Indian Journal of Biotechnology
Vol 1, October 2002, pp 350-356

Sequence Conservation in the Coat Protein Gene of Tobacco streak virus Isolates Causing Necrosis Disease in Cotton, Mung bean, Sunflower and Sunn-hemp in India

A I Bhat[1], R K Jain[5*], V Chaudhary[5], M Krishna Reddy[2], M Ramiah[3], S N Chattannavar[4] and A Varma[5]

[1]Indian Institute of Spices Research, Marikunnu, Calicut 673 012, India

[2]Indian Institute of Horticultural Research, Bangalore 560 089, India

[3]Tamil Nadu Agricultural University, Coimbatore 641 003, India

[4]University of Agricultural Sciences, Agricultural Research Station, Dharwad 580 005, India

[5]Advanced Center for Plant Virology, Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi 110 012, India

Received 4 April 2002; accepted 20 August 2002

Natural infection of *Tobacco streak virus* (TSV) in cotton, mung bean, sunflower and sunn-hemp, collected from different locations in India, was detected by reverse transcription-polymerase chain reaction (RT-PCR). The coat protein (CP) gene sequences of the six TSV isolates originating from different hosts and locations were amplified. The resulting amplicons were cloned and sequenced to assess molecular variability. The sequence and phylogenetic analyses revealed that the CP gene among TSV isolates collected from different hosts and locations was highly conserved (99-100%), suggesting a common origin.

Keywords: *Tobacco streak virus*, cotton, mung bean, sunflower, sunn-hemp, coat protein gene sequence

Introduction

Tobacco streak virus (TSV), belonging to genus *Ilarvirus* (Subgroup I) of family *Bromoviridae*, is characterized by icosahedral particles measuring 27-35 nm in diameter. Its genome consists of three single stranded positive sense RNA species of 2.9, 2.7 and 2.2 kb designated as RNA-1, RNA-2 and RNA-3 respectively. The coat protein, coded by RNA-3, is required for infectivity (Van Vloten-Doting, 1975; Brunt *et al*, 1996).

TSV is widely distributed in the North American and the Pacific regions where it infects *Asparagus officinalis, Dahlia* spp, *Glycine max, Gossypium herbaceum, Melilotus alba, Nicotiana tabacum, Phaseolus vulgaris, Rosa setigera* and *Trifolium pratense* causing a variety of symptoms (Brunt *et al*, 1996). Recently, TSV has also been found to occur in the Indian subcontinent causing serious diseases of sunflower and groundnut. In sunflower (*Helianthus annuus*), it causes extensive necrosis of the leaf lamina, petiole, stem and floral calyx with malformation of flowering head (Ramiah *et al*, 2001 a & b). The disease has become increasingly important in sunflower growing regions of Andhra Pradesh, Karnataka, Maharashtra and Tamil Nadu (Anon, 2001). The causal agent has been identified as TSV based on particle morphology, biological properties, serology and coat protein sequence identity (Ramiah *et al*, 2001 b; Bhat *et al*, 2002). In groundnut (*Arachis hypogaea*), the virus causes necrosis of stems and terminal leaflets and death of the affected plants (Reddy *et al*, 2002). The coat protein gene of the TSV isolate from sunflower shares 88% sequence identity with TSV-WC (X00435) and 58% identity with *Hydrangea mosaic virus* (U35145), which is the other member of subgroup I of *Ilarvirus* (Bhat, *et al*, 2002). The isolate from groundnut is also reported to have nearly 88% sequence identity with TSV-WC (Reddy *et al*, 2002). TSV appears to be spreading fast in India, as it has been found to also occur in cotton (*Gosspium hirsutum*), mung bean (*Phaseolus aureus*) and sunn-hemp (*Crotalaria juncea*) causing necrosis of veins, petiole, stem and other parts. Considering the seriousness of the diseases caused by TSV in a variety of crops in India, it is essential to examine diversity among the TSV isolates so that crop

*Author: for correspondence:
Tel: 5862134; Fax: 5823113
E-mail: rkjain@bic-iari.ren.nic.in

varieties with the durable resistance could be developed for minimizing the losses caused by TSV.

