Gene 638 (2018) 7–12

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Identification of novel RNA viruses in alfalfa (*Medicago sativa*): an *Alphapartitivirus*, a *Deltapartitivirus,* and a *Marafivirus*

Hyein Kim[1], Dongbin Park[1], Yoonsoo Hahn*

Department of Life Science, Research Center for Biomolecules and Biosystems, Chung-Ang University, Seoul 06974, South Korea

ARTICLE INFO

Keywords:
Alfalfa
Medicago sativa alphapartitivirus 1
Medicago sativa deltapartitivirus 1
Medicago sativa marafivirus 1
Partitiviridae
Tymoviridae

ABSTRACT

Genomic RNA molecules of plant RNA viruses are often co-isolated with the host RNAs, and their sequences can be detected in plant transcriptome datasets. Here, an alfalfa (*Medicago sativa*) transcriptome dataset was analyzed and three new RNA viruses were identified, which were named Medicago sativa alphapartitivirus 1 (MsAPV1), Medicago sativa deltapartitivirus 1 (MsDPV1), and Medicago sativa marafivirus 1 (MsMV1). The RNA-dependent RNA polymerases of MsAPV1, MsDPV1, and MsMV1 showed about 68%, 58%, and 46% amino acid sequence identity, respectively, with their closest virus species. Sequence similarity and phylogenetic analyses indicated that MsAPV1, MsDPV1, and MsMV1 were novel RNA virus species that belong to the genus *Alphapartitivirus* of the family *Partitiviridae*, the genus *Deltapartitivirus* of the family *Partitiviridae*, and the genus *Marafivirus* of the family *Tymoviridae*, respectively. The bioinformatics procedure applied in this study may facilitate the identification of novel RNA viruses from plant transcriptome data.

1. Introduction

Alfalfa (*Medicago sativa* L.), which is known as the "queen of the forages," is one of the most valuable crops used to feed sheep, dairy cattle, and other livestock (Samac et al., 2006). Alfalfa is cultivated in many countries, including the United States of America, Argentina, Canada, Russia, Italy, and China (Hu and Cash, 2009). Because of its ability to fix nitrogen, alfalfa is considered one of the most important crops ensuring sustainability in soil fertility (Vance et al., 1979). High throughput molecular studies such as transcriptomic analysis have been performed to understand the biology of alfalfa for productivity increases (Zhang and Wang, 2014), to evaluate the quality and yield of transgenic alfalfa (Gao et al., 2016), and to identify genes and molecular mechanisms associated with responses to lead stress (Xu et al., 2017) or aluminum stress (Liu et al., 2017).

A large number of viruses have been shown to infect plants, many of which usually manifest diseases with readily recognizable symptoms (Scholthof et al., 2011). However, some plant RNA viruses do not cause visible symptoms and are known as persistent viruses (Randall and Griffin, 2017). There are at least 12 DNA and RNA viruses reported to infect alfalfa; these viruses include Alfalfa dwarf virus, Alfalfa enamovirus 1, Alfalfa latent virus, Alfalfa leaf curl virus, Alfalfa mosaic virus, Bean leafroll virus, Bean yellow mosaic virus, Clover yellow mosaic virus, Groundnut rosette virus, Lucerne transient streak virus, Pea streak virus, and Red clover necrotic mosaic virus (http://www.genome.jp/virushostdb/3879; last accessed June 25, 2017) (Mihara et al., 2016).

Many plant viruses cause infected plants to develop disease symptoms. However, persistent plant RNA viruses are not well studied because they generally do not cause any apparent disease (Roossinck, 2010; Roossinck, 2013). Genomic molecules of these persistent RNA viruses can be isolated together with the host RNAs, and their sequences can be identified in the RNA-seq datasets (Eichmeier et al., 2016; Park and Hahn, 2017b; Park and Hahn, 2017a).

In this study, we analyzed alfalfa leaf transcriptome data collected for the identification of fall dormancy-related genes (Zhang et al., 2015) and identified genome sequences of three RNA viruses which were novel species of the genera *Alphapartitivirus*, *Deltapartitivirus*, and *Marafivirus*.

