Detection of ‘relict’ western lineage of citrus tristeza virus virulent genotype in declining Arunachal Wakro orange

Short Title: CTV in Arunachal Wakro Orange

Sibnarayan Datta^{1*}, Bidisha Das^{1@}, Raghvendra Budhauya^{1@#}, Reji Gopalakrishnan^{1§},
Vanlalh Muaka¹, Mukesh K Meghvansi^{1£}, Saifor Rahman², Sanjai K Dwivedi¹, Vijay
Veer¹

¹Entomology & Biothreat Management Division, Defence Research Laboratory (DRL-DRDO), Tezpur, Assam, INDIA

²Department of Horticulture, Government of Arunachal Pradesh, INDIA

@these authors contributed equally to this manuscript.

this author has deceased.

[§]Present address- Defence Bioengineering and Electromedical Laboratory (DEBEL-DRDO), Bangalore-560093, Karnataka, INDIA

[£] Present address- Bioprocess Technology Division, Defence Research & Development Establishment (DRDE-DRDO), Gwalior-474002, Madhya Pradesh, INDIA

*Email sndatta@drl.drdo.in (S Datta)

24 ABSTRACT

25 Eastern Himalayan foothills are known to have optimal agro-climatic conditions for
 26 production of quality citrus fruits including oranges. Among the citrus growing regions of
 27 eastern Himalayas, Wakro in the far eastern state of Arunachal Pradesh is known for its
 28 superior quality oranges, popular as the *Wakro orange* or the *Arunachal orange*, which has
 29 been included in Geographical Indication Registry by the Government of India. However,
 30 during last few years, Arunachal orange orchards have been experiencing severe
 31 infestation of aphid associated with rapid decline disease, causing catastrophe to the
 32 farmers as well as state economy. Therefore, in 2015, an intensive survey of severely
 33 affected orange orchards was carried out to investigate the aetiological factors of citrus
 34 decline. RNA samples extracted from leaf and aphid specimens collected from Wakro
 35 orchards were subjected to CTV detection through 3'-UTR specific RT-PCR.
 36 Subsequently, ORF1 and CP genetic regions were amplified and clonal-sequencing was
 37 performed. Although, BLAST search showed close homology of the present sequences
 38 with other virulent genotype VT sequences, detailed phylogenetic analysis demonstrated
 39 affinity and clustering of present sequences with VT sequences belonging to the 'western'
 40 lineage. This finding is considerably distinct from CTV sequences reported from citrus
 41 growing orchards in India and other neighbouring countries. Additionally, low diversity of
 42 CP gene sequences, recombination patterns and presence of sequence segments identical to
 43 the present ones in other CTV genotypes was also revealed. Collectively, these
 44 observations indicate pristine and primeval nature of present CTV sequences,
 45 corroborating well with the proposed origin of CTV in this part of the globe. We here
 46 report our finding of western lineage of CTV virulent genotype VT, which is distinct from
 47 CTV molecular epidemiology in other parts of India and discuss the implications of these
 48 findings.

49
 50 **Keywords:** CTV, CP gene, ORF1, genotype VT, Wakro Orange

51

52 INTRODUCTION

53 Citrus group of fruits include oranges, mandarins, grapefruits, lemons, limes etc. and are
 54 widely grown throughout the world in large quantities¹. However, production of citrus is
 55 largely restricted by the citrus tristeza virus (CTV), which is one of the economically most
 56 devastating viral pathogens of citrus. CTV is prevalent in almost all major citrus growing
 57 regions of the world and can infect almost all citrus species and hybrids and has caused a
 58 loss of more than 100 million citrus trees during the last century²⁻⁵. Long distance spread
 59 of CTV occurs via movement of vegetatively propagating material such as infected grafts
 60 (primary propagation), while local spread (secondary propagation) is facilitated by diverse
 61 aphid species in a semipersistent manner, of which the brown citrus aphid (BrCA,
 62 *Toxoptera citricida*) and the melon aphid (*Aphis gossypii*) are reported to be the most
 63 efficient vectors^{2,3,6}. CTV is not known to be spread by seeds^{3,4}.

64 CTV belongs to the genus *Closterovirus*, under family *Closteroviridae*, characterized by
 65 long flexuous rod-like virus particles (~2,000x 12 nm) composed of two types of coat
 66 protein (major & minor coat proteins), encapsulating a single stranded, monopartite,
 67 linear, positive sense genomic RNA of approximately 19.3 kilobases^{2,7,8}. The CTV gRNA
 68 consists of at least 12 open reading frames flanked by untranslated regions (UTRs) at both
 69 the 5' and 3' ends². The ORF1a segment of the ORF1 encodes a large polyprotein having
 70 papain-like protease (P-Pro), Methyltransferase (Mt), and Helicase (Hel) activities, while
 71 the ORF1b encodes RNA dependent RNA polymerase (RdRp), expressed through +1
 72 frameshift mechanism^{2,6,8}. ORFs 2–11 encode at least ten distinct proteins, namely p33,
 73 p6, HSP70h, p61, CPm, CP, p18, p13, p20 and p23, expressed through subgenomic
 74 RNAs^{2,6,8}. These structural and non-structural proteins are involved in capsid formation,
 75 virion assembly, suppression of plant RNA silencing mechanism, host range determination
 76 etc^{2,6}.

77 As per available ICTV taxonomy report, three distinct strains of CTV have been identified,
 78 namely the T30 (mild isolate), T36 (intermediate severity) and VT (severe isolate)⁸.
 79 However, availability of complete genome sequences and detailed phylogenetic analyses
 80 has led to identification of at least eight different CTV genotypes distributed throughout
 81 the globe, namely the T36, T30, T3, VT, B165, HA16-5, T68 and the resistance breaking
 82 RB⁹. Genotype of infecting CTV is considered as an important factor that determines
 83 magnitude of damage^{2,3}.

