*J. Crop Prot. 2025, 14 (4): 405-423*

Research Article

# Micro-evolution of tomato yellow leaf curl virus (TYLCV)-IL populations from wild *Datura stramonium* plants in Iran, the center of TYLCV diversity

**Mohammadreza Hosseinzadeh[1, 2], Sajad Astaraki[1], Judith K. Brown[3] and Masoud Shams-Bakhsh[1*]**

1. Department of Plant Pathology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran.
2. Department of Plant Protection, Bojnourd Branch, Islamic Azad University, Bojnourd, Iran.
3. School of Plant Sciences, University of Arizona, Tucson, AZ 85721 USA.

**Abstract:** Five of the seven recognized phylogeographical tomato yellow leaf curl virus (TYLCV; *Begomovirus coheni*) strains occur in Iran, the proposed TYLCV center of diversification, including the globally invasive, exotic tomato yellow leaf curl virus-Israel (TYLCV-IL). Here, the phylogeny, population structure, and microevolutionary patterns were studied for TYLCV-IL populations infecting wild Datura stramonium plants growing near commercial tomato fields in Bojnourd, Iran, from which TYLCV-IL haplotypes had previously been characterized. Genome variation of TYLCV in *D. stramonium* was maintained primarily by natural selection on viral coding regions involved in replication and the non-coding intergenic region involved in regulation of replication and transcription. The C4 ORF harbored signatures of host adaptation consistent with its known functions in long-distance movement and host plant gene silencing. Recombination analysis revealed predicted interspecies recombinants in *D. stramonium*, in the C1 and C1/C4 coding regions and intergenic region (IR). The IR nucleotide substitution rate of $4.02 \times 10^{-4}$ per site was indicative of extensive genome plasticity. Further, the high genomic diversity of TYLCV isolates associated with 'wild-host' populations was attributed to 'within-host structuring', evident by the presence of unique variants in each host plant. Finally, recombination and population structure analyses suggest that mutation and interspecies recombination have contributed to the differentiation of TYLCV-IL populations and subsequent genetic isolation by host.

**Keywords:** Begomovirus, Nucleotide substitutions, Plant virus evolution, Recombination

## Introduction

Whitefly-transmitted geminiviruses contain a single-stranded DNA (ssDNA) genome and are classified in the genus *Begomovirus* (*Geminiviridae*) (Frischmuth *et al*., 1997). The begomoviral genome encodes six open reading frames (ORFs), some partially overlapping and organized bidirectionally on the viral and complementary strands. A non-coding regulatory sequence of ~300 nucleotides, located between the C1 and V2/V1 ORFs, is referred to

Handling Editor: Naser Safaie

*Corresponding authors: shamsbakhsh@modares.ac.ir
Received: 31 December 2025, Accepted: 07 January 2026
Published online: 07 June 2026

as the intergenic region (IR). It contains a conserved stem-loop structure and a nonanucleotide sequence, TAATATTAC, comprising the predicted loop, conserved in all geminiviruses. The nonanucleotide is nicked by the begomovirus-encoded replication-associated protein (Rep) between the thymine and adenine residues, respectively, to initiate rolling-circle replication (Laufs *et al*., 1995; Navot *et al*., 1991). In monopartite begomoviruses, the V1 and V2 ORFs are arranged on the viral sense strand and encode the coat protein (CP) and a viral movement protein (MP; BC1 for bipartite begomoviruses), respectively (Rojas *et al*., 2005). Four ORFs are encoded on the complementary sense strand, C1, C2, C3, and C4, and the last ORF is nested within the C1 ORF. They encode a Rep, a transcription activator (TrAP), and a replication enhancer (REn) protein, respectively, while the C4 protein is involved in virus-mediated suppression of host plant-mediated post-transcriptional gene silencing and in virus-mediated movement (Jupin *et al*., 1994; Laufs *et al*., 1995; Wartig *et al*., 1997).

Begomoviruses replicate in the nucleus, where host proof-reading DNA polymerases also reside (Gutierrez, 1999). Despite this, moderate to high nucleotide substitution rates have been reported for several begomoviruses, including the monopartite TYLCV and the DNA-A component of the bipartite East African cassava mosaic virus, which were estimated at $2.88 \times 10^{-4}$ and $1.60 \times 10^{-3}$, respectively (Duffy and Holmes, 2008, 2009). RNA-directed DNA methylation (RdDM), the recruitment of an error-prone DNA polymerase, and deamination of bases during replication have been proposed as explanations for the unanticipatedly high mutation rates in these ssDNA viruses, which can approach those observed for certain RNA viruses (Duffy and Holmes, 2008; Xia and Yuen, 2005). Accumulation of mutations and recombination are primary drivers of begomovirus evolution (Ge *et al.*, 2007); however, an upper limit or threshold that preserves essential coding and non-coding sequence functions is anticipated (Domingo and Perales, 2019). The monopartite tobacco leaf curl virus (TLCV) infecting *Eupatorium makioni* Kawahara & Yuhara var. oppositifolium (Koidzumi) Kawahara & Yuhara was shown to exhibit within-host variability manifest as 9-14 nucleotides (nts) per AC1 ORF per individual host plant (Ooi *et al*., 1997). Tomato plants infected with the monopartite tomato yellow leaf curl China virus harbored highly heterogeneous populations, as reflected by single-nucleotide polymorphisms in the IR, C4, and C1 ORFs (1396 nt) (Ge *et al*., 2007). Similarly, high variability has been reported for monopartite cotton leaf curl begomoviruses and betasatellites recovered from fifteen wild and cultivated *Gossypium* species (Nawaz-ul-Rehman *et al*., 2012). Finally, analysis of predicted recombinant genomes of TYLCV-tomato leaf curl Mayotte virus has shown that individual host plants can harbor multiple recombinants (Urbino *et al*., 2013).

During 1970 to 1980, begomoviruses emerged for the first time as damaging plant pathogens, a phenomenon thought to be exacerbated by monoculture crop expansion, and the increasingly widespread cultivation of genetically-uniform varieties that supported year-round populations of the whitefly vector *Bemisia tabaci* (Genn.) cryptic species group (Hemiptera, Aleyrodidae) (Brown and Bird, 1992). Most economically important cryptic species are polyphagous, may have a unique host range, and occupy distinct micro-climate niches (Paredes-Montero *et al*., 2021). During this time, begomoviruses underwent rapid diversification and speciation (Brown and Czosnek, 2002), with genome variation manifest as nucleotide substitutions, recombination, and/or component (bipartite) reassortment (Padidam *et al*., 1999; Silva *et al*., 2012). Further, begomoviruses previously unknown to infect cultivated plants were transmitted from endemic wild plant reservoirs to crop species by the whitefly vector (Silva *et al*., 2012). Even so, few studies have addressed the population structure of endemic begomoviruses associated with infected wild-cultivated plant counterparts (Sobrinho *et al*., 2014).

Begomoviruses represent either those first identified in tomato crops following a host jump from a wild reservoir (Castillo-Urquiza *et al*. 2007; Ribeiro *et al*., 2003), or had been initially recognized in symptomatic wild plant host(s) before their discovery in cultivated plants (Castillo-Urquiza *et al*., 2010). Compared to virus populations in genetically uniform cultivated plant species, viruses associated with wild plant populations exhibit greater genetic variability and can serve as reservoirs or 'melting pots' that support ongoing virus-host evolution (Cooper and Jones, 2006; Roossinck, 2019), while also providing populations potentially amenable to a 'host species jump' that can result in virus host range shifts. In one such scenario, the genome sequences of bean golden mosaic virus and Macroptilium yellow spot virus associated with the respective wild and cultivated legume host species were highly similar, providing an example in which population structure was minimally influenced by different host associations (Sobrinho *et al*., 2014). This observation may have been due to naturally occurring, frequent intra-host transmission of these begomovirus species by the whitefly vector. In contrast, comparisons of the potyvirus hardenbergia mosaic virus (HarMV) genome sequences recovered from wild and cultivated host counterparts indicated that each host-specific population harbored distinct signals of genome variability (Webster *et al*., 2007), revealing virus-host-specific influences on HarMV evolution (Czosnek and Laterrot, 1997).

The objective of this study was to investigate the variability among TYLCV-IL isolates recovered from individual wild *Datura stramonium* L. (Hosseinzadeh *et al*., 2014) and those previously characterized isolates infected crops and weeds from the Middle East. In the latter study, four monophyletic TYLCV-IL sister clades recovered from tomato plants grouped based on 'local-geography' (Hosseinzadeh *et al*., 2014). Here, TYLCV-IL genome sequences from a community of wild *D. stramonium* plants were subjected to phylogenetic and recombination analyses, and population structure was compared within and between individual plants. Finally, genome sequences of TYLCV-IL isolated from *D. stramonium* and cultivated tomato plants were compared. Of particular interest was to evaluate the extent of small-scale 'within-' and 'among'-host genome variability harbored by the respective wild and cultivated host plant counterparts on a 'micro-evolutionary' scale for host-specified evolution of TYLCV-IL populations.

## Material and Methods

### Acquisition of full-length TYLCV genomic sequences

According to our previous study (Hosseinzadeh *et al*., 2014), we selected TYLCV genomes collected from *D. stramonium* plants and submitted them to the National Center for Biotechnology Information (NCBI) database (Table 1). To investigate the population structure of isolates infecting *D. stramonium* compared to other host plants, some isolates infected crops and weeds. Comprehensive details on the sequences (accession numbers, countries of origin, and hosts) are provided in Table S1.

