RESEARCH ARTICLE

Exploring the genetic diversity and recombination analysis of *Citrus tristeza virus* isolates prevalent in Northeast India

Trishnamoni Gautom[1,2] · Dibya Jyoti Hazarika[1] · Gunajit Goswami[1] · Madhumita Barooah[1] · Mohan Chandra Kalita[2] · Robin Chandra Boro[1]

Abstract

Citrus tristeza virus (CTV) is associated with three manifestations: citrus decline, stem pitting and seedling yellows of almost all citrus species and hybrids. Infected citrus plants harbour a large number of distinct CTV strains due to multiple and frequent inoculations by aphids and the fallible nature of its RNA polymerase. As such, it is important and necessary to identify different strains of CTV for proper management and disease control. In this study, we report the identification of 15 CTV isolates prevalent in three states of Northeast (NE) India: Assam, Meghalaya and Nagaland. A total of 50 samples were collected, out of which 15 showed CTV infection based on enzyme linked immuno-sorbent assay (ELISA) and reverse transcription-polymerase chain reaction (RT-PCR) assays. RT-PCR products of 674 bp and 407 bp corresponding to the coat protein (CP) gene and 5′ORF1a region respectively were sequenced and their phylogenetic analysis was carried out. Analysis of putative recombination events detected recombination in four isolates in the 5′ORF1a fragment and in one isolate in the CTV-CP gene. Furthermore, CP monomer structure of four CTV isolates was predicted by Phyre2 online structure prediction tool. Phylogenetic studies involving more number of CTV isolates reveal better understanding of the recombination or evolutionary pattern of the virus which would aid in developing advanced CTV diagnostic methods.

Keywords CTV · Coat protein · 5′ORF1a · Recombination · Monomer structure

Introduction

In terms of international trade, citrus is the highest value fruit crop grown globally. Citrus crops are cultivated in as many as 140 countries (D'Onghia et al. 2009) with 121 million tonnes of fruits produced annually (FAO 2017) and marketed worldwide. Citriculture is the second largest industry in India securing fifth position among the world's top citrus producing countries (FAO 2017). In India, citrus is the third most important horticultural crop after mango and banana

and is cultivated over an area of approximately 0.97 million ha with an annual production of 10 million tonnes (Ministry of Agriculture & Farmers Welfare 2016). It is generally grown commercially as a monoculture in majority of citrus growing areas. Therefore, any onset of epidemic to this crop may lead to enormous crop destruction. In the same context, its productivity has been greatly affected by a number of abiotic and biotic stresses. Among the biotic stresses, viruses and a fastidious greening pathogen cause significant yield loss in several commercial citrus species right from nursery level to fruit bearing stage. Among the viral diseases affecting citrus, 'tristeza' caused by CTV has the greatest impact worldwide (Moreno et al. 2008). In 1997, CTV affected more than one million trees in India and was declared as a century old problem in all citrus growing regions of the country (Ahlawat 1997).

CTV contains a single stranded, non-segmented, positive-sense RNA genome of approximately 20 kb and belongs to the genus, Closterovirus. The CTV genome is encapsulated by two capsid proteins of 25 kDa (CP) and 27 kDa (CPm) that coat 95% and 5% of the virion length,

✉ Robin Chandra Boro
robinboro@gmail.com

1 Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam 785013, India

2 Department of Biotechnology, Gauhati University, Guwahati, Assam 781014, India

respectively. The genome has 12 open reading frames (ORFs), potentially encoding at least 19 different proteins. The complete sequence of several CTV isolates have been reported, including the mild CTV isolates (Vives et al. 1999; Albiach-Martí et al. 2000; Suastika et al. 2001; Biswas et al. 2012b; Licciardello et al. 2015). The virus is phloem limited and is transmitted readily via infected graft and in a semipersistent manner by the main citrus visiting aphid species: *Toxoptera citricida*, *Aphis gossypi*, *Aphis spiraecola* and *Toxoptera aurantii* (Bar-Joseph and Lee 1989). The CTV is primarily associated with quick decline (QD) of citrus species grafted on sour orange rootstocks and stem pitting (SP) of grapefruit, pummelo and some sweet orange varieties irrespective of rootstock. A third disease, seedling yellows (SY) is also caused by CTV in young sour orange, lemon and grapefruit plants under experimental conditions. Damages caused by severe QD and SP isolates can be extremely devastating; however, trees with mild isolates show some productivity for a certain period (Bar-Joseph et al. 2002).

CTV is widespread in the citrus cultivating geographical zones of India viz. northwest, central northeast and south, with an estimated disease incidence of 10–90% (Ahlawat 1997; Biswas 2008). Novel pathogenicity traits are continuously being inculcated by recombination of RNA viruses which have been extensively documented (Chare and Holmes 2006). Therefore, knowledge on the genetic diversity of CTV and its geographical distribution plays a vital role in understanding the epidemiology of the virus resulting in proper diagnosis and development of long term management strategies. In this study, we have reported the distribution and characterization of CTV in Assam, Meghalaya and Nagaland states of NE India based on the CP gene and 5′ORF1a region, their putative recombination events and also made an attempt to predict the structure of the coat protein monomer.