84 Citrus species are among the economically most important horticulture crops and are
85 cultivated throughout the tropical and sub-tropical geographical zones of India^{7,10}. Among
86 them, mandarins (*Citrus reticulata* Blanco) particularly the *Nagpur mandarin* (Central
87 India), *Kinnow mandarin* (Northwestern India), *Coorg mandarin* (Southern India),
88 *Darjeeling mandarin*, *Sikkim mandarin* or *Khasi mandarin* (Northeastern India, NE India)
89 and the *Arunachal Orange/ Wakro Orange/ Arunachal Wakro Orange* (NE India) are
90 economically important cultivars. Among others, sub-tropical agro-climatic conditions of
91 eastern Himalayan regions (including Darjeeling hills, Sikkim, Arunachal Pradesh) of NE
92 India is known to be optimal for growing quality citrus fruits^{11,12}. Similar to other citrus-
93 growing regions, Indian citriculture is also severely threatened by CTV. CTV has killed
94 estimated one million citrus trees in India during the last century¹³. Percentage of disease
95 incidence varies widely throughout India^{7,13-15}. Some NE Indian orchards in Assam and
96 Meghalaya are reported to have even 100% incidence of CTV infection¹⁶. A number of
97 previous studies have reported prevalence of genotype VT or its recombinants, with
98 extensive diversity of coat protein (CP) gene in different citrus growing regions of India¹⁷⁻
99 ²⁰.

100 Among the NE Indian State of Arunachal Pradesh (AP), citrus fruits represent the largest
101 horticulture crop, of which the *Arunachal mandarin* or the *Arunachal Wakro orange* or
102 *Wakro orange* (named after place of its origin – *Wakro*, 28.113°N and 96.812°E, Lohit
103 district) accounts for 90% of total citrus production. *Wakro orange* is largely organic and
104 has certain unique features that make it superior from other mandarin oranges grown in
105 other parts of India. High vitamin-C and total soluble solid (TSS) content, medium acidity
106 sweet-sour taste, loose and easily removable medium thick peel, good size, better juice
107 content and an attractive orange colour of this variety is attributed to characteristic agro-
108 climatic conditions of AP²¹. Owing to its distinctive characteristics, *Wakro Orange* has
109 been registered under the *Geographical Indications of Goods (Registration and*
110 *Protection) Act, 1999* by the Government of India²¹.

111 Unfortunately, during the last few years, *Wakro orange* has been witnessing a sharp
112 decline in production, severely affecting the economy of orange growers as well as of the
113 state²². A recent survey conducted by the Arunachal Pradesh Horticulture Research and
114 Development Mission (APHRDM) found this “*Citrus Decline*” situation highly alarming,
115 where farmers were even forced to abandon hundreds of hectares of orchards, completely
116 engulfed by the rapidly spreading decline disease^{23,24}. After survey, APHRDM officials

117 suspected bacterial and/or viral infection, nutrient deficiency etc²³. Despite presence of
118 diverse aphid species including *T. aurantii* and *T. citricida*²¹ and indication of CTV
119 infection in several quick declining citrus orchards²⁵, molecular epidemiology of CTV in
120 this part of India has not been studied systematically. This dearth of information may be
121 largely attributable to location of *Wakro Orange* orchards in tough terrains in some of the
122 remotest parts of India.

123 Our present study was carried out with an aim to detect, identify and characterize CTV
124 associated with rapid decline of Wakro oranges in this region.

125

126 **MATERIALS AND METHODS**

127 *Sample collection*

128 During March 2015, a survey of severely declining orange orchards in the Wakro circle
129 (also known as the ‘orange bowl of the State’) of the Lohit district in Arunachal Pradesh,
130 India was conducted, based on request of district horticultural department. Sampling was
131 done from affected orchards at four sites namely *Karhe* (27°54.20’N 96°17.38’E; 365 m
132 ASL), *Cibi* (27°44.77’N 96°20.24’E; 345 m ASL), *Kangjang* (27°48.15’N 96°20.78’E;
133 460 m ASL), *Mawai* (27°48.15’N 96°20.75’E; 330 m ASL) and from one healthy
134 Government maintained healthy orchard at *Mawai* (27°49.66’N 96°19.91’E; 326 m ASL).
135 Location of the study site, and some glimpses of the affected orchards are shown in [Figure](#)
136 [1](#). Leaf samples along with aphids were collected from affected plants, immediately
137 preserved in RNAlater (Sigma-Aldrich, St. Louis, USA) and transported to laboratory for
138 further investigations.

139 *RNA extraction and detection of CTV RNA*

140 Approximately 1 mg of individual leaf sample and aphid pools (each composed of 10-15
141 aphids) were ground in ball mill (Retsch GmbH, Düsseldorf, Germany) and total RNA was
142 extracted from homogenate using standard TRI-reagent protocol (Sigma-Aldrich).
143 Following assessment of quality and quantity (denaturing agarose gels and
144 spectrophotometry), presence of enzyme inhibitors in RNA extracts was assessed by
145 reverse transcription PCR (eAMV RT, Sigma-Aldrich) using universal primers for 18S
146 RNA. Appropriate control reactions were included during each reaction.

Initially, CTV was detected using RT-PCR targeting amplification of highly conserved 3'-UTR of the CTV genome²⁶. Briefly, first round of RT-PCR was done with external primer pairs PEX1/PEX2, followed by second round of amplification with internal primer pairs PIN1/PIN2, generating an amplicon of approximately 132 bps, which indicated presence of CTV genome. Specificity of the 132 bp amplicons was verified by restriction digestion with *Alu* I, which yielded two fragments of approximately 86 bp and 46 bp, as expected. Restriction digestion pattern of the amplicon was predicted using an alignment of 3'-UTR region from 70 CTV genomes belonging to various genetic types, available in the NCBI GenBank.

Amplification, cloning and sequencing of CTV major CP and RdRp genetic regions

For amplification of CP gene sequence, we used previously described PCR primers- HCP₁ and HCP₂, which amplify complete CP ORF of 672 bps²⁷. The ORF1 encoding complete RdRp ORF (RNA dependent RNA polymerase) was amplified using two pairs of overlapping primers, namely SY17/SY18 and SY19/SY20, described earlier²⁸.

