High-Throughput Sequencing and Bioinformatic Analysis Reveal Presence of the Endogenous Pararetrovirus Tobacco vein clearing virus Genome in the Tomato (*Solanum lycopersicum*) Host Genome

Mahmood Othman Abass and Adnan A. Lahuf*
Department of Plant Protection, College of Agriculture, University of Kerbala, Karbala, Iraq,
*Email of corresponding author: adnan.lahuf@uokerbala.edu.iq

Abstract

Abass, M.O. and A.A. Lahuf. 2023. High-Throughput Sequencing and Bioinformatic Analysis Reveal Presence of the Endogenous Pararetrovirus Tobacco vein clearing virus Genome in the Tomato (*Solanum lycopersicum*) Host Genome. Arab Journal of Plant Protection, 41(1): 77-84. https://doi.org/10.22268/AJPP-41.1.077084

Endogenous pararetroviral sequences (EPRVs) are repetitive sequences that have been discovered mostly in the Kingdom Plantae, particularly in various species of the family Solanaceae. In this study, the draft genome of an endogenous retrovirus identified by next generation sequencing (NGS), was found integrated in the genome of *Solanum lycopersicum*. Results of the homology alignment revealed that the virus identified was Tobacco vein clearing virus (TVCV), is a member of the genus Solendovirus, family Caulimoviridae. It consists of a double-stranded DNA genome of 7,760 nucleotides in length. Additionally, it has four open reading frames (ORFs), which encodes Solendovirus typical conserved domains that comprise the putative coat protein (ORF1), putative cell-to-cell movement protein (ORF2), the polyprotein (ORF3), which comprises the aspartic protease, reverse transcriptase and RNase H, as well as the putative Trans-activator factor (ORF4). Sequence alignment analysis revealed that the Iraqi TVCV had 81.60% sequence identity to the INSDC Tobacco vein clearing virus (AF190123.1), that was only reported to be integrated in the genome of some species of Nicotiana spp. However, in the current study, TVCV genome was identified associated with genome of *S. lycopersicum*. This new fact was further verified through BLASTn analysis that confirmed the presence of TVCV genome associated with the genome of several cultivated and wild *S. lycopersicum* worldwide. In conclusion, the TVCV is the first EPRVs of Solendovirus members discovered in the *S. lycopersicum* identified from Iraq.
Keywords: Tobacco vein clearing virus; *Solanum lycopersicum* genome, next generation sequencing, NGS.

Introduction

Endogenous pararetroviral sequences (EPRVs) are retroelements double stranded DNA viruses that integrated in the nuclear genome of their hosts using reverse transcription for replication and thus are transmitted among host generations as a normal cellular gene. They are ubiquitous in the planta kingdom, and frequently descend from members of either the *Caulimoviridae* or *Geminiviridae* families (Diop *et al*., 2018; Sharma *et al*., 2020). It has been believed that the EPRVs are basically molecular fossils of those viruses existed millions of years ago. Therefore, they present important understanding into evolution of viruses through time (Aiewsakun & Katzourakis, 2015; Geering *et al*., 2014). They also significantly affect evolution of plants through adding novel protein-coding genes or exons in their genome hosts (Carrasco *et al*., 2019; Liu *et al*., 2010). Moreover, they are likely assessing in form the transcriptome of their hosts via providing supplies of non-coding regions and/or novel promoter regulatory elements (Grandi & Tramontano, 2018). Occasionally, the EPRVs, which represent the full viral genomes, maintain replication-competency. However, they can be reactivated due to abiotic or biotic stress and cause new infections (Chabannes *et al*., 2013; Richert-Pöggeler *et al*., 2003).

Tobacco vein clearing virus (TVCV) has 50 nm in diameter virions consists of a 45 kDa capsid protein and a 7767 bp dsDNA genome. Despite of its genome sequence and arrangement and being disrupted via a site-specific discontinuity, it is comparable to caulimoviruses in virion and genome features (Hull, 1984). In fact, TVCV is very closely similar in nucleotide sequence and genome organization to the Cassava vein mosaic virus (CsVMV), which has been suggested as the type member of a endogenous pararetroviruses subgroup, intermediate in genome composition between caulimoviruses and badnaviruses genera (de Kochko *et al*., 1998). This indicates clearly to the fact that TVCV is a EPRVs member. This virus was first identified in the tobacco species *Nicotiana edwardsonii* and was found to be substantially seed-transmitted to all progeny plants. However, it was not transmitted mechanically by grafting or insect to any of other members of the genus Nicotiana. Later, it was discovered that its genome is integrated to the genomic DNA of *N. edwardsonii*, *N. glutinosa*, *N. tabacum* and *N. rustica*. However, it was not found in the genomic DNA of *N. clevelandii* (Lockhart *et al*., 2000). Furthermore, the EPRVs have been reported to be integrated in the nuclear genomes of numerous plants such as tobacco (Gregor *et al*., 2004; Jakowitsch *et al*., 1999;), rice (Kunii *et al*., 2004), petunia (Richert-Pöggeler *et al.,* 2003) and potato (Hansen *et al*., 2005). However, in Iraq, studies regarding the EPRVs are few. Thus, in order to enhance our understanding of the EPRVs particularly the TVCV in economically important crops, such as *S. lycopersicum,* this study was conducted.