Abbreviations: CP, coat protein; MsAPV1, Medicago sativa alphapartitivirus 1; MsDPV1, Medicago sativa deltapartitivirus 1; MsMV1, Medicago sativa marafivirus 1; NCBI, National Center for Biotechnology Information; ORF, open reading frame; RdRp, RNA-dependent RNA polymerase; SRA, Sequence Read Archive

* Corresponding author at: Department of Life Science, Research Center for Biomolecules and Biosystems, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, South Korea.

E-mail address: hahny@cau.ac.kr (Y. Hahn).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2017.09.069
Received 30 July 2017; Received in revised form 13 September 2017; Accepted 29 September 2017
Available online 30 September 2017
0378-1119/

2. Materials and Methods

2.1. Alfalfa transcriptome dataset

The alfalfa transcriptome data analyzed in this study were downloaded from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). The data comprised about 32 gigabase pairs of RNA-seq reads obtained from alfalfa leaves (SRA accession number SRA057663) (Zhang et al., 2015).

2.2. Transcriptome assembly

A previously reported protocol, which was developed to identify spinach RNA viruses (Park and Hahn, 2017b; Park and Hahn, 2017a), was applied with modifications in this study. First, high quality sequence reads were collected using the Sickle program (version 1.33; http://github.com/najoshi/sickle) with the parameters "-q 30 -l 50." Then, the SPAdes Genome Assembler software (version 3.10.0; http://cab.spbu.ru/software/spades) was used to assemble the RNA-seq reads into contigs (Bankevich et al., 2012).

2.3. Identification of virus genome sequences

The RNA-dependent RNA polymerase (RdRp) motif sequences of known RNA viruses were obtained from the Pfam database (release 30.0; http://pfam.xfam.org). The accession numbers of the collected RdRp families are PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF08716, PF08717, and PF12426. There were 345 non-redundant RdRp motif sequences.

To identify viral RdRp-containing genomic segments in the assembled transcriptome contigs, a BLASTX search was performed against the RdRp motif sequences using the parameters "-e 1e-5 -max_target_seqs 1." Matched contigs longer than 1 kbp were collected. To confirm whether the candidate was a viral genomic segment and to identify closely related viruses, a BLASTX search of the non-redundant protein database was performed using each viral contig as a query at the NCBI web site (https://blast.ncbi.nlm.nih.gov/Blast.cgi). When the virus was predicted to have a multipartite genome, additional genomic segments were retrieved from the transcriptome contigs based on similarity to closely related virus proteins. Protein coding regions of the viral genomic segments were predicted by BLASTX searches against the viral protein sequences available at NCBI database and by using the ORF finder (https://www.ncbi.nlm.nih.gov/orffinder).

Domain analysis was performed using Pfam (http://pfam.xfam.org). A sequence logo was generated using WebLogo 3 (http://weblogo.threeplusone.com) (Schneider and Stephens, 1990; Crooks et al., 2004).

2.4. Phylogenetic analysis

Multiple sequence alignments of RdRp motif-containing protein sequences were generated by using the MUSCLE program (version 3.8.31; http://www.drive5.com/muscle) (Edgar, 2004). Phylogenetic trees were inferred from the aligned RdRp sequences using the neighbor-joining method implemented in the ClustalW2 software (version 2.1; http://www.clustal.org) (Larkin et al., 2007).

3. Results and discussion

3.1. Identification of novel viruses in alfalfa transcriptome data

Genome sequences of three novel virus species (Table 1) were discovered by analysis of the transcriptome data of alfalfa leaf (SRA accession number SRA057663) (Zhang et al., 2015). Sequence similarity searches and phylogenetic analyses showed that these viruses were new species of the genera *Alphapartitivirus*, *Deltapartitivirus*, and *Marafivirus*. The three viruses were named Medicago sativa alphapartitivirus 1 (MsAPV1), Medicago sativa deltapartitivirus 1 (MsDPV1), and Medicago sativa marafivirus 1 (MsMV1).

