Journal of Plant Diseases and Protection
https://doi.org/10.1007/s41348-024-00980-4

ORIGINAL ARTICLE

Cross-species substitution matrix comparison of *Tomato leaf curl New Delhi virus* (ToLCNDV) with medicinal plant isolates

Muhammad Arif[1]

Received: 20 June 2024 / Accepted: 1 August 2024

Abstract
The main objective of this study was to estimate and compare substitution matrixes of nucleotide frequencies for *Tomato leaf curl New Delhi virus* (ToLCNDV) with recently identified begomoviral isolates from two medicinal false daisy (*Eclipta prostrata*) and tomato (*Solanum lycopersicum*) plants. The ToLCNDV has become a significant limitation to vegetables production in many countries. A polymerase chain reaction was conducted to conserve the existence of begomoviral infection. The acquired amplicon was amplified using primers appropriate to the sequence in order to retrieve the full genome. The sequence analysis has confirmed the presence of ToLCNDV in symptomatic plants. The complete genome sequence having a 2.6–2.7 kb entire genome of ToLCNDV was obtained. An investigation of the phylogenetic and evolutionary history has verified the connection between this virus and other closely related viruses. The available nucleotide frequencies of codon regions (*A*, *T*/*U*, *C*, *G*) with newly isolates revealed 20–28% substitution matrixes. There was a minimal difference of nucleotide frequencies' with already submitted database of this virus. Substitution matrixes, which quantify the probability of nucleotide substitutions evolving over a period of time, offer valuable information about mutation patterns and the forces driving evolution. This comparative analysis enhanced the comprehension of the genetic diversity of ToLCNDV and its possible consequences on medicinal plants. It also assisted in the formulation of efficient control measures and the preservation of begomoviruses in medicinal plant biodiversity. The information presented here is highly valuable for understanding the ToLCNDV biology and epidemiology, and it would also assist in disease management in the future.

Keywords *Solanum lycopersicum* · *Eclipta prostrata* · Medicinal plants · Begomovirus · ToLCNDV · Full genome · Phylogenetic analysis · MEGA11

Introduction

Geminiviruses are insect-borne viruses that have emerged as a crucial group of plant viruses with substantial economic impact, leading to major crop and ornamental plant losses worldwide. These viruses infect a wide range of plants, including both monocotyledonous and dicotyledonous plants, primarily in tropical and sub-tropical regions (Arif et al. 2022). Within the family *Geminiviridae*, which comprises over 320 species, the largest genus is begomovirus (Zerbini et al. 2017) and the transmission of this genus is facilitated by an insect vector known as the whitefly (*Bemisia tabaci*) (Gilbertson et al. 2015).

The genome of begomovirus is comprised of one or two tiny single-stranded DNA components, each around 2.8 kb in size, encased between twin isometric particles (King et al. 2011). The distinction between old world (OW) and new world (NW) begomoviruses is evident in their genomic composition (Souza et al. 2022; Fiallo-Olivé and Navas-Castillo 2020). It is now acknowledged that begomoviruses with bipartite genomes are more common, while those with monopartite genomes have just recently been identified. Within the field of plant virology, the majority of monopartite viruses are associated with a specific kind of single-stranded DNA satellites known as betasatellites (Briddon and Mansoor 2008; Briddon et al. 2012; Arif 2024).

The ToLCNDV is categorized as a bipartite, unpaired DNA begomovirus belonging to the genus begomovirus in the family geminiviridae. It was initially detected during

✉ Muhammad Arif
muhammadarif@subu.edu.tr

[1] Department of Plant Protection, Faculty of Agriculture, Sakarya University of Applied Sciences, 54580 Arifiye, Sakarya, Türkiye

Published online: 13 August 2024

1995 in India, where it primarily affected the solanaceous crops. Subsequently, it inflicted significant damage upon cucurbit crops throughout the subcontinent both Pakistan and India (Janssen et al. 2022).

A multitude of biotic and abiotic conditions can cause damage to tomato and false daisy plants. Biotic factors encompass several organisms that might negatively impact plants, such as pests like aphids, whiteflies, and tomato hornworms, as well as diseases caused by bacteria, fungus, and viruses such as *Tomato leaf curl New Delhi virus* (ToLCNDV). These pests and pathogens can cause symptoms such as plant wilting, yellowing, stunted growth, and fruit decay. Abiotic variables encompass environmental stressors such as severe temperatures, drought, excessive rainfall, unfavorable soil conditions, and nutrient imbalances or toxicities. Insufficient sunlight, incorrect irrigation practices, and pollution can all have adverse effects on the health and productivity of both plants. Comprehending and controlling these elements are essential for sustaining robust crops and guaranteeing maximum yield.

Epidemics of tomato-infecting begomoviruses, i. e., *Tomato yellow leaf curl China virus* (TYLCV) and ToLCNDV in the Mediterranean territory, West Asia, and Central America, have caused substantial crop losses and subsequently the inadvertent introduction of productive insecticide-resistant specie of whitefly (*B. tabaci*) which identified as "*B*" biotype. Since the 1980s in sub-Saharan Africa, epidemics of a destructive *cassava mosaic virus* progressed through Uganda with speed of 20 km/year and later forward to adjoining countries including Tanzania and Kenya (Londoño et al. 2024; Shahriari et al. 2023).

The *E. prostrata*, a member of the *Asteraceae* family, is a prevalent weed found around waterways across the whole subcontinent. This plant often exhibits yellow vein symptoms, which are distinctive of several begomoviruses. Previously, the identification of a begomovirus (tentatively termed *E. prostrata yellow vein virus*) in relation to the condition has been founded, with a specific focus on DNA hybridization testing (Haider et al. 2006). Past reports have indicated that ToLCNDV is commonly found on weeds such as *Carica papaya, Hibiscus cannabinus*, and *E. prostrata* (Kumar et al. 2021).

Studying the variations in the spread of ToLCNDV among diverse host species can provide valuable epidemiological insights that can help anticipate the effectiveness of management strategies in regions with several crops. The retention of the virus in the whitefly is crucial not only for effectively controlling the virus's propagation over short distances, but also for understanding its spread over long distances (Janssen et al. 2022; Kil et al. 2020). The present work covers on an analysis of the host range of ToLCNDV, including the new and old isolates of this virus, its occurrence, and the interaction between the virus and its vector in crop plants belonging to the families *Asteraceae* and *Solanaceae*. This work also aims to comprehend the genetic variability and evolutionary connections among this viral isolates by investigating the nucleotide sequences of ToLCNDV and comparing them with those from the novel isolates. The comparative phylogenetic and evolutionary history along with evolutionary divergence was evaluated between the old isolates and new isolates of this begomovirus.

