ARCHIVES OF PHYTOPATHOLOGY AND PLANT PROTECTION
https://doi.org/10.1080/03235408.2022.2123601

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Genome sequencing of eggplant reveals *Eggplant mild leaf mottle virus* existence with associated two endogenous viruses in diseased eggplant in Iraq

Bahaa Khaffajah, Osamah Alisawi and Fadhl Al Fadhl

Plant Protection Department, Faculty of Agriculture, University of Kufa, Najaf, Iraq

ABSTRACT

Eggplant has recently suffered a serious threat from a virus infection that causes mosaic and fruit deformation in eggplant fields in Iraq. The infection ratio has registered between 1-80% in examined eggplant fields. The Illumina platform has been applied to sequencing genomic DNA and total RNAseq of symptomatic samples in addition to healthy genomic DNA. The analysis of an entire RNA data revealed *Eggplant mild leaf mottle virus* (EMLMV) as a causal agent of the symptoms. The sequence of EMLMV was composed of 9280 nt encodes ten protein domains and the transcripts per million (TPM) was 9. Two endogenous pararetroviruses registered, *Caulimovirus - SMe* and *SmelV* belong to *Caulimovirus* and *Florendovirus* genera with lengths of 8233 and 6198 nt respectively, and both were found interestingly in the RNA transcripts. Phylogeny has revealed EMLMV and endogenous elements' relations to related viruses.

ARTICLE HISTORY
Received 27 December 2021
Revised 31 August 2022
Accepted 7 September 2022

KEYWORDS
Whole DNA seq; total RNA seq; bioinformatics; eggplant genome; potyviruses

Introduction

Eggplant (*Solanum melongena* L.) belongs to the *Solanaceae* family that includes the genus *Solanum,* the largest of the *Solanaceae* with more than 1,000 species (D'Arcy 1991). Unlike potato and tomato, eggplant has been widely underrated as a target for molecular genetic study due to its lower production and consumption especially in Western countries (Wu et al. 2009). Whole genome sequencing has been recently achieved to construct a draft genome dataset of eggplant with 33,873 scaffolds that covered 74% of the genome (Hirakawa et al. 2014). Interestingly, Wei et al. (2020) reported high-quality reference genome of eggplant

CONTACT Osamah Alisawi osama.alisawi@uokufa.edu.iq Plant Protection Department, Faculty of Agriculture, University of Kufa, Najaf, Iraq.

using combined technologies of Illumina, Nanopore and 10X genomics in addition to Hi-C technology. The assembled genome has 12 chromosomes and a size of ~1.17 Gb, consisting of 36,582 genes and 70.09% (811.14 Mb) repetitive elements. Eggplant has considered as an important host for a wide range of viruses such as *Eggplant mosaic virus*, *Eggplant* mottled dwarf *virus*, *Tobacco ringspot virus*, *Cucumber mosaic virus*, *Potato virus Y*, *Eggplant severe mottle virus*, *Eggplant mottle crinkle virus* and *Alfalfa mosaic virus*. In Iraq, new disease exhibiting mosaic symptoms suspected to be caused by a virus/viral agent have been noticed since 1998 and then registered on eggplant fields in south of Baghdad. The symptoms were very distinguishable on leaf and fruit causing leaf mottling and fruit distortion. Earlier studies including host range and indicator plants suggested that symptoms belong to a virus named temporarily as *Eggplant blister mottled virus* (EBMV). The virus was characterized as a *Potyvirus* depending on its shape showing flexuous particles of 720 nm under electron microscope. To control the virus, further study has evaluated three products of Vit-org nutrient, 2-nitromethyl phenol, and Thuja extract to inhibit the virus in eggplant and the more efficient product was Vit-org nutrient that protected plants in post and pre-applications (Al-Ani et al. 2011). Al-Ani et al. (2011) mentioned that the virus was mechanically transmitted and showed necrotic local lesions on *Zinnia elegans* and *Gomphrena globose*, while *Chenopodium amaranticolor* showed systemic latent infection. Other plants like *Chenopodium quinoa*, *Chenopodium murale*, *Vigna unguiculata*, *Capsicum annum*, and *Phaseolus vulgaris* did not reflect any symptoms. The study above has reported that *Myzus persicae* could transmit the virus particles in non-persistent methods. In 2003, similar symptoms were observed on eggplant in the Jordan and Arava Valleys. The symptoms are caused by a viral agent named *Eggplant mild leaf mottle virus* (EMLMV) that is transmitted by mechanical inoculation and whitefly. The virus was characterized as ssRNA genome with sequence length of 9280 nt and has a large open reading frame encodes 3011 amino acids (Dombrovsky et al. 2012). Dombrovsky et al. (2013) also reported further facts around the virus confirming its filamentous shape (720 nm long), existence of cytoplasmic pinwheels and crystalline structures, close relationship to *Tomato mild mottle virus* (TomMMoV), and interestingly whitefly role in the virus transmission unlikely to *Potyviridae* members.

