Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 113 (2005) 132–142
Virus Research
www.elsevier.com/locate/virusres

Assessment of sequence diversity in the 5′-terminal region of *Citrus tristeza virus* from India

Avijit Roy [a], K.L. Manjunath [b], R.H. Brlansky [a,*]

[a] *University of Florida, IFAS, Plant Pathology Department, Citrus Research and Education Center, 700, Experiment Station Road, Lake Alfred, FL 33850-2299, USA*
[b] *University of Florida, IFAS, Plant Pathology Department, 1453 Fifield Hall, P.O. Box 110680, Gainesville, FL 32611-0680, USA*

Received 22 January 2005; received in revised form 22 April 2005; accepted 23 April 2005
Available online 15 June 2005

Abstract

Twenty-one *Citrus tristeza virus* (CTV) isolates from India were characterized, using genotype-specific multiple molecular markers (MMM) from the 54′-terminal region and two other overlapping primer pairs (CN487/489 and CN488/491) from ORF1a (697–1484 nucleotides (nt)). The 5′-terminal genotype-specific primer pairs amplified about 500 bases from the 5′-end of the CTV genomic RNA (gRNA). With the three different MMM, the VT genotype-specific primers amplified 19 Indian CTV isolates. The T30-specific primers amplified five isolates, and the T36 primer amplified only one isolate T36. All isolates were amplified with CN488/491 primers; however, only 20 isolates were amplified with CN487/489 pair. A phylogenetic tree, derived from the sequences of the different MMM primer-amplified products, placed all the isolates into four distinct genogroups. Three of these four groups were typified by the reference isolates T30, T36, and VT. The fourth group, represented by the isolate BAN-2, was considered as a new genogroup. A phylogenetic tree based on sequences of the CN487/491 amplified products and other published sequences placed all of the isolates in eight genogroups. Phylogenetic correlation over the three different regions sequences of these CTV isolates showed more sequence variability between 1082 and 1484 nt than between 1 and 500 or 697–1105 nt of the CTV gRNA. Based on three different 5′ regions sequences and phylogenetic analysis, it is hypothesized that isolates BAN-1, BAN-2, and B165 are three naturally occurring variants that add to the complexity of the CTV populations in India.

Keywords: *Citrus tristeza virus*; Genogrouping; Multiple molecular markers, Phylogenetic analysis; Recombination

1. Introduction

Citrus tristeza virus (CTV), the largest known (19.3 kb) monopartite, positive-stranded RNA virus of plants, causes one of the most economically important diseases of citrus (Bar-Joseph et al., 1989). The virus, a member of the genus *Closterovirus* within the family Closteroviridae, is phloem-limited, and is transmitted by infected buds and by various aphid species. *Toxoptera citricida* (Kirkaldy) and *Aphis gossypii* (Glover) are the most efficient aphid vectors and transmit CTV in a semipersistent manner (Bar-Joseph et al., 1989). CTV virions are filamentous flexuous particles, 2000 nm × 11 nm in size, with two coat proteins (CP, CPm) covering 95 and 5% of the particle length, respectively (Febres et al., 1996). The CTV genome is organized into 12 open reading frames (ORFs), encoding at least 19 protein products (Karasev, 2000; Karasev et al., 1995; Mawassi et al., 1996; Pappu et al., 1994) and two untranslated regions (UTRs) of about 107 and 273 nucleotides (nt) at the 5′ and 3′ termini, respectively (Karasev et al., 1995; Pappu et al., 1994). The 5′ proximal ORF 1a encodes a 349 kDa polyprotein that includes two cysteine papain proteinase-like (P-Pro) domains as well as methyltransferase-like (MT) and helicase-like (HEL) domains. ORF1b encodes an RNA-dependent RNA polymerase (RdRp)-like domain and is thought to allow the continued translation of the polyprotein by a + 1 frameshift. ORFs 2–11, located in the 3′ moiety of the genomic-RNA (gRNA), are translated via subgenomic RNAs (sgRNAs), which are 3′ co-terminal (Hilf

* Corresponding author. Tel.: +1 863 956 1151; fax: +1 863 956 4631.
E-mail address: rhby@crec.ifas.ufl.edu (R.H. Brlansky).

0168-1702/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2005.04.023

et al., 1995). Infected plants may contain defective RNA (dRNA) molecules, which may have both gRNA termini but lack variable portions of the central region (Ayllon et al., 1999; Mawassi et al., 1995).

Complete genomic sequences from CTV isolates T36 and T30 from Florida (Albiach-Marti et al., 2000; Karasev et al., 1995), VT from Israel (Mawassi et al., 1996), T385 from Spain (Vives et al., 1999), SY568 from California (Yang et al., 1999), NUagA from Japan (Suastika et al., 2001), and Qaha from Egypt (GenBank accession no. AY340974) have been determined. The gRNAs of these seven isolates vary from 19,226 to 19,302 nt in size. An unexpected sequence relationship between T36 and VT isolates of tristeza virus was reported (Mawassi et al., 1995, 1996). The 3′ half of the gRNA is relatively conserved (approximately 90% identity); however, less than 70% identities were found in the 5′-end. Hilf et al. (1999) classified various Florida CTV isolates in two sequence groups (VT and T36 types) based on hybridization with probes for the 5′- and 3′-terminal regions of T3, T30, and T36 CTV isolates. The authors had also classified four genotypes using three sets of molecular markers located in the 5′ half of the genome (Hilf and Garnsey, 2000). Further analysis suggested the existence of additional genotypes, which do not react with any of the multiple molecular markers (MMM) described by Hilf and Garnsey (2000). The MMM were successfully used to analyze subisolates transmitted by aphids from different CTV isolates (Brlansky et al., 2003). Recent work using MMM primer pairs in the regions of POL and k17 of CTV genome of Indian CTV isolates (Roy and Brlansky, 2004) showed that amplifications were not always genotype-specific. For example, the T3 genotype-specific primer amplified product of some isolates had more sequence similarity with either the T30 or SY568 genotypes rather than with the T3 genotype (Roy and Brlansky, 2004).

CTV belongs to the genus *Closterovirus* in the family Closteroviridae. Speciation in the genus *Closterovirus* is based on particle size, sequence and size of the coat protein gene, serological, and vector-specificities (Martelli et al., 2002). Even though CTV isolates exhibit a high degree of sequence diversity in the 5′ half of their genomes, they are placed under one genus based on the above criteria. Citrus plants can be infected with multiple genotypes of CTV, and proper identification is important for disease management. Genome fragments in different parts of the CTV genome have been utilized to classify CTV isolates based on PCR amplifications, sequencing, SSCP, hybridization, and other techniques.

Comparison of isolates may be accomplished by complete sequencing of the 19.3 kb CTV genome but this is time consuming and difficult. Lopez et al. (1998) developed another approach based on the analysis of total 1250 nt from the two terminal regions. Sequence comparison of multiple cDNA clones of both gRNA terminal regions from several CTV isolates showed that the 3′ untranslated region was highly conserved (97% nt) identity, whereas 5′ UTRs were less conserved (44–88%). The variability observed in the 5′ UTR and ORF1a separated the sequences into three groups, which represented the sequences of T36, VT, and clone 8 of isolate T317. Some isolates yielded clones that were placed into multiple groups, illustrating the presence of different sequence variants in infected plants (Lopez et al., 1998). Extended analysis of the variability of the 284 nt from 5′-terminal region of the CTV gRNA showed four conserved regions. These regions were used to develop four sets of primers to determine the group-specificity of tristeza virus (Ayllon et al., 2001). These results confirmed the presence of three types of CTV populations from at least 9 of the 12 countries from which samples were taken.