### Begomovirus genome sequences, pairwise distances, and phylogenetic analysis

Based on the BLASTn analysis (http://www.ncbi.nlm.nih.gov), representative TYLCV-IL isolates that were most closely related to the TYLCV-IL isolates recovered from *D. stramonium* were selected for the analyses and downloaded from the GenBank database (Table 1). Also, a sequence of Tomato leaf curl Iran virus (TLCIV) was selected as an outgroup and downloaded from GenBank.

Pairwise distance analysis was carried out using the PASC algorithm, available at the NCBI Genbank database (http://www.ncbi.nlm.nih.gov/sutils/pasc/). For phylogenetic analysis, the sequences determined and reference sequences were aligned using the Q-INS-i algorithm in

MAFFT (https://mafft.cbrc.jp/alignment/server) (Katoh *et al*., 2013). The alignment was edited with Gblocks (http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=gblocks) and the least stringent parameters (Castresana, 2000). The optimal base-substitution model was determined using MrModeltest 2 (Nylander, 2004). The Akaike-supported model, a general time-reversible model, was used to analyze among-site rate heterogeneity and estimate invariant sites (GTR + G + I). The Bayesian analysis was carried out using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003), with a random burn-in tree run of 1,000,000 generations and convergence set to 25%. The posterior probabilities (Larget *et al*., 1999) were estimated by the Markov chain Monte Carlo (MCMC) method with 50% majority rule.

**Recombination analysis**

Recombination analysis was carried out for TYLCV-IL genomes recovered from *D. stramonium* plants and selected closest relatives (Table 1) using the Recombination Detection Package RDP v.4, with default parameters. The RDP package includes a suite of programs: RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq. Statistically significant results across at least six of the seven programs were considered robust evidence of a predicted recombination event (Martin *et al*., 2010).

**Table 1** List of isolates and full-length sequences of *Datura stramonium*-associated *Tomato yellow leaf curl virus*-IL (TYLCV-IL) haplotypes from Bojnourd, Iran. Three reference begomoviral sequence and acronym, plant host and number, origin, collection year, GenBank Accession number, begomovirus genome size, and source or published reference.

| Virus haplotype | Acronym | Plant host /Plant number | Origin | Year of collection | GenBank accession number | Full-length genome size (nt) | Sequence source or reference |
|---|---|---|---|---|---|---|---|
| *Tomato yellow leaf curl virus*-IL TYLCV-IL[IR:BojD5-2:12] | TYLCV-IL[IR:BojD5-2:12 ] | Datura #5 | Bojnourd /Iran | 2011 | KC106635 | 2782 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD8-1:12] | TYLCV-IL[IR:BojD8-1:12] | Datura #8 | Bojnourd /Iran | 2011 | KC106636 | 2779 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD11-1:12] | TYLCV-IL[IR:BojD11-1:12] | Datura #11 | Bojnourd /Iran | 2011 | KC106637 | 2779 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD23-7:12] | TYLCV-IL[IR:BojD23-7:12] | Datura #23 | Bojnourd /Iran | 2011 | KC106638 | 2782 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD23-A:12] | TYLCV-IL[IR:BojD23-A:12] | Datura #23 | Bojnourd /Iran | 2011 | KC106640 | 2782 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD28-7:12] | TYLCV-IL[IR:BojD28-7:12] | Datura # 28 | Bojnourd /Iran | 2011 | KC106641 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD28-3:12] | TYLCV-IL[IR:BojD28-3:12] | Datura #28 | Bojnourd /Iran | 2011 | KC106642 | 2759 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD28-5:12] | TYLCV-IL[IR:BojD28-5:12] | Datura #28 | Bojnourd /Iran | 2011 | KC106643 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD28-2:12] | TYLCV-IL[IR:BojD28-2:12] | Datura #28 | Bojnourd /Iran | 2011 | KC106644 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD28-8:12] | TYLCV-IL[IR:BojD28-8:12] | Datura #28 | Bojnourd /Iran | 2011 | KC106645 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD132-1:12] | TYLCV-IL[IR:BojD132-1:12] | Datura #132 | Bojnourd /Iran | 2011 | KC106646 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IL [IR:BojD132-13:12] | TYLCV-IL[IR:BojD132-13:12] | Datura #132 | Bojnourd /Iran | 2011 | KC106647 | 2781 | (49) |
| *Tomato yellow leaf curl virus*-IR[IR:Iranshar:98] | TYLCV-IR[ IR:Iranshar:98] | Tomato | Iranshahr/Iran | 1998 | AJ132711 | 2771 | (67) |
| *Tomato yellow leaf curl virus*-Mild [Israel] | TYLCV-Mld | Tomato | Israel | 1994 | X76319 | 2790 | (96) |
| *Tomato yellow leaf curl virus*-Israel [Israel:rehovot] | TYLCV-IL | Tomato | Israel | 1991 | X15656 | 2787 | (5) |
| *Tomato leaf curl Iran virus*[Iran] | TLCIV | Tomato | Iranshahr/Iran | 2004 | NC005842 | 2763 | (88) |

**Nucleotide substitution rate estimates and mutational bias**

Analysis of the mean rate of nucleotide substitutions per site was carried out for TYLCV-IL coding regions (n = 6) and the corresponding IR sequence associated with each genome (n = 12 haplotypes). Mutation and insertion-deletion rates were determined using Maximum likelihood in MEGA5 with the T92 + G model of evolution (Tamura *et al.*, 2011).

**Gene flow and genetic differentiation among TYLCV-IL populations**

Population differentiation analyses (Hudson *et al*., 1992; Hudson, 2000) of the TYLCV genome sequences recovered from *D. stramonium* (Hosseinzadeh *et al*., 2014) plants, and previously reported TYLCV genomes associated with other plants (Table S1), based on KS, Z, and Snn parameters, implemented in the DnaSP v.5 software package, were calculated. Genetic divergence among populations by plant host was assessed using the KST statistic, as established methods (Tsompana *et al*., 2005). The FST algorithm was used to analyze between-population gene flow and genetic isolation (Hudson *et al*., 1992; Coskan *et al.*, 2022).

**Genetic variability and evolutionary determinants**

Genetic variability analysis between TYLCV-IL genome sequences isolated from *D. stramonium* were determined based on the number of polymorphic sites (S), total number of mutations (Eta), nucleotide diversity ($\pi$), number of haplotypes (h), mean number of nucleotide differences (k), haplotype diversity (Hd), Watterson's estimate of population mutation rate (total number of segregating sites (θ-w) and mutations, or θ-Eta). The Tajima's D dN/dS (Ka/Ks) ratio of neutrality test was used to identify the extent of deviation from assumptions associated with neutral evolution by analyzing the TYLCV open reading frames (n = 6) and non-coding IR sequences for twelve haplotypes (Table 1). Analysis was conducted with DnaSP v.5 to determine the best-fitting evolutionary model, the neutral mutation rate/site, and the transition/transversion ratio (Rozas *et al*., 2003).

## Results

**Begomovirus genome sequences, preliminary BLASTn identification, and pairwise distances**

Pairwise distance analyses of the TYLCV-IL sequences (Table 1), indicated that the TYLCV-IL sequences from *D. stramonium* shared > 94% nucleotide identity with reference TYLCV-IL sequences. The Rep binding site sequences or 'iterons' in the IR of all TYLCV-IL haplotypes (from *D. stramonium*) (Table 1) were similar to but unique from the IR of previously reported TYLCV-IL and TYLCV-IR isolates (Fig. 1). Differences within the IR consisted of the following modifications: **1**) adenine (**A**) replaced by thymine (**T**) in the predicted 5'-putative Rep binding motif e.g., plants #132 and #28; **2**) a 4-nt spacer consisting of **ATCC** at the 3'-putative Rep binding motif, compared to **ATC** e.g., plant #5; **3**), replacement of **T** with **A** at the third base position (5′-GG**A**GT-3′) e.g., plants #5, #28, and #132 (Table 1).

**Nucleotide substitution rate and mutational bias**

The estimated mean nucleotide substitution rate for TYLCV-IL genomes isolated from *D. stramonium* was $1.45 \times 10^{-4}$ substitutions per site. Substitution rates were $4.02 \times 10^{-4}$, $1.94 \times 10^{-4}$, $1.37 \times 10^{-4}$, and $8.07 \times 10^{-5}$ for IR, V1 (*cp*), C1 (*rep*), and C4 coding regions, respectively (Table 2). Based on the number and location of mutations in *D. stramonium*-TYLCV-IL haplotype genomes, the most common substitutions were C→T and G→A, indicating a bias toward transitions over transversions (Table 2). Also, the CG→AT shift occurred in parallel with the most common substitutions, which were C→T or G→A transitions.

