RESEARCH

Phytopathology Research

Open Access



Genetic variation and molecular evolution of tomato yellow leaf curl China virus and its betasatellite DNA isolates in China

Jingwen Yu^{1,3†}, Yan Xiong^{1†}, Yin Li¹, Meisheng Zhao¹, Xiaolong Yang¹, Yalu Wang⁴, Hongkun Huang⁴, Changyong Zhou², Wenkun Huang^{3*} and Ling Qing^{1,2*}

Abstract

Tomato yellow leaf curl China virus (TYLCCNV) and its betasatellite DNA isolates (TYLCCNB), seriously threaten tomato crop production in China. The present work aimed to analyze the genetic diversity and population structure of TYL-CCNV/TYLCCNB, collected from 168 leaf samples with apparent yellow and curly leaf disease symptoms in China. The study involves phylogenetic, recombination, and selection pressure analysis, based on the genome sequences of 57 TYLCCNV and 109 TYLCCNB isolates. It was found that the TYLCCNV/TYLCCNB populations collected from the same geographic regions exhibit a close relationship under phylogenetic analysis. The recombination analysis revealed 8 possible recombination sites in the TYLCCNV C1 and C4 genes, and 6 possible recombination sites in the TYLCCNB $\beta C1$ gene. The results showed that the TYLCCNV C4 gene was under positive selection pressure in the selection pressure analysis. Moreover, nucleotide and predicted amino acid sequence identities in C1 and C4 were significantly lower than other ORF region sequences. The lower gene flow and significant genetic differentiation between the geographic populations of Guangxi and Sichuan provinces suggested that environmental adaptation was an important evolutionary force in shaping the genetic structure of TYLCCNV/TYLCCNB. In addition, C1 and C4 ORFs of TYLCCNV were proved to be the major mutation regions in greenhouse and field inoculation experiments. A-rich region was the major mutant hot spot in the associated betasatellites such as TYLCCNB, TbCSB, and MYVB. A thorough investigation into the evolutionary factors affecting the population structure of TYLCCNV/TYLCCNB will provide vital information for systematic virus management.

Keywords Genetic variability, Molecular evolution, Tomato yellow leaf curl China virus, Tomato yellow leaf curl China betasatellite

[†]Jingwen Yu and Yan Xiong contributed equally to this study.

*Correspondence: Wenkun Huang wkhuang2002@163.com Ling Qing qling@swu.edu.cn ¹ Chongqing Key Laboratory of Plant Disease Biology, College of Plant Protection, Southwest University, Chongqing 400715, China ² National Citrus Engineering Research Center, Southwest University, Chongqing 400712, China ³ State Key Laboratory for Biology of Plant Diseases and Insect Pests, hotigitude of Plant Protection Chinese Academy of Arricultural Sciences

Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

⁴ Agricultural Ecology and Resources Protection Station of the Ministry of Agriculture and Villages, Beijing 100125, China

Background

Viruses of the Geminiviridae family pose an imminent threat to global food security via seriously damaging many economically important crops (such as cotton, tomato, maize, cassava, and wheat) in tropical and temperate regions of the world. The family Geminiviridae comprises 14 genera, with *Begomovirus* being the largest with over 400 species, based on their genome structure, host ranges, and insect vectors (Walker et al. 2021). Viruses in the genera Begomovirus have mono- or bipartite genomes, whereas the other 13 genera have only monopartite genomes (Ren et al. 2022). The geographic distribution of begomoviruses spread by the whitefly Bemisia tabaci to infect dicotyledonous plants are mainly classified in two subgroups: New World viruses (the Americas) and Old World viruses (Europe, Africa, Asia, and Australasia) (Harrison and Robinson 1999). The genomes of Old World begomoviruses are either monopartite or bipartite, while New World begomoviruses exhibit bipartite genomes (Zhou 2013).

Tomato yellow leaf curl China virus (TYLCCNV, genus Begomovirus) is a typical monopartite geminivirus, which appears associated with the betasatellite DNA TYL-CCNB in the field (Yin et al. 2001). It is one of the most damaging and threatening viruses for tomato production in China. The infection of TYLCCNV alone did not cause any obvious symptoms in *Nicotiana benthamiana*, N. glutinosa, N. tabacum, or tomato plants. However, when co-infection of TYLCCNV and TYLCCNB occurs, the hosts showed signs of dwarfing, leaf curling, yellow mosaic patterns, and stem deformation (Cui et al. 2004). The circular ssDNA molecules known as TYLCCNB are approximately half the size (~1.4 kb) of the genome of the helper virus (TYLCCNV) (Liu et al. 1998; Xie et al. 2013). TYLCCNB is dependent on TYLCCNV for plant mobility, insect transmission, and replication (Zhou et al. 2003; Settlage et al. 2005).

The ORF regions of the TYLCCNV genome contain six known genes. The virion-sense strand contains genes VI and V2, while the complementary-sense strand contains genes C1, C2, C3, and C4 (King et al. 2011). It has been demonstrated that VI encodes the viral coat protein (CP), involved in the encapsidation of viral genome, whereas V2 encodes the movement protein (MP), which is responsible for virus motility in plants and serves as a suppressor of RNA silencing (Briddon et al. 1990; Glick et al. 2009; Mubin et al. 2010). Previous studies also showed that the complementary sense gene C1 encodes the replication enhancer protein Rep (Hanley-Bowdoin et al. 2004). Interestingly, C2 encodes a transcriptional activator protein (TrAP), which activates the expression of coat protein (CP) and movement protein (MP) genes (Vanitharani et al. 2004). While C3 is a replication enhancer protein (REn), which promotes the accumulation of virus (Settlage et al. 2005). Additionally, it was shown that C4 protein determines disease symptoms (Rigden et al. 1994). In contrast to TYLCCNV, TYLCCNB encodes the β C1 protein in the complementary sense orientation. The β C1 protein is a movement protein, essential for the intercellular transmission of the virus (Zhou 2013). Furthermore, β C1 protein functions as an RNA silencing suppressor and is involved in transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) (Li et al. 2018). Recently, β V1 protein was identified to be conserved on the β -satellite and plays a role in the promotion of virulence during the co-infection by TYLCCNV/TYLCCNB (Hu et al. 2020).

The population structure and genetic diversity of plant viruses have been discovered to be highly connected with their outbreaks, geographical origin, host range, and transmission vectors (García-Arenal et al. 2001). The analysis of variation in TYLCCNV populations in different stages after infecting the host found that the level of TYLCCNV variation was similar to that of RNA viruses (Ge et al. 2007). Similarly, TYLCCNV exhibits a quasispecies structure identical to that of RNA viruses, which is one of the important reasons for the loss of resistance in crop varieties (Domingo et al. 1998). Furthermore, recombination, pseudo-recombination, and mutations were all significant contributors to the rapid mutation of TYLCCNV (Harrison and Robinson 1999). Additionally, it was observed that the homology of the virus, geo-ecological, and climatic conditions are highly correlated. However, few quantitative studies have been conducted to investigate the genetic structure and variation in TYLCCNV populations under controlled conditions, indicating that there is a lack of knowledge on genetic variation in TYL-CCNV in China.

The current study mainly focuses on TYLCCNV's molecular characteristics, biology, genetic diversity, and molecular variability during infection. The objectives of this study were as follows: (i) investigate the distribution of TYLCCNV/TYLCCNB in China; (ii) study the genetic diversity and genetic structure of TYLCCNV/TYLCCNB populations in different hosts based on sequences of different ORF regions; and (iii) characterize molecular variation among different begomoviruses under natural and experimental populations. These investigation of TYLCCNV will help trace the viruses origin and development, enhance our knowledge regarding its epidemiology, and provide a theoretical foundation for the management of virus disease.

