Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

Spatial distribution and genetic diversity of TYLCV in Saudi Arabia

Zafar Iqbal*

Central Laboratories, King Faisal University, PO Box 31982, Al-Ahsa, Saudi Arabia

Article Info

Article history:

Received: January 24, 2025
Accepted: March 23, 2025
Published: April 30, 2025

Use your device to scan and read the article online

Abstract

This study investigated the genetic diversity, phylogeny, and evolutionary dynamics of Tomato yellow leaf curl virus (TYLCV) across seven regions in Saudi Arabia. Analyzing 28 full-length TYLCV genomes, phylogenetic analysis revealed two distinct clades: one predominantly comprised of isolates from the Ahsa (Eastern province) region and the other encompassing isolates from Northern and Western regions. The Ahsa region exhibited significantly higher TYLCV prevalence and genetic diversity, harboring the most divergent isolates with high haplotype (Hd = 1.00) and nucleotide (π = 0.079) diversity. Conversely, regions like Jeddah and Hadasham showed lower diversity, suggesting less or stable viral populations. Genetic diversity analyses revealed high variation in coding regions like CP and Rep, which are under strong selective pressures and prone to recombination. Conversely, V2 displayed lower diversity, indicating purifying selection. Selection pressure analysis using dN/dS ratios indicated diversifying selection in C4 (2.20) and Rep (1.28). Single Likelihood Ancestor Counting identified one positively selected site in Rep. In contrast, Fast Unconstrained Bayesian AppRoximation identified multiple sites in C4 (8), TrAP (7), REn (6), and V2 (1), suggesting roles in host adaptation and immune evasion. A total of 32 credible recombination events, predominantly in Ahsa isolates, were identified using RDP and confirmed by GARD analysis. These events, involving both inter- and intraspecies recombination, play a crucial role in enhancing TYLCV genetic diversity and adaptability. The conservation of motifs in V2 and C4 indicated their essential roles in TYLCV function. In contrast, variations in ORFs like CP, Rep, TrAP, and REn among specific isolates may promote viral diversity and adaptation. This study demonstrates the crucial role of geographic and genetic factors, with Ahsa as a key hub for TYLCV diversity, in driving viral evolution and diversification. The findings emphasize the need to monitor regions with high viral diversity, like Ahsa, and develop strategies to manage TYLCV's swift spread.

Keywords: Begomovirus; Genetic diversity; Phylogenetic analysis; Recombination; Saudi Arabia; Tomato yellow leaf curl virus

1. Introduction

Tomato yellow leaf curl virus (TYLCV), is a single-stranded circular DNA virus classified under the largest genus *Begomovirus* within the family *Geminiviridae*. This monopartite begomovirus contains a ~2.8 kb genome that encodes four proteins: C1 (replication-associated protein, Rep), C2 (transcriptional activator protein, TrAP), C3 (replication enhancer protein, REn), and C4 (involved in symptom severity, host range, and movement) on the complementary strand. On the virion sense strand, TYLCV encodes the coat protein (CP [V1]) and a pre-coat protein (V2). Transmitted primarily by the whitefly (*Bemisia tabaci*), TYLCV causes tomato yellow leaf curl disease (TYLCD) in tomato plants around the globe. TYLCV is a highly devastating pathogen with a broad host range and can affect around 50 plant species, including tomatoes, cotton, chilies, tobacco, watermelon, potato, and cucurbits [1]. TYLCV is among the most economically damaging plant viruses, with reported 100% crop losses in various regions. In Saudi Arabia and other countries of Arabian Peninsula, TYLCV has been a major contributor to TYLCD, significantly limiting tomato production. It is now the biggest factor restricting tomato output worldwide and can result in significant losses. Notably, TYLCV is a seed-transmissible geminivirus. Seed infectivity rates ranged from 20% to 100%, as demonstrated in viruliferous whitefly-mediated transmission and agro-inoculation experiments [2]. This dual mode of transmission highlights its adaptability and the persistent challenge it poses to global agriculture.

TYLCV was first reported in Saudi Arabia during the 1970s, coinciding with its global emergence as a devastating tomato pathogen [3]. In the Eastern Province, specifically Ahsa, the virus was reported in the early 1990s [4]. Early outbreaks were linked to increasing tomato cultivation and the widespread presence of its vector, *B. tabaci*, particularly the invasive MEAM1 biotype, which thrives in Saudi Arabia's warm climate [5]. Over the decades, TYLCV has become endemic in major agricultural regions like Ahsa, Riyadh, Qassim, Tabuk, Hail, and Jazan [6]. Molecular studies revealed the co-circulation of multiple TYLCV strains in Saudi Arabia, some introduced

* Corresponding author.
E-mail address: zafar@kfu.edu.sa (Z. Iqbal).
Doi: http://dx.doi.org/10.14715/cmb/2025.71.4.7

via infected plant materials [7]. Additionally, the virus has been detected in both commercial and backyard tomato fields, demonstrating its adaptation to diverse agro-ecological zones. Early molecular investigations were performed through PCR, restriction fragment length polymorphism, and DNA sequencing. These studies identified TYLCV strains closely related to those from Egypt, Iran, and Israel [8]. However, recent studies uncovered unique Saudi-specific strains, indicating significant genetic evolution and the emergence of diverse variants in the region.

RNA viruses have long been a focus of evolutionary studies due to their error-prone RNA-dependent RNA polymerases, leading to high mutation rates. However, recent studies have demonstrated that certain DNA viruses exhibit rapid evolutionary rates, akin to RNA viruses [9, 10]. The rapid genetic diversification in ssDNA viruses arises from low-fidelity DNA polymerase activity and biochemical reactions like methylation and deamination [11, 12]. For geminiviruses, point mutations and recombination are key drivers of genetic variation. In Saudi Arabia, several factors may contribute to the genetic diversity of TYLCV in Saudi Arabia, including 1) vector population dynamics, 2) host plant diversity, 3) climatic conditions, and 4) agricultural practices. These factors are crucial for virus's adaptability to environmental pressures. They also influence its ability to overcome host resistance and persistence within agro-ecosystems [13].

Understanding the spatial and genetic diversity of TYLCV in Saudi Arabia is crucial for developing effective disease management strategies, particularly as the country strives to enhance agricultural productivity, especially in the Eastern province. By identifying the dominant TYLCV strains in specific regions and monitoring their evolution, researchers can contribute to the selection of resistant tomato cultivars. Additionally, this knowledge gap can guide the implementation of appropriate pest control measures. Moreover, tracking genetic diversity over time can help predict the emergence of more virulent strains, enabling timely interventions. This manuscript aims to explore the spatial and genetic diversity of TYLCV in Saudi Arabia through a comprehensive analysis that integrates bioinformatics approaches to elucidate the virus's distribution, genetic variability, and the key factors driving its evolution and spread.

2. Materials and Methods

2.1. Retrieval and alignment of TYLCV sequences

Full-length genome sequences of TYLCV were retrieved from NCBI GenBank (https://www.ncbi.nlm.nih.gov/), representing all full-length TYLCV sequences available in the database on November 11, 2024. Datasets containing both full-length nucleotide sequences and all encoded open reading frames (ORFs; V2, CP, Rep, TrAP, REn, and C4) were generated for each isolate categorized by its city of origin. Multiple sequence alignments (MSA) were then performed independently on each dataset using the Muscle program within MEGA11 software [14]. After alignment, all files were thoroughly reviewed and manually adjusted to ensure accuracy.

