BRIEF REPORT

Genetic diversity and evolution of two capsid protein genes of citrus tristeza virus isolates from China

Guan-Wei Wu · Min Tang · Guo-Ping Wang ·
Feng-Yin Jin · Zuo-Kun Yang · Li-Jing Cheng ·
Ni Hong

Abstract The genetic diversity and population structure of citrus tristeza virus (CTV) isolates from China were investigated based on partial sequences spanning the C-terminal end of p61 and the complete sequences of the CPm and CP genes. Phylogenetic analysis revealed five known groups (RB, T30, T36, HA and VT) and one new group (VI) consisting of only Chinese CTV isolates. Incongruent phylogenetic trees coupled with recombination analysis suggested several recombination events in the CPm gene. Positive selection was detected at codon 9 of CPm and codons 31, 41 and 68 of CP. The widespread CTV subpopulation AT-1 found in China has a unique amino acid insertion at the C-terminus of p61, which could increase CTV population complexity with implications for the evolutionary history of the virus. Our results suggest relevant roles for gene flow, purifying selection and recombination in shaping the CTV population in China.

Citrus tristeza virus (CTV), a member of the genus *Closterovirus* in the family *Closteroviridae*, is the most

G.-W. Wu · G.-P. Wang · N. Hong (✉)
National Key Laboratory of Agromicrobiology, Huazhong Agricultural University, Wuhan 430070, Hubei, China
e-mail: whni@mail.hzau.edu.cn

G.-W. Wu · M. Tang · G.-P. Wang · F.-Y. Jin · Z.-K. Yang ·
L.-J. Cheng · N. Hong
Key Laboratory of Crop Disease Monitoring and Safety Control in Hubei Province, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China

economically important virus of citrus worldwide. This virus can cause quick decline of sour orange (*Citrus aurantium* L.) or lemon [*C. limon* (L.) Burn. f.] rootstocks, stem pitting, regardless of the rootstock, and seedling yellows on sour orange, grapefruit (*C. paradisi* Macf.) or lemon seedlings. CTV is naturally transmitted by aphids in a semi-persistent manner [17]. CTV virions have two coat proteins of 25 (CP) and 27 (CPm) kDa, encapsidating about 97 and 3 %, respectively, of a single-stranded, positive-sense genomic RNA [5].

Phylogenetic analysis of nucleotide sequences of the most variable genomic region of CTV has allowed subdivision of isolates into seven genotypes: VT, T3, T30, T36, B165/T68, RB (trifoliate resistance breaking), and HA [6–9, 16, 21]. Although it has been determined that CTV populations consist of highly divergent molecular variants and recombinants derived from divergent variants [22, 27], the driving forces for CTV evolution are poorly understood. Previous studies indicated that some genomic regions of CTV isolates from different geographic origins were under negative selection, and showed frequent recombination [14]. Recently, an analysis of CP sequences from global CTV isolates showed moderate genetic variability and that the same recombination events could be detected in the CP of isolates from different geographic regions [23]. The genetic variability of the CPm gene has rarely been analyzed. There is only one report on the presence of two dominant groups (A1 and T30/R1) in Colombian field isolates based on CPm gene sequence analysis [18]. The CP and CPm proteins have multiple functions for CTV survival in hosts. The two coat proteins, combined with the heat shock protein homolog (HSP70h) and p61, are needed for virion assembly [27, 28], and CPm is possibly involved in aphid transmission [4, 5]. Therefore,

it is important to understand the forces driving evolution of these two genes.

China represents an old country for citrus production. It has been suggested that CTV could have originated from China [1]. Previous studies have shown a complex CTV population in China [10, 28, 29]. Novel severe strains, including RB and B165/T68-1, have been identified in recent years [6, 7, 21], highlighting the need to enhance our understanding of the population structure of CTV in China. In this study, we collected CTV isolates from four citrus-producing provinces in China and analyzed the genetic diversity and population structure of these isolates by sequence analysis of the two capsid genes (CPm and CP).

A total of 79 citrus samples were collected from four provinces (Hubei, Hunan, Jiangxi and Sichuan) in southern and western China. Most of the sampled citrus trees were asymptomatic, except for a few plants that showed reduced growth. Isolate AT-1 was obtained from a field isolate containing different variants by aphid transmission. Five additional isolates (S4, N4, S45, S39 and S36), maintained in an insect-proof greenhouse at the National Indoor Conservation Center of Virus-Free Germplasm of Fruit Crops (Wuhan, Hubei, China) [10, 28], were also included in this study.