https://doi.org/10.22268/AJPP-41.1.077084
© 2023 Arab Society for Plant Protection الجمعية العربية لوقاية النبات

Materials and Methods

Plant materials and DNA extraction

Leaf samples of various varieties of *Solanum lycopersicum* L. grown commonly in tomato frames in Karbala Province (32°36′52″ N, 44°1′27″ E), Iraq, were collected randomly. Genomic DNA was extracted from these leaves using the AccuPrep® Genomic DNA Extraction Kit (Bioneer/South Korea) following the manufacturer's instructions.

DNA sequencing and bio-informatic analysis

The genomic DNA extracted was used to prepare an Illumina paired-end shotgun library via the TruSeq Nano DNA Kit (Illumina, San Diego, California, USA). This library was sequenced at Macrogen, Inc. (Seoul, South Korea) using the NovaSeq 6000 Illumina platform (151 bp paired-end reads) (Zakeel *et al*., 2021). The raw reads obtained were aligned to the genome sequence of *S. lycopersicum* SL3.0 operating the Bowtie2 (V. 2.4.5) tool, and the unaligned reads were subsequently subjected to De Novo assembly procedure with SPAdes (V3.15.4) tool. The contigs produced were compared against the GenBank database of plant viruses (https://www.ncbi.nlm.nih.gov/genome/viruses/) obtained from NCBI using BLASTn and BLASTp. The reference sequence of the endogenous pararetrovirus of TVCV (GenBank accession number NC_003378.1) identified in this study was utilized in map to reference assembly procedure with the raw reads using Geneious Prime® (V2022.1.1) in order to verify our findings.

All the open reading frames (ORFs) of the EPRV identified were annotated via BLASTp tool against the NCBI non-redundant protein database. Sequence comparisons against plant endogenous pararetroviruses retrotransposon were accomplished operating multiple alignments tool in ClustalW algorithm available at molecular evolutionary genetics analysis (MEGA; V.10.1.5) software. A phylogenetic tree was built with the multiple alignment in neighbor-joining tree under bootstrap analysis (1000 replicates), using MEGA package (Tamura *et al*., 2011; Felsenstein, 1985). Additionally, the draft genome sequence of the detected TVCV and the reference sequence of the same virus (NC_003378.1) was BLASTn against the NCBI nucleotide collection (nr/nt) database to verify presence of this virus in other genomes of tomato hosts.

Results and Discussion

Identification and genome characterization of TVCV

An EPRV-like sequence was AT-rich (74.5%), and comparable to a previously identified EPRVs in *Nicotiana* sp. (Jakowitsch *et al*., 1999), uncovering up to 87.4% sequence pairwise identity and 98.1% query coverage of the NCBI reference sequence of Tobacco vein clearing virus (TVCV) (NC_003378.1; Figure 1). Additionally, there were 145,356 matching reads reached 0.278% percentage of the total raw reads. The average coverage of the TVCV sequence was 2825 X.

The final assemblage of the sequence of TVCV genome determined (GenBank accession# ON684329) using overlapping sequences was 7,760 bp comprising four open reading frames (ORFs): coat protein (CP), cell-to-cell movement protein (MP), polyprotein (POL) and Trans-activator protein (TAV). This draft genome and its putative proteins comprised the entire attributes estimated to exist in the endogenous pararetrovirus members that is represented in the reference genome of TVCV (Figure 2).

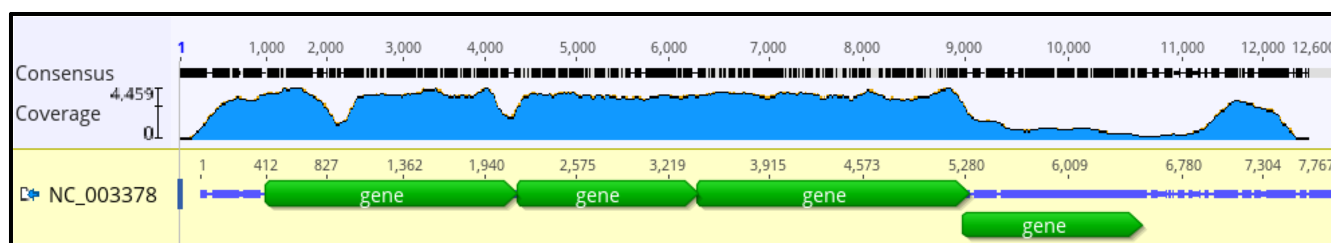

Figure 1. The raw reads coverage of the endogenous pararetroviral reference genome (TVCV; NC_003378.1).