Results

PCR detection of TYLCCNV and TYLCCNB isolates

A total of 168 leaf samples were collected from 11 crops and 11 weed species in three provinces of China, which exhibited typical symptoms of begomoviruses (leaf curling, yellowing, and growth retardation) in the field. These samples were used to detect and identify TYLCCNV/ TYLCCNB infection. PCR results revealed that TYL-CCNV was detected in 67 (39.9%, 67 of 168) samples, whereas TYLCCNB was observed in 46 (27.4%, 46 of 168) samples (Additional file 1: Table S1). While samples collected from two provinces were: Yunnan [TYL-CCNV (35.2%, 19 of 54), TYLCCNB (25.9%, 14 of 54)] and Sichuan [TYLCCNV (50.5%, 48 of 95), TYLCCNB (33.7%, 32 of 95)] (Additional file 1: Table S1). In addition, TYLCCNV/TYLCCNB were detected in only two crops (Nicotiana tabacum and Solanum lycopersicum L.) and three weeds (Ageratum conyzoides L., Malvastrum coromandelianum Garcke, and Malva sinensis Cavan.) (Fig. 1a). A total of 23 new TYLCCNV isolates and 10 new TYLCCNB isolates were sequenced and assembled based on genomic organization, geographic, and host origins (Fig. 1b, Additional file 1: Table S2 and Table S3).

Phylogenetic classification of TYLCCNV and TYLCCNB isolates

To clarify the relationships between TYLCCNV isolates and TYLCCNB isolates, the whole genome sequences of 57 TYLCCNV isolates (23 sequences obtained in this study and 34 corresponding sequences in the GenBank) and 109 TYLCCNBV isolates (10 sequences obtained in this study and 99 corresponding sequences in the GenBank) were used to construct phylogenetic trees using the software MEGA11, respectively. The 57 TYL-CCNV sequences were clustered into three phylogenetic groups, correlating to some extent with geographic origins (Group I and III: Yunnan & Sichuan, and Group II: Yunnan & Guangxi) (Fig. 1c). Notably, Guangxi isolates (n=5) and Sichuan isolates (n=15) did not cluster into one subgroup. Similarly, the 109 TYLCCNB sequences clustered into four phylogenetic groups, closely related to geographic origin (Groups I and IV: Yunnan & Guangxi, Group II: Yunnan, and Group III: Yunnan & Sichuan) (Fig. 1d). In addition, phylogenetic analyses of TYL-CCNV and TYLCCNB isolates revealed no association with their host origins. (Additional file 2: Figure S1).

Recombination analysis in TYLCCNV and TYLCCNB isolates

A crucial factor in the development and evolution of begomoviruses is recombination. The SplitsTree 4 v.4.14.6 analyses showed that TYLCCNV and TYL-CCNB sequences were linked to one another via various pathways to form a network structure, suggesting that the populations of TYLCCNV and TYLCCNB may have undergone multiple possible recombination events. Meanwhile, the constructed network showed that the TYLCCNV population was divided into three groups of isolates (Group I and II: Yunnan & Sichuan, and Group III: Yunnan & Guangxi) (Additional file 2: Figure S2a), and the TYLCCNB population was also distinguished into three groups of isolates (Groups I: Yunnan & Sichuan, Group II: Yunnan & Guangxi, and Group III: Yunnan) based on geographic origin (Additional file 2: Figure S2b).

The RDP4 algorithms identified multiple major putative recombination events which spread across the various coding regions. Recombination analysis revealed 12 clear recombinants among the 57 TYLCCNV isolates, with at least four of the seven recombination detection algorithms below the threshold P < 0.05 as acceptable. However, only 10 significant recombinants were detected in 109 TYLCCNB isolates. Interestingly, the majority of recombination breakpoints (67%) in TYLCCNCV isolates were identified in the C1 (position in nucleotides 1501-594) and C4 (position in nucleotides 2136-2437) gene regions, whereas recombination breakpoints (60%) in TYLCCNCB isolates were located in the $\beta C1$ (position in nucleotides 209-565) gene region (Table 1). These results indicated that the TYLCCNV genome regions of *C1* and *C4* and the TYLCCNB genome regions of β *C1* are recombination hotspots.

Sequence identity analysis in TYLCCNV and TYLCCNB isolates

The nucleotide sequence identities of 57 TYLCCNV and 109 TYLCCNB genomes among and within phylogenetic groups were identified based on geographic origin. The nucleotide sequence identities ranged from 78.4 to 99.8% among 57 TYLCCNV isolates, while nucleotide sequence identities between genome sequences of 109 TYLCCNB isolates ranged from 69.4 to 100%, indicating a high level of sequence diversity among isolates (Additional file 1: Table S4 and Table S5 and Additional file 2: Figure S3). To further analyze the degree of variation in the TYLCCNV and TYLCCNB genomes, we used the average pairwise diversity parameter (π) for all positions on the TYL-CCNV genome. The regions with the highest nucleotide variation occur in the 3' terminal of C1 and throughout the C4 region (Fig. 2a). However, the lowest nucleotide diversity was in the C2 region. Similarly, nucleotide variation was not evident in the $\beta C1$ and $\beta V1$ regions of the TYLCCNB genome (Fig. 2b).

Variations in TYLCCNV protein and nucleotide sequences were also analyzed. Overall, the predicted protein amino acid sequence identities (aa) of different gene regions of TYLCCNV isolates were significantly lower in



Fig. 1 a Typical symptoms caused by TYLCCNV in some collected samples. **b** Genome organization features of TYLCCNV and TYLCCNB. Neighbor-joining (NJ) phylogenetic tree constructed using MEGA11 based on **c** 57 TYLCCNV isolates and **d** 109 TYLCCNB isolates. Different geographic regions are represented (\bullet =Yunnan, \blacksquare =Sichuan, \blacktriangleleft =Guangxi). Tobacco curly shoot virus (TbCSV, NCBI accession number NC_003722 and TbCSB, NCBI accession number NC_004546) isolate served as an outgroup (\star)

| Table 1 | Putative recombination events among 57 Tomato | yellow leaf curl China viru | us (TYLCCNV) isolates | s and 109 Tomato y | ellow leaf/ |
|----------|---|-----------------------------|-----------------------|--------------------|-------------|
| curl Chi | na virus betasatellite (TYLCCNB) | | | | |

| | Event | Putative | Putative parent | | Breakpoint ^a | Dete | ection | netho | 1 ^b | | | |
|---------|-------|-------------|-----------------|-------------------|-------------------------|------|--------|-------|----------------|----|----|----|
| | | recombinant | Major | Minor | | R | G | В | м | с | S | т |
| TYLCCNV | 1 | MY432 | SC230 | Unknown (YN-2022) | 1206-2734 | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| | 2 | YN72 | YN4476 | Dali-198 | 520-2733 | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| | 3 | YN4534 | G102 | Y295 | 111-925 | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| | 4 | YN182 | 09–29 | Y295 | 1531-2727 | ++ | ++ | ++ | ++ | + | ++ | ++ |
| | 5 | SC65 | HG263 | Y8 | 176-1052 | ++ | - | + | + | ++ | + | + |
| | 6 | YN4467 | Y10 | YN4348 | 180–2,706 | + | + | - | ++ | ++ | ++ | ++ |
| | 7 | Y278 | YN48 | Unknown (09–29) | 1240-2694 | + | + | - | + | + | ++ | + |
| | 8 | Dali-198 | Y295 | YN204 | 2273-2680 | + | + | - | + | + | ++ | + |
| | 9 | YN204 | YM8 | YN182 | 2352-2726 | + | + | + | + | + | ++ | ++ |
| | 10 | HG263 | NN2 | Unknown (SC44) | 140-982 | + | - | - | + | + | + | + |
| | 11 | YM5 | Y11 | Unknown (YN182) | 2358-2685 | + | - | - | + | + | + | + |
| | 12 | Y10 | Y278 | Y38 | 712-952 | + | - | - | + | + | - | + |
| TYLCCNB | 1 | HG230 | YM5 | YN506 | 30-523 | ++ | - | ++ | ++ | ++ | ++ | ++ |
| | 2 | YN206 | Y5 | GX-BS | 43-313 | ++ | ++ | ++ | ++ | + | ++ | ++ |
| | 3 | Y92 | Y87 | Y263 | 635–1374 | _ | - | _ | + | ++ | ++ | ++ |
| | 4 | GX-BS | YN20 | Unknown (Y268) | 442-476 | - | ++ | - | + | + | - | + |
| | 5 | YN147 | YN38 | G102 | 120–978 | - | - | - | + | + | + | ++ |
| | 6 | Y330 | G103 | YN154 | 265-1341 | - | - | - | + | + | ++ | ++ |
| | 7 | YN506 | Unknown (G103) | YN466 | 1-1210 | + | + | + | + | + | + | + |
| | 8 | YN147 | Unknown (G103) | YN391 | 1271-1405 | - | + | + | + | - | + | + |
| | 9 | HG230 | G103 | YN383 | 982-1381 | + | + | + | + | + | ++ | + |
| | 10 | G103 | Y244 | Unknown (Y87) | 1210-1380 | + | + | + | + | - | + | + |