CTV detection was carried out by RT-PCR using the T36-CP primer set [11]. The primers CPm-F 5′-CGTCA-TATGAGCAGAGACGTG-3′ and CP-R 5′-TCAACGTGTGTTGAATTTCC-3′ [2, 11] were used for the amplification of an approximate 1540-bp fragment spanning the 3′ terminal region of the p61 gene (about 53 bp), the complete CPm (723 bp) and CP (672 bp) genes, and their internal non-coding region (about 92 bp).

Total RNA was extracted from leaf and phloem tissues using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. First-strand cDNA was synthesized using M-MLV reverse transcriptase (Promega, Madison, WI, USA) and the random primer hexadeoxyribonucleotide mixture pd(N)$_6$ (TaKaRa, Dalian, China) at 37 °C for 1 h. PCR reactions were carried out in a Mastercycler$^®$ gradient Thermocycler (Eppendorf, Hamburg, Germany) with the following conditions: one cycle at 94 °C for 3 min, then 35 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 95 s, with a final extension for 10 min at 72 °C. The products were gel-purified and cloned into the vector pMD18-T (TaKaRa, Dalian, China). At least three positive clones of each product were sequenced by a commercial service (GenScript Biological Science, Nanjing, China). All sequences obtained were edited to remove primer sequences.

Multiple nucleotide and amino acid sequence alignments were performed using the MUSCLE algorithm implemented in the program MEGA 5.2.2 [25]. Aligned sequences for CPm and CP were scanned using seven recombination detection programs (RDP, BOOTSCAN, SISCAN, SEQ, GENECONV, LARD and MAXCHI) implemented in the programs RDP 3.44 [13] and Recco [15] to detect putative recombination events. Recombinant sequences were subsequently excluded from phylogenetic analyses. Phylogenetic trees were inferred by the maximum-likelihood (ML) method implemented in MEGA 5.2.2, with 1000 bootstrap replicates and the general time-reversible model of nucleotide substitution and gamma distribution of heterogeneity. The genetic distances of intra- and inter- clusters were computed using MEGA 5.2.2 with 1000 bootstrap replicates. The fixed-effects likelihood (FEL), internal fixed-effects likelihood (IFEL) and random-effects likelihood (REL) methods available from the DATAMONKEY server (http://www.data monkey.org) were used for the identification of positively and negatively selected sites in the CPm and CP genes. Only concordant results using all three methods were considered.

A total of 32 citrus samples were positive for CTV by RT-PCR using the T36-CP primer set, and amplicons with the expected size of about 1540-bp from 26 individual samples were sequenced and submitted to GenBank with accession numbers KF144718 to 144771 (Supplementary Table 1). CTV from each sample was considered as an individual isolate. In total, 81 sequences were obtained from these isolates. Most of the isolates from Sichuan province comprised at least two different variants, while all four isolates from Jiangxi province and three isolates (N4, S4 and AT-1) from Hubei province had only one variant. Thus, 38 consensus sequences of clones with over 99 % identity within each isolate were generated and used for analysis.

These sequences were first scanned for putative recombination events in the CPm and CP genes. One recombination event involving the CP sequence of isolate SC-WZ-9 was identified (Suppl. Fig. S1A), and no recombination event was detected in CPm sequences. To avoid different alignments with contrasting evolutionary histories produced by recombination [19], the sequence of SC-WZ-9 was excluded from subsequent phylogenetic and population genetic analysis. Then, 58 CP sequences (37 obtained in this study and 21 retrieved from GenBank), and 60 CPm sequences (37 obtained in this study and 23 from GenBank) were used for analysis of phylogenetic relationships. The CP sequences of isolates A1 and R1 from Colombia were unavailable from GenBank, and thus were absent in the phylogenetic analysis of the CP gene.

