A New Procedure to Identify Plant RNA Viruses Associated with the Whitefly (*Bemisia tabaci*) Using Next-Generation Sequencing

Mohammed D. Salman and Adnan A. Lahuf*
Department of Plant protection, College of Agriculture, University of Kerbala, Kerbala, Iraq.
*Email of corresponding author: adnan.lahuf@uokerbala.edu.iq

Abstract

Salman, M.D. and A.A. Lahuf. 2022. A New Procedure to Identify Plant RNA Viruses Associated with the Whitefly (*Bemisia tabaci*) Using Next-Generation Sequencing. Arab Journal of Plant Protection, 40(2): 169-174. https://doi.org/10.22268/AJPP-040.2.169174

Next-generation sequencing (NGS) was applied to obtain the transcriptome data of the Iraqi whitefly (*Bemisia tabaci*) that was analyzed using some specific bioinformatic tools and programs to identify the accompanying plant RNA viruses. Seven different plant viruses were detected: *Alfalfa mosaic virus* (AMV), *Pittosporum cryptic virus-1* (PiCV1), *Grapevine leafroll-associated virus* (GLRaV), *Broad bean wilt virus* (BBWV), and *Zantedeschia mild mosaic virus* (ZaMMV), in addition to Tomato spotted wilt virus (TSWV). The highest quantity of viruses identified were AMV, BBWV, and ZaMMV, respectively through achievement of the highest viral sequence coverages. In this study, we report a new NGS-based procedure, which facilitates prompt and precise the identification and quantification of plant viruses in a pool of *B. tabaci* insect samples without the necessity for specific primers and application of the conventional PCR technique. The benefit of this method is the quick detection of the potential viruses transmitted by the whitefly, *B. tabaci* vector that affects countless plant hosts. However, Additional examinations are required to confirm these findings.

Keywords: Next generation sequencing, plant RNA viruses; whitefly vector; transcriptome data.

Introduction

The whitefly *Bemisia tabaci* (Gennadius) is an ambiguous species because of possessing nearly 39 morphologically very similar biotypes (Alemandri *et al.*,2015). The two most common and harmful biotypes are biotype B (Middle East-Asia Minor 1) and biotype Q (Mediterranean) (De Barro *et al.*, 2011). These two dangerous biotypes and others cause critical destruction of their plant hosts through phloem sap-sucking and excretion of honeydew resulting in fungal contamination of fruits and leaves surfaces that leads to plant tissues damage. Furthermore, they and other members of *B. tabaci* can transmit 200 plant viruses belonging to various genera such as *Begomovirus*, *Torradovirus*, *Ipomovirus*, *Carlavirus* and *Crinivirus* (Adnan *et al*., 2017; Jones, 2003, Polston *et al*., 2014). Seasonally, some of these viruses become epidemic as a result of different reasons including the high population density of the biotypes B and Q and this cause destructive effect to numerous economic crops (Gilbertson *et al*., 2015; Islam *et al.*, 2017; Legg *et al.*, 2011).

Next-generation sequencing (NGS) offers an effective methodology for detection and diagnosis of viral sequences in different hosts (Belliure *et al*., 2005). At present, analysis of the transcriptome and other type of NGS data to identify known or new viruses becomes a commonly used procedure to reveal viral sequences and assemble them to construct their complete-length genomes (Barba *et al*., 2014). Additionally, numerous viruses can be detected or discovered via analysis of the insect transcriptome data, which means these insects are possibly harboring or being co-infected with these viruses, but without producing any disease symptoms (Luan *et al*., 2014). Since there are no previous studies to determine of the virome of *B. tabaci* using the new molecular techniques, an attempt was conducted to develop a NGS workflow through exploring the transcriptome data of the Iraqi whitefly samples in order to identify and quantify the associated plant viruses.

