Iraqi Journal of Biotechnology, April 2020, Vol. 19, No. 1, 40-55

The Partial DNA Sequencing and Phylogenic Analysis of Tomato yellow leaf curl virus Isolated from Iraqi tomato

Basheer Nsaif Jasim[1] , **Ali A. Al-Salihy**[2] , **Ali Mohammad Moner**[2]

[1]Directorate of Diyala Agriculture, Ministry of Agriculture, Iraq
[2]Institute of Genetic Engineering and Biotechnology for Postgraduate Studies, University of Baghdad, Iraq.

Received: March 10, 2020 / **Accepted:** April 15, 2020 / **Published:** April 28, 2020

Abstract: Tomato yellow leaf curl virus (TYLCV) is one on the top ten supreme pathogens threatening tomato (*Solanum lycopersicum L.*) production worldwide. While TYLCV is a serious disease in tomato production areas of Iraq under field and plastic tunnels conditions, little is known about genomic features of these viruses from Iraq. Total DNA was extracted from infected tomatoes were collected from two growing areas of Iraq. A partial sequence of the TYLCV was carried out, and phylogenetic study was conducted to understand the genetic diversity of the TYLCV in these regions. Based on the DNA sequence and phylogenetic analyses, the two collected virus strains were confirmed as TYLCV strains. The two viral isolates that have been characterized in this study were shared 98.91% and 98.38 % nucleotide sequence identity with a previously characterized Iraqi and Lebanese TYLCV isolates, (JQ35499.1and EF051116.1) respectively. The result showed that Iraqi isolates shared highest sequence identity (<98 %) with group of TYLCV complete genome sequencing from Iran, Jordan, Lebanon and Israel include viral strains JQ231214.1, EF054893.1, AB116631.1, AY594174.1, AB110217.1, AB116629.1, AB116630.1 and AB116631.1.

Key words: TYLCV, *Begomoviruses*, *Solanum lycopersicum*.

Corresponding author: (Gmail: basheer.nsaif.alsaedy@gmail.com).

Introduction

Tomato (*Solanum lycopersicum L.*) is one of the most important grown vegetables worldwide due to its valuable content nutrient and it is on top list of every culinary dishes worldwide (1). In 2017 about 5 million hacters were harvested worldwide and the production estimate was about 182 million Mt in 2017 (2). In Iraq tomato fruits are very important for daily consumption due to high content of vitamins, minerals, essential amino acids, sugars and fibers. Tomato production areas are facing a strong threat by different disease pathogens. Huge investment with pest management valued up to 40%, are required for tomato production (3). *Begomoviruses* is on the top of the destructive tomato pathogens damaging tomato production areas around the world (4,5). Begomovirus genomes are classified into two types; monopartite and bipartite. The monopartite viruses consist of DNA-A only and they called Old World begomoviruses (6,7,8). Whereas, bipartite viruses have two segments of ssDNA genome called DNA-A (2.6–2.8 kb) and DAN-B (2.5–2.8 kb) and is encapsidated in a geminate particle of ~25–30 nm (Figure 1), which they referred as New World begomoviruses(9,10).

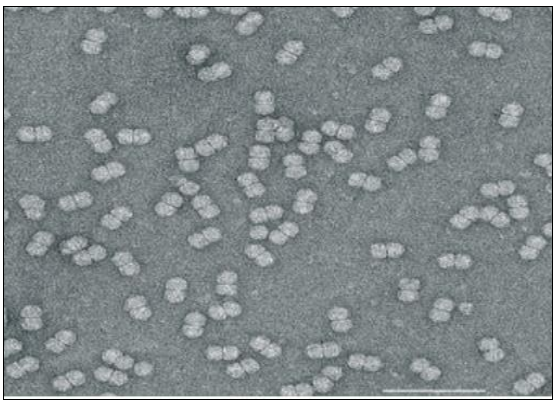


Figure (1): Viral particles (Virions) with twinned morphology of approximately 20x30nm in size (Xie *et al.,* 2013)(11).

Different strains of 60 distinct species of begomoviruses can be hosted tomato plant(12). Begomoviruses comprise viral species causing one of the most limiting pathogen for tomato production namely the tomato yellow leaf curl disease (TYLCD)(13). There are four causative agents for TYLCD in the Mediterranean region, which are currently cause severe losses on tomato yields, these are Tomato yellow leaf curl Sardinia virus (TYLCSV), Tomato yellow leaf curl virus (TYLCV), Tomato yellow leaf curl Malaga virus, and Tomato yellow leaf curl Axarquia virus (3,5,14).

All TYLCV isolates known today have a monopartite genome(15,16). TYLCV can be transmitted by whitefly vector *Bemisia tabaci*, (Homopetra: Aleyrodidae) which causes direct feeding damage and indirect damage as a TYLCV vector. Lost can be even higher on tomato in developing countries due to lack of knowledge on plant virus disease control(17). *Bemisia tabaci* can spread the virus between many hosts and this way may have an important role in the survival of virus between seasons as source of tomato infection (18).

A wide range of weeds found in the tomato fields or around are serving as alternative host for begomovirus, including both monocots and dicots. For instances, plants belonged to families; *Solanaceae, Acanthaceae, Asteraceae, Malvaceae, Brassicaceae, Fabaceae, Oxalidaceae, Pedaliaceae, Urticaceae* and which were frequently identified with TYLCV existence (19).

The tropical climate in many regions of Iraq allows year-round tomato cropping. Also, the existence of perennial host plants for both TYLCV and *Bemisia tabaci* allows the easy carry-over of TYLCD between growing seasons (20).

