VirusDis.
https://doi.org/10.1007/s13337-018-0500-2

ORIGINAL ARTICLE

Complete genome sequence and phylogenetic relationships of tobacco streak virus causing groundnut stem necrosis disease in India

Daliyamol[1] · A. Abdul Kader Jailani[1] · K. Vemana[2] · Anirban Roy[1] · M. Krishnareddy[3] · Kappei Kobayashi[4] · Bikash Mandal[1]

Received: 25 October 2018 / Accepted: 10 November 2018

Abstract Tobacco streak virus (TSV, genus *Ilarvirus* family *Bromoviridae*) is known to cause stem necrosis disease (SND) in groundnut (*Arachis hypogaea*) since 2000 in Southern India. The TSV isolate infecting groundnut so far has not been characterized based on the complete genome sequence. In this study, TSV was isolated from a naturally infecting groundnut plant in Kadiri, the hot-spot of the SND in southern India. During the *Kharif* season of 2014, groundnut plants in an experimental field were affected with chlorosis and necrosis in leaf, stem and buds. The cent percent of the 48 samples with these symptoms collected from the field tested positive for TSV in ELISA samples in this context. One isolate, GN-Kad was established from a single lesion on cowpea cv. C-152 through successive sap inoculation. Cloning and sequencing of coat protein gene (717 nucleotides) of the isolate showed high sequence identity (98–99%) with the TSV isolates reported from different crops in India. The isolate produced local necrotic rings or veinal necrosis following sap inoculation to cowpea (cultivars C-152, Pusa Komal, Pusa Sukomal and Krishi Kanchan), French bean and sunflower; whereas, it produced systemic chlorotic mottling symptoms in *Nicotiana benthamiana*. The three segments of the virus genome (RNA 1, RNA 2 and RNA 3) contained 3523, 2903 and 2232 nucleotides, respectively. The overall genome sequence (8639 nt) of the present isolate shared 77–99% of nucleotide sequence identity with that of the other seven isolates reported from Australia, India and USA. The GN-Kad shared very close phylogenetic relationship with the okra and pumpkin isolates reported from India. The present report is the first comprehensive study of the molecular characterization of TSV associated with the stem necrosis disease of groundnut.

Keywords Stem necrosis disease · Groundnut · Tobacco streak virus · Complete genome · India

Introduction

Groundnut (*Arachis hypogaea*), an important grain legume and oil seed crop in India, is seriously affected the stem necrosis disease (SND) caused by tobacco streak virus (TSV). The stem necrosis disease was first reported in Ananthapuramu district of Andhra Pradesh, India in 2000. Symptoms in groundnut first appear on young leaves as necrotic lesions and veinal necrosis. The necrosis later spreads to petiole and stem. Necrotic lesions on the stem later spread upwards killing the bud. Majority of the plants infected within a month after sowing die due to severe necrosis. In 2000, the SND affected nearly 225,000 ha resulting in an estimated crop loss of US $65 million [12]. TSV has emerged in India as an important virus infecting several other economically important crops such as bottle

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13337-018-0500-2) contains supplementary material, which is available to authorized users.

✉ Bikash Mandal
leafcurl@rediffmail.com

[1] Division of Plant Pathology, Advanced Centre for Plant Virology, Indian Agricultural Research Institute, New Delhi 110012, India

[2] Agricultural Research Station, Acharya NG Ranga Agricultural University, Ananthapuramu, Kadiri, Andhra Pradesh 515591, India

[3] Division of Plant Pathology, Indian Institute of Horticultural Research, Hessaraghatta, Bangalore 560089, India

[4] Plant Molecular Biology and Virology, Graduate School of Agriculture, Ehime University, Tarumi, Matsuyama, Ehime 790-8566, Japan

Published online: 27 November 2018

gourd, cotton, gherkin, mungbean, okra, pumpkin, safflower, soybean, sun-hemp and urdbean [1–4, 6–9, 11, 13].

TSV (family *Bromoviridae,* genus *Ilarvirs*) is a quasi-isometric plant virus of 30 nm size containing a segmented positive sense single-stranded RNA genome. The genome is divided into RNA 1 (3.5 kb), RNA 2 (2.9 kb), and RNA 3 (2.2 kb). The monocistronic RNA 1 encodes for RNA-dependent RNA polymerase that is responsible for virus replication. The dicistronic RNA 2 encodes the non-structural 2a protein necessary for viral replication and an overlapping 2b protein. The RNA 3 genome encodes two proteins, movement protein (MP) and coat protein (CP).

The association of TSV with groundnut in southern India has been reported based on the partial sequence MP and CP genes [12]. The complete genome sequence of TSV causing SND of groundnut is lacking. In this study, TSV (GN-Kad isolate) was isolated from a naturally infected groundnut plant from a hot-spot in Andhra Pradesh. The biological property of the GN-Kad isolate was determined through sap transmission to the different plant species. The complete genome sequence of all the three RNA segments was determined and the phylogenetic relationships were established with the globally known isolates of TSV.

Materials and methods

Virus source, sap transmission and confirmation

Naturally infected groundnut plant samples showing severe SND were collected from the Agricultural Research Station, Kadiri, Ananthapuramu, Andhra Pradesh in 2014. For the transmission of the virus, the symptomatic leaves of groundnut were homogenized in 0.01 M sodium phosphate buffer (pH 7.0) containing 0.01 M sodium sulfite and rub-inoculated cowpea (*Vigna unguiculata*) seedlings pre-dusted with Carborundum, 600 mesh. The virus was subsequently isolated from a single lesion and maintained on cowpea in a greenhouse at 25–28 °C.