Materials and methods

Procurement of sample

Suspected CTV infected *Citrus reticulata* Blanco leaf samples were collected from three states of NE India viz. Assam, Meghalaya and Nagaland showing vein clearing, leaf cupping and poor growth symptoms. The midrib along with the petiole of the collected leaf samples were cut into small pieces and dipped in tissue storage reagent, RNAlater solution (Sigma, USA), transported to lab and stored in – 80 °C until used for molecular analysis. A total of 50 samples were collected from the three states.

Double antibody sandwich enzyme linked immuno-sorbent assay (DAS-ELISA)

As a preliminary test to screen the collected samples for the presence of CTV infection, the suspected CTV infected samples were subjected to DAS-ELISA using CTV ELISA kit (BIOREBA AG, Switzerland) as per the manufacturer's protocol.

RNA isolation and synthesis of cDNA

Total plant RNA was isolated from the ELISA positive leaf samples using TRI-Reagent (Sigma, USA) as per the manufacturer's protocol. Synthesis of viral cDNA was achieved using SuperScript® III One-Step RT-PCR System with Platinum® *Taq* DNA polymerase (Invitrogen, USA) and CTV-CP and 5′ORF1a specific primers. The primers, CN488 (5′ TGTTCCGTCCTGSGCGGAAYAATT 3′) and CN491 (5′ GTGTARGTCCCRCGCATMGGAACC 3′) were used to amplify the 5′ end of ORF1a (407 nt) (Roy et al. 2005) and BC24 (5′ ATGGACGACGAAACAAAG AA 3′) and BC25 (5′ TCAACGTGTG TTGAATTTCC 3′) were used to amplify the entire CP gene (672 nt) (Çevik et al. 2013).

Cloning of CP and 5′ORF1a gene

The CTV-CP gene (672 bp) and 5′ORF1a region (407 bp) were cloned into the pGEM®-T Easy Vector System I (Promega, USA). Plasmids from the positive clones were isolated using Wizard® Plus SV Minipreps DNA Purification System (Promega, Madison, USA) and sequenced to obtain the nucleotide sequences. The sequencing was carried out using the vector derived primer according to the BigDye® Terminator v3.1 Cycle Sequencing Kit protocol in an ABI PRISM 3730xl automated DNA Sequencer (Applied Biosystems, USA). The sequence reads obtained after sequencing were assembled into contigs using CodonCode Aligner (CodonCode Corporation, USA). Sequence similarity tool BLAST was employed to find the similarity of the sequences with known CP and 5′ORF1a sequences available in NCBI. The sequences were deposited in GenBank of NCBI and accession numbers obtained.

Sequence analysis

Multiple sequence alignment of the CP gene and 5′ORF1a was performed using MEGA6 (Tamura et al. 2013). CLUSTALW 1.6 was employed as DNA paired matrix and alignment was performed using default parameters of MEGA6. The genomes of international isolates: VT

(Mawassi et al. 1996), Nuaga (Suastika et al. 2001), SY568 (Vives et al. 2005), T318A (Ruiz-Ruiz et al. 2006), T36 (Karasev et al. 1995), Qaha (GenBank Accession No. AY340974), Mexican isolate (GenBank Accession No. DQ272579), HA 16-5 (Melzer et al. 2010), T3 (GenBank Accession No. KC525952), NZRB-TH28 (Harper et al. 2010), T68-1 (GenBank Accession No. JQ965169) and T30 (Albiach-Martí et al. 2000) as well as Indian isolates: K27, K10, K5, Kpg2, Kpg3, D13, B165 and AG28 (Roy and Brlansky 2010; Biswas et al. 2012a; Tarafdar et al. 2013) were retrieved from GenBank and used as reference sequences for CP gene and 5′ORF1a region. Phylogenetic analysis of the sequences isolated for both CP gene and 5′ORF1a was also performed using the MEGA6 program (Tamura et al. 2013). The evolutionary history was inferred using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura 1980). The tree with the highest log likelihood (for CP: − 2824.1850 and for 5′ORF1a: − 2607.0037) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The analysis involved 35 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 663 positions in the final dataset for the CP phylogram and 401 for that of the 5′ORF1a region.

Potential recombination events in the CP gene and 5′ORF1a sequences of the isolated strains were identified using the recombination detecting program RDP4 (Martin et al. 2010). Recombination analysis of the aligned sequences of all the positive isolates for CP gene and 5′ORF1a region along with their reference sequences was performed using seven algorithms of RDP4 (RDP, Geneconv, BootScan, Maxchi, Chimera, SiScan and 3Seq). The same recombination events detected by two or more algorithms were considered to be evidence for putative recombination.

Structure prediction of the coat protein monomer

The three dimensional structure of the monomeric subunit was predicted from the amino acid sequence of the CP gene. The nucleotide sequence of the CP gene for each isolate was translated to obtain its amino acid sequence using the translate tool of the ExPASy server (http://web.expasy.org/translate). Translated sequences were subjected to multiple sequence alignment in MEGA6.0 using default parameters of the PAM protein weight matrix. MEGA6.0 was also used to calculate the pairwise distances among the protein sequences. Analyses were conducted using the Poisson correction model (Zuckerkandl and Pauling 1965). All positions containing gaps and missing data were eliminated and evolutionary analyses were conducted (Tamura et al. 2013). Sequences with maximum distances were selected for molecular modelling of the respective coat protein structures. Three dimensional structures were predicted using the Phyre2 online protein structure prediction tool (Kelley et al. 2015). Models were generated using the intensive mode of modelling and the quality of the models was checked using the Phyre2 investigator.