^a Recombination breakpoint

^b Seven algorithms were used: RDP (R), GENECONV (G), BootScan (B), Maximum χ -Square (M), Chimaera (C), SisScan (S), and 3Seq (T). Statistical significance for recombination: + indicates $10^{-6} < P$ value ≤ 0.05 , ++indicates P value $\leq 10^{-6}$, and - indicates not significant (P value > 0.05)



Fig. 2 Distribution of nucleotide diversity (π) along **a** 57 TYLCCNV whole-genome sequence and **b** 109 TYLCCNB whole-genome sequence. The nucleotide diversity (Y-axis) was plotted against nucleotide position (X-axis) using DnaSP6 with a 100-nucleotide (nt) sliding window and a 25-nt step size. C1 = Replication-associated protein, C2 = Transcriptional activator protein, C3 = Replication enhancer protein, C4 = Disease symptom determinants, V1 = Coat protein, V2 = Pre-coat protein, β C1 = Movement protein, β V1 = Promotion of virulence during the infection

C1 (88.42%) and C4 (83.29%) genes than in other gene regions (C2: 93.14%, C3: 92.97%, V1: 95.80%, V2: 94.04%) (Table 2). Notably, C2 was seen to have the highest nucleotide sequence identity (95.58%), but C4 had the lowest nucleotide sequence identity (90.73%). Meanwhile, the nucleotide sequence identity of the genes in different geographic populations showed a similar trend, ranging from 55.1% to 100% for C4 and 75.5% to 100% for C2 (Additional file 1: Table S4). In addition, synonymous nucleotide mutations were dominant in V1, whereas non-synonymous nucleotide mutations dominated in three genes (C1, C3, C4). Meanwhile, C3 and V1 each had one insertion and deletion event (InDels), and two InDels both occurred on the boundary of V2 (Additional file 1: Table S6). Interestingly, significant differences in protein and nucleotide sequence identity were found in $\beta C1$ and $\beta V1$ genes of TYLCCNB. The predicted protein amino acid sequence identity of β V1 (59.63%) and the nucleotide sequence identity of $\beta V1$ (63.41%) were significantly lower than that of β C1. Nevertheless, the types of β C1 and $\beta V1$ nucleotide variants were mainly non-synonymous nucleotide mutations, and only two InDels were detected in the region of the $\beta V1$ gene.

Differentiation of geographical populations

In order to clarify the degree of differentiation between different genetic regions of TYLCCNV and TYLCCNB isolates, pairwise comparisons of TYLCCNV and TYL-CCNB populations from Yunnan, Guangxi, and Sichuan were performed using three different parameters (Ks*, Z^* , and Snn) (Table 3). Overall, alignment tests of Ks*, Z^* , and Snn between TYLCCNV/TYLCCNB populations from different geographical regions showed significant genetic differentiation (P < 0.05). Additionally, gene flow in different gene coding regions of TYLCCNV and TYL-CCNB isolates was assessed using two parameters (Fst and Nm) of the population. These findings showed infrequent gene flow between Sichuan and Guangxi isolates in the coding regions of *C1*, *C2*, *C3*, *V1*, and *V2* genes, as indicated by |Fst|>0.33 and |Nm|<1. The coding region of the $\beta C1$ and $\beta V1$ genes was detected in TYL-CCNB isolates, which also indicated infrequent gene flow between Sichuan and Guangxi isolates. In conclusion, the data suggest significant genetic differentiation and rare gene flow between geographic groups.

Neutrality tests and selection pressure analysis

Nucleotide diversity and haplotype analyses were performed on six gene sequences from the TYLCCNV genome and the $\beta C1$ and $\beta V1$ gene sequences from TYLCCNB. Among all TYLCCNV and TYLCCNB isolates, the C4 gene had the lowest haplotype diversity (0.970 ± 0.014) , while the V1 gene had the highest haplotype diversity (0.999 ± 0.004) (Table 4). Haplotype nucleotide diversity analysis revealed that the nucleotide diversity of C2 (0.06435 ± 0.00392), C3 (0.05400 ± 0.00454) , and V1 (0.09616 ± 0.00292) were below 0.1. All eight gene sequences of the Yunnan isolates had negative Tajima's D values for the parameter neutrality test, indicating that the genome evolution of the Yunnan population of TYLCCNV followed a neutral evolutionary model. In addition, mean dN/dS ratio values were calculated for eight genes (C1, C2, C3, C4, V1, V2, $\beta C1$, and $\beta V1$), in which six genes (C1, C2, C3, V1, V2, and $\beta V1$) were found to be under negative or purifying selection (dN/dS<1) in each population. However, the *C4* and β *C1* genes were under positive selection because they had a dN/dS > 1 in all geographic populations.

Table 2 Sequence identities (ID), insertion or deletion events (InDels), and site nucleotide mutations in individual genes or proteins encoded by 57 Tomato yellow leaf curl China virus (TYLCCNV) isolates and 109 Tomato yellow leaf curl China virus betasatellite (TYLCCNB)

| Genomic region | Amino acids (aa) |) | | Nucleotide (nt) | Mutations ^a | | |
|----------------|------------------|--------|---------|-----------------|------------------------|-----|-----|
| | Length (aa) | ID (%) | InDels | Length (nt) | ID (%) | Syn | Non |
| C1 (Rep) | 358-361 | 88.42 | None | 1077-1086 | 93.09 | 144 | 310 |
| C2 (TrAP) | 134 | 93.14 | None | 405 | 95.58 | 55 | 95 |
| C3 (Ren) | 116-134 | 92.97 | I | 351-405 | 94.32 | 55 | 101 |
| C4 (SD) | 85–97 | 83.29 | None | 258-294 | 90.73 | 21 | 109 |
| V1 (CP) | 256 | 95.80 | II | 771 | 92.78 | 181 | 91 |
| V2 (Pre-CP) | 112-115 | 94.04 | III, IV | 339–348 | 92.68 | 66 | 70 |
| βC1 | 109-119 | 78.25 | None | 327-357 | 91.70 | 24 | 155 |
| βV1 | 19–131 | 59.63 | V, VI | 57–393 | 63.41 | 53 | 219 |

 $C1 = Replication-associated protein, C2 = Transcriptional activator protein, C3 = Replication enhancer protein, C4 = Disease symptom determinants, V1 = Coat protein, V2 = Pre-coat protein, <math>\beta C1 = Movement protein, \beta V1 = Promotion of virulence during the infection$

^a Number of nucleotide mutations. Syn = synonymous and Non = nonsynonymous

| Table 3 Measurement of genetic differentiation amo | ong 57 Tomato yellow | leaf curl China virus (TYL | CCNV) isolates and 1 | 09 Tomato |
|--|----------------------|----------------------------|----------------------|-----------|
| yellow leaf curl China virus betasatellite (TYLCCNB) | | | | |