The overall mean distance was 0.071 ± 0.006 and 0.095 ± 0.009 for the CP and CPm gene, respectively. These values are much lower than those previously observed for other seven genes (p33, HSP70h, p61, p18, p13, p20 and p23) in the 3′ region of the viral genome [14].
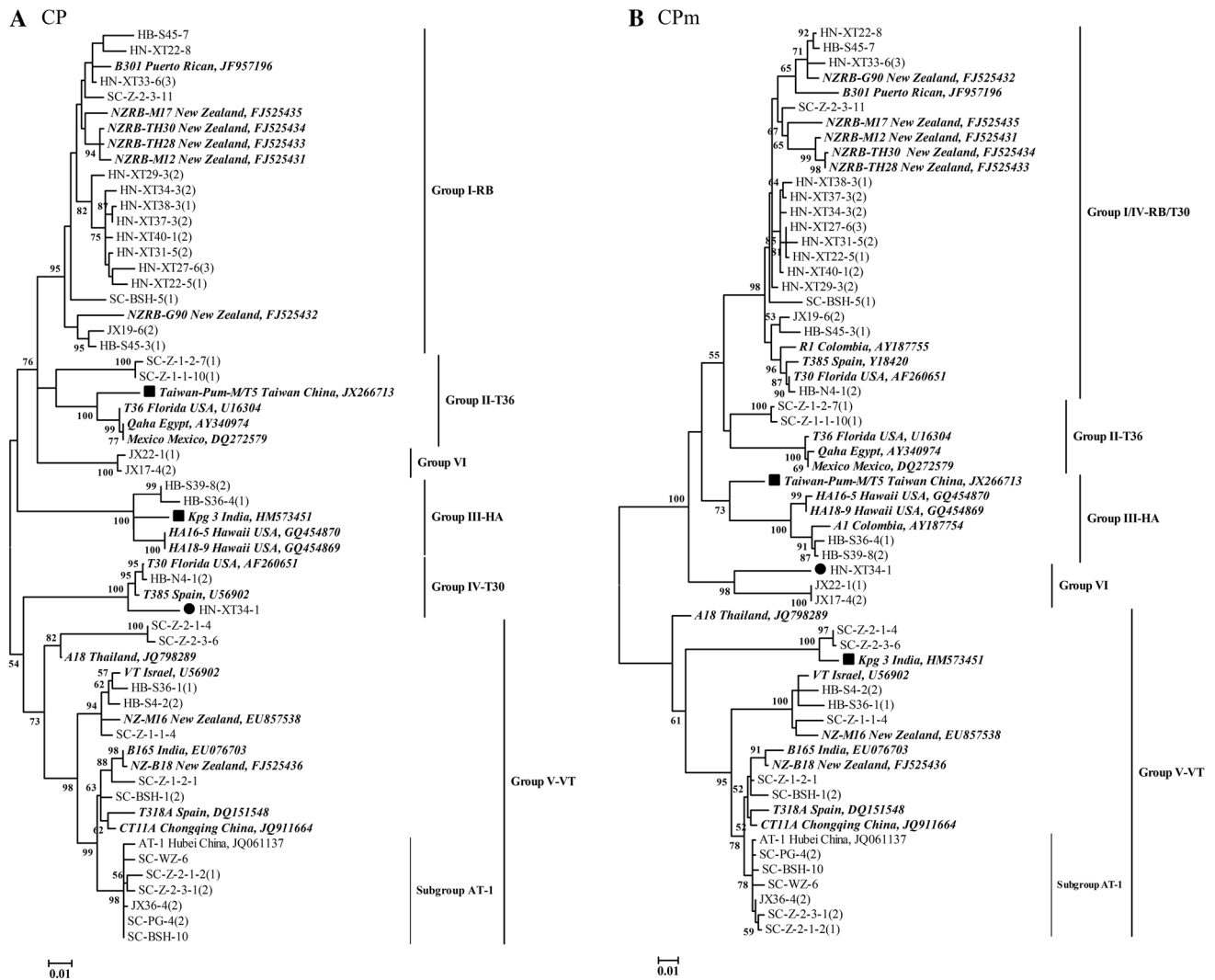
Fig. 1 Unrooted maximum-likelihood phylogenetic trees based on nucleotide sequences of CP (A) and CPm (B) of Chinese and representative global isolates of citrus tristeza virus. Bootstrap values (1000 replicates) are given at the branch nodes, and the values greater than 50 % are shown. Roman numerals I to VI represent individual groups. Names, geographical origins and GenBank accession numbers of reported isolates are indicated in bold and italic. The sequences from the same isolate having ≥99 % similarity are represented by one sequence, and their numbers are indicated in brackets. Chinese isolates obtained in this study are indicated by their province of origin (HB, Hubei; HN, Hunan; JX, Jiangxi; SC, Sichuan) followed by the isolate name and a clone number. Branch lengths are proportional to the genetic distance. The CP and CPm sequences of HN-XT-34-1, Taiwan-Pum-M/T5 and Kpg3, clustering in different groups in two phylogenetic trees, are indicated by a black circle and a black square, respectively