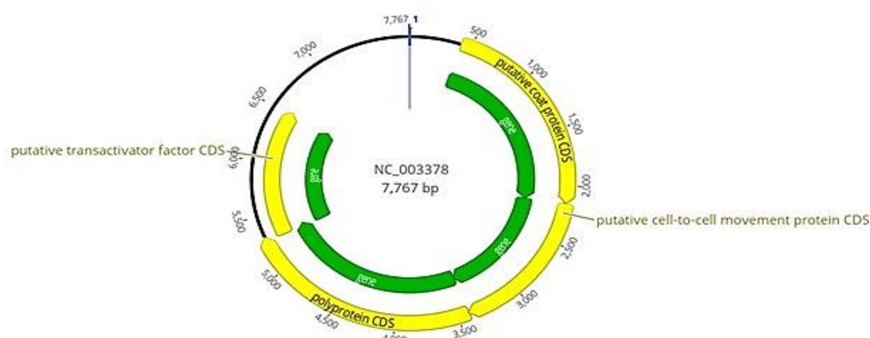

Figure 2. Genome map of a representative EPRV, Tobacco vein clearing virus (NC_003378.1). The preserved protein domains are identified in different reading frames. Genes (green) and coding sequences (CDS, yellow) of the representative viral genome of the EPRV are denoted in the chart.

The order of protein functions was conserved and showed similarity with the same sequences of the four proteins of TVCV. The putative sequence of amino acid identities of the four coding regions ranged from 63.40 to 82.83% while the quarry coverage ranged between 93 to 100%. For example, the CP amino acid sequence displayed 100 % quarry coverage and 63.40 % sequence similarity to the fragments of TVCV-CP sequence with accession number NP_569139.1 (Table 1). On the other hand, the MP protein domain showed 93 % quarry coverage and 75.89% identical amino acid sequence with TVCV-MP that has accession number NP_569140.1 (Table 2), while POL and TAV revealed 100 and 96 % quarry coverages as well as 82.83 and 64.00 similarity to those protein sequence with accession numbers NP_569141.1 and NP_569142.1 of TVCV-POL and TAV, respectively (Tables 3 and 4). These results were confirmed via the phylogenetic analysis that indicated clearly to the close relationship among the CP, MP, POL and TAV proteins of the Iraqi strain of the TVCV with the same proteins that belong as well to the TVCV (Figures 3, 4, 5, 6, respectively). It is worth mentioning that each of the protein-coding regions is either truncated or harbor some frameshifts and can consequently be believed translationally defective, which is a trait also realized previously with the EPRVs in *Nicotiana* spp. (Jakowitsch *et al*., 1999). The putative non-coding intergenic region (IGR) of the virus was identified in the full draft genome that was altogether 2276 bp on both terminals.

Verification presence of the TVCV genome in other global *S. lycopersicum* genomes:

To examine whether the TVCV genome is integrated in worldwide *S. lycopersicum* genomes, a BLASTn search of the tomato genome assembly was performed operating the full genome of the TVCV strain Iraq-1 (GenBank Accession Number ON684329) and the reference sequence of the same virus as the query. Significant hits were obtained (Table 5) that revealed the virus genome sequence was identified in all the 12 chromosomes of tomato host plant. This result confirms our finding of the TVCV in the tomato plant cultivated in Iraq. Although, the presence of TVCV in numerous tomato genome assembly has been described, it has not been reported worldwide in tomato plant. This is likely because the TVCV is an endogenous pararetrovirus that can only be identified through conducting a BLAST analysis (taxid:10239) data.

Table 1. Similarity between amino acid sequence of the putative coat protein of the Tobacco vein clearing virus strain Iraq-1 and other related worldwide amino acid sequences.

Description of the protein	Virus name	Percentage identity %	Quarry coverage %	Total score	Accession number	Country
putative coat protein	Tobacco vein clearing virus	63.40	100	431	NP_569139.1	USA
coat protein	Sweet potato vein clearing virus	40.70	68	192	AWA81913.1	USA
coat protein	Sweet potato vein clearing virus	39.69	68	184	YP_004300272.1	Dominican Republic
coat protein	Blueberry red ringspot virus	31.68	39	65.9	AFL90606.1	Poland
coat protein	Blueberry red ringspot virus	31.68	39	65.9	AFL90601.1	Poland
coat protein	Blueberry red ringspot virus	31.68	39	65.9	AFL90605.1	Poland
coat protein	Blueberry red ringspot virus	31.68	39	65.5	AFL90598.1	Poland
coat protein	Blueberry red ringspot virus	31.68	39	66.6	AFL90602.2	Poland
coat protein	Blueberry red ringspot virus	31.68	39	66.2	AFL90603.2	Poland
coat protein	Blueberry red ringspot virus	30.77	38	63.5	AGI44295.1	Serbia
TPA: coat protein	Bacopamonnieri virus 3	30.12	83	141	DAF42456.1	India
coat protein	chicory mosaic cavemovirus	29.56	80	97.8	QZH55166.1	Brazil
coat protein	Sweet potato collusive virus	28.63	66	91.7	YP_004347414.1	USA
coat protein	Epiphyllum virus 4	27.98	48	90.9	YP_010087806.1	USA
capsid protein	Blueberry red ringspot virus	27.66	58	66.2	AFK73388.1	South Korea

Table 2. Similarity between amino acid sequence of the putative cell-to-cell movement protein of the Tobacco vein clearing virus strain Iraq-1 and other correlated global amino acid sequences.