Results

Screening of the collected samples for CTV infection

A total of 50 samples were collected from three states of northeast India viz., Assam, Meghalaya and Nagaland. The samples were screened for the presence of CTV infection by DAS-ELISA and 15 samples tested positive for CTV. Of the infected samples, three each were from Meghalaya and Nagaland and rest of the samples belonged to Assam (Online Resource S1). CTV infection of the 15 isolates was further confirmed by RT-PCR. An amplicon size of 672 bp and 407 bp was observed with the CP specific primer and 5′ORF1a primer respectively (Online Resource S2 and S3). No amplification was observed with samples that were negative in DAS-ELISA. The sequences of CP gene and 5′ORF1a region were analysed by BLAST and all the sequences showed identity to severe strains of CTV deposited in NCBI. The CP gene and 5′ORF1a sequences were successfully deposited in the GenBank and accession numbers obtained (Table 1).

Phylogenetic analysis of the CTV isolates

For the CP gene, a phylogenetic tree was constructed using maximum likelihood method which generated nine groups (Fig. 1a). Although one-third of the CTV isolates of the present study were grouped in the Kpg3 (two isolates: TG1 and TG2) and K5 (three isolates: TG7, TG8, TG15) groups, however, the majority of the isolates from NE India formed two new clusters designated as New1 (six isolates: TG9, TG10, TG11, TG12, TG13 and TG14) and New2 (four isolates: TG3, TG4, TG5 and TG6) which did not group with any reference sequences used in the study. The Kpg3 group also included the Indian isolate D13 and the Hawaiian isolate HA16-5 while mild Florida isolate T30 remained as a single separate group to its nearest neighbour group Kpg2.

For the 5′ORF1a region, the phylogenetic tree was recorded with seven groups: Group I–Group VII (Fig. 1b). The isolates from NE India in the present study were scattered in four different groups. The isolates TG9, TG10, TG12, TG13, TG14 and TG15 belonged to Group IV;