Material and methods

Insect collection, RNA extraction and Illumina sequencing

In this study, different populations of whiteflies were collected and identified based on anatomical and morphological features by an insect taxonomist. One population was collected from eggplant fields near Ayn al-Tamr District, Karbala Province, Iraq. The insect populations were immediately placed in DNA/RNA Shield™ (Zymo Research, USA), frozen at -20℃ and transported to the BGI Hongkong, China (Dou *et al.*, 2021). Based on the reports provided by BGI company, the total RNA was extracted from whiteflies samples utilizing Agilent RNA 6000 nano reagents part 1 (Agilent Technologies, Germany) based on the manufacturer's instructions. The quantity and quality of RNA extracted were examined by Agilent 2100 and Fragment Analyzer. Preparation and High-throughput sequencing of the mRNA library was accomplished on an Illumina HiSeq 3000 platform.

Bioinformatic analysis

The FastQC software was operated to approve the raw reads quality (Babraham Bioinformatics, Cambridge, UK). These raw reads were then assembled utilizing SPAdes or/and Trinity programs to generate 232,446 contigs that were

https://doi.org/10.22268/AJPP-040.2.169174

scanned against a local virus sequence database, retrieved from NCBI viral sequences database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus? (Accessed in 25/7/2021) using BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (E-value < 0.01) in order to identify potential similar sequences to viral RNA sequences. The RefSeq of the candidate viruses detected were imported from NCBI Reference Sequence Database. Bowtie 2 program was functioned to RefSeq-based assembly search and matched contigs that are similar to viral sequences were selected and additional annotated and reported using the Geneious Prime® 2021.1.1.(Lahuf, 2021).

The viral sequences of the predicted conserved and functional domains were recognized via InterProScan program (https://www.ebi.ac.uk/interpro/search/sequence/). The fundamental motifs sequences of genes and other genetic regions identified were downloaded and aligned with several similar sequences using the ClustalW method embedded in MEGA v.10.1.5 that was also operated for building of the phylogenetic trees applying the neighbor-joining approach with 1000 bootstrap replicates (Kumar *et al*., 2016).

Results

A total of 61,183,994 paired ends reads with 100 nucleotide (nt.) length were generated by sequencing of the prepared mRNA libraries. Significant numbers of contigs were assembled that showed high similarity to various sequences of plant viruses. However, only those contigs of putative viral origins that were longer than 400 nucleotide lengths were selected. In total, there were seven transcript hits (length > 400 nt) to putative plant viruses were identified, herein a description of their features:

***Alfalfa mosaic virus* (AMV)**

In this study, 48,630 contigs (a collection of reading sequence joined to one another by overlap of their sequences) related to AMV were identified, representing the highest number of detected viral contigs. This means that AMV is possibly the supreme virus associated with *B. tabaci*. These contigs were assembled to one consensus contig in 631 nt. long that showed similarity to the 3`end UTR of AMV segment 1 (Figure 1). This finding was confirmed through the phylogenetic analysis that proved significant resemblances among the detected putative sequence of AMV and several strains of the same virus particularly those with the accession numbers MT669391.1 and MT596809.1 (Figure 2).

***Broad bean wilt virus* (BBWV)**

The second most frequently detected virus was BBWV with 25,922 contigs identified that assembled to consensus 569 nt. contig. This contigs was similar to the 5` UTR of the BBWV segment RNA1 (Figure 3). The phylogenetic analysis approved this identification through the significant similarities among the detected BBWV putative sequence and many other strains of the same virus mainly the two strains with accession numbers MN216351.1 and MN216349.1 (Figure 2).

Figure 1. The AMV contigs coverage. The upper bar is the 3`UTR of the viral genome (indicated with black arrow). The lower bars represent the contigs produced.

***Zantedeschia mild mosaic virus* (ZaMMV)**

Additionally, 20,099 contigs were matched to part of ZaMMV genome, representing the third highest number of detected contigs. In other words, the ZaMMV is probably one of top three most frequently detected viruses in the studied *B. tabaci* population. These contigs were assembled to produce one consensus contig consists of 268 nt. that displayed similarity to the NIP and CP genes of the same virus genome (Figure 4). This conclusion was based on the results of the phylogenetic analysis that showed high resemblances among the putative ZaMMV contig and various strains of the same virus mostly the Australian strain with accession numbers KT729506.1 (Figure 2).