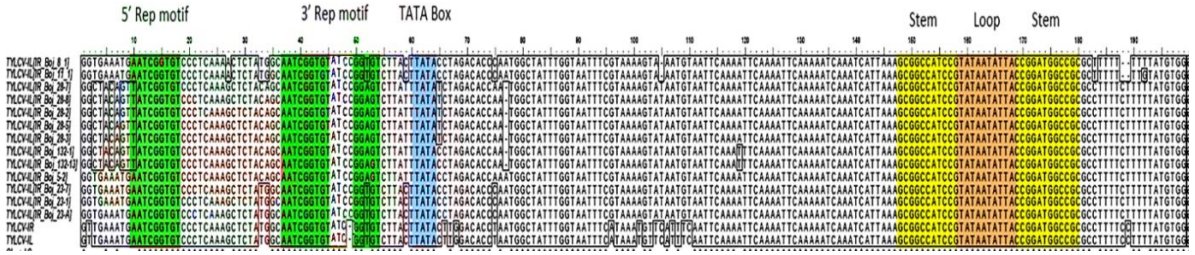


**Figure 1** Alignment of predicted Rep binding sites in the non-coding intergenic region of sequences of 12 Tomato yellow leaf curl virus-IL genomes recovered from *Datura stramonium* plants naturally-occurring near tomato fields in Bojnourd, Iran, and selected TYLCV-IL and TYLCV-IR isolates (Table 1). The stem (yellow) and loop (light brown) sequence of predicted stem-loop structure, location of the 5'- and 3'- putative regions of Rep-binding motifs (green) or 'iterons', and the TATA box (blue) of the Rep promoter are highlighted by different colors, as indicated.

**Table 2** The estimated nucleotide substitution rate and mutational bias analysis of *Tomato yellow leaf curl virus*-IL-*Datura stramonium* haplotypes based on the T92 + G model of evolution. The substitution rate for transitions is shown in bold and transversion substitutions are indicated by *italics*.

| | A | T | C | G |
|---|---|---|---|---|
| A | - | *5.90* | *4.09* | **12.30** |
| T | *5.90* | - | **12.30** | *4.09* |
| C | *5.90* | **17.74** | - | *4.09* |
| G | **17.74** | *5.90* | *4.09* | - |

## Recombination analysis

The recombination analysis for TYLCV-IL haplotypes (n = 12) and selected putative parental TYLCV genomes (Table 1) predicted five unique recombination events, each supported six or more algorithms (Table 3). A recombination breakpoint was detected for TYLCV-IL haplotype from plant #23, isolate KC106640 (TYLCV-IL [IR: Boj:23-A]). Breakpoints were located between nucleotide coordinate 2606 of C1 and 2728 in the IR, which were supported by six of seven RDP algorithms (Table 3). The isolate KC106643, TYLCV-IL [IR: Boj:28-5] from plant #28 was identified as the major parent. An analogous fragment was identified at the same genomic location in the TYLCV-IL haplotype from plant #8, KC106636 (TYLCV-IL [IR: Boj: 8-1]), identifying it as the minor parent. A third recombination event (data not shown) was associated with haplotype X76319 (TYLCV-MId), based on a breakpoint between coordinates 2589 in C1/C4 ORF and 2715 in the C1 ORF. The major parent was identified as a haplotype from *D. stramonium* plant #132, KC106646, TYLCV-IL [IR: Boj: 132-1], whereas the minor parent was identified as haplotype X15656 (TYLC-IL). The predominant regions for recombination were consistently localized in the IR, C1 ORF, and overlapping C1/C4 regions, all involved in replication and/or containing regulatory elements, i.e., IR, known to interact with host elements.

## Phylogenetic analysis

To avoid overestimating genetic variation due to elevated mean apparent substitution rates and genomic variability, the predicted recombinant regions of the TYLCV-IL haplotype sequences were removed before Bayesian analysis. In the Bayesian phylogeny, all of the TYLCV-IL haplotypes were grouped by plant host species (Fig. 2). The 'within-host' phylogenetic analysis identified six subpopulations of TYLCV-IL from *D. stramonium* plant cohorts. They are represented by haplotypes from plant #5 (KC106635), #8 (KC106636), #11 (KC106637), #23 (KC106638, KC106640,), #28 (KC106641, KC106643, KC106644, KC106645, KC106642), and #132 (KC106646, KC106647), respectively. Based on the results the evolution (Fig. 2) of *D. stramonium*-TYLCV-IL haplotypes has been shaped by host plant species-begomovirus interactions.

**Table 3** Putative recombination breakpoints detected by RDP for the *Datura stramonium*-associated *Tomato yellow leaf curl virus*-IL isolates from Bojnourd, Iran and selected GenBank reference sequences.

| Event | Recombinant isolate* | Parents | | Breakpoints[1] | | Detection methods[2] | P-value** |
|---|---|---|---|---|---|---|---|
| | | Major | Minor | Start | End | | |
| 1 | TYLCV-IL[IR:Boj:23-A] | TYLCV-IL[IR:Boj:28-5] | TYLCV-IL[IR:Boj:8-1] | 2606 | 2728 | GBMCST | $1.054\times10^{-2}$ |
| 2 | TYLCV-MId | TYLCV-IL [IR: Boj: 132-1] | TYLCV-IL | 2589 | 2715 | RGBMCST | $1.312\times10^{-10}$ |

* Virus names and their accession numbers are given in table1 and 2.
**The described P-value corresponds to the calculated P-value for the event in question, detected by the program in bold.
[1]. Numbering starts at the first nucleotide after the nicking site of the origin of replication and increases clockwise.
[2]. Recombination detection methods: RDP (R); GENECONV (G); BootScan (B); MaxChi (M); Chimaera (C); SiScan (S); 3Seq (3S).

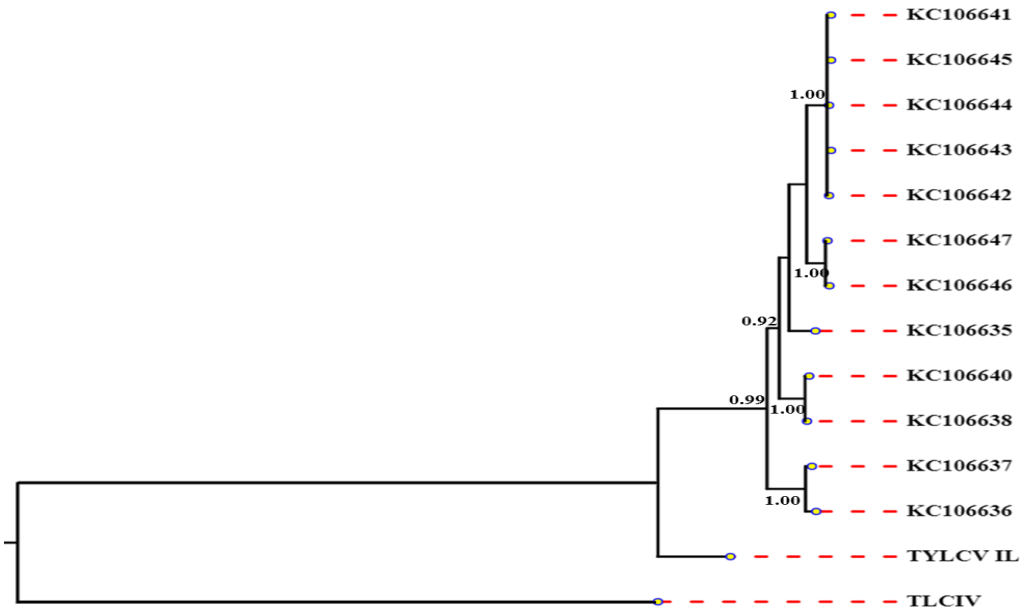


**Figure 2** Bayesian consensus tree of the *Datura stramonium* plant-isolated full genome *Tomato yellow leaf curl virus-IL* variants from Bojnourd, Iran and the selected begomoviruses (Table 1), under the GTR + G + I model. Bayesian posterior probability values, shown above the nodes. The burn-in phase was set at 25% of the converged runs. The tree was rooted with the *Tomato leaf curl Iran virus* (TLCIV) genome sequence.

## Gene flow and Genetic differentiation among TYLCV-IL populations

The best-fitting evolutionary model, relevant to the MEGA5 analysis, varied by genome region and was identified as GTR, T92 + G, T92, and JC (Table S2). Based on DnaSP assignments, analysis of population genetic structure revealed that the *D. stramonium*-TYLCV-IL genome haplotypes recovered from plants #23 and #28, respectively, showed greater genetic variability than haplotypes from plant #132 (Table 4). The indicators of genetic variability among twelve TYLCV-IL haplotypes recovered from the six *D. stramonium* plants were as follows: high nucleotide diversity ($\pi = 0.00888$), haplotype diversity (Hd = 0.985), mutation rate (Eta = 75), segregating sites (S = 75), average inter-genome nucleotide differences (k = 24.47), diversity based on segregating sites (θ-w = 0.0098), and total mutations (θ-Eta = 0.0098) (Table 4).

Genetic differentiation of TYLCV populations, analyzed pairwise, indicated the highest average and most significant pairwise nucleotide difference between the wild D. stramonium reservoir and populations associated with tomato plant hosts (Table 5), with KS = 4.106. The KST, Z, Snn, and FST values were significant, indicating the *D. stramonium* populations were distinct from those associated with a non-*D. stramonium* host (Table 5). The FST value was highest for *D. stramonium* population comparisons, with the 'crop' population (FST = 0.517) (Table 5). The Tajima's D neutrality test carried out to evaluate potential selection and/or demographic forces acting on V2 and C3 coding regions, indicated both ORFs had a negative dN/dS ratio (Table S2). for the C4 ORF the dN/dS ratio was 1.1663 or > 1, indicating positive selection that has contributed to TYLCV-IL haplotype micro-evolution in wild *D. stramonium* plants. Finally, the dN/dS ratio of > 1 for V1 and C1 ORFs among TYLCV-IL haplotypes from plants #23 and #28, respectively, indicated within-host selection of TYLCV-IL haplotypes unique to each *D. stramonium* plant, a pattern consistent with ongoing diversification driven by non-genetically uniform host plants.