Material and methods

Samples collections

The location of Dera Ghazi Khan is on the southeastern side of Pakistan which serves as border between Punjab and Baluchistan. The Fort Minro is scenic spot and hill station of district Dera Ghazi Khan. It is famous due to mountainous and natural landscapes. It has a warm and humid climate. The maximum area of this district is mountainous and spare. Various kind of crops, flowers, and ornamental plants are widely grown in this area.

Various plants, such as tomato and false daisy, exhibited common viral symptoms, such as yellowing of veins, thickening of veins, mottling, enations, mosaic patterns, and stunting, were observed. The farmers' grievances over viral infections on their crops proved invaluable in the collection and identification of these samples. A broad survey was conducted to find the diversity and re-allocation of begomoviruses in these fertile agricultural zones. The survey was carried out between 2021 and 2022 in five distinct county locations, namely Fort Minro (FM), Rakhigaj (RG), Khar (K), Nari (NA), and Demis Lake (DM). A total of 25 leaf samples, comprising both symptomatic and healthy specimens, were obtained from each plant species. The leaf samples, both symptomatic and healthy, were appropriately preserved in hermetically sealed polythene bags. The gathered samples were transported to the plant virology laboratory for further examination. To prevent the destruction of the plant virus, all of these samples were stored in a refrigerator at a temperature of − 80 °C.

Extraction of DNA, its amplification, sequencing, and blast analysis

The DNA was isolated from leaves showing virus symptoms, specifically 15 from tomato plants and 5 from false daisy plants, as well as from 5 healthy samples, using the cetyl trimethyl ammonium bromide (CTAB) method (Aboul-Maaty and Oraby 2019). The leaves showing symptoms from each sample were crushed into a fine powder by pulverizing them with a pre-chilled pestle and mortar in liquid nitrogen.

The tissue powder was immediately transferred to a 50-ml autoclaved polypropylene centrifuge tube containing 20 ml of CTAB extraction buffer heated to 65 °C. The extraction buffer consisted of CTAB at a concentration of 1.5%, Tris HCL of 100 mM and a pH of 8.0, Nacl of 1.4 M, EDTA of 20 mM and a pH of 8.0, β-mercaptoethanol of 2%, and polyvinylpyrolidone (PVP) of 2%. The complete tissue powder was thoroughly distributed in the extraction buffer by employing centrifugation and inverting the mixing tubes. The homogenate mixture was incubated at a temperature of 65 °C for 1 h in a water bath. This mixture was consistently mounted during the incubation period. Following incubation, a 15 ml quantity of a mixture comprising chloroform and isoamyl alcohol in a ratio of 24:1 was introduced into each tube. The mixture tubes were vigorously stirred till they underwent a metamorphosis into a deep green emulsion. Afterward, the mixture tubes were stirred for a period of 30 min using a rotary shaker. After the agitation, all tubes were then exposed to centrifugal shaking at a speed of 4000 RPM for a duration of 20 min. The liquid portion was moved to a fresh, sanitized tube with a capacity of 50 ml. Each tube was supplemented with 2 µl of RNase solution, which had a concentration of 10 mg/ml. The tubes were then placed in a water bath and incubated at a temperature of 37 °C for approximately 1 h. This stage was implemented at the discretion of the user and has the ability to be completed promptly following the purification processes. The chloroform, isoamyl alcohol, and centrifugation stages were reiterated subsequent to the RNAse treatment (Arif et al. 2018).

Each purified sample's DNA concentration per gram of leaf tissue was measured at a wavelength of 260 nm using a Biorad SmartSpec Plus spectrophotometer. The concentration of each sample was quantified as 50 µg/mL of extracted DNA. The whole nucleic acid was passed through a filter using an extraction buffer and then restored in a cold TE solution (containing 10 mM Tris, 1 mM EDTA, pH 8.0). The amplified product was sequenced in its entirety using degenerate primers. The complete genome sequence was acquired using PCR amplification, utilizing specific primers designed for begomoviral DNA AV494/Dep. The DNA extracts were subjected to PCR using the specified primer pairs to detect the presence of the target DNA.

The amplification method was carried out in a total volume of 25 µl. The PCR data were analyzed using 1% agarose gel electrophoresis, employing tris–acetate-EDTA (TAE) buffer and ethidium bromide dye. The amplified DNA was isolated from the agarose gel using a purification kit (QuikGene, Qiagen Inc., USA). The purified PCR product was ligated into the pGEM-*t*-vector and then inserted into *Escherichia coli* competent cells. The amplicon products from collected specimens were cleaved using sequence-specific forward and reverse primer pairs. Following the verification of the cloned gene from both species by restriction release fragmentation, random clones were commercially sequenced. The acquired sequences were submitted to the BLAST website for analysis. The acquired whole-genome sequence was selected for homology verification. To compare the obtained sequences with previously identified sequences in GenBank, a BLASTn search was performed on the NCBI non-redundant nucleotide database (http://www.ncbi.nlm.nih.gov/BLAST).

Comparative phylogenetic analysis and evolutionary history

In order to comprehend the lineage of begomoviral isolates together with their satellites, the corresponding isolates were retrieved from NCBI. An ancestral tree was constructed for obtained isolates of this virus using associated viral isolates of ToLCNDV. The maximum likelihood method was used to construct this phylogenetic tree. The bootstrap values for this ancestral tree were derived from maximum repeats that were used to represent evolutionary antiquity of the examined taxa. A nucleotide appraisal of these obtained sequences with their corresponding isolate was performed using the MEGA11 software (Tamura et al. 2013). The obtained begomoviral isolates sequences were matched on the basis of their frequencies expressed as percentages. The domains of the corresponding viral isolated were collected from NCBI in order to analyze the disparity in nucleotide frequencies with their respective codons. The sequence demarcation tool (SDT) with SimPlot was used to compare the sequences of fifteen TYLCNDV isolates, including four new isolates and eleven previously described isolates. The ClustalW method was used to construct the sequence alignment.