***Pittosporum cryptic virus*-1 (PiCV1)**

The PiCV1 was another plant virus detected through identification of 9,254 contigs mapped to part of RdRp gene and 3`UTR of its genome, representing the fourth highest quantity of contigs found. These contigs were assembled to create one consensus contig consists of 1396 nt. (Figure 5). This detection was based on the phylogenetic analysis that indicated clearly the high resemblance among the putative PiCV1 contig and various strains of the same virus especially that strain with accession numbers LR679767.1 (Figure 2).

***Grapevine leafroll-associated virus* (GLRaV)**

The GLRaV was also detected in the *B. tabaci* population by identifying 4,671 contigs covering the 3` end of its coat protein-like gene (Figure 6). These contigs were subsequently assembled into one consensus contig with 937 nt. Long, and the phylogenetic analysis confirmed this identification by placing it in the same clade and sharing the same descent with many other strains of the same virus, mainly those with accession numbers HQ442263 to 67.1 (Figure 2).

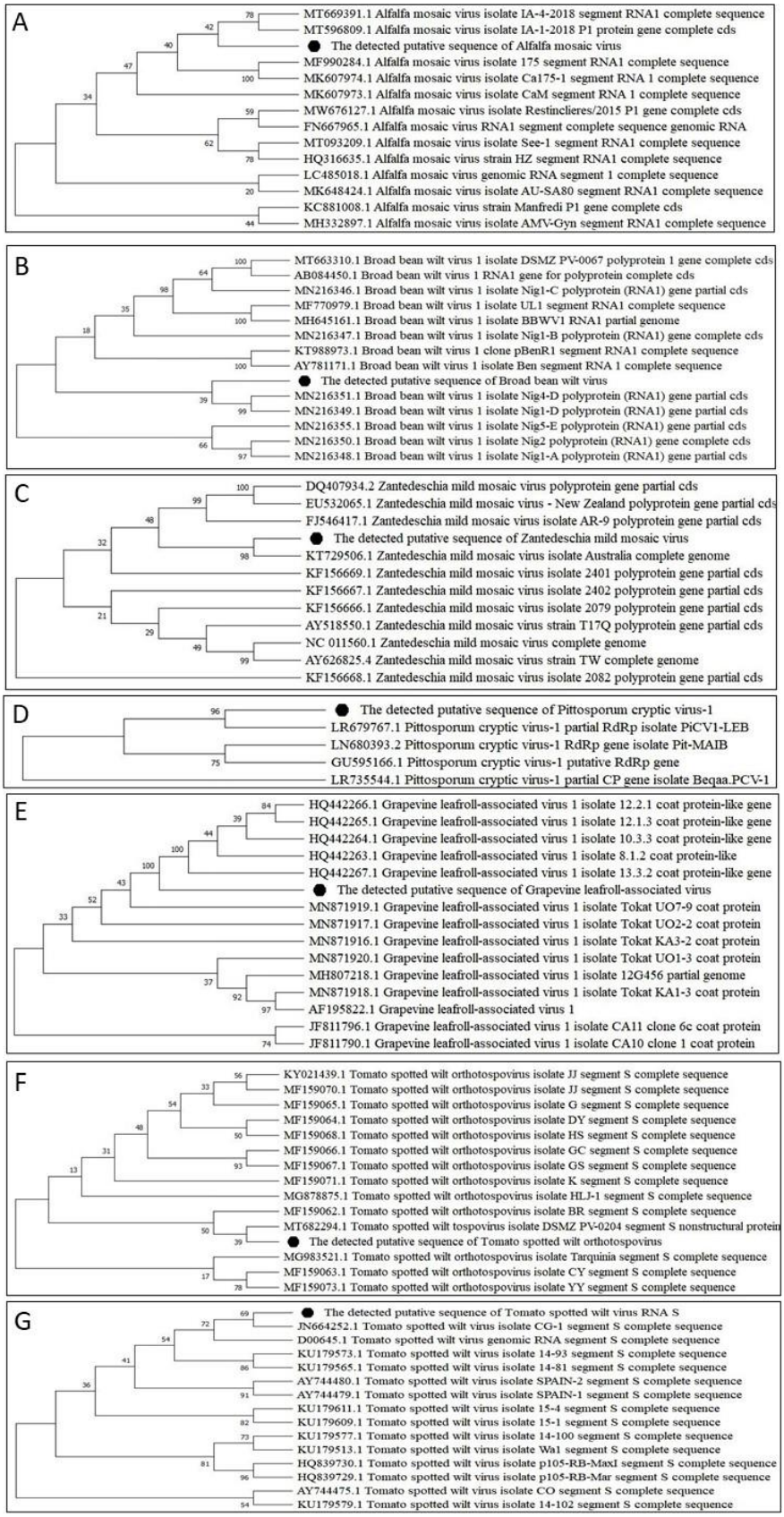

Figure 2. The phylogenetic correlation among the assembled putative AMV (A), BBWV (B) ZaMMV (C), PiCV1 (D) GLRaV (E), *Tomato spotted wilt orthotospovirus* (F) and TSWV (G) obtained in this study (marked with black dot) and other global GenBank strains.

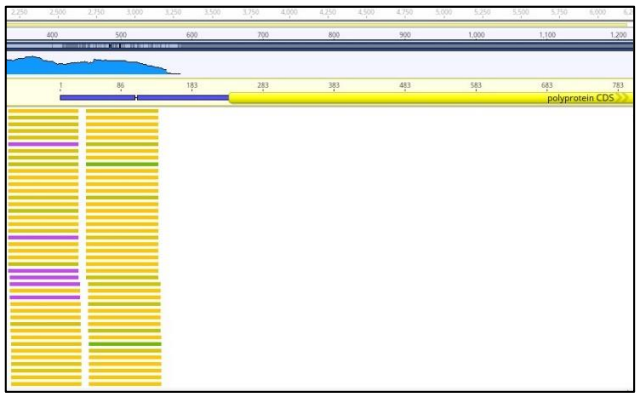

Figure 3. The BBWV contigs coverage. The upper bar is the 5`UTR of the viral genome (indicated with a black arrow). The lower bars represent the contigs created.