| Genomic region | Comparisons | Ks* (P value) | Z* (P value) | Snn (P value) | Fst | Nm |
|----------------|---|---------------------------------|---------------------------------|---------------------------------|---------|-------|
| C1 (Rep) | Yunnan (n = 37) versus Sichuan (n = 15) | 4.36113 (P = 0.0080**) | 6.09968 (<i>P</i> = 0.0050**) | 0.73007 (<i>P</i> = 0.0210*) | 0.05179 | 9.15 |
| | Yunnan (n = 37) versus Guangxi (n = 5) | 4.19892 (P = 0.0200*) | 5.68251 (P = 0.0130*) | 0.97619 (P = 0.0010**) | 0.25381 | 1.47 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 4.26561 (<i>P</i> = 0.0070**) | 3.95030 (P = 0.0050**) | 1.00000 (P = 0.0010**) | 0.36457 | 0.87 |
| C2 (TrAP) | Yunnan (n=37) versus Sichuan (n=15) | $2.96836 (P = 0.0000^{***})$ | 6.06650 (<i>P</i> = 0.0000***) | 0.74382 (<i>P</i> = 0.0030**) | 0.18244 | 2.24 |
| | Yunnan (n=37) versus Guangxi (n=5) | 3.16528 (P = 0.0650 ns) | 5.71014 (P = 0.0450*) | 0.94841 (<i>P</i> = 0.0000***) | 0.18413 | 2.22 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 2.60825 (<i>P</i> = 0.0000***) | 3.93507 (<i>P</i> = 0.0000***) | 1.00000 (<i>P</i> = 0.0000***) | 0.55018 | 0.41 |
| C3 (Ren) | Yunnan (n=37) versus Sichuan (n=15) | 2.95784 (P = 0.0010**) | 6.04114 (P = 0.0010**) | 0.70417 (P = 0.0390*) | 0.22590 | 1.71 |
| | Yunnan (n=37) versus Guangxi (n=5) | 3.03696 (<i>P</i> = 0.0050**) | 5.67271 (P = 0.0060**) | 0.98810 (<i>P</i> = 0.0000***) | 0.30921 | 1.12 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 2.61738 (P = 0.0000***) | 3.79147 (<i>P</i> = 0.0000***) | 1.00000 (P = 0.0010**) | 0.65776 | 0.26 |
| C4 (SD) | Yunnan (n = 37) versus Sichuan (n = 15) | 2.88070 (P = 0.0100*) | 6.08731 (<i>P</i> = 0.0000***) | 0.76122 (P = 0.0020**) | 0.04323 | 11.07 |
| | Yunnan (n=37) versus Guangxi (n=5) | 2.67718 (P = 0.1030 ns) | 5.71561 (<i>P</i> = 0.0420*) | 0.95238 (P = 0.0000***) | 0.18857 | 2.15 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 2.84760 (<i>P</i> = 0.0300*) | 3.99473 (P = 0.0020**) | 1.00000 (<i>P</i> = 0.0000***) | 0.31522 | 1.09 |
| V1 (CP) | Yunnan (n = 37) versus Sichuan (n = 15) | 3.86409 (P = 0.0010**) | 6.02408 (P = 0.0010**) | 0.74679 (<i>P</i> = 0.0240*) | 0.24873 | 1.51 |
| | Yunnan (n=37) versus Guangxi (n=5) | 3.98796 (<i>P</i> = 0.0050**) | 5.68138 (P = 0.0100*) | 0.95238 (<i>P</i> = 0.0000***) | 0.28916 | 1.23 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 3.38331 (<i>P</i> = 0.0000***) | 3.91271 (<i>P</i> = 0.0000***) | 0.95000 (P = 0.0040**) | 0.61660 | 0.31 |
| V2 (Pre-CP) | Yunnan (n = 37) versus Sichuan (n = 15) | 3.10102 (P = 0.0000***) | 6.01641 (<i>P</i> = 0.0000***) | 0.65865 (P = 0.1240 ns) | 0.24375 | 1.55 |
| | Yunnan (n=37) versus Guangxi (n=5) | 3.16259 (P = 0.0200*) | 5.71082 (P = 0.0310*) | 0.97616 (<i>P</i> = 0.0000***) | 0.23296 | 1.65 |
| | Sichuan (n = 15) versus Guangxi (n = 5) | 2.70925 (P = 0.0000***) | 3.85143 (P = 0.0000***) | 0.95000 (P = 0.0010**) | 0.61931 | 0.31 |
| βC1 | Yunnan (n=92) versus Sichuan (n=7) | 3.36093(<i>P</i> = 0.0000***) | 7.43235 (<i>P</i> = 0.0000***) | 0.98990 (<i>P</i> = 0.0000***) | 0.28495 | 1.25 |
| | Yunnan (n=92) versus Guangxi (n=10) | 3.33741 (<i>P</i> = 0.0000***) | 7.48659 (P = 0.0020**) | 0.98039 (<i>P</i> = 0.0000***) | 0.18406 | 2.22 |
| | Sichuan (n = 7) versus Guangxi (n = 10) | 2.94481 (P = 0.0000***) | 3.44883 (<i>P</i> = 0.0000***) | 0.94118 (P = 0.0010**) | 0.41920 | 0.69 |
| βV1 | Yunnan (n=92) versus Sichuan (n=7) | 1.77660 (<i>P</i> = 0.0020**) | 7.46902 (P = 0.0030**) | 0.97306 (<i>P</i> = 0.0000***) | 0.20215 | 1.97 |
| | Yunnan (n=92) versus Guangxi (n=10) | 1.70877 (<i>P</i> = 0.0000***) | 7.51238 (P = 0.0010**) | 0.90083 (P = 0.0000***) | 0.16380 | 2.55 |
| | Sichuan (n = 7) versus Guangxi (n = 10) | 2.52568 (<i>P</i> = 0.0000***) | 3.50130 (<i>P</i> = 0.0000***) | 0.88235 (<i>P</i> = 0.0040***) | 0.36354 | 0.88 |

Probability (P value) obtained by the permutation (PM) test with 1000 replicates. Fst = Standardized variance of allele frequencies across populations, Nm = Migration rate, C1 = Replication-associated protein, C2 = Transcriptional activator protein, C3 = Replication enhancer protein, C4 = Disease symptom determinants, V1 = Coat protein, V2 = Pre-coat protein, β C1 = Movement protein, β V1 = Promotion of virulence during the infection. * indicate 0.01 < P < 0.05, ** indicate P < 0.01, and ns = not significant (P > 0.05). Fst > 0.33 or Nm < 1 suggests infrequent gene flow; Fst < 0.33 or Nm > 1 suggests frequent gene flow

Molecular variation between different begomoviruses

Based on the current results, a mutation rate of 2.2×10^{-4} was found in 21 out of the 84 sequences obtained from the TYLCCNV population in N. glutinosa, while a mutation rate of 4.3×10^{-4} was found in 27 out of 65 sequences collected from N. benthamiana (Additional file 1: Table S7). Interestingly, the mutated bases were mainly occurred in the C1 and C4 regions (60.0%, 15 of 25 in N. glutinosa; 47.4%, 18 of 38 in N. benthamiana) (Fig. 3a). Analysis of the distribution of population mutations in different natural hosts of TYLCCNV revealed that mutant bases in YN48 (100%, 11 of 11), SC65 (100%, 8 of 8), and YM5 (100%, 6 of 6) populations were similarly found to be distributed within the C1 and C4 regions (Fig. 3b). Meanwhile, the base mutation types of TYL-CCNV were analyzed and the mutation types $(T \rightarrow A)$, $G \rightarrow T$, $T \rightarrow C$, $C \rightarrow T$, and $A \rightarrow G$) were identified in both N. glutinosa and N. benthamiana in the indoor populations of TYLCCNV, whereas the mutation types $(G \rightarrow T)$ and $C \rightarrow T$) were found in the natural populations of TYLCCNV (Solanum lycopersicum L., N. tabacum, and Malvastrum coromandelianum Garcke) (Fig. 3c, d).

The results of the indoor tests on TYLCCNB showed that 63 of 68 sequences obtained from N. glutinosa were mutated, with a mutation rate of 1.5×10^{-3} . In contrast, 63 of 67 sequences obtained from N. benthamiana were also mutated, with a mutation rate of 1.2×10^{-3} (Additional file 1: Table S8). Remarkably, the mutated bases were predominantly distributed in the A-rich region (67.0%, 89 of 133 in N. glutinosa; 82.9%, 87 of 105 in N. benthamiana) (Fig. 4a). Similarly, base mutations in natural populations of HG230 (42.1%, 8 of 19), YN48 (47.6%, 10 of 21), SC65 (57.1%, 16 of 28), and YM5 (73.9%, 17 of 23) were found to be distributed within the A-rich region (Fig. 4b). Meanwhile, mutation types $(T \rightarrow G, C \rightarrow A)$, $A \rightarrow T$, $G \rightarrow A$, $C \rightarrow T$, and $A \rightarrow G$) were observed both indoor populations of TYLCCNB in N. glutinosa and N. *benthamiana*, whereas the mutation types were $(G \rightarrow A)$ and $C \rightarrow T$) recorded both in *N. tabacum* and *Solanum lycopersicum* natural population (Fig. 4c, d).

