J. Gen. Plant Pathol. 67 : 73-77 (2001)

# Nucleotide Sequence of *Citrus Tristeza Virus* Seedling Yellows Isolate†

**Gede SUASTIKA[1,2], Tomohide NATSUAKI[1]*, Hirotsugu TERUI[1], Takeshi KANO[3], Hiroyuki IEKI[3] and Seiichi OKUDA[1]**

ABSTRACT

**The complete nucleotide sequence of a seedling-yellows-inducing isolate NUagA of *Citrus tristeza virus* (CTV) was determined. It consisted of 19302 nucleotides and contained 12 open reading frames (ORF) organized identically to those of previously sequenced isolates. This genome is the largest among the CTV genome sequenced so far ; it is 6 nucleotides (nt), 76 nt, 43 nt, and 53 nt longer than that of T36 (quick decline, Florida), VT (seedling yellows, Israel), T385 (mild, Spain), and SY568 (stem pitting, California), respectively. Sequence comparison of NUagA and the other isolates revealed approximately 90% identities throughout the 3′ half of the genome. The 5′ half of the genome was only about 70% identical to that of T36 but still high at about 90% to those of VT, SY568, and T385. Comparison of amino acid sequences on ORF1a encoding polyproteins, the most variable region, reflects the CTV isolate relationship ; NUagA is closely related to VT, SY568, and T385, but distantly related to T36.**

(Received May 29, 2000 ; Accepted November 16, 2000)

**Key words :** *Citrus tristeza virus*, seedling yellows, NUagA, *Closterovirus*, nucleotide sequence.

An economically devastating citrus disease, commonly known as tristeza, in most citrus-producing regions of the world is caused by *Citrus tristeza virus* (CTV). The virus is phloem-limited and naturally transmitted by several aphid species to a restricted range of rutaceous host plants[3]. CTV belongs to the genus *Closterovirus*. Its particles are long, flexuous rods of about 2,000 nm with two types of capsid proteins. A major capsid protein (CP) of 25 kDa constitutes about 95% of the particle length and a minor capsid protein (CPd) of 27 kDa covers only 5% of one end of the virion[7]. The CTV genome consists of a positive-sense, single-stranded RNA, approximately 20 kb in size[3].

CTV isolates considerably differ in their pathogenicity on various citrus species[16]. Decline and death of plants, stem pitting, and seedling yellows are the major, common symptoms of infected plants. The disease phenotype is thought to result from a complex mixture of related sequence variants because field trees are commonly exposed to multiple aphid-transmissible inocula throughout their life[4]. The field isolates may be separated by aphid or graft transmission and subcultured on different host species[4,9].

The complete genomic sequences of four subcultured CTV isolates—quick decline isolate T36 from Florida[10], seedling yellows isolate VT from Israel[15], mild isolate T385 from Spain[19], and stem pitting isolate SY568 from California[21]—have been determined previously. However, nothing is known about the genomic sequence of Japanese CTV isolates. Here we report the complete genomic sequence of CTV NUagA, a seedling-yellows-inducing isolate, originating from Japan[9], and compare the sequence to those reported for T36, VT, T385, and SY568.

NUagA is one of the most destructive isolates among the aphid transmitted-subculture collection of CTV in Japan[9]. The isolate was transmitted by *Aphis gossypii* Glover from a satsuma mandarin (*Citrus unshiu* Marc.) tree in 1986 and still maintains the ability to cause severe stunting and stem pitting in Mexican lime [*C. au-*

† The nucleotide sequence of the virus described in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under accession number AB046398.

[1] Faculty of Agriculture, Utsunomiya University, Mine-machi 350, Utsunomiya 321-8505, Japan

[2] Present address : Department of Plant Pests and Diseases, Faculty of Agriculture, Bogor Agricultural University, Darmaga Champus, Bogor 16680, Indonesia

[3] National Institute of Fruit Tree Science, Tsukuba 305-8605, Japan

* Corresponding author (E-mail : natsuaki@cc.utsunomiya-u.ac.jp)

*rantifolia* (Christm.) Swingle] and yuzu (*C. junos* Sieb ex Tanaka), and seedling yellows in sour orange (*C. aurantium* L.). The isolate has been propagated in satsuma mandarin plants under greenhouse conditions for viral RNA isolation.