Materials and Methods

Sample Collection and Extraction of Total DNA from Plant Tissues

Fifty samples of tomato plants showing yellow leaf curl symptoms were collected from tomato fields as following. From Baghdad 20 samples, Karbala 15 samples and Diyala 15 samples during October 2018. Control samples were also collected from the same locations from healthy leaf samples of the same tomato variety. All the leaf samples were collected when the tomato plants were at the early

flowering. Genomic DNA was isolated from samples according to the protocol of Wizard Genomic DNA Purification Kit, Promega USA as manufacturer's instructions. DNA purity was checked using Nanodrop spectrophotometer (Guangzhou, China) which was 1.97. Five microliters aliquots of PCR products were analyzed on 1% agarose gel (Biobasic, Canada) electrophoresis with loading dye at 100 volts for 75 minutes in 1X TAE-electrophoresis buffer and stained with ethidium bromide (10mg ml^{-1}). For image acquisitions, gels were visualized under UV light and documented by using Gel Imaging System (Major Science, Taiwan).

Detection of The Presence of TYLCV in Tomato Leaf Samples

The extracted DNA was used as a template for PCR using set of primers as shown in table (1). by Li *et al.*, (21) were used for specifying the virus as TYLCV. Finally, TYLCV-partial sequencing primers were designed using primer 3 software for TYLCV sequencing this study. Firstly, extracted DNA samples were amplified with degenerate primers according to Deng *et al.*, (22) for Begomovirus detection. Also, TYLCV specific primers designed.

Table (1): The primers used in this study for viral detection and partial sequencing.

Name of primer	Expected amplicon size	Sequence of primer 5' ⟶ 3'
Degenerate primers	520bp	TAATATTACCKGWKGVCCSC GGACYTTRCAWGGBCCTTCACA
Detecting primers	543bp	ACGCATGCCTCTAATCCAGTGTA CCAATAAGGCGTAAGCGTGTAGAC
Sequencing primers	976bp	TACTCACAGAGTGGGTAAGAG ATCAAGGTCCAACACAAGATAG

K: G or T, R: A or G, S: C or G, W: A or T, Y: C or T, B: C, G or T, V: A, C or G

PCR reaction was performed in the thermal-cycler (BioRad, USA) applying 25µL of PCR reaction mixture. This mixture contained a 2µl DNA template (100 ng/µl), 12.5 µl of Go Taq® green master mix 2X (Promega, USA) and 1 µl of each primer (10 pmol/µL) for each specific gene, up to the final volume 25 µl with nucleases free water. PCR thermocycler quantities were achieved by conventional PCR thermocycler program conditions for primer sets as listed in table (2).

Table (2): The PCR program for viral detection and partial sequencing sequencing

(A) Begomovirus detection using Deng primers (22, 23)			
Steps	**Temperature**	**Time**	**Cycle**
Initial Denaturation	94°C	5min	1
Denaturation	94°C	30 sec	30
Annealing	58°C	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	7 min	1
Hold	4°C	∞	

(B) Detecting TYLCV using Lee primers (Lee *et al*., 2012).			
Steps	**Temperature**	**Time**	**Cycle**
Initial Denaturation	94°C	3min	1
Denaturation	94°C	30 sec	30
Annealing	56°C	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	7 min	1
Hold	4°C	∞	
(C) Conditions for partial sequence using current study designed primers.			
Steps	**Temperature**	**Time**	**Cycle**
Initial Denaturation	95°C	4min	1
Denaturation	95°C	30 sec	30
Annealing	60°C	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	1 min	1
Hold	4°C	∞	

Samples giving rise to 543bp product were then subjected to PCR with designed primers "TYLCV-Partial-Seq" which was used to amplify 940 bp. PCR reactions were optimized for 25 µl and the final concentrations of reaction components were: 12.5x master mix, 1 µl of 10 µM of each complementary and viral-sense primers and 2 µl of DNA were used as target templates. PCR cycle parameters for Deng primers were as follows: one cycle (4 min at 94°C), 30 cycles (30 sec. at 94°C, 30 sec. at 58°C and 30 sec.at 72°C), followed by one cycle at 72°C for 10 min in thermocycler (BioRad, USA). PCR cycle parameters for TYLCV-Partial-Seq primers were as follows: one cycle (4 min at 94°C), 30 cycles (30 sec. at 94°C, 30 sec. at 58°C and 30 sec.at 72°C), followed by one cycle at 72°C for 10 min in thermocycler (BioRad, USA).

Sequencing

PCR products resulting from TYLCV-Partial-Seq detecting primers were sent for Sanger sequencing using ABI3730XL, automated DNA sequencer, by Macrogen Corporation–Korea. The results were analyzed using Geneious software. Sequences were compared to other related sequences available in the GenBank database using BLASTn.

Multiple sequence alignment

In order to determine the identity of the viral nucleotide sequences, a comparison was made using the BLASTn algorithm against GenBank's "refseq-genomic" database (24). According to this analysis, thirty-seven TYLCV isolate sequences were selected that exhibited higher identity with the two isolated virus sequences in Baghdad and Karbala provinces (Table 5). Nucleotide sequences were blasted against sequences similarity using the BLASTn algorithm (http://www.ncbi.nlm.nih.gov). Multiple sequence alignment was performed for the two isolates and 37 closely related full sequences of TYLCV isolates by using (Multiple Alignment using Fast Fourier Transform) MAFFT v7.450 software. Phylogenetic tree for the two partially sequenced TYLCV isolates were constructed using the maximum-likelihood principle softwares (PHYML and MrBayes) and bootstrap values of 1000 replicates (25,26).

Protein Structure and Function Prediction

Sequences were translated to amino acids using Geneious software with standard setting then a web servers named RaptorX Property and ExPASy were used for predicting structure property of a protein sequence(27). RaptorX Property based on protein sequence or sequence profile that derived from several sequences alignment of sequence homologs in a protein family. Additionally, this server uses advanced technology for machine learning called DeepCNF (Deep Convolutional Neural Fields) by which researchers can model the property label correlation among adjacent residues. ExPASy was used for protein analysis to predict certain properties about a protein(27).