Description of the protein	Virus name	Percentage identity %	Quarry coverage %	Total score	Accession number	Country
putative cell-to-cell movement protein	Tobacco vein clearing virus	75.89	93	507	NP_569140.1	USA
movement protein	Sweet potato vein clearing virus	41.27	93	251	AWA81914.1	Fiji
movement protein	Sweet potato vein clearing virus	40.51	93	244	YP_004300273.1	Peru
TPA: movement protein	Bacopa monnieri virus 3	33.67	93	171	DAF42457.1	India
movement protein	Cauliflower mosaic virus	32.72	50	80.1	AGT42087.1	Iran
movement protein	Cauliflower mosaic virus	31.33	56	81.3	BCW03857.1	Italy
ORF1 movement protein	Atractylodes mild mottle virus	31.07	41	82.8	YP_009165746.1	South Korea
movement protein	Cauliflower mosaic virus	30.80	56	80.9	AGT42150.1	Iran
movement protein	Cauliflower mosaic virus	30.40	56	80.9	AHA91306.1	Iran
movement protein	Cauliflower mosaic virus	30.40	56	80.5	AHA91293.1	Iran
movement protein	Carnation etched ring virus	30.29	48	79.7	CAH59634.1	India
movement protein	Cauliflower mosaic virus	29.95	50	80.1	AII00786.1	Czech
movement protein	Cauliflower mosaic virus	29.95	50	80.1	AII80282.1	Czech
movement protein	Cauliflower mosaic virus	29.95	50	80.1	BAO53438.1	Japan
movement protein	Soybean Putnam virus	29.67	49	81.3	YP_006607888.1	USA

Table 3. Similarity between amino acid sequence of the putative polyprotein of the Tobacco vein clearing virus strain Iraq-1 and other correlated global amino acid sequences.

Description of the protein	Virus name	Percentage identity %	Quarry coverage %	Total score	Accession number	Country
polyprotein	Tobacco vein clearing virus	82.83	100	939	NP_569141.1	USA
replicase	Sweet potato vein clearing virus	60.79	99	721	AWA81915.1	USA
replicase	Sweet potato vein clearing virus	60.63	99	721	YP_004300274.1	Peru
TPA: replicase	Bacopa monnieri virus 3	53.26	98	588	DAF42458.1	India
replicase	chicory mosaic cavemovirus	41.77	99	399	QZH55167.1	Brazil
replicase	Sweet potato collusive virus	41.07	98	391	YP_004347415.1	Portugal
aspartic protease/ reverse transcriptase	Cassava vein mosaic virus	40.90	98	392	NP_056848.1	USA
replicase	Epiphyllum virus 4	38.59	98	380	YP_010087807.1	USA
putative reverse transcriptase	Strawberry vein banding virus	34.00	98	296	ALF37644.1	China
putative reverse transcriptase	Strawberry vein banding virus	34.00	98	295	UCJ01210.1	Canada
ORF V protein	Strawberry vein banding virus	34.00	98	295	CCG14716.1	China
putative reverse transcriptase	Strawberry vein banding virus	34.00	98	295	UCJ01189.1	Canada
ORFV protein	Strawberry vein banding virus	33.85	98	295	ARO77050.1	China
putative reverse transcriptase	Strawberry vein banding virus	33.85	98	295	UCJ01196.1	Canada
hypothetical protein	Strawberry vein banding virus	33.59	98	296	NP_043933.1	USA

Table 4. Similarity between amino acid sequence of the putative polyprotein of the Tobacco vein clearing virus strain Iraq-1 and other correlated global amino acid sequences.

Description of the protein	Virus name	Percentage identity %	Quarry coverage %	Total score	Accession number	Country
putative Trans-activator factor	Tobacco vein clearing virus	64.00	96	434	NP_569142.1	USA
putative inclusion body protein	Sweet potato vein clearing virus	48.46	55	209	AWA81917.1	USA
putative inclusion body protein	Sweet potato vein clearing virus	46.70	55	204	YP_004300276.1	Peru
putative inclusion body protein	Bacopa monnieri virus 3	29.63	65	114	DAF42459.1	India
putative Trans-activator factor	chicory mosaic cavemovirus	26.04%	61%	64.3	QZH55168.1	Brazil