Endogenous pararetroviruses (EPRVs) have colonized different plant genomes and increasingly discovered based on a massive progress of genome sequencing techniques. Most of these elements belong to the family *Caulimoviridae* and their genera, especially *Caulimovirus*, *Badnavirus, Petuvirus* and *Florendovirus* (Alisawi 2019). Like other *Solanaceae* members, the eggplant genome has been interested in

diversity research alongside their ancestral sources. The genome was sequenced and assembled multiple times for genes and non-coding regions studies (Barchi et al. 2019). Currently, three inserted *Badnavirus* elements have been characterized in eggplant genomes within a large-scale search for co-transcripts. The insertion along with Ty-1 copia happened into a late blight resistance gene (*R1*) of eggplant and wild relatives and disrupted *R1* orthologs (Serfraz et al. 2021). In this paper, we report the complete sequence of a new isolate of *Eggplant mild leaf mottle virus* infecting eggplant in Iraq, and novel endogenous pararetroviruses found within eggplant genome.

Materials and methods

Virus survey and plant material

The survey and sample collection were conducted between September 1st and October 15th 2020. Eggplant open and greenhouses fields in South of Baghdad were surveyed for infection ratio, and collecting typical symptomatic leaves from infected plants showed mosaic, stunt, and fruit deformation. Young apical leaves of infected and non-infected eggplant variety Barcelona (Fito seeds company, Spain) were harvested. The samples were put in plastic bags and brought to the Molecular Plant Virus Lab in the Plant Protection Department, Faculty of Agriculture, University of Kufa.

DNA and RNA extraction

From the same leaf of infected plant, two samples were collected for DNA and RNA extractions, while one typical healthy sample was collected for DNA extraction. To extract the DNA, the young leaves (about 2 g) of the infected and healthy eggplant were subjected to cetyl-trimethylammonium bromide (CTAB) method (Doyle and Doyle 1990), with minor modifications (The incubation of preheated CTAB buffer was for 45 min and the precipitated DNA was spun down at 735 g for 5 min).

For RNA extraction, the infected leaf sample was put in an Eppendorf tube and immersed in RNALatter solution, and sent to DNA-link Company, Republic of Korea, together with the extracted DNA samples. The total RNA was extracted following the company instruction.

NGS sequencing

The NGS library preparation was attempted in the company, using TruSeq DNA Library prep kit and TruSeq total RNA library prep kit

for DNA and RNA sequencing respectively. The extracted DNA samples were sequenced at DNA-link Company, Republic of Korea, using Novaseq6000 2×150bp reads technique and application WGS (PCR Free550) based on the manufacturer's procedure. While the quality of RNA sample was checked with a 2100 Expert Bioanalyzer (Agilent), and then sequenced using NovaSeq6000, 2×101PE to obtain the total RNA seq. The coverage of genome sequencing was computed by the following formula: (Number of fastq reads × read length (150 bp))/Eggplant genome size (1.17 Gb). The Sequence Read Archive (SRA) accession numbers of all the NGS sequencing libraries were SRR21032823 for the total RNA seq of the infected sample, and SRR21031625 and SRR21031555 for infected and virus-free genomes respectively.

RepeatExplorer pipeline

To check DNA viruses within raw reads, RepeatExplorer pipeline was applied to identify virus sequences among clusters of repetitive DNA blocks reported (Novák et al. 2013). The clusters were firstly extracted from Archive file, and then Basic Local Alignment Search Tool (Altschul et al. 1990) and Repbase dataset (Jurka et al. 2005) were utilized to identify contigs of each cluster. Further, the extracted sequences were mapped to previously identified virus sequences from DPVweb (https://www.dpvweb.net/) (Adams and Antoniw 2005), and also from the NCBI.

Map to reference

The mappers used in map to reference runs were Geneious and Geneious RNA to map the DNA and RNASeq data (Sensitivity: Medium-Low Sensitivity) to the reference sequences respectively, and the iteration was up to 5 times. The RNA raw reads were mapped against 30 virus sequences that were expected to infect eggplant using Geneious prime version 11 (Kearse et al. 2012), and then the consensus sequence was extracted. Additionally, all plant virus sequences (5040 elements) were taken from GenBank and then concatenated to create only one representative sequence (76, 145,671 nt) and mapped it against the whole RNA reads. The extracted endogenous pararetroviruses from RepeatExplorer pipeline were also mapped against the whole DNA reads of infected and non-infected samples. The outcomes showed in a report that has a number of assembled reads and total used reads. The analysis of the pathogenic RNA virus was computed based on the transcripts per million (TPM) value of the reference virus sequence as the read

count per kilobase of sequence (RPK) divided by a million (Bester et al. 2021). The DNA data was utilized to calculate copy numbers (Number of assembled reads × read length/reference sequence length) and genome proportions (Number of assembled reads/numbers of total NGS reads × 100) of the endogenous pararetroviruses (Mustafa et al. 2018).