2.2. Phylogenetic analysis

To infer phylogenetic relationships among the retrieved TYLCV sequences, maximum likelihood (ML) analysis was performed. The best-fit nucleotide substitution model was determined using MEGA11 based on MSAs. This model was then used for ML tree construction, with nodal support assessed using 1000 bootstrap replicates. The resulting tree was visualized and edited using ChiPlot (https://chiplot.online/).

2.3. Evolutionary Probability of TYLCV

To infer the evolutionary probability (EP) distribution of nucleotide variations across TYLCV genomic sites, MEGA 11 was employed. All 28 full-length TYLCV-aligned sequences, along with CLCuMuV AJ496287 as an outgroup, were utilized to construct an ML tree. This ML tree was then used to estimate EP values using the Bayesian method with the Tamura-Nei model and a Gamma distribution of 2. Subsequently, OriginPro (v2024) was used to visualize the EP distribution in a scatter plot (https://www.originlab.com/).

2.4. TYLCV population structure assay

DNA polymorphism analysis was conducted using DnaSP v.6.12 (Rozas et al., 2017). Nucleotide diversity (π), a measure of average nucleotide differences per site, was calculated for all datasets. Significant differences in mean π values were assessed using 95% bootstrap confidence intervals. A window size of 100 nucleotides and a step size of 25 nucleotides were used, allowing for the identification of regions with significant variation.

Neutrality tests, including Fu and Li's D (FLD) and Tajima's D (TD), were performed on all datasets. These tests evaluated deviations from neutrality based on the mean number of pairwise differences and segregating sites.

2.5. Inference of selection pressure

To assess potential selection pressures acting on TYLCV-encoded ORFs, the ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated using two complementary approaches. Initial dN/dS estimations were performed using MEGA11 with standard parameters. To further investigate site-specific selection, the Datamonkey web server (www.datamonkey.org) was utilized, employing the Fast Unconstrained Bayesian AppRoximation (FUBAR) and Single Likelihood Ancestor Counting (SLAC) methods within the HyPhy package. Both analyses were conducted using the HyPhy package available on the Datamonkey web server (www.datamonkey.org). Prior to FUBAR and SLAC analyses, each ORF nucleotide dataset was aligned, stop codons were excised, and nucleotides gaps were removed. Additionally, DnaSP was used to further evaluate selection pressures within the TYLCV population.

2.6. Recombination analysis

Potential recombination events within the TYLCV genomes were investigated the Recombination Detection Program (RDP v.5.5) [15]. RDP analysis utilized seven different algorithms (BOOTSCAN, CHIMAERA, GENECONV, RDP, MAXCHI, SISCAN, and 3SEQ) to identify potential recombination breakpoints with default detection thresholds and a Bonferroni-corrected *p*-value of 0.05. Only recombination events detected by at least four algorithms were considered credible. Subsequently, to validate the RDP-inferred recombination breakpoints, GARD analysis was conducted using a Beta-Gamma site-to-site variation model with four rate classes and normal run mode.

2.7. Motif prediction

To identify conserved functional motifs within the aligned TYLCV ORF sequences, the MEME suit was employed (https://meme-suite.org/meme/tools/meme) [16]. For robust and specific motif identification, MEME was configured to allow only one motif per sequence. Additionally, the total numbers of identified motifs in the dataset were limited to ten. This prioritizes the discovery of high-confidence, non-redundant motifs potentially associated with the TYLCV function.

2.8. Statistical analysis

All the potential recombination breakpoints using the RDP software were inferred with a stringent Bonferroni-corrected *p*-value cutoff of 0.05 to minimize the false positives.

The ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated using two complementary approaches to ensure robustness. First, initial dN/dS estimations were conducted using the Nei-Gojobori method in MEGA11, with a *p*-distance threshold of ≤ 0.05 to account for sequence divergence. Second, to identify sites under positive selection, significance thresholds were set at $p \leq 0.05$ for the SLAC method and a posterior probability of $p \geq 1.5$ for the FUBAR method.

3. Results

3.1. Distribution of TYLCV

A total of 28 full-length TYLCV sequences were included in this study (Fig. 1). These sequences originated from seven distinct regions across Saudi Arabia. The Ahsa region had the highest number of TYLCV sequences (13), followed by Jeddah (5) and Hadasham (5). In contrast, only one full-length TYLCV sequence was reported from each of the Hail, Qassim, Riyadh, and Jazan regions.

3.2. Phylogenetic analysis

To elucidate the phylogeographic spread of TYLCV in Saudi Arabia, phylogenetic analysis was performed on 28 full-length viral sequences (Fig. 2). This analysis revealed two major distinct phylogenetic clades, indicating a clear pattern of regional clustering rather than segregation based on host species. The first major clade was further subdivided into five subclades. Isolates from the geographically proximal regions of Qassim, Riyadh, Hail, and Tabuk clustered within the first two subclades. While isolates from Jeddah and Hadasham formed the remaining three subclades, reinforcing the observed regional structuring of TYLCV populations.

The second major clade also comprised five subclades, but all isolates within this clade originated exclusively from the Ahsa region. Notably, the analysis identified the oldest, most recent, and most divergent TYLCV isolate as originating from Ahsa. This strongly suggested that Ahsa region has played a critical role in TYLCV diversification and evolution within Saudi Arabia.

3.3. Evolution Probability of TYLCV

The EP plot highlights regions of high and low nucleotide diversity across the TYLCV genome (Fig. 3A). The pairwise comparison plot provides insights into the population's genetic diversity and evolutionary dynamics. The presence of multiple nucleotides at a site indicates genetic variation, while patterns of nucleotide substitution may reflect selective pressures. Off-diagonal scatter plots depict pairwise comparisons of nucleotide frequencies, revealing correlation and substitution biases among nucleotides. Notably, a distinct pattern emerges in the relationship between different nucleotide substitutions (T-A, T-C, T-G), with clustered dots indicating specific relationships. The first 600-800 nucleotides exhibit a positive correlation between nucleotide substitutions, while the middle region shows no specific relationship. In contrast, a negative correlation is observed after the middle region (2000-2800 nucleotides).

Regions with denser dot clusters suggested higher variation, potentially indicating genetic recombination or selection hotspots. The nucleotide frequency distribution histograms provide insights into the overall nucleotide composition of the viral population. The x-axis in each subplot represents the genomic site or nucleotide frequency. The y-axis represents the respective frequency values or site indices, offering insights into nucleotide distribution and substitution dynamics within the dataset.

3.4. Genetic Diversity Indices

3.4.1. GDIs in TYLCV and its regional populations

The genetic diversity of TYLCV populations in different regions of Saudi Arabia was analyzed based on key parameters. These include the number of sequences, insertion-deletion (InDel) sites, haplotype diversity (Hd), segregating sites (S), total mutations (Eta), number of haplotypes (h), average number of nucleotide differences (k), nucleotide diversity (π), and two measures of population mutation rate (θ-Eta and θ-W) (Table 1, Fig. 3B).