**Table 4** Estimated genetic variability for *Datura stramonium-Tomato yellow leaf curl virus*-IL isolates from Bojnourd, Iran.

| Population(s) | No. of sequences | Sequence length | S[1] | Eta[2] | k[3] | $\pi$[4] | h[5] | Hd[6] |
|---|---|---|---|---|---|---|---|---|
| *D. stramonium* plant 132 | 2 | 2782 | 1 | 1 | 1 | 0.00030 | 2 | 1 |
| *D. stramonium* plant 23 | 2 | 2782 | 2 | 2 | 1 | 0.00023 | 2 | 1 |
| *D. stramonium* plant 28 | 5 | 2781 | 3 | 3 | 1.200 | 0.00043 | 4 | 0.900 |
| *D. stramonium* plants 23, 132 and 28 | 9 | 2781-2 | 40 | 40 | 15.72 | 0.00570 | 8 | 0.972 |
| *D. stramonium* plants 5, 8, 11, 23, 28 and 132 | 12 | 2779-2782 | 75 | 75 | 24.47 | 0.00888 | 11 | 0.985 |

1. Total number of segregating sites.
2. Total number of mutations.
3. Average number of nucleotide differences between sequences, or Tajima's estimate of the population mutation rate, θ.
4. Nucleotide diversity.
5. Haplotype number.
6. Haplotype diversity.
7. Waterson's estimate of population mutation rate based on total number of segregating sites.
8. Waterson's estimate of population mutation rate based on total number of mutations.

**Table 5** Genetic differentiation among *Tomato yellow leaf curl virus* genomes associated with different host plants.

| Populations | KS* | KST* | P value* | Z* | P value* | Snn | P value* | FST |
|---|---|---|---|---|---|---|---|---|
| Datura Vs Tomato | 4.071 | 0.038 | 0.0000 *** | 7.5870 | 0.0000 *** | 1 | 0.0000 *** | 0.343 |
| Datura Vs Weed | 3.930 | 0.183 | 0.0000 *** | 4.1450 | 0.0000 *** | 0.95830 | 0.0010 ** | 0.509 |
| Datura Vs Other crops | 4.106 | 0.171 | 0.0000 *** | 4.6863 | 0.0000 *** | 0.93548 | 0.0000 *** | 0.517 |
| Weed Vs Tomato | 4.527 | 0.005 | 0.0110 * | 7.7459 | 0.0020 ** | 0.91149 | 0.0000 *** | 0.050 |
| Tomato Vs Other crops | 4.522 | 0.009 | 0.0010 ** | 7.8530 | 0.0000 *** | 0.86079 | 0.0020 ** | 0.050 |
| Weed Vs Other crops | 5.385 | 0.019 | 0.0870 ns | 4.9460 | 0.0800 ns | 0.67930 | 0.0620 ns | 0.060 |

KS *, KST *, Z * and Snn are test statistics of genetic differentiation. FST, coefficient of gene differentiation, which measures inter-population diversity.

## Discussion

Five strains of tomato yellow leaf curl virus (TYLCV) are extant in Iran, including the most widespread and damaging, TYLCV-IL (Brown and Idris, 2006), with predicted endemism in Israel and Iran, TYLCV-IR, TYLCV-Bou, and TYLCV-Ker strains, all endemic to Iran, and TYLCV-OM, which is endemic to Iran and Oman (Bananej *et al*., 2004; Khan *et al*., 2008; Lefeuvre *et al*., 2010). Although the host range might vary among TYLCV strains, *D. stramonium* is most probably a universal wild host of TYLCV-IL (Diaz-Pendon *et al*., 2010; Hosseinzadeh *et al*., 2014). Also, naturally-occurring reservoirs of TYLCV-IL have been identified in the Compositae, Leguminosae, Malvaceae, Plantaginaceae, and Solanaceae (Abraham *et al*., 2021; Ioannou *et al*., 1987).

At least seven TYLCV strains have evolved in Middle East (https://www.worldatlas.com/webimage/countrys/me.htm) and are recognized as pathogens of tomato crops worldwide. They are represented by 'type' genome sequences: TYLCV-Bou [IR-Gen29-06] GU076454, TYLCV-IR [IR-Ira-98] AJ132711, TYLCV-Kah [IR-Kah-07] EU635776, TYLCV-Ker [IR-Hor32-06] GU076442, TYLCV-Mld [IL-93] X76319, and TYLCV-OM [OM-Alb22-05] FJ956700. Based on the phylogeographical analysis of isolates extant in Iran and the Middle East-Arabian Peninsula, five of those strains are endemic to Western Asia (Hosseinzadeh *et al*., 2014).

In this study, within-host TYLCV-IL isolates from each *D. stramonium* plant were grouped into six phylogeographical monophyletic groups (Fig. 2), including predicted recombinants. Although several TYLCV-IL strains are distributed worldwide, most have not spread beyond their respective centers of diversification (Marchant *et al*., 2023). Those TYLCV-IL isolates originating in southern Iran show greater genetic variability than their northeastern counterparts, apparently due to geographical isolation (Hosseinzadeh *et al*., 2014). In Oman, TYLCV-OM isolates are geographically isolated and exhibit greater genomic variability than their Iranian relatives (Khan *et al*., 2008).

Newly emerging recombinants were discovered, several of which have invaded tomato-producing areas worldwide. An analogous recombinant region identified in TYLCV-IL Iran isolates from D. stramonium was localized to the IR, C1, and C1/C4 overlapping coding regions. These findings have been reported previously for other TYLCV isolates (Bananej *et al*., 2004; Berrie *et al*., 2001; Hosseinzadeh *et al*., 2014; Padidam *et al*., 1999). In particular, the IR has been reported as a common hotspot for begomovirus recombination (Berrie *et al*., 2001; Padidam *et al*., 1999). Because the IR contains the origin of replication (*ori*) and regulatory elements that interact with host factors during replication, this region comprises likely targets involved in begomovirus evolution and host adaptation. Although coding regions may be less susceptible to recombination (Lefeuvre *et al*., 2007), predicted recombination hotspots have been detected in CP, *Rep*, and C4 (Berrie *et al*., 2001; Hosseinzadeh *et al*., 2014; Idris and Brown, 2002). Here, predicted TYLCV recombinants were identified in *D. stramonium*, comprising co-infecting TYLCV isolates herein (Tables 1 and 2), including a strain and another species of TYLCV.

Micro-evolutionary analyses of TYLCV-IL, the predominant and most invasive TYLCV strain infecting tomato crops worldwide (Fig. 2 and Table 4), suggest its potential to evolve uniquely structured populations within individual plants. Furthermore, in *D. stramonium*, TYLCV-IL genomes exhibit a high 'RNA-like substitution rate,' harbor extensive genetic variation (Table S2), and show within-host predicted recombination. Further, in this wild reservoir, the consensus sequence that represents divergent mutants was maintained. A similar result has been reported for different TYLCV species ($2.88 \times 10^{-4}$), East African cassava mosaic virus [$1.60 \times 10^{-3}$ for DNA A] (Duffy and Holmes, 2008, 2009), and several RNA viruses ($1 \times 10^{-3}$ to $1 \times 10^{-5}$) (Jenkins *et al*., 2002). The high nucleotide substitution rate

within the TYLCV-IL IR indicates that positive selection is acting on the conserved Rep-binding 'iterons' (Font *et al*., 2007), thereby preserving the essential functionality ascribed to these repeated sequences. Although the TYLCV-IL IR harbors extensive genetic variability that allows for relaxed, purifying selection, essential promoter sequences and the *ori* must also be maintained for virus survival (Ge *et al*., 2007). Begomovirus evolution has been reported to rely primarily on mutations occurring unevenly throughout the virus genome (Duffy and Holmes, 2008b; Ge *et al*., 2007; Silva *et al*., 2012). This pattern is consistent with the substitution pattern reported for other TYLCV species, which also involved G-to-A and C-to-T substitutions (Duffy and Holmes, 2008). The dN/dS analysis revealed ORF regions of *D. stramonium*-associated TYLCV-IL variants with a dN/dS ratio of less than one, indicating that through purifying selection, the integrity of proteins has been preserved. And although C4 showed positive selection 'among' host-specified begomovirus populations (Table S2), the 'within' and/or 'among' host population comparisons revealed both showed strong negative and positive selection for C1 and V1 ORFs. The begomovirus CP is a multi-functional protein required for encapsidation, movement, and whitefly vector transmission (Noris *et al*., 1998). Evidence of synonymous substitutions in the *cp* and C1 ORFs indicated purifying selection, consistent with maintaining capsid integrity. Finally, positive selection in V1 may be suggestive of ongoing-evolution between TYLCV-IL variants and the whitefly vector, with virus variants vying *in planta* and/or in the vector itself for preferential transmission.