Results and Discussion

Begomovirus Detection in The Symptomatic Leaf Sample

TYLCV is predominant begomovirus in cultivated tomato growing areas in Iraq. TYLCV diversity and genetic evolution are very important issues in recent agricultural studies. Due to the frequently emergence of TYLCV in many tomato growing regions of Iraq, tomato samples showing typical TYLCV symptoms have been collected from different regions to understand TYLCV diversity and genetic evolution.

DNA-based detection of begomovirus with degenerate primers Deng A/B indicated that 30% of tomato samples were infected by begomovirus (Figure 2). The expected PCR product size (~530 bp) was amplified in 19 of the 50 samples. Disease incidence was recorded in 33.33%, 50% and 26.67% of Baghdad, Karbala and Diyala regions respectively (Table 3).

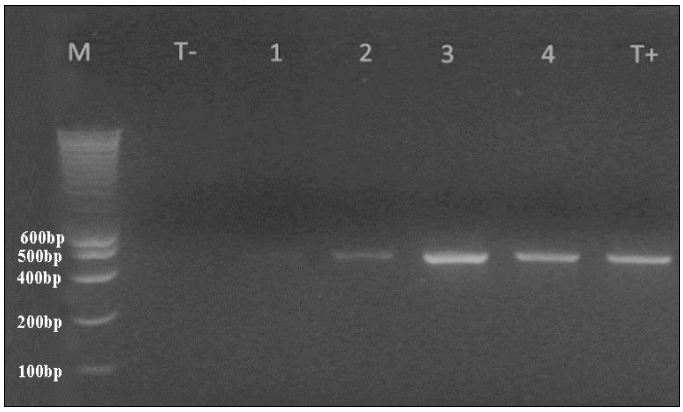


Figure (2): Detection of Begomoviruses by PCR using Deng primers and 100 bp Ladder on 1% agarose gel at 75 volts for 90 minutes. (M = ladder, T- Negative control, 1,2,3,4, tomato samples, T+ positive control.

Table (3): Percentage of samples infected by begomoviruses in three regions of Iraq.

Regions	Number of Samples	Number of Samples positive to PCR	Percent disease incidence (%)
Baghdad	15	5	33.33%
Karbala	20	10	50%
Diyala	15	4	26.67%
Total	50	15	

In order to determine whether these positive samples are TYLCV isolates or not, a TYLCV specific primers (TYLCV/Li) were used for this purpose. Results found that Baghdad and Karbala samples were positive for TYLCV test while no bands shown with Diyala samples figure (3). Li primers can precisely identify TYLCV while other begomovirus would not be recognized and gave negative results. Expected PCR product size was 543bp.

Following, Baghdad and Karbala samples were tested with designed primers (TYLCV-Partial-primers) and the PCR products were exited and sent for sequencing (Figure 4).

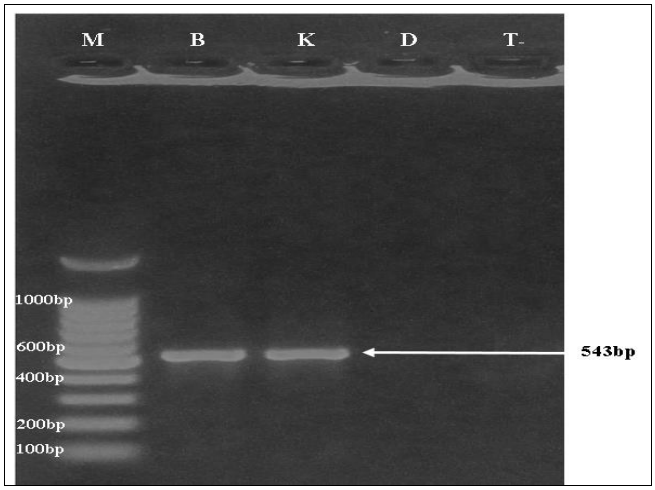


Figure (3): Detection of TYLCV by PCR using Li primers and 100 bp Ladder on 1% agarose gel at 75 volts for 90 minutes. (M = ladder, B- Virus isolated from Baghdad, K- Virus isolated from Karbala, D- Virus isolated from Diyala, T- Negative control.

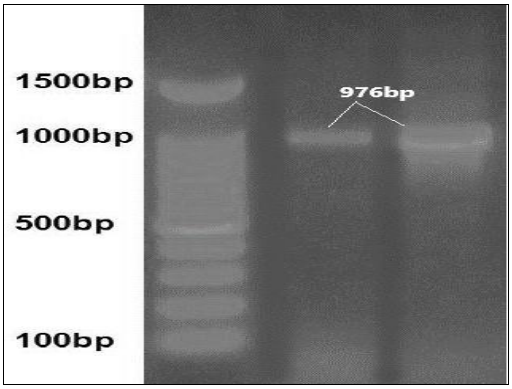


Figure (4): Detection of TYLCV by PCR using four sets of designed primers and 100 bp Ladder on 1% agarose gel at 75 volts for 90 minutes. (M = ladder, B- Virus isolated from Baghdad, K- Virus isolated from Karbala.