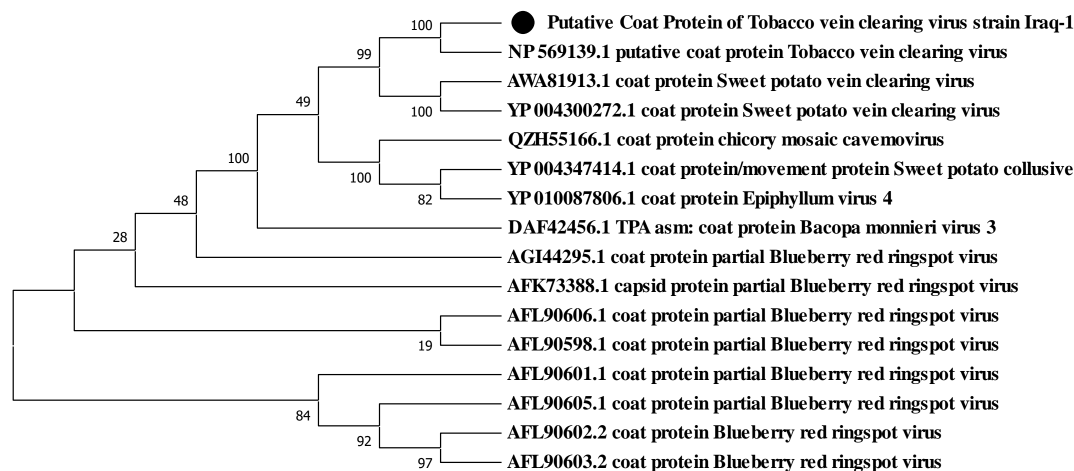

Figure 3. Phylogenetic tree of the putative coat protein of Tobacco vein clearing virus (TVCV) strain Iraq-1 (indicated with a black dot) and different global coat protein sequences of various viruses including the TVCV. This tree was built based on the alignment of the amino acid sequences using MEGA-X applying neighbour-joining method.

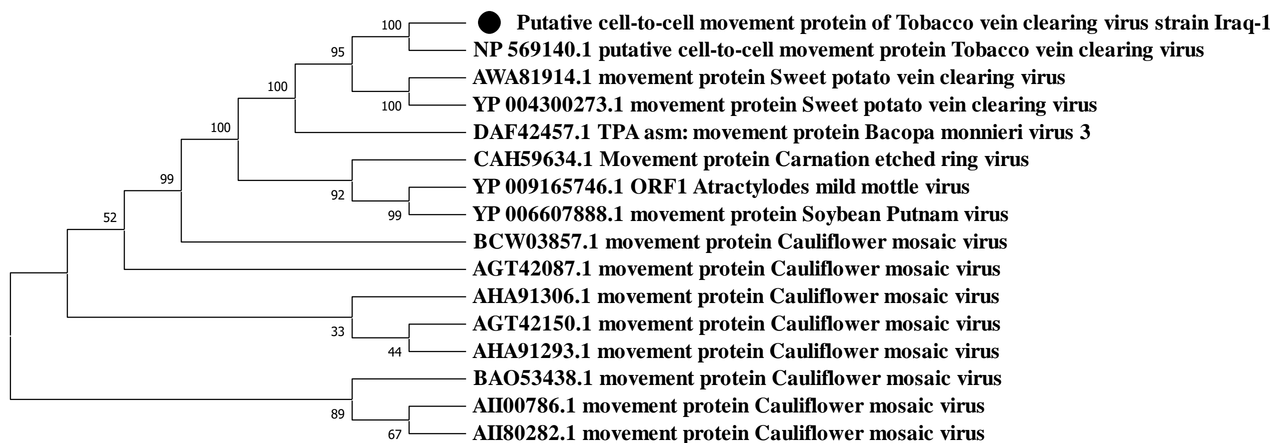

Figure 4. Phylogenetic tree constructed using the amino acid sequences of various viral movement proteins, obtained from GenBank database, including that belongs to TVCV strain Iraq-1 identified in the present study (indicated with a black dot). Phylogenetic distances were calculated operating MEGA-X employing the neighbor-joining method.