Figure 4. The ZaMMV contigs coverage. The upper bar is the reference sequence of the viral genome (marked with black arrow). The lower bars represent the contigs produced.

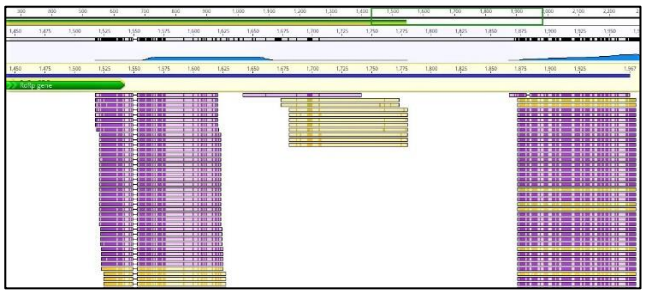

Figure 5. The PiCV1 contigs coverage. The upper bar is the reference sequence of the PiCV1 (marked with black arrow). The lower bars represent the contigs generated.

Figure 6. The GLRaV contigs coverage. The upper bar is the reference sequence of the GLRaV CP like gene (marked with the black arrow). The lower bars represent the contigs collected.

Tomato spotted wilt orthotospovirus* and *Tomato spotted wilt virus

The last two viruses documented were *Tomato spotted wilt orthotospovirus* and *Tomato spotted wilt virus* that belong to the same genus. This study revealed the presence of 1283 and 464 contigs to these viruses, respectively. The first group of contigs covered the non-coding region between the non-structural protein gene and nucleocapsid protein gene (Figure 7), and can be assembled to one consent contig with 437 nt. Long, and showed significant similarity with many strains of Tomato spotted wilt orthotospovirus, particularly the MT682294.1 strain (Figure 2). On the other hand, the second group of contigs were mostly similar to hairpin structure repeat region of Tomato spotted wilt virus RNA S (TSWV) (Figure 8). Their consensus one contig was 261 nt. long, which also displayed high similarity with those of TSWV preserved in the NCBI-GenBank, and the closest strain was JN664252.1 (Figure 2). It should be mentioned here that ICTV do not recognize Tomato spotted wilt orthotospovirus as a separate species from TSWV.