Following amplification, PCR products were purified, cloned into vectors using T/A cloning kit (Fermentas, Villinius, Lithuania) and were used to transform XL1-Blue competent bacterial cells. For each amplicon 5-10 clones were randomly selected, cultured and plasmid DNA purified (Sigma-Aldrich). Subsequently, plasmid DNA was amplified using vector specific universal primers (M13F and M13R) and sequenced (AgriGenome Labs Pvt Ltd., Kochi, India).

Molecular evolutionary analysis

After trimming the vector sequences from sequence data, alignment, manual editing and assembly (for overlapping amplicons) were done using Bioedit version 7.2.5²⁹. Sequence alignments were subsequently analysed through *Elimdupes* program (<https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html>) to identify identical or near identical sequences among clonal sequences, based on inter-sequence percent nucleotide identity (PNI). From each group of identical or near identical sequences (PNI ≥ 98), one representative sequence was selected for further analyses and submission to the GenBank. Integrity of protein coding nucleotide sequences was verified through *ORFfinder* program (www.ncbi.nlm.nih.gov/orffinder/). Sequences were then BLAST (BLASTn, Megablast) searched³⁰ to recognize and retrieve homologous sequences

178 available in the GenBank non-redundant (nr) database. Additionally, well-characterized
179 CTV sequences enlisted in the ICTV species list⁸ and reviewed in recent literature^{2,7} were
180 also retrieved from GenBank for further analyses.

181 Sequences were aligned using MUSCLE algorithm³¹ followed by molecular evolutionary
182 analyses using MEGA version 6³². Best fitting nucleotide substitution model was
183 determined from lowest Bayesian Information Criterion (BIC) score. Evolutionary
184 distances were calculated using Maximum Composite Likelihood method and
185 phylogenetic relatedness was determined employing Neighbour-Joining (NJ) algorithm. To
186 keep the phylogenetic tree uncluttered, among similar or nearly similar GenBank
187 sequences, identical or near identical sequences arising from same site/time/study were
188 kept to minimum. Consistencies of tree branches were computed by 1000 bootstrap
189 replications of the datasets.

190 To avert ambiguities in sequence relatedness, which arise from recombination events, we
191 reconstructed Neighbour Network (NeighborNet or NN), using *SplitsTree* version 4.14.1³³,
192 with Kimura-2 parameter distance correction algorithm. Data gaps and parsimony
193 uninformative sites were excluded from the calculations. NN reconstruction algorithms
194 accommodate recombination events and represent sequence relatedness in more accurate
195 topologies³⁴.

196 To examine evidence of recombination events in present sequences, aligned datasets were
197 analysed through different recombination detection algorithms, namely RDP,
198 GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ, incorporated
199 in RDP version 4.72³⁵. Computations were done on default settings with standard
200 *Bonferroni* correction. Recombination events with $P < 0.01$ were further verified by
201 analysis of breakpoint plots and phylogenetic trees (Unweighted Pair Group Method with
202 arithmetic Mean; UPGMA) generated by the program.

203

204 **RESULTS**

205 ***Detection, amplification, cloning and sequencing of CTV gRNA from Citrus plants***

206 Initially, leaf RNA extracts from five different orchards (4 affected and 1 healthy) and
207 RNA extracts from four aphid pools (collected from affected orchards) were subjected to
208 3'-UTR RT-PCR. Amplicons of expected size were repeatedly detectable in leaf RNA

209 extracts from affected orchards, but not in extracts from healthy orchard (Figure 2A).
 210 Additionally, amplicons were also detected in three of four aphid pools, collected from
 211 different declining orchards. Specificity of the 132 bp amplicons was confirmed by
 212 digestion with restriction endonuclease *AluI*, which yielded two fragments of 86 bp and 46
 213 bp (Figure 2B), and also by direct sequencing of the 132 bp amplicons. Complete CP and
 214 RdRp coding genetic regions could be reproducibly amplified in three of four leaf RNA
 215 extracts that were positive by 3'-UTR RT-PCR. Nevertheless, despite repeated attempts,
 216 we could not amplify these genetic regions in one of the 3'UTR RT-PCR positive leaf
 217 RNA extracts or in any of the aphid pool extracts.

218 Subsequent to clonal-sequencing of the CP and RdRp amplicons, alignment of
 219 homologous sequences and screening through *Elimdupes* program, one representative
 220 sequence from each group of identical clonal-sequences was selected for further analyses.
 221 Finally, three representative CP gene sequences (S21, S24 & S28) and two representative
 222 RdRp gene sequences (S5 & S9) were analyzed in details.

223 Nucleotide sequences generated in this study are available in the NCBI GenBank under
 224 accession numbers - KY882459 through KY882476.

225 ***Diversity of the sequences and results of BLAST analysis***

226 Among the complete CP gene sequences, S21 and S24 were found to be virtually identical,
 227 sharing a PNI of 99% among themselves. While sequence S28 was significantly divergent
 228 from S21 and S24, sharing a PNI score of only 93% (S28 vs. S21 & S24, $P < 0.01$). On the
 229 other hand, the two representative RdRp sequences S5 and S9 shared a PNI of 98% among
 230 themselves.

231 In BLASTn search, CP sequences S21 and S24 were found to be highly homologous (99%
 232 identity at 100% query coverage) to a number of sequences isolated from different parts of
 233 India as well as other countries, viz. eastern and NE India (HM573451, GQ392063,
 234 GQ475551- GQ475553, LT576375), western India (GQ475563, GQ475564), Southern
 235 India (EU869297), Northern India (JN974902, FJ001829), Uruguay (KU900351),
 236 Argentina (EU579381), South Africa (KU883267), Reunion Islands (AY660010) and
 237 Malaysia (HQ012413). Conversely, sequence S28 was found to be highly homologous
 238 (99% identity at 100% query coverage) to sequences recently isolated from different parts

239 of NE India (GQ475549, KR185332; KR185331, KR185329, KR105771; KR259640,
240 KR259641, KR080487, LN997804) as well as from New Zealand (AY896563).