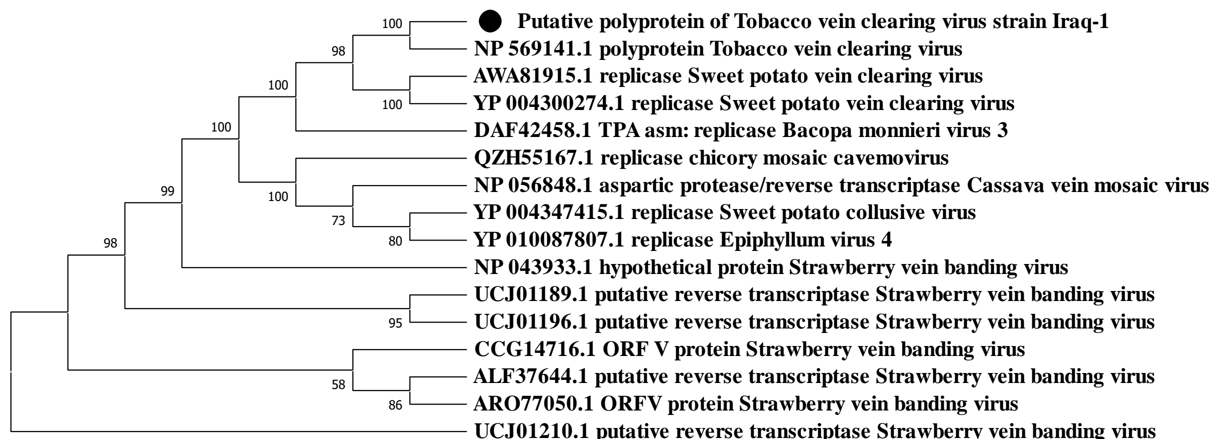

Figure 5. Phylogenetic tree of the putative polyprotein of Tobacco vein clearing virus (TVCV) strain Iraq-1 (indicated with a black dot) and diverse related protein sequences of several viruses comprising the TVCV. This tree was constructed based on the alignment of the amino acid sequences using MEGA-X applying neighbour-joining method.

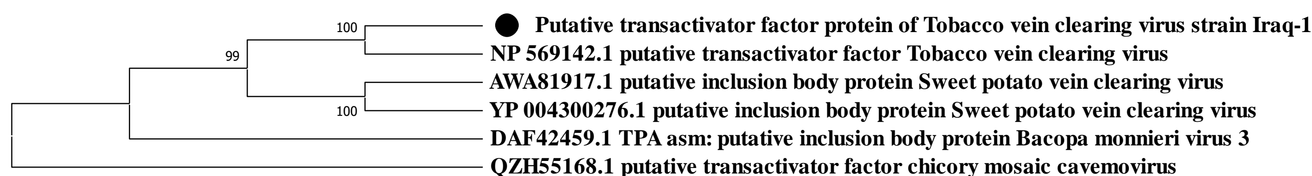

Figure 6. Phylogenetic tree assembled utilizing the amino acid sequences of several viral Trans-activator factor proteins, imported from GenBank database, including that of TVCV strain Iraq-1 (indicated with a black dot). Phylogenetic distances were calculated using MEGA-X and applying the neighbor-joining method.

Table 5. Similarity analysis of the full genome TVCV strain Iraq-1 and related worldwide genome assembly of tomato genome.

Chromosome	Percentage identity %	Quarry coverage %	Total score	Accession number
genome assembly, chromosome: 2	92.80	96	6975	OU640345.1
genome assembly, chromosome: 10	91.78	97	8299	OU640353.1
genome assembly, chromosome: 9	91.29	99	7998	OU640352.1
genome assembly, chromosome: 7	91.19	97	7969	OU640350.1
genome assembly, chromosome: 12	91.09	97	7908	OU640355.1
genome assembly, chromosome: 4	90.72	97	7332	OU640347.1
genome assembly, chromosome: 8	90.67	98	7812	OU640351.1
genome assembly, chromosome: 5	90.53	97	7055	OU640348.1
genome assembly, chromosome: 1	90.35	97	7714	OU640344.1
genome assembly, chromosome: 3	90.18	97	7581	OU640346.1
genome assembly, chromosome: 11	89.82	98	7513	OU640354.1
genome assembly, chromosome: 6	89.43	97	8475	OU640349.1

In this study, a single stretch of 7,760 bp of a nucleotide sequence corresponds to the putative EPRV-like sequence was reported. The revealed endogenous viral element was related to the complete genome of the TVCV with 81.60% nucleotides identity. Furthermore, the collected sequence comprises four open reading frames (ORFs) that share 63.40 % – 82.83% amino acid similarity with proteins members of the TVCV. These coding regions were flanked by altogether 2276 bp homologous to the IGR on both sides and revealed no internal stop codon. Similar to other identified endogenous pararetrovirus sequences (EPRVs), TVCV was found to be integrated in the genome of *S. lycopersicum* and displayed low activity through detection related sequences through deep sequencing technology (NGS). Although, the genomic DNA of TVCV was reported to be hybridized into genomic DNA of *Nicotiana* spp. (Lockhart *et al*., 2000), it has not been recorded to be integrated into other plant genomes. Thus, this is first report of the TVCV integration in the genome of *S. lycopersicum*. The results of current study provide additional evidence for the extensive invasion of the EPRVs represented by the TVCV into the genomes of solanaceous plants. Thus, the integrating viral sequences into the host genome in nature could influence results of virus surveys using conventional PCR for detection, and the survey results end up as an exaggeration of reality.