Phylogenetic analysis

The maximum likelihood (ML) method was applied to find a phylogeny model using MEGA 11 (Tamura et al. 2013). The Geneious prime V. 11 (Kearse et al. 2012; http://www.geneious.com) was used for alignment and optimized manually. Then, ClustalW alignment was used for extracting sequences of all aligned lengths. The phylogenetic tree has been reconstructed using General Time Reversible (GTR). Bayesian phylogeny inference was applied with Bayesian inference of phylogeny (MrBayes 3.2.6) (Huelsenbeck and Ronquist 2001). Twelve virus sequences were applied in the phylogeny study of pathogenic viruses, and PVY was the outgroup member, while eleven endogenous viruses were used to reconstruct the tree and CaMV was the outgroup member (Figure 1).

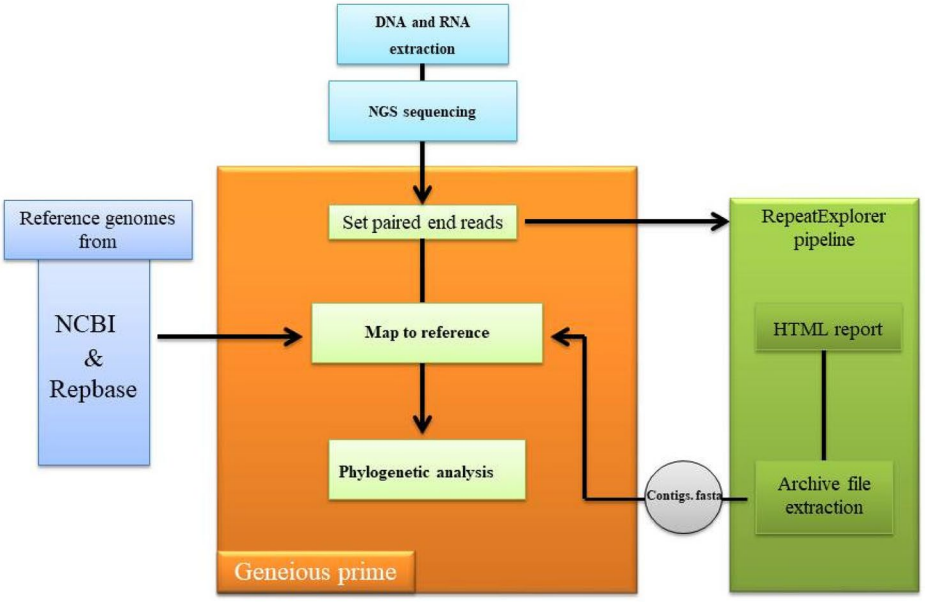

Figure 1. Flowchart shows the outline of the protocols of next-generation sequencing and bioinformatics techniques. DNA and RNA extractions have processed toward NGS and then analyzed through Geneious prime and RepeatExplorer pipeline. The reference genome sequences have been taken from NCBI and Repbase databases.

Results

Symptomatic plants showed a wide range of infection ratios between 1-80% in most of the eggplant fields of South of Baghdad. The NGS Illumina platform produced high quality reads with about 184,331,646, 142,934,264 and 64,631,484 for infected, non-infected DNA and infected RNA samples respectively. The WGS was performed on infected and non-infected eggplant at a high sequencing depth, approximately 23.6x and 18.3x coverage, respectively. The total RNA reads were successfully mapped against *Eggplant mild leaf mottle virus* (Accession number HQ840786.2) with 2,656,616 assembled reads. The consensus sequence with 9280 nt length was extracted and deposited in Genbank under accession number MW503940 as Iraq isolate. The virus sequence encodes ten protein domains starting with peptidase_S76 (P1) followed by Peptidase_C6 (HC-Pro), PRK11664 (P3), DEAH_box_HrpB (6K1), HrpA (C1), Poty-pp (6K2), Peptidase_C4 (NIa Vpg), RdRP-1 (NIa Pro), RNA_dep_RNAP (NIb) and Poty-coat (CP). Interestingly, read coverage was 71801, and read depth was higher in the first third part of the sequence that included P1, HC-Pro and P3 domains than other regions (Figure 2). The transcripts per million (TPM) was 9. The pairwise alignment has revealed 91.4% similarity between Iraq and Israel isolates. Also, the pairwise identities of nucleotide and amino acid sequences of the ten