Phylogenetic analysis based on CP nucleotide sequences revealed six major groups. Based on their placements related to reference sequences of CTV isolates [7], these groups were designated as I (RB), II (T36), III (HA), IV (T30), V (VT), and VI, respectively (Fig. 1). Group VI is a newly identified cluster consisting of only two Chinese isolates. The sequences from Chinese CTV isolates segregated into all six groups, suggesting no correlation between the genetic relationship and geographic origin of CTV isolates. The majority (41.7 %) of CP sequences belonged to group RB, followed by group VT (32.0 %), groups T36 and HA (both 7.8 %), group T30 (5.8 %) and the new group VI (4.9 %). The average genetic distance over sequence pairs within each of the six CP-derived groups ranged from 0.005 ± 0.003 to 0.041 ± 0.006, and the average genetic distance between groups ranged from 0.064 ± 0.008 to 0.104 ± 0.014 (Supplementary Table 2).

In the phylogenetic tree generated from CPm nucleotide sequences, all CTV sequences from China segregated into five groups. These were identified based on the relative positions of the same reference isolates in the CP-derived tree. Very low genetic distance (0.029 ± 0.004) was found between clusters RB and T30, which were thus designated as a single group (RB/T30). Except for isolate HN-XT34-1,

Table 1 Nucleotide diversity and neutrality tests calculated for the CPm and CP of CTV isolates on the basis of their geographical origins

Gene	Position	Geographic origin	Nt[a]	Nucleotide diversity and Fst Value[b]	dN[c]	dS[c]	dN/dS[c]	Fu & Li's D[d]	Fu & Li's F[d]
CPm	1–720	Worldwide	115	0.07619	0.01608	0.37602	0.043	−1.26945	−1.04387
		Other countries	58	0.07753	0.01545	0.38912	0.040	0.41111	0.40584
		China	57	0.07310 (0.02204)*	0.01564	0.35967	0.043	−0.52031	−0.49653
CP	1-669	Worldwide	655	0.06764	0.01911	0.28970	0.066	−8.00050*	−5.01659*
		Other countries	587	0.06794	0.01906	0.29019	0.066	−7.85858*	−4.99034*
		China	68	0.05918 (0.06654)*	0.01767	0.24442	0.072	−1.17394	−1.127336

[a] Nt = number of nucleotides analyzed, and all available and correct sequences in GenBank and obtained in this study are included

[b] Nucleotide diversity = average nucleotide substitutions per site between sequence pairs. Fst values between subpopulations from China and other countries to estimate the extent of gene flow are in parentheses. An absolute value of Fst less than 0.33 suggests infrequent gene flow

* The P values of Ks* Z*and Snn were significant (P < 0.01)

[c] dN, average number of nonsynonymous substitutions per nonsynonymous site; dS, average number of synonymous substitutions per synonymous site; dN/dS, average ratio between nonsynonymous and synonymous substitutions in sequence pairs

[d] * Significant values that reject the null hypothesis of selective neutrality; P, 0.02

and as indicated by black circles in Fig. 1, the CP and CPm sequences of most isolates obtained in this study showed the same positions in both the CP and CPm phylogenetic trees. The CP sequence of HN-XT34-1 fell into group T30, while its CPm sequence fell into group VI. In addition, the CP and CPm sequences of isolates Taiwan-Pum and Kpg 3 retrieved from GenBank also showed incongruent positions in both phylogenetic trees (as indicated by black squares in Fig. 1). The average genetic distance over sequence pairs ranged from 0.027 ± 0.004 to 0.050 ± 0.006 within each of the five CPm-derived groups, and from 0.073 ± 0.009 to 0.146 ± 0.017 between groups (Supplementary Table 2).

It was interesting to find that isolates from three provinces (Sichuan, Hubei and Jiangxi) in China formed a separate subgroup in group VT irrespective of their CP and CPm sequences. Isolates in this subgroup had an insertion of three amino acids (Q^{537th}-N^{538th}-R^{539th}) at the C-terminal end of the p61 protein compared with most isolates in other groups (Suppl. Fig. S2). Three reference isolates (NZ-B18, T318A and Kpg 3) in group VT had an insertion of two amino acids (Q^{537th}-N^{538th} or Q^{537th}-S^{538th}) at the same position. This subgroup was designated as AT-1. A HinfI/RFLP analysis of the AT-1 CP gene showed that it contained a single sequence variant and belonged to the HinfI/RFLP group III (data not shown), a prevalent group in China [10, 29].