Genetic recombination analysis

The genetic recombination analysis of newly isolates was carried out in RDP.4 with slight modifications as per samples and sequence requirement. Recombination detection analysis leveraged the aligned sequences both obtained and available in NCBI database. In order to identify recombination events, a step-down correction was employed with a maximum acceptable p-value of 0.05. This was executed using nine different methods (RDP, Chimaera, BootScan, 3Seq, GENECONV, MaxChi, SiScan, LARD, and PhylPro) accessible in RDP 4, with all default parameters (Al-Ali et al. 2023).

Quantifications of the degree of evolutionary divergence between obtained sequences

The analyses for quantifications of degree of evolutionary divergence between obtained sequences of this plant virus were performed with the maximum composite likelihood

model. This analysis encompassed a total of 20 nucleotide sequences. The pairwise deletion option was used to eliminate all instances of uncertain places in each sequence pair. The completed dataset consisted of a total of 3400 bp of this virus. The evolutionary comparisons were made using MEGA11. The displayed data represent the frequency of base substitutions per site between sequences. The standard error estimate(s) displayed above the diagonal and were derived using a bootstrap technique involving standard replicates (Tamura et al. 2021).

Results

Begomoviruses and its related satellites are of significant concern in contemporary agriculture due to their rapid ability to infect new hosts. These begomoviruses infect about 400 different host species, including both cultivated and non-cultivated plants. However, these uncultivated plants, predominantly found in their natural habitat, such as weeds, hedges, and shrubs, primarily serve as reservoirs and carriers for the development and evolution of new plant viruses. Typically, insect vectors are capable of completing their life cycle even without their original host, allowing the virus to transfer from the insect vector's body to a non-cultivated plant.

A PCR assay was conducted to ascertain the existence of begomoviral infection from the collected specimens of both species exhibiting symptoms. The specimens with a greater DNA concentration were chosen for subsequent PCR amplification. The DNA concentration in the samples collected from the Fort Minro county was better. The specimens from Fort Minro and its environs were chosen for additional investigation. After performing PCR, all samples underwent gel electrophoresis with a 1% gel to ascertain the dimensions of the amplicons. From the Fort Minro county in Mari area, symptomatic samples of tomato plants were taken and provided a 550 base pair amplicon size. In addition, false daisy plants from the Khar area yielded a 730 bp amplicon size. Following that, all of these samples were subjected to PCR analysis in order to confirm the precision of the amplicon size. The amplified products were isolated using a purification kit (QuikGene, Qiagen Inc., USA). The eluted products were sent to Beijing Genomics Institute (BGI, Shenzhen, China) for sequencing.

The ToLCNDV isolates with a DNA-A amplicon size of 550 bp were identified from tomato plants in multiple locations as shown in Table 1. The infecting false daisy plants yielded an amplicon size of 730 bp. The sequencing results revealed that the amplicon clone has a size of 550 bp, which matches the previously submitted sequences of ToLCNDV in the NCBI database. A 730 base pair DNA fragment was acquired from each infected sample of false daisy. These fragments were separated and recovered using gel electrophoresis and the gel extraction instructions provided by OMEGA Bio-Tek. The quantification, cloning, and sequencing of all the retrieved products were performed. Obtained sequences showed 99% identity with their Chinese and Indian counterparts. The sequence alignment of the acquired sequence revealed that the nucleotide sequence similarity varied between 69 and 98% among samples from all locations. For amplification of full-length DNA-A of suspected begomovirus, primer pairs were designed from the obtained sequences. Begomoviral DNA-A-F (5′-CTT CATCGTTTCTCAGCATCAT-3′) and DNA-A-R (5′-CAC TTGCACACGATCTCTAAGA-3′) were used to amplify full-length DNA A and DNA B. PCR by TempliPhi™ kit (GE Healthcare) was conducted to amplify full-length DNA A and DNA B with the protocol described in material and method section and by following the manufacturer's protocol. Amplified PCR products were allowed for digestion with four enzymes including BamH1, Xba1, Pst1, and EcoR1. Digested PCR products with Pst1 were comprised of about 2.6 kbp fragment, while 2.7 kbp fragment was obtained from Xba1 digestion. All these products were cloned by cloning (Thermo Scientific FastDigest protocol) and then sequenced from a commercial company.

Sequence alignment and BLASTn analysis revealed that 2.6 kb fragment sequence demonstrated maximum resemblance with already deposited begomoviruses sequences of DNA A component. Full-length DNA A was comprised of 2680 nucleotides and contains six open reading frames. The analysis of sequences of DNA A of Fort Minro isolates showed maximum nucleotide sequence identity of about 98–99% with TYLCV Yunnan isolates infecting tomato (AJ495812 AJ566746).

Table 1 The number of begomoviral isolates identified from medicinal plants

Serial no	Location	Host	Amplicon size	Virus isolate
1	Fort Minro (FM)	Tomato	550*DNA-A	ToLCNDV-FM
2	Nari (NA)	Tomato	550*DNA-A	ToLCNDV-NA
3	Demis Lake (DM)	Tomato	550*DNA-A	ToLCNDV-DM
4	Rakhigaj (RG)	False daisy	730*DNA-A	ToLCNDV-RG

*is indicating the amplicon size of DNA-A of both medicinal plants from selected zones

A tree of the acquired nucleotide sequences of ToLCNDV from tomato and false daisy plants was generated using the neighbor-joining method and a sequence alignment as shown in Fig. 1 and 2. Each sequence is accompanied by its corresponding geographical origin and host. The node number represents the bootstrap value, which is determined by 1000 repeats. The sequence of infected plants collected in this investigation is indicated by a circle. Our isolates have maximum resemblance with the Indian isolates as because of migration of insect vector.

Our results suggest that viruses isolated from tomato and false daisy plants are an isolate of ToLCNDV. Our results indicate that ToLCNDV might be spreading quickly from these locations to another locations and nearby major cropping zones.