Total nucleic acids were extracted from phloem tissue of virus-propagating plants with phenol/chloroform, and double-stranded (ds) RNA was purified by two cycles of CF-11 cellulose column chromatography[20]. In analysis on 5% polyacrylamide gel electrophoresis, a high molecular dsRNA of approximately 20 kb was detected from infected samples but not from a healthy control. In addition to the 20-kb dsRNA, several low molecular dsRNAs, which are presumably a nested set of subgenomic RNAs[13] or defective RNAs[14], were also detected (data not shown). Only the high molecular dsRNA was eluted from the gels and used for cDNA synthesis.

Purified dsRNA was reverse-transcribed by priming with random hexanucleotides using a First-Strand cDNA Synthesis Kit (Amersham Pharmacia Biotech, USA). Specific primers derived from the known sequence of CTV genomes[10,15] were used to obtain clones by amplification through a polymerase chain reaction (PCR). Cloning of the terminal regions was performed using the 5′ RACE System (Gibco BRL, USA).

Fragments of cloned cDNA were sequenced using a Thermo Sequenase Kit (Amersham Pharmacia Biotech, USA) on a sequencer (Shimadzu DNA Sequencer DSQ-1000L, Japan) according to the manufacturer's directions. Verifying sequences in both directions of all the cDNA from at least three independent clones gave a similar sequence.

After assembling the obtained fragment sequences (Fig. 1), the complete genome sequence of NUagA was 19302 nt in length. This genome is the largest of the CTV genomes sequenced so far, that is 6 nt, 76 nt, 43 nt, and 53 nt longer than the T36, VT, T385, and SY568 genomes, respectively. These size differences are mainly due to several gaps in a separated region of open reading frame (ORF) 1a, and small gaps within other ORFs and noncoding regions. Some of the gaps within ORF1a consist of a number of 1- or 2-nt deletions or insertions expected to cause a shift of subsequent amino acids, but generally gaps were 3-nt deletions or insertions that did not alter subsequent amino acids. The 6-nt insertion at the intergenic region between the p27 and p25 genes of T385, which was proposed to be an important sequence in a recombinant mechanism[19], was not found in the genome of NUagA, nor in those of the other isolates.

The NUagA genomic RNA was organized identically to those of the other sequenced isolates[10,15,19,21] that contain 12 ORFs and untranslated regions (UTR) at the 5′ and 3′ termini (Fig. 1). The translation products of each ORF were similar in size to corresponding ORF products of the other CTV isolates.

The 5′ proximal ORF1a of 9351 nt encodes a 349-kDa polyprotein containing three functional domains characteristic of papain-like protease (PRO), methyltransferase (MTR) and helicase (HEL). The N-terminal portion of the polyprotein consists of two domains of PRO, each of which conserved the putative catalytic residues of cysteines at positions 404 and 889, and histidines at positions 565 and 949. The polyprotein was cleaved at amino acids G/G 485 and G/G 969 to give a proximal PRO (53.4 kDa), distal PRO (54.0 kDa), and 241.6-kDa C-terminal fragments consisting of the MTR and HEL domains. The MTR domain, encompassing amino acids 1033 through 1299, was found to contain all the conserved motifs typical of positive-strand RNA viral type I MTRs[1], and the HEL domain, which encompasses amino

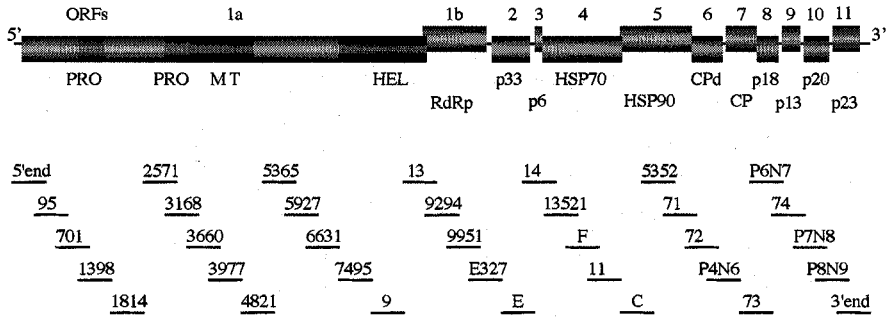


Fig. 1. Schematic representation of the CTV genome organization and the strategy for cloning and sequencing of NUagA genome. The clones were prepared by RT-PCR using appropriate primers based on the T36 and VT sequences, except that clone numbers 3168, 5365, and 13521 were from RT-PCR using primers based on the NUagA sequence. Clones of the 5′- and 3′-terminal regions were obtained by using the 5′ RACE System (Gibco BRL). The clones are represented by the lines below the genome scheme, and the clone numbers are indicated above the lines.

acids from 2694 to 3101, also conserved all seven typical motifs of type I helicases[1].

