNA helicase motifs GAVGSGKST and PTR were found conserved in the N-terminal region of the CIP at amino acid postions 1249-1257 and 1274-1277, respectively. Three RNA-dependent RNA polymerase (RdRp) consensus motifs (Domier et al., 1987), YCHADGS at amino acid positions 2528-2534, GNNSGQPSTVVDNTLMV at amino acid positions 2591-2607, and NGDDL at amino acid positions 2634-2638, were conserved in the large NIb of TW-TN3 polyprotein.

Comparison of TW-TN3 Gene Products with those of other ZYMV Isolates

Amino acid identities among the gene products of TW-TN3 and those of the other three reported ZYMV isolates are summarized in Table 2. The most variable protein was found to be the P1 protein (59.0-93.2%). The other proteins showed high degrees of conservation with amino acid identities ranging from 87.2-100% among four ZYMV isolates. Our results are similar to those of Shukla et al. (1994), who concluded that among potyviral gene

Table 1. Comparison of cleavage sites, recognized by the NIa protease, in the polyprotein of four ZYMV isolates originating in California.

Cleavage site	Virus isolate[a]	Amino acid position									
		P7	P6	P5	P4	P3	P2	P1	P1´	P2´	P3´
P3/6K1	TW	A	E	V	V	A	T	Q	A	K	S
	CA	V	E	V	V	S	T	Q	A	K	S
	RU	A	E	I	V	T	P	Q	A	K	S
	S	G	E	V	V	T	A	Q	A	K	S
6K1/CIP	TW	G	E	T	V	R	L	Q	G	L	E
	CA	G	E	T	V	R	L	Q	G	L	E
	RU	G	E	T	V	R	L	Q	G	L	E
	S	G	E	T	V	R	L	Q	G	L	E
CIP/6K2	TW	I	K	S	V	V	L	Q	S	K	Q
	CA	I	K	S	V	V	L	Q	S	K	Q
	RU	I	K	S	V	V	L	Q	S	K	Q
	S	I	K	S	V	V	L	Q	S	K	Q
6K2/NIa-Vpg	TW	N	E	P	V	R	V	E	S	K	K
	CA	N	E	P	V	R	V	E	S	K	K
	RU	N	E	P	V	R	V	E	S	K	K
	S	N	E	P	V	R	V	E	S	K	K
NIa-Vpg/NIa-Pro	TW	N	E	H	V	E	L	E	S	K	S
	CA	N	E	H	V	E	L	E	S	K	S
	RU	N	E	H	V	E	L	E	S	K	S
	S	N	E	H	V	E	L	E	S	K	S
NIa-Pro/NIb	TW	T	S	G	V	E	T	Q	S	K	R
	CA	T	S	G	V	E	T	Q	S	K	R
	RU	N	S	G	V	E	T	Q	S	K	R
	S	T	S	G	V	E	T	Q	S	K	R
NIb/CP	TW	G	D	T	V	M	L	Q	S	G	T
	CA	G	D	T	V	M	L	Q	S	G	T
	RU	R	D	T	V	M	L	Q	S	D	T
	S	G	D	T	V	M	L	Q	S	D	T
Consensus sequence		-	-	-	V	-	-	Q,E	S,A,G	-	-

[a]CA, GenBank accession number L31350; Reunion Island (RU, L29569); Singapore (S, AF014811); Taiwan (TW-TN3, AF343979).

Table 2. Comparison of nucleotide identities of genomic sequences and amino acid identities of gene products of ZYMV TW-TN3 with those of three reported ZYMV isolates originating in California.