Sequencing

The expected PCR fragment size (976 bp) was amplified in samples from Baghdad and Karbala provinces, whereas there were no products obtained from Diyala province. These two fragments have been sequenced and subjected to multiple alignment with thirty-seven other previously published full sequences of TYLCVs. The results showed that Baghdad isolate has

similarity identity with these isolates ranged from 98.38% to 95.90%. These isolates includes for instance, EF051116.1 (Tomato yellow leaf curl virus isolate Ra3), LN846609.1 (Tomato yellow leaf curl virus - Il recombinant IS76, isolate 49-23), GQ861427.1 (Tomato yellow leaf curl virus isolate TYLCV-Mld-JO: Ju:08), HF548825.1 (Tomato yellow leaf curl virus strain Mild-Sweden-Imported), KY971326.1 (Tomato yellow leaf curl virus isolate Florida_ 19.1), X347162.1 (Tomato yellow leaf curl virus isolate TYLCV_IR_ KIRT12_2009), KX347166.1 (Tomato yellow leaf curl virus isolate TYLCV_ IR_KIRT20_2012) and AJ132711.1 (Tomato yellow leaf curl virus-Iran complete genome, isolate Iran). Whereas, the results showed that Karbala isolate has similarity identity with these isolates ranged from 98.91% to 95.87%. These isolates include for instance, JQ354991.1 (Previously identified Iraqi Tomato yellow leaf curl virus), AJ519441.1 (Tomato yellow leaf curl virus - Mild -Spain), JQ231214.1 (Tomato yellow leaf curl virus isolate Kermanshah), AJ489258.1 (Tomato yellow leaf curl virus-Almeria, isolate Almeria), KX347155.1 (Tomato yellow leaf curl virus isolate TYLCV_IR_KIRT04_2006), KX347162.1 (Tomato yellow leaf curl virus isolate TYLCV_IR_KIRT12_ 2009), KX347166.1 (Tomato yellow leaf curl virus isolate TYLCV_IR_ KIRT20_2012) and AJ132711.1 (Tomato yellow leaf curl virus-Iran complete genome, isolate Iran). Also, the results have indicated that the fragments sequenced include regions of V1, V2 and C3 genes of TYLCV genome.

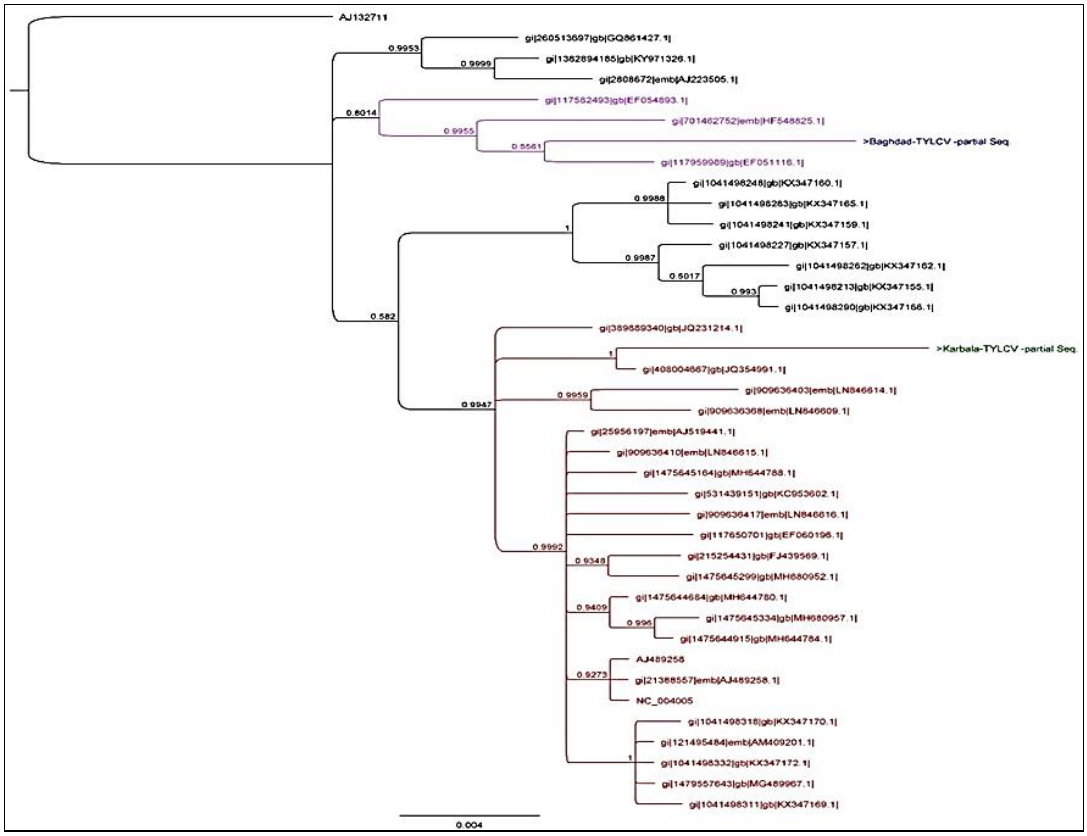
Phylogenetic Analysis

Applying MrBayes and PhyLM programs had generated phylogenetic tree showed identical positions topology respecting the two isolates identified in this study, so only one tree created with MrBayes software has been put in the results section (Figure 5). TYLCV-Baghdad-IRAQ and TYLCV-Karbala-IRAQ isolates are grouped in two separated clades. TYLCV-Baghdad-IRAQ isolate was very close to other isolate denoted EF051116.1 characterized in Lebanon. Also, data showed Baghdad TYLCV isolate was close to other TYLCV isolate from Jordan placed in a clade in another group. On the other hand, TYLCV-Karbala-IRAQ isolate was neighbor to previously fully sequenced and identified TYLCV in Iraq (JQ354991.1) and it shares 98.91% identity similarity. Also, they have been placed near Iranian TYLCV isolate (JQ231214.1) which has 98.15% identity similarity.