241 On the other hand, in BLASTn analysis, RdRp sequences- S5 and S9 were found to be
242 highly similar (97-98% identity at 100% query coverage) to CTV genotype VT sequences
243 from USA (KC517493, EU937519, KC517494; KC517492; KU361339), Greece
244 (KC262793), China (JQ911664, KU720382, JQ061137), Spain (DQ151548), Italy
245 (KJ790175, KC748392), Japan (AB046398) and Israel (U56902). Interestingly, present
246 sequences were reasonably divergent from two distinct complete genome CTV sequences
247 isolated from India, namely the Kpg3 from Eastern India (HM573451, 94% identity at
248 100% query coverage) and the B-165 from southern India (EU076703; 86% identity at
249 100% query coverage).

250 *Phylogenetic affinities of the CP and RdRp genetic regions*

251 In the CP gene phylogenetic tree, sequences from various parts of India and other countries
252 segregated into eight distinct clades previously classified as genogroups I through VIII³⁶.
253 Clustering of present sequences S21 and S24 corroborated with BLASTn results,
254 demonstrating their affinity with group VII of CP gene sequences (Figure 3). Conversely,
255 sequence S28 along with two other sequences (AY896563 from New Zealand &
256 KR259640 from Sikkim, NE India) acquired an intermediate position between group I
257 sequence clade (which comprise sequences mostly isolated from NE India) and group VIII
258 sequence (KC590504 from Tirupati, south India) (Figure 3). This intermediate position of
259 sequence S28 was more prominent in NN analysis (Figure 4). Although clustering pattern
260 of CP gene sequence in NN analysis was largely analogous to NJ phylogenetic tree,
261 extensive networking within and among various CP gene groups was evident in NN
262 analysis (Figure 4).

263 On the other hand, RdRp sequences corresponding to S5 and S9 phylogenetically
264 clustered with well-characterized CTV genotype VT ‘western’ lineage sequences from
265 different countries, while maintaining considerable distance from the unique VT sequence
266 isolate (Kpg3; HM573451) from another part of NE India (Figure 5). These results largely
267 substantiated with BLASTn results. Notably, NN reconstructed with RdRp sequences was
268 substantially simpler, having scarcer evidences of inter and intragenotype networks (Figure
269 6), sharply contrasting the CP gene NN (Figure 4).

270 ***Analyses of recombination***

271 Among the two RdRp sequences generated in this study, sequence S5 had evidence of
 272 recombination (detected by Bootscan, MaxChi, SiScan and 3Seq with a probability of
 273 1.346×10^{-02} , 3.033×10^{-06} , 2.166×10^{-06} and 1.510×10^{-02} , respectively) between the present
 274 sequence S9 (minor parent) and a sequence resembling an isolate from China (major
 275 parent JQ061137, belonging to the 'Asian' lineage of genotype VT) (Figure 7a).
 276 Interestingly, evidence of sequence S9 as the major parent was detected in all the four
 277 reference sequences belonging to genotype T68 (EU076703, India; FJ525436, New
 278 Zealand; JQ965169, USA and KC333868, South Africa), while as minor parent in both the
 279 reference sequences belonging to genotype T3 (EU857538, New Zealand and KC525952,
 280 USA) included in the present analysis.

281 On the other hand, analysing the CP gene sequence dataset, RDP program could not detect
 282 any evidence of recombination within sequences S21 and S24. However, isolate S28 was
 283 found to have a sequence fragment which had evidence of recombination (detected by
 284 SiScan and 3Seq with a probability of 1.243×10^{-10} and 1.336×10^{-02} , respectively) between
 285 a sequence isolated from Reunion Island (major parent, AY660010, belonging to group
 286 VII) and a sequence (isolate AG28) from Tinsukia, NE India (minor parent, KC590498,
 287 belonging to group I) (Figure 7b). Nevertheless, considering the remote origin of sequence
 288 AY660010, and its intimate phylogenetic relatedness with sequence S21 (Figure 3), it is
 289 logical to interpret that an immediate ancestor (probably originating in NE India) of these
 290 two sequences might have actually been the major parent.

291 **DISCUSSION**

292 The NE Indian biogeographic zone is an important part of the Indo-Burma biodiversity
 293 hotspot region, where the great Himalayan Mountains meets the Peninsular India. NE
 294 India comes under the *Hindustan (Indian) centre of origin of cultivated plant*, proposed by
 295 Nikolai Vavilov, and is considered the native place for numerous tropical and sub-tropical
 296 fruits including citrus^{37,38}. It is also proposed that CTV has coevolved with its citrus
 297 hosts^{2,38}. CTV is now ubiquitous and is the most devastating pathogen of citrus². Similar to
 298 other citrus-growing regions of the world, CTV is highly prevalent in sub Himalayan
 299 regions of NE India, where it has severely affected and wiped-out several orchards^{7,39}.
 300 Nevertheless, our understanding of CTV epidemiology in NE India is mostly restricted to
 301 accessible orchards in States of Assam, Meghalaya and Sikkim. To the best of our

302 knowledge, there is no reliable literature on molecular epidemiology of CTV in the state of
303 Arunachal Pradesh and this study is the first to report CTV molecular epidemiology in one
304 of the oldest and remotest citrus-growing regions of India.