TSV was detected in the field as well as experimental samples by direct antigen coating enzyme linked immunosorbent assay (DAC-ELISA) [5] using polyclonal antibodies (PAbs) (Bioreba, Reinach, Switzerland). Further, the association of TSV was confirmed by cloning and sequencing of CP gene. For this, RT-PCR was performed with forward (5′-atgaatactttgatccaagg-3′) and reverse (5′-tcagtcttgattcaccag-3′) primers using one-step Qiagen RT-PCR kit (Qiagen Inc., Chatsworth, CA, USA). Amplification was performed in an automated PCR machine (GeneAmp® PCR system 9700, PE Applied Biosystems) programmed for one cycle of cDNA synthesis at 42 °C for 45 min, followed by denaturation of reverse transcriptase at 95 °C for 2 min and 35 cycles of denaturation at 94 °C for 30 s, annealing at 48 for 45 s, and extension at 72 °C for 1 min, followed by a cycle of final extension at 72 °C for 10 min. The amplicon was cloned into pGEM-T Easy vector (Promega, Madison, WI, USA).

Analysis of host responses

In order to know the host responses to the TSV-GN Kad isolate, 10 seedlings each of the plant species were mechanically inoculated as previously described. Symptoms were recorded and both inoculated and non-inoculated leaves were tested by RT-PCR using TSV-CP specific primers.

Amplification of TSV genome

The complete genome of TSV was amplified with eight pairs of primers, which were prepared based on the sequence information available in the NCBI database (Table 1).The details of the primer pairs along with the expected PCR products from all the three RNAs (RNA 1, 2 and 3) of TSV and their position and orientation have been shown as bars in Table 2 and Supplementary Fig. 1, respectively.

Total RNA was extracted from symptomatic leaf samples using RNeasy Plant Mini Kit (Qiagen, Chatsworth, CA, USA). The RNA concentrations of the samples were estimated by Nanodrop (Thermo scientific, Waltham, MA, USA). The RNA was used to synthesize double-stranded cDNA using the cDNA Synthesis System kit (Roche, Indianapolis, IN, USA). The reverse transcription was performed using a Superscript III Reverse Transcription Kit (Invitrogen, USA). PCR amplification was carried out using fragment specific forward and reverse primers in a total volume of 20 µl using *Taq* DNAzyme II (Thermo Scientific, USA) and *Pfu* DNA polymerase (NEW England Biolabs, USA) Amplification of all RNA fragment was obtained in 35 cycles of the following profile: 94 °C for 2 min, 94 °C for 30 s, annealing at the temperature appropriate to primers (50,52 or 54 °C) and elongation at 72 °C for 2 min with a final extension step at 72 °C for 10 min.

Rapid amplification of cDNA end (RACE)

A 5′-end cDNA synthesis was performed using RLM RACE kit (Invitrogen, USA). For the first strand synthesis, approximately 1.5–3.0 µg (total volume, 20 µl) of each viral RNA of TSV-GN Kad was denatured for 2 min at 70 °C in a reaction volume containing 1 µl of gene-specific primer (BM 704F and BM 766R) and 5 µl of dH2O. After cooling on ice for 10 min, the following first strand cDNA

Table 1 Complete genome sequences of RNA 1, RNA 2 and RNA 3 of TSV isolate/strains used for comparison and phylogenetic analysis

Isolate/strain	Genomic fragments					
	RNA 1	Accession no.	RNA 2	Accession no.	RNA 3	Accession no.
TSV-GN	3523	KT766195	2903	KT766196	2208	MH895294
Okra-India	3523	FJ561302	2903	FJ561303	2213	FJ561304
Pumpkin-India	3523	FJ561299	2903	FJ561300	2213	FJ561301
Squash-USA	3525	KM504246	2898	KM504247	2211	KM504248
Soybean-USA	3491	FJ403375	2911	FJ403376	2216	FJ403377
Tobacco-USA	3482	JX073656	2911	JX073657	2216	JX073658
Sunflower-Australia	3512	JX463334	2922	JX463335	2215	JX463336
Aster-Australia	3481	JX463337	2901	JX463338	2208	JX463339
AMV	3476	NC003464	2979	NC003465	2056	NC003480

Table 2 List of primers used in RT PCR to amplify the genome of TSV-GN isolate

Fragments	Primer	Primer size (bp)	Primer sequence (5′–3′)	PCR product size (bp)
RNA 1	5′R1- BM 765R(5′RACE)	22	GACGTTYAACTTCTGGAAATCA	290
	F1-BM 704 F	22	ATGGATTCTCGTTCATTACCCA	1852
	R1-BM 766 R	22	TCTCTTCCATCATTTCCATGTC	
	F2-BM 767 F	21	GTGAAGTGATTTACACGGTCC	1802
	R2-BM 493 R	23	GCATCTCCTTTAAAGGAGGCATT	
RNA 2	5′R2-BM 768R(5′RACE)	20	GTRTAGTCCACGTCAAGATC	270
	F3-BM 705 F	22	ATGGATTCCGTTATAAAGAACC	1652
	R3-BM 769 R	19	CATCGCCACTGGCGGCAAC	
	F4-BM 770 F	20	CGAMGAACTGGTGATGCATG	1523
	R4-BM 495 R	20	GCATCTCCATTTGGAGGCAT	
RNA 3	F5-BM 496 F	24	GTATTCTCCGAGCTTTAGATACCA	1200
	R5-BM 825 R	21	GACATGGCGTTGGATGGATGG	
	F6-BM 826 F	22	GTGAGACGAGTATTAAGTRGAT	1200
	R6-BM 497 R	22	GCATCTCCTATAAAGGAGGCAT	

synthesis mixture [1 × first strand reaction buffer, 0.5 mM DTT, 1 mM dNTP, 0.8 units RNase inhibitor, 0.2 units AMV (TaKaRa)] was added and incubated at 42 °C for 1 h. For amplification of cDNA, 48 µl of the PCR reaction mixture [1 × *Taq* polymerase buffer (TaKaRa), 2 mM dNTP, 5 mM MgCl2, 0.5 µM Universal primer mix (UPM; Clontech), 0.1 µM, gene specific primers, and EX-*Taq*DNA polymerase (TaKaRa)] was added to the 2 µl of the first strand cDNA solutions. PCR cycling consisted of 94 °C for 40 s, 52 °C for 40 s, and 68 °C for 10 min (35 cycles). The last cycle was followed by a prolonged extension for 20 min at 68 °C to complete synthesis of the cDNA products. The amplified DNAs were purified using Promega SV Gel Kit, (Promega, USA).