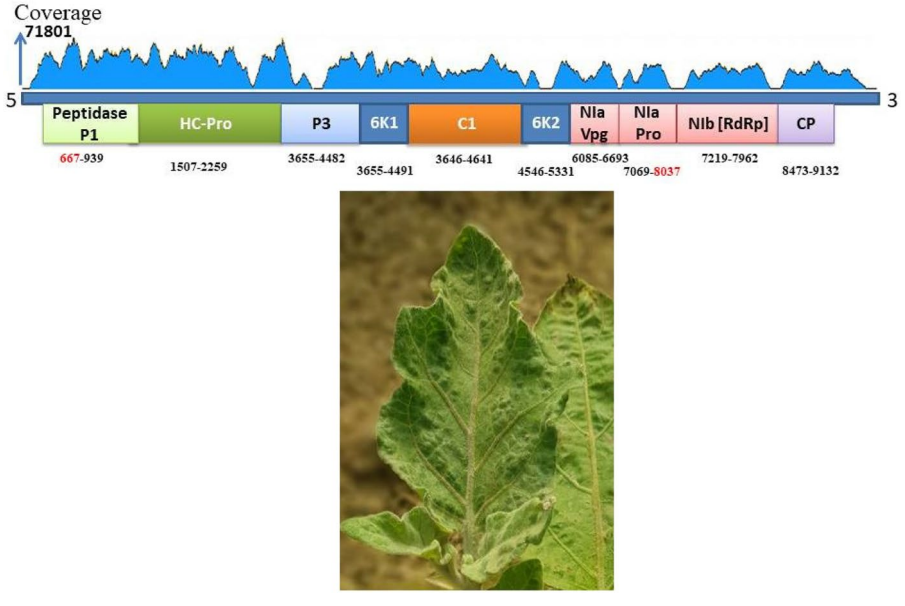

Figure 2. Typical symptoms on eggplant leaf showing mosaic and blistered regions. Schematic design of EMLMV sequence showing genome organization and viral proteins. The blue curve above the EMLMV genome shows assembled reads coverage over the whole sequence. The positions of cleavage sites of protein domains have showed different start position of P1 and end position of Nia Pro.

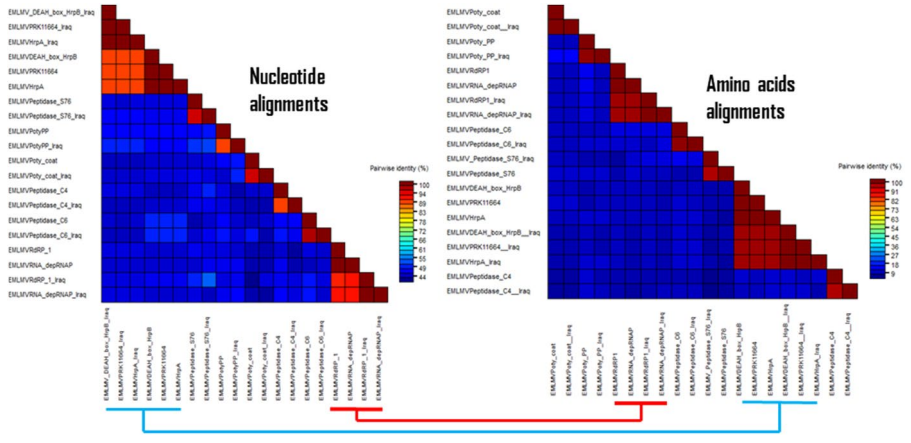

Figure 3. The pairwise similarities of the ten protein domains between EMLMV Iraq isolate and EMLMV (Accession number HQ840786.2). The identities plot of nucleotide and amino acids sequences aligned by ClustalW and presented by Sequence Demarcation Tool (SDTv1.2) software.

protein domains showed higher identities in amino acids than nucleotide sequences (Figure 3). Both Iraq and Israel isolates shared similar positions of the all protein domain cleavage sites except for Peptidase_S76 and RdRP-1. The putative cleavage site of Peptidase_S76 was identified at start position 667 instead of 574, while RdRP-1 was proposed at end position 8037 instead of 7980, unlike the EMLMV Israel isolate based on comparison of the two isolate sequences (Table 1).

For endogenous pararetroviruses, RepeatExplorer pipeline outcomes showed these elements positioned in two out of 225 clusters of repetitive sequence, each cluster representing a particular genus. The cluster 51 belongs to *Caulimovirus* genus and named *Caulimovirus-SMe* following Repbase dataset regulation (RepBase v.21 released in 01-25-2021; https://www.girinst.org/2021/vol21/issue1/Caulimovirus-SMe.html), and the full length was 8233 nt that encoded five protein domains MP, RT-LTR, RVT, RT-RNase, and RNase-H. The assembled

Table 1. The cleavage sites of protein domain positions for both EMLMV-Iraq and EMLMV-Israel isolates.

Protein domains	EMLMV-Iraq	**EMLMV-Israel** (Acc. no HQ840786.2)
Peptidase_S76 (P1)	667–939	574–939
Peptidase_C6 (HC–Pro)	1507–2259	1507–2259
PRK11664(P3)	3655–4482	3655–4482
DEAH_box_HrpB (6K1)	3655–4491	3655–4491
HrpA (C1)	3646–4641	3646–4641
Poty-pp (6K2)	4546–5331	4546–5331
Peptidase_C4 (NIa Vpg)	6085–6693	6085–6693
RdRP-1 (NIa Pro)	7069–8037	7069–7980
RNA_dep_RNAP (NIb)	7219–7962	7219–7962
Poty-coat (CP)	8473–9132	8473–9132

reads were 500,080 and 363,140 for infected and non-infected genomes respectively. Genome proportion of this integrant was 0.25% in healthy DNA reads, while it was 0.27% in infected DNA reads. Copy numbers were 9111 and 6616 copies in infected and non-infected DNA reads respectively (Figure 4). Whereas, the second cluster 53 belongs to *Florendovirus* genus and named as *SmelV* following Geering et al. (2014), and deposited also in RepBase dataset (RepBase v.21 released in 01-25-2021; https://www.girinst.org/2021/vol21/issue1/SmelV.html). The complete sequence of *SmelV* was 6198 nt and encoded four protein domains RT-LTR, RVT, RT-RNaseH, and RNase-H. The assembled reads were 363,415 and 284,198 for infected and non-infected genomes respectively. Genome proportion was 0.19% in both DNA reads, and copy numbers were 8795 copies in infected DNA reads and 6877 copies in healthy DNA reads. Interestingly, however, both endogenous pararetroviruses found also in the RNA transcripts, the assembled reads against *Caulimovirus-SMe* were 4252, and the transcripts per million (TPM) was 0.017, while the assembled reads were 11,807 against *SmelV*, and the transcripts per million (TPM) was 0.06 (Figure 5).