305 Previous studies have demonstrated an evolutionary asymmetry between the 5' and 3'
306 halves of CTV genome, where the 5' (including the ORF1) is highly conserved and
307 encodes genotype-specific phylogenetic signals while the 3' (including the CP) shows
308 relatively high genetic diversity^{2,40,41}. Based on these observations, a ~400 nucleotide
309 region of ORF1a (corresponding to partial P-Pro domain) has been widely used for
310 determining CTV genotypes^{7,47}. In the present study, we amplified and analysed a
311 comparatively longer genetic fragment (1860 bases corresponding to partial ORF1a &
312 ORF1b coding regions) and found that phylogeny reconstructed with this genetic region is
313 highly concordant with phylogenies derived from CTV complete genome sequences.
314 Nevertheless, analyses of this genetic region in the present isolates demonstrated the
315 presence of virulent genotype VT, substantiating with previous studies from the Indian
316 subcontinent^{7,14-20,36,42}. Interestingly however, present isolates clustered with the 'western'
317 lineage clade of VT sequences in the phylogeny. Since Wakro region geo-physically
318 connects NE Indian part of Indian subcontinent with China and other parts of Southeast
319 Asia, phylogenetic affiliation of present sequences to either 'Indian' or 'Asian' lineage was
320 therefore highly anticipated. The *Western* and the *Asian* lineages signify evolution and
321 global dispersal of two distinct variants of genotype VT in Western and Asian countries
322 respectively², the *Indian* lineage has been proposed to accommodate the decline inducing
323 recombinant strain 'Kpg3' (GenBank accession HM573451), which do not cluster within
324 either of the two lineages in phylogenies^{2,7}.

325 Among extant CTV genotypes distributed globally, highest genetic diversity and widest
326 geographical distribution of VT advocates its ancient origin. Analyses based on mixed
327 effects model of evolution (MEME) of the *ORF1a*, further strengthens earliest origin of
328 this genotype, providing additional cue that *western* lineage of VT is the most ancestral
329 type known so far, from which all other lineages, including the *Asian* lineage have evolved
330 through 'episodic diversification events'². Considering this geographical region as the
331 proposed centre of origin of CTV, *Western* lineage CTV sequences found in present study
332 appears to represent the original 'relict' CTV genotype. In such a scenario, natural spatio-
333 temporal dispersal of this original virus to southeast Asia and peninsular India might have
334 resulted in evolution of *Asian* and *Indian* lineages, respectively. In contrast, phylogenetic

335 clustering of present sequences with genotype VT sequences from western countries amply
336 support its relatively recent introduction to the western countries. Earlier, it has been
337 proposed that CTV evolved and diverged into major clades in this region followed by their
338 dispersal across Southeast Asia, Indian subcontinent and islands of Indian Ocean long ago,
339 while their dissemination to Europe, Africa and Americas took place only recently (~400-
340 500 years ago)^{3,6,44}.

341 Additionally, owing to stability of CTV genomes (evolutionary rates being the slowest
342 among RNA viruses), similarity of genome sequences in distant geographical regions have
343 been suggested to indicate origin and dispersal of CTV^{3,6,44}. Interestingly, one of the ORF1
344 sequences (S5) in this study was found to be recombinant, involving a major parent similar
345 to *Asian* lineage genotype VT sequence (JQ061137, China) and present isolate S9
346 sequence as minor parent. This finding of an *Asian* lineage sequence as major recombinant
347 partner may either be suggestive of remnants of the '*Asian*' lineage evolved in this
348 geographical region or might be a relatively recent introduction, which cannot be
349 concluded from the present results. Notably, in recombination analysis, present sequence
350 S9 was found to be involved in all the genotype T68 and genotype T3 reference sequence
351 included in the present study.

352 On the other hand, previous studies on CP gene sequences have revealed circulation of
353 diverse phylogroups in different citrus-growing regions, which do not necessarily correlate
354 with the genotype determined by analysis of ORF1 sequences^{17,38,45}. Such incongruence
355 between different genetic regions of CTV genome has been attributed to frequent
356 recombination events or gene flow among distinct genetic types occurring particularly in
357 the 3' half of the CTV genome^{7,18,19,38,45}. Our findings of NN analyses of ORF1 and CP
358 genetic regions showing sharply discrete pattern of reticulate evolution strongly support
359 these previous observations.

360 Nevertheless, analysing CP gene variability, previous studies have reported eight globally
361 circulating CP genetic variants (genogroups I to VIII), of which seven (I, III, IV, V, VI,
362 VII, and VIII) are present throughout Indian subcontinent while five (I, III, V, VI and VII)
363 are present in different parts of NE India^{7,36,43}. In contrast, CP gene sequence diversity
364 observed in present samples was limited to only two genogroups (I & VII), which were
365 related to sequences from other parts of sub Himalayan NE India (Darjeeling, Kalimpong
366 and Sikkim). We also detected a recombinant CP sequence involving a group VII sequence

(resembling present sequence S21) accepting a large fragment from a group I sequence (similar to isolate AG28, GenBank KC590498) reported from Tinsukia, Assam^{7,39}. However, we could not identify any genogroup VI CP gene sequence, which represents a clade of phylogenetically related sequences exclusively from different parts of NE India. Taken together, limited genetic diversity, evidence of recombination with sequences only from nearby geographical regions and remnants of the present sequences in other genotypes indicate the antiquity of present CTV sequences from Wakro and provides credible evidence to support prevailing hypotheses on origin of CTV.

Apart from CTV genotype, transmitting vector aphid species can also significantly influence the spatio-temporal dynamics and magnitude of infection. Although, BrCA (*T. citricida*) is the most frequent and efficient vectors associated with CTV transmission in India^{13,15}, other aphid vectors (*T. aurantii*, *A. gossypii*, *Myzus persicae*, *Dactynotus jaceae*, *A. craccivora*) have also been recognized to transmit CTV⁴⁶. During present survey, we could detect *T. aurantii* in high densities in all the declining orchards (situated at an altitude of 330-426 m). These findings are different from a previous study from Darjeeling, where *T. aurantii* was dominant in the citrus orchards at high altitude (>500 m), whereas *T. citricida* was predominant in the orchards of lower altitude⁴⁷. *T. aurantii* has earlier been shown to be highly adaptable to wider host range and higher altitudes, particularly in the sub-tropical hills of NE India^{48,49}. Remarkably, we could reproducibly amplify CTV specific 132 bp 3'-UTR genetic region in aphid pools, we could not amplify other genetic fragments (CP or the ORF1) from the aphid pools to implicate their role in transmission. Even though exactly similar observation of amplification of smaller viral RNA segments (130-162 nts) while non-amplification of larger segments⁵⁰ have been reported from mosquito pools, reason for such phenomenon remains to be investigated.