Cloning and sequence analysis

Each of the purified amplified fragments was ligated in pGEM-T Easy Vector (Promega, USA) and *Escherichia coli* DH5α was transformed with the recombinant vector. The clones were confirmed by colony-PCR and plasmid DNA was prepared using a QIAprep Spin Miniprep Kit (QIAGEN, USA). The clones were sequenced in both the directions using dye terminator cycle sequencing (Applied Biosystems 3130 Genetic Analyzer, USA).

The sequences of the cloned fragments were assembled using BioEdit (http://bioedit.software.informer.com). Multiple sequence alignment was performed in the ClustalW program. Phylogenetic analyses based on both nucleotide and predicted amino acids sequences were performed with MEGA 6 package using the maximum parsimony method with 1000 bootstrap replicates and visualized in the tree view (http://www.megasoftware.net/mega4). Percent

sequence identities was calculated using MegAlignTM. Nucleotides (nt) and amino acids (aa) sequences representing RNA 1, RNA 2, and RNA 3 of TSV-GN isolate were compared with other TSV strains reported in the GenBank database (Table 4). Apple mosaic virus (AMV) was used as an out-group in the phylogenetic tree.

Results

Field detection and typing of the isolate

During *Kharif* 2014, groundnut plants were observed exhibiting various types of symptoms in the experimental field at Kadiri, Ananthapuramu district, Andhra Pradesh. The symptoms included chlorosis and necrotic lines or spots on leaves, necrotic stem and bud, proliferation of axillary leaves and death of plants (Fig. 1). The pods from the symptomatic plants showed necrotic lesions on the immature shell. The 48 samples showing these symptoms reacted positively in the DAC-ELISA with the PAbs to TSV, however, none of these samples reacted with the PAbs to groundnut bud necrosis virus (GBNV) (Table 3). TSV was successfully transmitted from one of the groundnut samples showing high ELISA reading through sap inoculation to cowpea cv. 152. The typical necrotic rings developed following sap inoculation on cowpea at 3 dpi (days post inoculation). Subsequently, a pure culture of TSV was established by transferring the virus from a single lesion and maintained on cowpea as a TSV-GN-Kad isolate.

RT-PCR with the CP gene based primers resulted in amplification of ~ 750 bp fragment from the symptomatic leaf tissues of cowpea, whereas no such amplification was obtained in the healthy leaf. Cloning and sequence analysis showed that the amplified product was 717 nucleotides long coding for a protein of 238 amino acids, which was similar to the size of CP gene of the TSV isolates characterized from India and Japan. Comparative sequence analysis revealed that the CP gene of TSV-GN-Kad isolate shared 98–99% identity at nucleotide level with that of the TSV isolates reported from other crop species and locations in India. The CP gene sequence of 51 isolates of TSV available in the GenBank was used to generate phylogenetic relationships between the isolates (Table 4). Of these, only one isolate originated from Dahlia in Japan and rest of the isolates were from India infecting various crops like pulses, vegetables, oilseeds and ornamentals. The phylogenetic analysis revealed that the TSV Indian isolates along with Japan isolate showed a completely different