Complex evolutionary relationships among numerous plant viruses and their hosts have been discovered previously via endogenous viral elements (Katzourakis & Gifford, 2010). Most of these endogenous elements are functionally inactivated in their hosts. However, some of them have recalled their function in limited number of hosts leading to acquiring particular functions that is beneficial to their hosts (Holmes, 2011; Mette *et al*., 2002). This is by insertion of the viral element into or nearby a gene, which can modify the activity and function of the affected gene via a diversity of mechanisms such as inducing gene silencing and adding novel transcriptional regulatory motifs (Kashino-Fujii *et al*., 2018). On the other hand, some of them have the ability to be activated and cause infection to their host plant, as this process has been documented in three endogenous viruses, namely: Petunia vein clearing virus, Banana streak virus and Tobacco vein clearing virus. However, these types of EPRVs were activated as a result of exposure to environmental stress conditions (Alisawi, 2019; Geering *et al*., 2014). Further investigations are needed to explore other EPRVs in tomato genome and possibility of presence the TVCV in other plant hosts.

الملخص

عباس، محمود عثمان وعدنان عبد الجليل لهوف. 2023. تقنية التسلسل عالية الإنتاجية وتحليل المعلوماتية الحيوية يكشف عن وجود مجين الفيروس الداخلي *Tobacco vein clearing virus* ضمن جينوم الطماطم/البندورة *Solanum lycopersicum*. مجلة وقاية النبات العربية، 41(1): 77–84.
https://doi.org/10.22268/AJPP-41.1.077084

إنّ تسلسل الفيروسات الداخلية العكسية Endogenous pararetroviruses (EPRVs) هو تسلسل متكرر تمّ تشخيصه في المملكة النباتية تحديداً في العديد من أنواع العائلة الباذنجانية (Solanaceae). لقد تمّ في هذه الدراسة تشخيص جينوم فيروس داخلي ضمن جينوم الطماطم/البندورة (*Solanum lycopersicum*) باستعمال تقنية الجيل التالي لتحديد التسلسل. أشارت نتائج التشابه إلى أنّ الفيروس المشخص كان *Tobacco vein clearing virus* (TVCV)، وهو أحد أنواع الجنس *Solendovirus* التابع للعائلة *Caulimoviridae*. يتكون جينوم هذا الفيروس من شريط DNA مزدوج طوله 7760 زوج قاعدي (نيوكليوتيد) يتضمن أربعة أطر قراءة مفتوحة (open reading frames)، أي جينات تشفّر المناطق المحافظة النموذجية لجينوم أنواع الجنس *Solendovirus*، يشفر الإطار الأول للغلاف البروتيني (putative coat protein)، والإطار الثاني لبروتين الحركة (putative cell-to-cell movement protein)، أما الإطار الثالث للبروتينات المتعددة (polyprotein) التي تشمل بروتين (aspartic protease) وأنزيم النسخ العكسي (reverse transcriptase) وبروتين RNase H، وأخيراً الإطار الرابع الذي يشفّر لبروتين عامل (Trans-activator factor). لقد كشف تحليل تشابه التسلسلات أن السلالة العراقية لفيروس TVCV هو أكثر تشابهاً بنسبة 81.60% مع سلالة الفيروس نفسه INSDC، ذات الرمز AF190123.1، والتي كانت الوحيدة المسجَّلة لهذا الفيروس، ومتداخلة فقط ضمن جينوم أنواع مختلفة من نبات التبغ *Nicotiana* spp. وعلى الرغم من ذلك، أشارت نتائج الدراسة الحالية وبوضوح إلى أن هذا الفيروس مرتبط بجينوم عائل جديد غير معروف سابقاً وهو الطماطم/البندورة. لقد تمّ تأكيد هذه الحقيقة عن طريق إجراء تحليل التشابة للنيوكليوتيدات (BLASTn analysis) الذي أشار إلى أنّ الفيروس الداخلي TVCV كان موجوداً ضمن جينوم العديد من نباتات الطماطم/البندورة البرية والمزروعة في العالم. واعتماداً على النتائج التي تمّ التوصّل إليها في هذه الدراسة فإن مجين فيروس TVCV هو أول مجين فيروس داخلي من الجنس *Solendovirus* يُكتشف ضمن جينوم الطماطم/البندورة في العراق.

كلمات مفتاحية: Tobacco vein clearing virus، جينوم *Solanum lycopersicum*، تقنية الجيل التالي لتحديد التسلسل (NGS).

عناوين الباحثين: محمود عثمان عباس وعدنان عبد الجليل لهوف*، قسم وقاية النبات، كلية الزراعة، جامعة كربلاء، العراق. *البريد الإلكتروني للباحث المراسل: adnan.lahuf@uokerbala.edu.iq

References

Aiewsakun, P. and A. Katzourakis. 2015. Endogenous viruses: connecting recent and ancient viral evolution. Virology, 479-480: 26-37. https://doi.org/10.1016/j.virol.2015.02.011

Alisawi, O.N. 2019. Virus integration and tandem repeats in the genomes of Petunia. Ph.D. dissertation. University of Leicester. UK. 216 pp.