The MMM approach (Hilf and Garnsey, 2000) is one method currently used for CTV isolate classification. This method utilizes genotype-specific primers from three different regions of the genome. CTV isolates are classified into four groups (T3, T30, T36, and VT) based on the presence or absence of RT-PCR amplifications. The purpose of the present study was to find the sequence variability of 21 CTV isolates from India in the 5′-terminal region using universal and the genotype-specific MMM primer pairs (Hilf et al., 1999). The relationship of the Indian CTV isolates to the different CTV genotypes (Ayllon et al., 2001; Lopez et al., 1998) was studied by the sequencing and phylogenetic analyses.

2. Materials and methods

2.1. Virus isolates

The 21 CTV isolates from India used in this study (Table 1), from three different citrus growing areas of India, have been previously described (Roy and Brlansky, 2004). All of the isolates, except for CTV-B, CTV-D, CTV N, and CTV-P, are maintained at the Exotic Citrus Pathogen Collection at Beltsville, MD, in Mexican lime [*Citrus aurantifolia* (Christm.) Swing.] and in sweet orange [*C. sinensis* (L.) Osb.] plants. The work on isolates CTV-B, D, N, and P was done at Indian Agricultural Research Institute, New Delhi, India.

2.2. Nucleic acid extraction and reverse transcription-polymerase chain reaction (RT-PCR)

Total RNA from plants infected with each of the 21 Indian CTV isolates was extracted using RNeasy Plant Mini Kit (QIAGEN, Valencia, CA) following the manufacturer's protocol. The MMM primer pairs used to amplify the 5′-terminal regions were derived from the analogous sites within the genomes of T30 (nt 6–26 and 580–600), T36 (nt 1–20 and 481–500), and VT (nt 1–22 and 471–492) isolates (Hilf et al., 1999). Other regions of ORF1a were amplified with universal primers, CN487/489 (^{697}5′-GCG TTG GAT GAT ATC CTT CGC TGG-3′$^{720/1082}$5′-AAT TRT TCC GCS CAG GAC GGA ACA-3′^{1105}) and CN488/491 (^{1082}5′-TGT TCC GTC CTG SGC GGA AYA ATT-3′$^{1105/1461}$5′-GTG TAR GTC CCR CGC ATM GGA ACC-3′^{1484}) primer pairs. Nucleotide

Table 1
Genogrouping of *Citrus tristeza virus* (CTV) isolates based on RT-PCR amplification by MMM, phylogenetic analysis of sequences of two regions of the CTV genome (from 1 to 500 nt and 697 to 1105 nt)

Isolates[a]	MMM-based genogroups[b]	Genogrouoping[c] based on sequence from nt 1–500		Genogrouping based on sequence from nt 697–1484	
		(Dendrogram)	Accession no.[d]	(Dendrogram)	Accession no.
BAN-1	II	II	AY295920	VII	AY572216
BAN-2	II	II	AY295917	IV	AY572217
	III	IV	AY295912		
CTV-D	III	III	AY295905	III	AY572218
CTV-N	II	III	AY295918	III	AY572219
	III	III	AY295914		
CTV-P	II	III	AY295919	III	AY572220
	III	III	AY295913		
B165	III	III	AY295911	VIII	AY572201
B166	III	III	N/S[e]	III	AY572202
B167	III	III	N/S	III	AY572203
B184	III	III	AY295907	III	AY572204
B194	II	II	AY295916	III	AY572205
	III	III	AY295910		
B195	III	III	AY295906	III	AY572206
B219	I	I	AY295915	III	AY572207
	III	III	AY295909		
B220	III	III	N/S	III	AY572208
B221	III	III	N/S	III	AY572209
B222	III	III	N/S	III	AY572210
B224	III	III	N/S	III	AY572211
B225	III	III	N/S	III	AY572212
B225.1	III	III	N/S	III	AY572213
B226	III	III	AY295908	III	AY572214
B227	III	III	N/S	III	AY572215
CTV-B	?[f]	N.A[g]	–	III[h]	AY572221

[a] Isolates used in this study.
[b] Genogrouping based on multiple molecular marker used for RT-PCR amplification.
[c] Genogroup designation based on phylogenetic analysis of nucleotide sequences.
[d] Accession no. indicates the isolate-specific nucleotide sequences submitted in the Gene Bank.
[e] N/S: gene sequences not submitted in GenBank.
[f] '?': group not determined since no amplification was obtained.
[g] N.A: not amplified.
[h] Only partial sequence (697–1105) was obtained using primers CN488 and CN491. Shaded area shows the differences between the genogroups based on MMM and phylogenetic analysis.

numbering is based on the genome of T36 isolate (NC 001661). Reverse transcription-polymerase chain reaction was conducted as described previously (Roy and Brlansky, 2004; Roy et al., 2003). The PCR products were resolved on 0.8% agarose gels.

2.3. Nucleotide sequence analysis

The nucleotide sequences of the 5′-terminal region were determined for 21 Indian CTV isolates. Genotype-specific cDNAs from MMM primer pairs and CN487/489, CN488/491, and CN487/491 primer pairs amplified RT-PCR products were eluted using a gel purification kit (QIAGEN, Valencia, CA) and were directly ligated into a pGEM-T Easy Vector (Promega, Madison, WI) according to the manufacturer's protocol. The nucleotide sequence of the inserts was determined at the University of Florida DNA Sequencing Core Laboratory using ABI Prism Big Dye Terminator cycle sequencing protocols (Applied Biosystems, Perkin-Elmer Corp., Foster City, CA). The dye terminator method was used to analyze the fluorescent-labeled extension products in an Applied Biosystems Model 3100 genetic analyzer (Perkin-Elmer Corp.). Four to five clones from the PCR amplified products of each isolate were used to analyze the sequence variation within the isolate.

2.4. Phylogenetic analysis

The nucleotide sequences of the selected clones were used for searching identical sequences in the database with the programs of the GCG package (Devereux et al., 1984) and BLAST. Amino acid (aa) translation was done using the Expasy Translate Tool (http://us.expasy.org/tools/dna.html). Nucleotide and aa sequences were aligned using the program

Table 2
Percent nucleotide (above diagonal) and amino acid (below diagonal) identities between the selected Indian CTV isolates (A–E) and type isolates (F–H) based on 5′-terminal region (nt 1–500) of the genomic RNA

Categorys	Isolate/MMM/genogroup[a]	Genogroup[b]	Percent identity of nucleotide sequence							
			A	B	C	D	E	F	G	H
A	B219/T36/I	I	*	68	72	72	75	94	69	71
B	B194/T30/II	II	67	*	89	87	83	69	98	85
C	CTV-N/T30/II	III	70	90	*	97	84	72	89	94
D	B194/VT/III	III	69	90	97	*	86	72	87	96
E	BAN-2/VT/III	IV	72	83	83	85	*	75	84	83
F	T36/T36/I	I	72	51	53	51	54	*	69	70
G	T30/T30/II	II	68	97	92	93	86	51	*	86
H	VT/VT/III	III	65	86	92	93	80	50	88	*
			Percent identity of amino acid sequence							

The nucleotide identity was based on the 5′ UTR and the partial region of ORF1a. The amino acid identity was based on the partially translated region (108–500 nt) of ORF1a. The asterisks are used for the homologous identities of A with A and B with B which would be 100% and also are used as a diagonal for dividing the table so that two sets of data can be presented in the same table.