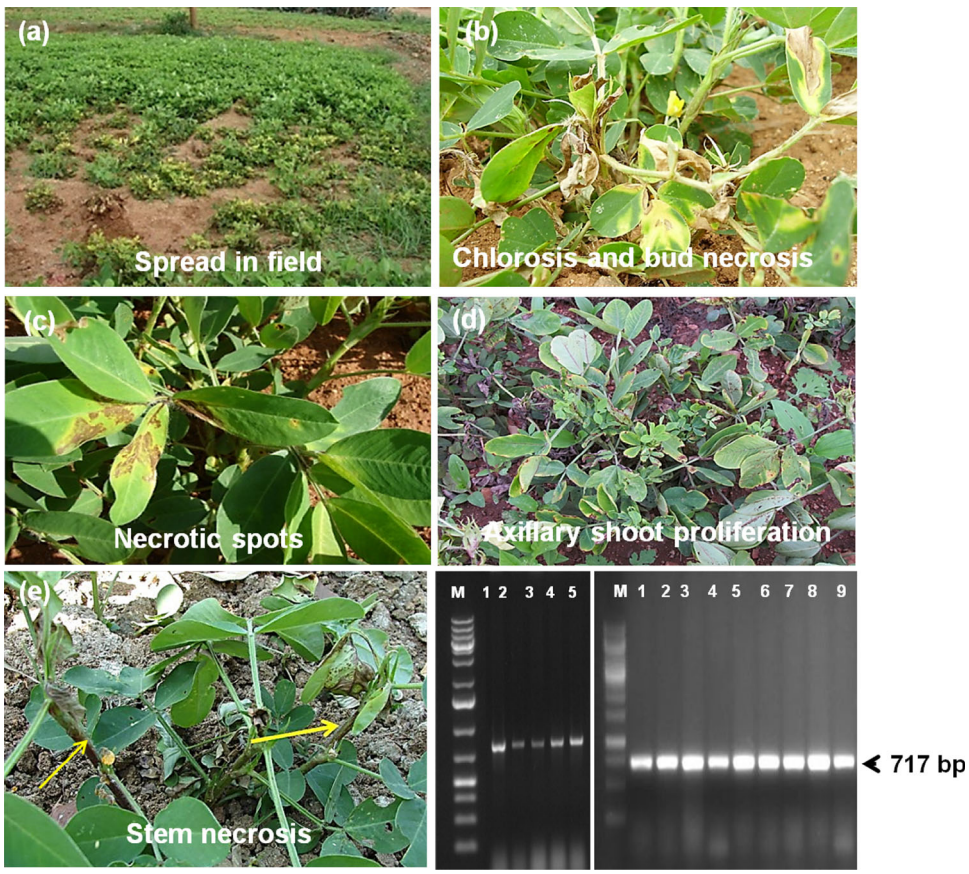

Fig. 1 Distribution and different symptom expression of tobacco streak virus during the *Kharif* season, 2014 at Kadiri, Ananthapuramu, Andhra Pradesh. **a** TSV infected field view, **b** chlorosis and bus necrosis, **c** necrotic spots on leaves, **d** axillary shoot proliferation, **e** stem necrosis, **f** RT-PCR of tested field samples with CP primer

Table 3 Detection of tobacco streak virus (TSV) and groundnut bud necrosis virus (GBNV) in groundnut field samples collected during the *Kharif* season, 2014 at Kadiri, Ananthapuramu, Andhra Pradesh, India by DAC-ELISA

Symptoms	No. of samples positive/No. tested	DAC-ELISA (OD)[1]	
		TSV	GBNV
Necrotic spots on leaves	12/12	1.7–2.61	0.02–0.09
Top bud necrosis	10/10	1.5–1.7	0.03–0.04
Axillary shoot proliferation with chlorotic small leaflets	11/11	0.8–1.2	0.05–0.08
Groundnut pod samples with necrotic lesions on shell	15/15	1.2–1.4	0.04–0.06
Positive control	–	2.80	2.60
Negative control	–	0.09	0.08

[1]OD values were recorded 1 h after adding substrate p nitrophenyl phosphate

Table 4 Details of coat protein gene sequence of tobacco streak virus isolates reported from India and Japan

Sl. no	Host	location	Isolate code	Accession number	Sl. no	Host	Location	Isolate code	Accession number
1.	Niger seed	Bidar	TSV-NG Bid	DQ864458	27.	Sunflower	Bidar	TSV-SF Bid	AY606067
2.	Sunflower	Hyderabad	TSV-SF hyd	AY061930	28.	Sunflower	Kurnool	TSV-SF Kur	AY501479
3.	Sunflower	Bangalore	TSV-SF Ban	AY061929	29.	Okra	Gulbarga	TSV-OK Gul	AY510130
4.	Sunflower	Aurgabad	TSV-SF Aur	AY061928	30.	Groundnut	Raichur	TSV-GN Rai	AY510128
5.	Sunflower	Bijapur	TSV-SF Bij	DQ864460	31.	Sunflower	Davenagree	TSV-SF Dav	AY510125
6.	Soybean	Osmanabad	TSV-SB Osma	DQ864457	32.	Cotton	Parbhani	TSV-CO Par	AY505082
7.	Okra	Coimbatore	TSV-OK Coim	DQ864456	33.	Cotton	Jalna	TSV-CO Jal	AY505081
8.	Okra	Raichur	TSV-OK Rai	DQ864455	34.	Sunflower	Pune	TSV-SF Pun	AY505077
9.	Groundnut	Kurnool	TSV-GN Kur	DQ864453	35.	Sunflower	Parbhani	TSV-SF Par	AY505076
10.	Groundnut	Anantapur	TSV-GN Ana	DQ864452	36.	Sunflower	Osmanabad	TSV-SF Osma	AY505075
11.	Sunflower	Gulburga	TSV-SF Gul	DQ864450	37.	Sunflower	Buldana	TSV-SF Bul	AY505074
12.	Sunflower	Sholapur	TSV-SF Sho	DQ864449	38.	Sunflower	Jalna	TSV-SF Jal	AY505073
13.	Sunflower	Beed	TSV-SF Bee	DQ864448	39.	Sunflower	Tirupati	TSV-SF Tiru	GU355899
14.	Sunflower	Latur	TSV-SF Lat	DQ864446	40.	Cowpea	Coimbatore	TSV-CP Coim	DQ058079
15.	Jasmine	Kadapa	TSV-JM Kap	KC996726	41.	Urd bean	Coimbatore	TSV-UB Coim	DQ225172
16.	Jasmine	Kurnool	TSV-JM Kur	KC996725	42.	Chilli	Coimbatore	TSV-CH Coim	AY590139
17.	Sunflower	Coimbatore	TSV-GN Coim	GQ167767	43.	Groundnut	Kadiri	TSV-GN Kad	FJ355949
18.	Cluster bean	Tirupati	TSV-CB Tiru	JQ269831	44.	Parthenium	Jalna	TSV-PN Jal	AY940157
19.	Onion	Kadapa	TSV-ON Kap	JQ269830	45.	Cotton	Raichur	TSV-CO Rai	AY940154
20.	Wild marigold	Tirupati	TSV-WM Tiru	FJ447358	46.	Marigold	Anantapur	TSV-MG Ana	AY510129
21.	Groundnut	Bangalore	TSV-GN Ban	HM622157	47.	Onion	Coimbatore	TSV-ON Coim	JX294487
22.	Onion	Kurnool	TSV-ON Kur	HM131490	48.	Sunflower	Hindupur	TSV-SF Hind	KC628720
23.	Okra	Dawalwadi	TSV-OK Dawa	AY501481	49.	Soybean	Coimbatore	TSV-SB Coim	DQ518916
24.	Sunflower	Haveri	TSV-SF Hav	AY606072	50.	Sunflower	Coimbatore	TSV-SF Coim	EU368963
25.	Sunflower	Bagalkot	TSV-SF Bag	AY606071	51.	Dahlia	Japan	TSV-DL Jap	LC030108
26.	Sunflower	Raichur	TSV-SF Rai	AY606069					