GDI results revealed significant variation in genetic diversity across regions (Table 1). The Ahsa region exhibited the highest genetic diversity, suggesting it as a potential hotspot for TYLCV evolution in Saudi Arabia. In contrast, the Hadasham and Jeddah regions showed lower genetic diversity, indicating more conserved viral lineages. The complete TYLCV dataset (TYLCV-all) exhibited substantial genetic diversity, characterized by 72 InDel sites, 860 segregating sites, and 1,000 total mutations. Haplotype di-

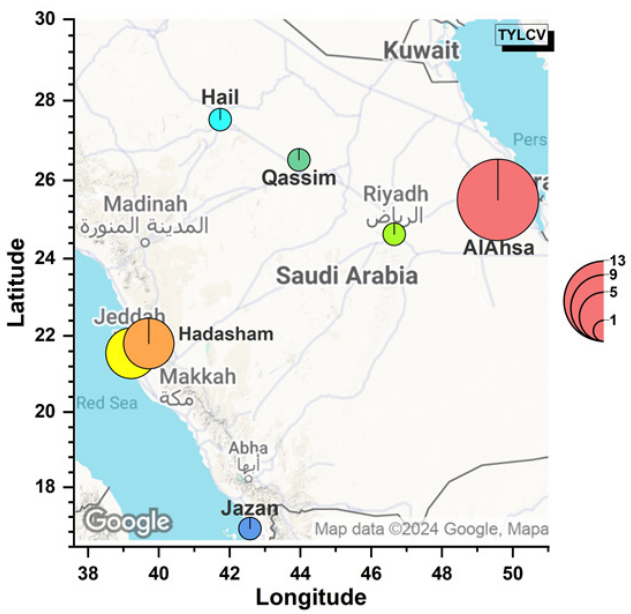

Fig. 1. Geographical distribution and population density map of TYLCV in Saudi Arabia.

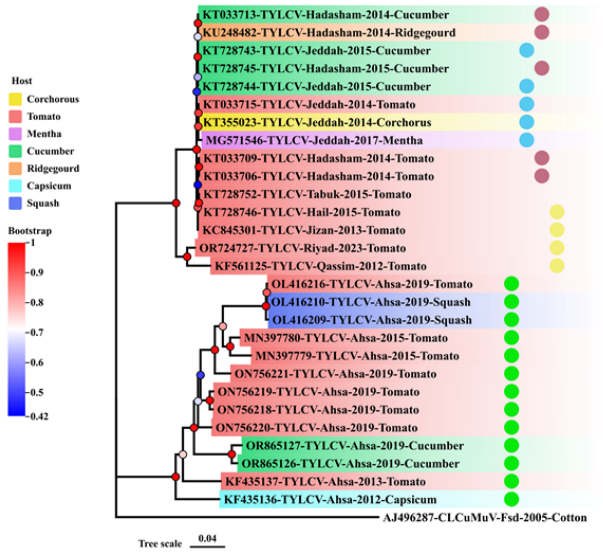

Fig. 2. Phylogenetic tree of TYLCV isolates from Saudi Arabia. A maximum likelihood phylogenetic tree was constructed using MEGA11 with 1000 bootstrap replicates (indicated by the blue-red scale). Isolates from different hosts are represented by distinct colors: tomato (red), cucumber (green), squash (blue), Corchorus (yellow), capsicum (cyan), menthe (violet), and ridge gourd (brown). Likewise, geographic regions are represented by distinct color circles: Ahsa (green), Hadasham (red), Jeddah (blue), and others (yellow). All related isolates included in the analysis were retrieved from the NCBI database and were identified by their accession numbers.

versity (Hd) was high at 0.921, with an average of 296.61 nucleotide differences (k) among 19 haplotypes. Nucleotide diversity (π) was calculated to be 0.110, with θ-Eta and θ-W values of 0.096 and 0.082, respectively (Table 1, Fig. 3B).

The TYLCV-Ahsa subset displayed the highest level of diversity, with 102 InDel sites, 624 segregating sites, and 693 total mutations. All 13 sequences were unique haplotypes (Hd = 1.00), with an average of 215.82 nucleotide differences. Nucleotide diversity (π) was 0.079, with θ-Eta and θ-W values of 0.082 and 0.074, respectively.

In contrast, the TYLCV-Jeddah and TYLCV-Hadasham subsets exhibited low levels of diversity. The TYLCV-Jeddah subset had 13 InDel sites, 11 segregating sites, and 11 total mutations, with an Hd of 0.80 and an average of 5.40 nucleotide differences among three haplotypes. Nucleotide diversity (π) was 0.002, with θ-Eta and θ-W values of 0.002. Similarly, the TYLCV-Hadasham subset had seven InDel sites, six segregating sites, and six total mutations, with an Hd of 0.60 and an average of 3.6 nucleotide differences among two haplotypes. Nucleotide diversity (π) was 0.0013, with θ-Eta and θ-W values of 0.0011 and 0.001, respectively.

The TYLCV-Misc subset, comprising the remaining sequences, showed moderate diversity, with 41 InDel sites, 182 segregating sites, and 186 total mutations. Haplotype diversity was 0.783, with an average of 91.9 nucleotide differences among four haplotypes. Nucleotide diversity (π) was 0.033, with θ-Eta and θ-W values of 0.032 and 0.031, respectively.

3.4.2. GDIs in TYLCV-encoded ORFs

Genetic diversity was analyzed across six TYLCV-encoded ORFs: V2, CP, Rep, TrAP, REn, and C4, with results summarized in Table 2.

3.5. Nucleotide diversity across entire nucleotide position

The nucleotide diversity (π) plots for TYLCV popula-

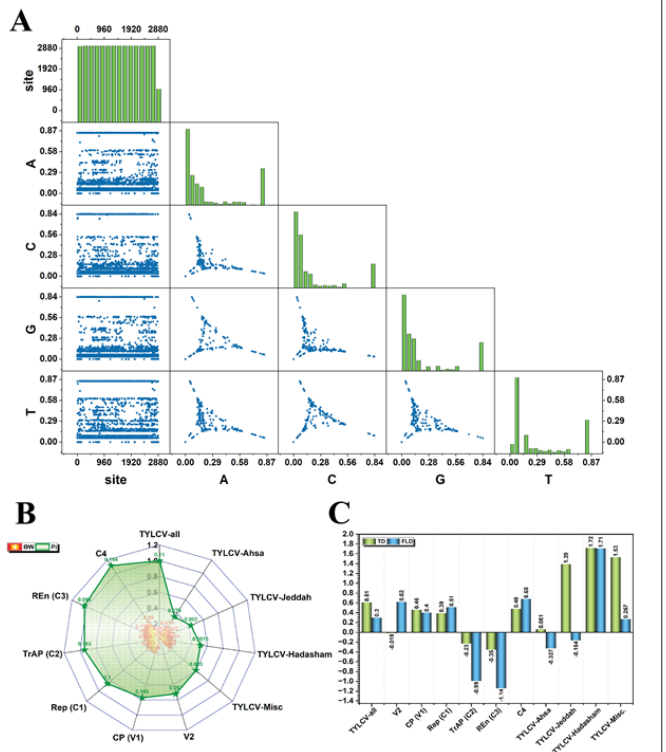

Fig. 3. A) Scatter plot matrix of TYLCV isolates showing pairwise nucleotide diversity at each nucleotide position. The x-axis represents the nucleotide position along the genome, while the y-axis indicates nucleotide diversity. Each dot represents a pairwise comparison between two isolates. The color of the dots corresponds to the nucleotide type (A, C, G, T). The histograms on the diagonal show the nucleotide frequency distribution at each position. **B)** Genetic diversity parameters (π) and Watterson's theta (Θw) estimated for the complete TYLCV population (TYLCV-All), regional populations within Saudi Arabia, and individual ORFs encoded by TYLCV. **C)** Estimation of Tajima's D (TD) and Fu & Li's D (FLD) genetic diversity attributes in complete TYLCV population (TYLCV-All), regional populations within Saudi Arabia, and individual ORFs encoded by TYLCV.