Table 1 Attributes of the CTV isolates prevalent in NE India

| Isolate code	GeneBank accession number for CP	Protein_ID	Group for CP gene	GeneBank accession number for 5'ORF1a	Protein_ID	Group for 5'ORF1a	Collection site	Pathotype	Collection year	Host
TG1	KY436740	AUB45085	Kpg3	KY612348	AUR45203	I	Jorhat, Assam	Severe	2015	*Citrus reticulata* Blanco
TG2	KY436741	AUB45086	Kpg3	KY612349	AUR45204	VII	Jorhat, Assam	Severe	2015	*Citrus reticulata* Blanco
TG3	KY436742	AUB45087	New2	KY612350	AUR45205	I	Tinsukia, Assam	Severe	2016	*Citrus reticulata* Blanco
TG4	KY436743	AUB45088	New2	KY612351	AUR45206	I	Tinsukia, Assam	Severe	2016	*Citrus reticulata* Blanco
TG5	KY436744	AUB45089	New2	KY612352	AUR45207	VII	Boko, Assam	Severe	2016	*Citrus reticulata* Blanco
TG6	KY436745	AUB45090	New2	KY612353	AUR45208	VII	Boko, Assam	Severe	2016	*Citrus reticulata* Blanco
TG7	KY436746	AUB45091	K5	KY612354	AUR45209	VII	Mukokchang, Nagaland	Severe	2015	*Citrus reticulata* Blanco
TG8	KY436747	AUB45092	K5	KY612355	AUR45210	III	Mukokchang, Nagaland	Severe	2015	*Citrus reticulata* Blanco
TG9	KY436748	AUB45093	New1	KY612356	AUR45211	IV	Boko, Assam	Severe	2015	*Citrus reticulata* Blanco
TG10	KY436749	AUB45094	New1	KY612357	AUR45212	IV	Boko, Assam	Severe	2015	*Citrus reticulata* Blanco
TG11	KY436750	AUB45095	New1	KY612358	AUR45213	VI	Boko, Assam	Severe	2015	*Citrus reticulata* Blanco
TG12	KY436751	AUB45096	New1	KY612359	AUR45214	IV	Umsning, Meghalaya	Severe	2016	*Citrus reticulata* Blanco
TG13	KY436752	AUB45097	New1	KY612360	AUR45215	IV	Umsning, Meghalaya	Severe	2016	*Citrus reticulata* Blanco
TG14	KY436753	AUB45098	New1	KY612361	AUR45216	IV	Umsning, Meghalaya	Severe	2016	*Citrus reticulata* Blanco
TG15	KY436754	AUB45099	K5	KY612362	AUR45217	IV	Mukokchang, Nagaland	Severe	2015	*Citrus reticulata* Blanco
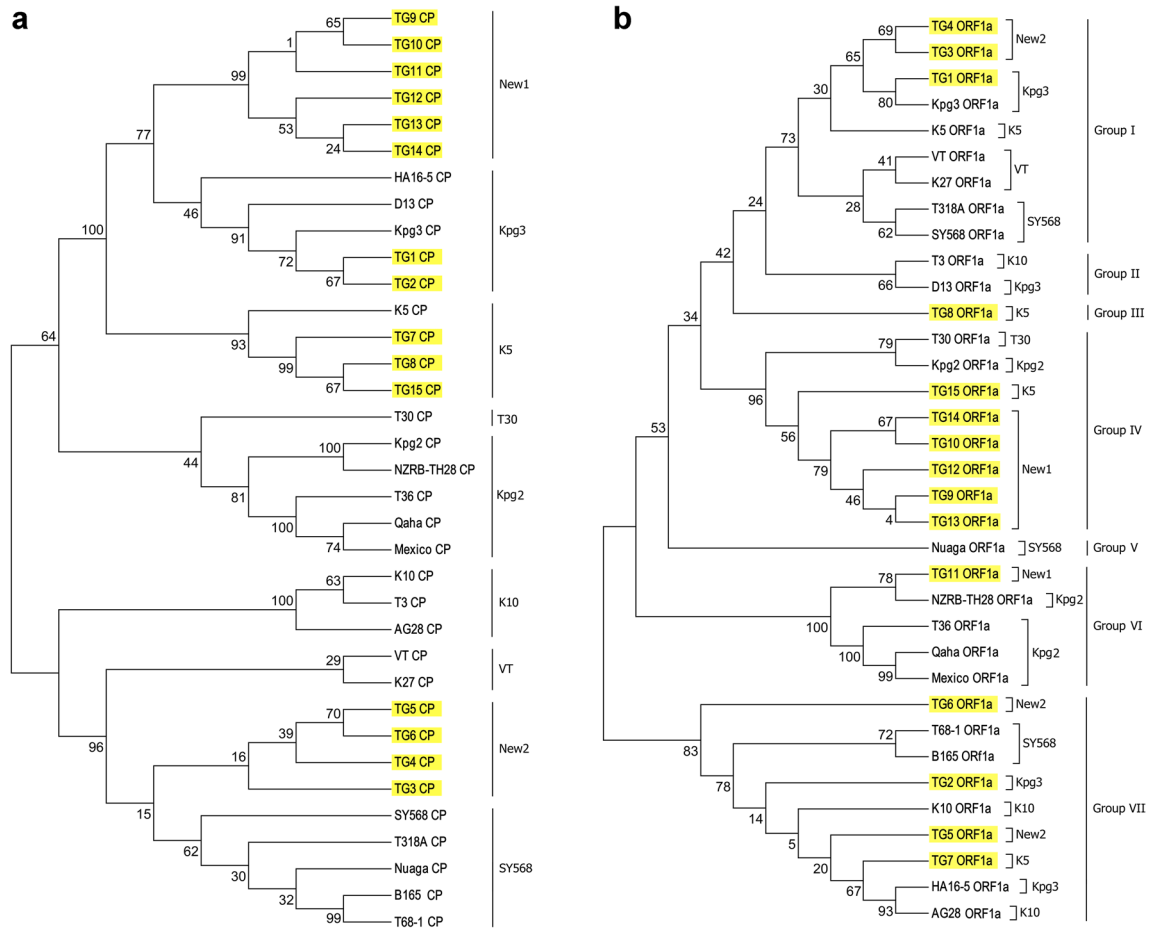
Fig. 1 Phylogenetic analysis of CTV isolates based on the sequences of **a** CP gene and **b** 5′ORF1a fragment using maximum likelihood method with 1000 bootstrap repetitions. The bootstrap values are denoted at the branch nodes. The CTV isolates of our study are highlighted in yellow showing genetic diversity among them and also with the reference isolates

isolates TG2, TG5, TG6 and TG7 were included in Group VII and isolates TG1, TG3, TG4 in Group I. The isolate TG8 from Nagaland formed a separate group designated as Group III with no other members. The isolate TG11 from Assam was grouped with T36 from Florida, Mexico isolate, Qaha from Egypt and NZRB-TH28 from New Zealand as Group IV.

Recombination analysis

The CP gene of the isolate TG5 was found to be a recombinant with T318A (major parent) and AG28 (minor parent) as revealed by the RDP4 program. A recombination event was detected at sequence position from 90 to 567 nucleotides of CP gene, supported by average P value of 2.816×10^{-2} in GENECONV and 3.356×10^{-1} in BootScan detection method of RDP4. For 5′ORF1a sequence, isolates TG6, TG7, TG8 and TG15 showed recombination events, supported by more than one detection method implemented in RDP4. Isolates TG6 and TG8 were confirmed as potent

recombinants at 5′ORF1a region by all the employed detection methods viz. RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan and 3Seq. D13 was found to be the major parent for both the isolates, while K10 and AG28 were detected as minor parents for TG6 and TG8, respectively. Isolates TG7 and TG15 were also detected as recombinant at the 5′ORF1a region by five detection methods for both TG7 (BootScan, MaxChi, Chimera, SiScan and 3Seq) and TG15 (RDP, BootScan, MaxChi, Chimera, and 3Seq). The AG28 isolate was detected as the major parent for TG7 and as the minor parent for TG15, while K10 served as the minor parent for TG7 and Kpg2 was found to be the major parent for TG15. The recombination breakpoints for all the recombination events along with the average P-values for all the algorithms are presented in Table 2.