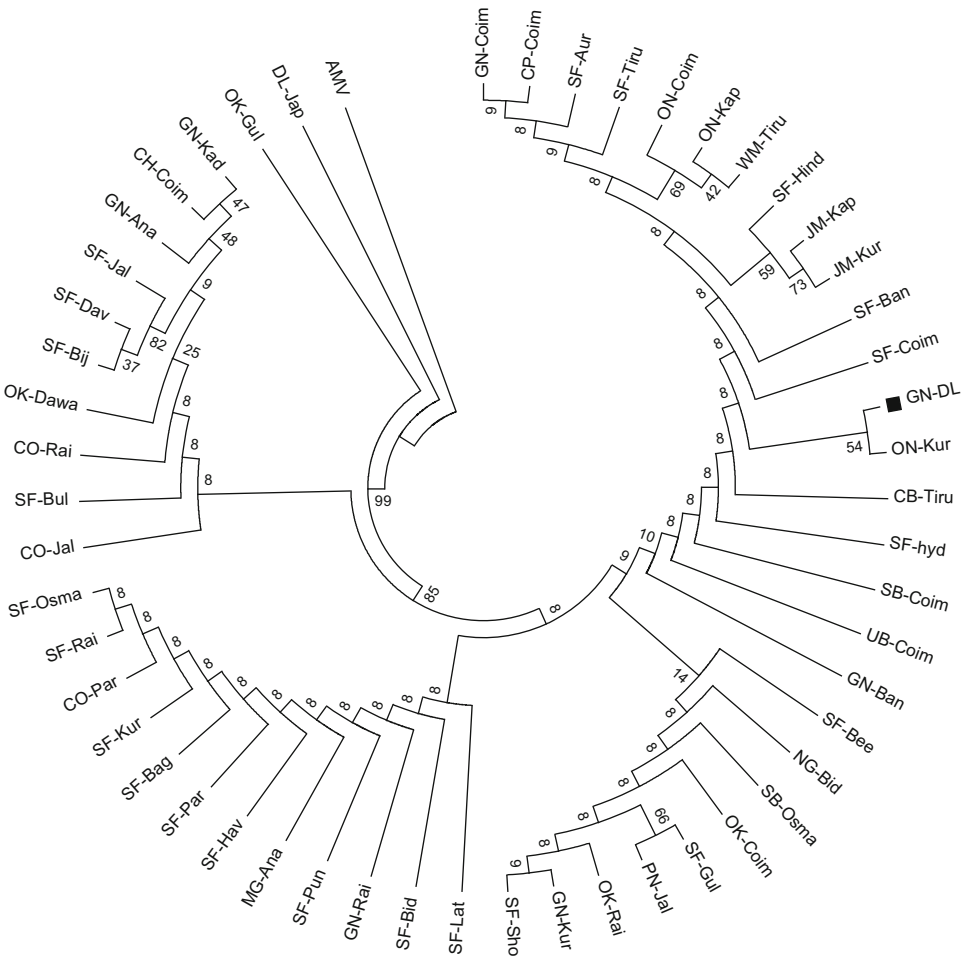

Fig. 2 Phylogenetic analysis of the TSV-GN Kad isolate with the other 51 TSV isolates reported worldwide based on the nucleotide sequence (717 nt) of coat protein gene. Phylogenetic tree was inferred using maximum parsimony method conducted in the MEGA 6 software

phylogenetic divergence. The TSV-GN-Kad isolate shared more similarity with TSV isolates originating from Onion from Kurnool (TSV-ON-Kur), wheras isolates from groundnut from Anantapur and kadiri formed a separate sub-clade (Fig. 2).

Host reactions

The host reactions of TSV-GN-Kad isolate was analysed following sap inoculation in four cultivars (cv) of cowpea (*V. unguiculata*), French bean (*Phaseolus vulgaris*) (Fig. 3), *Nicotiana benthamiana* and sunflower (*Helianthus annuus*) (Table 5). The cowpea cultivars C-152 and Pusa Komal produced numerous small necrotic rings on the inoculated leaves at 3–4 dpi (Fig. 3a, b). Pusa Sukomal and Krishi Kanchan produced chlorotic necrotic symptoms (Fig. 3c, d). Vein necrosis was observed in case of French bean followed by whole leaf necrosis (Fig. 3e). In sunflower, systemic mottle mosaic and necrotic symptoms were observed. Systemic mild mottling was observed in case of only *N. benthamiana* (Fig. 3f). In RT-PCR, the virus was detected only in the inoculated leaves of cowpea and French bean but not in newly emerged leaf indicating no systemic infection. In case of *N. benthamiana* the virus was detected in both inoculated and systemic leaves (Fig. 3h).

Genomic properties of TSV-GN isolate

The three each overlapping fragments were amplified for RNA 1 (290, 1852 and 1802 bp) and RNA 2 (270, 1652 and 1523 bp) and two overlapping fragments for RNA 3 (1200 and 1200 bp) including the terminal regions (Supplementary Table 2). The whole genome sequence comprising of the three segments of RNA (RNA 1, RNA 2 and RNA 3) of the TSV-GN-Kad isolate was submitted to the NCBI database (Table 1).

