Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/6DWMY

ENDOGENOUS PARARETROVIRUSES (EPRVs) IN LETTUCE GENOME ARE LESS EXPRESSED BUT PROBABLY SHARE CO-INFECTION

AHMED MUHAMMAD AL-SAHI
Plant Protection Department - College of Agriculture - Tikrit University - Salah Al-Din – Iraq.

OSAMAH NADHIM ALISAWI
Plant Protection Department - Faculty of Agriculture - University of Kufa - Najaf – Iraq.

MAADH ABDULWAHAB AL-FAHAD
Plant Protection Department - College of Agriculture - Tikrit University - Salah Al-Din - Iraq

Abstract:

Lettuce (Lactuca sativa L.) is widely planted and considered a valuable crop in Iraq. Next generation technology (NGS) and bioinformatics analysis were used to detect endogenous Para retroviruses in the genome of lettuce. In this study, two virus elements were identified, belong to two genera, Caulimovirus and Florendovirus, which were named Caulimovirus-LSa, and LsatiV respectively. The full length of Caulimovirus-LSa, and LsatiV were 7827, 7205 bp respectively. The Caulimovirus-LSa encodes seven coding domains; movement protein (MP), Cauli-AT, Cauli-DNA-bind, Peptidase-A3, RT-LTR, RNase-H and RT-RH. The LsatiV has five domains of RNaseH (RH), RT-RH, reverse transcriptase (RT and RVT) and movement protein (MP). Interestingly the sequence of LsatiV was inverted from 3' to 5', unlike Caulimovirus-LSa. Transcripts per million values (TPM) were 4.3 and 4.45 for Caulimovirus-LSa, and LsatiV respectively, which considered low in expression. However, the resemblance of almost 60% between Peptidase, RT and RH domains of Dahlia mosaic virus (DMV) and Caulimovirus-LSa probably indicate co-infection alongside pathogenic virus or viroid agents that might be caused symptoms in lettuce. The phylogenetic tree confirms the close relationships of both viruses to endogenous viruses from Taraxacum officinale that belong interestingly to the same family Astraceae.

Keyword: Lettuce genomes, bioinformatics, endogenous pararetroviruses, next generation sequencing (NGS).

1. INTRODUCTION:

Lettuce (Lactuca sativa), belongs to the family Asteraceae, which is the largest family among plants, as it includes about 30,000 species. The crop is also an extremely valuable winter crop in Iraq and world due to its nutritional value (Gomes et al., 2014). The plant hosts many plant viruses like Mirafiori lettuce big vein virus (MiLBVV), Lettuce mosaic virus (LMV), Cucumber mosaic virus (CMV), Lettuce chlorosis virus (LCV) and Lettuce big vein-associated virus (LBVaV) (Hadad et al., 2019; Ertunç and Zelyut 2019). The viruses belong to different virus groups and transmit through seeds and insects to cause highly impact diseases with significant losses in lettuce fields (Dinant and Lot, 1992). Endogenous pararetroviruses (EPRVs) are increasingly discovered in most plant genomes due to sequencing techniques discovery. These viral elements belong to eight genera belonging to the family Caulimoviridae; Soymovirus, Cavemovirus, Solendovirus, Petuvirus, Caulimovirus, Tungrovirus, Rosadnavirus, and Badnavirus. In addition, a newly discovered genus called Florendovirus was recently characterized and existed with highly amounts in flowering plants. The DNA of these elements generally has a molecular weight of 7.2-9.2 kilobases (Geering et al., 2010: Geering and Hull

2012: Diop et al., 2018). Endogenous viruses have a double-stranded DNA that replicate through reverse transcriptase phase to turn from DNA into mRNA and then return to the form of DNA to integrate within the genome of the host. Most of these elements are strongly silenced; however, three EPRVs have the ability to break the dormancy phase and activate to be capable to start infection such as Banana streak virus and Tobacco vein clearing virus and Petunia vein clearing virus (Gayral et al., 2010; Geering et al., 2014; Alisawi 2019). Recently, new data have been confirmed the ability of further EPRVs to be expressed and probably share the infection alongside with pathogenic viruses (Khaffajah et al., 2022). The aim of this study was to identify endogenous pararetroviruses that have never been studied before in lettuce genomes and figure out a possible role in infection using next-generation sequencing (NGS) and bioinformatics tools.