Nucleotide identity (%)				Amino acid identity (%)									
Isolate[a]	5´UTR	3´UTR	ORF	P1	HC	P3	6K1	CIP	6K2	NIa/VPg	NIa/Pro	NIb	CP
CA	83.3	95.2	93.6	93.2	96.1	98.0	100	97.3	96.2	98.4	99.2	98.3	97.8
RU	71.0	90.4	81.6	62.9	93.2	89.9	96.2	94.8	98.1	94.7	95.1	96.3	90.7
S	61.6	95.7	82.9	59.0	89.5	94.2	98.1	96.1	92.5	97.4	97.1	87.2	91.0

[a]CA, GenBank accession number L31350; Reunion Island (RU, L29569); Singapore (S, AF014811); Taiwan (TW-TN3, AF343979). The nucleotide identities and amino acid identities were calculated by the GAP program of GCG package (version 10.0, 1999, Genetics Computer Group, Madison, WI., USA).

products, the P1 protein is most variable and the NIb protein is the most conserved. This investigation found that the NIb protein of TW-TN3 shares 98.3 and 96.3% amino acid identities with CA and RU isolates, respectively. However, it only shares an 87.2% amino acid identity with the NIb protein of the S isolate. This low identity apparently results from two regions, GRSKHLLQTLDGGV PAEQIKQGSIQEGL at amino acid positions 7140-7223 and PKFVWMTLTMSSTVKSQVSVKL at amino acid positions 7602-7667. Both are completely different from those of TW-TN3, CA, and RU isolates and apparently result from sequence errors. The NIa-Vpg proteins, NIa-Pro proteins, and CIPs of the four ZYMV isolates were also highly conserved with amino acid identities of 99.2-94.7%. However, the HC-Pro proteins shared lower degrees (96.1-89.5%) of amino acid identity.

Comparison of the 5´ and 3´ Untranslated Regions with those of other ZYMV Isolates

Comparison of the 5´ UTR of TW-TN3 with those of ZYMV isolates CA, RU, and S showed that they share 83.3%, 71.0%, and 61.6% nucleotide identities, respectively. This relatively low identity reflects the diversity of the 5´ UTRs of ZYMV isolates from different areas of the world. Such sequence variability in the 5´ UTR of ZYMV is useful for strain differentiation and phylogenetic analysis. The 3´ UTR of TW-TN3, however, exhibited higher degrees of homology with nucleotide identities of 95.2, 90.4, and 95.7% when compared with those of CA, RU, and S, respectively. Although the 3´ UTRs of potyviruses are heterogeneous in size (Turpen, 1989), the high degree of identity among these ZYMV isolates reflects the common roles involved in the initiation of minus-strand RNA synthesis and prevention of exonucleolytic degradation (Bryan et al., 1992).

Phylogenetic Analysis of P1 Protein

Shukla and Ward (1988; 1989) concluded that amino acid identities of CPs between different potyviruses ranged from 38% to 71%, whereas between strains of each species, they ranged from 90% to 99%. For P1 protein, the amino acid sequences of G2 and G7 strains of *Soybean mosaic virus* (SMV) share a 93.5% identity (Jayaram et al., 1992). However, only a 66.7% identity was found in the P1 proteins of *Papaya ringspot virus* YK (PRSV YK) and HA (Wang and Yeh, 1997). Therefore, the threshold of sequence similarity of CPs for grouping strains of the same potyvirus is obviously not applicable to the P1 protein. Similar results were obtained when the P1 proteins were compared between strains of PVY (Tordo et al., 1995). The amino acid identities of the P1 coding regions of 12 PVY isolates ranged from 72.8-100%, and these isolates can be classified into three geographical groups. In this investigation, the amino acid and nucleotide sequences of the P1 proteins of seven ZYMV isolates were compared, including TW-TN3, RU, S, and four US isolates (CA, MD, SV, and FL/AT) (Wisler et al., 1995) (Figure 3A). The four US isolates were found to be more conserved, with nucleotide identities of 96.7-98.0% and amino acid identities of 95.3-96.3%. The Taiwan TW-TN3 isolate was found to have closer relationships with the four US isolates, with nucleotide identities of 93.4-94.2% and amino acid identities of 92.3-93.9%. However, P1 proteins of ZYMV isolates from RU and S shared low amino acid identities of 56.2-62.9% with those of TW-TN3 and US isolates. These results indicate that TW-TN3 isolate and US isolates can be classified in the same group if the nucleotide and amino acid identities are set at 90% as a threshold. On the other hand, the RU isolate and S isolate are distinct from the Taiwan and US isolates and should be classified as two different individual groups.