The TYLCV-IL iteron sequence, 5′-GGTGT-3′, at nucleotide coordinates 2627-2631, and two directly-repeated sequences consisting of the three-nucleotide spacer, 5′-GGTGTATCG GTGT-3′, occurring on the virus-sense strand between coordinates 2645-2657, immediately upstream of the TATA box of the Rep promoter. Independently replicating begomovirus variants that have a single nucleotide change in the Rep binding motif have been reported (Behjatnia *et al.*, 2001; Chatterji *et al*., 1999). Alignment of the IR for *D. stramonium*-derived TYLCV-IL isolates revealed that regions within the 5′ and 3′ Rep binding motif differed from tomato-infecting TYLCV-IL counterparts (Fig.). Although these changes occurred consistently between the wild and cultivated host-specified TYLCV-IL isolates, their functional importance is not yet understood.

Positive selection acting on the C4 ORF is expected to reflect an adaptive evolution scenario in which sequences become fixed in a virus population over time. The *D. stramonium*-TYLCV-IL haplotypes are therefore posited to have evolved to counter host-specific defenses conferred by the C4 protein suppressor of host plant PTGS (Li *et al*., 2020). Increased virus-host adaptation could result in larger relative population sizes that drive adaptation and increase the likelihood of host jumps. For example, tobacco leaf curl virus (TLCV), a begomovirus of *Eupatorium makinoi* endemic to Japan, is predicted to have established in the latter host, following a host-jump attributed to C4 protein-mediated host-range determination (Yahara *et al.*, 1998). Selection pressure can lead to host adaptation through increased or decreased rates of replication, host defense evasion, increased vector transmission efficiency/competency, and speciation (Nigam, 2021). Finally, scenarios observed here for TYLCV-IL are consistent with a "niche-filling model", which purports a role for host interactions in shaping virus evolution (Simmonds *et al*., 2019). This study provides evidence that *D. stramonium*-TYLCV-IL interactions have shaped distinct begomovirus populations on an individual plant basis by individual host-specified selection of TYLCV-IL populations or 'quasispecies'. Finally, gene flow and genetic differentiation analyses have shown that *D. stramonium* plants support host-specific populations.

## Conclusions

A high mutation rate is a primary driving force behind variation in begomovirus populations

(Lima *et al.*, 2017), and host plant-imposed selection can shape virus population-host interactions (Nigam *et al*., 2020). The observed within- and among-host plant variability associated with TYLCV-IL populations evolving in a wild, genetically variable host suggests that D. stramonium acts as an important 'melting pot' for TYLCV-IL diversification in Iran and likely elsewhere. Each plant harbored a uniquely structured virus population with the potential to contribute to evolutionary plasticity conducive to survival and plant-to-plant spread.

A better understanding of the evolutionary dynamics of begomoviruses associated with wild plant reservoirs could benefit efforts to develop TYLCV-IL-resistant tomato and other crop species. Resistance-breaking by begomoviruses in commercial tomato varieties has been linked to the global expansion of monoculture agriculture, which has favored the planting of genetically uniform cultivars over expansive areas. Deciphering genome- and population-level interactions between coevolving quasi-species at the wild-cultivated plant host interface, particularly in centers of plant host and/or virus diversification, may provide important insights into quasi-species. This diversification favors the emergence of resistance-breaking strains through repeated encounters with genetically uniform crop varieties, leading to virus-host coevolutionary scenarios that could ultimately inform the development of durable resistance.

**Acknowledgments**

The authors from Iran acknowledge financial support from Tarbiat Modares University, Tehran, Iran.

**Conflicts of Interest**

The authors declare they have no conflicts of interest.

**Author Contributions**

Conceptualization (MSB, MH); research, data curation (MH, SA), and preparation of the draft manuscript (MH, SA, MSB, JKB), supervision and resources (MSB); final reviewing and editing (MH, MSB, JKB). All authors approve of the authorship, have read and discussed the results and conclusions, and approve of the content in the final version.

**References**

Abraham, P., Banwo, O. O., David Kashina, B. and Alegbejo, M. D. 2021. Identification of weed hosts of *Tomato yellow leaf curl virus* in field-grown tomato in Sudan Savanna, Nigeria. International Journal of Horticultural Science and Technology, 8: 235-246. https://doi.org/10.22059/ijhst.2021.306752.381.

Bananej, K., Kheyr-Pour, A., Salekdeh, G. H. and Ahoonmanesh, A. 2004. Complete nucleotide sequence of Iranian *Tomato yellow leaf curl virus* isolate: further evidence for natural recombination amongst begomoviruses. Archives of Virology, 149: 1435-1443. https://doi.org/10.1007/s00705-004-0308-9.

Behjatnia, S. A. A., Dry, I. B. and Rezaian, M. A. 2001. Sequence divergence in new strains of *Tomato leaf curl virus* resulting in replication specificity. Australasian Plant Pathology, 30: 337-342. https://doi.org/10.1071/AP01042.

Berrie, L. C., Rybicki, E. P. and Rey, M. E. 2001. Complete nucleotide sequence and host range of *South African cassava mosaic virus*: further evidence for recombination amongst begomoviruses. Journal of General Virology, 82: 53-58. https://doi.org/10.1099/0022-1317-82-1-53.

Brown, J. and Idris, A. 2006. Introduction of the exotic monopartite *Tomato yellow leaf curl virus* into west coast Mexico. Plant Disease, 90: 1360-1360. https://doi.org/10.1094/PD-90-1360A.

Brown, J. K. and Bird, J. 1992. Whitefly-transmitted geminiviruses and associated disorders in the Americas and the Caribbean Basin. Plant Disease, 76: 220-225. https://doi.org/10.1094/PD-76-0220.

Brown, J. K. and Czosnek, H. 2002. Whitefly transmission of plant viruses. Advances in Botanical Research, 36: 65-100. https:// doi.org/10.1016/S0065-2296(02)36059-2.

Castillo-Urquiza, G. P., Alfenas-Zerbini, P., Beserra-Junior, J. E. A., Mizubuti, E. S. G., Varsani, A., Martin, D. P. and Zerbini, F. M. 2010. Genetic structure of tomato-infecting begomovirus populations in two tomato-growing regions of Southeastern Brazil. In: Program and Abstracts, 6th International Geminivirus Symposium and 4th International ssDNA Comparative Virology Workshop, Guanajuato, Mexico.

Castillo-Urquiza, G. P., Beserra Júnior, J. E. A., Alfenas-Zerbini, P., Varsani, A., Lima, A. T. M., Barros, D. R. and Zerbini, F. M. 2007. Genetic diversity of begomoviruses infecting tomato in Paty do Alferes, Rio de Janeiro state, Brazil. Virus Reviews and Research, 12: 233.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution, 17: 540-552. https://doi.org /10.1093/oxfordjournals.molbev.a026334.

Chatterji, A., Padidam, M., Beachy, R. N. and Fauquet, C. M. 1999. Identification of replication specificity determinants in two strains of *Tomato leaf curl virus* from New Delhi. Journal of Virology, 73: 5481-5489. https://doi.org/10. 1128/jvi.73.7.5481-5489.1999.

Cooper, I. and Jones, R. A. 2006. Wild plants and viruses: under-investigated ecosystems. Advances in Virus Research, 67: 1-47. https ://doi.org/10.1016/s0065-3527(06)67001-2.

Coskan, S., Morca, A. F., Akbaş, B., Celik, A. and Santosa, A. 2022. Comprehensive surveillance and population study on *Plum pox virus* in Ankara Province of Turkey. Journal of Plant Diseases and Protection, 129: 981-991. http://dx.doi.org/10.1007/s413 48-022-00597-5.

Czosnek, H. and Laterrot, H. 1997. A worldwide survey of *Tomato yellow leaf curl viruses*. Archives of Virology, 142: 1391-1406. https://doi.org/10.1007/s007050050168.

Diaz-Pendon, J. A., Canizares, M. C., Moriones, E., Bejarano, E. R., Czosnek, H. and Navas-Castillo, J. 2010. *Tomato yellow leaf curl virus*: menage a trois between the virus complex, the plant and the whitefly vector. Molecular Plant Pathology, 11: 441-450. https://doi.org/10.1111/j.1364-3703. 2010.00618.x.

Domingo, E. and Perales, C. 2019. Viral quasispecies. PLOS Genetics, 15: e1008271. https://doi.org/10.1371/journal.pgen.1008271.

Duffy, S. and Holmes, E. C. 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. Journal of Virology, 82: 957-965. https://doi.org/10.1128/JVI.01929-07.

Duffy, S. and Holmes, E. C. 2009. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. Journal of General Virology, 90: 1539-1547. https://doi.org/10.1099/vir.0.009266-0.

Font, M. I., Rubio, L., Martínez-Culebras, P. V. and Jordá, C. 2007. Genetic structure and evolution of natural populations of viruses causing the tomato yellow leaf curl disease in Spain. Virus Research, 128: 43-51. https://doi.org/10.1016/j.virusres.2007.04.003.

Frischmuth, T., Engel, M., Lauster, S. and Jeske, H. 1997. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, *Sida*-infecting bipartite geminiviruses in Central America. Journal of General Virology, 78: 2675-2682. https:// doi.org/10.1099/0022-1317-78-10-2675.

Ge, L., Zhang, J., Zhou, X. and Li, H. 2007. Genetic structure and population variability of *Tomato yellow leaf curl China virus*. Journal of Virology, 81: 5902-5907. https://doi.org/10.1128/jvi.02431-06.