The ORF1a-encoded polyprotein was the most variable protein both in length and amino acid sequence among isolates (Table 1). Alignment of the amino acid sequence in this polyprotein of NUagA (3133 aa) gave a relatively high homology, close to 90% with that of VT (3108 aa), T385 (3116 aa), and SY568 (3116 aa), but somewhat lower (about 70%) with that of T36 (3125 aa) (Table 1). In spite of the high amino acid sequence variability in ORF1a among the CTV isolates, the PRO, MTR, and HEL domains were conserved (data not shown).

ORF1b of NUagA potentially codes for a protein of 477 amino acids with a calculated molecular weight of 54.3 kDa. ORF1b of CTV has a sequence motif characteristic of RNA-dependent RNA polymerase (RdRp) for positive-strand RNA viruses[6]. ORF1b was postulated to be expressed via a +1 ribosomal frameshifting mechanism as a large fusion protein with ORF1a[1,8,10]. The 5′ region of NUagA ORF1b overlaps the 3′ region of ORF1a by 75 nt. This overlapping region was the same as those of VT but differ from those of T36, T385, and SY568 in which 123 nt, 77 nt, and 127 nt were overlapped, respectively. The slippery sequence GGGUUU at the overlapping ORF1a/1b region and the stop codon at the frameshift point in *Beet yellows virus* (BYV)[1] are not present in NUagA, nor in those of the other isolates[10,15,19,21]. As in the other CTV isolates, the UAG terminator codon in ORF1a of BYV was replaced by a codon for Arg (CGG) in NUagA. Ribosomal pausing at this rare codon was suggested to perform a stop function[6]. It is interesting to note that the comparison of ORF1b of NUagA with that of T36 gave only 79% nucleotide sequence identity, but the amino acid identity was still high (90%) (Table 1).

Table 1. Nucleotide and amino acid sequence identities (%) between the different ORFs and UTRs of CTV isolate NUagA compared to the corresponding regions in isolates T36, VT, T385 and SY568[a]

| NUagA genome | Isolates | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | T36 | | VT | | T385 | | SY568 | |
| | nt | aa | nt | aa | nt | aa | nt | aa[b] |
| 5′UTR | 57 | — | 96 | — | 73 | — | 94 | — |
| ORF1a (pol) | 72 | 69 | 93 | 90 | 89 | 88 | 95 | 92 |
| ORF1b (RdRp) | 79 | 90 | 97 | 97 | 91 | 96 | 88 | 92 |
| ORF2 (p33) | 85 | 86 | 95 | 93 | 85 | 88 | 86 | 90 |
| ORF3 (p6) | 88 | 92 | 95 | 92 | 92 | 96 | 92 | 96 |
| ORF4 (HSP70) | 89 | 94 | 97 | 95 | 88 | 93 | 88 | 94 |
| ORF5 (HSP90) | 87 | 90 | 94 | 95 | 88 | 90 | 87 | 90 |
| ORF6 (CPd) | 90 | 96 | 95 | 97 | 89 | 95 | 89 | 95 |
| ORF7(CP) | 93 | 95 | 97 | 97 | 93 | 95 | 98 | 99 |
| ORF8 (p18) | 93 | 96 | 96 | 95 | 94 | 98 | 97 | 92 |
| ORF9 (p13) | 90 | 92 | 94 | 97 | 90 | 91 | 97 | 96 |
| ORF10 (p20) | 92 | 97 | 97 | 98 | 93 | 95 | 99 | 100 |
| ORF11 (p23) | 91 | 91 | 91 | 96 | 90 | 91 | 98 | 99 |
| 3′UTR | 97 | — | 97 | — | 97 | — | 98 | — |

a) Accession numbers are U16304 (T36), U56902 (VT), Y18420 (T385) and AF001623 (SY568).

b) nt, nucleotide; aa, amino acid.