Figure 7. The coverage contigs of *Tomato spotted wilt orthotospovirus*. The upper bar is the reference sequence of the detected virus (marked with black arrow). The lower bars represent the contigs identified.

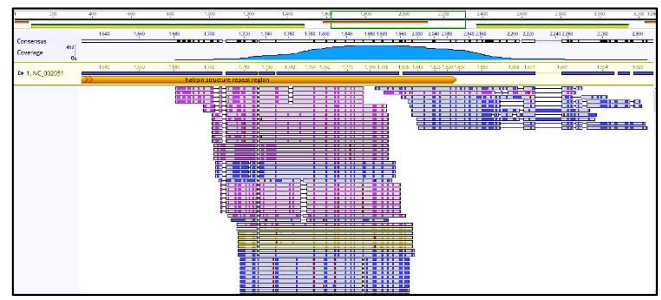

Figure 8. The coverage contigs of TSWV. The upper bar is the reference sequence of the identified virus (marked with black arrow). The lower bars represent the contigs documented.

Discussion

Generally, viruses infecting plant hosts are difficult to detect accurately and rapidly using traditional techniques. However, in the last two decades, NGS technologies and bioinformatics analysis have been applied in various aspects of microorganisms detection and discovery including viruses

which infect different hosts such as plants (Barba *et al.*, 2014), insects (Kobayashi *et al*., 2017), fungi (Chiapello *et al.*, 2020) and humans (Chiu, 2013). Different types of data are produced by the NGS technique including the transcriptome sequencing data that has been used successfully for the identification of plant and insect viruses through viral small RNA analysis (Zografidis *et al.*, 2015).

There are more than 1000 whitefly species belonging to over 1200 genera that were reported, some of them transmit plant viruses (Anon, 2001). However, only two genera of whiteflies (*Bemisia* and *Trialeurodes*) were reported as a vectors of numerous plants viruses. In the genus Bemisia, only one species, *B. tabaci* was found to be a vector that acquire plant viruses during feeding process and transmit them to healthy plants. This species gained its importance as virus vector and as an economic pest of field crops, fruits, vegetables and ornamental plants causing significant losses in several regions with tropical, subtropical, arid and Mediterranean environments (De Barro, 1995; Jones, 2003).

In the current study, several partial RNA sequences derived from plant viruses were identified in *B. tabaci* through transcriptome analysis, most of them in high quantity. Based on this high quantity of sequencing reads, a suggestion can be raised that these identified viruses are constantly expressed in the *B. tabaci* population. It should be mentioned however, that all the associated viruses found in whiteflies are not reported previously to be vectored by whiteflies from infected to healthy plants, but their detection in whiteflies suggest their existence in the host on which the whiteflies were feeding on. Ghosh *et al.* (2019), however, have found a new RNA virus Pepper whitefly-borne vein yellows virus, which belongs to the genus Polerovirus, usually transmitted by the aphid *M. persicae*, was found to be transmitted by the whitefly *B. tabaci*, which is a well-known vector of several DNA and RNA plant viruses. Nevertheless, whether a virus is only associated with a whitefly species or vectored by such species require further investigations.

Acknowledgments

High appreciation is expressed to Prof. Dr. Junmin Li (Institute of Plant Virology, Ningbo University, China) for his advice to improve this manuscript. This work was supported by the National Natural Science Foundation of China (U20A2036).

الملخص

سلمان، محمد وعدنان لهوف. 2022. طريقة جديدة لتشخيص الفيروسات النباتية ذات الجينوم RNA المرتبطة بالذبابة البيضاء (*Bemisia tabaci*) باستعمال تسلسل الجيل التالي (NGS). مجلة وقاية النبات العربية، 40(2): 169–174. https://doi.org/10.22268/AJPP-040.2.169174