Table 1. The landscape of genetic diversity indices in TYLCV and its regional population.

Virus Population	No. of seq	InDel sites	S	Eta (h)	Eta (s)	Hd	k	h	Θ-Eta
TYLCV-all	28	72	860	1000	128	0.921	296.61	19	0.096
TYLCV-Ahsa	13	102	624	693	292	1.00	215.82	13	0.082
TYLCV-Jeddah	5	13	11	11	6	0.80	5.40	3	0.002
TYLCV-Hadasham	5	7	6	6	0	0.60	3.6	2	0.0011
TYLCV-Misc	5	41	182	186	96	0.783	91.9	4	0.032

Table 2. Landscape of genetic diversity indices in TYLCV-encoded ORFs.

ORF	No. of seq	InDel sites	S	Eta (h)	Eta (s)	Hd	k	h	Θ–Eta
V2	28	0	68	73	13	0.775	18.72	12	0.058
CP	28	11	268	292	62	0.79	103.86	14	0.098
Rep (C1)	28	16	321	377	74	0.794	106.53	16	0.091
TrAP (C2)	28	0	126	153	61	0.79	36.99	14	0.097
REn (C3)	28	1	122	146	60	0.783	34.19	13	0.094
C4	28	9	110	134	23	0.79	38.72	14	0.138

tions from Misc, Ahsa, Jeddah, and Hadasham regions, as well as a combined dataset (TYLCV-all), reveal distinct patterns of genetic variation (Fig. 4). TYLCV-all, TYLCV-Misc, and TYLCV-Ahsa exhibited moderate variability across most of the genome, with notable peaks near the Rep, CP, and REn regions, suggesting higher evolutionary pressure in these functional areas. However, with one principal difference, TYLCV-all showed higher genetic diversity in the CP region than the other two. Conversely, TYLCV-Jeddah and TYLCV-Hadasham displayed significantly lower nucleotide diversity, indicating reduced genetic variability, possibly due to localized outbreaks or limited gene flow between populations.

The nucleotide diversity (π) plots of ORFs showed that specific genomic regions showed higher nucleotide diversity levels than the genome average. These regions include the N-ter (0.06%) and C-ter (0.08%) regions V2, middle region of CP (0.15%), the N-ter region and initial middle region of Rep (0.15%), N-ter of REn (0.14%), and C-ter of C4 (0.15%), (Fig. 4).

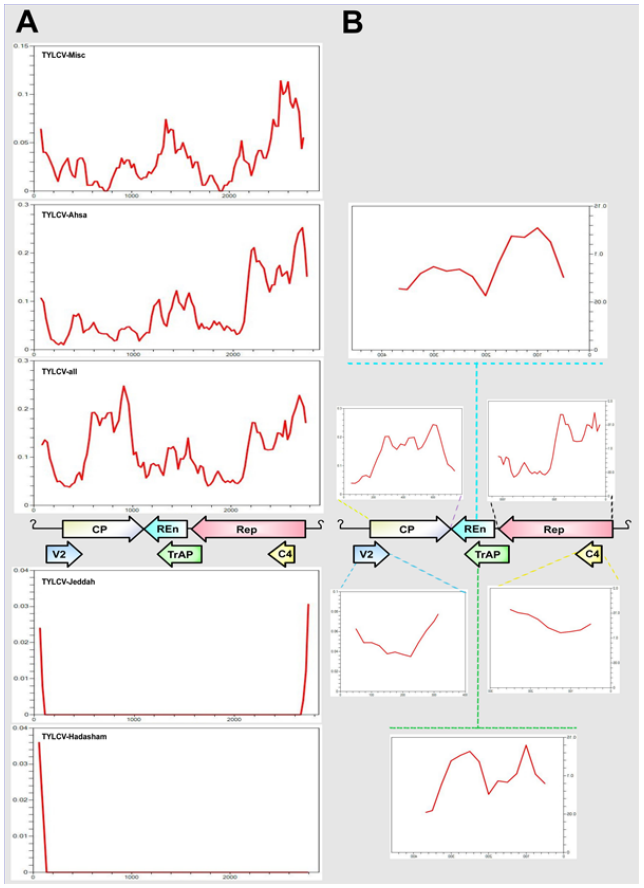

Fig. 4. Nucleotide diversity analysis of the TYLCV population in Saudi Arabia. (A) Nucleotide diversity across the entire genome of TYLCV. (B) Nucleotide diversity within individual TYLCV-encoded ORFs. The linear genome organization of TYLCV is displayed in the center, highlighting all annotated ORFs. The x-axis represents nucleotide positions along the genome, while the y-axis indicates the nucleotide diversity values (π), providing insights into the genetic variation across the viral genome and its functional regions.

3.6. Estimation of selection pressure

To assess the selective pressures acting on different TYLCV-encoded ORFs, a combination of phylogenetic and codon-based substitution models (dN/dS, FUBAR, and SLAC) were employed. The best-fitting nucleotide substitution model for each ORF was determined using MEGA11 (Table 3).

The dN/dS ratio analysis revealed that purifying selection predominates across most TYLCV ORFs, suggesting that these regions are functionally constrained. However, the C4 and Rep ORFs exhibited elevated dN/dS ratios (2.20 and 1.28, respectively), suggesting potential diversifying selection. Conversely, the V2 and CP ORFs displayed the lowest dN/dS ratios (0.302 and 0.39, respectively), indicating strong purifying selection.

SLAC analysis identified a single positively selected site in the Rep. FUBAR analysis, on the other hand, revealed a higher number of positively selected sites in the C4 (8 sites), TrAP (7 sites), and REn (6 sites) ORFs, while V2 had only one. Conversely, the Rep, CP, and REn ORFs exhibited a greater number of sites under negative selection, as identified by both SLAC and FUBAR analyses. However, the number of negatively selected sites identified by SLAC was significantly lower compared to FUBAR (Table 3).

The TD and FLD statistics were calculated to assess the neutrality and evolutionary dynamics of the TYLCV genome, including its individual ORFs and regional populations (Fig. 3C). These results collectively highlight the complex interplay of evolutionary forces acting on TYLCV across different genomic regions and geographic populations.

The V2 showed signs of neutrality with TD and FLD values of –0.019 and 0.62, respectively. The CP showed evidence of balancing selection, as indicated by positive TD (0.46) and FLD (0.40) values. Similarly, the Rep displayed moderate neutrality with TD and FLD values of 0.39 and 0.51, respectively. The TrAP suggested purifying selection with a slightly negative TD value (–0.23) and a more pronounced negative FLD value (–0.99). The REn showed the strongest evidence of purifying selection, with the most negative TD (–0.35) and FLD (–1.14) values among all ORFs. Finally, the C4 ORF showed balancing selection, with positive TD (0.68) and FLD (0.061) values.