[a] The isolate names are followed by the MMM primer used for amplification and grouping based on PCR amplifications.

[b] Grouping is based on analysis of actual sequences. Sequences with over 94% identity are placed in the same genogroup. There was one sequence that belonged to Category A (B219 using T36 MMM), three belonged to category B (B194, BAN-1, and BAN-2 using T30 MMM), two belonged to category C (CTV-P and CTV-N using T30 MMM), one belonged to category E (BAN-2 using VT MMM), and the rest belonged to category D (CTV-D, CTV-N, CTV-P, B165, B166, B167, B184, B194, B195, B219, B220, B221, B222, B224, B225, B226, and B227 using VT MMM). SY568 and NUagA isolates were not included in this table.

Clustal-X (Thompson et al., 1997). Percent identity matrices, and nt and aa sequence alignments for each isolate were generated using GeneDoc (Nicholas and Nicholas, 1997). The phylogenetic relationships of all the Indian isolates with isolate T3 (GenBank accession no. AY986484), and the seven fully sequenced CTV isolates [T30 (GenBank accession no. AF260651), T36 (Accession no. U16304), T385 (Accession no. Y18420), VT (Accession no. U56902), SY568 (Accession no. AF001623), NUagA (Accession no. AB046398), Qaha from Egypt (Accession no. AY340974)], were deduced using the Bootstrap Neighbor-Joining (N-J) methods in the Phylip formatted labels node tree. A 'phb' file was created in the program Clustal-X and the tree was visualized using the program TreeView (Win32) Version 1.6.6. The 5′-terminal sequences of gRNA of the 12 CTV isolates and eight clones under EMBL accession numbers from AJ401190 to AJ401201 (Ayllon et al., 2001) and Y16838 to Y16845 (Lopez et al., 1998) were aligned with all the Indian isolates using Clustal-X (Thompson et al., 1997) and were compared with MMM primer pairs amplified cloned products using GeneDoc (Nicholas and Nicholas, 1997). Dendrograms were displayed using the TreeView. In order to detect putative recombinant isolates, the phylogeny of gene segments was determined using PHYLPRO (Weiller, 1998). The genes were scanned using 40 nucleotide-sliding windows. Each genetic group consensus sequence was used as the reference sequence for the recombinant isolates.

3. Results

3.1. Analysis of CTV isolates using multiple molecular markers at the 5′-end of the genome by sequencing

Three pairs of genotype-specific MMM primer pairs for the 5′-terminal region of CTV described by Hilf et al. (1999) were used for analysis of Indian isolates (Table 1). RT-PCR amplified products were obtained from 20 of the Indian isolates but not from the isolate CTV-B which did not give an amplification product with any of the three primer pairs; 19 isolates were amplified by the VT-specific MMM, 5 by T30 and 1 by T36-specific MMM (Table 1). Two pairs of genotype-specific primers amplified each of the isolates BAN-2, CTV-N, CTV-P, B194, and B219, indicating the presence of mixtures of different CTV genotypes in these plants. Throughout this paper, the amplified products are designated as genogroups I–III which represents T36, T30, and VT isolate-specific genogroups, respectively.

The 5′-terminal sequences of 20 Indian CTV isolates differed from each other by up to 69% (Table 2) and grouped them into four different genotypes (Fig. 1). Most of the amplification products had sequences belonging to their expected genogroups. However, the sequences of BAN-2 amplified by VT MMM, and those of CTV-N and CTV-P amplified by T30 MMM belonged to other groups. The isolates BAN-1, BAN-2, and B194 amplified with the T30 MMM had 97–98% nt identity with the T30 group while the isolates CTV-N and CTV-P amplified by the same T30 MMM had only 89% nt sequence identity with CTV isolates in that group. The two latter sequences were closely related to those in the VT group showing about 94–96% nt identity with the VT and SY568 and NUagA isolates. The nt identity of CTV-N and CTV-P was over 97% with several Indian isolates in the VT group and the isolate SY568 which suggested that these two T30-5′ amplification products are more closely related to the California isolate SY568 than to either the Japanese NUagA (96%) or the Israeli VT (94%) isolates. B219 was the only Indian CTV isolate that produced an amplification product with the T36-5′-terminal region genotype primers and showed 94% nt sequence identity with T36 decline isolate as compared to 69–71% nt identity with the two other groups (Table 2).