RNA 1

Genomic sequences showed that the total length of RNA 1 of TSV-GN isolate was 3523 nt long and was similar in length to the sequence of known isolates of TSV. RNA 1 contained one long open reading frame (ORF) starting at

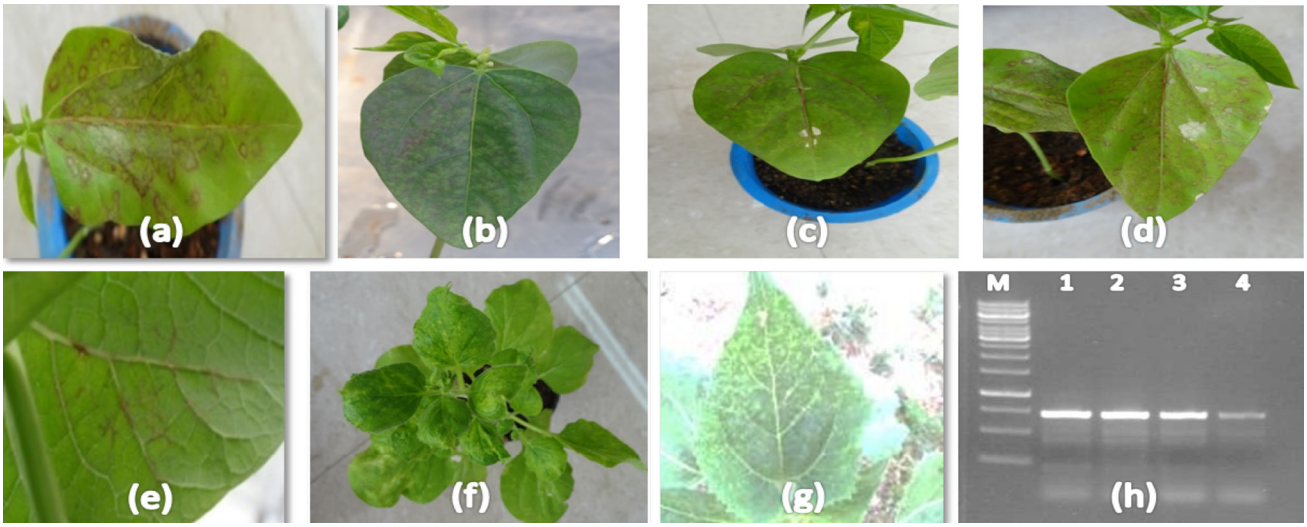

Fig. 3 Symptoms of tobacco streak virus (TSV-GN) isolate in **a** *V. unguiculata* cv. C-152, **b** *V. unguiculata* Pusa Komal, **c** *V. unguiculata* Pusa sukomal, **d** *V. unguiculata* Krishi Kanchan, **e** *P. vulgaris*, **f** *N. benthamiana*, **g** *H. annuus* and **h** RT-PCR of the inoculated plants (lane 1-Cowpea; lane 2-French bean, lane 3-Tobacco and lane 4-Sunflower sample

Table 5 Reaction of various host species to TSV-GN isolate following sap inoculation under the greenhouse conditions

Family	Species	TSV-GN	
		Symptoms[a] Ino/sys leaf	PCR[b] results Ino/sys leaf
Fabaceae	*V. unguiculata c-152*	NLL/ns	±
	V. unguiculata Pusa Komal	NLL/ns	±
	V. unguiculata Pusa sukomal	CL/ns	±
	V. unguiculata KrishiKanchan	CL/ns	±
	P. vulgaris	VN/ns	±
Solanaceae	*N. benthamiana*	M/M	+/+
Malvaceae	*H. annuus*	NL, VN/ns	+/+

[a]Symptoms symbols on inoculated (Ino) and upper non inoculated (sys) leaves: *NLL* necrotic local lesions, *CL* chlorotic lesions, *VN* veinal necrosis, *M* mosaic, *NL* necrotic lesion, *ns* no symptoms

[b]Local and systemic infections were confirmed by RT-PCR. +positive by RT-PCR, −negative by RT-PCR

position 38 nt and extending until 3316 nt encoding the 1a-protein (1092 aa). The 5′ (37 nt) and 3′ (207 nt) untranslated regions (UTR) were also similar in size among the other isolates of TSV reported from India (Supplementary Table 1). A comparison of the RNA 1 nucleotide and predicted amino acid sequences showed a high degree of identity between TSV-GN-Kad isolate and the other Indian isolates from Okra and pumpkin and also from squash from USA (Table 5). TSV-GN-Kad isolate showed 92–93% nt sequence and 97–98% aa identity with pumpkin-India and Okra-India isolates. Interestingly, TSV-GN-Kad 5′UTR of RNA 1showed 100% identity with that of TSV isolate from squash-USA, sunflower and aster isolate from Australia.

RNA 2

The total length of RNA 2 was 2903 nt containing two ORFs, ORF 2a and ORF 2b. The UTR at 5′ and 3′ are 41 and 141 nt long, respectively (Supplementary Table 1), which shared 93 and 100% nucleotide identity with okra and pumpkin isolates from India, respectively. The 5′UTR of RNA 2 of TSV-GN-Kad shared only 73% similarity with the sunflower isolate from Australia. The ORF 2a and 2b were overlapping (ORF 2a from 42 to 2420 nt and ORF 2b from 2164 to 2763 nt) encoding a replicase protein and a suppressor protein of 799 and 199 aa, respectively. A comparison of the RNA 2 sequence showed that TSV-GN-Kad had a higher sequence identity with the other Indian isolate. However, the total length of the two overlapping

ORFs was same for all TSV isolates from India and slightly shorter from the isolates reported from the other countries. The lowest nucleotide and amino acid identity was shared with the sunflower isolate from Australia. A shift in nucleotide position of ORF 2a and ORF 2b is evident in all Indian isolates from the other isolates reported worldwide (Supplementary Table 1).

RNA 3

The RNA 3 segment of TSV-GN-Kad was 2232 nt long containing two ORFs, ORF 3a and ORF 3b. The ORF 3a started with ATG at position 212 and terminated with TGA at position 1084 nt, encoding a movement protein (MP). The ORF 3b (1208–1924 nt) encoded the coat protein (CP) of 26.5 kDa. The untranslated regions at 5′ and 3′ were 211 nt and 288 nt long, respectively.