Three dimensional structure prediction

CTV isolates were selected for the prediction of 3D structures based on the pairwise distances among the protein

Table 2 Recombination events in 5′ORF1a fragment and CTV coat protein sequence of CTV isolates as detected by recombination-detecting program RDP4

Isolate code	Major parent	Minor parent	Recombination breakpoints	Recombination detection methods	p value
For 5′ORF 1a fragment					
TG6	D13	K10	Beginning—20 nucleotides Ending—220 nucleotides	RDP	2.066×10^{-2}
				GENECONV	1.063×10^{-2}
				BootScan	2.008×10^{-4}
				MaxChi	6.503×10^{-7}
				Chimera	4.185×10^{-3}
				SiScan	1.078×10^{-8}
				3Seq	9.011×10^{-6}
TG7	AG28	K10	Beginning—283 nucleotides Ending—391 nucleotides	BootScan	3.247×10^{-2}
				MaxChi	8.225×10^{-3}
				Chimera	4.305×10^{-3}
				SiScan	4.228×10^{-3}
				3Seq	6.236×10^{-4}
TG8	D13	AG28	Beginning—42 nucleotides Ending—133 nucleotides	RDP	8.644×10^{-3}
				GENECONV	5.501×10^{-3}
				BootScan	6.472×10^{-4}
				MaxChi	4.093×10^{-3}
				Chimera	1.963×10^{-3}
				SiScan	2.757×10^{-3}
				3Seq	3.032×10^{-5}
TG15	Kpg2	AG28	Beginning—320 nucleotides Ending—43 nucleotides	RDP	8.833×10^{-3}
				BootScan	7.255×10^{-3}
				MaxChi	3.036×10^{-4}
				Chimera	2.965×10^{-3}
				3Seq	2.023×10^{-4}
For CP gene					
TG5	T318A	AG28	Beginning—90 nucleotides Ending—567 nucleotides	GENECONV	2.816×10^{-2}
				BootScan	3.356×10^{-1}

sequences using MEGA6.0. Protein sequences of four isolates (TG4, TG6, TG11 and TG15) showed maximum pairwise distances among themselves (Online Resource S4), hence were selected for structure generation. Models for monomeric subunit of coat protein were generated using the Phyre2 online protein modelling tool and final models were generated with > 95% alignment confidence with the respective templates for 85% residues of isolate TG6, 78% residues of isolate TG15, 71% residues of isolateTG4 and 70% residues of isolate TG11. The final models were generated based on the pre-determined cryo-electron microscopic structure of the coat protein molecule (chain A) of filamentous flexible virus pepmv (pepino2 mosaic virus; PDB ID: C5N1) and coat protein (chain B) structure of watermelon mosaic potyvirus (PDB ID: C5N1). Predicted final models of TG4, TG6, TG11 and TG15 are shown in Fig. 2. Around 50–55% of the sequence in the final models was folded into α-helices and 7–8% was folded into β-sheets. In order to validate the quality of the generated models, Phyre2 investigator was used to assess the quality of the predicted models
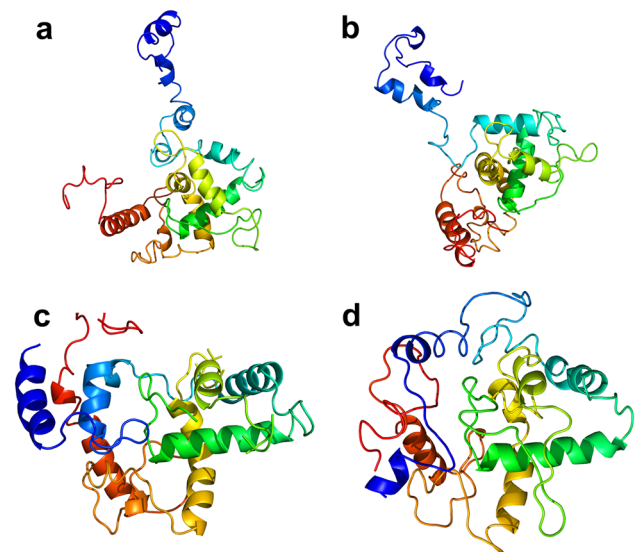
Fig. 2 Predicted 3D structures of the coat protein monomer of isolates TG4 (**a**), TG6 (**b**), TG11(**c**) and TG15 (**d**)

with respect to the corresponding most identical template (Kelley et al. 2015). ProQ2 quality check indicated average quality of the predicted models. Clash analysis was performed via Phyre2 investigator to observe the side chain placement and the results indicated > 95% of the amino acids in the selected sequences formed backbone correctly with good side chain placement. This result was also supported by rotamer analysis. Structure validation was also performed through Ramachandran analysis and all the amino acids were found to be in the good or allowed regions, suggesting the correct structural conformation. Disorder prediction was made by the knowledge-based Disopred method included in Phyre2 investigator, which suggested that all the four models had no or negligible disorders. The results of Phyre2 investigation were shown in Online Resource S5.