Materials and Methods

Virus Isolates

Samples were collected from necrosis disease affected plants of cotton, mung bean, sunflower, and sunn-hemp from different parts of the country (Table 1). The presence of TSV in these samples was confirmed by direct antigen coated-ELISA (DAC-ELISA) using TSV polyclonal antiserum developed at Indian Agricultural Research Institute, New Delhi (Bhat *et al*, 2001), and finally three isolates from sunflower and one each from cotton, mung bean and sunn-hemp were selected for further studies.

Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

Total RNAs extracted from ELISA positive necrosis affected tissues from different hosts using R Neasy kit (Qiagen Inc, Chatsworth, CA, USA) were used as template for RT-PCR. RT and PCR were performed in the same tube without any buffer changes as described earlier (Pappu *et al*, 1993). The primer pair derived from the CP gene sequence of TSV (Bhat *et al*, 2002) was used to prime the amplification. The genome sense primer 5′ ATGAATACTTTGATCCAAGG 3′ was derived from the beginning of the first 20 bases of the coding region. The genome antisense primer, 5′ TCAGTCTTGATTCACCAG 3′ represented last 18 bases of the coding region of the CP gene. The PCR reaction (100 µl) contained 200 ng each of the primers, 16 to 20 units RNasin (MBI Fermentas Inc), 10 units Omniscript reverse transcriptase (Qiagen Inc), 2.5 units Taq Polymerase (Qiagen Inc), 1×PCR buffer (Qiagen Inc), 10 m*M* Dithiothreitol, 1×'Q' solution (Qiagne Inc) and 10 µ*M* each of the dNTPs (Qiagen Inc). PCR mix (46 µl) containing the above components was added to the tubes containing the template RNA (54 µl) resulting in a final reaction volume of 100µl. Amplification was performed in an automated thermal cycler (Power Block II, Ericomp Inc, San Diego, CA, USA) programmed for one cycle of 42°C for 45 min for cDNA synthesis and 40 cycles of amplification with the following parameters: 30 Sec of denaturation at 94°C, 1 min of annealing at 48°C, and 1 min of extension at 72°C; this was followed by one cycle of final extension for 60 min at 72°C.

cDNA Cloning and Sequencing of the CP Gene

Following PCR, reaction products (10µl) were analysed by 1% agarose gel elctrophoresis in Tris–acetate EDTA (TAE) buffer containing ethidium bromide. DNA was visualised and photographed using a UV transilluminator and a gel documentation apparatus (UVP Image Store 5000, Germany). One kb ladder (MBI Fermentas Inc) was used as a size standard. The ca. ~700 bp band was excised and purified using Qiax II gel extraction kit (Qiagen Inc,). The purified PCR product was ligated into p GEM-T

Table 1 — Source of coat protein nucleotide and amino acid sequences of *Tobacco streak virus* isolates originating from different hosts and locations used for comparisons

Original host	Location State/Country*	Acronym Used	GenBank accession number
Tobacco	Wisconsin, USA	TSV-WC	X00435
Sunflower	Andhra Pradesh, India	TSV-SF-Ap	AY061930 (This Study)
Sunflower	Karnataka, India	TSV-SF-Ka	AY061929 (This Study)
Sunflower	Maharashtra, India	TSV-SF-Mh	AY061928 (This Study)
Sunflower	Tamil Nadu, India	TSV-SF-Tn	AF400664 (This Study)
Mung bean	Tamil Nadu, India	TSV-MB-Tn	AF515823 (This Study)
Cotton	Maharashtra, India	TSV-CT-Mh	AF515824 (This Study)
Sunn-hemp	Karnataka, India	TSV-SH-Ka	AF515825 (This Study)

* Except for the isolate from tobacco (Cornelissen *et al*, 1984), all the isolates used in this study were from India.

Easy vector (Promega, Madison, WI, USA) and competent *Escherichia coli* (strain DH5α) were transformed by following standard molecular biology procedures (Sambrook & Russell, 2001). Recombinant clones were identified by restriction endonuclease digestion, and selected clones were sequenced at the automated DNA sequencing facility at the Department of Biochemistry, University of Delhi, South Campus, New Delhi, India.