The overall TYLCV population showed a slight excess of intermediate-frequency alleles, suggesting potential population expansion or balancing selection, as shown by TD and FLD values of 0.61 and 0.30, respectively (Fig. 3C). The Hadasham population exhibited the highest posi-

Table 3. Sequence variability analysis of ORFs encoded by TYLCV.

ORF	Best Model	Mean distance (d)	dN	dS	dN/dS	Selection sites*			
						SLAC		FUBAR	
						NS	PS	NS	PS
TYLCV-all	TN93+G	--	--	--	--				
V2	T92+G	0.061±0.012	0.038±0.006	0.126±0.029	0.302	2	0	7	1
CP	TN93+G	0.199±0.017	0.10±0.008	0.257±0.022	0.39	10	0	24	0
Rep (C1)	HKY+G	0.125±0.011	0.11±0.009	0.086±0.013	1.28	27	1	81	0
TrAP (C2)	GTR+G	0.114±0.016	0.084±0.013	0.164±0.029	0.51	4	0	7	7
REn (C3)	GTR+G	0.104±0.013	0.073±0.011	0.17±0.028	0.43	7	0	13	6
C4	TN93+G	0.123±0.027	0.178±0.018	0.081±0.024	2.20	1	0	3	8

Abbreviation: Tomato yellow leaf curl virus (TYLCV), non-synonymous (dN), synonymous (dS), Tamura-Nei (TN93), Tamura 3-parameter (T92), general time reversible (GTR), Hasegawa Kishino Yano (HGY), gamma distribution (G), invariable (I), Fast Unconstrained Bayesian AppRoximation (FUBAR), single likelihood ancestor counting (SLAC), positively selected sites (PS), negatively selected sites (NS), and not calculated due to just a few sequences (NC).

* To detect positive selection, the GTR model was employed. Statistical significance was determined using a threshold of $p \leq 0.01$ for the SLAC method and posterior probability ≤ 1.5 for the FUBAR method.

tive TD (1.72) and FLD (1.71) values, suggesting low levels of polymorphism and a balance in selection leading to a reduction in population size. In contrast, the Jeddah and Ahsa populations displayed negative FLD values (–0.327 and –0.164, respectively). This indicated that these populations had an excess of low-frequency polymorphism and are expanding under natural and purifying selection. The Miscellaneous population showed positive TD (1.53) and FLD (0.267) values, further supporting population expansion or balancing selection.

3.7. Recombination analysis

3.7.1. RDP

RDP analysis of TYLCV-all identified 59 potential recombination events across 28 isolates, of which only 32 were deemed credible (Table S1). Notably, TYLCV isolate (ON756220, ON756221, OR865127, MN397779, KF435136, KF435137) from Ahsa region harbored more than one recombination event in their genomes. While all the remaining isolates showed just one recombination event in their genomes. The identified recombination events varied in their breakpoint locations and parental sequences. Notably, isolate OL416209 (isolated from squash in 2019 from Ahsa) was found to be major parent of most isolates. Some events involved major recombination breakpoints within the coding regions of the viral genome, while others occurred in non-coding regions. The parental sequences involved in these recombination events were diverse, suggesting a complex pattern of genetic exchange within the TYLCV population.

To find genomic regions with significantly higher or lower recombination rates in TYLCV genomes, this study analyzed recombination breakpoints within sliding windows across the genome (Fig. 5A). Likelihood ratios were calculated, and statistically significant regions were determined through permutation testing. This permutation test randomly shuffles recombinant fragments while maintaining the same number of variable nucleotide positions. This effectively controls the inherent sequence variability within the alignment. This approach accounts for sequence variability and generates a density map highlighting regions in the TYLCV alignment where recombination is more likely to occur.

Global and local recombination hot and cold spots were identified based on breakpoint frequencies across the TYLCV genome (Fig. 5B). 'Global' hotspots were defined as regions with a higher number of breakpoints than 95% of the sliding windows across the entire genome. While 'local' hotspots were more stringent, requiring a higher breakpoint frequency than 99% of windows at that specific position. Our analysis revealed global recombination hotspots situated within the Rep and C4 regions (Fig. 5B). Conversely, local hotspots were observed in conserved regions, the beginning of the CP gene, and the middle of the Rep gene. In contrast, four local cold spots were identified: in the C-terminus of CP (750 nt), the middle of the REn

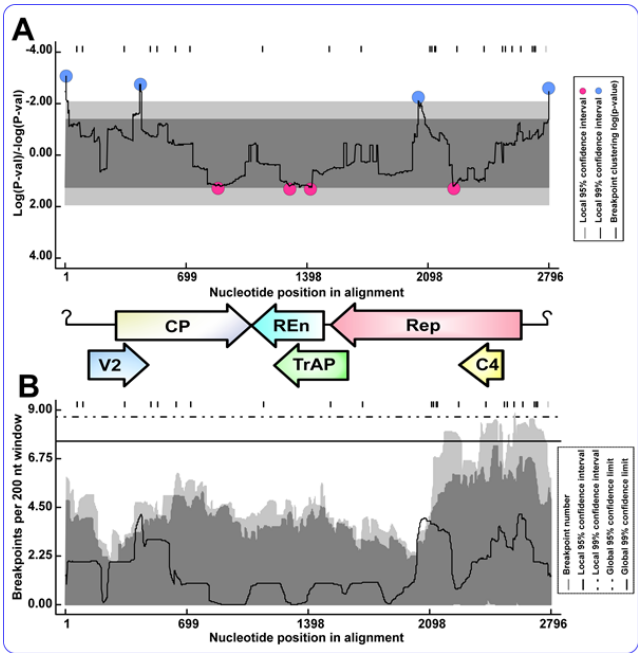

Fig. 5. Recombination breakpoint distribution plot of TYLCV datasets from Saudi Arabia. A) Statistical log converted *p*-value with local cold spot (blue circle) and local hotspot (red circle), and (**B**) recombination breakpoint distribution across TYLCV genomes in a 200 bp window. All the recombination events were identified using 95% and 99% permutation tests and *p*-value distribution for recombination frequency across the TYLCV genome. Light and dark grey ribbons represent the 95% and 99% confidence intervals, respectively, as determined by permutation tests. TYLCV linear genome organization is depicted in the middle.

(1200 nt) and TrAP (1300 nt), the middle of C4 (2500 nt), and the middle and N-terminus of the Rep gene (2100 nt and 2400 nt, respectively) (Fig. 5A). These findings represent the first statistically supported recombination hot and cold spots reported for TYLCV in Saudi Arabia.

3.7.2. GARD

Recombination analysis using GARD, largely consistent with RDP results, identified numerous potential recombination breakpoints in the TYLCV-all, TYLCV-Ahsa, TYLCV-Jeddah, TYLCV-Hadasham, and TYLCV-Misc datasets. However, only a subset of these breakpoints was considered credible based on specific statistical criteria (Fig. S1).