Discussion

Although the first report on the occurrence of tristeza disease of citrus in NE India dates back to 1989 (Bhagbati et al. 1989), till date very few research has been focused on the citrus decline due to CTV in this part of the country. Meanwhile, it is worth mentioning that one of the centres for the origin of different citrus species is the Northeastern part of India. In this study, we collected mandarin orange (*Citrus reticulata* Blanco) leaf samples from three states of NE India to analyze the prevalence of CTV infection. Mandarin orange farms were included in the study due to their widespread cultivation in this region and also mandarin ranking the sixth in the world fruit production (FAO 2017). Moreover, CTV does not produce much detectable symptoms on mandarin orange (Biswas et al. 2012a) and the symptoms of this disease often mask with symptoms of other bacterial diseases like citrus canker and citrus greening as well as poor growth due to nutrient deficiency making its detection difficult. Therefore, we adopted ELISA followed by RT-PCR using the CP specific primer for diagnosis of CTV infection. The error-prone nature of RNA polymerases, selection pressures along with super infection of field trees with divergent CTV variants, genetic drift after transmission to new hosts or recombination results in frequent genetic variations in CTV isolates (Martín et al. 2009). Therefore, gaining an insight to the newly evolved genetic structure of viral populations could provide better understanding about not only the outbreaks of new epidemics but also about virulence changes in current isolates thus resulting in proper diagnosis and development of long term management strategies.

The most variable region of the CTV genome is the 5′ region with more variability at the ORF1a than the ORF1b and ORF2 (Roy et al. 2005). Phylogenetic analysis of the 5′ORF1a region and the CP gene was carried out by comparing corresponding sequences of available international and Indian isolates. On the basis of the CP gene sequences, isolates from Jorhat, Assam were grouped with the Indian isolate Kpg3 from Darjeeling Hills and those from the state of Nagaland were grouped with the Indian isolate K5 from Kalimpong. Previously reported isolates from Jorhat, Assam and Mukokchang, Nagaland were grouped with K5 isolate (Tarafdar et al. 2013; Kashyap et al. 2015) which differed from our study in case of the Jorhat isolates. Interestingly, we also came across two distinct groups referred to as New1 and New2 which exclusively included the remaining isolates of our present study with no reference CP sequences. However, previously reported isolates from Tinsukia, Assam were grouped with K5 and Kpg3 (Tarafdar et al. 2013; Kashyap et al. 2015), but Tinsukia isolates, TG3 and TG4 in our study formed one of the distinct groups New2. It is also worth mentioning that two different sets of samples collected from two nearby farms from the same site (Boko, Assam) a year apart were grouped in different groups: isolates TG9, TG10, TG11 in New1 and TG5, TG6 in New2 group. This is the first report on the presence of CTV infection in the Boko region of Assam and Umsning region of Meghalaya. Although, there have been reports on the prevalence of CTV infection in other parts of Meghalaya, however, those isolates were grouped in K5 and VT groups (Tarafdar et al. 2013). This variation in the distribution of isolates in the phylogram suggests that CTV population is evolving continuously and spreading to new areas posing economic threat and concern to the local citrus growers. In this context, during our survey on mandarin orchards, we came across many low income citrus growers who are planning to switch to some other fruit crops after facing productivity loses from their orchards.

The phylogenetic tree based on the sequences of 5′ORF1a region showed that isolate TG15, a member of K5 group in the CP phylogram, was included in Group IV along with other New1 members except for TG11 which is differently placed in Group VI. The Meghalaya isolates from our study and that from a previous study were placed with Kpg2 and T30 isolates based on their 5′ORF1a region (Tarafdar et al. 2013). However, a previous study reported isolates from Tinsukia to group together with the Hawaiian isolate HA16-5 in 5′ORF1a phylogram which in our study were placed with Indian isolates Kpg3, K5 and Israeli isolate VT (Tarafdar et al. 2013). The Jorhat isolates also showed a similar pattern signifying that the genetic variability exists within the isolates prevalent in NE India which could be due to any of the above mentioned factors. There have been no previous reports on phylogenetic analysis of 5′ORF1a region of CTV isolates from Nagaland. Our study also ascertained that the variation within the CP gene and the 5′ORF1a region are two independent events.

In the CP and 5′ORF1a-based phylograms, the CTV isolates occupied a dissimilar position indicating the possibility of recombination events (Fig. 1). To further investigate the recombination events in CTV isolates, we used the 672 bp sequence of CP gene and 407 bp sequence of 5′ORF1a and analysed in RDP4. Putative recombination events of four isolates (TG6, TG7, TG8 and TG15) in 5′ORF1a sequence and that of one isolate (TG5) in CP gene was observed. Further analysis revealed that except for the isolate TG6, recombination events in other four isolates involving both 5′ORF1a and CP gene, AG28 was found to be one of the parents. Interestingly, the AG28 isolate was earlier reported from Assam (Tarafdar et al. 2013) which in our study displayed putative parental sequences of CTV isolates from Nagaland and Assam indicating a possible transmission by aphid vectors. Isolate TG5 from Assam, showing putative recombination with major parent, Spanish isolate T318A did not fall in the same group considering the CP gene phylogram. Although isolates TG1, TG3 and TG4 were also grouped with T318A in the CP phylogram, no putative recombination events were found within their CP gene (Table 2). This is the first instance where the CP gene of an isolate from Assam was predicted to have evolved from the recombination of a Spanish and an Indian isolate. It was also evident that most of the recombination events occurred in the 5′ORF1a region, with only one recombination event detected in the CP gene, suggesting that the 5′ORF1a region could be a recombination hotspot in the CTV genome which was also reported earlier (Biswas et al. 2012a). Recombination event in CTV isolates of distant geographical regions could be due to selection pressure and multiple infections by successive aphids (Wu et al. 2015). Recombination event in CTV may help balance the effect of deleterious mutation and may lead to greater genome diversity assisting the virus in adapting to new environments (Simon and Bujarski 1994).