Sequence Analyses

Sequence data were compiled using Seqaid Version 3.6 (Rhoads & Roufa, 1985). Multiple sequence alignments were made using CLUSTAL W (Thompson *et al*, 1994). Sequence phlylograms were constructed using PHYLIP package (Bootstrap analysis with 1000 replicates), and unrooted trees were generated using TREEVIEW software (Page, 1996). The CP nucleotide and amino acid sequences of other TSV isolates used for comparison are given in Table 1. The BLAST programme (Altschul *et al*, 1990) was used to identify related sequences available from the GenBank data base.

Results and Discussion

The CP gene of all the six isolates was successfully amplified by RT-PCR, giving a product of the expected size (ca. ~700 bp). The identity of amplicons was confirmed by sequencing. The nucleotide sequences of the TSV isolates were deposited at GenBank and their accession numbers are listed in Table1. The sequenced region in all the six isolates contained a single open reading frame of 717 bases that could potentially code for a protein of 238 amino acids (Fig.1). This was in agreement with a previous report of TSV CP gene from sunflower collected from Tamil Nadu (Bhat *et al*, 2002).

Comparative sequence analyses showed that CP gene sequences of all the six isolates of TSV from cotton, mung bean, sunflower and sunn-hemp were highly conserved, with sequence identity varying from 99-100% both at the nucleotide and amino acid levels. This also explains close serological affinity observed between these isolates. Only at two positions (amino acid 184 and 236) conserved amino acid residues aspartic acid and asparagine were substituted by glycine and serine in sunflower isolate

Table 2 — Differences among the coat protein amino acid sequence of the *Tobacco streak virus* (TSV) isolates

Amino acid position/ majority residue	Isolate*		
	TSV-WC	TSV-SF-Mh	TSV-MB-Tn
19/T	A		
23/F	S		
26/N	-		
33/F	I		
39/V	M		
42/G	N		
47/A	N		
48/A	T		
49/A	V		
66/F	Y		
73/M	V		
74/T	P		
75/A	L		
77/I	L		
81/P	S		
89/T	A		
92/K	R		
112/I	V		
131/N	G		
132/S	A		
137/Y	F		
184/D		G	
198/I	V		
208/D	E		
236/N			S