Phylogeny study of pathogenic virus confirmed close relationship between Iraq and Israel isolates, and to *Tomato mild mottle virus* (Figure 6). For endogenous elements, *Caulimovirus-SMe* showed high similarity to endogenous members of *Petunia*, the other plant from *Solanaceae* (*Caulimovirus-PAx* and *Caulimovirus-PIn*) and also to *Cauliflower mosaic*

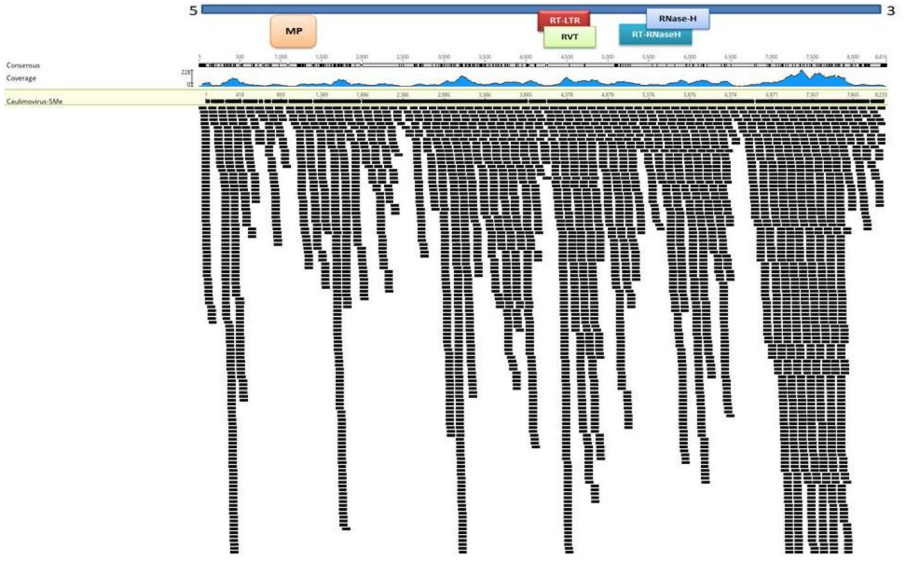

Figure 4. The whole sequence of *Caulimovirus-SMe* showing five protein domains, and the mapped reads against the *Caulimovirus* member with very little coverage.

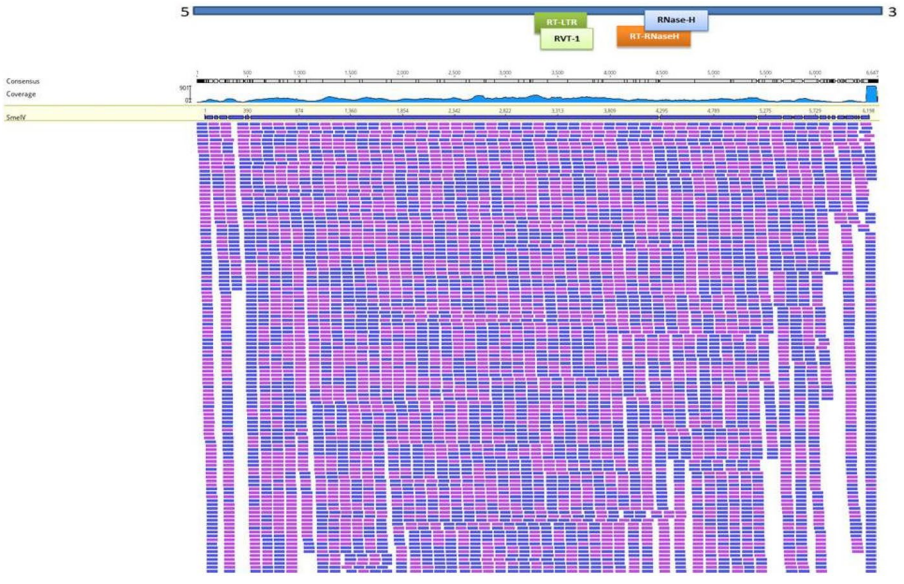

Figure 5. The whole sequence of *Florendovirus* member *SmelV* with four protein domains, and the mapped reads showing little coverage against this endogenous pararetrovirus.