Determining the structure of the CTV coat protein is important to investigate the virulence and diagnosis. Since, no structural representation is available till date for the coat protein of CTV, therefore, comparative modeling approach was tried to design the 3D structure of the coat protein monomeric subunit. The challenge of template-based modeling depends on the correct template recognition and generation of correct sequence-template alignments (Peng and Xu 2010).When sequence identity is below 30%, it is difficult to identify the perfect template and generate precise structural models (Sánchez et al. 2000; Chakravarty et al. 2008). Our sequences showed very low homology, therefore we shifted to a low homology protein threading using Phyre2 as this server finds multiple templates to predict the final model (Kelley et al. 2015). The four sequences selected for the generation of 3D subunits resulted in four different conformations due to the diversity in their amino acid sequences. Our models were generated with high confidence level and

acceptable results for Ramachandran analysis and other parameters; however, wet lab validation would be required to determine the proper monomeric structure of the coat protein. This computational approach was used only to gain a hypothetical insight into the structural morphology of the coat protein monomer. Molecular docking approach may also be used to assemble the complete virus particle to check if our monomeric models could form the complete viral capsid.

The perennial nature of citrus, prolonged infection of CTV for several years together with repeated and continuous infestation by aphid vectors and recombination phenomenon could be responsible for the complex CTV population. Gaining an insight into the genetic diversity and geographical distribution of CTV population is pre-requisite in understanding the epidemiology and developing long term management strategy against the virus. Phylogenetic studies involving more number of CTV isolates will reveal better understanding of the recombination or evolution pattern of the virus which may be helpful in developing advanced CTV diagnostic methods.

Compliance with ethical standards

References

Ahlawat YS (1997) Viruses, greening bacterium and viroids associated with citrus (Citrus species) decline in India. Indian J Agric Sci 67:51–57

Albiach-Martí MR, Mawassi M, Gowda S, Satyanarayana T, Hilf ME, Shanker S, Almira EC, Vives MC, López C, Guerri J, Flores R, Moreno P, Garnsey SM, Dawson WO (2000) Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. J Virol 74:6856–6865. https://doi.org/10.1128/JVI.74.15.6856-6865.2000

Bar-Joseph M., Lee RF (1989) *Citrus tristeza virus*. In: AAB descriptions of plant viruses No. 353. http://www.dpvweb.net/dpv/showadpv.php?dpvno=353. Retrieved on 7th Jan 2019

Bar-Joseph M, Che X, Mawassi M, Gowda S, Satyanarayana T, Ayllón M a, Albiach-martí MR, Garnsey SM, Dawson WO (2002) The continuous challenge of *Citrus tristeza virus* molecular research. In: Fifteenth IOCV conference, pp 1–7

Bhagbati KN, Ahlawat YS, Chakraborty NK, Borthakur BC (1989) Distribution of greening, tristeza and mosaic diseases of citrus in North-Eastern States of India. Indian Phytopathol 42:552–555

Biswas KK (2008) Molecular diagnosis of Citrus tristeza virus in mandarin (*Citrus reticulata*) orchards of Darjeeling hills of West Bengal. Indian J Virol 19:26–31

Biswas KK, Tarafdar A, Diwedi S, Lee RF (2012a) Distribution, genetic diversity and recombination analysis of *Citrus tristeza virus* of India. Virus Genes 45:139–148. https://doi.org/10.1007/s11262-012-0748-3

Biswas KK, Tarafdar A, Sharma SK (2012b) Complete genome sequence of mandarin decline *Citrus tristeza virus* of the Northeastern Himalayan hill region of India: comparative analyses determine recombinant. Arch Virol 157:579–583. https://doi.org/10.1007/s00705-011-1165-y

Çevik B, Yardimci N, Korkmaz S (2013) The first identified *Citrus tristeza virus* isolate of Turkey contains a mixture of mild and severe strains. Plant Pathol J 29:31–41. https://doi.org/10.5423/PPJ.OA.09.2012.0141

Chakravarty S, Godbole S, Zhang B, Berger S, Sanchez R (2008) Systematic analysis of the effect of multiple templates on the accuracy of comparative models of protein structure. BMC Struct Biol 8:31. https://doi.org/10.1186/1472-6807-8-31

Chare ER, Holmes EC (2006) A phylogenetic survey of recombination frequency in plant RNA viruses. Arch Virol 151:933–946. https://doi.org/10.1007/s00705-005-0675-x

D'Onghia A, Djelouah K, Roistacher C (eds) (2009) *Citrus tristeza virus* and toxoptera citricidus: a serious threat to the Mediterranean citrus industry—IAM Bari: CIHEAM (Centre International De Hautes Etudes Agronomiques Méditerranéennes). In: Série B studies and research, no. 65, Options Méditerranéennes, pp 13–18