A

	TW-TN3 1	CA 2	MD 3	SV 4	FL/AT 5	RU 6	S 7
TW-TN3 1		**94.0**	**93.4**	**93.8**	**94.2**	65.7	64.1
CA 2	**93.2**		**96.7**	**96.9**	**97.4**	64.3	63.8
MD 3	**92.7**	**95.3**		**97.3**	**97.9**	63.9	62.7
SV 4	**92.3**	**95.7**	**95.7**		**98.0**	64.3	62.8
FL/AT 5	**93.9**	**95.8**	**96.7**	**96.3**		67.2	64.8
RU 6	62.9	60.3	60.0	59.9	61.0		74.2
S 7	59.0	57.4	56.3	56.2	57.7	69.3	

B

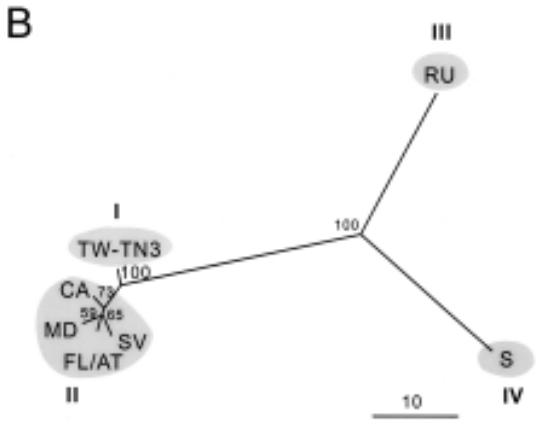

Figure 3. A, nucleotide identities (above diagonal) and amino acid identities (below diagonal) of the P1 proteins of seven ZYMV isolates. The nucleotide and amino acid identities over 95% are grouped in the black box, and the 90-95% identities are grouped in the gray box. B, an unrooted minimum distance tree of P1 Proteins of ZYMV isolates from Taiwan (TW-TN3), United States (CA, MD,SV, and FL/AT), Reunion Island (RU), and Singapore (S). The phylogenetic tree was established by the PHYLIP software package (Kimura, 1980). Numbers indicating the percentage of 200 bootstrap replicates are shown at the appropriate nodes. The scale to measure genetic distances is indicated at the lower right.

Although, TW-TN3 was classified closer to US isolates, it shared nucleotide and amino acid identities lower than 95%, indicating that TW-TN3 isolate can be further considered as a different genotype distinct from the US isolates. Lin et al. concluded that different genotypes of ZYMV isolates share nucleotide identities of CP lower than 95% (Lin et al., 2000). If the same 95% identity threshold was applied to P1 proteins, TW-TN3 could be considered a different genotype from the US isolates.

The evolutionary relationships of the P1 protein of TW-TN3 with those of other ZYMV isolates were further analyzed by phylogenetic analysis (Figure 3B). The P1 proteins of the seven ZYMV isolates were compared and grouped into four genotypes based on the 95% nucleotide identity as a threshold. This result was similar to the phylogenetic analysis of ZYMV CP, which revealed four genotypes, including genotype I (4 Taiwan isolates), genotype II (3 US isolates, 1 Israel isolate, and 1 Taiwan isolate), genotype III (Reunion Island isolate), and genotype IV (Singapore isolate) (Lin et al., 2000). In this investigation, the phylogenetic tree based on P1 proteins indicated that TW-TN3 was classified closer to the US isolates, agreeing well with the CP analysis that the distance relationships of genotype I were closer to genotype II. Thus, phylogenetic analyses based on the P1 proteins and CPs indicated that the ZYMV of Taiwan isolates and US isolates may have evolved from the same ancestor. On the other hand, the variation in P1 proteins may reflect the variation in symptom development, host range, and geographical distribution (Lee and Wong, 1998; Tordo et al., 1995). It is apparent that these parameters for the ZYMV isolates from Reunion Island and Singapore are quite different from those for Taiwan and US isolates.