Likewise, many TYLCV isolates from Iraq neighbors like Iran and Jordan and far Arabic and non-Arabic countries such as Lebanon, Egypt, Israel, Spain and Mauritius were found to have identity similarities ranged from 98% to 97% with our TYLCV isolates. This result agrees with a study conducted by Al-Waeli *et al.*,(28), where Iraqi isolates showed high nucleotide sequences identity with JQ354991.1 TYLCV isolate. The results of this study serve as a vital information for understanding the variation between two isolates in respecting to TYLCV source and ancestors. Moreover, it is crucial for applying integrated management strategies for choosing better resistant tomato varieties that can be grown in Iraq.

Genetic differentiation between TYLCV Iraqi isolates may be explained to a low gene flow was between southern and centric Iraq TYLCV subpopulations. Full length viral genome sequencing of TYLCV isolates from different Iraq regions is essential to distinguish virus strain(s) in Iraq. Data generated through this research suggested that TYLCV-Baghdad-IRAQ and TYLCV-Karbala-IRAQ isolates being genetically similar with TYLCVs found neighbor countries like Iran, Jorden or areas that have high rates of vegetable trafficking with Iraq through the last two decades by which high gene flow and variation might occur(29).

Also, the unstable general situation in Iraq after USA occupation in 2003, had made Iraqi boarders freely opened for agricultural goods which had potential effects on Iraqi agriculture. Likewise, high rate of insect infestation such as whiteflies can have a notorious role in spreading and evolving TYLCV.



Our phylogenetic analysis showed that there was no relationship between the two isolates and isolates from distal geographical origin, in contrast with the results of Shirazi *et al.*, (30). It is noteworthy to say that Baghdad isolate is more related with isolates from Jordan, Iran and Lebanon in comparison with Karbala isolate which is more related with previous Iraqi isolate and other European isolates (Figure 5).

Similarity observed with Baghdad and Lebanon, Jordan and Iran TYLCV isolates might be elucidated to the high gene flow among these isolates (29) in contrast with other TYLCV isolates from other countries. High rates of agricultural crops exportations between

highly reported of TYLCV incidence areas and Iraq could have very potential effects on tomato production in Iraq.

SNP Identifying

The most important type of genetic changes influencing TYLCV is a single-nucleotide polymorphism (SNP). SNP is substitution of a single nucleotide that occurs at a specific position in the genome that happens at level higher than 1% in the population. They are commonly occurring in introns rather than exons in the DNA which can be used as DNA-based markers helping researchers in molecular studies. While these SNPs occur within exons or in a regulatory region before a gene, they may have very direct role in altering gene function. SNPs were called using Geneiouse prime 2019 v.04 software with minimum 2 coverage per site. SNPs analysis indicates that the two TYLCV isolates have several SNPs placed in their genomes. These SNPs are located in regions started before V1 gene and within V2 and C3 genes of TYLCV genome. SNPs that have been colored red in table (4) may explained the variation between two isolates and illustrating high disease incidents in tomato farms in some tomato growing regions. For example, SNP in alignment position 113 indicated that Karbala isolate had high percentages polymorphisms with nucleotide 39 sequences (C) (84.60%) in comparison to Baghdad isolate which has less percentages polymorphisms with nucleotide (T) (15.40%). Also, in alignment position number 221,422, 466 and 553 found that Karbala isolates had higher percentages of nucleotide polymorphisms than found in Baghdad isolate (61.50% and 94.90%) respectively in 39 nucleotide sequences (A, T, C and G). On the other hand, high percentage of polymorphisms in Baghdad isolate had seen only in positions 859 and 865 (71.80% and 92.30%). The results showed that high incidences of changing bases C and G into T base in positions (43, 80,183,188,539 and 736). Also, the two isolates have C base instead of T in positions (164, 170 and 555). Consequently, high percentages of many polymorphisms in V1 region that belongs to Karbala isolates may have increased the virus virulently and made "resistant tomato varieties" vulnerable to TYLCV.

Table (4): Percentage of SNPs identified in the partially sequenced TYLCV Iraqi isolates.

Local TYLV	Alignment position	Number of sequences	Polymorphism type	Percentages	Base
---	43	39	SNP	87.20%	C
B K	43	39	SNP	12.80%	T
---	80	39	SNP	30.80%	C
B K	80	39	SNP	69.20%	T
---	109	39	SNP	17.90%	A
B K	109	39	SNP	82.10%	G
K	113	39	SNP	84.60%	C
B	113	39	SNP	15.40%	T
B K	164	39	SNP	87.20%	C
---	164	39	SNP	10.30%	T
B K	170	39	SNP	92.30%	C
---	170	39	SNP	5.10%	T
---	183	39	SNP	17.90%	C

Local TYLV	Alignment position	Number of sequences	Polymorphism type	Percentages	Base
B K	183	39	SNP	82.10%	T
---	188	39	SNP	23.10%	G
B K	188	39	SNP	76.90%	T
---	209	39	SNP	48.70%	A
B K	209	39	SNP	51.30%	G
K	221	39	SNP	61.50%	A
B	221	39	SNP	35.90%	T
B K	245	39	SNP	89.70%	A
---	245	39	SNP	10.30%	T
---	308	39	SNP	10.30%	C
B K	308	39	SNP	89.70%	T
B K	320	39	SNP	89.70%	C
---	320	39	SNP	7.70%	G
B K	326	39	SNP	89.70%	C
---	326	39	SNP	10.30%	T
---	350	39	SNP	12.80%	A
B K	350	39	SNP	87.20%	G
B K	394	39	SNP	92.30%	C
---	394	39	SNP	7.70%	T
B	422	39	SNP	5.10%	C
K	422	39	SNP	94.90%	T
B	466	39	SNP	5.10%	C
K	466	39	SNP	94.90%	T
B K	515	39	SNP	94.90%	A
---	515	39	SNP	5.10%	C
---	533	39	SNP	10.30%	A
B K	533	39	SNP	89.70%	G
---	539	39	SNP	87.20%	G
B K	539	39	SNP	12.80%	T
B K	541	39	SNP	94.90%	A
---	541	39	SNP	5.10%	T
K	553	39	SNP	94.90%	G
B	553	39	SNP	5.10%	T
B K	555	39	SNP	94.90%	C
---	555	39	SNP	5.10%	T
B K	562	39	SNP	87.20%	A
---	562	39	SNP	12.80%	C
---	572	39	SNP	5.10%	C
B K	572	39	SNP	92.30%	G
K	577	39	SNP	5.10%	C
B	577	39	SNP	92.30%	T
K	578	39	SNP	84.60%	C
B	578	39	SNP	15.40%	G
K	581	39	SNP	92.30%	A
B	581	39	SNP	7.70%	G
B	601	39	SNP	35.90%	C
K	601	39	SNP	64.10%	T
B	665	39	SNP	12.80%	A
K	665	39	SNP	87.20%	T
---	678	39	SNP	20.50%	A