A nucleotide comparison was made between our obtained isolates, which are shown in bold, with their corresponding isolates retrieved from the NCBI database. The nucleotide comparison data were acquired from MEGA11. The frequencies are expressed as percentages. The * sign indicates the total number of base pairs for each individual viral isolate. An analysis of nucleotide frequencies in the sequences of *Tomato leaf curl New Delhi virus* (ToLCNDV) allows for the identification of both conserved and changeable genomic areas. This study facilitates comprehension of the genetic variability and evolutionary trends of the virus. These discoveries are

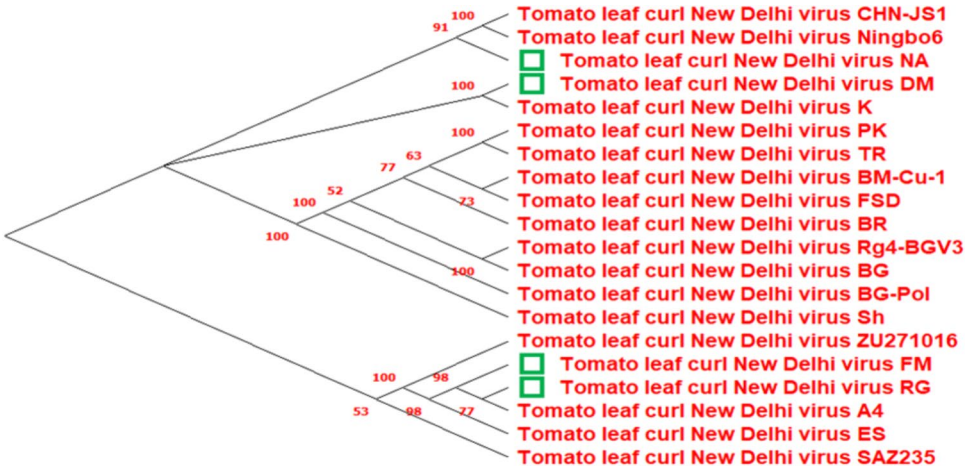

Figure 1 This ancestral tree illustrates the evolutionary lineage of this virus with other isolates. The obtained sequences, including NA, DM, FM, and RG isolates, were compared with all other similar sequences retrieved from the NCBI database

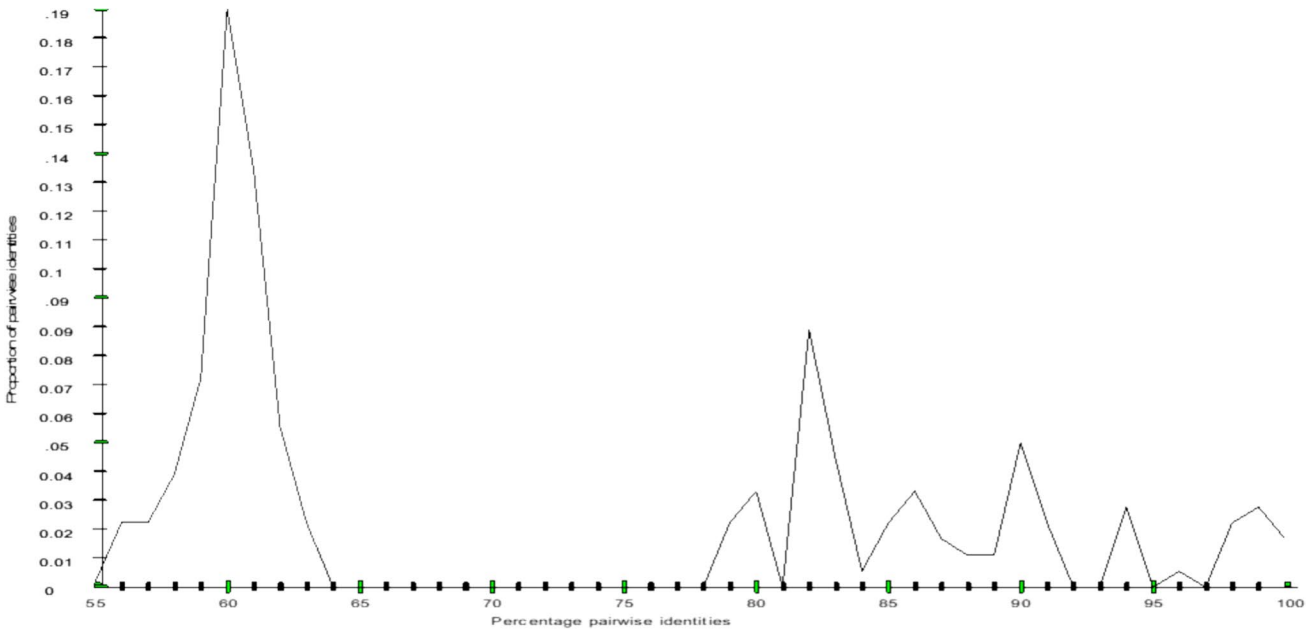

Fig. 2 The proportion of pairwise identities of newly isolates of ToLCNDV with already submitted database in NCBI. This proportion of identities was carried out to map the nucleotide identities differences of this viral isolate

essential for the development of efficient diagnostic tools, precise therapies, and resilient crop types. Average was calculated by sequence length of each isolate (both new and old) with their respective codon regions.

The analysis of nucleotide diversity between the different nations revealed minimal differentiation between our isolated group and groups from other Asian countries. However, it was not possible to compute the nucleotide diversity for Pakistan due to the availability of only one sequence. Remarkably, this distinctive sequence, when viewed from a phylogenetic perspective, does not exhibit significant divergence from the other isolates of ToLCNDV. Almost 20 isolates of this begomovirus with one outgroup were used for analyzing nucleotide divergence and conducting comparisons. The collected viral isolates were comprised of 2692–2740 bp as shown in Table 2. The average nucleotide frequencies after comparisons were 29.7 T(*U*), 21.5 (*C*), 28.5 (*A*), and 20.2 (*G*). There was minimal difference between our isolates and other retrieved isolates from NCBI.

Each entry represents the chance of a substitution (*r*) occurring from one base (row) to another base (column). The determination of substitution pattern and rates was conducted using MEGA11, with some adjustments made to meet the unique needs of the specimens. The bold type denotes the rates of various transitional substitutions, whereas the italic font denotes the rates of transversional substitutions.

The estimation of substitution pattern and rates was conducted. The rates of various impermanent substitutions are indicated in bold, whereas the rates of transversional substitutions are indicated in italics. When evaluating instantaneous r values, it is important to take into account their relative values. To simplify, the sum of the r values is set to 100. The nucleotide frequencies are as follows: $A = 28.54\%$, $T/U = 29.72\%$, $C = 21.53\%$, and $G = 20.21\%$ as shown in Table 3. To estimate ML values, an automatic computation of a tree topology was performed. The highest log likelihood achieved during this computation was −22,428.879. There were a total of 20 nucleotide sequences included in this analysis. The completed dataset consisted of a total of 3400 locations.