| Table 4 | Genetic parameters, | neutrality test, an | d selection pre | essure on 57 | Tomato yellow | leaf curl Chi | na virus (TYLCCNV) | isolates and |
|---------|------------------------|---------------------|-----------------|--------------|----------------|---------------|--------------------|--------------|
| 109 Tom | ato yellow leaf curl C | hina virus betasate | ellite (TYLCCNF | 3) subpopula | tions based or | n geographic | origin | |

| Genomic region ^a | Population | N ^b | Haplotype diversity | Nucleotide diversity | Tajima's D ^c | dN/dS ^d |
|-----------------------------|------------------|----------------|---------------------|-----------------------|-------------------------|--------------------|
| C1 (Rep) | All (n=57) | 54 | 0.998±0.004 | 0.10369±0.00959 | -0.40650 (ns) | 0.42 |
| | Yunan (n = 37) | 36 | 0.998 ± 0.007 | 0.09762±0.01258 | – 0.55573 (ns) | 0.42 |
| | Sichuan (n = 15) | 14 | 0.990 ± 0.028 | 0.12311±0.01582 | 0.39777 (ns) | 0.44 |
| | Guangxi (n=5) | 5 | 1.000 ± 0.126 | 0.01971 ± 0.00305 | -0.38767 (ns) | 0.44 |
| C2 (TrAP) | All (n=57) | 49 | 0.993 ± 0.005 | 0.06435 ± 0.00392 | – 1.38557 (ns) | 0.37 |
| | Yunan (n = 37) | 32 | 0.991 ± 0.009 | 0.06336 ± 0.00420 | – 0.97510 (ns) | 0.37 |
| | Sichuan (n = 15) | 14 | 0.990 ± 0.028 | 0.04889 ± 0.01240 | – 1.47299 (ns) | 0.41 |
| | Guangxi (n=5) | 5 | 1.000 ± 0.126 | 0.02519 ± 0.00373 | -0.25279 (ns) | 0.28 |
| C3 (Ren) | All (n=57) | 44 | 0.987 ± 0.007 | 0.05400 ± 0.00454 | – 1.55238 (ns) | 0.39 |
| | Yunan (n = 37) | 33 | 0.992 ± 0.009 | 0.07747 ± 0.00327 | -0.42140 (ns) | 0.33 |
| | Sichuan (n = 15) | 13 | 0.971 ± 0.039 | 0.04846 ± 0.01419 | – 1.33768 (ns) | 0.27 |
| | Guangxi (n=5) | 5 | 1.000 ± 0.126 | 0.02815 ± 0.00600 | -0.37214 (ns) | 0.23 |
| C4 (SD) | All (n=57) | 41 | 0.970 ± 0.014 | 0.12779±0.01968 | 0.38725 (ns) | 1.25 |
| | Yunan (n = 37) | 27 | 0.943 ± 0.032 | 0.11601 ± 0.02509 | -0.11698 (ns) | 1.29 |
| | Sichuan (n = 15) | 11 | 0.952 ± 0.040 | 0.17128 ± 0.03323 | 1.29297 (ns) | 1.29 |
| | Guangxi (n=5) | 4 | 0.900 ± 0.161 | 0.01565 ± 0.00321 | 0.46060 (ns) | 3.10 |
| V1 (CP) | All (n=57) | 55 | 0.999 ± 0.004 | 0.09616 ± 0.00292 | -0.40929 (ns) | 0.09 |
| | Yunan (n = 37) | 35 | 0.997 ± 0.007 | 0.09282 ± 0.00498 | -0.41862 (ns) | 0.09 |
| | Sichuan (n = 15) | 15 | 1.000 ± 0.024 | 0.06206 ± 0.01684 | -0.88493 (ns) | 0.09 |
| | Guangxi (n = 5) | 5 | 1.000 ± 0.126 | 0.03826 ± 0.00820 | -0.52106 (ns) | 0.07 |
| V2 (Pre-CP) | All (n=57) | 51 | 0.987 ± 0.010 | 0.10064 ± 0.00497 | -0.57899 (ns) | 0.17 |
| | Yunan (n = 37) | 31 | 0.968 ± 0.022 | 0.09765 ± 0.00865 | -0.73616 (ns) | 0.19 |
| | Sichuan (n = 15) | 15 | 1.000 ± 0.024 | 0.06526 ± 0.01677 | -0.69765 (ns) | 0.15 |
| | Guangxi (n = 5) | 5 | 1.000 ± 0.126 | 0.03448 ± 0.00710 | 0.30976 (ns) | 0.07 |
| βC1 | All (n=109) | 70 | 0.982 ± 0.005 | 0.11688 ± 0.00325 | -0.59699 (ns) | 2.68 |
| | Yunan (n=92) | 55 | 0.975 ± 0.007 | 0.11421 ± 0.00307 | -0.09960 (ns) | 2.79 |
| | Sichuan (n = 7) | 7 | 1.000 ± 0.076 | 0.06736 ± 0.01154 | -0.00819 (ns) | 0.20 |
| | Guangxi (n = 10) | 9 | 0.978 ± 0.054 | 0.08472±0.02861 | – 1.45465 (ns) | 0.23 |
| βV1 | All (n=109) | 31 | 0.922±0.012 | 0.12021±0.23330 | – 1.55761 (ns) | 0.35 |
| | Yunan (n=92) | 23 | 0.904 ± 0.016 | 0.10909 ± 0.11022 | -0.03154 (ns) | 0.31 |
| | Sichuan (n=7) | 7 | 1.000 ± 0.076 | 0.07423 ± 0.07003 | 0.34393 (ns) | 0.36 |
| | Guangxi (n = 10) | 8 | 0.956 ± 0.076 | 0.09820 ± 0.14104 | – 1.49921 (ns) | 0.46 |

^a C1 = Replication-associated protein, C2 = Transcriptional activator protein, C3 = Replication enhancer protein, C4 = Disease symptom determinants, V1 = Coat protein, V2 = Pre-coat protein, βC1 = Movement protein, βV1 = Promotion of virulence during the infection

^b Number of haplotypes

^c ns = not significant (P > 0.05)

^d Nonsynonymous/synonymous ratio (dN/dS) < 1 (negative selection), dN/dS = 1 (neutral evolution), and dN/dS > 1 (positive selection)

In order to further clarify the betasatellite population variation of different begomoviruses, indoor inoculation tests were conducted on TbCSB and MYVB populations, respectively. Mutated bases were mainly concentrated in the A-rich region (66.7%, 6 of 9 in *N. glutinosa*; 24.5%, 13 of 53 in *N. benthamiana*) for TbCSB population (Additional file 2: Figure S4a and Additional file 1: Table S9). While in the MYVB population, the mutated bases were also predominantly occurred in the A-rich region (91.8%, 112 of 122 in *N. glutinosa*; 79.4%, 143 of 180 in *N. benthamiana*) (Additional file 2: Figure S4b and Additional file 1: Table S10). The mutation types $(T \rightarrow A, G \rightarrow T, G \rightarrow C, A \rightarrow T, and A \rightarrow G)$ were observed in TbSCB population collected from both *N. glutinosa* and *N. benthamiana*, whereas mutation types $(G \rightarrow T, C \rightarrow A, A \rightarrow T, G \rightarrow A, C \rightarrow T, and A \rightarrow G)$ were found in MYVB populations both in *N. tabacum* and *Solanum lycopersicum* (Additional file 2: Figure S4c, d).



Fig. 3 a Distribution of mutations in IR-C1 regions of TYLCCNV populations in *N. glutinosa* and *N. benthamiana*, and **b** distribution of mutations in IR-C1 regions of TYLCCNV in field populations. **c** The analysis of genome variation of TYLCCNV populations in *N. glutinosa* and *N. benthamiana*, and **d** genome variation of TYLCCNV populations in field populations.



Fig. 4 a Distribution of mutations in SCR-βC1 regions of TYLCCNB populations in *N. glutinosa* and *N. benthamiana*, and **b** distribution of mutations in SCR-βC1 regions of TYLCCNB in field populations. **c** The analysis of genome variation of TYLCCNB populations in *N. glutinosa* and *N. benthamiana*, and **d** genome variation of TYLCCNB populations in field populations.