GARD analysis identified 917 potential recombination breakpoints using 26198 models in TYLCV-all dataset. However, only 8 were deemed credible at ΔC-AIC (vs. null model) = 2149.67; ΔC-AIC (vs. the single tree multiple partition) = 927.86. These breakpoints were clustered around nucleotide positions 500, 1100, 1200, 1550, 2150, 2500, and 2800 (Fig. S1). Likewise, in TYLCV-Ahsa dataset, a total of 646 potential breakpoints using 20377 models were revealed and 9 were considered credible at ΔC-AIC = 1681.70: ΔC-AIC (vs. the single tree multiple partition) = 200065. These breakpoints were primarily found around nucleotide positions 500, 750, 1200, 1700, 2150, 2500, 2590, and 2750 (Fig. S1).

In contrast, the TYLCV-Jeddah, TYLCV-Hadasham, and TYLCV-Misc datasets exhibited significantly fewer credible breakpoints. The TYLCV-Jeddah population had only 2 credible breakpoints at nucleotide positions 100 and 2750. The TYLCV-Hadasham population had 2 credible breakpoints at nucleotide positions 18 and 70 (Fig. S1). The TYLCV-Misc population had 4 credible breakpoints at nucleotide positions 50, 2100, 2500, and 2700.

These findings suggest that recombination may play a role in the genetic diversity of TYLCV, particularly in the TYLCV-all and TYLCV-Ahsa datasets. However, the lower number of credible breakpoints in the other datasets indicates that recombination may not be as significant a factor in these populations.

3.8. Conserved Sequence Motif Analysis

Motif analysis, conducted on both DNA and protein sequences, revealed extensive conservation of regulatory motifs across TYLCV-encoded ORFs, suggesting their crucial roles in viral function (Table S2). Notably, ORFs such as V2 and C4 exhibited high conservation of motif number and position across all isolates, highlighting their essential roles. However, some variations were observed in other ORFs, including CP, Rep, TrAP, and REn, in specific isolates. For instance, isolates MN397780 (Ahsa) and OR724727 (Riyadh) lacked certain motifs in their protein and DNA sequences. The former did not contain motif number 3 in protein and motif number 4 in DNA, while the latter did not have motif number 3 and 4 in its proteins. Additionally, isolate MN397780 lacked motif 5 in the Rep region, while isolates KF435136 (Ahsa) and KT728743 (Jeddah) exhibited variations in the arrangement of motifs within the Rep and REn regions, respectively. Isolate KF435136 had motif number 5 in between motif numbers 1 and 2, while in isolate KT728743, motif number 4 was substituted with motif number 5. Likewise, in REn of two isolates, OR724727 (from Riyadh) and KF561125 (from Qassim), motif number 3 in protein was substituted with motif number 5. These variations in motif distribution and composition may contribute to the genetic diversity and functional adaptations of TYLCV isolates.

4. Discussion

The present study provides a comprehensive exploration of the genetic diversity, spatial distribution, and evolutionary dynamics of TYLCV in Saudi Arabia, revealing a complex interplay of factors that shape its genetic landscape. These findings contribute valuable insights into how geographic distribution, genetic diversity, recombination, and selective pressures influence viral evolution and adaptation.

The analysis of 28 full-length TYLCV sequences revealed distinct regional patterns of genetic diversity by grouping them into two major clades, suggesting regional clustering rather than host-specific adaptation. The first clade comprised sequences from the northern and western regions of Saudi Arabia, while the second clade included sequences predominantly from Ahsa region. This highlights Ahsa as a hotspot for TYLCV diversity, with a high number of diverse isolates, including the oldest and most divergent sequences. Such diversity likely results from recurrent introductions, high vector density, and frequent recombination events, emphasizing Ahsa's role as a significant source of TYLCV variation in Saudi Arabia. Additionally, Ahsa, a major agricultural hub in Saudi Arabia, showed notably high genetic diversity of TYLCV, likely driven by a combination of interacting factors. The region's extensive tomato cultivation, coupled with large and persistent whitefly (*Bemisia tabaci*) populations, created optimal conditions for continuous viral transmission and spread. Furthermore, the limited adoption of resistant tomato varieties and potentially inconsistent implementation of vector control measures may have facilitated the virus's persistence and diversification. In contrast, regions such as Jeddah and Hadasham exhibited lower genetic diversity, indicating more stable and less diverse viral populations. Additionally, lower diversity might suggest a recent introduction or a strong bottleneck effect – as Hadasham is an isolated area with limited agriculture, leading to lower crop diversity and restricted tomato cultivation, which affects TYLCV distribution and vector dynamics. Furthermore, the presence of even limited diversity in Hadasham allows us to examine the selective pressures acting on a less complex viral population, potentially revealing the key drivers of TYLCV evolution in the absence of extensive recombination. Previous studies based on genetic analysis of CP have shown the presence of genetic variability within the TYLCV in Saudi Arabia [17]. This study did not directly assess the impact of host plant diversity on TYLCV evolution. However, host variability may influence viral adaptation, selection pressures, and recombination events. Different crop species can act as reservoirs or selective environments, potentially shaping viral genetic diversity. Future studies incorporating host-specific analyses could offer deeper insights into the role of plant diversity in TYLCV evolution and epidemiology.

The spatial distribution of TYLCV isolates reflects an evolutionary trajectory intricately shaped by local agro-ecological conditions [18], and likely influences the virus's adaptation to different hosts and environments. This aligns with global observations where certain regions serve as

focal points for viral emergence and diversification due to conducive environmental, agricultural, and vector-related factors (Z. Iqbal unpublished data). Phylogenetic analysis corroborated these findings as it classified TYLCV sequences into two major clades, suggesting that local environmental factors, farming practices, and limited gene flow significantly influence TYLCV phylogeography. This regional clustering is consistent with prior studies reporting localized evolution in Jeddah, Jizan, and Qasim [8]. Moreover, evidence of isolates grouping with those from Boushehr and Iran, alongside a distinct monophyletic clade, indicates potential multiple introductions or divergent evolutionary trajectories of TYLCV in Saudi Arabia [19].

The TYLCV genome showed moderate to high genetic variation, with the CP and Rep showing the highest diversity. This suggests these regions are under strong selective pressure or prone to recombination. In contrast, the V2 exhibited lower diversity, indicating strong purifying selection maintains its functional integrity. These results are consistent with observations in other begomoviruses, such as ChiLCV, PaLCuV, PeLCV, ToLCPalV, and ToLCNDV, where prior studies reported similar non-random patterns of genetic diversity within viral ORFs [10, 20-22]. The uneven distribution of genetic variation across the TYLCV genome suggests that various factors influence genetic variability at different locations and intensities. This non-random pattern is characteristic of begomoviruses, with regions encoding critical functional proteins experiencing higher evolutionary pressures due to their roles in host-virus interactions and adaptation to new hosts [21, 23, 24]. This observation aligned with previous research demonstrating that regions encoding critical functional proteins often undergo increased evolutionary pressure to facilitate adaptation to new hosts and overcome host defenses [25].