Notably, most CTV isolates from Hunan province fell into group RB. The results obtained here confirm the presence of RB-like CTV isolates in China [28].

Groups RB and VT consist of CTV isolates from China and other countries. The close genetic relationships amongst isolates in these groups indicate that several events of long-distance migration (gene flow) could have

occurred between China and other countries. Three permutation-based statistical tests (Ks*, Z* and Snn) and the Fst statistics implemented in the program DnaSP 5.1 [12] were further used to assess the genetic differentiation and gene flow level between subpopulations. Normally, an absolute value of Fst >0.33 suggests infrequent gene flow, while an absolute value of <0.33 suggests frequent gene flow [12]. All available CP and CPm sequences deposited in GenBank were used for gene flow and natural selection analysis (Table 1). We considered two subpopulations: CTV isolates from China and isolates from other countries. Both subpopulations gave P-values <0.01 for the three tests and Fst values of 0.02204 and 0.06654 for the CPm and CP gene, respectively, suggesting low genetic differentiation in these two genes and a relatively higher gene flow in the CPm than that in CP. Long-distance gene flow or migration has also been described for CTV isolates from other countries [22], suggesting that frequent genetic exchange could have occurred over time between CTV populations from different geographical locations.

The rate of synonymous substitutions per synonymous site (dS) and the rate of nonsynonymous substitutions per nonsynonymous site (dN) in the CPm and CP genes were calculated using the program DnaSP 5.1. The values of dN/dS for the CPm and CP genes of the worldwide population and two subpopulations were lower than one, indicating purifying selection (Table 1). The higher value of dN/dS for CP (0.066) compared with CPm (0.044) suggests that CPm is under stronger selection pressure than CP. To evaluate the importance of natural selection in shaping the population structure of CTV, Fu and Li's D and Fu and Li's F tests were used to test the mutation neutrality hypothesis for three subpopulations. The positive values for CPm of the subpopulation consisting of isolates from other

Table 2 Recombination events detected in the 1540-bp fragment sequences of CTV isolates by using seven programs packaged in the software RDP3