Discussion

Begomoviruses are highly notorious for affecting many plants worldwide. Researchers in China have identified

99 species of begomoviruses with 1651 isolates, extensively spread over 32 administrative regions at the provincial level (Li et al. 2022). However, the knowledge regarding the distribution of TYLCCNV, an important monopartite begomovirus in China is still limited. Many factors, such as virus transmission through international trade, changes in vector populations, genetic recombination, novel farming practices, and variations in weather patterns, have been identified as potential drivers for the emergence of viral outbreaks (Jones 2009). Here, we examined 168 samples, including 67 TYLCCNV positive samples and 46 TYLCCNB positive samples. Notably, TYLCCNB was not detected in 31.3% (21 out of 67) of the TYLCCNV positive samples, due to the fact that TYLCCNV might be accompanied by heterologous satellites, such as TbCSB (Qing and Zhou 2009; Zhou 2013). In addition, we studied the occurrence and distribution of TYLCCNV and betasatellite (TYLCCNB) isolates, predominantly distributed in the southwestern region of China (Yunnan, Sichuan, and Guangxi provinces).

Previous studies showed that geographic isolation is an important factor in the genetic structure of viral populations (Sun et al. 2021). Our results suggest that TYLCCNV/TYLCCNB populations are associated with geographic origins in China, independent of host plants. Unexpectedly, Guangxi isolates and Sichuan isolates among TYLCCNB isolates were not clustered into one subgroup. In addition, low levels of gene flow and significant genetic differentiation between different ORF regions were found in TYLCCNV/TYLCCNB populations of Sichuan and Guangxi isolates, suggesting that geographic isolation factors are responsible for influencing TYLCCNV/TYLCCNB population structure. These results are consistent with SCSMV and rice stripe mosaic virus geographical driven adaptation (Liang et al. 2016; Yang et al. 2018). Interestingly, it was also found that TYLCCNB, as a virus satellite, tends to co-evolve with its helper virus TYLCCNV.

The ability of viruses to evolve and become more environmentally adapted is largely dependent on recombination (Lin et al. 2014; Lefeuvre et al. 2019). In the present study, 12 recombination events were detected in TYL-CCNV isolates. Remarkably, 8 recombination events had recombination breakpoints distributed in the coding regions of the C1 and C4 genes. Similarly, 10 recombination events were detected in TYLCCNB isolates, of which 6 recombination breakpoints were distributed in the coding region of the $\beta C1$ gene. In addition, recombination events were also supported by split network analysis. Notably, the presence of cross-infection in TYLCCNV and TYLCCNB isolates may promote recombination between different viral species or isolates and facilitate the generation of new viral strains or variants (Moradi and Mehrvar 2019).

Natural selection is an important evolutionary mechanism and a key of driver variation in viral populations. The purifying selection can effectively drive variation in viral populations by increasing the rate of elimination of genetically deleterious mutations and the formation of a stable population genetic structure (Moradi and Mehrvar 2019). Here, the *C4* gene was found to be under positive selection, whereas the five TYLCCNV genes (*C1, C2, C3, V1,* and *V2*) and the β *V1* gene in TYLCCNB were under negative or purifying selection, suggesting that they might play an important role in the adaptation of TYLCCNV and TYLCCNB to environmental changes. Moreover, the neutrality tests of different ORF regions in Yunnan isolates were negative, indicating that the population was in a state of expansion.

Begomoviruses genome sequence mutations may lead to amino acid changes that affect virus particle formation, replication, and host range as well as virus-induced symptoms formation (Yaakov et al. 2011). Based on previous studies, the TYLCCNV population is a quasispecies, much like RNA virus populations (Ge et al. 2007). In this study, we analyzed the genetic structure and variability of TYLCCNV and TYLCCNB populations in natural and indoor infections with different hosts. It was found that TYLCCNB $(1.2 \times 10^{-3} - 1.5 \times 10^{-3})$ had a higher mutation rate during viral genome replication compared to TYL-CCNV $(2.2 \times 10^{-4} - 4.3 \times 10^{-4})$, which might be related to the fact that TYLCCNB is a determinant of viral symptom formation (Cui et al. 2004; Saunders et al. 2004; Briddon and Stanley 2006). Furthermore, the distribution of base mutations showed that the mutations in TYLCCNV were concentrated in C1 and C4, indicating that it was less selectively constrained. Meanwhile, the distribution of base mutations in the three begomoviruses (TYL-CCNB, TbCSB, and MYVB) revealed that the mutations were mainly concentrated in the A-rich region. In future studies, we plan to reintroduce these mutations into the TYLCCNV and TYLCCNB genomes to elucidate their functional consequences.

Conclusions

The data from this study involves the molecular genetic variation, population genetic structure, and evolutionary drivers of the impact of the TYLCCNV/TYLCCNB isolates at the level of viral genes or gene fragments as well as at the level of the genome. We found that the TYL-CCNV population was mainly distributed in south-western regions of China. Genetic diversity analysis revealed a co-evolutionary relationship between TYLCCNV and TYLCCNB isolates. In addition, the molecular variation of TYLCCNV and its accompanying satellite TYLCCNB was characterized in both indoor and natural field populations, and the coding regions of the *C1* and *C4* genes were found to be the major mutated regions in TYL-CCNV, which were hypothesized to be possibly related to their functions (viral replication as well as symptom formation related). Indoor population analysis of different companion satellites TYLCCNB, TbCSB, and MYVB and natural population analysis of TYLCCNB revealed that the A-rich region was the main mutated region. It was also shown that TYLCCNB has a higher mutation frequency and higher mutation rate than TYLCCNV. The present research results will provide important information for epidemiological studies and reliable diagnostic methods for healthy tomato programs.

Methods

Sample collection and virus detection

During 2008-2012, 168 samples were collected from 11 crops (N. tabacum, Carica papaya, Ipomoea aquatica, Capsicum annuum, Lactuca sativa var. angustata, Solanum lycopersicum, Beta vulgaris, Ipomoea batatas, Glycine max, Vigna unguiculata, and Phaseolus vulgaris) and 11 weeds (Ageratum conyzoides, Malvastrum coromandelianum Garcke, Alternanthera philoxeroides Griseb, Mirabilis jalapa, Malva sinensis Cavan, Solanum nigrum, Sigesbeckia orientalis, Datura stramonium, Petunia hybrida (Hook.) E. Vilm, Dysphania ambrosioides, and Emilia sonchifolia) that showed typical symptoms of begomoviral infection in different regions of three provinces (Yunnan, Sichuan, and Guangxi) in China. Fresh leaf genomic DNA was extracted by the conventional CTAB method (Yan et al. 2008). Subsequently, TYLCCNV and TYLCCNB sequences were amplified using specific primers and whole genome sequence information was obtained by sequencing (Additional file 1: Table S11). Information and locations of the collected samples are shown in Additional file 1: Table S1.

Sequence alignment and phylogenetic analysis

The whole genome sequences and six ORFs regions (V1, V2, C1, C2, C3, and C4) of 57 TYLCCNV isolates (23 from this study and 34 from GenBank) were used for sequence analysis. Among them, 37 from Yunnan, 5 from Guangxi, and 15 from Sichuan (Additional file 1: Table S2). Likewise, the whole genome sequences and $\beta C1$ and $\beta V1$ gene regions of 109 TYLCCNB isolates (10 in this study and 99 GenBank) were sequence analyzed, of which 92 were from Yunnan, 10 from Guangxi, and 7 from Sichuan (Additional file 1: Table S3). The full TYL-CCNV/TYLCCNB whole genome sequences were compared multiple times using the CLUSTALW algorithm in MEGA11 (Kumar et al. 2016). Phylogenetic trees were constructed through neighbour-joining (NJ) method in MEGA11 using the two aligned nucleotide datasets of the TYLCCNV/TYLVVNB genome sequences, respectively. Pairwise identities between and among phylogenetic groups were calculated using BioEdit version 7.1.9 and SDT v1.2 (Hall 1999).

Test for recombination signal

The TYLCCNV and TYLCCNB whole genome sequence datasets were tested for the presence of recombination signals using SplitsTree4 v.4.13.1 (Huson and Bryant 2006). The parsing phylogenetic network was constructed based on 1000 bootstrap pseudo-replicates to validate the statistical confidence of specific nodes. Subsequently, evidence of recombination was further analyzed using seven programs (RDP, GENECONV, MaxChi, Chimaera, BOOTSCAN, SISCAN, and 3Seq) in the software RDP v.4.16 (Martin et al. 2015). A putative recombination analysis was considered significant if it was supported by at least four of the seven different methods and the associated *P* value was less than 1×10^{-6} (Chinnaraja et al. 2013; Lin et al. 2014).