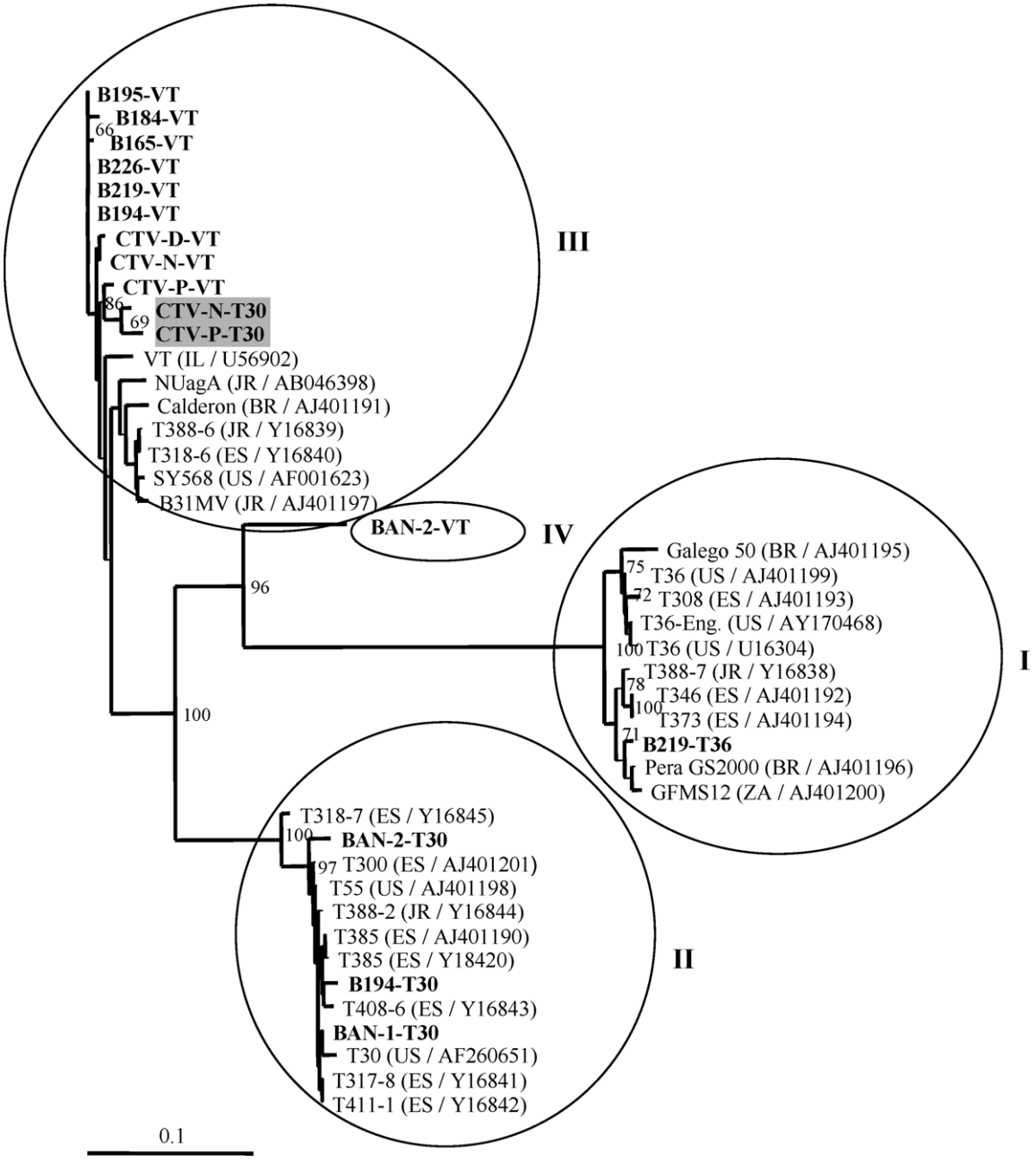

Fig. 1. Phylogenetic dendrograms of the 266 nucleotides sequence of the 5′-terminal region of Indian and other published CTV isolates. Previously published CTV isolates are shown with isolate number followed by geographical origin and accession number. The abbreviations are BR, Brazil; ES, Spain; IL, Israel; JR, Japan; US, United States; ZA, South Africa. The Indian isolates used in this study were amplified by using genotype-specific primers. The names of the Indian isolates and the genotype-specificity of primers used in RT-PCR amplifications are shown in bold letters. Shaded amplified product occupied T30 cluster although it amplified with VT genotype-specific primers NB [VT-5′ amplified sequences from all the Indian isolates were not included in the phylogram because of 99–100% sequence identity among them].

The VT-5′ primer amplified 492 nt from the VT isolate and 500 nt from 19 Indian isolates (Table 1). These amplification products had 94–97% nt sequence identity (except the isolate BAN-2-VT) with VT, SY568 and NUagA isolates (Table 2 and data not shown). The BAN-2 isolate did not show closer identity with any of the three groups. It had only 83–84% nt and 80–82% aa identity with the VT, NUagA, and SY568 isolates and 86% nt, and 84–85% aa identity with other Indian isolates. A sequence homology of more than 92% was used to place the isolates into a genogroup.

3.2. Analysis of CTV isolates based on sequences of nucleotides 697–1484 from ORF1a

The two overlapping primer pairs, CN487/CN489 (697–1105 nt) and CN488/CN491 (1082–1484 nt) amplified

Table 3
Percent nucleotide identities among the Indian CTV isolates and selected published full-length sequences based on sequences from nt 697–1105 (above the diagonal) and nt 1082–1484 (below the diagonal)

Category	Isolates	Genogroup[a] (409nt/403nt)	Percent nucleotide identity of CN487/489 primer pair amplified product									
			A	B	C	D	E	F	G	H	I	J
A	T36	I/I	*	78	73	76	77	89	77	78	76	77
B	T30	II/II	82	*	88	90	91	79	92	90	91	92
C	VT	III/III	80	90	*	93	94	76	88	89	94	95
D	CTV-N	III/III	82	92	94	*	96	78	89	90	96	96
E	B195	III/III	82	91	95	96	*	79	90	91	97	98
F	BAN-2	IV/IV	91	83	81	84	83	*	78	78	78	79
G	T3	V/V	81	89	87	90	90	81	*	91	90	91
H	NUagA	VI/VI	81	90	87	88	88	83	88	*	91	92
I	BAN-1	III/VII	81	90	90	92	92	82	87	92	*	97
J	B165	III/VIII	80	86	83	85	85	81	83	84	84	*
			Percent nucleotide identity of CN488/491 primer pair amplified product									

Indian isolates with over 94% identity at the nucleotide level are represented by one type isolate. The asterisks are used for the homologous identities of A with A and B with B which would be 100% and also are used as a diagonal for dividing the table so that two sets of data can be presented in the same table.

[a] Genogroups based on sequences of two different regions of CTV genome; 697–1105 and 1082–1484 nt. Shaded areas show differences in genotyping based on the sequences of two regions described above. [There were four Indian CTV isolates in category D (BAN-1, CTV-D, CTV-N, and CTV-P), 15 in E (B165, B166, B167, B184, B194, B195, B219, B220, B221, B222, B224, B225, B225.1, B226, and B227), one in F (BAN-2) from CN487/489 genogroup. There were three Indian CTV isolates in category D (CTV-D, CTV-N, and CTV-P), 14 in E (B166, B167, B184, B194, B195, B219, B220, B221, B222, B224, B225, B225.1, B226, and B227), one in I (BAN-1), and one in J (B165) CN488/491 genogroup.]

RT-PCR products from most of the 21 Indian isolates. The only exception was that no product was obtained with the isolate CTV-B using the CN487/CN489 primer pair. Table 3 summarizes the nt identity data for all the Indian isolates from these product sequences. Except for the BAN-2 isolate, all other Indian CTV isolates had a 96–98% nt identity between themselves and were closely related (93–95%) to the VT isolates in the region of nt 697–1105. However, the subsequent overlapping sequences (nt 1082–1484) of the same isolates showed some interesting differences. Most isolates showed similar nt identity with the VT isolate in this region. But the sequences of BAN-1 and B165 which showed only 90% nt and 83% nt identity, respectively, with the VT isolate in the sequence of nt 1082–1484. The sequence of the BAN-2 isolate in this region had a 79–80% identity with the other Indian CTV isolates and 88–89% identity with Qaha and T36 isolate (Table 3).

Since the sequence divergence in the 1082–1484 nt region was much higher than the 697–1105 nt region, further analysis was done by sequencing the 788 nt products (697–1484 nt) amplified by CN487/491 pair for all the isolates (except CTV-B). Complete sequencing of cDNA clones was done to check if the sequences in two regions belonged to same RNA strand or not. Analysis of these sequences showed that the above sequence divergences were actually present in the same RNA strands and did not result from amplification of different virus populations from same plant. Out of 20 isolates, 17 isolates had 96% nt sequence identity with California SY568 isolate and a 95% nt identity with the Israeli VT isolate. The unique BAN-2 isolate shared 90% nt identity with T36. Sequence differences in the last 403 nt in the BAN-1 and B165 isolates lowered its nt identity from 96–98 to 91–94 and 83–86%, respectively, with other Indian isolates.

3.3. Phylogenetic relationships of CTV sequences at the 5′ proximal end

A phylogenetic tree was constructed using the nucleotide sequences of the genotype-specific 5′-terminal region (1–500 nt) from the Indian CTV isolates and other sequences available in the database using Clustal-X and Treeview programs (data not shown). Since a larger number of sequences were available in the database for the 5′-terminal 266 nucleotides, another phylogenetic tree was developed using only this region (Fig. 1). Four main clusters of CTV are identified in both the dendrograms. Most of the amplification products from the MMM genotype-specific primers amplified from only the specific genotypes, as expected. The T30-5′-specific primer produced amplification products from five of the Indian isolates. Three of these five isolates, BAN-1, BAN-2, and B194 formed a cluster along with the mild isolates T30 and T385 (Fig. 1) and belonged to genogroup II. Two other isolates, CTV-N, and CTV-P, formed a different cluster with the Californian SY568, Japanese NUagA, and Israeli VT isolate sequences (Fig. 1, genogroup III). Except for BAN-2 isolate, amplification products from the VT-5′ primer grouped together in the same branch with the VT isolate (Fig. 1, genogroup III). The sequence of isolate BAN-2 was unique and was placed in between the branch of the T36 and T30 isolates (Fig. 1, genogroup IV). The B219 was the only isolate that gave an amplification product with T36-5′ primer and was clustered with the T36 isolate (Fig. 1, genogroup I).