The nucleotide identity analysis of several isolates from ToLCNDV was presented, together with the four previously reported isolates from medicinal plants (Figs. 1, 2, and 3). The nucleotide sequence identity of the novel isolates of ToLCNDV varies between 91.20% and 100%. Out of the twenty ToLCNDV isolates, isolates that were reported earlier and four out of the all new isolates include an additional 19 nucleotides (TTCTTTCTAGGTGTGCCCC) in the intergenic region, as shown in Fig. 4. There are four nucleotide variants preceding the additional 19 nucleotides. The three isolates that were previously reported, along with four new isolates, contained the nucleotide sequence "G/CCTT" before the 19 additional nucleotides. In contrast, the remaining isolates had the sequence "AAA(A)." An investigation was conducted using the RDP4 tool to identify potential recombination events. At least four approaches in the RDP4 program discovered fourteen recombination events in the twenty worldwide ToLCNDV full-length sequences. The newly found isolates in this study were determined to be possible recombinants. Furthermore, ToLCNDV effectively separates FM, NA, DM, and RG, which are the primary parents used to generate various recombinants. The data indicate the presence of microevolution in the ToLCNDV populations in Dera Ghazi Khan, as seen in Fig. 4 and 5. Moreover, it has been demonstrated that genetically

Table 2 Comparison of nucleotide frequencies of available sequences of ToLCNDV

Virus/Isolate name	*T*(*U*)	*C*	*A*	*G*	Total(bp*)
ToLCNDV CHN isolate	30.6	22.4	28.6	18.4	2692*
ToLCNDV PK isolate	29.1	21.5	27.5	22.0	2739*
ToLCNDV ZU isolate	30.7	21.4	29.7	18.2	2686*
ToLCNDV BM isolate	28.7	21.7	27.4	22.2	2740*
ToLCNDV PL isolate	28.2	21.3	27.7	22.8	2731*
ToLCNDV SH isolate	28.9	18.3	30.5	22.2	2676*
ToLCNDV TR isolate	28.8	21.5	27.5	22.1	2721*
ToLCNDV NG isolate	30.7	22.3	28.7	18.3	2693*
ToLCNDV FM isolate	30.6	21.6	29.7	18.2	2684*
ToLCNDV NA isolate	30.1	22.4	28.4	19.1	2695*
ToLCNDV DM isolate	30.5	21.8	28.7	18.9	2682*
ToLCNDV RG isolate	30.6	21.6	29.6	18.1	2684*
ToLCNDV RG-4 isolate	28.8	21.3	27.1	22.8	2737*
ToLCNDV BG isolate	28.6	21.1	27.2	23.0	2734*
ToLCNDV SA isolate	30.5	22.4	29.0	18.1	2694*
ToLCNDV ES isolate	30.5	21.5	29.7	18.2	2684*
ToLCNDV A4 isolate	30.6	21.6	29.7	18.1	2685*
ToLCNDV BR isolate	29.1	21.3	27.4	22.2	2738*
ToLCNDV FSD isolate	29.0	21.5	27.6	21.9	2741*
Average	29.7	21.5	28.5	20.2	2706*

A nucleotide comparison was performed between our obtained isolates, which are shown in bold, with their corresponding isolates retrieved from the NCBI database

Table 3 Maximum likelihood estimate of substitution matrix

	A	*T/U*	*C*	*G*
A	–	*7.57*	*5.48*	**8.83**
T/U	*7.27*	–	**11.67**	*5.15*
C	*7.27*	**16.12**	–	*5.15*
G	**12.46**	*7.57*	*5.48*	–

The bold type denotes the rates of various transitional substitutions, whereas the italic font denotes the rates of transversional substitutions

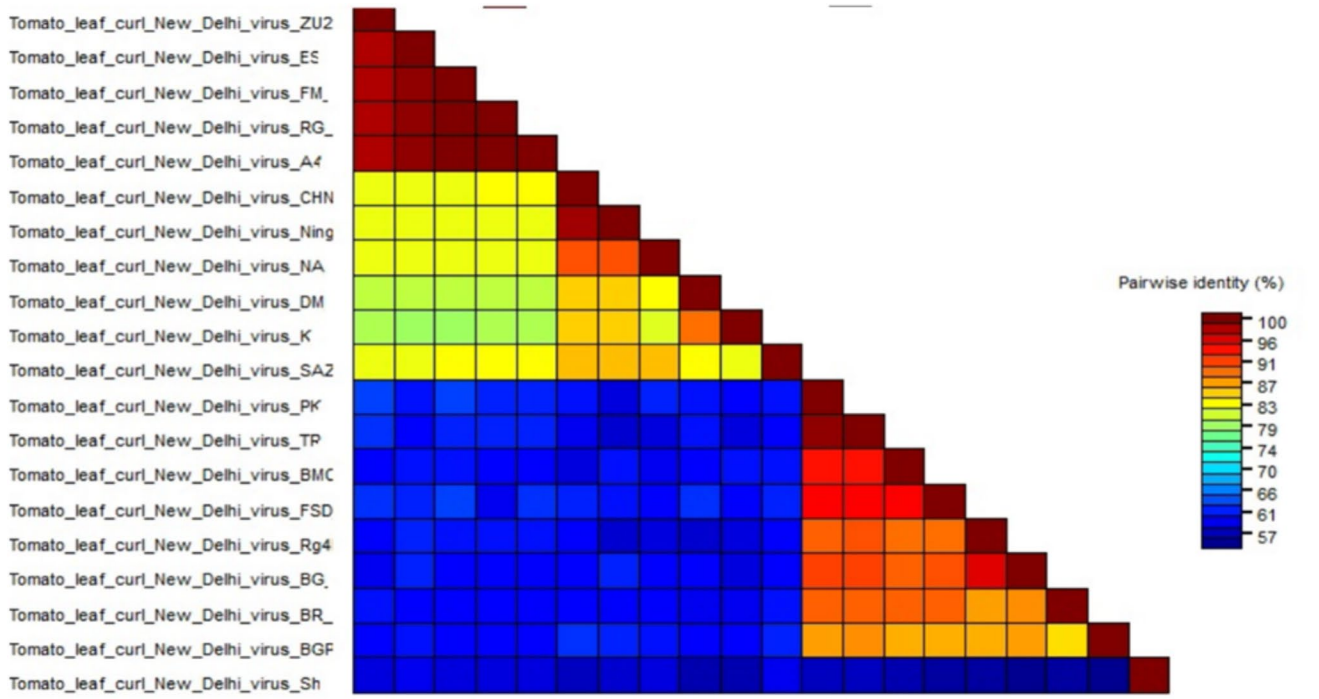

Figure 3 Comparative sequence analysis was conducted using the whole-genome sequences of *Tomato yellow leaf curl New Delhi virus* (ToLCNDV) isolates via sequence demarcation tool (SDT) obtained from two distinct medicinal plants in Dera Ghazi Khan. A comparative sequencing analysis was performed utilizing the whole-genome sequences of *Tomato yellow leaf curl New Delhi virus* (ToLCNDV) isolates from two different medicinal plants in Dera Ghazi Khan. This investigation was carried out using the sequencing demarcation tool. The objective of this investigation was to detect genetic variations and commonalities among the isolates. The data yield vital insights into the virus's evolutionary process and its ability to adapt to various host plants