*No difference was observed in other isolates; - indicates deletion

```
TSV         ATGAATACTTTGATCCAAGGTCCAGACCATCCATCCAACGCCATGTCTTCGCGTGCTAAC
TSV-SF-Ap   **************************************************C***A*****
TSV-SF-Ka   **************************************************C***A*****
TSV-SF-Mh   **************************************************C***A*****
TSV-SF-Tn   **************************************************C***A*****
TSV-MB-Tn   **************************************************C***A*****
TSV-CT-Mh   **************************************************C***A*****
TSV-SH-Ka   **************************************************C***A*****

TSV         AACCGCTCAAATAA---CAGCAGATGCCCAACTTGCATTGATGAGTTGGACGCTATGGCC
TSV-SF-Ap   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-SF-Ka   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-SF-Mh   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-SF-Tn   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-MB-Tn   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-CT-Mh   *******TT**C**CAA******************TT*C**C********T**AG*A**G
TSV-SH-Ka   *******TT**C**CAA******************TT*C**C********T**AG*A**G

TSV         AGGAATTGTCCCGCCCATAATACCGTGAACACTGTTTCACGACGCCAGCGGCGTAATGCC
TSV-SF-Ap   ***GG***C*****T**CGC*C***C************G*****T********A******
TSV-SF-Ka   ***GG***C*****T**CGC*C***C************G*****T********A******
TSV-SF-Mh   ***GG***T*****T**CGC*C***C************G*****T********A******
TSV-SF-Tn   ***GG***C*****T**CGC*C***C************G*****T********A******
TSV-MB-Tn   ***GG***C*****T**CGC*C***C************G*****T********A******
TSV-CT-Mh   ***GG***C*****T**CGC*C***C************G*****T********A******
TSV-SH-Ka   ***GG***C*****T**CGC*C***C************G*****T********A******

TSV         GCTAGAGCTGCCGCGTATAGAAACGCGAATGCTAGAGTACCGCTACCGCTTCCTGTGGTA
TSV-SF-Ap   ****************T***G************C**A*GA*CGC***AA**********G
TSV-SF-Ka   ****************T***G************C**A*GA*CGC***AA**********G
TSV-SF-Mh   ****************T***G************C**A*GA*CGC***AA**********G
TSV-SF-Tn   ****************T***G************C**A*GA*CGC***AA**********G
TSV-MB-Tn   ****************T***G************C**A*GA*CGC***AA**********G
TSV-CT-Mh   ****************T***G************C**A*GA*CGC***AA**********G
TSV-SH-Ka   ****************T***G************C**A*GA*CGC***AA**********G

TSV         TCGGTTTCCCGCCCTCAAGCGAAGGCGTCGTTGAGGTTACCCAACAATCAAGTTTGGGTA
TSV-SF-Ap   C************************A*A********A*C************************
TSV-SF-Ka   C************************A*A********A*C************************
TSV-SF-Mh   C************************A*A********A*C************************
TSV-SF-Tn   C************************A*A********A*C************************
TSV-MB-Tn   C************************A*A********A*C*************G*********
TSV-CT-Mh   C************************A*A********A*C*************T*********
TSV-SH-Ka   C************************A*A********A*C************************

TSV         ACTCGCAAAGCGAGTGAATGGTCTGCAAAGACTGTTGATACCAACGATGCTATCCCCTTC
TSV-SF-Ap   ********************************CA***************************
TSV-SF-Ka   ********************************CA***************************
TSV-SF-Mh   ********************************CA***************************
TSV-SF-Tn   ********************************CA***************************
TSV-MB-Tn   ********************************CA***************************
TSV-CT-Mh   ********************************CA***************************
TSV-SH-Ka   ********************************CA***************************

TSV         AAAACCATAGTCGAGGGGATTCCCGAAATCGGTGCTGAGACGAAGTTTTTCCGTCTCTTG
TSV-SF-Ap   **G*****************************AA*T*G************A*******C*A
TSV-SF-Ka   **G*****************************AA*T*G************A*******C*A
TSV-SF-Mh   **G*****************************AA*T*G************A*******C*A
TSV-SF-Tn   **G*****************************AA*T*G************A*******C*A
TSV-MB-Tn   **G*****************************AA*T*G************A*******C*A
TSV-CT-Mh   **G*****************************AA*T*G************A*******C*A
TSV-SH-Ka   **G*****************************AA*T*G************A*******C*A
```