The downstream ORFs 2 and 3 potentially encode proteins consisting of 303 and 51 amino acids with a calculated molecular mass of 33 kDa and 6 kDa, respectively. Pairwise alignment of the amino acid sequences of NUagA with those of the other isolates gave identity values ranging from 86% to 96% (Table 1). The functions of these two proteins are still unknown. Previous reports for T36[10,17] and VT[15] also failed to align the ORF2-encoded, 33-kDa protein and the ORF3-encoded, small, hydrophobic, 6-kDa protein with any other protein in the databases.

ORFs 4 and 5 encoded proteins of 594 and 536 amino acids approximately equivalent to 65 kDa and 61 kDa, respectively. As in other CTV isolates, the ORF4-encoded protein of NUagA had significant sequence homology to HSP 70, which has an important role as a movement protein of the virus as suggested previously[2]. The CTV ORF5-encoded, 61-kDa protein has been reported to be distantly related to HSP90[17]. The function of this protein is still unclear. Alignment of the amino acid sequences in these two chaperone-like proteins between NUagA and the other isolates demonstrated high conservation with at least 90% identity (Table 1).

ORFs 6 and 7 of NUagA encoded 240- and 223-amino-acid proteins with a calculated molecular mass of 27 kDa and 25 kDa, respectively. The ORF7-encoded, 25-kDa protein of CTV was identified as the coat protein (CP)[17], and the ORF6-encoded, 27-kDa protein was identified as a diverged copy of the CP (CPd)[7]. A comparison of these two proteins between NUagA and the other isolates demonstrated conservation of the CP and CPd with amino acid identities of at least 95%. The arginine residue located at amino acid 137 for the CP and 152 for the CPd, and the aspartic acid residue located at amino acid 178 for CP and 193 for CPd, all of which are believed to play an important role in the stabilization of the virus particles[5], were preserved in NUagA.

The remaining four ORFs downstream of the CP (ORFs 8-11) potentially encoded proteins of 167, 119, 182 and 209 amino acids with a calculated molecular weight of 18 kDa, 13 kDa, 20 kDa and 23 kDa, respectively. Amino acid alignments of these four ORF proteins in NUagA with the other isolates resulted in identities ranging from 91% to 100% (Table 1).

The 3′ UTR of the NUagA genome extends along 273

nt. Alignment of this sequence with the 3′ UTR in T36 (273 nt), VT (272 nt), T385 (273 nt) and SY568 (273 nt) showed a high degree of conservation with identity values of 97% for each. In our experiment using the 5′ RACE method, the 5′ end of the minus-strand of NUagA dsRNA was found to be GG. However, the 3′ end of the T36, VT, T385 and SY568 consensus sequences were reported to be CCA. Some sequence heterogeneity in the 3′ terminus of CTV Spain isolates[12] was also reported with a minor discrepancy in the exact 3′ terminus of the genomic RNA obtained by PCR amplification from the polyadenylated plus strands of the dsRNA of CCAT, CCAG, or CCCGA. Regardless of this heterogeneity, the 3′ terminal double C is a common characteristic in the genomes of CTV[18], *Beet yellow stunt virus* (BYSV)[11], and BYV[1] and may be important for viral RNA replication.

The length of the 5′ UTR of NUagA was 107 nt. Unlike the 3′ UTR, the 5′ UTR seemed to be polymorphic. Pairwise comparison of the NUagA 5′ UTR with the corresponding regions in VT and SY568 showed a high degree of identity (at least 94%) and somewhat lower (73%) with that of T385. In contrast, a comparison with the same region in T36 gave a nucleotide identity of only 57%. Based upon the nucleotide sequence variability in the 5′ end of CTV genome, Lopez *et al.*[12] classified isolates into three groups in which T36 and VT were the isolate type for group I and II, respectively. Various Spanish isolates belonged to group III. The 5′ UTR sequence of NUagA likely shares its characteristics with group II.