2. PROCEDURE

2.1. Plant material and DNA sequencing

Two samples of slightly symptomatic lettuce leaves showing curling and yellowing were collected for DNA and RNA extractions. The samples were put in an Eppendorf tube and immersed in RNALatter solution, and sent to DNA-Link Company, Republic of Korea. Total DNA and RNA were extracted following the company's instructions. For DNA and RNA sequencing, TruSeq DNA Library prep kit and TruSeq whole RNA library prep kit were used for NGS library preparation in the company. Based on the manufacturer's instructions, DNA samples were examined using Novaseq6000 2x150bp reads technique and application WGS (PCR Free550). In order to obtain the total RNA sequence, the quality of the RNA sample was checked with an Agilent 2100 Expert Bioanalyzer, then sequenced with NovaSeq6000, 2x101PE.

2.2. Graph-based read clustering with Repeat-Explorer

RepeatExplorer pipeline was used to explore and characterize EPRV clusters and repetitive DNA sequences in last-generation sequencing data (Novák et al., 2013). Clustering by RepeatExplorer2 (Galaxy 2.3.8.1) was applied, and the selected taxon and protein domain database was viridplantae version 3.0. The generated clusters were then imported into the Repbase dataset (Jurka et al., 2005), and the Basic Local Alignment Search Tool (Altschul et al., 1990). In addition, these sequences were aligned to previously published plant virus sequences (DPVweb) (Adams and Antoniw 2005). Through sequence alignment, the suggested viruses were identified at the genus level.

2.3. Map to reference

Raw Illumina reads were aligned against the identified viruses, and a report showed the number of assembled reads, total used reads, and frequently overlapped reads. As a result of the data, we calculated copy numbers and genome proportions as follows: 1-Copy number: number of assembled reads x read length/reference sequence length. 2-Genome proportion: number of aligned reads / total NGS reads x 100 (Mustafa et al., 2018). As a measure of transcripts per million (TPM), the read count per kilobase of the

EPRV sequences was divided by a million to calculate the RNA analysis (Bester et al., 2021).

2.4. Phylogenetic analysis

The maximum likelihood (ML) method was applied to MEGA 11 (Tamura et al., 2013) to create a robust phylogeny model. The sequences were aligned using a Clustal W alignment (about 7000 bp for each). The tree reconstructed with General Time Reversible (GTR). The phylogeny tree of Caulimovirus group was constructed using 10 EPRVs, with Cacao swollen shoot Ghana virus as the outgroup. 17 EPRVs were applied to build the tree of Florendovirus group, and Cauliflower mosaic virus (CaMV) was the outgroup member.