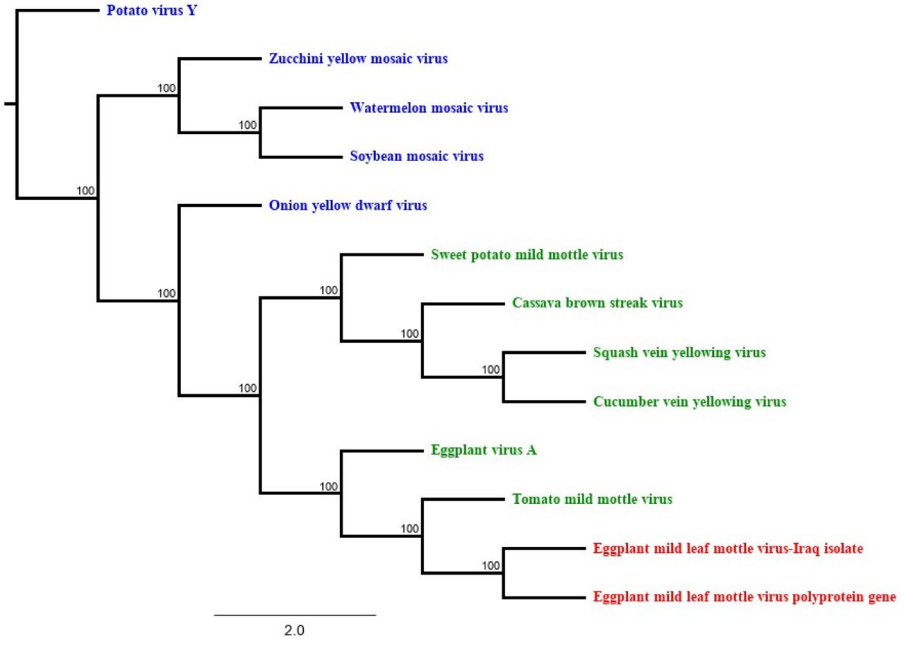

Figure 6. *Eggplant mild leaf mottle virus* (EMLMV) neighbor-joining tree reconstructed based on complete genome sequences of eleven related virus isolates deposited at GenBank, and Potato virus Y was the outgroup member. Scale bar = 2.0.

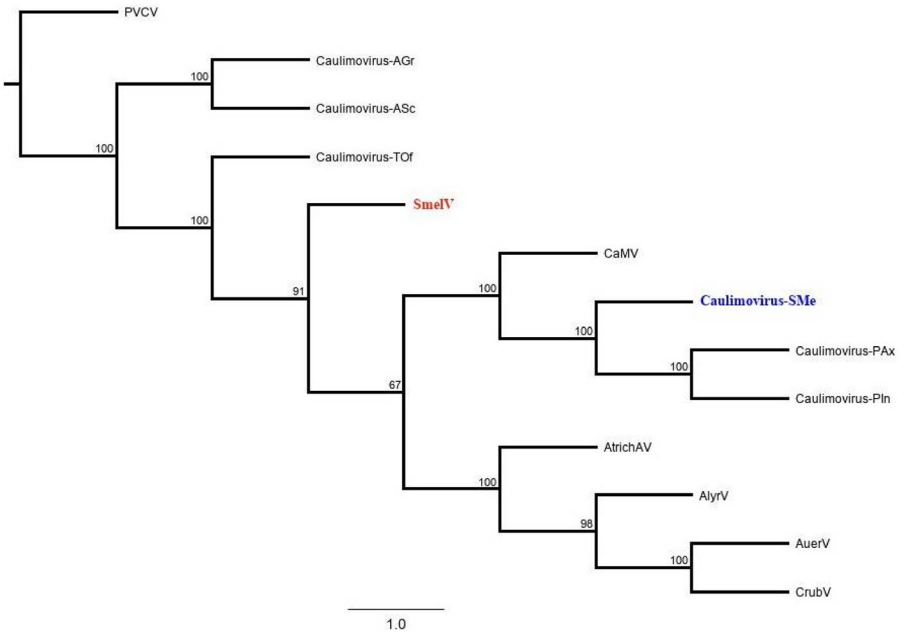

Figure 7. Neighbor-joining phylogenetic tree of *Caulimovirus-SMe* and *SmelV* based on full genome sequences of ten related EPRVs deposited in Repbase database. *Petunia vein clearing virus* (PVCV) was the outgroup member. Scale bar = 1.0.

virus, the typical member of *Caulimoviridea*. While, *SmelV* was clearly linked to other florendovirus members (Figure 7).