The RNA 3 ORFs were compared with the other seven isolates reported from the various parts of the world. The ORF 3a and 3b showed > 99.5% similarity with tobacco and soybean isolates of TSV from USA (Supplementary Table 2). Nucleotide and amino acid sequence identity was compared with a total 16 isolates from all over the world. The analysis of 5′ and 3′ UTR region revealed that TSV-GN-Kad isolate shared 100% nt identity with the okra and pumpkin isolates from India, whereas with tobacco and soybean isolate, it shared only 91 and 78% for 5′ and 3′ UTR respectively. Surprisingly, the 3a and 3b ORFs shared 100% sequence identity with the soybean and tobacco isolates from USA. The Indian isolates however shared 87–90% nucleotide and 89–93% aa identity in the 3a and 3b ORFs, respectively. The lowest identity was found with the sunflower isolate from Australia (Data not shown).

Phylogenetic relationships

A total of seven complete genome sequence of TSV isolates were found in the database reported from India, USA and Australia. To determine the phylogenetic relationships of the present isolate, the maximum parsimony based phylogenetic trees were constructed for RNA 1, RNA 2 and RNA 3 (Fig. 4). The phylogenetic analysis of RNA 1 segments demonstrated that TSV-GN-Kad isolate was closely related to the Indian isolate reported from okra and pumpkin (Fig. 4a). The same relations were obtained in case of RNA 2 and RNA 3 nt sequences (Fig. 4b, c). TSV isolates from USA from tobacco and soybean formed an entirely new sub-clade and apart from that TSV sunflower isolate from Australia formed a different clade from others. The similar relationship was evident with the RNA 2 and RNA 3 phylogenetic analysis. Phylogenetic trees obtained based on amino acid sequences of the ORFs of RNA 1, RNA 2 and RNA 3 of TSV isolates showed the same groupings as obtained on the basis of nucleotide sequences (data not shown).

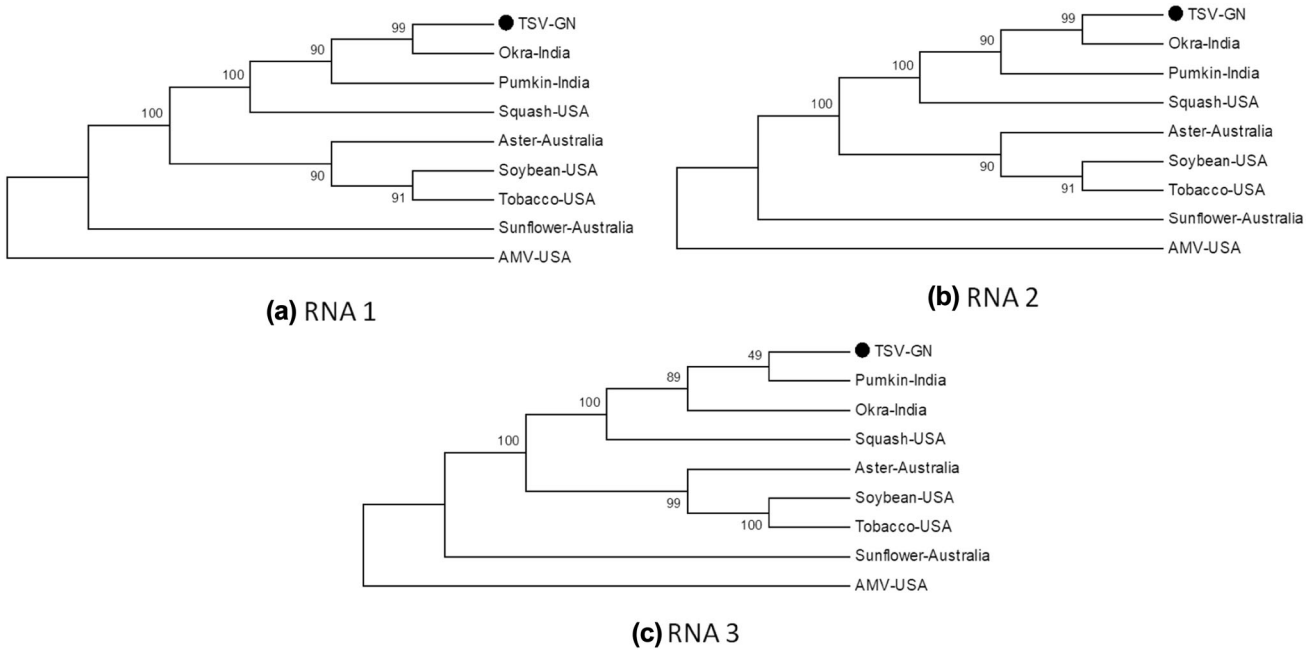

Fig. 4 Maximum parsimonius tree of the complete nucleotide sequences of RNA 1 (**a**), RNA 2 (**b**) and RNA 3 (**c**) of TSV-GN isolate. Number on the node represents the percent bootstrap values. Values lower than 50% have not been shown. AMV obtained from GenBank has been used as the out-group

Discussion

TSV is known to occur in as many as 26 countries. In India, TSV was first recorded in necrosis disease of sunflower and SND of groundnut in 1999–2000 from Andhra Pradesh [10, 12]. Subsequently, the virus was identified infecting several crops and weeds in many southern states of India [8]. SND has emerged as a serious problem in the groundnut production area in southern India. Ananthapuramu district (Andhra Pradesh) is the largest producer of groundnut, where nearly 7 lakh ha is under the crop during the rainy season. High incidence up to 30.0–40.0% has been recorded in certain pockets Ananthapuramu district depending on the stage of infection and dry spell period. The association of TSV with the SND of groundnut was confirmed based on the partial sequence of CP and MP genes [12]. The present study for the first time provides comprehensive characterization based on the biology and complete genome sequence of TSV isolated from a field in Kadiri, the hot-spot of SND.