In the previous studies (Ayllon et al., 2001; Lopez et al., 1998) 94 clones of the 5′ proximal end of the CTV gRNA clustered into three groups. We expanded this work here by sequencing 125 clones from 25 amplified products of the 20 Indian CTV isolates. Most of the isolates were assigned

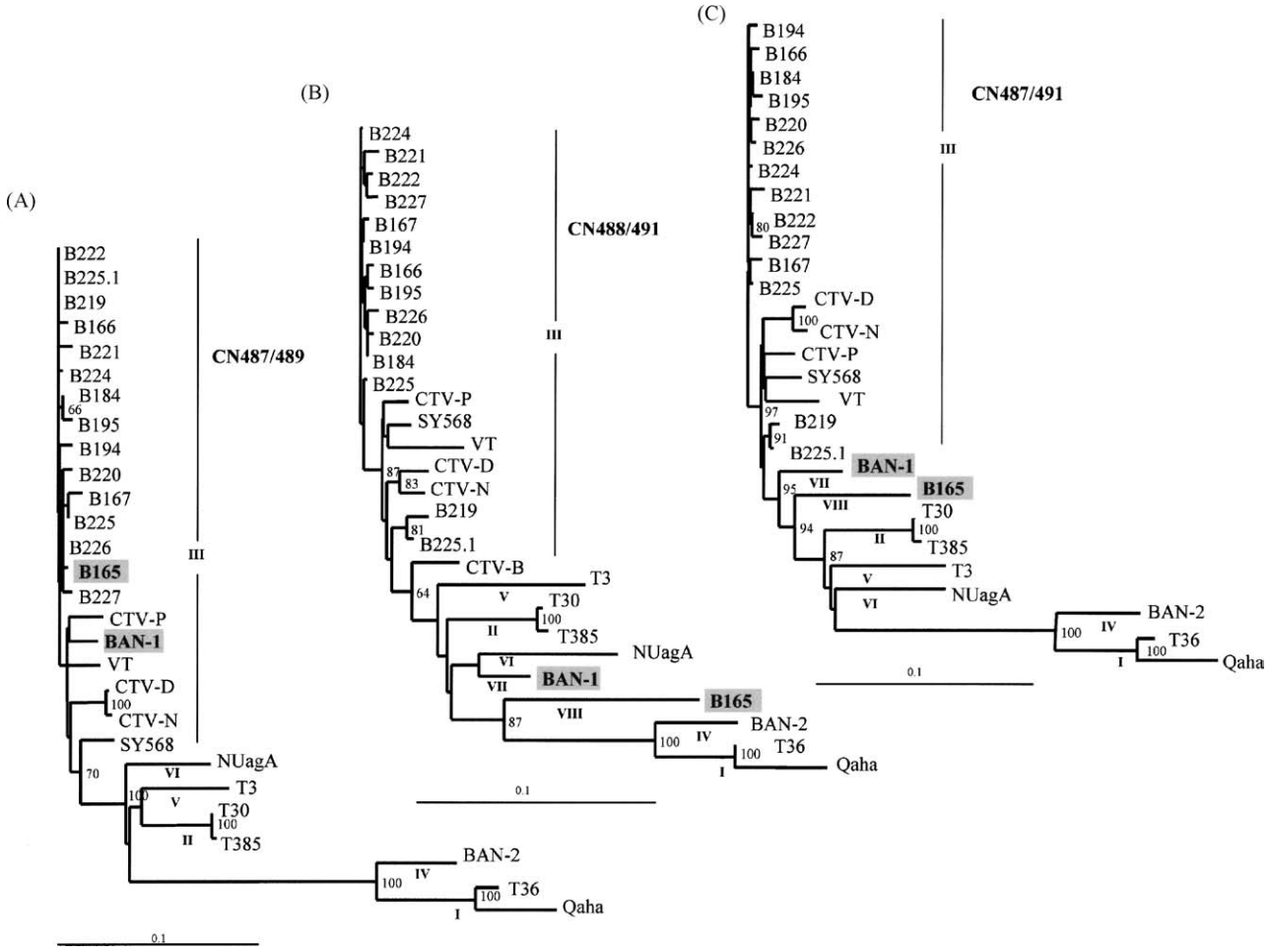

Fig. 2. Phylogenetic dendrograms based on nucleotide sequences of different regions of CTV gRNA from ORF1a: (A) 409 bases, 697–1105 nt; (B) 403 bases, 1082–1484 nt; (C) 788 bases, 697–1484 nt. Sequences with about 95% nucleotide identity are grouped together (groups designated I–VIII).

to one of the three groups, except for the BAN-2 VT-5′ sequence, which was placed in a new fourth group. Maximum variability in the nucleotide sequence of CTV genomes has been reported in the 5′ untranslated region (Ayllon et al., 2001; Lopez et al., 1998). Phylogenetic analysis using 266 nucleotide sequences from all of the experimental isolates (from GenBank data) of Lopez et al. (1998) and Ayllon et al. (2001) and the Indian CTV isolates under study displayed four distinct clusters (Fig. 1). The analysis of either 266 or 500 nt sequences from 5′ proximal end showed the presence of four genogroups in the Indian CTV population instead of three groups as previously reported (Ayllon et al., 2001; Lopez et al., 1998).

3.4. Phylogenetic relationship of Indian CTV isolates based on 788 nucleotides (697–1484 nt) in the 5′ region

A 788-nucleotide region (697–1484 nt) of CTV genomes was analyzed by using two sets of primers amplifying overlapping regions. Only six genogroups were displayed in the phylogram when the sequences of 409 nt (697–1105 nt; with primers CN487/CN489) from the first half of the 788 nt region were compared. But the sequences of amplified products in the second half of the region (nt 1082–1484; primers CN487/491) suggested that all the isolates could be clustered into eight genogroups (Fig. 2). The isolates BAN-1 and B165 represented two new branches in the phylogenetic tree (Fig. 2). Although isolate CTV-B occupied the same genogroup with SY568 and VT isolates, this isolate was not amplified with genotype-specific 5′-terminal MMM primers or with CN487/489 primer pairs, which indicated sequence diversity in the first half of the 788 nt. Differences in genogrouping in the dendrograms based on sequences of these two overlapping regions indicated possible recombinations in some of the isolates.