3.2. Medicago sativa alphapartitivirus 1 (MsAPV1)

One of the three new viruses discovered in the alfalfa transcriptome showed sequence similarity with previously reported plant RNA viruses of the genus *Alphapartitivirus* of the family *Partitiviridae*. Subsequent phylogenetic analysis confirmed that this virus was a member of the genus *Alphapartitivirus*. Therefore, the virus was named Medicago sativa alphapartitivirus 1 (MsAPV1).

MsAPV1 has two genomic segments, RNA1 and RNA2. The RNA1 segment is 1922 nucleotides (nt) in length and has an open reading frame (ORF) for a RdRp comprising of 586 amino acids (aa) (Table 1). The MsAPV1 RdRp showed 68–69% aa sequence identities with RdRps of the most closely related viruses, including Vicia faba partitivirus 1 (VfPV1) (Blawid et al., 2007), Rose partitivirus (RoPV) (Phelan and James, 2016), and Arabidopsis halleri partitivirus 1 (AhPV1) (Kamitani et al., 2016) (Table 2 and Supplementary Fig. S1). RdRps of two other viruses, Bipolaris maydis partitivirus 1 (BmPV1) (Deng et al., 2017) and Rosellinia necatrix partitivirus 2 (RnPV2) (Chiba et al., 2013), also showed more than 50% aa sequence identity with the MsAPV1 RdRp. These viruses were members of the genus *Alphapartitivirus* of the family *Partitiviridae* (Nibert et al., 2014).

The RNA2 segment of MsAPV1 is 1707 nt in length and encodes a 491-aa-long coat protein (CP). The MsAPV1 CP also showed strong aa sequence identities with CPs of four of the above-mentioned alphapartitiviruses: RoPV, 64% (374/481); BmPV1, 51% (242/478); AhPV1, 42% (201/477); and RnPV2, 36% (178/494) (Supplementary Fig. S2). The VfPV1 CP was not found in the NCBI Protein database. The partitivirus CPs showed lower sequence identities than the RdRps because CPs are known to evolve faster than RdRps (Crawford et al., 2006).

For the phylogenetic analysis, multiple sequence alignments of the RdRp and CP sequences of MsAPV1 and other related viruses were performed (see Supplementary Figs. S1 and S2). Phylogenetic trees from RdRps and CPs confirmed that MsAPV1 is a novel member of the genus *Alphapartitivirus* (Fig. 1A). The phylogenetic tree from RdRps (see Fig. 1A, left panel) showed that MsAPV1 formed a strong clade with VfPV1, RoPV, and AhPV1, which were isolated from plants of three different orders of the rosids clade: Fabales (alfalfa and *Vicia*), Rosales (rose), and Brassicales (*Arabidopsis*) (Blawid et al., 2007; Kamitani et al., 2016; Phelan and James, 2016). Interestingly, the next most closely related viruses, BmPV1 and RnPV2, were isolated from the fungi *Bipolaris maydis* and *Rosellinia necatrix*, respectively (Chiba et al., 2013; Deng et al., 2017). The close relationship among plant-infecting and fungus-infecting alphapartitiviruses suggests that MsAPV1 and its closely related plant-infecting viruses might have originated from fungal hosts by horizontal transmission (Nibert et al., 2014).

In contrast to the RdRp tree, the CP tree (see Fig. 1A, right panel) revealed that MsAPV1 was closer to the fungus-infecting BmPV1 than to the plant-infecting AhPV1. The discrepancy between RdRp and CP tree topologies can be explained by the reassortment mechanism in which different viruses exchange genomic segments during co-infection (McDonald et al., 2016). This process may be very common among partitiviruses because their genomic segments are packaged in separate virion particles (Pan et al., 2009).

3.3. Medicago sativa deltapartitivirus 1 (MsDPV1)

The second virus discovered in the alfalfa transcriptome dataset also consisted of two RNA segments. Each of these segments showed sequence similarities with the corresponding segments of viruses of the genus *Deltapartitivirus* (family *Partitiviridae*), including Citrullus lanatus cryptic virus (ClCV) (Xin et al., 2017), Pepper cryptic virus 1 (PepCV1) (Sabanadzovic and Valverde, 2011), Pittosporum cryptic virus 1