TSV symptoms in groundnut are very similar to that caused by GBNV. Therefore, the estimation of prevalence of TSV in groundnut based on the visual observation is erroneous. ELISA based on the PAbs, however provides correct indication of the association of TSV or GBNV with the necrosis disease of groundnut. In the present study, chlorosis, necrosis and death of plants were observed in the field and ELISA testing of such plants showed presence of only TSV and not GBNV. The necrotic spots were also observed on the shell of the pods from the infected plants, which previously also was documented as a differentiating symptom of TSV from GBNV [12]. The present isolate of TSV was established from a single lesion and the comparison of the host reactions showed that the virus isolate caused local necrosis on French bean and several cultivars of cowpea. Biologically, the present isolate was similar to the previous one described from groundnut, which was classified as pathotype-1 [12].

Although, TSV is widely prevalent in many countries and economically an important virus, the complete genome sequence is available only for seven isolates originating from Australia, India and USA. In India, TSV has been identified mostly based on CP gene from a variety of plant species. The present and the previous studies showed that the CP gene of TSV occurring in India is highly conserved [1–4, 6, 7, 9, 11, 14]. The complete genome sequence was determined only for two isolates from okra and pumpkin from Bangalore, India. This study for the first time determined the complete genome of TSV associated with the SND of groundnut. The genome organization of the present isolate was same as the other isolates. The present and the other two Indian isolates had identical length of RNA 1 (3523 nt) and RNA 2 (2903 nt), whereas the isolates reported from USA and Australia had variable length of RNA 1 (3476–3525 nt) and RNA 2 (2898–2979 nt). The RNA 3 of the present isolate (2208 nt) was shorter by three nucleotides from the two other Indian isolates (2211 nt). The length of RNA 3 of the USA and Australian isolates was more variable (2056–2216 nt). The complete genome of the present isolate comprising of three RNAs contained 8639 nt, which shared very close sequence identity (97–99%) with that of two Indian isolates and a squash isolates from USA, whereas the other USA and Australian isolates shared significantly diverse sequence identity (77–87%). The comparison of various the genes and UTRs of all the three segments showed that the present isolates were very similar to the Indian isolate but varied considerable from the tobacco isolate from USA and sunflower isolates from Australia. TSV in India isolated from okra, pumpkin and groundnut (present one) shared a close phylogenetic relationship with the squash isolate from USA compared to the isolates obtained from aster, tobacco, soybean and sunflower in USA and Australia. The present study showed that the genome sequence of TSV infecting groundnut, okra and pumpkin in India is highly conserved.

Acknowledgements Authors are thankful to the Department of Science and Technology, Government of India for supporting the Indo-Japan Joint Research Project (DST/INT/JSPS/P-222/2016).

References

1. Arunkumar N, Lakshmi NM, Usha BZ, Ravi KS. Natural occurrence and distribution of *Tobacco streak virus* in south India. Indian J Plant Prot. 2006;34:54–8.
2. Bhat AI, Jain RK, Chaudhary V, Krishna Reddy M, Ramaiah M, Chattannavar SN, Varma A. Sequence conservation in the coat protein gene of *Tobacco streak virus* isolates causing necrosis disease in cotton, mungbean, sunflower and sunn-hemp in India. Indian J Biotechnol. 2002;1:350–6.
3. Bhat AI, Jain RK, Kumar A, RamiahM VA. Serological and coat protein sequence studies suggest that necrosis disease on sunflower in India is caused by strain of *Tobacco streak ilarvirus*. Arch Virol. 2002;147:651–8.
4. Chander Rao S, Prasada Rao RDVJ, Manoj Kumar V, Divya S, Raman M. First report of *Tobacco streak virus* infecting safflower (*Carthamustinctorius*) in Maharashtra, India. Plant Dis. 2003;87:1396.
5. Clark MF, Bar-Joseph M. Enzyme immunosorbent assays in plant virology. Methods Virol. 1984;2:51–85.
6. Jain RK, Bag S, Awasthi LP. First report of natural infection of *Capsicum annuum* by *Tobacco streak virus* in India. Plant Pathol. 2005;54:257.
7. Krishnareddy M, Salil J, Samuel DK. Outbreak of *Tobacco streak virus* causing necrosis of cucumber and gherkin in India. Plant Dis. 2003;87:1264.
8. Krishnareddy M, Vemana K, Kumar S. In: Mandal B, et al., editors. A century of plant virology in India. Singapore: Springer; 2017. p. 271–84.

9. Ladhalakshmi D, Ramiah M, Ganapathy T, Krishna Reddy M, Khabbaz SE. First report of the natural occurrence of *Tobacco streak virus* on blackgram (*Vigna mungo*). Plant Pathol. 2006;55:569.
10. Prasad Rao RDVJ, Reddy AS, Chander Rao S, Varaprasad KS, Thirumala DK. *Tobacco streak* Ilarvirusas causal agent of sunflower necrosis disease in India. J Oilseeds Res. 2000;17:400–1.
11. Ravi KS, Buttgereitt A, Kitkaru AS, Deshmukh S, Lesemann DE. Sunflower necrosis disease from India is caused by an *Ilarvirus*related to *Tobacco streak virus*. Plant Pathol. 2001;50:800.
12. Reddy AS, Prasada Rao RDVJ, Thirumala-Devi K, Reddy SV, Mayo MA, Roberts I, Satyanarayana T, Subramaniam K, Reddy DVR. Occurrence of Tobacco streak virus on peanut (*Arachis hypogaea*) in India. Plant Dis. 2002;86:173–8.
13. Vemana K, Jain RK. New experimental hosts of tobacco streak virus and absence of true seed transmission in leguminous hosts. Indian J Virol. 2010;21:117–27.
14. Vinodkumar S, Nakkeeran S, Malathi VG, Karthikeyan G, Amala Balu P, Mohankumar S, Renukadevi P. Tobacco streak virus: an emerging threat to cotton cultivation in India. Phytoparasitica. 2017;45:729–43.