Calculation of population genetic parameters

Based on phylogenetic groups (geographic distribution) of TYLCCNV and TYLCCNB, population genetic parameters were calculated for different ORFs regions using the software DnaSP version 5.10.01 (Librado and Rozas 2009). InDel analyses were calculated manually based on the aligned sequences of TYLCCNV and TYLCCNB isolates. To investigate the extent and distribution of genetic variation among the 57 TYLCCNV isolates and 109 TYL-CCNB isolates, nucleotide diversity was estimated based on the average number of nucleotide differences per site, with a sliding window adjusted to 100 nt and a step size of 25 nt. Meanwhile, haplotypic diversity (h) and nucleotide diversity (π) were calculated for the TYLCCNV and TYLCCNB isolates, respectively (Nei 1987). Neutrality tests were examined using Tajima's D method (Tajima 1989). Genetic differentiation between TYLCCNV and TYLCCNB populations was assessed using three ranking statistics, Ks*, Z*, and Snn (Hudson et al. 1992; Hudson 2000). The null hypothesis of no genetic differentiation was rejected if the *P* value < 0.05. In addition, the degree of gene flow between the TYLCCNV and TYLCCNB populations was analyzed using the standardized variance of Fst (subpopulation fusion index) and Nm (number of migrants) (Sun et al. 2021). The gene flow was considered to have occurred infrequently if |Fst|>0.33 or |Nm| < 1. When |Fst| < 0.33 or |Nm| > 1, there is a high frequency of gene flow. To assess the strength of selection pressure in TYLCCNV and TYLCCNB different ORFs regions, we calculated the ratio of non-synonymous (dN) to synonymous (dS) substitution rates ($\omega = dN/dS$) using DnaSP version 5.10.01 software.

Molecular variation between different viruses

Infectious clones of TYLCCNV and TYLCCNVB were mixed in equal proportions at the same concentration with TYLCCNV-Y10 (Y10A and Y10B) isolates, TbCSV-Y35 (Y35A and Y35β) isolates, and MYVB-Y47 (Y47A and Y47β) isolates, respectively. Subsequently, they were inoculated into the phloem of N. benthamiana and N. glutinosa. The TYLCCNV-Y10 isolate and TbCSV-Y35 (Y35A and Y35B) inoculated plants leaves were collected at 60 dpi and 120 dpi for N. benthamiana and N. glutinosa, respectively, whereas MYVB-Y47 (Y47A and Y47 β) isolate inoculated plants leaves were collected at 30 dpi, 60 dpi, and 120 dpi. N. benthamiana and N. glutinosa were tested separately using specific primers. Sequence splicing and processing were performed with the aid of DNAStar software (Version 7.0, Madison, Wis., USA), and multiple sequence comparisons were performed using the DNAStar Clustal V method (Jia et al. 2008). Accordingly, the molecular variant sites were identified by comparison with Y10A, Y10B, Y35A, Y35B, Y47A, and Y47B. The TYL-CCNV and TYLCCNVB sequences of the viruses were obtained from naturally infected TYLCCNV tomato, tobacco, and malvastrum, and the mutant clones of all populations of TYLCCNV-Y10 (Y10A and Y10β) were calculated by comparing them with the primary sequences by following the method of Ge et al., respectively (Ge et al. 2007). The percentage (ratio of the total number of mutated clones to the total number of clones) and mutation frequency (ratio of the total number of mutated bases to the total number of sequenced bases) were calculated for all populations of TYLCCNV-Y10 (Y10A and Y10β) as an indicator of the genetic diversity of the viral populations and the level of population variability.

Abbreviations

| C1 (Rep) | Replication-associated protein |
|-------------|---|
| C2 (TrAP) | Transcriptional activator protein |
| C3 (Ren) | Replication enhancer protein |
| C4 (SD) | Disease symptom determinants |
| DnaSP | DNA sequence polymorphism |
| MEGA7 | Molecular evolutionary genetics analysis |
| MYVB | Malvastrum yellow vein betasatellite |
| NCBI | National Center for Biotechnology Information |
| ORFs | Open read frames |
| PTGS | Post-transcriptional gene silencing |
| RDP4 | Recombination detection program version |
| SDT | Sequence demarcation tool |
| TbCSB | Tobacco curly shoot betasatellite |
| TGS | Transcriptional gene silencing |
| TYLCCNV | Tomato yellow leaf curl China virus |
| TYLCCNB | Tomato yellow leaf curl China betasatellite |
| V1 (CP) | Coat protein |
| V2 (Pre-CP) | Pre-coat protein |
| βC1 | Movement protein |
| βV1 | Promotion of virulence during the infection |

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42483-025-00312-w.

Additional file 1: Table S1. PCR detection of Tomato vellow leaf curl China virusin typical gemini viruses samples collected in China between 2008 and 2012. Table S2. Information for Tomato yellow leaf curl China virus isolates used in this study. Table S3. Information for Tomato yellow leaf curl China beta satellite was used in this study. Table S4. Nucleotide sequence identities in Tomato yellow leaf curl China virus genomes based on geographic populations. Table S5. Nucleotide sequence identities in Tomato yellow leaf curl China virus betasatellite based on geographic populations. Table S6. Insertions/Deletions events in individual proteins from 57 Tomato yellow leaf curl China virus isolates and 109 Tomato yellow leaf curl China virus betasatellite. Table S7. Genetic structure and variation of TYLCCNV populations. Table S8. Genetic structure and variation of TYLCCNB populations. Table S9. Genetic structure and variation of TbCSB populations. Table S10. Genetic structure and variation of MYVB populations. Table S11. Primers were used for PCR detection of DNA viruses in 168 samples.

Additional file 2: Figure S1. Neighbor-joiningphylogenetic tree constructed using MEGA11 based on a 57 Tomato yellow leaf curl China virus isolates and b 66 Tomato yellow leaf curl China virus betasatellite. Different colours are represented host. Figure S2. Split network analysis of a 57 Tomato yellow leaf curl China virus isolates and b 109 Tomato yellow leaf curl China virus betasatellite isolates. Figure S3. Nucleotide sequence identities in of a 57 Tomato yellow leaf curl China virus whole-genome sequence and b 109 Tomato yellow leaf curl China virus betasatellite whole-genome sequence. Figure S4. a Distribution of mutations in SCR-βC1 regions of TbCSB populations in *N. glutinosa* and *N. benthamiana*, and b distribution of mutations in SCR-βC1 regions of MYVB in field populations. c The analysis of genomic variation of TbCSB populations in *N. glutinosa* and *N. benthamiana*, and d genomic variation of MYVB

Acknowledgements

We thank Dr. Muhammad Ayaz from Anhui Academy of Agricultural Sciences, Hefei, China for critically reading the manuscript.

Author contributions

LQ and WH conceived the manuscript. JY and YX wrote the manuscript. YL, MZ, XY, and CZ revised the manuscript. YW and HH collected experimental samples and analyzed the data. All authors read and approved the final manuscript.

Funding

This work was supported by Innovation Research 2035 Pilot Plan of Southwest University (SWU-XDZD22002) and National Key Research and Development Program (2022YFC2602200, 2021YFC2600404).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 September 2024 Accepted: 13 January 2025 Published online: 15 April 2025