تمّ استعمال تقنية تسلسل الجيل التالي (NGS) للحصول على بيانات النسخ (Transcriptome Data) من عينات حشرة الذبابة البيضاء العراقية (*Bemisia tabaci*) والتي تمّ تحليلها باستخدام بعض أدوات وبرامج المعلومات الحيوية المتخصّصة بتحديد الفيروسات النباتية ذات الحمض النووي الريبي (RNA) المصاحبة للحشرة. أظهرت نتائج عملية التحليل وجود سبعة فيروسات نباتية مختلفة: *Alfalfa mosaic virus* (AMV)، *Pittosporum cryptic virus-1* (PiCV1)، *Grapevine leafroll-associated virus* (GLRaV)، *Broad bean wilt virus* (BBWV)، *Zantedeschia mild mosaic virus* (ZaMMV)، بالإضافة إلى فيروس الذبول المتبقع للبندورة *Tomato spotted wilt virus* (TSWV). كانت أكبر كمية من تسلسلات الفيروسات التي تم تحديدها تعود الى AMV، BBWV و ZaMMV، على التوالي. طبّق في هذه الدراسة أسلوب جديد قائم على تقنية تسلسل الجيل التالي (NGS)، والذي سهّل عملية الكشف والتشخيص الفوري والدقيق للفيروسات النباتية نوعياً وكمياً في مجموعة من عينات حشرات الذبابة البيضاء *B. tabaci* دون الحاجة إلى بادئات محدَّدة (Specific Primers) وتطبيق تقنية الـ PCR التقليدية. وتتمثل فائدة هذا الأسلوب في الكشف السريع للفيروسات المحتملة المنقولة بواسطة الذبابة البيضاء أو المصاحبة لها والتي تؤثر على عددٍ لا يكاد يحصى من العوائل النباتية. ومع ذلك، لابدَّ من إجراء فحوصات إضافية لتأكيد هذه النتائج.

كلمات مفتاحية: تسلسل الجيل التالي (NGS)، الفيروسات النباتية ذات الحمض الريبي، الذبابة البيضاء، بيانات النسخ.

عناوين الباحثين: سلمان، محمد وعدنان لهوف*، قسم وقاية النبات، كلية الزراعة، جامعة كربلاء، العراق. *البريد الالكتروني للباحث المراسل: adnan.lahuf@uokerbala.edu.iq

References

Adnan, M., W. Zheng, W. Islam, M. Arif, Y. Abubakar, Z. Wang and G. Lu. 2017. Carbon catabolite repression in filamentous fungi. International Journal of Molecular Science, 19(1): 48. https://doi.org/10.3390/ijms19010048

Anon. 2001. Crop Protection Compendium, Global Module, 3rd edn. CAB International CD-Rom Database, UK.

Alemandri, V., C.G. VaghiMedina, A.D. Dumon, E.B. ArguelloCaro, M.F. Mattio, S. GarciaMedina, P.M. LopezLambertini and G. Truol. 2015. Three members of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) cryptic species complex occur sympatrically in Argentine horticultural crops. Journal of Economic Entomology, 108(2): 405-413. https://doi.rog/10.1093/jee/tov017

Barba, M., H. Czosnek and A. Hadidi. 2014. Historical perspective, development and application of next-generation sequencing in plant virology. Viruses, 6: 106-136. https://doi.org/10.3390/v6010106

Belliure, B., A. Janssen, P.C. Maris, D. Peters and M.W. Sabelis. 2005. Herbivore arthropods benefit from vectoring plant viruses. Ecological Letters, 8(1): 70-79. https://doi.org/10.1111/j.1461-0248.2004.00699.x

Chiapello, M., J. Rodríguez-Romero, M.A. Ayllón and M. Turinam. 2020. Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. Virus Evolution, 6: veaa058. https://doi.org/10.1093/ve/veaa058

Chiu, C.Y. 2013. Viral pathogen discovery. Current Opinion in Microbiology, 16: 468-478. https://doi.org/10.1016/j.mib.2013.05.001

De Barro, P.J. 1995. *Bemisia tabaci* biotype B, A Review of its Biology, Distribution and Control, 2nd edition, CSIRO Division of Entomology, Technical Paper No. 36. CSIRO, Canberra, Australia.