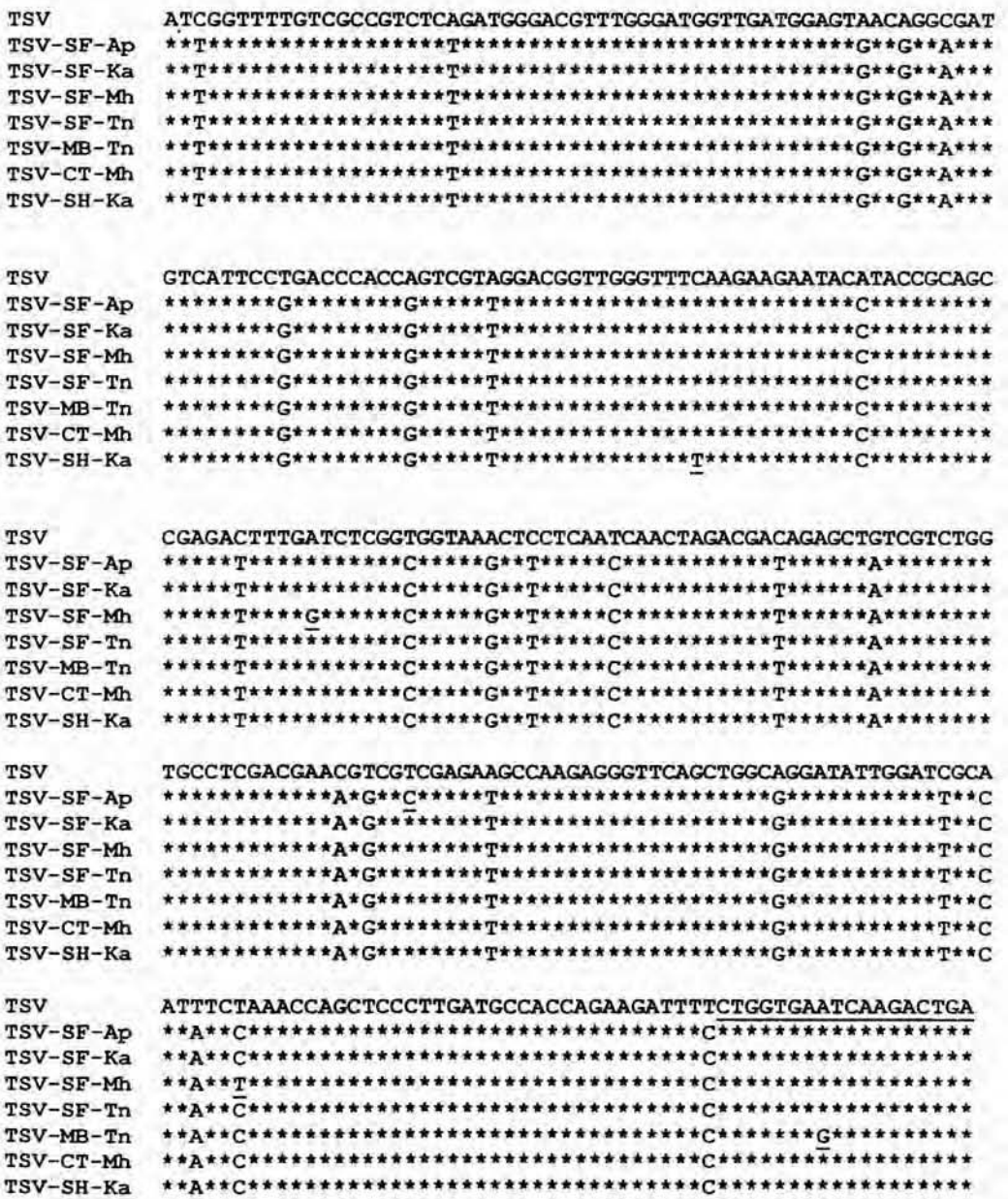

Fig. 1 — Multiple alignment of coat protein genes (shown as DNA) of *Tobacco streak virus* (TSV) isolates from India. Sequences for comparison were obtained from GenBank (Benson *et al*, 1999). The alignment was generated by CLUSTAL W (Thompson *et al*, 1994. Asterisk indicates identity and a dash indicates deletion at a given position. The nucleotide differences with in TSV isolates from India are underlined. The primer sequences used for amplification are underlined. The TSV-WC isolate (Cornelissen *et al*, 1984) was used as representative isolate for comparison. Designations given to each TSV isolate used in the study were given in Table 1.

from Maharashtra (TSV-SF-Mh) and mung bean isolate from Tamil Nadu (TSV-MB-Tn), respectively (Table 2). The role of these amino acid substitutions in serological affinity and symptom expression remains to be examined. The sequence identity with the groundnut isolate (Reddy *et al*, 2002) could not be compared due to the non- availability of its sequence in the GenBank. However, as the groundnut isolate was also collected from sunflower growing regions it is expected to have close sequence identity with the other isolates. By contrast, a higher level of sequence divergence (up to 12%) was observed when Indian TSV isolates were compared with an isolate from USA (Cornelissen *et al*, 1984). Sequence divergence

was predominantly clustered near the amino terminal region of coat protein (amino acids 19-49), with scattered substitution in the remaining portion of the sequence (Table 2).

The results of the multiple alignments were used to generate a dendrogram illustrating phylogenetic relationship. Dendrogram based on nucleotide as well as deduced amino acid sequences of the CP gene showed that the TSV isolates originating from different hosts and locations in India formed one cluster, distinct from the TSV isolate from USA. Highly conserved CP gene among the Indian TSV isolates suggests their common origin.

The virus seems to have been introduced to the country recently, as the diseases caused by TSV are so obvious and serious that these would not have been missed if these had been prevalent in the region earlier. Since high rates of seed transmission of TSV have been reported for soybean (up to 90 %) and common bean (26%) (Frison *et al*, 1990), there is a possibility of its introduction through the seeds of some susceptible crop. Seed-borne infection of TSV in sunflower has been detected by ELISA in two genotypes, PAC1091 and KBSH44. The incidence was higher in KBSH44 (100 positive out of 369; A405:0.11-0.51) than in PAC1091 (6 positive out of 240; A405:0.16-0.46) (Unpublished observations). But, its role in seed transmission needs to be examined.