FAO (2017) Citrus fruit—fresh and processed statistical bulletin 2016. Food & Agriculture Organisation of the United Nations (FAO), Rome

Harper SJ, Dawson TE, Pearson MN (2010) Isolates of *Citrus tristeza virus* that overcome *Poncirus trifoliata* resistance comprise a novel strain. Arch Virol 155:471–480. https://doi.org/10.1007/s00705-010-0604-5

Karasev AV, Boyko VP, Gowda S, Nikolaeva OV, Hilf ME, Koonin EV, Niblett CL, Cline K, Gumpf DJ, Lee RF, Garnsey SM, Lewandowski DJ, Dawson WO (1995) Complete sequence of the *Citrus tristeza virus* RNA genome. Virology 208:511–520. https://doi.org/10.1006/VIRO.1995.1182

Kashyap A, Nath PD, Acharjee S, Biswas KK (2015) Prevalence of *Citrus tristeza virus* in North Eastern region of India and molecular characterization of its isolates. Indian J Hortic 72:206–211. https://doi.org/10.5958/0974-0112.2015.00040.7

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120. https://doi.org/10.1007/BF01731581

Licciardello G, Scuderi G, Ferraro R, Giampetruzzi A, Russo M, Lombardo A, Raspagliesi D, Bar-Joseph M, Catara A (2015) Deep sequencing and analysis of small RNAs in sweet orange grafted on sour orange infected with two *Citrus tristeza virus* isolates prevalent in Sicily. Arch Virol 160:2583–2589. https://doi.org/10.1007/s00705-015-2516-x

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462–2463. https://doi.org/10.1093/bioinformatics/btq467

Martín S, Sambade A, Rubio L, Vives MC, Moya P, Guerri J, Elena SF, Moreno P (2009) Contribution of recombination and selection to molecular evolution of *Citrus tristeza virus*. J Gen Virol 90:1527–1538. https://doi.org/10.1099/vir.0.008193-0

Mawassi M, 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. https://doi.org/10.1099/0022-1317-77-9-2359

Melzer MJ, Borth WB, Sether DM, Ferreira S, Gonsalves D, Hu JS (2010) Genetic diversity and evidence for recent modular recombination in Hawaiian *Citrus tristeza virus*. Virus Genes 40:111–118. https://doi.org/10.1007/s11262-009-0409-3

Ministry of Agriculture & Farmers Welfare (2016) Horticultural statistics at a glance 2015. Oxford University Press, Oxford

Moreno P, Ambrós S, Albiach-Martí MR, Guerri J, Peña L (2008) *Citrus tristeza virus*: a pathogen that changed the course of the citrus industry. Mol Plant Pathol 9:251–268. https://doi.org/10.1111/j.1364-3703.2007.00455.x

Peng J, Xu J (2010) Low-homology protein threading. Bioinformatics 26:i294–i300. https://doi.org/10.1093/bioinformatics/btq192

Roy A, Brlansky RH (2010) Genome analysis of an orange stem pitting *Citrus tristeza virus* isolate reveals a novel recombinant genotype. Virus Res 151:118–130. https://doi.org/10.1016/j.virusres.2010.03.017

Roy A, Manjunath KL, Brlansky RH (2005) Assessment of sequence diversity in the 5′-terminal region of *Citrus tristeza virus* from India. Virus Res 113:132–142. https://doi.org/10.1016/j.virusres.2005.04.023

Ruiz-Ruiz S, Moreno P, Guerri J, Ambrós S (2006) The complete nucleotide sequence of a severe stem pitting isolate of *Citrus tristeza virus* from Spain: comparison with isolates from different origins. Arch Virol 151:387–398. https://doi.org/10.1007/s00705-005-0618-6

Sánchez R, Pieper U, Melo F, Eswar N, Martí-Renom MA, Madhusudhan MS, Mirković N, Šali A (2000) Protein structure modeling for structural genomics. Nat Struct Biol 7:986–990. https://doi.org/10.1038/80776

Simon AE, Bujarski JJ (1994) RNA–RNA recombination and evolution in virus-infected plants. Annu Rev Phytopathol 32:337–362

Suastika C, Natsuaki T, Terui H, Kano T, Ieki H, Okuda S (2001) Nucleotide sequence of *Citrus tristeza virus* seedling yellows isolate. J Gen Plant Pathol 67:73–77. https://doi.org/10.1007/PL00012992

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

Tarafdar A, Godara S, Dwivedi S, Jayakumar B, Biswas K (2013) Characterization of *Citrus tristeza virus* and determination of genetic variability in North–east and South India. Indian Phytopath. 66:302–307

Vives MC, Rubio L, Navas-Castillo J, Albiach-Mart MR, Dawson WO, 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

Vives MC, Rubio L, Sambade A, Mirkov TE, Moreno P, Guerri J (2005) Evidence of multiple recombination events between two RNA sequence variants within a *Citrus tristeza virus* isolate. Virology 331:232–237. https://doi.org/10.1016/j.virol.2004.10.037

Wu G-W, Tang M, Wang G-P, Jin F-Y, Yang Z-K, Cheng L-J, Hong N (2015) Genetic diversity and evolution of two capsid protein genes of *Citrus tristeza virus* isolates from China. Arch Virol 160:787–794. https://doi.org/10.1007/s00705-014-2281-2

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, p 97166