3. RESULTS

The total number of clean DNA reads amounted to about 72.929.460 short reads with a length of 151 bp and the RNAseq reads were about 53.111.688 with a length of 101 bp. The RepeatExplorer pipeline revealed that DNA sequences of the genome of lettuce have two clusters of EPRVs. Both clusters extracted and mapped against the whole clean reads of DNA to obtain the full sequence of each EPRV (Fig 1).The analysis showed 20954 reads were assembled against Caulimovirus member, while 21120 reads were assembled against Florendovirus to form the full sequence of each EPRV with possible coded domains. Two virus elements were identified belong to two genera, Caulimovirus and Florendovirus, which were named Caulimovirus-LSa (https://www.girinst.org/2022/vol22/issue9/Caulimovirus-LSa.html), and LsatiV (https://www.girinst.org/2022/vol22/issue9/LsatiV.html), following Repbase dataset regulation and Geering et al. (2014) respectively. The full length of Caulimovirus-LSa, and LsatiV were 7827, 7205 bp respectively. The Caulimovirus-LSa encodes almost full set of protein domains with seven coding regions; movement protein (MP), Cauli-AT, Cauli-DNA-bind, Peptidase-A3, RT-LTR (reverse transcriptase), RNase-H and RT-RH (Fig 2). The LsatiV has five domains of RNaseH (RH), RT-RH, reverse transcriptase (RT and RVT) and movement protein (MP) (Fig 3). Interestingly, the sequence of LsatiV was inverted from 3' to 5', unlike Caulimovirus-LSa. Both sequences were found in RNA transcripts with 7634 reads assembled against Caulimovirus-LSa and 7287 reads mapped to LsatiV. The domains of Caulimovirus-LSa have expressed and covered the whole group of its proteins, while LsatiV showed expression for the RH, RT-RH and MP domains. Transcripts per million values (TPM) were 4.3 and 4.45 for Caulimovirus-LSa and LsatiV respectively. The Peptidase, RT and RH domains of Caulimovirus-LSa shared almost 60% similarity with same domains in Caulimovirus-T Of and Dahlia mosaic virus (DMV) as confirmed by dot plot analyses (Fig 4). The phylogenetic tree confirms the close relationships of both viruses to endogenous viruses from Taraxacum officinale that belong interestingly to the same family Astraceae (Fig 5).

Fig 1: The whole sequence of Caulimovirus-LSa (A) and LsatiV (B) that mapped against the lettuce DNA reads with full coverage.

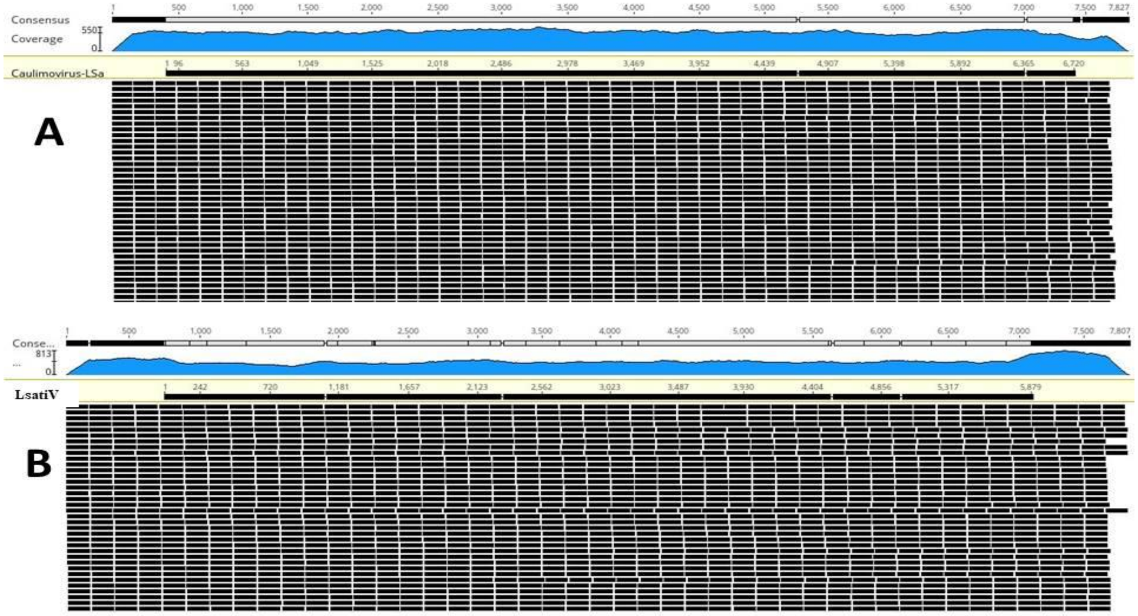


Fig 2: The whole sequence of Caulimovirus-LSa encodes seven protein domains and found in RNA transcripts of lettuce.

