RNA-Seq data analysis reveals various viral sequences associated with genome of date palm (*Phoenix dactylifera*)

Adnan A. Lahuf

Abstract

RNA-Seq data analysis can provide an unlimited contribution in increasing our knowledge in viruses and their interaction with the hosts. In current study, transcriptomic analyses of 20 publicly libraries of date palm (*P. dactylifera*) fruit was conducted for identification the associated plant viruses. Although, the group of viruses *Grapevine leafroll-associated virus* (GLRaV) and *Pittosporum cryptic virus-1* (PiCV-1) have been identified previously, another thirty-seven different plant viruses have been discovered through the detection of partial nucleotide sequences of their RdRp and Replicase genes via a workflow analysis developed in this study. The viral sequences detected were almost identical (100%) at the nucleotide level with numerous viral isolates deposited at GenBank. However, very rare viral infection has been reported previously on *P. dactylifera* trees. Thus, an investigation for determining virome of the date palm tree is demanded.

1. Introduction

Olive is an evergreen wooden Date palm (*Phoenix dactylifera* L.) tree is one of the most important and oldest fruit trees spreading in the tropic and subtropical regions particularly in the Middle East and North Africa [1,2]. Historically, it has been believed to be originated in the Mesopotamia area, which is at present south of Iraq [3,4]. Production of date palm in Iraq was 614584 tons in 2018, which accounts for 7.07% of global production [5]. However, the palm tree suffers from numerous and various pests such as the lesser date moth (*Batrachedra amydraula*), Dubas Bug (*Ommatissus binotatus*) and inflorescence rot, caused by *Mauginiella scaetae*, *Fusarium moniliforme* and *Thielaviopsis paradoxa* that produce extensive damage leading to low productivity and poor quality of the fruits causing great economic losses [6,7]. On the other hand, viral infection has not reported yet on date palm trees in Iraq or worldwide [8].

Numerous biological, serological, and molecular tests have been developed and utilized successfully for the detection and characterization of viruses infecting plants including fruit trees. However, the bioassays have less feasible applications due to relatively high consumption of time and effort in addition to problems related to reliability and individual assessment [9]. Furthermore, despite the serological approaches, mainly the enzyme-linked immunosorbent assay (ELISA) and its adjustments, empower economy and rapid parallel identification of several viruses, they are limited in sensitivity besides scarce immunogenicity of some viruses in producing antibodies which are lack dependability for detection of those particular viruses [10,11]. Consequently, several specific and sensitive molecular methods such as Polymerase chain reaction, Reverse transcription polymerase chain reaction, Real-time Polymerase chain reaction, Loop-mediated isothermal amplification etc. have been applied for detection the majority of plant viruses. They also were found to be more sensitive than ELISA and employed successfully in simultaneously multiplex detection of several viruses [12,13] However, these methods are very specific and time-consuming as well as they are strongly relying on availability of the sequence databases of pathogens including viruses for designing amplification primers. Additionally, the available sequence data of poorly analyzed viruses could lead to insufficient diagnosis by providing false either positive or negative results [9,14]. These obstacles have highlighted the demanding for developing new molecular approaches. One of these modern approaches is high-throughput sequencing (HTS) or also known as next generation sequencing (NGS) technology. This technology is able to produce high-throughput sequence data without the requirement of any previous knowledge regarding of pathogens genome data [15]. Moreover, it demands relatively less time for the accomplishment of a sequence run and in silico analysis of the obtained data comparing with others molecular methods. Nowadays, this
technology has facilitated the identification of enormous known and new viruses in various infected hosts [16]. However, previous NGS studies of different date palm tissues have not dealt with viral populations that could be associated with these tissues. Thus, the objective of this study was to reveal viruses in fruit tissues of date palm through developing a particular bioinformatics workflow.

2. Materials and Methods

2.1. Plant materials and library preparation

To determine possibility of association viruses with date palm tissues, different gene expression libraries of date palm fruits at different development stages were selected. Detailed information regarding these plant tissues and libraries preparation can be found in the previous study [17]. In brief, the study had two experiments. In the first experiment, fruits of several date palm trees from the Khenezi and Khalas cultivars were sampled at five various growth stages (45, 75, 105, 120, and 135 days post-flowering) in 2014. Extraction of RNA was conducted from fruits of each cultivar and libraries were made according to the manufacturer’s instruction of TruSeq library preparation Kit. The sequencing was accomplished via an Illumina HiSeq 2500 sequencer producing paired end (2 x 101 bp) reads. In the second experiment, the fruits were sampled on khalal stage from many trees of different cultivars in 2016. The RNA was extracted and the libraries were made using Nextera library preparation kit. The paired-end reads (2 x 76 bp) were collected through a NextSeq (Illumina) sequencer.