The computer program PHYLPRO was utilized to graphically display the sequence relationships among the CTV isolates studied. The program determines pairwise distances of all sequences, evaluates which patterns of distances in regions agree, and detects recombinant sequences. Results showed no evidence of recombination when the 5′-end

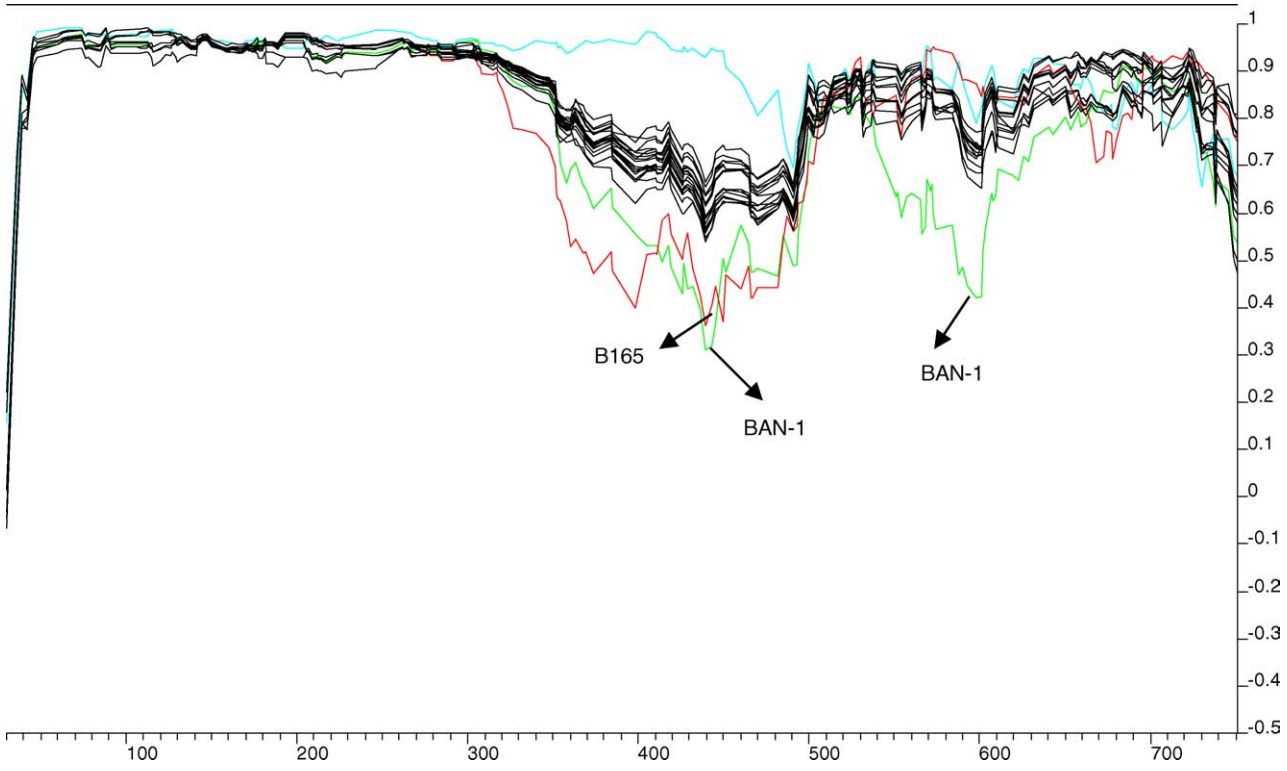

Fig. 3. Phylogenetic correlation profiles over the entire length of aligned 5′-end sequence (697–1484 nt) of 20 Indian CTV isolates. Phylogenetic correlation (*y*-axis) was obtained at each variable site from pairwise distance analysis of all aligned sequences by using the program PHYLPRO, with a fixed window of 40 bp. Numbers under low phylogenetic correlation areas indicated possible recombination signals by arrows. In the graph *x*-axis is represented the value of correlation coefficient [colours black lines represent isolates from genogroup III, aqua for genogroup IV, green for genogroup VII, and red for genogroup VIII].

sequences (697–1484 nt) of the genogroup III were analyzed (data not shown). Similarly, no recombination events were observed when the isolate BAN-2 (genogroup IV) sequences were plotted together with the genogroup III isolates (data not shown). Recombination events were indicated when isolates BAN-1 and B165 were compared either individually or together (Fig. 3) with the isolates of genogroups III and IV. The presence of two sharp downward peaks at 439th and 599th nt corresponds to 0.325 and 0.425 correlation coefficients indicating recombination for the isolate BAN-1. The isolate B165 also showed recombination at the 439th nt position. These data suggest the occurrence of recombination events in this region of the genome among the Indian CTV isolates. When other CTV sequences from the GenBank were added and compared; T3, T36, and Qaha, isolates were identified as recombinants in the same region of the genome along with the isolates BAN-1 and B165 (Fig. 4). The isolates T36 and Qaha showed recombination at the 551st nt position whereas with T3 recombination occurred at 74th and 150th nt positions. The above data probably explains the reasons for differences observed in the phylogenetic association of BAN-1 and B165 with the VT group. This was based on sequences in one region (697–1105 nt; Fig. 2A) and their placement in six separate groups and on sequences of another region (1082–1484 nt; Fig. 2B) with the placement into eight groups.

4. Discussion

In our previous work (Roy and Brlansky, 2004), the genotype profiles of 21 Indian CTV isolates were determined based on the POL and k17 region using the genotype-specific MMM primer pairs (Hilf and Garnsey, 2000; Hilf et al., 1999); 19 of 21 isolates contained the VT genotype either alone or in combination with 2 other genotypes (T3 and T30), the exceptions being isolates BAN-1 and CTV-B (Roy and Brlansky, 2004). In the current study, the amplified products obtained from the 5′-terminal region using the genotype-specific MMM, CN487/489, CN488/491, and CN487/491 primer pairs were sequenced and analyzed. Genotype grouping was accomplished by phylogenetic analysis.

Comparisons of sequences of the 500 nt from the 5′-terminal region of all the Indian CTV isolates, provided additional evidence that more than one genotype may exist within an isolate. Sequences from BAN-2 (genogroups II and IV), B194 (genogroups II and III), and B219 (genogroups I and III) isolates clustered in different groups. In the previous studies, (Ayllon et al., 2001; Lopez et al., 1998) maximum variability was found within the 5′ UTR of all the published CTV genome sequences. Although isolates CTV-N and CTV-P were amplified with the T30-5′ primer, comparative analysis of their sequences did not place them in the T30 isolate-specific genogroup II, but placed them in the