Gutierrez, C. 1999. Geminivirus DNA replication. Cellular and Molecular Life Sciences, 56: 313-329. https://doi.org/10. 1007/s000180050433.

Hosseinzadeh, M. R., Shams-Bakhsh, M., Osaloo, S. K. and Brown, J. K. 2014. Phylogenetic relationships, recombination

analysis, and genetic variability among diverse variants of *Tomato yellow leaf curl virus* in Iran and the Arabian Peninsula: further support for a TYLCV center of diversity. Archives of Virology, 159: 485-497. https://doi.org/10.1007/s00705-013-1851-z.

Hudson, R. R., Slatkin, M. and Maddison, W. P. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics, 132: 583-589. https://doi.org/10.1093/genetics/132.2.583.

Hudson, R. R. 2000. A new statistic for detecting genetic differentiation. Genetics, 155: 2011-2014. https://doi.org/10.1093/genetics/155.4.2011.

Idris, A. M. and Brown, J. K. 2002. Molecular analysis of *Cotton leaf curl virus*-Sudan reveals an evolutionary history of recombination. Virus Genes, 24: 249-256. https://doi.org/10.1023/A:1015380600089.

Ioannou, N., Kyriakou, A. and Hadjinicolis, A. 1987. Host range and natural reservoirs of *Tomato yellow leaf curl virus*. Technical Bulletin, 85: 1-8. https://doi.org/10.22004/ag.econ.316090.

Jenkins, G. M., Rambaut, A., Pybus, O. G. and Holmes, E. C. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. Journal of Molecular Evolution, 54: 156-165. https://doi.org/10.1007/s00239-001-0064-3.

Jupin, I., De Kouchkovsky, F., Jouanneau, F. and Gronenborn, B. 1994. Movement of *Tomato yellow leaf curl geminivirus* (TYLCV): involvement of the protein encoded by ORF C4. Virology, 204: 82-90. https://doi.org/10.1006/viro.1994.1512.

Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30: 772-780. https://doi.org/10.1093/molbev/mst010.

Khan, A. J., Idris, A. M., Al-Saady, N. A., Al-Mahruki, M. S., Al-Subhi, A. M. and Brown, J. K. 2008. A divergent isolate of *Tomato yellow leaf curl virus* from Oman with an associated DNAβ satellite: an evolutionary link between Asian and the Middle Eastern virus–satellite complexes. Virus Genes, 36: 169-176. https://doi.org/10.1007/s11262-007-0163-3.

Larget, B. and Simon, D. L. 1999. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution, 16: 750–759. https://doi.org/10.1093/oxfordjournals.molbev.a026160.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. and Gronenborn, B. 1995. *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of *Tomato yellow leaf curl virus*. Proceedings of the National Academy of Sciences USA, 92: 3879-3883. https://doi.org/10.1073/pnas.92.9.3879.

Lefeuvre, P., Lett, J. M., Reynaud, B. and Martin, D. P. 2007. Avoidance of protein fold disruption in natural virus recombinants. PLOS Pathogens, 3: e181. https://doi.org/10.1371/journal.ppat.0030181.

Lefeuvre, P., Martin, D. P., Harkins, G., Lemey, P., Gray, A. J., Meredith, S., Lakay, F., Monjane, A., Lett, J. M., Varsani, A. and Heydarnejad, J. 2010. The spread of *Tomato yellow leaf curl virus* from the Middle East to the world. PLOS Pathogens, 6: e1001164. https://doi.org/10.1371/journal.ppat.1001164.

Li, Z., Du, Z., Tang, Y., She, X., Wang, X., Zhu, Y., Yu, L., Lan, G. and He, Z. 2020. C4, the pathogenic determinant of *Tomato leaf curl Guangdong virus*, may suppress post-transcriptional gene silencing by interacting with BAM1 protein. Frontiers in Microbiology, 11: 851. https://doi.org/10.3389/fmicb.2020.00851.

Lima, A. T. M., Silva, J. C., Silva, F. N., Castillo-Urquiza, G. P., Silva, F. F., Seah, Y. M., Mizubuti, E. S. G., Duffy, S. and Zerbini, F. M. 2017. The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evolution, 3: vex005. https://doi.org/10.1093/ve/vex005.

Marchant, W. G., Mugerwa, H., Gautam, S., Al-Aqeel, H., Polston, J. E., Rennberger, G., Smith, H. A., Turechek, B., Adkins, S. and

Brown, J. K. 2023. Phylogenomic and population genetics analyses of extant *Tomato yellow leaf curl virus* strains on a global scale. Frontiers in Virology, 3: 1221156. https://doi.org/10.3389/fviro.2023.1221156.

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

Navot, N., Pichersky, E., Zeidan, M., Zamir, D. and Czosnek, H. 1991. *Tomato yellow leaf curl virus*: a whitefly-transmitted geminivirus with a single genomic component. Virology, 185: 151-161. https://doi.org/10.1016/0042-6822(91)90763-2.

Nawaz-ul-Rehman, M. S., Briddon, R. W. and Fauquet, C. M. 2012. A melting pot of old world begomoviruses and their satellites infecting a collection of *Gossypium* species in Pakistan. PLOS ONE, 7: e40050. https: //doi.org/10.1371/journal.pone.0040050.

Nigam, D., LaTourrette, K. and Garcia-Ruiz, H. 2020. Mutations in virus-derived small RNAs. Scientific Reports, 10: 9540. https: //doi.org/10.1038/s41598-020-66374-2.

Nigam, D. 2021. Genomic variation and diversification in begomovirus genome in implication to host and vector adaptation. Plants, 10: 1706. https://doi.org/10.3390/plants10081706.

Noris, E., Vaira, A. M., Caciagli, P., Masenga, V., Gronenborn, B. and Accotto, G. P. 1998. Amino acids in the capsid protein of *Tomato yellow leaf curl virus* that are crucial for systemic infection, particle formation, and insect transmission. Journal of Virology, 72: 10050-10057. https://doi.org/10.1128/jvi.72.12.10050-10057.1998.

Nylander, J. A. A. 2004. MrModeltest ver. 2. Evolutionary Biology Centre, Uppsala University, Sweden.

Ooi, K., Ohshita, S., Ishii, I. and Yahara, T. 1997. Molecular phylogeny of geminivirus infecting wild plants in Japan. Journal of Plant Research, 110: 247-257. https://doi.org/10.1007/BF02509313.

Padidam, M., Sawyer, S. and Fauquet, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology, 265: 218-225. https://doi.org/10.1006/viro.1999.0056.

Paredes-Montero, J. R., Haq, Q. I., Mohamed, A. A. and Brown, J. K. 2021. Phylogeographic and SNPs analyses of *Bemisia tabaci* B mitotype populations reveal only two of eight haplotypes are invasive. Biology, 10: 1048. https://doi.org/10.3390/biology10101048.

Ribeiro, S. G., Ambrozevicius, L. P., Avila, A. C., Bezerra, I. C., Calegario, R. F., Fernandes, J. J., Lima, M. F., De Mello, R. N., Rocha, H. C. and Zerbini, F. M. 2003. Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. Archives of Virology, 148: 281-295. https://doi.org/10.1007/s00705-002-0917-0.

Rojas, M. R., Hagen, C., Lucas, W. J. and Gilbertson, R. L. 2005. Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. Annual Review of Phytopathology, 43: 361-394. https://doi.org/10.1146/annurev.phyto.43.040204.135939.

Ronquist, F. and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19: 1572-1574. https://doi.org/10.1093/bioinformatics/btg180.

Roossinck, M. J. 2019. Viruses in the phytobiome. Current Opinion in Virology, 37: 72-76. https://doi.org/10.1016/j.coviro.2019.06.008.

Rozas, J., Sanchez-Del Barrio, J. C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19: 2496-2497. https://doi.org/10.1093/bioinformatics/btg359.

Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Júnior, B. T., Assunção, I. P., Lima, G. S. A., Pio-Ribeiro, G., Mizubuti, E. S. G. and Zerbini, F. M. 2012. Species diversity, phylogeny and genetic variability of begomovirus populations infecting

leguminous weeds in Northeastern Brazil. Plant Pathology, 61: 457-467. http://dx.doi.org/10.1111/j.1365-3059.2011.02543.x.

Simmonds, P., Aiewsakun, P. and Katzourakis, A. 2019. Prisoners of war—host adaptation and its constraints on virus evolution. Nature Reviews Microbiology, 17: 321-328. https://doi.org/10.1038/s41579-018-0120-2.

Sobrinho, R. R., Xavier, C. A. D., Pereira, H. M. B., Lima, G. S. A., Assuncao, I. P., Mizubuti, E. S. G., Duffy, S. and Zerbini, F. M. 2014. Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. Journal of General Virology, 95: 2540-2552. https://doi.org/10. 1099/vir.0.067009-0.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28: 2731-2739. https://doi.org/10.1093/molbev/msr121.

Tsompana, M., Abad, J., Purugganan, M. and Moyer, J. W. 2005. The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. Molecular Ecology, 14: 53-66. https://doi.org/10.1111/j.1365-294X.2004.02392.x.

Urbino, C., Gutierrez, S., Antolik, A., Bouazza, N., Doumayrou, J., Granier, M., Martin, D. P. and Peterschmitt, M. 2013. Within-host dynamics of the emergence of *Tomato yellow leaf curl virus* recombinants. PloS ONE, 8: e58375. https://doi.org/10.1371/journal.pone.0058375.