2.2. Processing of raw data and de novo assembly

A 20 raw Illumina sequencing reads available at Sequence Read Archive (SRA) database under the project accession number PRJNA505138 were downloaded. All RNA-seq analyses were executed using Samtools software in the Ubuntu 18.04 operation system. Trimmmomatic software was applied for trimming the raw paired-end reads of each library using the following parameters: ILLUMINACLIP: 2:30:10, TRAILING and LEADING: 3, SLIDINGWINDOW: 4:20, MINLEN: 20. The reads produced were then de novo assembled utilizing two different programs, Trinity and VelvetOptimiser following their default parameters.

2.3. Mapping and viral contigs identification

In order to identify, the viruses associated with date palm, the de novo assembled contigs were mapped using Bowtie2 software to the complete reference sequence of date palm obtained from Date Palm Genomic Resource Database (DRDB) available at http://drdb.big.ac.cn/home [18]. The unmapped contigas were extracted and converted to FASTA files via Samtools (v.1.7) programs. Subsequently, the NCBI BLAST+ blastn software was utilized with the algorithm of MEGABLAST and cut-off E-value 1e-5 for identification of the viral sequences in the extracted FASTA files. The MEGABLAST search was implemented against complete reference sequences of plant viruses and viroids downloaded from http://www.ncbi.nlm.nih.gov/genome/viruses. Based on calculating the number of RNAs contigs that aligned to the reference sequences of numerous plant virus were accomplished for determining the scaffolds, coverage, and average depth parameters.

3. Results and Discussion

3.1. Identification of different viruses from date palm mRNA transcriptome

A comprehensive investigation was accomplished for determining viral population associated with fruits of date palm trees through transcriptome analysis. Publicly available transcriptome data (accession number PRJNA505138) gathered from various stages of fruit growth of the date palm (P. dactylifera) cultivars Khenezi, Khalas, and others were examined for identification the viruses and viroids associated. The transcriptome analysis was comprised 20 libraries selected randomly from total 55 libraries and reanalyzed according to the workflow displayed in Figure (1). Although, previous transcriptome analysis [17] discovered the viruses group Grapevine leafroll-associated virus and Pittosporum cryptic virus-1, a total of 985 assembled contigs were obtained that belonged to additional thirty-seven plant viruses (Figure 2). However, numerous contigs associated with two viroids Citrus exocortis viroid and Citrus exocortis Yucatan viroid were identified. Their sequences deposited have been erroneously annotated leading to false identification [19]. Thus, they were neglected. Of thirty-seven recognized viruses associating with date palm fruits, Alfalfa mosaic virus was the dominant virus reached 17.40256400% of total viral contigs followed by Pepper chlorotic spot virus and Plant associated genomovirus with 8.66256400 and 7.55256400% respectively (Table 1).
Worth to mention that the *Alfalfa mosaic virus* was also the most common virus that was detected in 19 libraries followed by *Papaya mosaic virus*, *Banana bract mosaic virus*, *Turnip mosaic virus* and *Tobacco virus 2* found in 18 libraries. On the other hand, *Garlic virus A*, *Pittosporum cryptic virus* and *Bell pepper endornavirus* were identified in 17 libraries (Table 1). Additionally, the analysis revealed that the plant viruses identified belonged to 17 different families. The most common families were Betaflexiviridae and Caulimoviridae containing almost half of viruses detected.

---

**Figure 1.** The schematic workflow of the transcriptome analysis implemented for plant virus identification in date palm fruits.

**Figure 2.** The contigs percentages of plant viruses identified in transcriptome data of date palm fruits.
Table 1. Percentages of plant viruses contigs identified in libraries of date palm fruits transcriptome.