When the complete genomic sequence of NUagA was compared with those of the other isolates, a high average nucleotide identity of about 90% was shown with VT, T385 and SY568, but a low identity of about 80% with T36. The distribution of nucleotide differences along the genome was variable depending upon the isolates compared. The sequence heterogeneity between NUagA and T36 increased from ORF5 toward the 5′ end of the genome (Table 1), reaching an identity value as low as 57%, a value greater than the genomic deviation among different viruses of most other groups. This pattern was also noticed between VT and T36[15] and between T385 and T36[19].

In comparison with VT, NUagA had a relatively symmetrical distribution of nucleotide differences along its genome. The identity values were greater than 90% in all ORFs and UTRs (Table 1). This genetically close relationship likely reflects a shared biological characteristic of the isolates, both of which cause seedling yellows disease syndrome. But, it is still unclear which part of the genome is responsible for symptom expression. Comparison of NUagA with T385 and SY568 showed a somewhat uneven distribution of nucleotide divergences. The nucleotide identity values were close to 90% in most ORFs, and only in ORF2 was the value as low as 85% (Table 1). Overall, NUagA, when compared with T385 and SY568, had nucleotide divergence pattern similar to that between NUagA and VT, in which sequence differences in the 5′ half and the 3′ half of genome were relatively similar.

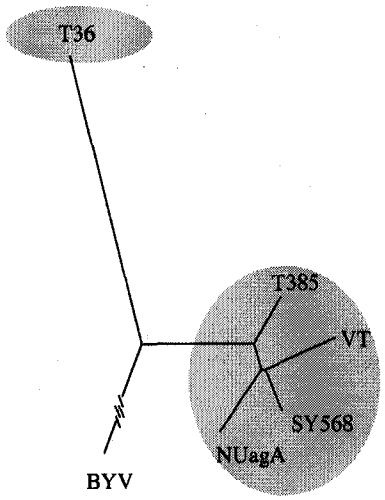


Fig. 2. Unrooted phylogenetic tree constructed from amino acid sequences of ORF1a-encoded polyprotein of CTV isolates using Clustal W[18]. Branch lengths are drawn to scale corresponding to amino acid sequence divergence. All of the bootstrap values based on 1000 replicates were more than 60%. BYV (accession no. AF056575) used as the outgroup.

The relationship of NUagA with other isolates was confirmed by the amino acid sequence identity of their ORF1a-encoded polyproteins, which is considered to be a reliable region for phylogenetic analysis because of their high sequence variability. More than 90% sequence identity between the polyprotein of NUagA and that of SY568, T385 and VT places these isolates in one cluster in phylogenetic tree (Fig. 2). This phylogenetic proximity was not evident between NUagA and T36.

Collectively, the molecular relationship between NUagA and the other isolates confirmed that CTV is a virus with a high level of genetic complexity. We hope that the data in this paper will improve our understanding of the exact genetic variants of CTV populations present in citrus from around the world.

This research was supported in part by Grant-in-Aid for Scientific Research C 09660042 & 12660041 from the Ministry of Education, Science, Sports and Culture of Japan.