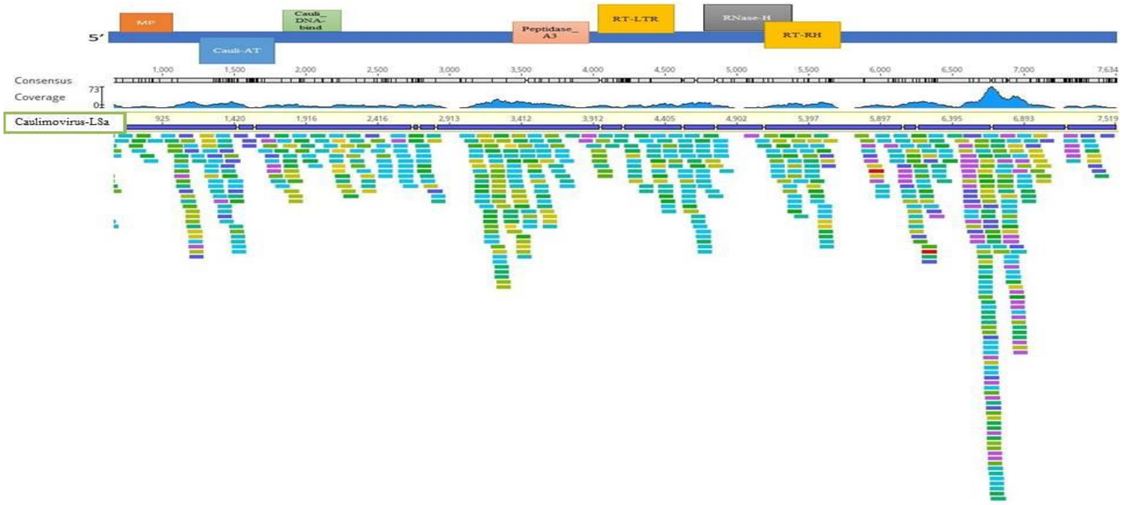


Fig 3: The whole sequence of LsatiV with five protein domains that partially found in RNA transcripts with low expression.