| No. | The virus identified                        | No. of Libraries | Contigs Percentage |
|-----|--------------------------------------------|------------------|-------------------|
| 1.  | Alfalfa mosaic virus                       | 19               | 17.40256400%      |
| 2.  | Pepper chlorotic spot virus                | 4                | 8.66256400%       |
| 3.  | Plant associated genomovirus               | 10               | 7.55256400%       |
| 4.  | Garlic virus A                             | 17               | 5.11256400%       |
| 5.  | Pittosporum cryptic virus                  | 17               | 4.91256400%       |
| 6.  | Broad bean wilt virus                      | 2                | 4.81256400%       |
| 7.  | Papaya mosaic virus                        | 18               | 4.81256400%       |
| 8.  | Cherry twisted leaf associated virus        | 5                | 3.96256400%       |
| 9.  | Bell pepper endornavirus                   | 17               | 3.39256400%       |
| 10. | Grapevine leafroll-associated virus        | 3                | 3.18256400%       |
| 11. | Maize-associated picornavirus              | 3                | 2.78256400%       |
| 12. | Banana bract mosaic virus                  | 18               | 2.37256400%       |
| 13. | Turnip mosaic virus                        | 18               | 2.37256400%       |
| 14. | Tobacco virus 2                            | 18               | 2.37256400%       |
| 15. | Cherry virus B                             | 7                | 1.66256400%       |
| 16. | Phlox virus S                              | 4                | 1.66256400%       |
| 17. | Cherry virus A                             | 6                | 1.56256400%       |
| 18. | Cherry virus Trakiya                       | 2                | 1.56256400%       |
| 19. | Apple stem pitting virus                   | 6                | 1.46256400%       |
| 20. | Banana bract mosaic virus                  | 15               | 1.36256400%       |
| 21. | Sweet potato chlorotic fleck virus         | 4                | 1.36256400%       |
| 22. | Ullucus tymovirus 1                        | 7                | 1.26256400%       |
| 23. | Banana streak UA virus                     | 3                | 1.15256400%       |
| 24. | Apricot latent virus                       | 3                | 1.05256400%       |
| 25. | Citrus tristeza virus                      | 5                | 1.05256400%       |
| 26. | Cherry rusty mottle associated virus        | 2                | 0.95256400%       |
| 27. | Grapevine vein clearing virus              | 1                | 0.95256400%       |
| 28. | Peach chlorotic mottle virus               | 2                | 0.95256400%       |
| 29. | Piper yellow mottle virus                  | 3                | 0.95256400%       |
| 30. | Apple chlorotic leaf spot virus            | 3                | 0.85256400%       |
| 31. | Cherry necrotic mottle virus               | 2                | 0.85256400%       |
| 32. | Potato virus S                             | 1                | 0.85256400%       |
| 33. | Banana streak virus                        | 1                | 0.75256400%       |
| 34. | Grapevine roditis leaf discoloration-associated virus | 1 | 0.75256400% |
| 35. | Asian prunus virus 2                       | 1                | 0.65256400%       |
| 36. | Banana streak OL virus                     | 1                | 0.65256400%       |
| 37. | Banana streak VN virus                     | 1                | 0.65256400%       |
| 38. | Citrus leaf blotch virus                   | 1                | 0.65256400%       |
| 39. | Cowpea mild mottle virus                   | 1                | 0.65256400%       |

3.2. De novo assembly of viral reads associated with the transcriptomic data and their quantity

The de novo assembled contigs of viruses obtained from total and each library were assessed. The contig percentage of viruses identified in all libraries selected was 0.0000256103% while in each library was diverse, ranging from 0.0000294% to 0.000988% (Figure 3). Among the 20 libraries analyzed, the highest contigs quantity of viruses was found in the SRR10122010 library (0.000988%) followed by the SRR10121956 and SRR10121959 (0.000927% and 0.000330% respectively). On the other hand, the lowest percentages were discovered in the SRR10121957 and SRR10122006 libraries (0.0000395% and 0.0000294% respectively). The number of viruses in each library appeared to be not correlated with the size of the library. Thus, the viruses number calculated was diverse in each individual library. For instance, the highest number of viruses identified was found in the SRR10121959 and SRR10121969 libraries, which comprised 21 and 22 various viruses respectively. Conversely, the lowest
numbers of viruses were detected in libraries of SRR10122006 (4 viruses) and SRR10121957 (6 viruses; Figure 4). Interestingly, the Alfalfa mosaic virus, Garlic virus A, Papaya mosaic virus, Bell pepper endornavirus, Banana bract mosaic virus, Turnip mosaic virus, and Tobacco virus 2, that were commonly identified in most of the libraries (17-19 libraries), their contigs were among the highest percentages discovered. Furthermore, the RNA reads of the groups of viruses Grapevine leafroll-associated virus and Pittosporum cryptic virus formerly reported [17] were relatively abundant in the libraries analysed. Moreover, the analysed data showed that the viruses percentages identified in each library were various and Alfalfa mosaic virus was the highest in most libraries.

Figure 3. The contigs percentages of viruses identified in their libraries.