Local TYLV	Alignment position	Number of sequences	Polymorphism type	Percentages	Base
B K	678	39	SNP	79.50%	G
B K	696	39	SNP	94.90%	C
---	696	39	SNP	5.10%	T
K	734	39	SNP	79.50%	G
B	734	39	SNP	20.50%	T
---	736	39	SNP	5.10%	G
B K	736	39	SNP	94.90%	T
---	744	39	SNP	7.70%	G
B K	744	39	SNP	92.30%	T
---	762	39	SNP	10.30%	A
B K	762	39	SNP	89.70%	G
---	783	39	SNP	48.70%	A
B K	783	39	SNP	51.30%	G
---	792	39	SNP	7.70%	A
B K	792	39	SNP	92.30%	G
B K	804	39	SNP	94.90%	C
---	804	39	SNP	5.10%	T
---	847	39	SNP	15.40%	A
B K	847	39	SNP	84.60%	G
B K	858	39	SNP	74.40%	G
---	858	39	SNP	25.60%	T
K	859	39	SNP	28.20%	A
B	859	39	SNP	71.80%	G
K	865	39	SNP	7.70%	A
B	865	39	SNP	92.30%	G

B: Baghdad isolate. K: Karbala isolate. ---: References

Protein Structure and Function Prediction

Analysis of the amino acid composition of both proteins revealed that leucine and valine are the most predominant amino acids (8.5%) in Baghdad isolate while, serene and arginine counted for 7.8% and 7.1% predominant amino acids (Table 5). The instability index of both proteins are 40.48 and 60.11, which suggests that the protein would be unstable in solution. Grand average of hydropathicity index (GRAVY) were 0.129 and 0.060 indicated that the proteins are hydrophobic(31).

The amino acid sequences of Baghdad and Karbala had 31and 29 negatively charged residues (Aspartic acid+Glutamic acid) and 19 and 23 positively charged residues (Arginine+Lysine) as shown in table (5). The molecular formula of the protein found as $C_{1548}H_{2402}N_{400}O_{414}S_{18}$ and $C_{1482}H_{2357}N_{411}O_{394}S_{26}$.

Also, analysis of the physicochemical characteristics of partially sequenced TYLCV isolates by ExPASy program revealed that the Baghdad isolate protein is more basic than Karbala isolate (estimated isoelectric point [pI] = 9.59 and 8.66 respectively) with a molecular weight of 33817.61Da for Baghdad protein TYLCV isolate and 33070.09Da for Karbala protein TYLCV isolate (Table 6).

Table (5A): Amino acids of partially sequenced TYLCV Baghdad isolate.

Partially sequenced Baghdad TYLCV isolate			
10	20	30	40
AGFEIFFRVW	MDENIKKQNH	TNQVMFFLVR	DRRPYGSSPM
50	60	70	80
DFGQVFNMFD	NEPSTATVKN	DLRDRFQVMR	KFHATVIGGP
90	100	110	120
SGMKEQALVK	RFFKINSHVT	YNHQEAAKYE	NHTENALLLY
130	140	150	160
MACTHASNPV	YATMKIRIYF	YDSISNNLYF	ISASVTFIVF
170	180	190	200
SSTSYNTSTA	LITLLIDITP	ILSKYLITCL	NTLKKRPVGC
210	220	230	240
NVVQIRKLRK	HLIPNTFLML	WLNLIMEMMS	WFIRNGLWLC
250	260	270	280
SVILKRGLFI	SQIKTPFSAG	AVMSSPVRES	MIVAVEVEVV
290	300		
AATVVYTLTP	YWFLLGYLGG		

Table (5 B): Amino acids of partially sequenced TYLCV Karbala isolate.

(B) Partially sequenced Karbala TYLCV isolate			
10	20	30	40
AVQSIIFLGV	WMDENYQEAE	SHSGHVLFWS	VIEGPMEAAQ
50	60	70	80
WILDRFLICS	IMSPVRQPRM	ICVIGFKGNF	MQQLLVGPLE
90	100	110	120
RNRHLRDFLK	LTVMLIIIRR	QPSTRTILKT	PCYCIWHVRM
130	140	150	160
PLIQCMQLKY	ASISMIQYLI	NKIYMLYHEF	LLHLLCFQVH
170	180	190	200
HTIHDQLLLH	CLKLHQDYLN	TELHIILLRN	DQSEAVMSSK
210	220	230	240
FGSENICESP	LPSCCGILSE	WKCRGSLEMA	AGCVLLSNRG
250	260	270	280
DCLSPRKRHS	LPDEQVPLCV	NPLLQLRWRY	EQPQSRSTRL

Table (6A): Shows Result of Physicochemical and Functional Characterization of predicted protein for Baghdad isolate

Properties	Value
Sequence length	293
Molecular weight (Da)	33817.61
Theoretical pI	9.59
Extinction coefficients (M-1 ·cm-1)	45630
Instability index	40.48
Aliphatic index	93.11
GRAVY index	0.129
Estimated half-life	4.4 hours (mammalian reticulocytes, *In vitro*). >20 hours (yeast, *In vivo*). >10 hours (*Escherichia coli*, *In vivo*)
Total number of positively charged residues (Arg + Lys)	31
Total number of negatively charged residues (Asp + Glu)	19

Table (6B): Shows Result of Physicochemical and Functional Characterization of predicted protein for Karbala isolate.