Wartig, L., Kheyr-Pour, A., Noris, E., De Kouchkovsky, F., Jouanneau, F., Gronenborn, B. and Jupin, I. 1997. Genetic analysis of the monopartite tomato yellow leaf curl geminivirus: roles of V1, V2, and C2 ORFs in viral pathogenesis. Virology, 228: 132-140. https://doi.org/10.1006/viro.1996.8406.

Webster, C. G., Coutts, B. A., Jones, R. A. C., Jones, M. G. K. and Wylie, S. J. 2007. Virus impact at the interface of an ancient ecosystem and a recent agroecosystem: studies on three legume-infecting potyviruses in the Southwest Australian floristic region. Plant Pathology, 56: 729-742. https://doi.org/10.1111/j.1365-3059.2007.01653.x.

Xia, X. and Yuen, K. Y. 2005. Differential selection and mutation between dsDNA and ssDNA phages shape the evolution of their genomic AT percentage. BMC Genetics, 6: 20. https://doi.org/10.1186/1471-2156-6-20.

Yahara, T., Ooi, K., Oshita, S., Ishii, I. and Ikegami, M. 1998. Molecular evolution of a host-range gene in geminiviruses infecting asexual populations of *Eupatorium makinoi*. Genes & Genetic Systems, 73: 137-141. https://doi.org/10.1266/ggs.73.137.

**Supplementary table 1.** List of the full-length sequences of *Tomato yellow leaf curl virus* from several hosts used for gene flow and genetic differentiation among populations categorized based on host plant.

| Accession | Host | Country | Accession | Host | Country |
|---|---|---|---|---|---|
| KY284012 | *Solanum lycopersicum* | India | GU076447 | *Solanum lycopersicum* | Iran |
| MT551610 | *Solanum lycopersicum* | India | KX347163 | *Solanum lycopersicum* | Iran |
| MT551618 | *Solanum lycopersicum* | India | KX347165 | *Solanum lycopersicum* | Iran |
| MT551621 | *Solanum lycopersicum* | India | KX347166 | *Solanum lycopersicum* | Iran |
| MT551616 | *Solanum lycopersicum* | India | OQ319127 | *Solanum lycopersicum* | Iran |
| MT551617 | *Solanum lycopersicum* | India | JQ414025 | *Solanum lycopersicum* | Iran |
| MT551620 | *Solanum lycopersicum* | India | JQ928347 | *Solanum lycopersicum* | Iran |
| MT551615 | *Solanum lycopersicum* | India | JQ928348 | *Solanum lycopersicum* | Iran |
| MT551611 | *Solanum lycopersicum* | India | EU085423 | *Solanum lycopersicum* | Iran |
| MT551612 | *Solanum lycopersicum* | India | JQ231214 | *Solanum lycopersicum* | Iran |
| MT551613 | *Solanum lycopersicum* | India | GU076451 | *Solanum lycopersicum* | Iran |
| MT551614 | *Solanum lycopersicum* | India | AJ132711 | *Solanum lycopersicum* | Iran |
| MT551619 | *Solanum lycopersicum* | India | EU635776 | *Solanum lycopersicum* | Iran |
| MN397780 | *Solanum lycopersicum* | Saudi Arabia | GU076441 | *Solanum lycopersicum* | Iran |
| KT355023 | *Corchorus olitorius* | Saudi Arabia | GU076448 | *Solanum lycopersicum* | Iran |
| KT033715 | *Solanum lycopersicum* | Saudi Arabia | GU076449 | *Solanum lycopersicum* | Iran |
| KT033709 | *Solanum lycopersicum* | Saudi Arabia | GU076452 | *Solanum lycopersicum* | Iran |
| KF435137 | *Solanum lycopersicum* | Saudi Arabia | GU076453 | *Solanum lycopersicum* | Iran |
| KF561125 | *Solanum lycopersicum* | Saudi Arabia | MH507499 | *Carica papaya* | Iran |
| KT728745 | *Cucumis sativus* | Saudi Arabia | GU076442 | *Solanum lycopersicum* | Iran |
| KT728744 | *Cucumis sativus* | Saudi Arabia | GU076443 | *Solanum lycopersicum* | Iran |
| KT728743 | *Cucumis sativus* | Saudi Arabia | GU076454 | *Solanum lycopersicum* | Iran |
| KC845301 | *Solanum lycopersicum* | Saudi Arabia | KT990213 | *Solanum lycopersicum* | Iran |
| KT728752 | *Solanum lycopersicum* | Saudi Arabia | KY825714 | *Brassica rapa* | Ira |
| KT728746 | *Solanum lycopersicum* | Saudi Arabia | MF536415 | *Physalis divaricata* | Iran |
| MN397779 | *Solanum lycopersicum* | Saudi Arabia | MZ911862 | *Cyamopsis tetragonoloba* | Iran |
| KT033713 | *Cucumis sativus* | Saudi Arabia | GU076450 | *Solanum lycopersicum* | Iran |
| MG571546 | *Mentha longifolia* | Saudi Arabia | JQ928346 | *Solanum lycopersicum* | Iran |
| KT033706 | *Solanum lycopersicum* | Saudi Arabia | ON254272 | *Solanum lycopersicum* | Iraq |
| KU248482 | *Ridge gourd* | Saudi Arabia | OP771625 | *Solanum lycopersicum* | Iraq |
| KF435136 | *Capsicum annuum* | Saudi Arabia | MT583814 | *Solanum lycopersicum* | Iraq |
| ON756220 | *Solanum lycopersicum* | Saudi Arabia | JQ354991 | *Solanum lycopersicum* | Iraq |
| ON756221 | *Solanum lycopersicum* | Saudi Arabia | EF054894 | *Solanum lycopersicum* | Jordan |
| ON756218 | *Solanum lycopersicum* | Saudi Arabia | EF158044 | *Cucumis sativus* | Jordan |
| ON756219 | *Solanum lycopersicum* | Saudi Arabia | EU143745 | *Cucumis sativus* | Jordan |
| KX347156 | *Solanum lycopersicum* | Iran | GQ861427 | *Solanum lycopersicum* | Jordan |
| KX347162 | *Solanum lycopersicum* | Iran | EF054893 | *Solanum lycopersicum* | Jordan |
| KX347155 | *Solanum lycopersicum* | Iran | EF433426 | *Cucumis sativus* | Jordan |
| KX347157 | *Solanum lycopersicum* | Iran | JX444575 | *Solanum lycopersicum* | Jordan |
| KX347158 | *Solanum lycopersicum* | Iran | MG930479 | *Solanum lycopersicum* | Jordan |
| KX347159 | *Solanum lycopersicum* | Iran | GQ861426 | *Solanum lycopersicum* | Jordan |
| KX347160 | *Solanum lycopersicum* | Iran | JX131286 | *Sinapis arvensis* | Jordan |
| KX347161 | *Solanum lycopersicum* | Iran | MG930477 | *Solanum lycopersicum* | Jordan |
| KX347164 | *Solanum lycopersicum* | Iran | MF766109 | *Solanum lycopersicum* | Jordan |
| OQ319123 | *Ocimum basilicum* | Iran | MG930476 | *Solanum lycopersicum* | Jordan |
| OQ319125 | *Vicia faba* | Iran | KJ830842 | *Solanum lycopersicum* | Kuwait |
| OQ319126 | *Euphorbia sp.* | Iran | KR108214 | *Solanum lycopersicum* | Kuwait |
| OQ319130 | *Dracocephalum moldavica* | Iran | KJ830841 | *Solanum lycopersicum* | Kuwait |
| GU076440 | *Solanum lycopersicum* | Iran | JF451352 | *Solanum lycopersicum* | Kuwait |
| GU076444 | *Solanum lycopersicum* | Iran | OL890677 | *Solanum lycopersicum* | Kuwait |
| GU076445 | *Solanum lycopersicum* | Iran | OL890678 | *Solanum lycopersicum* | Kuwait |
| GU076446 | *Solanum lycopersicum* | Iran | OL890679 | *Solanum lycopersicum* | Kuwait |

**Continued supplementary table 1.**

| Accession | Host | Country |
|---|---|---|
| OM691684 | *Solanum lycopersicum* | Kuwait |
| OL890666 | *Solanum lycopersicum* | Kuwait |
| HG969254 | *Solanum lycopersicum* | Oman |
| KF229722 | *Solanum lycopersicum* | Oman |
| KF229721 | *Solanum lycopersicum* | Oman |
| MF996518 | *Solanum lycopersicum* | Oman |
| KF229726 | *Solanum lycopersicum* | Oman |
| KF229725 | *Solanum lycopersicum* | Oman |
| KF229724 | *Solanum lycopersicum* | Oman |
| KF229723 | *Solanum lycopersicum* | Oman |
| MT800821 | *Cucurbita pepo* | Pakistan |
| MT800822 | *Cucurbita pepo* | Pakistan |
| MT800823 | *Cucurbita pepo* | Pakistan |
| MT800824 | *Cucurbita pepo* | Pakistan |
| MT800838 | *Solanum melongena* | Pakistan |
| MT800839 | *Solanum melongena* | Pakistan |
| MT800840 | *Solanum melongena* | Pakistan |
| KX710157 | *Cyamopsis tetragonoloba* | Pakistan |
| MG210483 | *Solanum lycopersicum* | Pakistan |
| MG210484 | *Solanum lycopersicum* | Pakistan |
| ON864378 | *Solanum lycopersicum* | Syria |
| ON864380 | *Solanum lycopersicum* | Syria |
| ON864382 | *Solanum lycopersicum* | Syria |
| ON864372 | *Solanum lycopersicum* | Syria |
| ON864374 | *Solanum lycopersicum* | Syria |
| ON864376 | *Solanum lycopersicum* | Syria |
| AJ812277 | *Solanum lycopersicum* | Turkey |
| OM691682 | *Solanum lycopersicum* | Kuwait |