Figure 4. Number of viruses in each library analysed.
extracted the raw reads and calculated the contigs assembled of several viruses from the date palm transcriptomes. Furthermore, the analysis revealed the possible correlation among these viruses. For instance, there was no predominant single virus in any library. On the contrary, various viruses belonging to different families were found and this refers to a complex infection that can lead to a competition among them [21] resulting decrease their density and consequence influence. However, it is believed that the interaction between communities and populations of viruses, as well as the environmental changes are possibly going to control them. Additionally, this analysis details an effective application of in silico workflow on plant transcriptome data to determining a list of possible viruses and viroids that infect date palm cultivar is advantageous to select an effective method for regulate the transcriptional machinery of the host and this may lead to cause severe disease symptoms in the upcoming future.

Table 2. Percentage of viruses identified in each library.

| viruses                                      | SRR101 21970 | SRR101 21972 | SRR101 22006 | SRR101 22007 | SRR101 22008 | SRR101 22009 | SRR101 22010 |
|----------------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Alfalfa mosaic virus                          | 9.67         | 19.35        | 0            | 21.05        | 16.66        | 32.55        | 29.85        |
| Apple stem pitting virus                      | 1.61         | 0            | 0            | 0            | 0            | 0            | 0            |
| Apricot latent virus                          |              |              |              |              |              |              |              |
| Asian prunus virus 2                          | 0            | 3.22         | 0            | 0            | 0            | 0            | 0            |
| Banana bract mosaic virus                     | 3.22         | 6.45         | 0            | 5.26         | 2.77         | 2.32         | 2.98         |
| Bell pepper endornavirus                      | 1.61         | 3.22         | 0            | 5.26         | 5.55         | 4.65         | 2.98         |
| Cherry twisted leaf-associated virus           | 0            | 3.22         | 0            | 0            | 0            | 0            | 0            |
| Cherry necrotic rusty mottle virus            | 0            | 0            | 0            | 5.26         | 5.55         | 0            | 0            |
| Cherry virus A                                | 3.22         | 0            | 0            | 0            | 0            | 0            | 0            |
| Cherry virus B                                | 3.22         | 0            | 0            | 0            | 0            | 0            | 0            |
| Cherry virus Trakiya                          | 0            | 0            | 0            | 0            | 0            | 4.65         | 5.97         |
| Citrus tristeza virus                         | 0            | 0            | 0            | 2.63         | 0            | 2.32         | 0            |
| Cowpea mild mottle virus                      | 0            | 0            | 20           | 0            | 0            | 0            | 0            |
| Garlic virus A                                | 3.22         | 3.22         | 0            | 5.26         | 8.33         | 9.30         | 10.44        |
| Grapevine leafroll-associated virus           | 0            | 0            | 0            | 0            | 0            | 4.65         | 0            |
| Maize-associated picornavirus                 | 0            | 0            | 0            | 0            | 0            | 9.30         | 11.94        |
| Papaya mosaic virus                           | 8.06         | 9.67         | 0            | 5.26         | 8.33         | 9.30         | 0            |
| Peach chlorotic mottle virus                  | 0            | 0            | 0            | 5.26         | 5.55         | 0            | 0            |
| Pittosporum cryptic virus                     | 3.22         | 0            | 0            | 10.52        | 11.11        | 4.65         | 2.98         |
| Plant associated genomovirus                  | 19.35        | 25.80        | 0            | 0            | 0            | 0            | 0            |
| Phlox virus S                                 | 0            | 0            | 40           | 0            | 0            | 0            | 0            |
| Sweet potato chlorotic fleck virus            | 0            | 0            | 20           | 0            | 0            | 0            | 0            |
| Tobacco virus 2                               | 37.09        | 6.45         | 0            | 10.52        | 5.55         | 2.32         | 8.95         |
| Turnip mosaic virus                           | 4.83         | 9.67         | 0            | 5.26         | 8.33         | 9.30         | 10.44        |
| Ullucus tymovirus 1                           | 1.61         | 9.67         | 0            | 7.89         | 11.11        | 4.65         | 13.43        |
| Yam mosaic virus Y                            | 0            | 0            | 0            | 5.26         | 5.55         | 0            | 0            |

In the present study, RNA-seq analyses of available transcriptome data of date palm fruits were accomplished. Revising previous literature, this analysis with previous study [20] are probably the first investigations of viruses and viroids associated with date palm trees. This study also in comparison with other molecular and conventional approaches [9]. (Jo et al.,2015 was efficiently extracted the raw reads and calculated the contigs assembled of several viruses from the date palm transcriptomes. Furthermore, the analysis revealed the possible correlation among these viruses. For instance, there was no predominant single virus in any library. On the contrary, various viruses belonging to different families were found and this refers to a complex infection that can lead to a competition among them [21] resulting decrease their density and consequence influence. However, it is believed that the interaction between communities and populations of viruses, as well as the environmental changes are possibly going to regulate the transcriptional machinery of the host and this may lead to cause severe disease symptoms in the upcoming future. Determining a list of possible viruses and viroids that infect date palm cultivar is advantageous to select an effective method for control them. Additionally, this analysis details an effective application of in silico workflow on plant transcriptome data to unveil the viral communities in date palm trees. Indubitably, forthcoming studies for assessing the virome of different tissues and cultivars of date palm trees benefiting the NGS technique are required for confirmation these outcomes.