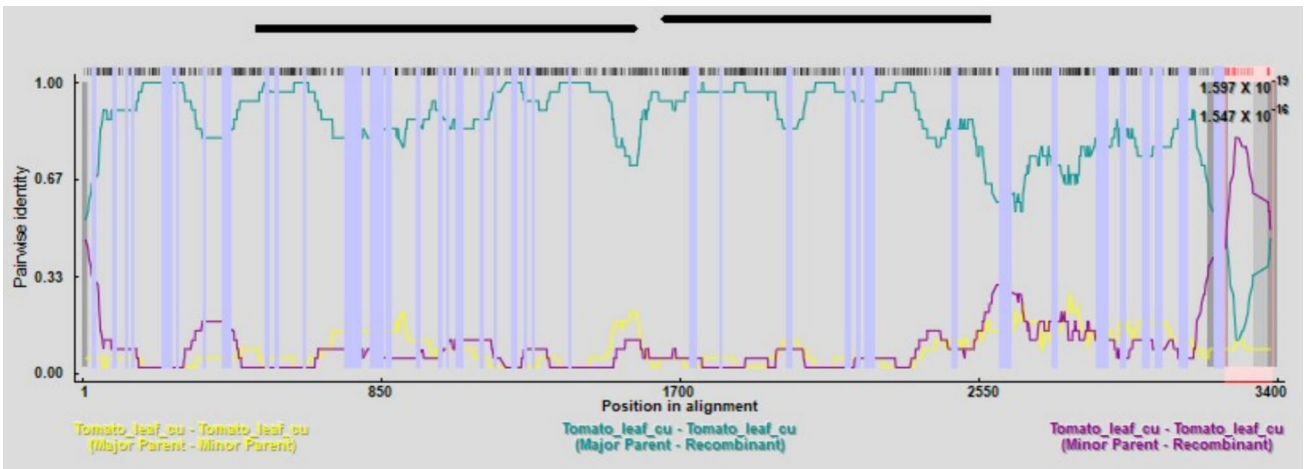

Figure 4 Recombinants have been identified using the recombination detection program (RDP v 4) among different isolates of *Tomato yellow leaf curl New Delhi virus* (ToLCNDV). The event numbers are displayed. The major and minor parent recombinants are displayed in color form. The donated dataset includes the major and minor parents' details and their corresponding p-values

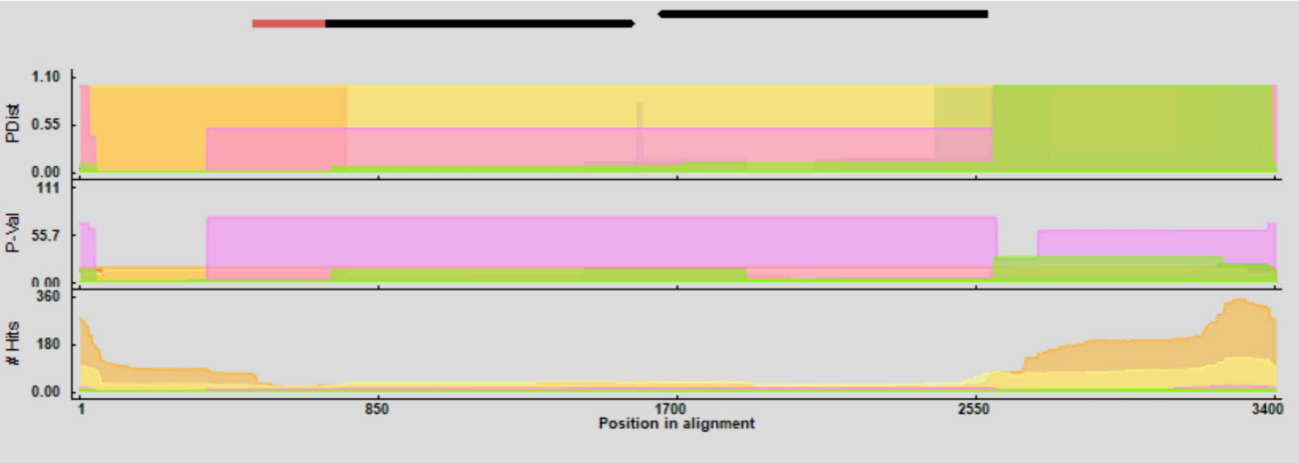

Figure 5 Pairwise alignments of recombinants of *Tomato yellow leaf curl New Delhi* virus among its corresponding *p*-values. The events numbers show the complete genome sequences of this virus along with submitted database in NCBI

recombined ToLCNDV isolates possess a distinct ecological advantage compared to their original viral forms.

Discussion

Human activities like agriculture, which commonly bring huge populations of genetically homogeneous hosts into contact with potential pathogens, are a common cause of disease emergence events. Although viruses are accountable for around 50% of the emergence of plant diseases, there is a dearth of thorough data regarding the transmission of viruses across agro-ecological boundaries. Moreover, our comprehension of the multitude of viruses present in the natural environment is significantly restricted (Bernardo et al. 2018; Riaz et al. 2023; Ahmad et al. 2019).