## LITERATURE CITED


1. Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Maiss, E., Frotschl, R., Lunina, N.A. and Atabekov, J.G. (1994). Beet yellows closterovirus : complete genome structure and identification of a leader papain-like thiol protease. Virology 198 : 311-324.
2. Agranovsky, A.A., Folimonov, A.S., Folimonova, S.Y., Morozov, S.Y., Schiemann, J., Lesemann, D. and Atabekov, J.G. (1998). Beet yellows closterovirus HSP70-like protein mediates the cell-to-cell movement of a potexvirus transport-deficient mutant and a hordeivirus-based chimeric virus. J. Gen. Virol. 79 : 889-895.
3. Bar-Joseph, M. and Lee, R.F. (1989). Citrus tristeza virus. ABB Descriptions of Plant Viruses No. 353.
4. Broadbent, P., Brlansky, R.H. and Indsto, J. (1996). Biological characterization of Australian isolates of citrus tristeza virus and separation of subisolates by single aphid transmissions. Plant Dis. 80 : 329-333.
5. Dolja, V.V., Boyko, V.P., Agranovsky, A.A. and Koonin, E.V. (1991). Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses : two families with distinct patterns of sequence and probably structure conservation. Virology 184 : 79-86.
6. Dolja, V.V., Karasev, A.V. and Koonin, E.V. (1994). Molecular biology and evolution of closteroviruses : sophisticated build-up of large RNA genomes. Annu. Rev. Phytopathol. 32 : 261-285.
7. Febres, V.J., Ashoulin, L., Mawassi, M., Frank, A., Bar-Joseph, M., Manjunath, K.L., Lee, R.F. and Niblett, C.L. (1996). The p27 protein is present at one end of citrus tristeza virus particles. Phytopathology 86 : 1331-1335.
8. Jelkmann, W., Fechtner, B. and Agronovsky, A.A. (1997). Complete genome structure and phylogenetic analysis of little cherry virus, a mealybug-transmissible closterovirus. J. Gen. Virol. 78 : 2067-2071.
9. Kano, T., Koizumi, M. and Ieki, H. (1992). Biological properties of aphid-transmitted subcultures of citrus tristeza virus in Japan. Bull. Fruit Tree Res. Stn. 23 : 145-154.
10. Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K.C., Gumpf, D.J., Lee, R.F., Garnsey, S.M., Lewandowski, D.J. and Dawson, W.O. (1995). Complete sequence of the citrus tristeza virus RNA genome. Virology 208 : 511-520.
11. Karasev, A.V., Nikolaeva, O.V., Mushegian, A.R., Lee, R.F. and Dawson, W.O. (1996). Organization of the 3′-terminal half of beet yellow stunt virus genome and implications for the evolution of closteroviruses. Virology 221 : 199-207.
12. López, C., Ayllón, M.A., Navas-Castillo, J., Guerri, J., Moreno, P. and Flores, R. (1998). Molecular variability of the 5′- and 3′-terminal regions of citrus tristeza virus RNA. Phytopathology 88 : 685-691.
13. Mawassi, M., Gafny, R., Gagliardi, D. and Bar-Joseph, M. (1995). Populations of citrus tristeza virus contain smaller-than-full-length particles which encapsidate subgenomic RNA molecules. J. Gen. Virol. 76 : 651-659.
14. Mawassi, M., Mietkiewska, E., Hilf, M.E., Ashoulin, L., Karasev, A.V., Gafny, R., Lee, R.F., Garnsey, S.M., Dawson, W.O. and Bar-Joseph, M. (1995). Multiple species of defective RNAs in plants infected with citrus tristeza virus. Virology 214 : 264-268.
15. Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G. and Bar-Joseph, M. (1996). Unusual sequence relationships between two isolates of citrus tristeza virus. J. Gen. Virol. 77 : 2359-2364.
16. Moreno, P. and Guerri, J. (1997). Variability of citrus tristeza closterovirus (CTV) : methods for differentiating isolates. *In* Filamentous Viruses of Woody Plants (Monet, P.L., ed.). pp. 97-108, Research Signpost, Trivandum, India.
17. Pappu, H.R., Karasev, A.V., Anderson, E.J., Pappu, S.S., Hilf, M.E., Febres, V.J., Eckloff, R.M.G., McCaffery, M., Boyko, V., Gowda, S., Dolja, V.V., Koonin, E.V., Gumpf, D.J., Cline, K.C., Garnsey, S.M., Dawson, W.O., Lee, R.F. and Niblett, C.L. (1994). Nucleotide sequence and organization of eight 3′ open reading frames of the citrus tristeza closterovirus genome. Virology 199 : 35-46.
18. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Improving the sensitivity of progressive multiple sequence alignment through seqence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22 : 4673-4680.
19. Vives, M.C., Rubio, L., López, C., Navas-Castillo, J., Albiach-Marti, M.R., Dawson, W.O., Guerri, J., Flores, R. and Moreno, P. (1999). The complete genome sequence of the major component of a mild citrus tristeza virus isolate. J. Gen. Virol. 80 : 811-816.
20. Wang, W.Q., Natsuaki, T., Okuda, S. and Teranaka, M. (1988). Comparison of cucumber mosaic virus isolates by double-stranded RNA analysis. Ann. Phytopathol. Soc. Jpn. 54 : 536-539.
21. Yang, Z., Mathews, D.M., Dodds, J.A. and Mirkov, T.E. (1999). Molecular characteristization of an isolate of citrus tristeza virus that causes severe symptoms in sweet orange. Virus Genes 19 : 131-142.