Properties	Value
Sequence length	283
Molecular weight (Da)	33070.09
Theoretical pI	8.66
Extinction coefficients (M-1 ·cm-1)	45795
Instability index	60.11
Aliphatic index	108.13
GRAVY index	0.060
Estimated half-life	4.4 hours (mammalian reticulocytes, *In vitro*). >20 hours (yeast, *In vivo*). >10 hours (*Escherichia coli*, *In vivo*)
Total number of positively charged residues (Arg + Lys)	29
Total number of negatively charged residues (Asp + Glu)	23

Secondary Structure Analysis

The Raptor X and SOPMA were applied for secondary structure calculations (helix, sheets, and coils) of the hypothetical protein of Baghdad partially sequenced TYLCV isolate. The alpha helix was found to be the most predominant (40%), followed by random coil (29%) and extended strand (25.33%) and beta turn was found as (5.67%). different results were obtained from the rest servers as shown in figure (6).

Figure (6): Predication of the Baghdad partially sequenced TYLCV protein Secondary Structure by SOPMA Server.

Homology Modeling

The 3 dimensional structures of the identified proteins are not available in its native forms and computationally predicted by homology modeling. Owing to such circumstances where the experimental structure of the target protein is unavailable, computational programs may help in determining the biological functions of such un-characterized proteins(32). Homology modeling predicts the 3D structure of an assumed protein sequence build principally with respect to its alignments to one or more proteins of known structure. To perform the homology modeling: Raptor X program and SOPA were used to initiate 3D model of the partially sequenced of two isolates of TYLCV. Interestingly, only protein belongs to Baghdad isolates was found to have homology known structure (Figure 7) whereas, no homology known structure was found for Karbala isolate. Hypothetical protein has two domains and mainly consists of beta turns and alpha helix structures. Illustration may be made for this results, that Karbala isolate which has many variations in the genome as represented by high rate of SNPs, may be representing a new evolved TYLCV by which explanation can be made for high rate of TYLCV infection incidences in many Iraqi regions.

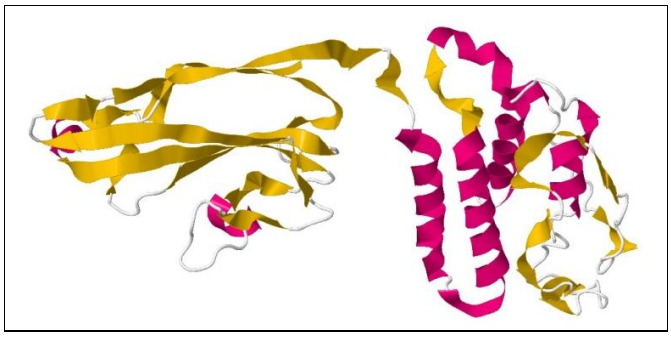


Figure (7): Baghdad partially TYLCV protein 3D by RaptorX Property program.

Conclusion

Several studies have indicated that TYLCV has becomes very infectious and destructive on tomato plants. on the effects of breeding systems on molecular evolution. However, some of them have not been clearly verified by empirical data, and numerous questions remain. Genomic data have also partly unveiled the complexity of breeding systems, especially in asexual or presumably asexual species. Promising prospects include (1) analysis of the rate and pattern of transition to selfing/asexuality using densely sampled phylogenies with appropriate breeding system distributions combined with genome-wide molecular data, (2) distinguishing between the different forms of selection with a better characterization of the fitness effect of mutations, (3) explicitly accounting for the possible association between breeding system shifts and non-equilibrium demo- graphic dynamics (e.g., bottlenecks in selfers, clone turnover in asexual). A large theoretical corpus has already been developed, and thanks to the increasing availability of genomic data, qualitative patterns are now rather well described and partly

understood. Another challenge in the future is also to make our predictions and tests more quantitative.