References

- Briddon R, Stanley J. Subviral agents associated with plant single-stranded DNA viruses. Virology. 2006;344:198–210.
- Briddon RW, Pinner MS, Stanley J, Markham PG. Geminivirus coat protein gene replacement alters insect specificity. Virology. 1990;177:85–94. https://doi. org/10.1016/0042-6822(90)90462-z.
- Chinnaraja C, Viswanathan R, Karuppaiah R, Bagyalakshmi K, Malathi P, Parameswari B. Complete genome characterization of *Sugarcane yellow leaf virus* from India: evidence for RNA recombination. Eur J Plant Pathol. 2013;135:335–49. https://doi.org/10.1007/s10658-012-0090-6.
- Cui XF, Tao XR, Xie Y, Fauquet CM, Zhou XP. A DNAβ associated with *Tomato Yellow Leaf Curl China Virus* is required for symptom induction. J Virol. 2004;78:13966–74. https://doi.org/10.1128/jvi.78.24.13966-13974.2004.
- Domingo E, Baranowski E, Ruiz-Jarabo CM, Martín-Hernández AM, Sáiz JC, Escarmís C. Quasispecies structure and persistence of RNA viruses. Emerg Infect Dis. 1998;4:521–7. https://doi.org/10.3201/eid0404.980402.
- García-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. Annu Rev Phytopathol. 2001;39:157–86. https://doi. org/10.1146/annurev.phyto.39.1.157.
- Ge LM, Zhang JT, Zhou XP, Li HY. Genetic structure and population variability of tomato yellow leaf curl China virus. J Virol. 2007;81:5902–7. https://doi. org/10.1128/jvi.02431-06.
- Glick E, Zrachya A, Levy Y, Mett A, Gidoni D, Belausov E, et al. Interaction with host SGS3 is required for suppression of RNA silencing by *tomato yellow leaf curl virus* V2 protein (vol 105, pg 157, 2007). Proc Natl Acad Sci U S A. 2009;106:4571. https://doi.org/10.1073/pnas.0900927106.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Proceeding of the nucleic acids symposium series. Oxford; 1999. p. 95–8.
- Hanley-Bowdoin L, Settlage SB, Robertson D. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Mol Plant Pathol. 2004;5:149–56. https://doi.org/10.1111/j.1364-3703.2004.00214.X.
- Harrison BD, Robinson DJ. Natural genomic and antigenic variation in whiteflytransmitted geminiviruses (Begomoviruses). Annu Rev Phytopathol. 1999;37:369–98. https://doi.org/10.1146/annurev.phyto.37.1.369.
- Hu T, Song Y, Wang YQ, Zhou XP. Functional analysis of a novel βV1 gene identified in a geminivirus betasatellite. Sci China-Life Sci. 2020;63:688–96. https://doi.org/10.1007/s11427-020-1654-x.
- Hudson RR. A new statistic for detecting genetic differentiation. Genetics. 2000;155:2011–4.
- Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. Mol Biol Evol. 1992;9:138–51.
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23:254–67.
- Jia W-Z, Li Z, Zhao L, Lun Z-R. Genetic variation and clustal analysis of Trichomonas vaginalis cysteine proteases. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi Chin J Parasitol Parasit Dis. 2008;26:191–6, 202.
- Jones RAC. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Res. 2009;141:113–30. https://doi.org/10. 1016/j.virusres.2008.07.028.
- King AM, Lefkowitz E, Adams MJ, Carstens EB. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Amsterdam: Elsevier; 2011.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4. https:// doi.org/10.1093/molbev/msw054.
- Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A. Evolution and ecology of plant viruses. Nat Rev Microbiol. 2019;17:632–44.
- Li FF, Yang XL, Bisaro DM, Zhou XP. The βC1 protein of geminivirus-betasatellite complexes: a target and repressor of host defenses. Mol Plant. 2018;11:1424–6. https://doi.org/10.1016/j.molp.2018.10.007.
- Li FF, Qiao R, Wang ZQ, Yang XL, Zhou XP. Occurrence and distribution of geminiviruses in China. Sci China-Life Sci. 2022;65:1498–503. https://doi.org/10.1007/s11427-022-2125-2.
- Liang SS, Alabi OJ, Damaj MB, Fu WL, Sun SR, Fu HY, et al. Genomic variability and molecular evolution of Asian isolates of sugarcane streak mosaic virus. Adv Virol. 2016;161:1493–503. https://doi.org/10.1007/ s00705-016-2810-2.

- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2. https://doi.org/10. 1093/bioinformatics/btp187.
- Lin YH, Gao SJ, Damaj MB, Fu HY, Chen RK, Mirkov TE. Genome characterization of sugarcane yellow leaf virus from China reveals a novel recombinant genotype. Adv Virol. 2014;159:1421–9. https://doi.org/10.1007/ s00705-013-1957-3.
- Liu YL, Cai JH, Li DL, Qin BX, Tian B. Chinese tomato yellow leaf curl virus—a new species of geminivirus. Sci China Ser C-Life Sci. 1998;41:337–43. https://doi.org/10.1007/bf02882731.
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015. https://doi.org/10.1093/ve/vev003.
- Moradi Z, Mehrvar M. Genetic variability and molecular evolution of Bean common mosaic virus populations in Iran: comparison with the populations in the world. Eur J Plant Pathol. 2019;154:673–90. https://doi.org/10. 1007/s10658-019-01690-6.
- Mubin M, Amin I, Amrao L, Briddon RW, Mansoor S. The hypersensitive response induced by the V2 protein of a monopartite begomovirus is countered by the C2 protein. Mol Plant Pathol. 2010;11:245–54. https://doi.org/10.1111/j.1364-3703.2009.00601.x.
- Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987. p. 512.
- Qing L, Zhou XP. Trans-replication of, and competition between, DNA β satellites in plants inoculated with *Tomato yellow leaf curl China virus* and *Tobacco curly shoot virus*. Phytopathology. 2009;99:716–20. https://doi. org/10.1094/phyto-99-6-0716.
- Ren YX, Tao XR, Li DW, Yang XL, Zhou XP. ty-5 Confers broad-spectrum resistance to geminiviruses. Viruses-Basel. 2022. https://doi.org/10.3390/v1408 1804.
- Rigden JE, Krake LR, Rezaian MA, Dry IB. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. Virology. 1994;204:847–50. https:// doi.org/10.1006/viro.1994.1606.
- Saunders K, Norman A, Gucciardo S, Stanley J. The DNA β satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (βC1). Virology. 2004;324:37–47. https://doi.org/10. 1016/j.virol.2004.03.018.
- Settlage SB, See RG, Hanley-Bowdoin L. Geminivirus C3 protein: replication enhancement and protein interactions. J Virol. 2005;79:9885–95. https:// doi.org/10.1128/jvi.79.15.9885-9895.2005.
- Sun SR, Chen JS, He EQ, Huang MT, Fu HY, Lu JJ, et al. Genetic variability and molecular evolution of maize yellow mosaic virus populations from different geographic origins. Plant Dis. 2021;105:896–903. https://doi.org/10. 1094/pdis-05-20-1013-re.
- Tajima F. The effect of change in population size on DNA polymorphism. Genetics. 1989;123:597–601.
- Vanitharani R, Chellappan P, Pita JS, Fauquet CM. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. J Virol. 2004;78:9487–98. https://doi. org/10.1128/jvi.78.17.9487-9498.2004.
- Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, et al. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Adv Virol. 2021;166:2633–48. https://doi.org/10.1007/s00705-021-05156-1.
- Xie Y, Zhao LL, Jiao XY, Jiang T, Gong HR, Wang B, et al. A recombinant begomovirus resulting from exchange of the C4 gene. J Gen Virol. 2013;94:1896–907. https://doi.org/10.1099/vir.0.053181-0.
- Yaakov N, Levy Y, Belausov E, Gaba V, Lapidot M, Gafni Y. Effect of a single amino acid substitution in the NLS domain of *Tomato yellow leaf curl virus*-Israel (TYLCV-IL) capsid protein (CP) on its activity and on the virus life cycle. Virus Res. 2011;158:8–11. https://doi.org/10.1016/j.virusres.2011. 02.016.
- Yan M, Wei G, Pan X, Ma H, Li W. A method suitable for extracting genomic DNA from animal and plant—modified CTAB method. Agric Sci Technol Hunan. 2008;9:39–41.
- Yang X, Chen B, Zhang T, Li ZB, Xu CH, Zhou GH. Geographic distribution and genetic diversity of rice stripe mosaic virus in Southern China. Front Microbiol. 2018. https://doi.org/10.3389/fmicb.2018.03068.

- Yin QY, Yang HY, Gong QH, Wang HY, Liu YL, Hong YG, et al. Tomato yellow leaf curl China virus: monopartite genome organization and agroinfection of plants. Virus Res. 2001;81:69–76. https://doi.org/10.1016/s0168-1702(01) 00363-x.
- Zhou XP. Advances in understanding begomovirus satellites. Annu Rev Phytopathol. 2013;51:357–81 (VanAlfen NK (ed)).
- Zhou XP, Xie Y, Tao XR, Zhang ZK, Li ZH, Fauquet CM. Characterization of DNAβ associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. J Gen Virol. 2003;84:237–47. https://doi.org/10.1099/vir.0.18608-0.