Recombinant isolate[a]	Parental isolates[a]		Break point (nt)[b]		P-values determined by seven different programs						
	Major	Minor	Beginning	Ending	RDP	GeneConv	BootScan	MaxChi	Chimaera	Siscan	3Seq
T30	NZ-B18	**HN-XT29-3** (98.3 %)	5	774	3.004×10^{-48}	5.406×10^{-101}	1.439×10^{-114}	3.388×10^{-45}	2.090×10^{-12}	1.344×10^{-71}	–
T385	**SC-Z-1-2-1**	**HN-XT40-1** (97.9 %)	5	774	2.505×10^{-48}	1.491×10^{-100}	1.649×10^{-111}	3.588×10^{-45}	2.863×10^{-12}	1.145×10^{-72}	–
HB-N4-1	**SC-Z-1-2-1**	**HN-XT29-3** (98.1 %)	5	774	3.656×10^{-46}	1.334×10^{-100}	1.193×10^{-115}	4.273×10^{-45}	4.781×10^{-12}	1.371×10^{-71}	–
Qaha	**SC-WZ-9**	Mexico (99.5 %)	3	750	3.702×10^{-50}	2.487×10^{-49}	8.356×10^{-51}	9.509×10^{-15}	2.730×10^{-15}	6.405×10^{-24}	3.149×10^{-55}
T36	**SC-WZ-9**	Mexico (99.9 %)	3	747	8.934×10^{-52}	7.220×10^{-53}	4.768×10^{-55}	6.961×10^{-15}	6.576×10^{-15}	1.166×10^{-24}	1.577×10^{-55}
SC-Z-1-1-10	T318A	HA18-9 (93.8 %)	25	744	1.237×10^{-38}	2.361×10^{-83}	2.122×10^{-85}	4.739×10^{-40}	3.260×10^{-11}	4.155×10^{-55}	–
SC-Z-1-2-7	T318A	HA18-9 (93.9 %)	25	744	6.387×10^{-39}	3.588×10^{-84}	2.976×10^{-85}	3.483×10^{-40}	3.264×10^{-11}	1.656×10^{-55}	–
T318A	**JX19-6**	HA18-9	25	754	4.391×10^{-37}	1.724×10^{-76}	2.551×10^{-78}	8.100×10^{-39}	2.949×10^{-10}	7.932×10^{-55}	–
VT	**SC-Z-1-2-7**	Mexico	15	744	5.304×10^{-43}	9.226×10^{-91}	1.812×10^{-93}	6.861×10^{-42}	2.217×10^{-11}	1.381×10^{-55}	–
NZ-B18	**HN-XT-29-3**	**HB-N4-1**	5	774	7.205×10^{-46}	1.333×10^{-100}	1.193×10^{-115}	4.273×10^{-45}	4.518×10^{-12}	8.492×10^{-74}	–
B165	**SC-Z-1-2-7**	Mexico	18	744	1.414×10^{-43}	2.217×10^{-90}	3.076×10^{-93}	1.160×10^{-41}	6.570×10^{-12}	9.785×10^{-56}	–
NZ-M16	**SC-Z-1-2-7**	Mexico	18	744	2.349×10^{-42}	1.088×10^{-90}	1.926×10^{-93}	8.171×10^{-42}	3.477×10^{-11}	6.057×10^{-56}	–
A18	**SC-Z-1-2-7**	Mexico	17	744	3.959×10^{-42}	5.976×10^{-91}	1.897×10^{-93}	3.718×10^{-42}	2.501×10^{-11}	6.844×10^{-56}	–
AT-1	**SC-Z-1-2-7**	Mexico	18	744	1.232×10^{-41}	1.412×10^{-90}	3.196×10^{-93}	6.272×10^{-42}	2.358×10^{-11}	9.894×10^{-56}	–
CT11A	**SC-Z-1-2-7**	Mexico	18	744	2.520×10^{-42}	2.312×10^{-90}	3.836×10^{-93}	8.184×10^{-42}	9.668×10^{-12}	5.421×10^{-56}	–
SC-Z-1-2-1	**SC-Z-1-2-7**	Mexico	18	744	1.615×10^{-42}	1.960×10^{-90}	3.608×10^{-93}	8.894×10^{-42}	6.602×10^{-12}	3.360×10^{-55}	–
SC-Z-2-1-2	**SC-Z-1-2-7**	Mexico	18	744	1.889×10^{-41}	1.959×10^{-90}	3.608×10^{-93}	8.894×10^{-42}	2.349×10^{-11}	2.616×10^{-56}	–
HB-S36-1	**SC-Z-1-2-7**	Mexico	18	744	8.713×10^{-43}	1.201×10^{-90}	3.010×10^{-93}	5.269×10^{-42}	2.360×10^{-11}	4.215×10^{-55}	–
JX36-4	**SC-Z-1-2-7**	Mexico	18	744	3.537×10^{-39}	2.056×10^{-90}	4.502×10^{-93}	6.277×10^{-42}	2.360×10^{-11}	4.886×10^{-55}	–
SC-BSH-1	**SC-Z-1-2-7**	Mexico	18	744	1.311×10^{-42}	1.587×10^{-90}	2.719×10^{-93}	8.178×10^{-42}	4.212×10^{-12}	2.692×10^{-55}	–

Table 2 continued

Recombinant isolate[a]	Parental isolates[a]		Break point (nt)[b]		P-values determined by seven different programs						
	Major	Minor	Beginning	Ending	RDP	GeneConv	BootScan	MaxChi	Chimaera	Siscan	3Seq
SC-BSH-10	**SC-Z-1-2-7**	Mexico	18	744	1.410×10^{-41}	1.662×10^{-90}	3.395×10^{-93}	7.468×10^{-42}	2.335×10^{-11}	5.451×10^{-55}	–
SC-PG-4	**SC-Z-1-2-7**	Mexico	18	744	1.232×10^{-41}	1.412×10^{-90}	3.196×10^{-93}	6.272×10^{-42}	2.358×10^{-11}	4.886×10^{-55}	–
SC-WZ-6	**JX19-6**	Mexico	18	747	3.618×10^{-37}	8.349×10^{-32}	2.570×10^{-85}	5.042×10^{-20}	3.137×10^{-10}	1.201×10^{-55}	–
SC-Z-2-3-1	**SC-Z-1-2-7**	Mexico	18	744	1.127×10^{-41}	1.959×10^{-90}	3.608×10^{-93}	8.894×10^{-42}	2.349×10^{-11}	2.748×10^{-56}	–
HB-S4-2	**SC-Z-1-2-7**	Mexico	18	744	9.023×10^{-43}	7.836×10^{-91}	1.706×10^{-93}	5.763×10^{-42}	2.217×10^{-11}	4.490×10^{-56}	–
SC-Z-2-1-4	**HN-XT33-6**	**HB-S36-4**	1, 1488	762, X	1.577×10^{-31}	2.323×10^{-67}	1.932×10^{-67}	9.326×10^{-37}	–	1.570×10^{-57}	–
SC-Z-1-1-4	**HN-XT33-6**	Mexico	1, 1488	761, X	7.736×10^{-32}	3.627×10^{-76}	3.241×10^{-74}	1.105×10^{-38}	7.431×10^{-10}	1.799×10^{-56}	–