Conclusions
The present analysis showed that the date palm (P. dactylifera) is more likely being infected with different plant viruses by detecting partial sequences of thirty-seven diverse viruses through a pipeline developed in this study. Hence, an examination for determine virus population of the date palm tree is necessitated.

Acknowledgements
Author would like to express their high gratitude to the department of plant protection, Agriculture College, University of Kerbala for supplying the facilities employed in this study.
References

[1] Al-Sweedi, T.M.M. (2016). Study the effect of some environmental factorial and chemical pesticides on different stages of Dubas bugs Ommatiusis binotatus lybicus Deberg. on two categories of the Date palm Phoenix dactylifera L. under field conditions. Journal of Kerbala for Agricultural Sciences 3 (3):1-19.

[2] Aubied, I.A. and Hamzah, H.A. (2019). Effect of Pollen Grains and Growth Regulator NAA on Some Fruit Characterization of Date Palm Phoenix Dactylifera L. cultivar Sultani. Al-Qadisiyah Journal for Agriculture Sciences 9 (1):136-142.

[3] Chao C. T. and Krueger R. R. (2007). The date palm (Phoenix dactylifera L.): overview of biology, uses, and cultivation. HortScience 42 (5): 1077-1082.

[4] Salman, A. H.; Hassan, W. F.; Mohamed, I. K.; Moter, H.A. (2017). The effect of chemical fertilization time on NPK pannaes contents and yield of new date palm breeding for three cultivars. Al-Qadisiyah Journal for Agriculture Sciences 7(1):41-48.

[5] Dreebe, H. A. & Abdul Razak, N.A. (2020). The Impact of Corruption on Agriculture Sector in Iraq: Econometrics Approach. IOP Conference Series: Earth and Environmental Science. 553 (1).

[6] Barbandi, A. (2000). Date palm technologies and prospects, Arab Center for Studies of Arid Zones and Dry Lands (ACSAD), Damascus, Syria.

[7] Al-Awsi A. I., Al-Saad I. H. and Al-Hadithy A. A. (2008). Training needs for Hait date palms farmers related to the identification and control of certain date palm pests. Anbar Journal of Agricultural Sciences 6 (1):1-10.

[8] Jassim S. A. A. and Limoges R. G. (2014). Date palm tree’s defense mechanisms from viral infection and solar ultraviolet radiation. Advances in Microbiology 4: 1-5.

[9] Mohammed, M.A. (2020). Structural, Optical, Electrical and Gas Sensor Properties of ZrO2 Thin Films prepared by Sol-Gel Technique. Neuroquantology, 18(3), 22-27. doi: 10.14704/nq.2020.18.3.nq20146

[10] Hazzouri, K.M.; Gros-Balthazard, M.; Flowers, J.M., et al. (2019). Genome-wide association mapping of date palm fruit traits. Nature Communications 10:716-725. doi: 10.1038/s41467-019-10731-8.

[11] He Z., Zhang C., Liu W., Lin Q., Wei T., Aljohi H.A., Chen W.H. and Hu S. (2017). DRDB: An Online Date Palm Genomic Resource Database. Frontiers Plant Science 8:1889. doi: 10.3389/fpls. 2017.01889.

[12] Di Serio F., Flores R., Verhoeven J. Th. J., Li S. F., Pallas V., Randles J. W., Sano T., Vidalakis G. and Owens R. A. (2014). Current status of viroid taxonomy. Archive Virology 159:3467–3478. DOI 10.1007/s00705-014-2200-6.

[13] Maachi, A.; Nagata, T. and Silva, J.M.F. (2020). Date palm virus A: first plant virus found in date palm trees. Virus Genes 56:792-795. https://doi.org/10.1007/s11262-020-01801-0.

[14] Roossinck, M.J. (2005). Symbiosis versus competition in plant virus evolution. Nature Review Microbiology 3(12):917-924. doi: 10.1038/nrmicro1285. PMID: 16322740.