In conclusion, present study reveals the molecular epidemiology of CTV from a place which comes under a geographical region, believed to be the origin of the citrus fruits. Due to its remoteness, this region has remained largely unreached and thus maintains pristine conditions in terms of biodiversity. Our results suggesting the circulation of an unanticipated genetic type of CTV with less interaction with CTV sequences only from neighbouring geographical areas indicate relict nature of present CTV sequences. However, we acknowledge that complete genome sequence would be required for better understanding the evolution and diversification of CTV.

399 ACKNOWLEDGEMENT

400 This research was supported by institutional intramural grants from the Defence Research
401 & Development Organization (DRDO), Ministry of Defence, Government of India. The
402 authors gracefully thank Dr. BM Mishra, Ex-Deputy Commissioner (Lohit District),
403 Government of Arunachal Pradesh, officials of the Horticulture department (AP) and the
404 orange growers of Wakro for their support and co-operation during field survey.

405

406 AUTHOR CONTRIBUTION

407 S.D. and V.V designed the study. S.D., R.G., V.M., M.K.M & S.R. performed field
408 surveys, collected samples and performed identifications. S.D., B.D., R.B. performed
409 laboratory experiments. S.D. analysed data, wrote the main manuscript, prepared the
410 figures and tables. S.K.D. & V.V. overall supervised the research and critically reviewed
411 the manuscript. All the authors except R.B. (deceased) reviewed the manuscript.

412

413 REFERENCES

- 414 1. FAO. Citrus fruit fresh and processed statistical bulletin.
415 <http://www.fao.org/economic/est/est-commodities/citrus-fruit/en/> (2016).
- 416 2. Harper, S. J. Citrus tristeza virus: evolution of complex and varied genotypic
417 groups. *Front. Microbiol.* **4**, 93 (2013).
- 418 3. Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J. & Peña, L. Citrus tristeza
419 virus: a pathogen that changed the course of the citrus industry. *Mol. Plant.*
420 *Pathol.* **9**: 251-268 (2008).
- 421 4. Bar-Joseph M., Dawson, W. O. Citrus tristeza virus (eds. Mahy, B.W.J. & Van
422 Regenmortel, M.H.V.) *In* Encyclopaedia Virology 3rd edition 1, 520–525
423 (Academic Press, 2008).
- 424 5. Lee, R.F. & Bar-Joseph, M. Tristeza. *In* Compendium of citrus diseases, 2nd
425 edition
426 61–63 (APS Press, 2000).
- 427 6. Albiach-Martí, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E.,
428 Shanker, S., et al. Sequences of citrus tristeza virus separated in time and space are
429 essentially identical. *J. Virol.* **74**, 6856-6865 (2000).

- 430 7. Biswas, K.K., Palchoudhury, S. & Ghosh, D.K. Closterovirus in India:
431 Distribution, Genomics and Genetic Diversity of Citrus Tristeza Virus (eds.
432 Mandal, B., et al.). In A Century of Plant Virology in India (Springer Nature,
433 2017).
- 434 8. Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M., Boscia, D., Candresse, T.,
435 Coutts R. H. A., et al. Family *Closteroviridae*. In Ninth Report of the International
436 Committee on Taxonomy of Viruses, Andrew, M.Q.K., Adams, M.J., Carstens,
437 E.B., Leftkowitz, E.J., pp. 987-1001. (Elsevier Academic Press, 2012).
- 438 9. Kleynhans, J., Pietersen, G. Comparison of multiple viral population
439 characterization methods on a candidate cross-protection Citrus tristeza virus
440 (CTV) source. *J. Virol. Methods*. **237**, 92-100 (2016).
- 441 10. ICAR- All India Coordinated Research Project on Fruits 2016-17.
442 <https://icar.org.in/dare-icar-annual-reports> (2017).
- 443 11. Barbhuiya, A.R., Khan, M.L. & Dayanandan, S. Genetic structure and diversity of
444 natural and domesticated populations of Citrus medica L. in the Eastern Himalayan
445 region of Northeast India. *Ecol. Evol.* **6**: 3898–3911 (2016).
- 446 12. Hynniewta, M., Malik, S.K. & Rao, S.R. Genetic diversity and phylogenetic
447 analysis of Citrus (L) from north-east India as revealed by meiosis, and molecular
448 analysis of internal transcribed spacer region of rDNA. *Meta. Gene.* **2**, 237–251
449 (2014).
- 450 13. Ahlawat, Y.S. Viruses, greening bacterium and viroids associated with citrus
451 (Citrus species) decline in India. *Indian J. Agric. Sci.* **67**, 51–57 (1997).
- 452 14. Biswas, K.K. Molecular characterization of Citrus tristeza virus isolates from the
453 Northeastern Himalayan region of India. *Arch. Virol.* **155**, 959–63 (2010).
- 454 15. Biswas, K.K. Molecular diagnosis of Citrus tristeza virus in mandarin (Citrus
455 reticulata) orchards of hills of West Bengal. *Indian J. Virol.* **19**: 26-31 (2008).
- 456 16. Biswas, K.K., Tarafdar, A., Sharma, S. K., Singh, J. K., Dwivedi, S., Biswas, K., *et*
457 *al.* Current status of Citrus tristeza virus incidence and its spatial distribution in
458 citrus growing geographical zones of India. *Indian J. Agri. Sci.* **84**: 8–13 (2014).
- 459 17. Biswas, K.K., Tarafdar, A., Dwivedi, S. & Lee, R. F. Distribution, genetic diversity
460 and recombination analysis of Citrus tristeza virus of India. *Virus Genes* **45**, 139–
461 48 (2012).