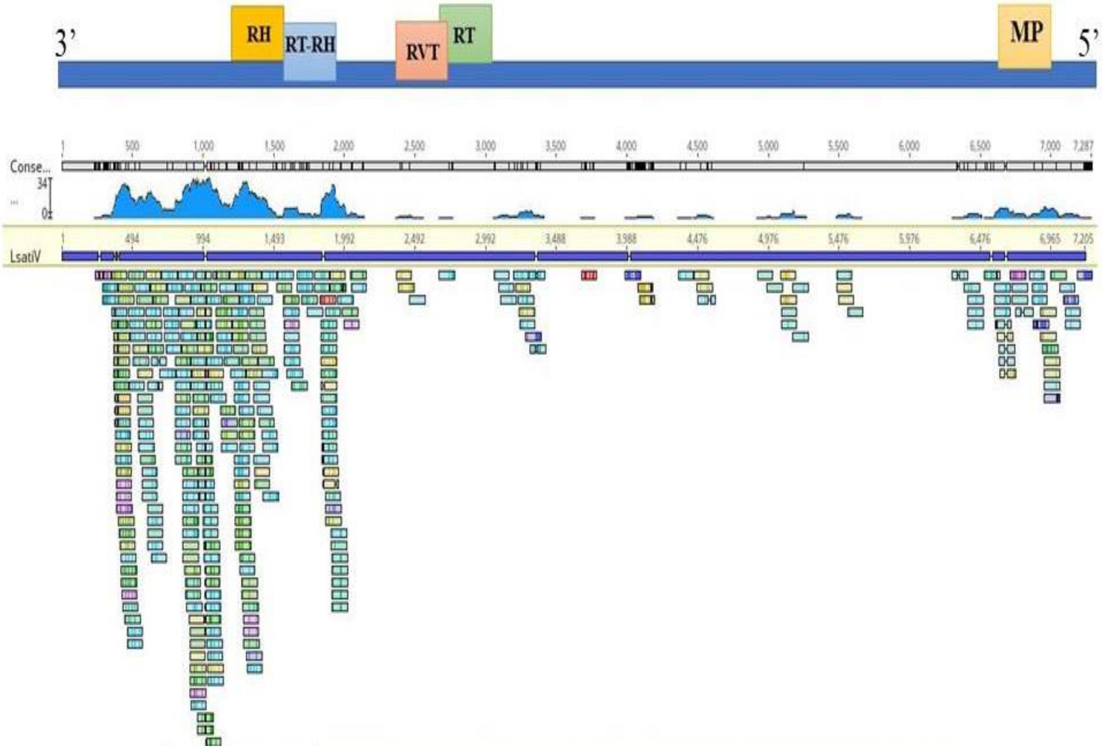


Fig 4: Dot plot analyses shows similarities of Peptidase, RT and RH domains of Caulimovirus-LSa and both Caulimovirus-TOf (Right) and Dahlia mosaic virus (Left).

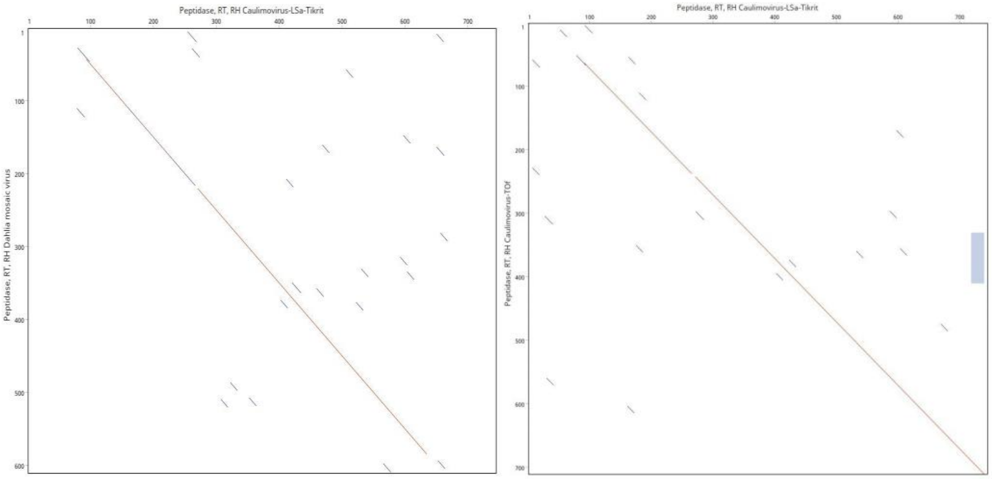
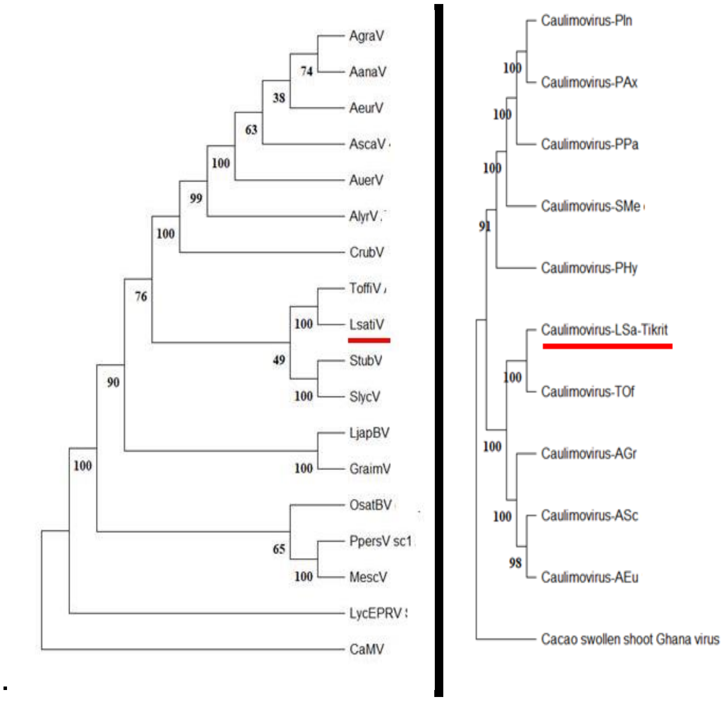