Discussion

Whole genome sequencing technique has made a massive progress in genomic research collecting a huge package of data that covered host genome sequences. The technique is considered as unbiased to detect pathogenic elements with no suspected agents have been predicted. In this method, researchers have not used specific antibodies, primers and virus-specific reagents for detecting expected pathogenic factors (Adams et al. 2009). This protocol gave us a complete image of involved viruses in a complex infection that was probably investigated in the infected genome unlike other techniques that look for one suspected virus (Jones et al. 2017). Applying NGS technology that has been an ongoing method alongside bioinformatics to calculate genomic elements was dramatically very useful to capture such unexpected pathogenic viruses and finding accurate statistics for their copy numbers and genome proportions (Huggett et al. 2015; Goodrich et al. 2016). Eggplant in Iraq has suffered a wide distribution symptom such as stunting, mosaic and fruit deformation reached 80% infection ratio, and belonged earlier to a plant virus called temporarily as *Eggplant blister mottled virus* with no evidence

about its exact identity. Multiple attempts like electron microscopy and serological tests tried to figure out more details regarding this pathogen, and they reached only to the level of the virus family, *Potyvirideae*. In this research, we characterized the full sequence of Iraqi isolate of the casual virus *Eggplant mild leaf mottle virus* for the first time, ending a puzzle associated with such eggplant infection. Notably, phylogeny analysis confirmed that Iraq and Israel isolates have shared high similarity sequence and genome organization, suggesting both came from one ancestor. The TPM value reflected virus intensity within infected cells that existed in a high concentration, and also indicated virus particles that infected the host genome. Moreover, copy number and genome proportion values could be easily calculated via NGS, and considered as a good mark for variation of endogenous and also pathogenic elements inside genomes, and this approach applied successfully to obtain correct number of virus copies (Yang et al. 2017; Mustafa et al. 2018; Alisawi 2019). The repetitive nature and high copy number of endogenous elements within plant genomes have complicated next-generation sequencing approach to assemble the complete genome even in the best assembly programs (Schnable et al. 2009). The endogenous pararetroviruses have variable sequence and structure in every individual host genome; however, the hosts have developed strategies to containment amplification, fixation and virulence of EPRVs such as epigenetic silencing and sequence fragmentation (Schmidt et al. 2021). The whole DNA and RNA sequencing alongside with metagenomics, palaeovirology and bioinformatics have clarified the massive role of viruses on host genomes, and revealed the expected origin of viral retroelements. The impact of DNA polymerases in the cell has secured the genome amplification, however, the role of RNA splicing, ribosomal RNA, RNA-dependent RNA polymerases and telomerases can be encountered as well (Gilbert 1986; Heslop-Harrison 2000; Krupovic et al. 2018). Interestingly, the infectious retrolements and virion forming should be highly rated for their effect on the horizontal transfer, emerging diseases and host genome (Krupovic and Koonin 2017).

Recently, many novel endogenous pararetroviruses have been increasingly discovered over different plant species. However, their exact function still has not been understood very clearly, and more works need to be done to draw a complete image around these elements of the story. Some integrants could rise from host genome to be infective factors such as *Petunia vein clearing virus, Banana streak virus, and Tobacco vein clearing virus* (Richert-Pöggeler et al. 2003). In this context, we confirmed that pararetroviruses are susceptive to be expressed even in very low copies like what we found in eggplant genome. Alisawi (2019) reported that caulimoviruses in *Petunia* could be transcribed

unlike to florendoviruses that showed very weak bands using Reverse-Transcriptase-PCR with no data about interaction between such integrants and episomal PVCV. Further, *Florendovirus* members have been represented in expressed sequence tag (EST) libraries, and noteworthy, early analysis of potato genome of EST databases showed pararetrovirus-like sequences in EST databases, particularly in callus tissues .The expression of those integrants has dependent activity based on host effect and specificity (Hansen et al. 2005; Geering et al. 2014). This research topic needs to be extended to figure out more details about interaction, expression, host specificity of the integrants and episomal viruses in such hosts. Finding little reads of CMV and TMV in the symptomatic leaf is answering why PCR method is not matching the correct virus that caused eggplant infection, and results were mistakenly reported. Phylogeny study of EPRVs presented a confirmation of *Caulimovirus-SMe* and *SmelV* taxonomy.

Conclusion

This study revealed that *Eggplant mild leaf mottle virus* was the causal agent of mosaic disease on eggplant in Iraq using whole genome sequencing and bioinformatics techniques. The result will help agronomists to diminish disease caused by the virus in major fields of eggplant. Further, two novel endogenous pararetroviruses have been found in the eggplant genome as well as in RNA transcripts. This finding provided more affirmations about probably the real act of these integrants that still not fully understood in the genome biology and activity alongside with pathogenic viruses. Phylogeny study revealed our virus of interest and associated EPRVs relationships among related viruses.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Osamah Alisawi http://orcid.org/0000-0002-8344-1113

References

Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N. 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol Plant Pathol. 10(4):537–545.

Adams MJ, Antoniw JF. 2005. DPVweb: An open access internet resource on plant viruses and virus diseases. Outlook Pest Man. 16(6):268–270.

Al-Ani RA, Adhab MA, Hassan KA. 2011. Antiviral activity of Vit-org, 2-nitromethyl phenol and Thuja extract against *Eggplant blister mottled virus* (EBMV). Afr J Microbiol Res. 5(21):3555–3558.

Al-Ani RA, Adhab MA, Ismail KA. 2011. *Eggplant blister mottled virus* (EBMV): A possible new potyvirus characterized from Iraq. J Gener Mol Virol. 3(3):49–54.