**Supplementary Table 2.** The evolutionary parameters measured for the full-length viral genomes, open reading frame (ORF), and intergenic region (IR) of *Tomato yellow leaf curl virus*-IL isolates associated with *Datura stramonium* plants from Bojnourd, Iran.

| Data set | Parameter | Full genome | V1 | V2 | C1 | C4 | C2 | C3 | IR |
|---|---|---|---|---|---|---|---|---|---|
| *Datura* 23 | Substitution best-fit model[1] | T92 | - | - | T92 | - | JC | - | - |
| | Neutral mutation rate | $1.05 \times 10^{-5}$ | - | - | $2.60 \times 10^{-5}$ | - | $6.75 \times 10^{-5}$ | - | - |
| | Sequence length(nt) | 2782 | 777 | 351 | 1056-1074 | 309 | 408 | 405 | 314 |
| | Sequence no. | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 |
| | Ts/Tv[2] | 0 | - | - | 0 | - | 0.50 | - | - |
| | dN[3] | - | - | - | 0.0142 | - | 0.0026 | - | - |
| | dS[4] | - | - | - | 0.0106 | - | 0 | - | - |
| | dN/dS | - | - | - | **1.3490** | - | - | - | - |
| | Tajima's D | - | - | - | -1.0548 | - | -1.0548 | - | - |
| *Datura* 28 | Substitution best-fit model[1] | T92 | T92 | - | GTR | JC | - | - | - |
| | Neutral mutation rate | $3.17 \times 10^{-5}$ | $7.51 \times 10^{-4}$ | - | $5.42 \times 10^{-5}$ | $9.43 \times 10^{-5}$ | - | - | - |
| | Sequence length(nt) | 2781 | 777-840 | 351 | 1074 | 309 | 408 | 405 | 313 |
| | Sequence no. | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| | Ts/Tv[2] | 2.0 | 0.67 | - | 0 | 0.50 | - | - | - |
| | dN[3] | - | 0.0706 | - | 0.0044 | 0 | - | - | - |
| | dS[4] | - | 0.0434 | - | 0 | 0.01483 | - | - | - |
| | dN/dS | - | **1.6263** | - | - | 0 | - | - | - |
| | Tajima's D | - | -1.6843 | - | -1.2371 | -1.0062 | - | - | - |
| Three *Datura* plant (5,23 and 28) | Substitution best-fit model[1] | T92 | T92 | JC | T92 | GTR | JC | T92 | JC |
| | Neutral mutation rate | $1.08 \times 10^{-4}$ | $2.00 \times 10^{-4}$ | $5.26 \times 10^{-5}$ | $8.16 \times 10^{-5}$ | $1.99 \times 10^{-5}$ | $4.52 \times 10^{-5}$ | $6.08 \times 10^{-5}$ | $3.14 \times 10^{-4}$ |
| | Sequence length(nt) | 2781-2782 | 777-840 | 351 | 1056-1074 | 309 | 408 | 405 | 313-314 |
| | Sequence no. | 22 | 22 | 23 | 23 | 23 | 23 | 23 | 23 |
| | Ts/Tv[2] | 1.62 | 0.79 | 0.50 | 3.68 | 0 | 0.50 | 3.02 | 0.50 |
| | dN[3] | - | 0.0729 | 0.0033 | 0.0195 | 0 | 0.0065 | 0.0034 | - |
| | dS[4] | - | 0.0682 | 0.0163 | 0.0464 | 0.0098 | 0.0228 | 0.0231 | - |
| | dN/dS | - | **1.0690** | 0.2052 | 0.4215 | 0 | 0.2860 | 0.1477 | - |
| | Tajima's D | - | -1.4952 | **1.8144** | **1.5127** | -1.1609 | **0.6986** | **1.9984** | **2.5149** |
| Six *Datura* plants (5,8,11,23,28 and 132) | Substitution best-fit model[1] | T92 + G | T92 | JC | T92 | JC | JC | T92 | T92 |
| | Neutral mutation rate | $1.45 \times 10^{-4}$ | $1.94 \times 10^{-4}$ | $5.68 \times 10^{-5}$ | $1.37 \times 10^{-4}$ | $8.07 \times 10^{-5}$ | $6.11 \times 10^{-5}$ | $7.39 \times 10^{-5}$ | $4.02 \times 10^{-4}$ |
| | Sequence length(nt) | 2779-2782 | 777-840 | 351 | 1056-1074 | 309 | 408 | 405 | 311-314 |
| | Sequence no. | 26 | 26 | 27 | 27 | 27 | 27 | 27 | 27 |
| | Ts/Tv[2] | 1.52 | 0.88 | 0.50 | 3.01 | 0.50 | 0.50 | 5.01 | 0.63 |
| | dN[3] | - | 0.0742 | 0.0090 | 0.0290 | 0.0162 | 0.0096 | 0.0068 | - |
| | dS[4] | - | 0.0826 | 0.0131 | 0.0868 | 0.0139 | 0.0326 | 0.0675 | - |
| | dN/dS | - | 0.8979 | 0.6880 | 0.3347 | **1.1663** | 0.2939 | 0.1014 | - |
| | Tajima's D | - | -1.7053 | **1.1446** | -0.1570 | -1.4126 | -0.1929 | **0.9956** | **1.0424** |

1. Abbreviations- GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura- Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.
2. Transition/Transversion ratio (R).
3. Mean non-synonymous substitutions per non-synonymous site.
4. Mean synonymous substitutions per synonymous site.

# ریز فرگشت جمعیت‌های سویه IL ویروس پیچیدگی برگ زرد گوجه-فرنگی از گیاهان وحشی تاتوره در ایران مرکز تنوع این ویروس

**محمدرضا حسین‌زاده[۲،۱]، سجاد آسترکی[۱]، جودیت کی. براون[۳] و مسعود شمس‌بخش[۱*]**

۱- گروه بیماری‌شناسی گیاهی، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران.
۲- گروه گیاه‌پزشکی، واحد بجنورد، دانشگاه آزاد اسلامی، بجنورد، ایران.
۳- دانشکده علوم گیاهی، دانشگاه آریزونا، توسان، آریزونا ۸۵۷۲۱، ایالات متحده آمریکا.
پست الکترونیکی نویسنده مسئول مکاتبه: shamsbakhsh@modares.ac.ir
دریافت: ۱۰ دی ۱٤۰٤؛ پذیرش: ۱۸ دی ۱٤۰٤

**چکیده:** پنج تا از هفت استرین شناخته شده ویروس پیچیدگی برگ زرد گوجه‌فرنگی (TYLCV) ازجمله سویه خطرناک اسرائیلی (TYLCV-IL) در ایران که به‌نظر می‌رسد مرکز تنوع این ویروس است، وجود دارد. در این بررسی، فیلوژنی، ساختار جمعیت و الگوهای ریز فرگشت هاپلوتیپ‌های جمعیت‌های TYLCV-IL آلوده‌کننده گیاهان وحشی تاتوره *Datura stramonium* که در اطراف مزارع گوجه‌فرنگی تجاری در بجنورد رشد می‌کردند، مورد مطالعه قرار گرفت. تنوع ژنومی TYLCV در تاتوره در درجه اول با انتخاب طبیعی در مناطق رمزکننده پروتئین همراه با همانندسازی و منطقه بین‌ژنی غیر کدکننده درگیر در تنظیم همانندسازی و رونویسی رخ داده است. ژن C4 دارای سیگنال‌های سازگاری با میزبان و عملکردهای شناخته شده در حرکت ویروس در مسیر طولانی و خاموشی ژن گیاه میزبان است. تجزیه‌وتحلیل نوترکیبی نوترکیب‌های بین‌گونه‌ای ویروس را در میزبان تاتوره در مناطق کدکننده C1 و C1/C4 و منطقه بین‌ژنی (IR) نشان داد. نرخ جایگزینی نوکلئوتید $4.02 \times 10^{-4}$ IR در هر سایت نشان‌دهنده انعطاف‌پذیری ژنوم بود. علاوه‌بر این، تنوع ژنومی بالای جدایه‌های TYLCV مرتبط با جمعیت‌های میزبان وحشی به تغییرات در میزبان نسبت داده شد که با وجود انواع منحصربه‌فرد در هر گیاه میزبان مشهود است. درنهایت، نوترکیبی و تجزیه‌وتحلیل ساختار جمعیت نشان می‌دهد که جهش و نوترکیبی بین‌گونه‌ای به تمایز جمعیت‌های TYLCV-IL و جداسازی ژنتیکی بعدی توسط میزبان کمک کرده است.

**واژگان کلیدی:** بگوموویروس، جایگزینی‌های نوکلئوتیدی، تکامل ویروس گیاهی، نوترکیبی