[a] Recombinants and parental isolates from this study are indicated in bold

[b] The analyzed sequences included the C-terminal end of the p61 gene (32 bp), the complete CPm gene (723 bp), a partial CP gene (652 bp), and a non-coding region (about 92 bp) between CPm and CP. The beginning points at the nucleotide positions 1, 3, 5, 17, 18, and 25 are located in the p61 gene; The ending points at the nucleotide positions 744 (aa 238), 747 (aa 239), and 750 (stop codon) are located inside of the CPm region, and those at the nucleotide positions 774 and 762 are located in the inter non-coding region. The beginning point at the nucleotide positions 1488 is located in the CP region. X indicates an unknown recombination site outside of the CP gene

countries suggested a possible population bottleneck (Table 1). Although the values for both CP and CPm from the Chinese subpopulation were non-significant, significant negative values were observed for the CP gene of the global population, suggesting population expansion. Therefore, demographic events may have contributed to CTV evolution.

For the CP gene, three positively selected sites (1.3 %) corresponding to codons 31, 41 and 68, and 67 negatively selected sites (30.0 %) were identified by both the FEL and IFEL methods at a significance level of 0.1. For the CPm gene, two positively selected sites (0.8 %) at positions 9 and 102, and 95 negatively selected sites (39.6 %) were identified. Positive selection at codon 102 has been reported previously [7]. It is known that CPm is involved in aphid transmission and virion assembly, and CP is related to virion assembly and viral RNA protection [4, 24]. Thus, the biological significance of those positively selected sites needs to be investigated further.

Initially, only one recombination event was found in the CP gene of isolate SC-WZ-9, and we used the appropriate 1498-bp fragment sequences for further investigations of recombination. The incongruent positions of isolate HN-XT34-1 in the CP- and CPm-based phylograms suggested a recombination event (Fig. 1). Further analysis confirmed the presence of a recombination event in HN-XT34-1, occurring between the C-terminus of CPm and the internal, non-coding region, with putative parental sequences from the RB and T30 groups (Suppl. Fig. S1B). In addition, putative recombination events were detected within 27 1498-bp sequences using RDP (Table 2). Sequence alignments showed that most of the recombination events occurred in AT-rich regions (Suppl. Fig. S3). Although the exact parents for these recombinants were not identified, most of the recombinants, irrespective of their geographic locations, had a putative parental sequence from Mexico or from isolates of the HA group, further supporting an ancient recombination event that occurred before the recombinants spread worldwide, as previously suggested for the CP gene [23]. Most of the recombination events occurred in the CPm region, with only two recombination events detected in the CP gene, suggesting that the CPm gene could be a recombination hotspot in the CTV genome. Most recombination breakpoints (except for two events in SC-Z-2-1-4 and SC-Z-1-1-4) were located outside of the CPm and CP regions, which could explain why only one recombination event in CP and no recombination events in CPm were detected when using the CPm or CP gene individually for the test. Previous reports have indicated that the VT strain can be more effectively transmitted by aphids than strains T36 and T30 [3, 26], and our results show that recombinant isolates belonged mainly to the VT

group, providing a possible explanation for the expansion of the VT cluster. The high gene flow between distant geographical regions, the possible selection pressure and multiple infections by successive aphid inoculations could have contributed to the recombination events between variant sequences [20].

In conclusion, this study represents a comprehensive analysis of the genetic diversity and population structure of CTV isolates in China. Our results also revealed the balance between strong purifying selection and high genetic variability in both the CPm and CP genes, which could be related to their biological functions. The structure of the CTV population in China is not associated with geographic location and is mainly shaped by long-distance gene flow, negative selection and frequent recombination.