- 462 18. Biswas, K.K., Tarafdar, A., Sharma, S. K. Complete genome of mandarin decline
463 Citrus tristeza virus of Northeastern Himalayan hill region of India: comparative
464 analyses determine recombinant. *Arch. Virol.* **157**: 579–83 (2012).
- 465 19. Sharma, S.K., Tarafdar, A., Khatun, D., Kumari, S. & Biswas, K.K. Intrafarm
466 diversity and evidence of genetic recombination of Citrus tristeza virus isolates in
467 Delhi region of India. *J. Plant Biochem. Biotechnol.* **21**: 38–43 (2012).
- 468 20. Ghosh, D. K., Aghave, B., Roy, A. & Ahlawat, Y.S. Molecular cloning,
469 sequencing and phylogenetic analysis of coat protein gene of a biologically
470 distinct CTV isolate occurring in central India. *J. Plant Biochem. Biotechnol.* **18**,
471 105–108 (2009).
- 472 21. Geographical Indications Journal, volume 61, Geographical Indications Registry,
473 Government of India. November 21, 2014. [http://www.ipindia.nic.in/journal-](http://www.ipindia.nic.in/journal-gi.htm)
474 [gi.htm](http://www.ipindia.nic.in/journal-gi.htm)
- 475 22. Business Standard. Orange growth hit in Arunachal Pradesh. January 11, 2015.
476 [https://www.business-standard.com/article/pti-stories/orange-growth-hit-in-](https://www.business-standard.com/article/pti-stories/orange-growth-hit-in-arunachal-pradesh-115011100455_1.html)
477 [arunachal-pradesh-115011100455_1.html](https://www.business-standard.com/article/pti-stories/orange-growth-hit-in-arunachal-pradesh-115011100455_1.html)
- 478 23. Arunachal Govt portal; Citrus decline is alarming. Published on State Portal of
479 Arunachal Pradesh
480 [http://arunachalpradesh.nic.in/csp_ap_portal/pdf/Documents/citrus-decline-](http://arunachalpradesh.nic.in/csp_ap_portal/pdf/Documents/citrus-decline-alarming-arunachal.pdf)
481 [alarming-arunachal.pdf](http://arunachalpradesh.nic.in/csp_ap_portal/pdf/Documents/citrus-decline-alarming-arunachal.pdf)
- 482 24. The Assam Tribune; Citrus decline hits orange bowl of Arunachal. Guwahati,
483 Monday, January 19, 2015.
484 <http://www.assamtribune.com/scripts/mdetails.asp?id=jan1915/oth052>
- 485 25. ICAR-NEH Annual Report 2014-15. ICAR Research Complex for NEH Region
486 Umroi Road, Umiam, Meghalaya. www.icarneh.ernet.in and <http://kiran.nic.in>
- 487 26. Olmos, A., Cambra, M., Esteban, O., Gorris, M. T. & Terrada, E. New device and
488 method for capture, reverse transcription and nested PCR in a single closed-tube.
489 *Nucleic Acids Res.* **27**, 1564-1565 (1999).
- 490 27. Huang, Z., Rundell, P.A., Guan, X. & Powell, C.A. Detection and Isolate
491 Differentiation of Citrus tristeza virus in infected trees based on Reverse
492 Transcription-Polymerase Chain Reaction. *Plant Dis.* **88**, 625-629 (2004).
- 493 28. Vives, M.C., Rubio, L., Sambade, A., Mirkov, T.E., Moreno, P. & Guerri, J.
494 Evidence of multiple recombination events between two RNA sequence variants
495 within a Citrus tristeza virus isolate. *Virology.* **331**, 232–237 (2005).

496 29. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and
497 analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98
498 (1999).

499 30. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local
500 alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

501 31. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high
502 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

503 32. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. MEGA 6: molecular
504 evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729
505 (2013).

506 33. Huson, D.H. & Bryant, D. Application of Phylogenetic Networks in
507 Evolutionary Studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

508 34. Woolley, S. M., Posada, D. & Crandall, K.A. A comparison of phylogenetic
509 network methods using computer simulation. *PLoS One* **3**: e1913 (2008).

510 35. Martin, D.P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4:
511 Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**,
512 vev003 (2015).

513 36. Tarafdar, A., Godara, S., Dwivedi, S., Jayakumar, B.K. & Biswas, K.K.
514 Characterization of Citrus tristeza virus and determination of genetic variability in
515 North-east and South India et al. *Indian Phytopath.* 2013, **66**, 302–307 (2013).

516 37. Hummer, K.E. & Hancock, J.F. Vavilovian Centers of Plant Diversity:
517 Implications and Impacts. *Hortscience.* **50**, 780–783 (2015).

518 38. Martin, S., Sambade, A., Rubio, L., Vives, M. C., Moya, P., Guerri, P., Elena, S.
519 F. & Moreno, P. Contribution of recombination and selection to molecular
520 evolution of Citrus tristeza virus. *J. Gen. Virol.* **90**, 1527–1538 (2009).

521 39. Biswas, K.K., Pal Choudhuri, S. & Godara, S. Decline of Mandarin Orange Caused
522 by Citrus tristeza virus in Northeast India: Conventional and Biotechnological
523 Management Approaches. *J. Agri. Eng. Food Tech.* **3**, 236–241 (2016).

524 40. Hilf, M.E., Mavrodieva, V.A., Garnsey, S.M. Genetic marker analysis of a global
525 collection of isolates of Citrus tristeza virus: characterization and distribution of
526 CTV genotypes and association with symptoms. *Phytopathology.* **95**, 909–917
527 (2005).