Fig 5: The neighbor-joining tree of both members of endogenous pararetroviruses reconstructed based on complete genome sequences of related EPRVs



4. DISCUSSION

The nuclear genome of the host contains endogenous pararetrovirus (EPRV) sequences that have a significant impact on the plant. With high throughput sequencing techniques, we investigated the nature, abundance and organization of EPRVs in lettuce genome using bioinformatic tools. There has been a lack of understanding of integrants in plants compared to animals, despite their presence in plant genomes since MYA 20-34 (Geering et al., 2014). In a way that is very similar to retrotransposons, endogenous pararetroviruses integrate into host genomes by hitchhiking on retrotransposons. This hybrid between the two is potentially capable of undergoing full transposition (Hohn 1994; Froissart et al., 2005; Richert-Pöggeler et al., 2003; Gregor et al., 2004, Staginnus et al., 2007). As a result of their integration into genomes, EPRVs have a huge impact on the host genome, including methylation status and chromosomal rearrangements (Hohn et al., 2008). In this study, Lactuca sativa has been added as a new host plant that involves the integrants of pararetroviruses within its genome. Further, we highlighted the genome arrangement and phylogeny of these viral elements. When taking into account the complete genome size of different EPRVs that clearly distinguished by their genome proportions, arrangement and copy numbers. Although, the two viral elements belong to different genera of Caulimoviridae, they displayed in the genome with consistent number of assembled reads so far, and shared nearly equal TPM values. However, in Caulimovirus-LSa, the Peptidase, RT and RH domains have shared almost 60% similarity with same domains of Dahlia mosaic virus (DMV) that existed in Dahlia genome (Astraceae). The endogenous copies of Dahlia mosaic virus (D10) were characterized earlier in the genome of Dahlia from different countries. In the Dahlia genome (Asteraceae), three sequences of Dahlia mosaic virus-D10 found as endogenous pararetroviruses in cultivated and wild Dahlia spp. in Mexico, New Zealand and Lithuania, suggesting that plant pararetroviruses probably emerged, co-existed, and evolved together (Eid et al., 2011). The three EPRVs possess the structure and organization typical of Caulimovirus species, showing high identity among many open reading frames (ORFs) at the nucleotide level (Almeyda et al., 2014). Studying repeatomes in synthetic and apomictic hybrids of Asteraceae, two pararetrovirus clusters were detected in apomicts with a higher deviation than in synthetic hybrids. This suggests an overabundance of pararetroviruses in the apomicts as the only detectable difference between examined hybrids (Zagorski et al., 2020). It is interesting to note that despite the low level of expression in the analyzed Caulimovirus-LSa, it may be enough to initiate infection or to cause co-infection with pathogenic viruses in the leaves that showed mild symptoms. The hypothesis we propose is not similar to that of the three known viruses (Petunia vein clearing virus, Tobacco vein clearing virus, and Banana streak virus), which have completed genome transcription to initiate infection, but may be related (Jakowitsch et al., 1999; Ndowora et al., 1999; Richet-Pöggeler et al., 2003). Eid and Pappu (2013) reported that the DNA amplicons produced by direct PCR from sap extracts were more intense for Dahlia mosaic virus and Dahlia common mosaic virus which are known as episomal caulimoviruses. Comparatively, the DvEPRS (the endogenous copy) amplicon was significantly less intense than that of the amplicon of dahlias, which has an internal transcribed spacer region. Recently, two EPRVs have been found in Taraxacum officinale genome, called Caulimovirus-TOf and ToffiV (Alisawi et al., 2022) ,and interestingly, the phylogeny analyses showed high similarity between Caulimovirus and Florendovirus members of Taraxacum and Lactuca genomes that belong to the same family. Further, the genome organization of florendoviruses in Taraxacum and Lactuca genomes was inverted; indicating similar approach of integration might be occurred for both viral elements. The results confirmed the importance of the element for showing how genomes from the same family participate in such elements as EPRVs.