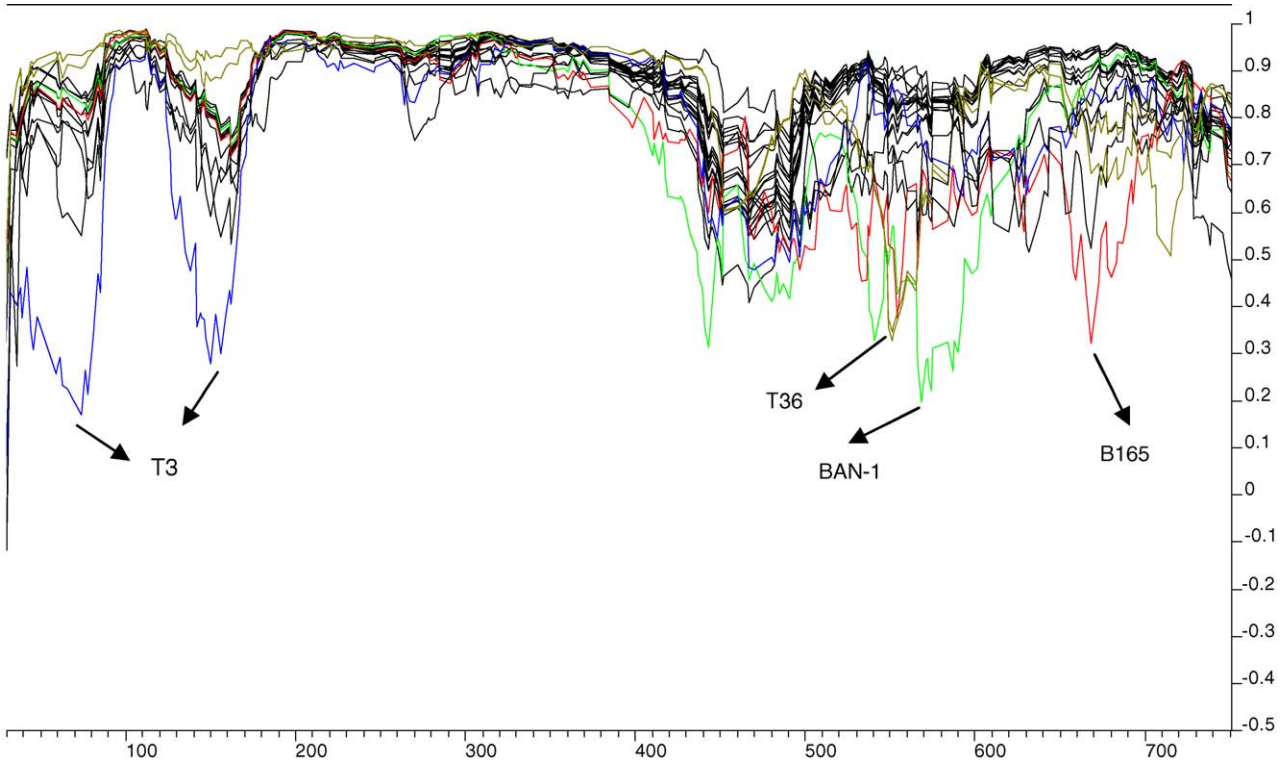

Fig. 4. Phylogenetic correlation profiles over the entire length of aligned 5′-end sequence (697–1484 nt) of all the Indian and seven completely sequenced CTV isolates obtained using the PHYLPRO program with a fixed window of 40 bp. Arrows indicating the recombinant signals for T3, T36, BAN-1, and B165 isolates.

VT isolate-specific genogroup III. Similar amplification of non-targeted regions by MMM primers has been reported previously (Roy and Brlansky, 2004). These results emphasize the limitations of classifying CTV isolates based on PCR amplification without sequencing.

Except for the isolate BAN-2, all of the other isolates amplified with the VT-5′ primer pairs occupied the genogroup III in the phylogenetic tree (Fig. 1). Even though isolate BAN-2 was amplified by the VT-5′ primer, the sequence of the amplified product did not align with VT genogroup or any of the other genogroups, and was placed in a separate genogroup (IV). This genogroup was positioned between the T36 and T30 genogroups. Sequencing of the CTV gRNA of other parts of 5′ region of other isolates would probably increase the complexity of the current grouping and may lead to identification of new genogroups.

The amplified products of the Indian isolates belonging to VT group contained eight nucleotides more than the original VT isolate. In the Indian VT type isolates, the 5′ UTR was longer with addition of two nucleotides in 83rd and 107th positions (not shown). A further addition of six nucleotides (at 177th to 179th and 419th to 421st nucleotide positions) did not introduce any shift in the frame of ORF1a in the Indian isolates belonging to VT group. However the aa sequences in the ORF1a showed only 92–93% identity with the original VT isolate. The maximum sequence identity at the 5′ UTR and part of ORF1a of Indian VT isolates sequences with SY568 placed them phylogenetically closer to SY568 rather than to the VT isolate.

In previous work on the variability of the 5′-terminal region, Lopez et al. (1998) grouped nine CTV isolates into three groups (I–III) on the basis of 5′ UTR and the 5′ proximal 159 nucleotides of ORF1a. Further analysis of the variability of the 5′ UTR and part of ORF1a of 58 cDNA clones of another 15 CTV isolates supported the three previously established groups (Ayllon et al., 2001), which included CTV isolates from Brazil, California, Corsica, Costa Rica, Cuba, Dominican Republic, Florida, Israel, Japan, South Africa, Spain, and Taiwan. CTV was first reported from India in 1958 (Vasudeva and Capoor, 1958); however, no Indian isolates sequence comparisons were included in these studies. On the basis of the present sequencing data of the 5′-terminal region of CTV gRNA from India, there is an increase in the complexity of the previous pattern due to the addition of the BAN-2-VT sequence in the CTV genotype classification.

BAN-2 is an interesting CTV isolate because the CPG of the isolate shared the closest identity with the Australian PB61 isolate (Roy et al., 2002) but the identified k17 region of ORF 1a was as a mixture of T3, T30, and VT genotypic isolates sequences (Roy and Brlansky, 2004). However, the overlapping region of RdRp and p33 gene had either a VT or an intermediate sequence of T30 and VT genotypic isolates (Roy and Brlansky, 2004). The BAN-2-VT-5′ sequence had partial similarity with isolates T30, VT, SY568, and NUagA at the

5′ UTR regions of the gRNA and a maximum aa relatedness with T30 (86%), and placed the isolate in a new genogroup in the phylogenetic tree. Based on sequence data from various regions of the gRNA, BAN-2 is a mixture of strains. Amplification with any of the genotype-specific 5′ MMM was not a predictor of the sequence homology. Similar results were found in the POL and k17 primer amplified sequencing data (Roy and Brlansky, 2004). Although the amplified sequence of BAN-2 isolate from ORF1a (697–1484 nt) shared the maximum relatedness with T36 genotype (89–90%), it was positioned in a different clade in the phylogenetic tree.

The 5′ region of the CTV genome is the most variable and a maximum of 70% identities among the nucleotide sequences have been found (Mawassi et al, 1995, 1996). The variable 5′region includes ORFs 1a, 1b, and 2. The genome size of ORF1a is ∼9.4 kb and more variability occurs in this region than ORFs 1b and 2. Comparison and analysis of different regions of ORF1a were used in this study in an attempt to standardize the CTV genogrouping. In an earlier study, we showed the presence of five genogroups in 21 Indian CTV isolates from the k17 region of ORF1a (∼4850–5230 nt) and the overlapping region of ORFs 1b and 2 (Roy and Brlansky, 2004). In this current study, an analysis of the 5′-end (1–500 nt) MMM amplified products also grouped all the isolates into four genogroups (Fig. 1). The number of genogroups increased from four to six when we analyzed the amplified products of CN487/489 primer pairs (697–1105 nt). The two additional clades contained the sole members isolate T3 from Florida and the Japanese isolate NUagA, respectively. When we analyzed the 1082–1484 nt-amplified products, the number of genogroups increased from six to eight (Fig. 2B) and an additional two Indian isolates BAN-1 and B165 were separated. Further analysis of the 697–1484 nt products did not change the genogrouping (Fig. 2C). Genogrouping was affected by the particular region from the 5′ half of the CTV genome.

Phylogenetic analysis of the CN487/491 amplified region showed the existence of three new genotypes (BAN-1, BAN-2, and B165) in the Indian CTV population. Mutations are the basis of all the known genotypes where we see both silent (synonymous) and nonsynonymous mutations. Further analysis of the 697–1484 nt using PHYLPRO showed that BAN-1 and B165 CTV isolates were evolved by recombination (Fig. 3). This finding supports the existence of three new sequence types among the studied isolates and explains the topology of the phylogenetic tree (Fig. 2). These data suggest that in addition to mutations, recombination may have also played an important role in evolution of CTV genotypes.

Isolates under study cause a wide range of mild, moderate and severe symptoms in different host plants (Roy and Brlansky, 2004). Symptom expression on different citrus hosts might depend on the relative titer of different sequence types of the mixture (Ayllon et al., 2001). Correlation of symptom expression in different hosts to a particular sequence does not mean that 5′-end sequences under study are directly involved in the pathogenesis. Regions in the genome responsible for symptom development may have co-evolved with the 5′-terminal end sequence. In a previous study (Satyanarayana et al., 1999), it was found that an engineered T36 isolate RNA replicon, had reduced replication levels in *Nicotiana benthamiana* protoplasts when the T36 5′ UTR was substituted individually by the isolates T30 (type II) or VT (type III). This result indicated that the T36 replicase interacts less efficiently with heterologous 5′ UTRs.