- 528 41. Roy, A., Brlansky, R.H. Genome analysis of an orange stem pitting citrus tristeza
529 virus isolate reveals a novel recombinant genotype. *Virus Res.* **151**, 118–130
530 (2010).
- 531 42. Singh, J.K., Tarafdar, A., Sharma, S.K. & Biswas, K.K. Evidence of
532 Recombinant Citrus tristeza virus Isolate Occurring in Acid Lime cv. Pant Lemon
533 Orchard in Uttarakhand Terai Region of Northern Himalaya in India. *Indian J.*
534 *Virol.* **24**, 35–41 (2013).
- 535 43. Palchoudhury, S., Ghimiray, S., Biswas, M.K. & Biswas, K.K. Citrus tristeza virus
536 Variants and their Distribution in Mandarin Orchards in Northeastern Himalayan
537 Hill Region of India. *Int. J. Curr. Microbiol. App. Sci.* **6**: 1680–1690 (2017).
- 538 44. Silva, G., Marques, N. & Nolasco, G. The evolutionary rate of citrus tristeza virus
539 ranks among the rates of the slowest RNA viruses. *J. Gen. Virol.* **93**, 419–429
540 (2012).
- 541 45. Rubio, L., Ayllón, M. A., Kong, P., Fernández, A., Polek, M., Guerri, J., Moreno,
542 P. & Falk, B.W. Genetic variation of Citrus tristeza virus isolates from California
543 and Spain: evidence for mixed infections and recombination. *J. Virol.* **75**, 8054–
544 8062 (2001).
- 545 46. Verma, P.M., Rao, D.G. & Capoor, S.P. Transmission of tristeza virus by *Aphis*
546 *craccivora* Koch and *Dactynotus jaceae* (L.) *Indian J. Entomol.* **27**, 67–71 (1965).
- 547 47. Ghosh, A., Das, A., Lepcha, R., Majumdar, K. & Baranwal, B.K. Identification and
548 distribution of aphid vectors spreading Citrus tristeza virus in Darjeeling hills
549 and Dooars of India. *J. Asia Pacific Entomol.* **18**, 601–605 (2015).
- 550 48. Agarwala, B.K. & Bhattacharya, S. Anholocycly in tropical aphids: population
551 trends and influence of temperature on development, reproduction, and survival of
552 three aphid species (Homoptera: Aphidoidea). *Phytophaga* **6**, 17–27 (1994).
- 553 49. Roychoudhuri, D.N. Aphids of North-east India and Bhutan. pp. 76–78
554 (Zoological Society of India, 1980).
- 555 **50.** Ravanini, P., Huhtamo, E., Ilaria, V., Crobu, M.G., Nicosia, A.M., Servino, L., et
556 al. Japanese encephalitis virus RNA detected in *Culex pipiens* mosquitoes in Italy.
557 *Euro. Surveillance* **17**, 20221 (2012).

558

559

560 **FIGURE LEGENDS**

561 **Figure 1.** (a) Map showing the extreme northeast Indian state of Arunachal Pradesh (in
562 light green and indicated by an arrowhead). Neighbouring countries are denoted by alpha-2
563 country codes: MM, Myanmar; BD, Bangladesh; BT, Bhutan; NP, Nepal; PK, Pakistan.
564 (b) Map of Arunachal Pradesh showing the Lohit district (in orange colour). Panels (c) to
565 (i) show fruits from diseased trees, affected orchards, aphid colonies on leaves and affected
566 leaves with disease symptoms, observed during field survey. Blank base maps were
567 obtained from the *d-maps* website (<https://d-maps.com/index.php?lang=en>) and annotated
568 using Microsoft Office PowerPoint 2007 and Microsoft Windows Paint softwares.

569 **Figure 2.** (A) Agarose gel showing 132 bp amplicon from CTV 3'-UTR RT-PCR. Lanes
570 N, No Template Control (NTC); L1 to L5, RNA from leaf extracts; A1 to A3, RNA
571 extracts from aphid pools. Marker, M1 is 100 bp DNA ladder. (B) Agarose gel showing
572 *Alu* I digested products of 132 bp amplicons. Lanes U, uncut amplicon; L1 to L4,
573 amplicons from leaf extracts; A1 to A3, amplicons from aphid pools. Marker, M2 is
574 pUC19 DNA/MspI digested; M3 is 50 bp DNA ladder. Gel images were acquired on an
575 UVP BioSpectrum 410 imaging system (UVP, Cambridge, UK) with automatic exposure
576 setting. Colour of the original gel image was inverted using Microsoft Windows Paint and
577 annotated using Microsoft Office PowerPoint 2007.

578 **Figure 3.** Phylogenetic tree reconstructed for CP (coat protein) gene sequences generated
579 in the present study along with 81 other GenBank sequences, submitted by other
580 researchers. Sequences from the present study (S21, S24 & S28) are marked by black solid
581 squares. CP clades were classified as groups I to VIII following a previous literature
582 ([Tarafdar et al., 2013](#)). Evolutionary distances were calculated using the Maximum
583 Composite Likelihood (MCL) method and phylogenetic relatedness was recreated using
584 the Neighbor-Joining (NJ) method. Percentage of bootstrap support for each cluster
585 (>75%) is shown at nodes.

586 **Figure 4.** Neighbour Net of CP gene sequences reconstructed using the dataset previously
587 used for reconstruction of NJ tree in Figure 3. Position of the three CP gene sequence
588 generated in the present study (S21, S24 & S28) is indicated by red arrowheads.

589 **Figure 5.** Phylogenetic tree reconstructed from 1860 bases of 5' half of CTV genome
590 (corresponding to partial ORF1a & ORF1b coding regions) from two sequence generated

591 in this study (S5 & S9) and 52 others well characterized complete genomes retrieved from
592 the GenBank. Sequences from the present study are marked by black solid rounds.
593 Evolutionary distances were calculated using the Maximum Composite Likelihood (MCL)
594 method and phylogenetic relatedness was recreated using the Neighbor-Joining (NJ)
595 method. Percentage of bootstrap support for each cluster (>75%) is shown at the respective
596 node. Genotypes and lineages are mentioned on the right-hand side.

597 **Figure 6.** Neighbour Net reconstructed from 1860 bases of 5' half of CTV genome
598 (corresponding to partial ORF1a & ORF1b coding regions) using the dataset previously
599 used for reconstruction of NJ tree in Figure 5. Position of the two-sequences generated in
600 the present study (S5 & S9) is indicated by red arrowheads.

601 **Figure 7.** Recombination patterns observed in the sequences S5 (a) and S28 (b), generated
602 in the present study. Each figure is followed by BootScan plot showing the breakpoints
603 corresponding to the recombination event. Plots represent % of permuted trees on the y
604 axis.

605

Figure 1.

Figure 2.

611

612 **Figure 3.**

613

614 **Figure 4.**

615

616 **Figure 5.**

617

618

619 **Figure 6.**

620

621 **Figure 7.**