5. CONCLUSION

The complete sequences of two endogenous pararetroviruses were found in whole genomic DNA and RNA reads. The genome organization and copy numbers of each integrant have been analyzed, and phylogenetically studied. The viral element was present in RNA transcripts with a low expression rate. However, the high similarity of particular domains of Caulimovirus-LSa and Dahlia mosaic virus suggests a possible role for the integrant in pathogenic infection.

REFERENCES

1) Adams M, Antoniw J, Fauquet C. (2005): Molecular criteria for genus and species discrimination within the family Potyviridae. Arch. Virol, 150(3), 459-479.

2) Altschul S, Gish W, Miller W, Myers E, Lipman D (1990): Basic local alignment search tool. J. Mol. Biol, 215(3), 403-410.

3) Alisawi, O. N. (2019): Virus integration and tandem repeats in the genomes of Petunia. Doctoral dissertation. University of Leicester. UK.

4) Alisawi, O. N., Salih, R. H. M., & Heslop-Harrison, P. (2022): Variable ratios of two endogenous pararetroviruses in the genome of Taraxacum. Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition), 41, 270-278.

5) Almeyda, C. V., Eid, S. G., Saar, D., Samuitiene, M., & Pappu, H. R. (2014): Comparative analysis of endogenous plant pararetroviruses in cultivated and wild Dahlia spp. Virus genes, 48(1), 140-152.

6) Bester, R., Cook, G., Breytenbach, J. H., Steyn, C., De Bruyn, R., & Maree, H. J. (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. Virology Journal, 18(1), 1-19.

7) Dinant, S. and Lot, H. (1992): Lettuce Mosaic Virus: a review. Plant Pathol. 41, 528–542.

8) Diop, S.I.; A.D. Geering; F.Alfama-Depauw; M. Loaec; Teycheney P.-Y.and Maumus F. (2018): Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae. Scientific Reports, (8): 572-80.

9) Eid, S., Saar, D. E., Druffel, K. L., & Pappu, H. R. (2011): Plant pararetroviral sequences in wild Dahlia species in their natural habitats in Mexican mountain ranges. Plant pathology, 60(2), 378-383.

10) Eid, S., and H. R. Pappu (2014): Expression of endogenous para-retroviral genes and molecular analysis of the integration events in its plant host Dahlia variabilis." Virus genes, 48.1: 153-159.

11) Ertunç, F., & Zelyut, F. R. (2019): Virus diseases of lettuce in Ankara province. International Journal of Agriculture Forestry and Life Sciences, 3(2), 202-206.

12) Froissart R., Roze D., Uzest M., Galibert L., Blanc S. & Michalakis Y. (2005): Recombination every day: abundant recombination in a virus during a single multicellular host infection. PLoS Biology, 3, 389-95.

13) Gayral, P. et al. (2010): Evolution of endogenous sequences of Banana streak virus: what can we learn from banana (Musa sp.) Evolution. J. Virol. 84, 7346–7359.