Alisawi ON. 2019. Virus integration and tandem repeats in the genomes of Petunia [Doctoral dissertation]. UK: University of Leicester.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Barchi L, Pietrella M, Venturini L, Minio A, Toppino L, Acquadro A, Andolfo G, Aprea G, Avanzato C, Bassolino L, et al. 2019. A chromosome-anchored eggplant genome sequence reveals key events in *Solanaceae* evolution. Sci Rep. 9(1):1–13.

Bester R, Cook G, Breytenbach JH, Steyn C, De Bruyn R, Maree HJ. 2021. Towards the validation of high-throughput sequencing (HTS) for routine plant virus diagnostics: measurement of variation linked to HTS detection of citrus viruses and viroids. Virol J. 18(1):1–19.

D'Arcy WG. 1991. The *Solanaceae* since 1976 with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M, Estrada N, editors. Solanaceae III. London: Royal Botanic Gardens/Kew; p. 75–138.

Dombrovsky A, Sapkota R, Lachman O, Antignus Y. 2012. *Eggplant mild leaf mottle virus* (EMLMV), a new putative member of the genus *Ipomovirus* that harbors an HC-Pro gene. Virus Genes. 44(2):329–337.

Dombrovsky A, Sapkota R, Lachman O, Pearlsman M, Antignus Y. 2013. A new aubergine disease caused by a whitefly-borne strain of *Tomato mild mottle virus* (TomMMoV). Plant Pathol. 62(4):750–759.

Doyle J, Doyle J. 1990. Isolation of DNA from small amounts of plant tissues. BRL Focus. 12:V15.

Geering AD, Maumus F, Copetti D, Choisne N, Zwickl DJ, Zytnicki M, ... , Teycheney PY 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nat. Commun. 5(1):1–11.

Gilbert W. 1986. The RNA world superlattices point ahead. Nature. 319(6055):618–618.

Goodrich D, Tao X, Bohrer C, Lonczak A, Xing T, Zimmerman R, Zhan Y, Scott Jr RT, Treff NR. 2016. A randomized and blinded comparison of qPCR and NGS-based detection of aneuploidy in a cell line mixture model of blastocyst biopsy mosaicism. J Assist Reprod Genet. 33(11):1473–1480.

Hansen CN, Harper G, Heslop-Harrison JS. 2005. Characterisation of pararetrovirus-like sequences in the genome of potato (Solanum tuberosum). Cytogenet. Genome Res. 110(1-4):559–565.

Heslop-Harrison JS. 2000. RNA, genes, genomes and chromosomes: repetitive DNA sequences in plants. In: Ettore Olmo, Carlo Alberto Redi (Eds.), Chromosomes today. Vol. 13; p. 45–56.

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, et al. 2014. Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. DNA Res. 21(6):649–660.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17(8):754–755.

Huggett JF, O'Grady J, Bustin S. 2015. qPCR, dPCR, NGS–a journey. Biomol Detect Quantif. 3:A1–A5.

Jones S, Baizan-Edge A, MacFarlane S, Torrance L. 2017. Viral diagnostics in plants using next generation sequencing: computational analysis in practice. Front Plant Sci. 8:1770.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110(1–4):462–467.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647–1649.

Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering ADW. 2018. Ortervirales: New virus order unifying five families of reverse-transcribing viruses. J Virol. 92:1–5.

Krupovic M, Koonin EV. 2017. Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. J. Virol. 91:e210–e217.

Mustafa SI, Schwarzacher T, Heslop-Harrison J. 2018. Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. Mitochondrial DNA Part A. UK: Taylor and Francis Online, 1–14.

Mustafa SI. 2018. Mitochondrial and repetitive DNA defining the sheep genome landscape [Doctoral dissertation]. University of Leicester, 298 pages.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 29(6):792–793.

Richert-Pöggeler KR, Noreen F, Schwarzacher T, Harper G, Hohn T. 2003. Induction of infectious *Petunia vein clearing (pararetro) virus* from endogenous provirus in petunia. Embo J. 22(18):4836–4845.

Schmidt N, Seibt KM, Weber B, Schwarzacher T, Schmidt T, Heitkam T. 2021. Broken, silent, and in hiding: Tamed endogenous pararetroviruses escape elimination from the genome of sugar beet (Beta vulgaris). Ann Bot. 128(3):281–299.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science. 326(5956):1112–1115.

Serfraz S, Sharma V, Maumus F, Aubriot X, Geering AD, Teycheney PY. 2021. Insertion of badnaviral DNA in the late blight resistance gene (R1a) of brinjal eggplant (*Solanum melongena*). Front Plant Sci. 12

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.

Wei Q, Wang J, Wang W, Hu T, Hu H, Bao C. 2020. A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. Hortic Res. 7(1):1–15.

Wu F, Eannetta NT, Xu Y, Tanksley SD. 2009. A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet. 118(5):927–935.

Yang H, Liu M, Jiang H, Zeng Y, Jin L, Luan T, Deng Y, He N, Zhang G, Zeng X. 2017. Copy number variation analysis based on gold magnetic nanoparticles and fluorescence multiplex ligation-dependent probe amplification. J Biomed Nanotechnol. 13(6):655–664.