References

1. Atta S, Zhou CY, Zhou Y, Cao MJ, Wang XF (2012) Distribution and research advances of *Citrus tristeza virus*. J Integr Agr 11:346–358
2. Barzegar A, Rahimian H, Sohi HH (2010) Comparison of the minor coat protein gene sequences of aphid-transmissible and -nontransmissible isolates of *Citrus tristeza virus*. J Gen Plant Pathol 76:143–151
3. Brlansky RH, Roy A, Damsteegt VD (2011) Stem-pitting *Citrus tristeza virus* predominantly transmitted by the brown citrus aphid from mixed infections containing non-stem-pitting and stem-pitting isolates. Plant Dis 95:913–920
4. Chen AYS, Walker GP, Carter D, Ng JCK (2011) A virus capsid component mediates virion retention and transmission by its insect vector. Proc Natl Acad Sci USA 108:16777–16782
5. Febres VJ, Ashoulin L, Mawassi M, Frank A, Bar-Joseph M, Manjunath KL, Lee RF, Niblett CL (1996) The p27 protein is present at one end of Citrus tristeza virus particles. Phytopathology 86:1331–1335
6. 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
7. Harper SJ (2013) *Citrus tristeza virus*: Evolution of complex and varied genotypic groups. Front. Microbiol 4:93. doi:10.3389/fmicb.2013.00093
8. Hilf ME, Karasev AV, Albiach-Marti MR, Dawson WO, Garnsey SM (1999) Two paths of sequence divergence in the *Citrus tristeza virus* complex. Phytopathology 89:336–342
9. Hilf ME, Mavrodieva VA, Garnsey SM (2005) Genetic marker analysis of a global collection of isolates of *Citrus tristeza virus*: characterization and distribution of CTV genotypes and association with symptoms. Phytopathology 95:909–917
10. Jiang B, Hong N, Wang GP, Hu J, Zhang JK, Wang CX, Liu Y, Fan XD (2008) Characterization of *Citrus tristeza virus* strains from southern China based on analysis of restriction patterns and sequences of their coat protein genes. Virus Genes 37:185–192
11. 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

12. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

13. 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

14. Martin 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

15. Maydt J, Lengauer T (2006) Recco: recombination analysis using cost optimization. Bioinformatics 22:1064–1071

16. 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

17. Moreno P, Ambros S, Albiach-Marti MR, Guerri J, Pena L (2008) Plant diseases that changed the world—Citrus tristeza virus: a pathogen that changed the course of the citrus industry. Mol Plant Pathol 9:251–268

18. Oliveros-Garay OA, Martinez-Salazar N, Torres-Ruiz Y, Acosta O (2009) CPm gene diversity in field isolates of *Citrus tristeza virus* from Colombia. Arch Virol 154:1933–1937

19. Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. J Mol Evol 54:396–402

20. Roy A, Brlansky RH (2009) Population dynamics of a Florida *Citrus tristeza virus* isolate and aphid-transmitted subisolates: identification of three genotypic groups and recombinants after aphid transmission. Phytopathology 99:1297–1306

21. 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

22. Rubio L, Ayllón MaA, Kong P, Fernández A, Polek M, Guerri J, Moreno P, Falk BW (2001) Genetic variation of *Citrus tristeza virus* isolates from California and Spain: evidence for mixed infections and recombination. J Virol 75:8054–8062

23. Rubio L, Guerri J, Moreno P (2013) Genetic variability and evolutionary dynamics of viruses of the family *Closteroviridae*. Front Microbiol 4:151. doi:10.3389/fmicb.2013.00151

24. Satyanarayana T, Gowda S, Ayllón MA, Dawson WO (2004) *Closterovirus* bipolar virion: evidence for initiation of assembly by minor coat protein and its restriction to the genomic RNA 5′ region. Proc Natl Acad Sci USA 101:799–804

25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

26. Velazquez-Monreal JJ, Mathews DM, Dodds JA (2009) Segregation of distinct variants from *Citrus tristeza virus* isolate SY568 using aphid transmission. Phytopathology 99:1168–1176

27. Weng Z, Barthelson R, Gowda S, Hilf ME, Dawson WO, Galbraith DW, Xiong Z (2007) Persistent infection and promiscuous recombination of multiple genotypes of an RNA virus within a single host generate extensive diversity. PLoS ONE 2:e917

28. Wu GW, Pan S, Wang GP, Tang M, Liu Y, Yang F, Hong N (2013) The genotypes of *Citrus tristeza virus* isolates from China revealed by sequence analysis of multiple molecular markers. Arch Virol 158:231–235

29. Zhou Y, Zhou CY, Song Z, Liu KH, Yang FY (2007) Characterization of *Citrus tristeza virus* isolates by indicators and molecular biology methods. Agric Sci China 6:573–579