A new standard to distinguish different strains of viruses in the genus *Potyvirus* has to be established when more sequence information for different P1 proteins becomes available. Nevertheless, the diversity of the P1 protein should be taken into consideration regarding taxonomy of potyviruses. Also the P1 gene can be used as a target for development of specific probes or for heteroduplex mobility assay (HMA) to distinguish strains of potyviruses. Furthermore, the complete genome sequence of ZYMV TW-TN3 provides an important base for construction of the in vitro and in vivo infectious clones at the cDNA level. The infectious clones of ZYMV TW-TN3 can be used as tools, through recombination at the cDNA level, to study the differences of nucleotides and amino acids among different strains for their roles in biological functions.

Acknowledgments. This study was supported in part by the NSC Grants 86-2313-B005-111-B14 and 87-2312-B005-005 of the Republic of China on Taiwan.

Literature Cited

Bazan, J.F. and R.J. Fletterick. 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases structural and functional implications. Proc. Natl. Acad. Sci. USA **85:** 7872-7876.

Bryan, G.T., R.C. Gardner, and R.L.S. Forster. 1992. Nucleotide sequence of the coat protein gene of a strain of clover yellow vein virus from New Zealand conservation of a stem-loop structure in the 3′ region of potyviruses. Arch. Virol. **124:** 133-146.

Carrington, J.C., S.M. Cary, T.D. Parks, and W.G. Dougherty. 1989. A second proteinase encoded by a plant potyvirus genome. EMBO J. **8:** 365-370.

Dayhoff, M.O., W.C. Barker, and L.T. Hunt. 1983. Establishing homology in protein sequences. Methods Enzymol. **91:** 524-545.

Domier, L.L., J.G. Shaw, and R.E. Rhoads. 1987. Potyviral proteins share amino acid sequence homology with picorna como and caulimoviral proteins. Virology **158:** 20-27.

Dougherty, W.G. and B.L. Semler. 1993. Expression of virus-encoded proteinases: Functional and structural similarities with cellular enzymes. Microbiol. Rev. **57:** 781-822.

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

Gallie, D.R., D.E. Sleat, J.W. Watts, P.C. Turner, and T.M. Wilson. 1987. A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression in vivo. Nucleic Acids Res. **15:** 8693-8711.

Gonsalves, D. and M. Ishii. 1980. Purification and serology of papaya ringspot virus. Phytopathology **70:** 1028-1032.

Gubler, U. and B.J. Hoffman. 1983. A simple and very efficient method for generating complementary DNA libraries. Gene **25:** 263-270.

Jayaram, C., J.H. Hill, and W.A. Miller. 1992. Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the rsv resistance gene. J. Gen. Virol. **73:** 2067-2077.

Jobling, S.A. and L. Gehrke. 1987. Enhanced translation of chimeric messenger RNA containing a plant viral untranslated leader sequence. Nature **325:** 622-625.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111-20

Lee, K.C. and S.M. Wong. 1998. Variability of P1 protein of zucchini yellow mosaic virus for strain differentiation and phylogenetic analysis with other potyviruses. DNA Sequence **9:** 275-293.

Lin, S.S., R.F. Hou, and S.D. Yeh. 2000. Heteroduplex mobility and sequence analyses for assessment of variability of zucchini yellow mosaic virus. Phytopathology **90:** 228-235.

Lin, S.S., R.F. Hou, C.H. Huang, and S.D. Yeh. 1998. Characterization of zucchini yellow mosaic virus (ZYMV) isolates collected from Taiwan by host reactions, serology, and RT-PCR. Plant Prot. Bull. **40:** 163-176.

Lisa, V. and H. Lecoq. 1984. Zucchini Yellow Mosaic Virus. CMI/AAB Descriptions of Plant Viruses, No. 282.