The inferred GDIs demonstrated significant variability within the TYLCV populations, with Ahsa displaying the highest diversity. This is reflected in the region's high haplotype diversity (Hd = 1.00) and nucleotide diversity ($\pi = 0.079$), compared to the more conserved populations in Jeddah (Hd = 0.80) and Hadasham (Hd = 0.60). The high genetic diversity observed in Ahsa is likely driven by recurrent introductions, a dense vector population, or recombination events, commonly linked to viral evolutionary hotspots [26]. Notably, the TYLCV-All dataset revealed distinct peaks in nucleotide diversity in functionally significant regions, such as CP, Rep, and REn. This observation is consistent with prior studies on begomoviruses, which identified Rep and CP as regions with higher nucleotide variability [10, 12].

Selection pressure analyses using dN/dS ratios, SLAC, and FUBAR revealed a complex interplay of evolutionary forces acting on TYLCV ORFs. Most ORFs, including V2 and CP (dN/dS = 0.302 and 0.39), are under strong purifying selection, constraining deleterious mutations. In contrast, C4 (dN/dS = 2.20) and Rep (dN/dS = 1.28) exhibited diversifying selection, likely reflecting adaptations to diverse hosts or mechanisms to evade immune responses [10, 27]. These patterns are consistent with findings in other viruses [28] and begomoviruses such as ChiLCV [21], ToLCPalV [22], and PeLCV [10]. This observation was further supported by FUBAR and SLAC analyses. Both analyses revealed the presence of positively selected sites predominantly in C4, TrAP, and REn. This highlights these regions as potential candidates for further study, as they may represent key determinants of viral adaptability and fitness.

Recombination emerged as a critical mechanism shaping TYLCV genetic diversity in Saudi Arabia. The identification of 32 credible recombination events, primarily within the Ahsa dataset, underscores the significance of this region as a hub for genetic exchange. GARD analysis further identified several recombination breakpoints across different TYLCV populations, with many clustering in the same region as determined by RDP, recognized as recombination "hotspots" [29]. The inferred recombination events involved parental sequences from different viruses, indicating a complex pattern of genetic exchange including both inter- and intraspecies recombination. The rolling circle replication mechanism in begomoviruses predisposes them to frequent, non-random recombination events [30, 31]. These recombination dynamics align with global observations of geminiviruses, where such events drive genetic innovation and the evolution of resistance-breaking phenotypes [8].

Motif analysis revealed conserved regulatory motifs across TYLCV-encoded ORFs, highlighting their critical role in viral function. Begomoviruses harbor conserved motifs within their protein sequences, essential for viral replication, movement, and pathogenesis. These motifs, such as those in the Rep protein, are highly conserved across different geminivirus genera and are crucial for viral fitness [32]. Mutations within these motifs can severely impact viral fitness, emphasizing their role in viral evolution. Computational and comparative genomic analyses have revealed that these motifs are dynamic evolutionary signatures that facilitate viral adaptation to changing host environments [33]. The intricate conservation of specific motif patterns suggests selective pressures maintaining critical functional domains while allowing for subtle genetic variations that enable viral survival [34] and adaptation to different hosts and environments [35].

The genetic variations (mutations and recombination) identified in TYLCV, particularly within key genes like CP, Rep, V2, and C4, hold significant implications for host-virus interactions and disease management. These variations can influence viral fitness, replication efficiency, and the evasion of plant immune responses. For example, mutations in CP can affect vector transmission, while changes in Rep may contribute to recombination, driving viral adaptability and the emergence of resistance-breaking strains [31]. Variations in V2 and C4, known for their roles in host RNA silencing suppression and symptom development, can further impact the virus's ability to overcome plant defenses. Understanding these variations is crucial for developing resistant cultivars through genome editing or marker-assisted selection, and for designing targeted interventions like CRISPR-based genome editing or RNAi to disrupt critical viral functions. Future research should prioritize functional validation of these genetic variations and their interactions with host defense pathways, ultimately guiding the development of innovative and sustainable crop protection strategies.

5. Conclusions

This study provides a comprehensive understanding of TYLCV's genetic diversity, evolution, and the molecular mechanisms driving its dynamics in Saudi Arabia, reveal-

ing the influence of geography, host range, and recombination. Our findings, including conserved motifs, recombination hotspots, and selective pressures, enhance our understanding of TYLCV evolution and inform management strategies. The significant genetic diversity identified in the Ahsa region highlights the need for targeted monitoring and containment strategies to mitigate the spread of potentially more virulent strains. Critically, these insights, particularly the identification of diversifying selection, can guide the development of durable resistance. Moving forward, integrating ecological, molecular, and epidemiological data is crucial for a complete understanding of TYLCV's evolution and its impact on global food security. To effectively manage TYLCV, especially in diverse regions, we advocate for a multi-pronged approach encompassing resistant varieties, integrated pest management for whiteflies, crop rotation, strict quarantine, early detection, controlled greenhouse cultivation, and comprehensive farmer training programs. These integrated strategies are essential for minimizing crop losses and mitigating the impact of TYLCV on tomato production.

Abbreviations

TYLCV - Tomato yellow leaf curl virus; CP - Coat Protein; Rep – Replication-associated protein; ORF - Open Reading Frame; REn – replication enhancer protein; TrAP – Transcriptional Activator protein; dN/dS - Non-Synonymous to synonymous mutation ratio; SLAC - Single Likelihood Ancestor Counting; FUBAR - Fast Unconstrained Bayesian Approximation; RDP - Recombination Detection Program; GARD - Genetic Algorithm for Recombination Detection

Hd - Haplotype Diversity; π - Nucleotide Diversity.

Conflict of Interests

The author has no conflicts with any step of the article preparation.

Consent for publications

The author read and approved the final manuscript for publication.

Ethics approval and consent to participate

No human or animals were used in the present research.

Informed Consent

The authors declare that no patients were used in this study.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' contributions

Zafar Iqbal: Research design, analysis, writeup, Proofreading.

Funding

The authors extend their gratitude to the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Kingdom of Saudi Arabia, for funding the publication of this work (Grant KFU250788).