In summary, four to eight genogroups of CTV isolates were identified based on sequences from different regions of ORF1a. The degrees of variation and genetic relationship with other reported CTV isolates were shown by phylogenetic analysis. The sole member of the fourth genogroup, the BAN-2 isolate, proposed that a new genotype is present in India. Detection of recombinant events in the sequence of some isolates suggests that recombination may have played a significant role in evolution of CTV isolates since it is common to find mixtures of genotypes in infected plants (e.g., isolates BAN-1, BAN-2, and others in this study).

Acknowledgements

The authors are grateful to Dr. Padma Ramachandran, Indian Agricultural Research Institute, New Delhi, for providing the BAN-1 and BAN-2 CTV isolates at the Exotic Citrus Pathogen Collection at Beltsville, MD. The authors acknowledge the financial assistance from the USDA Specific Cooperative Agreement Exotic Citrus Disease Grant during this investigation. This research was supported by the Florida Agricultural Experiment Station, and approved for publication as Journal Series No. R-09958.

References

Albiach-Marti, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shankar, S., Almira, E.C., Vives, M.C., Lopez, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., Dawson, W.O., 2000. Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. J. Virol. 74, 6856–6865.

Ayllon, M.A., Rubio, L., Moya, A., Guerri, J., Moreno, P., 1999. The haplotype distribution of two genes of *Citrus tristeza virus* is altered after host change in aphid transmission. Virology 255, 32–39.

Ayllon, M.A., Lopez, C., Navas-Castillo, J., Garnsey, S.M., Guerri, J., Flores, R., Moreno, P., 2001. Polymorphism of the 5′ terminal region of *Citrus tristeza virus* (CTV) RNA: incidence of three sequence types in isolates of different origin and pathogenicity. Arch. Virol. 146, 27–40.

Bar-Joseph, M., Marcus, R., Lee, R.F., 1989. The continuous challenge of *Citrus tristeza virus* control. Annu. Rev. Phytopathol. 27, 291–316.

Brlansky, R.H., Damsteegt, V.D., Howd, D.S., Roy, A., 2003. Molecular analyses of *Citrus tristeza virus* subisolates separated by aphid transmission. Plant Dis. 87, 397–401.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programmes for the VAX. Nucl. Acids Res. 12, 387–395.

Febres, V.J., Ashoulin, L., Mawassi, M.E., Frank, A., Bar-Joseph, M., Manjunath, K.L., Lee, R.F., Niblett, C.L., 1996. The p27 protein is present at the end of the *Citrus tristeza virus* particles. Phytopathology 86, 1331–1335.

Hilf, M.E., Garnsey, S.M., 2000. Characterization and classification of *Citrus tristeza virus* isolates by amplification of multiple molecular markers. In: Proceedings of the 14th Conference on International Organ Citrus Virology, Brazil 1998, IOCV, Riverside, CA, pp. 18–27.

Hilf, M.E., Karasev, A.N., Albiach-Marti, R.M., Dawson, W.O., Garnsey, S.M., 1999. Two paths of sequence divergence in the *Citrus tristeza virus* complex. Phytopathology 89, 336–342.

Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D., Niblett, C.L., Garnsey, S.M., 1995. Characterization of *Citrus tristeza virus* subgenomic RNAs in infected tissue. Virology 208, 576–582.

Karasev, A.V., 2000. Genetic diversity and evolution of closteroviruses. Annu. Rev. Phytopathol. 38, 293–324.

Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaev, O.N., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K., Gumpf, D., Lee, R.F., Garnsey, S.M., Lewandowski, D.J., Dawson, W.O., 1995. Complete sequence of *Citrus tristeza virus* genome. Virology 208, 511–520.

Lopez, C., Ayllon, M.A., Navas-Castillo, J., Guerri, J., Moreno, P., Flores, R., 1998. Molecular variability of the 5′- and 3′-terminal regions of *Citrus tristeza virus* RNA. Phytopathology 88, 685–691.

Martelli, G.P., Agranovsky, A.A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R.H., Dolja, V.V., Falk, B.W., Gonsalves, D., Jelkmann, W., Karasev, A.V., Minafra, A., Namba, S., Vetten, H.J., Wisler, G.C., Yoshikawa, N., 2002. The family Closteroviridae revised. Arch. Virol. 147, 2039–2044.

Mawassi, M.E., Karasev, A.V., Mietkiewska, E., Gafny, R., Lee, R.F., Dawson, W.O., Bar-Joseph, M., 1995. Defective RNA molecules associated with *Citrus tristeza virus*. Virology 208, 383–387.

Mawassi, M.E., Mietkiewska, E., Gofman, R., Yang, G., Bar-Joseph, M., 1996. Unusual sequence relationships between two isolates of *Citrus tristeza virus*. J. Gen. Virol. 77, 2359–2364.

Nicholas Karl, B., Nicholas Hugh, Jr., B., 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. www.psc.edu/biomed/genedoc.

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., Cline, K.C., Garnsey, S.M., Dawson, W.O., Lee, R.F., Niblett, C.L., 1994. Nucleotide sequence and organization of eight 3′ open reading frames of the *Citrus tristeza closterovirus* genome. Virology 199, 35–46.

Roy, A., Ramachandran, P., Brlansky, R.H., 2002. Molecular characterization of Indian *Citrus tristeza virus* isolates. Phytopathology 92, 71.

Roy, A., Ramachandran, P., Brlansky, R.H., 2003. Grouping and comparison of Indian *Citrus tristeza virus* isolates based on coat protein gene sequences and restriction analysis patterns. Arch. Virol. 148, 707–722.

Roy, A., Brlansky, R.H., 2004. Genotype classification and molecular evidence for the presence of mixed infections in Indian *Citrus tristeza virus* isolates. Arch. Virol. 149, 1929.

Satyanarayana, T., Gowda, S., Boyko, V.P., Albiach-Martí, M.R., Mawassi, M., Navas Castillo, J., Karasev, A.V., Dolja, V., Hilf, M.E., Lewandowski, D.J., Moreno, P., Bar-Joseph, M., Garnsey, S.M., Dawson, W.O., 1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. U.S.A. 96, 7433–7438.

Suastika, G., Natsuaki, T., Terui, H., Kano, T., Ieki, H., Okuda, S., 2001. Nucleotide sequence of *Citrus tristeza virus* seedling yellow isolate. J. Gen. Plant Pathol. 67, 73–77.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 24, 4876–4882.

Vasudeva, R.S., Capoor, S.P., 1958. Citrus disease in Bombay State. Bull. Plant Prot. 6, 91.

Vives, M.C., Rubio, L., Lopez, C., Navas-Castillo, J., Albiach Marti, M.R., Dawson, W.O., Guerri, J., Flores, R., Moreno, P., 1999. The complete genome sequence of the major component of a mild *Citrus tristeza virus* isolate. J. Gen. Virol. 80, 811–816.

Weiller, G.F., 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol. Biol. Evol. 15, 326–335.

Yang, Z.N., Mathews, D.H., Doods, J.A., Mirkov, T.E., 1999. Molecular characterization of an isolate of *Citrus tristeza virus* that causes severe symptoms in sweet orange. Virus Genes 19, 131–142.