Lisa, V., G. Boccardo, G. D'Agostino, G. Dellavalle, and M. D'Aquilio. 1981. Characterization of a potyvirus that causes zucchini yellow mosaic. Phytopathology **71:** 667-672.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1989. Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Springs Harbor Laboratory, New York.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463-5467.

Shukla, D.D. and C.W. Ward. 1988. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. J. Gen. Virol. **69:** 2703-2710.

Shukla, D.D., C.W. Ward, and A.A. Brunt. 1994. The Potyviridae. CAB International, Wallingford, Oxon, UK.

Shukla, D.S. and C.W. Ward. 1989. Identification and classification of potyviruses on the basis of coat protein sequence data and serology. Arch. Virol. **106:** 171-200.

Tordo, V.M.J., A.M. Chachulska, H. Fakhfakh, M. Le Romancer, C. Robaglia, and S. Astier-Manifacier. 1995. Sequence polymorphism in the 5´NTR and in the P1 coding region of potato virus Y genomic RNA. J. Gen. Virol. **76:** 939-949.

Turpen, T. 1989. Molecular cloning of a potato virus Y genome nucleotide sequence homology in non-coding regions of potyviruses. J. Gen. Virol. **70:** 1951-1960.

Verchot, J., E.V. Koonin, and J.C. Carrington. 1991. The 35-kDa protein from the amino-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. Virology **185:** 527-535.

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.

Wisler, G.C., D.E. Purcifull, and E. Hiebert. 1995. Characterization of the P1 protein and coding region of the zucchini yellow mosaic virus. J. Gen. Virol. **76:** 37-45.

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. Virology **143:** 260-271.

Yeh, S.D., F.J. Jan, 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 ringspot virus RNA. J. Gen. Virol. **73:** 2531-2541.

矮南瓜黃化嵌紋病毒台灣分離株之基因體核酸序列及其基因組成分析

[1]林詩舜　[1]侯豐男　[2]葉錫東

[1] 國立中興大學農業生物科技學研究所

[2] 國立中興大學植物病理學系

矮南瓜黃化嵌紋病毒台灣分離株（ZYMV TW-TN3）之全長基因體核酸序列經由五個互補去氧核糖核酸之選殖株全部解序完畢，其全長序列不包含 poly (A) 為 9591 個核苷酸。經電腦分析全長序列顯示其基因體核酸可轉譯出一個大的轉譯架構，此轉譯架構可產生一個含有 3080 胺基酸的複合大蛋白體。將已發表之加州、雷優尼島（Reunion Island）及新加坡 ZYMV 分離株與 TW-TN3 做比較，彼此之間的 5´ 非轉譯區與 3´ 非轉譯區之核酸相似度分別為 61.6-83.3% 及 90.4-95.7%。而彼此之間的 P1 蛋白是差異性最大，其胺基酸相同度為 59-93%。分析 TW-TN3 與其它已知序列之六個 ZYMV 分離株之 P1 蛋白演化關係，顯示 P1 蛋白可分成四個主要之基因型。TW-TN3 屬於第一基因型，美國分離株屬於第二基因型，雷優尼島與新加坡分離株分別屬於第三與第四基因型。由演化關係圖中顯示第一基因型與第二基因型的親源關係較為接近，因此台灣與美國分離株可能由共同祖先所演化而來。分析四個 ZYMV 分離株之 NIa 蛋白裂解酵素切位，其主要之裂解位置在 Q-S 及 E-S 雙胜肽序列，並且推測其 NIa 辨識序列為 V-x-x-(Q, E)-(S, A, G)。總結 TW-TN3 所轉譯的基因組成由 5´ 至 3´ 依序為 Vpg/5´ 引導序列/P1 (36 kDa)/HC Pro (52 kDa)/P3 (40 kDa)/6K1 (6 kDa)/CIP (71 kDa)/6K2 (6 kDa)/NIa-Vpg (22 kDa)/NIa-Pro (27 kDa)/NIb (60 kDa)/CP (31 kDa)-3´ 非轉譯區–poly(A)。

關鍵詞：ZYMV分離株；全長基因體核酸序列；基因組成。