References

1. Shanmukhi, C.; Reddy, M.; Dorajee, A.; Rao, A. and Babu, P. (2018). Flowering and yield in processing tomato varieties as influenced by planting density and fertigation. *Journal of Pharmacognosy and Phytochemistry*, 7(2):3481–3485.
2. FAOSTAT Crop, n.d., Area harvested, Yield and ProductionQuantity, Food and Agriculture Organization of the United Nations (FAO), viewed 20 January 2020,< http://www.fao.org/faostat/en/#data/QC>.
3. Leite, G. L. D. and Fialho, A. (2018). Protection of Tomatoes Using Bagging Technology and Its Role in IPM of Arthropod Pests. In Sustainable Management of Arthropod Pests of Tomato (pp. 305-311). Academic Press.
4. Glick, E.; Levy, Y. and Gafni, Y. (2009). The viral etiology of tomato yellow leaf curl disease - A Review. *Plant Protection Science*, 45:81-97.
5. Diaz-Pendon, J.A.; Canizares, M.C.; Moriones, E.; Bejarano, E.R.; Czosnek, H. and Navas-Castillo, J. E. S. U. S. (2010). Tomato yellow leaf curl viruses: ménage à trois between the virus complex, the plant and the whitefly vector. *Molecular plant pathology*, 11(4): 441-450.
6. Seal, S.; vandenBosch, F. and Jeger, M. (2006). Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Critical reviews in plant sciences*, 25:23-46.
7. Brown, J. K. (2009). Plant resistance to viruses:Geminiviruses. In B. W. Mahy and M. H. VAN Regenmortel, Desk encyclopedia of plant and fungal virology (52-58). London: Elsevier Academic press, UK. 613 pp.
8. Fondong, V. (2013). Geminivirus protein structure and function. *Molecular Plant Pathology*, 14: 635–649.
9. Brown, JK.; Zerbini, FM.; Navas-Castillo, J.; Moriones, E.; Ramos-Sobrinho, R.; Silva, *et al*. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology*, 16(10): 1593-1619.
10. Briddon, R. W.; Martin, D. P.; Roumagnac, P.; Navas-Castillo, J.; Fiallo-Olivé, E.; Moriones, E., *et al*. (2018). Alphasatellitidae: A new family with two subfamilies for the classification of geminivirus-and nanovirus-associated alphasatellites. *Archives of virology*, 163(9), 2587-2600.
11. Xie, Y.; Jiao, X.; Zhou, X.; Liu, H.; Ni, Y. and Wu, J. (2013). Highly sensitive serological methods for detecting tomato yellow leaf curl virus in tomato plants and whiteflies. *Virology journal*, 10(1) 142.
12. European Food Safety Authority (EFSAPLH), (2013). Scientific opinion on the risks to plant health posed by Bemisia tabaci species complex and viruses it transmits for the EU territory. EFSA Journal, 11: 3162.
13. Noris, E. and Miozzi, L. (2015). Real-time PCR protocols for the quantification of the begomovirus tomato yellow leaf curl Sardinia virus in tomato plants and in its insect vector. In Plant Virology Protocols (pp. 61-72). Humana Press, New York, NY.
14. European Food Safety Authority (EFSAPLH), (2014). Scientific opinion on the pest categorisation of Tomato yellow leaf curl virus and related viruses causing tomato yellow leaf curl disease in Europe. EFSA Journal, 12: 3850.
15. Czosnek, H. (2007). Interactions of Tomato yellow leaf curl virus with its whitefly vector. Dordrecht, the Netherlands: Springer, 157–170.
16. Basak, J. (2016). 'Tomato Yellow Leaf Curl Virus: A Serious Threat to Tomato Plants World Wide'. *Plant Pathology and Microbiology'*, 7:4. doi: 10.4172/2157-7471.1000346.
17. Caciagli, P.; Medina Piles, V.; Marian, D.; Vecchiati, M.; Masenga, V.; Mason, G. *et al*. (2009). Virion stability is important for the circulative transmission of Tomato yellow leaf curl Sardinia virus by Bemisia tabaci, but virion access to salivary glands does not guarantee transmissibility. *Journal of Virology*, 83: 5784–5795.
18. De Barro, P. J.; Liu, S. S.; Boykin, L. M. and Dinsdale, A. B. (2011). Bemisia tabaci: a statement of species status. *Annual review of entomology*, 56: 1-19.

19. Venkataravanappa, V.; Reddy, C.; Shankarappa, K. and Reddy, M. (2019). Association of Tomato Leaf Curl New Delhi Virus, Betasatellite, and Alphasatellite with Mosaic Disease of Spine Gourd (Momordica dioica Roxb. Willd) in India. *Iranian J. Biotech*, 17(1): 17–29.
20. Reddy, R.; Colvin, J.; Muniyappa, V. and Seal, S. (2005). Diversity and distribution of begomoviruses infecting tomato in India. *Archives of virology*, 150(5), 845-867.
21. Li, C. B.; Cui, Y. L.; Zhang, L. Y. and Li, C. Y. (2012). Molecular detection of tomato yellow leaf curl virus. Beijing Vegetable Research Center. *Beijing Academy of Agriculture and Forestry Science*, 34 (3): 366-370.
22. Deng, D.; McGrath, P. F.; Robinson, D. J. and Harrison, B. D. (1994). Detection and differentiation of whitefly-transmitted geminiviruses in plants and vector insects by the polymerase chain reaction with degenerate primers. *Annals of Applied Biology*, 125(2): 327-336.
23. Rienzie, K.; Wickramaarachchi, W; De Costa, D. M. and Wijesooriya, W. M. G. U. (2016). Molecular detection and characterisation of begomovirus causing bean yellowing disease in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*, 44(3).
24. Vaca-Vaca, J.C.; Morales-Euse, J.; RIVERA-TORO, D.M. and Lopez-Lopez, K. (2019). First report of begomoviruses infecting pepper (Capsicum spp.) Crops in Colombia. *Colombian Biological Act*, 24 (3): 452-462.
25. Katoh, K., Misawa, K., Kuma, K. I. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14): 3059-3066.
26. Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3): 307-321.
27. Wang, S.; Ma, J.; and Xu, J. (2016). proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics*, 32(17): i672-i679.
28. Al-Waeli, M.; Dizadji, A.; Mossahebi, G. H.; and Ahangaran, A. (2018). Phylogenetic analysis and genetic variation of Tomato yellow leaf curl virus based on the V1 gene in Iraq. *Genetic Engineering and Biosafety Journal*, 7(1): 1-11.
29. Yang, X.; Zhou, M.;Qian, Y.; Xie, Y. and Zhou, X. (2014). Molecular variability and evolution of a natural population of Tomato yellow leaf curl virus in Shanghai, China. *Journal of Zhejiang University-Science B*, 15(2):133-142.
30. Shirazi, M.; Mozafari, J.; Rakhshandehroo, F. and Shams-Bakhsh, M. (2014). Genetic diversity, host range, and distribution of Tomato yellow leaf curl virus in Iran. *Acta Virologica*, 58:128-136.
31. Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1): 105-132.
32. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J., *et al*. (2017). GenBank. *Nucleic Acids Res*., 45(D1): D37-D42.