Carrasco, J.L., J.A. Sánchez-Navarro and S.F. Elena. 2019. Exploring the role of cellular homologous of the 30K-superfamily of plant virus movement proteins. Virus Research, 262: 154-161. https://doi.org/10.1016/j.virusres.2018.02.015

Chabannes, M., F.C. Baurens, P.O. Duroy, S. Bocs, M.S. Vernerey and M. Rodier-Goud. 2013. Three infectious viral species lying in wait in the banana genome. Journal of Virology, 87(15): 8624-3867. https://doi.org/10.1128%2FJVI.00899-13

de Kochko, A., B. Verdaguer, N. Taylor, R. Carcamo, R. N. Beachy and C. Fauquet. 1998. Cassava vein mosaic virus (CsVMV), type species for a new genus of plant double stranded DNA viruses? Archives of Virology, 143(5): 945-962. https://doi.org/10.1007/s007050050344

Diop, S.I., A.D. Geering, F. Alfama-Depauw, M. Loaec, P.Y. Teycheney and F. Maumus. 2018. Tracheophyte genomes keep track of the deep evolution of the *Caulimoviridae*. Scientific Report, 8(1): 572-582. https://doi.org/10.1038/s41598-017-16399-x

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, 39(4): 783-791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Geering, A.D., F. Maumus, D. Copetti, N. Choisne, D.J. Zwickl and M. Zytnicki. 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nature Communications, 5(1): 5269. https://doi.org/10.1038/ncomms6269

Grandi, N. and E. Tramontano.2018. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. Frontiers in Immunology, 9: 2039. https://doi.org/10.3389/fimmu.2018.02039

Gregor, W., M.F. Mette, C. Staginnus, M. Matzke and A.J.M. Matzke. 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiology, 134(3): 1191-1199. https://doi.org/10.1104%2Fpp.103.031112

Hansen, C.N., G. Harper and J.S. Heslop-Harrison. 2005. Characterization of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). Cytogenetic and Genome Research, 110: 559-565. https://doi.org/10.1159/000084989

Holmes, E.C. 2011. The evolution of endogenous viral elements. Cell Host and Microbe, 10(4): 368-377. https://doi.org/10.1016/j.chom.2011.09.002

Hull, R. 1984. Caulimovirus group. CAB/AAB Descriptions of Plant Viruses, no. 295.

Jakowitsch, J., M.F. Mette, J.van der Winden, M.A. Matzke and A.J.M. Matzke. 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. The Proceedings of the National Academy of Sciences, 96: 13241-13264. https://doi.org/10.1073%2Fpnas.96.23.13241

Kashino-Fujii, M., K. Yokosho, N. Yamaji, M. Yamane, D. Saisho, K. Sato and Ma, J.F. 2018. Retrotransposon insertion and DNA methylation regulate aluminum tolerance in European barley accessions. Plant Physiology**,** 178(2): 716-727. https://doi.org/10.1104/pp.18.00651

Katzourakis A. and R.J. Gifford. 2010. Endogenous viral elements in animal genomes. PLOS Genetics, 6(11): e1001191. https://doi.org/10.1371/journal.pgen.1001191

Kunii, M., M. Kanda, H. Nagano, I. Uyeda, Y. Kishima and Y. Sano. 2004. Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. BMC Genomics, 5: 80. https://doi.org/10.1186/1471-2164-5-80

Liu, H., Y. Fu, D. Jiang, G. Li, J Xie and J. Cheng. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. Journal of Virology**,** 84 (22): 11876-11887. https://doi.org/10.1128/jvi.00955-10

Lockhart, B.E., J. Menke, G. Dahal and N.E. Olszewski. 2000. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. Journal of General Virology, 81(6): 1579-1585. https://doi.org/10.1099/0022-1317-81-6-1579

Mette, M., T. Kanno, W. Aufsatz, J. Jakowitsch, J. Van der Winden and M. Matzke. 2002. Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. EMBO Journal, 21(3): 461-469. https://doi.org/10.1093%2Femboj%2F21.3.461

Richert-Pöggeler, K.R., F. Noreen, T. Schwarzacher, G. Harper and T. Hohn. 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogeeous provirus in petunia. EMBO Journal, 22: 4836-4845. https://doi.org/10.1093%2Femboj%2Fcdg443

Sharma, V., P. Lefeuvre, P. Roumagnac, D. Filloux, P.Y. Teycheney and D.P. Martin. 2020. Large-scale survey reveals pervasiveness and potential function of endogenous geminiviral sequences in plants. Virus Evolution**,** 6(2): veaa071. https://doi.org/10.1093/ve/veaa071

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10): 2731-2739. https://doi.org/10.1093/molbev/msr121

Zakeel, M.C.M., D.W.A. Geering, J.E. Thomas and O.A. Akinsanmi. 2021. Characterization of novel endogenous geminiviral elements in macadamia. BMC Genomics, 22: 858-868. https://doi.org/10.1186/s12864-021-08174-0

Received: June 13, 2022; Accepted: July 05, 2022

تاريخ الاستلام: 2022/6/13؛ تاريخ الموافقة على النشر: 2022/7/5