The present study based on CP gene of six TSV isolates originating from different hosts and locations has provided only a partial genetic make up of naturally occurring TSV population. Sequence comparisons of other ORF of RNA 3 (3a- the movement protein gene) or the intergenic region would be required to ascertain the diversity among isolates. It remains to be demonstrated that whether sequence differences could be applied to resolve pathotypes and serotypes. A number of additional TSV isolates from different hosts and locations will have to be examined to establish correlation between molecular typing and serotypes and symptom types. Such correlation has been established in *Prunus necrotic ringspot virus* by comparing CP gene and movement protein gene sequences of different serotypes and pathotypes (Hammond &Crosslin, 1998; Vaskova *et al*, 2000).

Acknowledgement

Authors are thankful to Dr D V Singh, Head, Division of Plant Pathology, IARI, New Delhi for his continuous encouragement throughout the study. They also thank S Chander Rao, Directorate of Oilseeds Research, Hyderabad and H M Venkatesh, Syngenta India Ltd, Aurangabad for supplying necrosis infected samples. Financial support from the World Bank by way of National Agricultural Technology Project to Indian Council of Agricultural Research is greatly acknowledged. A I Bhat is supported by a NAAS-TATA Young Scientists' Research Fellowship of National Academy of Agricultural Sciences, India.

References

Altschul S F *et al*, 1990. Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

Anon, 2001. Annual Report (2000-2001). Directorate of Oilseeds Research, Hyderabad , India. Pp 182.

Benson D A *et al*, 1999. GenBank. *Nucleic Acids Res*, **27**, 12-17.

Bhat A I *et al*, 2002. Serological and coat protein nucleotide sequence studies suggest that necrosis disease on sunflower in India is caused by a strain of *Tobacco streak ilarvirus*. *Arch Virol*, **147**, 651-658.

Bhat A I *et al*, 2001. Development of serological based assays for the diagnosis of sunflower necrosis disease. *Annal Plant Prot Sci*, **9**, 292-296.

Brunt A A *et al*, 1996. Viruses of Plants. CAB International, Wellingford, U K .Pp 1484.

Cornelissen B J C *et al*, 1984. Complete nucleotide sequence of tobacco streak virus RNA3. *Nucleic Acids Res*, **12**, 2427-2436.

Frison E A *et al* (Eds), 1990. FAO/IBPGR Technical Guidelines for the safe movement of legume germplasm. Food and Agricultural Organization of the United Nations, Rome/International Board for Plant Genetic Resources, Rome.

Hammond R W & Crosslin James M,1998. Virulence and molecular polymorphism of Prunus necrotic ringspot virus isolates. *J Gen Virol*, **79**, 1815-

Page R D M, 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci*, **12**, 357-358.

Pappu S S *et al*, 1993. A polymerase chain reaction method adapted for selective cloning of 3' non translated regions of potyviruses: application to dasheen mosaic virus. *J Virol Methods*, **41**, 9-20.

Ramiah M *et al*, 2001a. Isolation of an isometric virus causing sunflower necrosis disease in India. *Plant Dis*, **85**, 443.

Ramiah M *et al*, 2001b.Partial characterization of an isometric virus causing sunflower necrosis disease. *Indian Phytopathol*, **54**, 246-250.

Reddy A S *et al*, 2002. Occurrence of *Tobacco streak virus* on Peanut (*Arachis hypogaea*) in India. *Plant Dis*, **86**, 173-178.

Rhoads D D & Roufa D S, 1985. Emetine resistance of Chinese hamster cells: structure of wild type and mutant ribosomal protein S14 messenger RNA species. *Mol Cell Biol*, **5**, 1655-1659.

Sambrook J & Russell D W, 2001. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

Thompson J D *et al*, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Van Vloten-Doting L, 1975. Coat protein is required for infectivity of *Tobacco streak virus* : Biological equivalence of the coat proteins of *Tobacco streak* and *Alfalfa mosaic viruses*. *Virology*, **65**, 215-225.

Vaskova D *et al*, 2000.Variability and molecular typing of the woody-tree infecting prunus necrotic ringspot ilarvirus. *Arch Virol*, **145**, 699-709.