14) Geering, A. D. W., & Hull, R. (2012): Genus Badnavirus in virus taxonomy classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. MQ King, MJ Adams, EB Carstens, and EJ Lefkowitz, eds. Elsevier Academic Press, San Diego, CA, 438-440.

15) Geering A.D.;T.Scharaschkin and P.Y.Teycheney (2010): The classificatio and nomenclature of endogenous viruses of the family Caulimoviridae. Archives of Virology, 155: 123-31.

16) Geering, A. D., Maumus, F., Copetti, D., Choisne, N., Zwickl, D. J., Zytnicki, M. ... & Teycheney, P. Y. (2014): Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nature Communications, 5(1), 1-11.

17) Gregor W., Mette M.F., Staginnus C., Matzke M.A. & Matzke A.J. (2004): A distinct endogenous pararetrovirus family in Nicotiana tomentosiformis, a diploid progenitor of polyploid tobacco. Plant Physiology, 134, 1191-9.

18) Gomes, D.P.; D.F. de Carvalho W.S. de Almeida, L.O.Medici and J.G.M.Guerra. (2014): Organic carrot-lettuce intercropping using mulch and different irrigation levels. Journal of food, Agriculture & Environment, 12(1): 323-328.

19) Hadad, L., Luria, N., Smith, E., Sela, N., Lachman, O., & Dombrovsky, A. (2019): Lettuce chlorosis virus disease: a new threat to cannabis production. Viruses, 11(9), 802.

20) Hohn T. (1994): Recombination of a plant pararetrovirus: Cauliflower mosaic virus. In: Homologous Recombination and Gene Silencing in Plants (pp. 25-38). Springer.

21) Hohn T., Richert-Pöggeler K.R., Staginnus C., Harper G., Schwarzacher T., Teo C.H., Teycheney P.-Y., Iskra-Caruana M.-L. & Hull R. (2008): Evolution of integrated plant viruses. In: Plant Virus Evolution (pp. 53-81). Springer.

22) Jakowitsch, J., Mette, M.F., van der Winden, J., Matzke, M.A. and Matzke, A.J. (1999): Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc. Natl. Acad. Sci. USA, 96, 13241–13246.

23) Khaffajah, B., Alisawi, O., & Al Fadhl, F. (2022): Genome sequencing of eggplant reveals Eggplant mild leaf mottle virus existence with associated two endogenous viruses in diseased eggplant in Iraq. Archives of Phytopathology and Plant Protection, 1-14.

24) Mustafa S (2018): Mitochondrial and repetitive DNA defining the sheep genome landscape (Doctoral dissertation, University of Leicester).

25) Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E. and Lockhart, B. (1999): Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. Virology, 255, 214–220.

26) Novák, P.; P. Neumann; J .Pech; J. Steinhaisl and Macas J. (2013): RepeatExplorer: a Galaxybased web server for genome-wide characterization of eukaryotic repetitive elements from next- generation sequence reads. Bioinformatics, (29): 792-3.

27) Richert-Pöggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G. and Hohn, T. (2003): Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. EMBO J. 22, 4836–4845.

28) Staginnus C., Gregor W., Mette M.F., Teo C.H., Borroto-Fernandez E.G., Machado M.L.d.C., Matzke M. & Schwarzacher T. (2007): Endogenous pararetroviral sequences in tomato (Solanum lycopersicum) and related species.(Research article)(Clinical report). BMC Plant Biology, 7, 24-40.

29) Tamura, K.; G. Stecher; D. Peterson; A. Filipski and Kumar, S. (2013): MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution, 30: 2725-2729.

30) Zagorski, D., Hartmann, M., Bertrand, Y. J., Paštová, L., Slavíková, R., Josefiová, J., & Fehrer, J. (2020): Characterization and dynamics of repeatomes in closely related species of Hieracium (Asteraceae) and their synthetic and apomictic hybrids. Frontiers in plant science, 11, 591053.