References

1. Cao X, HuangM, Wang S, Li T, Huang Y (2024) Tomato yellow leaf curl virus: Characteristics, influence, and regulation mechanism. Plant Physiol Biochem 213: 108812. doi: https://doi.org/10.1016/j.plaphy.2024.108812
2. Kil E-J, Kim S, Lee Y-J, Byun H-S, Park J, Seo H, Kim C-S, Shim J-K, Lee J-H, Kim J-K, Lee K-Y, Choi H-S, Lee S (2016) Tomato yellow leaf curl virus (TYLCV-IL): a seed-transmissible geminivirus in tomatoes. Sci Rep 6: 19013. doi: 10.1038/srep19013
3. Mazyad H, Omar F, Al-Taher K, Salha M (1979) Observations on the epidemiology of tomato yellow leaf curl disease on tomato plants. Plant Disease Reporter 63 (8): 695-698. doi:
4. Salama E, Al-Beshr A, Abou-Ghadir M (1994) Identification and characterization of tomato viruses in the Eastern Province of Saudi Arabia. Egy J Phytopathol 22 (1): 125-141. doi: 10.5555/19961001120
5. Paris TM, Johnston N, Strzyzewski I, Griesheimer JL, Reimer B, Malfa K, Allan SA, Martini X (2024) Tomato yellow leaf curl virus manipulates *Bemisia tabaci*, MEAM1 both directly and indirectly through changes in visual and volatile cues. PeerJ 12: e17665. doi: 10.7717/peerj.17665
6. Ajlan A, Ghanem G, Abdulsalam K (2007) Tomato yellow leaf curl virus (TYLCV) in Saudi Arabia: identification, partial characterization and virus-vector relationship. Arab J Biotechnol 10 (1): 179-192. doi: 10.5555/20083073933
7. Al-Saleh MA, Al-Shahwan IM, Brown JK, Idris AM (2014) Molecular characterization of a naturally occurring intraspecific recombinant begomovirus with close relatives widespread in southern Arabia. Virol J 11: 103. doi: 10.1186/1743-422x-11-103
8. Sohrab SS, Yasir M, El-Kafrawy SA, Al-Zahrani HSM, Mousa MAA, Bakhashwain AA (2016) Phylogenetic relationships, recombination analysis and genetic variability of Tomato yellow leaf curl virus infecting tomato in Jeddah, Saudi Arabia. Plant Omics 9 (1): 90-98. doi: 10.3316/informit.888014274173365
9. Duffy S, Shackleton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9: 267-276. doi: 10.1038/nrg2323
10. Iqbal Z, Shafiq M, Sattar MN, Ali I, Khurshid M, Farooq U, Munir M (2023) Genetic diversity, evolutionary dynamics, and ongoing spread of pedilanthus leaf curl virus. Viruses 15 (12): 2358. doi: doi.org/10.3390/v15122358
11. Duffy S, Holmes EC (2009) Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. J Gen Virol: vir.0.009266-009260. doi: 10.1099/vir.0.009266-0
12. Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM, Mizubuti ESG, Duffy S, Zerbini FM (2017) The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evol 3 (1): vex005. doi: 10.1093/ve/vex005
13. Mabvakure B, Martin DP, Kraberger S, Cloete L, van Brunschot S, Geering ADW, Thomas JE, Bananej K, Lett J-M, Lefeuvre P, Varsani A, Harkins GW (2016) Ongoing geographical spread of Tomato yellow leaf curl virus. Virology 498: 257-264. doi: http://dx.doi.org/10.1016/j.virol.2016.08.033
14. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38 (7): 3022-3027. doi: 10.1093/molbev/msab120
15. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B (2021) RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol 7

(1): veaa087. doi: 10.1093/ve/veaa087
16. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME Suite: tools for motif discovery and searching. Nucleic Acids Res 37 (suppl_2): W202-W208. doi: 10.1093/nar/gkp335
17. Nigam D, Muthukrishnan E, Flores-López LF, Nigam M, Wamaitha MJ (2023) Comparative genome analysis of Old World and New World TYLCV reveals a biasness toward highly variable amino acids in coat protein. Plants (Basel) 12 (10). doi: 10.3390/plants12101995
18. Rezk A, Alhudaib K (2017) Genetic diversity of tomato yellow leaf curl-like viruses in Saudi Arabia. J Plant Sci 12 (1): 5-16. doi: 10.3923/jps.2017.5.16
19. Sattar MN, Almaghasla MI, Tahir MN, El-Ganainy SM, Chellappan BV, Arshad M, Drou N (2024) High-throughput sequencing discovered diverse monopartite and bipartite begomoviruses infecting cucumbers in Saudi Arabia. Front Plant Sci 15: 1375405. doi: 10.3389/fpls.2024.1375405
20. Mar TB, Xavier CAD, Lima ATM, Nogueira AM, Silva JCF, Ramos-Sobrinho R, Lau D, Zerbini FM (2017) Genetic variability and population structure of the New World begomovirus Euphorbia yellow mosaic virus. J Gen Virol 98 (6): 1537-1551. doi: https://doi.org/10.1099/jgv.0.000784
21. Mishra M, Verma RK, Pandey V, Srivastava A, Sharma P, Gaur R, Ali A (2022) Role of diversity and recombination in the emergence of chilli leaf curl virus. Pathogens 11 (5): 529. doi: 10.3390/pathogens11050529
22. Sattar MN, Khurshid M, El-Beltagi HS, Iqbal Z (2022) Identification and estimation of sequence variation dynamics of Tomato leaf curl Palampur virus and betasatellite complex infecting a new weed host. Biotechnol Biotechnol Equip 36 (1): 609-619. doi: 10.1080/13102818.2022.2112911
23. Sobrinho RR, Xavier CAD, Pereira HMdB, Lima GSdA, Assunção IP, Mizubuti ESG, Duffy S, Zerbini FM (2014) Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. J Gen Virol 95: 2540-2552. doi: 10.1099/vir.0.067009-0
24. Srivastava A, Pandey V, Sahu AK, Yadav D, Al-Sadi AM, Shahid MS, Gaur RK (2022) Evolutionary dynamics of begomoviruses and its satellites infecting papaya in India. Front Microbiol 13: 879413. doi: 10.3389/fmicb.2022.879413
25. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL (2005) Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Ann Rev Phytopathol 43: 361-394. doi: 10.1146/annurev.phyto.43.040204.135939
26. Lefeuvre P, Moriones E (2015) Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. Curr Opinion Virol 10: 14-19. doi: 10.1016/j.coviro.2014.12.005
27. Seal SE, VandenBosch F, Jeger MJ (2006) Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. Crit Rev Plant Sci 25. doi: 10.1080/07352680500365257
28. Garcia-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Ann Rev Phytopathol 39: 157-186. doi: 10.1146/annurev.phyto.39.1.157
29. Lefeuvre P, Lett J-M, Varsani A, Martin DP (2009) Widely conserved recombination patterns amongst single stranded DNA viruses. J Virol 83: 2697-2707. doi: 10.1128/jvi.02152-08
30. Prasanna HC, Rai M (2007) Detection and frequency of recombination in tomato-infecting begomoviruses of South and Southeast Asia. Virol J 4: 111. doi: 10.1186/1743-422X-4-111
31. Martin DP, Biagini P, Lefeuvre P, Golden M, Roumagnac P, Varsani A (2011) Recombination in Eukaryotic Single Stranded DNA Viruses. Viruses 3 (9): 1699-1738. doi: https://doi.org/10.3390/v3091699
32. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibanez JT, Hanley-Bowdoin L (2010) Functional analysis of a novel motif conserved across geminivirus Rep proteins. J Virol 85 (3): 1182-1192. doi: 10.1128/jvi.02143-10
33. Lefeuvre P, Martin DP, Hoareau M, Naze F, Delatte H, Thierry M, Varsani A, Becker N, Reynaud B, Lett JM (2007) Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. J Gen Virol 88 (12): 3458-3468. doi: 10.1099/vir.0.83252-0
34. Muhire BM, Golden M, Murrell B, Lefeuvre P, Lett J-M, Gray A, Poon AYF, Ngandu NK, Semegni Y, Tanov EP, Monjane AL, Harkins GW, Varsani A, Shepherd DN, Martin DP (2014) Evidence of pervasive biologically functional secondary-structures within the genomes of eukaryotic single-stranded DNA viruses. J Virol 88 (4): 03031-03013. doi: 10.1128/jvi.03031-13
35. Arguello-Astorga G, Ascencio-Ibanez JT, Dallas MB, Orozco BM, Hanley-Bowdoin L (2007) High frequency reversion of geminivirus replication protein mutants during infection. J Virol 81 (20): 11005-11015. doi: 10.1128/jvi.00925-07