Environmental factors including temperature, relative humidity, and rainfall significantly correlate with the intensity of plant viruses throughout the world. Rainfall and humidity play a positive role in the spread of viral diseases, and vector is the third influential factor for dispersal. A disease predictive mathematical model has the ability to forecast the intensity of plant viruses in upcoming seasons can be helpful for virologists to manage the viral infection in the agricultural system (A. Ahmad et al. 2017a, b; Islam et al. 2018; Ahmad et al. 2017a, b). The ToLCNDV is a begomovirus that has significant economic importance due to its ability to inflict severe damage to tomato (*S. lycopersicum*) and other crops. It is particularly widespread in several Asian countries. The development of ToLCNDV in the old world (OW) is threatening several commercially significant species, and reports from the past 10 years have suggested that the virus has spread to other vegetables, flowers, and fiber crops (Moriones et al. 2017; Pratap et al. 2011). Here in this study, False daisy and tomato plants were severely infected by ToLCNDV in different locations of Dera Ghazi Khan. A comprehensive survey was carried out at five distinct locations within the hilly regions of Fort Minro and its environs between the years 2021 and 2022. Tomato and false daisy plants in these places display symptoms such as yellow mosaic leaves, leaf curling, enations, stunted development, and deformation. Intensive scrutiny of affected fields has shown evidence that begomoviruses are capable of infecting these plants. The presence of a large population of insect vector (*B. tabaci*) on infected plants or neighboring plants was indicative of begomovirus infection. Plants exhibiting characteristic viral symptoms were characterized molecularly and confirmed the presence of one devastating begomoviral infection.

The ToLCNDV is a bipartite begomovirus that comprises two genomic components, DNA-A and DNA-B, which are approximately 2.8 kb in size and have identical dimensions. ToLCNDV is a unique bipartite begomovirus found in the "New World" (NW) that stands out because it possesses an AV2 gene in its DNA-A component. The DNA-A and DNA-B components possess an identical region with a comparable sequence, which facilitates the replication of DNA-B through a replication initiator protein encoded by DNA-A (Hanley-Bowdoin et al. 2013; Ruiz et al. 2017; Murshed et al. 2024). This study additionally reveals the indigenous presence of ToLCNDV on tomato and false daisy plants for the first time in this geographical region; these hosts could potentially have a significant impact on the dissemination of the virus in these areas. A 550 and 730 bp DNA-A and 2.6–2.8 kb DNA-B were obtained from both plants after characterization. Nevertheless, the application of begomovirus universal primers Begomo-F and Begomo-R

led to the successful amplification of DNAs of the anticipated size. Subsequently, the nucleotide sequence of the DNA amplicons was determined. The BLASTn analysis findings indicated a close similarity between the recovered sequence of the isolate and the DNA-A and DNA-B of ToLCNDV.

The virus exhibits a broad spectrum of hosts, encompassing pepper, potato, tomato, and cucurbit plants. The ToLCNDV virus has been documented to cause severe symptoms, specifically yellow mosaic, on several types of cucurbit vegetables in their natural environment. These symptoms are distinct from the less severe symptoms observed in this study related to the virus infection. Additional investigations into the virulence of global strains of the virus appear to be important (Yazdani-Khameneh et al. 2016; Fortes et al. 2016). The phylogenetic analysis validated the results obtained from the database searches and revealed a grouping with other ToLCNDV isolates, supported by a strong bootstrap value. The ICTV has proposed many taxonomic criteria for demarcating species under the genus *begomovirus*. The most crucial criterion is the threshold of 89% nucleotide sequence identity between the DNA-A components of begomoviruses. Hence, our findings undeniably demonstrated the existence of ToLCNDV in the tomato and counterfeit daisy specimens.

Evidence has shown that ToLCNDV most likely arises from genetic recombination between a ToLCNDV-type isolate and other begomoviruses. Recombination events occur frequently within and between species of begomoviruses, resulting from mixed infections within the same host. These recombination events are considered to be more influential than mutations in driving the evolution of begomoviruses. They contribute to increased genetic diversity and enable the viruses to adapt to new agro ecological conditions, including infecting new hosts (Panno et al. 2019; Agnihotri et al. 2018). Thus far, our isolates have exhibited a remarkably low level of genetic diversity within major and minor parents, with a value of 0.015 ± 0.001. Additionally, the genetic diversity between Indian isolates and isolates from other Asian regions is even lower, with a value of 0.009 ± 0.002. The genetic diversity between Indian and other Asian populations, as well as between our isolates, was found to be higher (0.020 ± 0.003 and 0.012 ± 0.003, respectively) compared to the genetic diversity between Indian and Asian isolates.

Conclusion

This work involved estimating and comparing the substitution matrix of nucleotide frequencies of ToLCNDV using MEGA11. The analysis focused on newly obtained begomoviral isolates from two medicinal plants. The findings of this study demonstrate that ToLCNDV exhibits the highest incidence of plant infection in sub-continent regions. This can be attributed to the initial outbreak of ToLCNDV in tomato plants, which then spread to other regions within a year. In order to prevent the transmission of this virus and the introduction of new pathogens or pests to this region and other sub-continent regions, proper crop management, advanced molecular techniques for early pathogen detection and monitoring, and genetic resistance strategies and stricter phytosanitary measures at international borders must be implemented. The understanding of ToLCNDV biology, epidemiology, and disease control strategies will be greatly aided by the information provided in this study.

Author contributions Muhammad Arif helped in conceptualization, designing of experiment, methodology, analysis, resources, supervision, execution, writing initial and final draft.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The author did not receive support from any organization for the submitted work.

Data availability All data are available within the manuscript.

Declarations

Conflict of interest Authors declare that they have no financial or competing interest.

Human or animal rights This work does not involve in the study of humans and animals.

References

Aboul-Maaty NA-F, Oraby HA-S (2019) Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bullet Natl Res Centre 43(1):1–10

Agnihotri AK, Mishra SP, Tripathi RC, Ansar M, Srivastava A, Tripathi IP (2018) First natural co-occurrence of tomato leaf curl New Delhi virus DNA-A and chili leaf curl betasatellite on tomato plants (*Solanum lycopersicum* L.) in India. J Gen Plant Pathol 84:414–417

Ahmad A, Yasin NA, Ibrahim A, Shahzadi I, Gohar M, Bashir Z et al (2017a) Modelling of cotton leaf curl viral infection in Pakistan and its correlation with meteorological factors up to 2015. *Climate and Development*, pp 1–6

Ahmad S, Raqeeb A, Qasim M (2017b) Comparative analysis of pathogenicity, phylogenetic relationship and host fungal rice blast pathogen *Magnaporthe grisea*. J Bacteriol Mycol Open Access 5(5):Article 00152