De Barro, P.J., S.S. Li, L.M. Boykin and A.B. Dinsdale. 2011. *Bemisia tabaci*: a statement of species status. Annual Review of Entomology, 56(1): 1-19. https://doi.org/10.1146/annurev-ento-112408-085504

Dou, X., S. Liu, V. Soroker, A. Harari and R. Jurenka. 2021. Novel RNA viruses from the transcriptome of pheromone glands in the pink bollworm moth, *Pectinophora gossypiella*. Insects, 12(556): 1-13. https://doi.org/10.3390/insects12060556

Ghosh, S., S. Kanakala, G. Lebedev, S. Kontsedalov, D. Silverman, T. Alon, N. Mor, N. Sela, N. Luria, A. Dombrovsky, M. Mawassi, S. Haviv, H. Czosnek and M. Ghanim. 2019. Transmission of a new *Polerovirus* infecting pepper by the whitefly *Bemisia tabaci*. Journal of Virology, 93: e00488-19. https://doi.org/10.1128/JVI.00488-19

Gilbertson, R.L., O. Batuman, C.G. Webster and S. Adkins. 2015. Role of the insect supervectors *Bemisia tabaci* and Frankliniella occidentalis in the emergence and global spread of plant viruses. Annual Review of Virology, 2(1): 67-93. https://doi.org/10.1146/annurev-virology-031413-085410

Islam, W., J. Zhang, M. Adnan, A. Noman, M. Zainab and J. Wu. 2017. Plant virus ecology: a glimpse of recent accomplishments. Applied Ecology and Environmental Research, 15(1): 691-705. https://doi.org/10.15666/aeer/1501_691705

Jones, D.R. 2003. Plant viruses transmitted by whiteflies. European Journal of Plant Pathology, 109(3): 195-219. https://doi.org/10.1023/A:1022846630513

Kobayashi, D., H. Isawa, R. Fujita, K. Murota, K. Itokawa, Y. Higa, Y. Katayama, T. Sasaki, T. Mizutani and S. Iwanage. 2017. Isolation and characterization of a new iflavirus from *Armigeres* spp. mosquitoes in the Philippines. Journal of General Virology, 98: 2876-2881. https://doi.org/10.1099/jgv.0.000929

Kumar, S., G. Strcher and K. Tamura. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular and Biological Evolution, 33: 1870-1874. https://doi.org/10.1093/molbev/msw054

Lahuf, A.A. 2021. RNA-Seq data analysis reveals various viral sequences associated with genome of date palm (*Phoenix dactylifera*). Earth and Environmental Science, 735 012027. https://doi.org/10.1088/1755-1315/735/1/012027

Legg, J.P., S.C. Jeremiah, H.M. Obiero, M.N. Maruthi, I. Ndyetabula, G. Okao-Okuja, H. Bouwmeester, S. Bigirimana, W. Tata-Hangy, G. Gashaka, G. Mkamilo, T. Alicai and K.P. Lava. 2011. Comparing the regional epidemiology of the cassava mosaic and Cassava brown streak virus pandemics in Africa. Virus Research, 159(2): 161-170. https://doi.org/10.1016/j.virusres.2011.04.018

Luan, J.B., X.W. Wang, J. Colvin and S.S. Liu. 2014. Plant-mediated whitefly-begomovirus interactions: research progress and future prospects. Bulletin of Entomological Research, 104(3): 267-26. https://doi.org/10.1017/S000748531400011X

Polston, J.E., P.J. De Barro and L.M. Boykin. 2014. Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. Pest Management Science, 70(10): 1547-52. https://doi.org/10.1002/ps.3738

Zografidis, A., F.V. Nieuwerburgh, A. Kolliopoulou, K. Apostolou-Karampelis, S.R. Head, D. Defore, G. Smagghe and L. Swevers. 2015. Viral small-RNA analysis of Bombyx mori larval midgut during persistent and pathogenic cytoplasmic polyhedrosis virus infection. Journal of Virology, 89: 11473-11486. https://doi.org/10.1128/JVI.01695-15

Received: December 24, 2021; Accepted: March 26, 2022

تاريخ الاستلام: 2021/12/24؛ تاريخ الموافقة على النشر: 2022/3/26