Ahmad S, Wang S, Wu W, Yang K, Zhang Y, Tumukunde E et al (2019) Functional analysis of peptidyl-prolyl cis-trans isomerase from aspergillus flavus. Int J Mol Sci 20(9):2206

Al-Ali E, Al-Hashash H, Akbar A, Al-Aqeel H, Al-Shayji N, Alotaibi M et al (2023) Genetic recombination among tomato yellow leaf curl virus isolates in commercial tomato crops in Kuwait drives emergence of virus diversity: a comparative genomic analysis. BMC Res Notes 16(1):71

Arif M, Lin W, Lin L, Islam W, Jie Z, He Z et al (2018) Cotton leaf curl Multan virus infecting Hibiscus sabdariffa in China. Can J Plant Path 40(1):128–131

Arif M, Farooq S, Alasmari A, Alshehri MA, Hashem M, Alamri S et al (2022) Molecular study of geminiviruses with its complex biology, host-vector interactions, and increasing diversity. J King Saud Univ-Sci 34(4):102051

Arif M (2024) Cap-snatching mechanism-mediated unveiling of the transcriptional initiation sites of two distinct begomoviruses. J Plant Dis Prot. https://doi.org/10.1007/s41348-024-00941-x

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D et al (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. ISME J 12(1):173–184

Briddon R, Ghabrial S, Lin N, Palukaitis P, Scholthof K, Vetten H (2012) Satellites and other virus-dependent nucleic acids. In: Virus Taxonomy-Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, New York, pp 1209–1219

Briddon R, Mansoor S (2008) Beta ssDNA satellites. Encycl Virol 3:314–321

Fiallo-Olivé E, Navas-Castillo J (2020) Molecular and biological characterization of a New World mono-/bipartite begomovirus/deltasatellite complex infecting *Corchorus siliquosus*. Front Microbiol 11:1755

Fortes IM, Sánchez-Campos S, Fiallo-Olivé E, Díaz-Pendón JA, Navas-Castillo J, Moriones E (2016) A novel strain of tomato leaf curl New Delhi virus has spread to the Mediterranean basin. Viruses 8(11):307

Gilbertson RL, Batuman O, Webster CG, Adkins S (2015) Role of the insect supervectors *Bemisia tabaci* and *Frankliniella occidentalis* in the emergence and global spread of plant viruses. Annual Rev Virol 2:67–93

Haider M, Tahir M, Latif S, Briddon R (2006) First report of Tomato leaf curl New Delhi virus infecting Eclipta prostrata in Pakistan. Plant pathol 55(2):285

Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S (2013) Geminiviruses: masters at redirecting and reprogramming plant processes. Nat Rev Microbiol 11(11):777–788

Islam W, Lin W, Islam SU, Arif M, Li X, Yang Y et al (2018) Genetic diversity of begomoviruses in Pakistan captured through a vector based survey. Microb Pathog 118:91–97

Janssen D, Simón A, Boulares M, Ruiz L (2022) Host species-dependent transmission of tomato leaf curl New Delhi virus-ES by *Bemisia tabaci*. Plants 11(3):390

Kil E-J, Vo TTB, Fadhila C, Ho PT, Lal A, Troiano E et al (2020) Seed transmission of tomato leaf curl New Delhi virus from zucchini squash in Italy. Plants 9(5):563

King AM, Lefkowitz E, Adams MJ, Carstens EB (2011) In: Virus taxonomy: ninth report of the international committee on taxonomy of viruses: Elsevier

Kumar R, Tiwari RK, Jeevalatha A, Siddappa S, Shah MA, Sharma S et al (2021) Potato apical leaf curl disease: current status and perspectives on a disease caused by tomato leaf curl New Delhi virus. J Plant Dis Prot 128:897–911

Londoño A, Barrón-Contreras J, Cruz-Hernández A, Rivera-Bustamante R, Bañuelos-Hernández B (2024) Emergence, diversity, and epidemiological prospects of viruses infecting tomato crops. Viral Diseases of Field and Horticultural Crops. Elsevier, pp 369–377

Moriones E, Praveen S, Chakraborty S (2017) Tomato leaf curl New Delhi virus: an emerging virus complex threatening vegetable and fiber crops. Viruses 9(10):264

Murshed A, Alnoud MA, Ahmad S, Khan SU, Alissa M, Alsuwat MA et al (2024) Genetic Alchemy unveiled: microRNA-mediated gene therapy as the Artisan craft in the battlefront against hepatocellular carcinoma—a comprehensive chronicle of strategies and innovations. Front Genet 15:1356972

Panno S, Caruso A, Troiano E, Luigi M, Manglli A, Vatrano T et al (2019) Emergence of tomato leaf curl New Delhi virus in Italy: estimation of incidence and genetic diversity. Plant Pathol 68(3):601–608

Pratap D, Kashikar AR, Mukherjee SK (2011) Molecular characterization and infectivity of a Tomato leaf curl New Delhi virus variant associated with newly emerging yellow mosaic disease of eggplant in India. Virology Journal 8:1–13

Riaz MW, Arif M, Mehmood Y, Guldur ME, Zhou A, Alnemar SK et al (2023) Molecular characterization and evolutionary history analysis of a novel Kudzu mosaic virus (KMV) infecting the Kudzu (*Pueraria montana*). J King Saud Univ-Sci 35(11):102990

Ruiz L, Simon A, Velasco L, Janssen D (2017) Biological characterization of Tomato leaf curl New Delhi virus from Spain. Plant Pathol 66(3):376–382

Shahriari Z, Su X, Zheng K, Zhang Z (2023) Advances and prospects of virus-resistant breeding in tomatoes. Int J Mol Sci 24(20):15448

Souza JO, Melgarejo TA, Vu S, Nakasu EY, Chen L-F, Rojas MR et al (2022) How to be a successful monopartite Begomovirus in a bipartite-dominated world: emergence and spread of Tomato mottle leaf curl virus in Brazil. J Virol 96(18):e00725-e722

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol biol evol 30(12):2725–2729

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38(7):3022–3027

Yazdani-Khameneh S, Aboutorabi S, Shoori M, Aghazadeh A, Jahanshahi P, Golnaraghi A et al (2016) Natural occurrence of tomato leaf curl New Delhi virus in Iranian cucurbit crops. The Plant Pathol J 32(3):201

Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, Navas-Castillo J et al (2017) ICTV virus taxonomy profile: geminiviridae. J Gen Virol 98(2):131–133