Diseases Caused by Viruses

Identification and Complete Genome Sequencing of Alfalfa Virus S Diagnosed in Alfalfa Plants (*Medicago sativa*) from Washington State, U.S.A.

Lev G. Nemchinov,^{1,†} Samuel Grinstead,² Brian M. Irish,³ and Jonathan Shao¹

¹ USDA/ARS, Beltsville Agricultural Research Center, Molecular Plant Pathology Laboratory, Beltsville, MD 20705

² USDA/ARS, Beltsville Agricultural Research Center, National Germplasm Resources Laboratory, Beltsville, MD 20705

³ USDA/ARS Plant Germplasm Introduction Testing and Research Unit, Prosser, WA 99350

The complete genomic sequence of the virus has been deposited in GenBank under the accession number MT094142. Raw sequencing reads were deposited in NCBI Sequence Read Archive under BioProject number PRJNA641699.

Funding: This research was supported by the United States Department of Agriculture, Agricultural Research Service, CRIS project 8042-21000-300-00D. Plant Dis. 104:3271, 2020; published online as <https://doi.org/10.1094/PDIS-06-20-1374-PDN>. Accepted for publication 16 July 2020.

Alfalfa virus S (AVS) is a new species in the genus *Allexivirus*, family *Alphaflexiviridae*. It was recently identified in alfalfa plants obtained from Sudan, Northern Africa (Nemchinov et al. 2017). In 2019, the virus was also diagnosed in alfalfa samples acquired from China (Nemchinov and Grinstead 2020). Prior to this report, the virus has not been identified in the United States. Alfalfa plants used in this study originated from the Prosser, WA, area. Leaves displaying chlorotic symptoms with a mosaic pattern were collected from 10 different plants and pooled together in two samples (three to five leaves in each) for RNA extraction. Total RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific). Libraries were prepared from total RNA using the poly(A) selection protocol, and high-throughput sequencing (HTS) of paired-end reads (2 × 150) was performed on a HiSeq Illumina Platform by GENEWIZ (South Plainfield, NJ). As many as 116,965 of 70,213,076 (0.16%) quality-trimmed raw reads (CLC quality scores 0.05; reads with more than one ambiguous nucleotide were removed) were detected in one of the samples. They were assembled de novo by Qiagen's CLC Genomics Workbench (version 20.0) using automatic bubble and word size and a minimum contig length of 175 bp.

HTS reads that mapped to the reference genome of AVS (KY696659) employing the CLC's maximal exact match algorithm were then mapped back to the contigs assembled from the data. A complete nucleotide sequence of the AVS-US isolate was obtained by reference-guided assembly using CLC Genomics Workbench, MegAlign tool (version 15.2.0) of the Lasergene suite (DNA-STAR, Madison, WI), and 5' rapid amplification of cDNA ends (SMARTer 5' RACE kit, Takara Bio) with the virus-specific primer 5'-GATGCCTAACCCTTCTAGCG-3' (nt 270 to 289, reverse). A 3'-terminal poly(A) tail was detected in the assembled viral genome. The identity of the viral genome was additionally confirmed by RT-PCR with virus-specific primers (forward, 5'-TCGCTGCCTTTGACTTCTT-3', nt 7,795 to 7,814; reverse, 5'-GGACGGTGGTGGTGATTATT-3', nt 8,023 to 8,043) followed by Sanger sequencing of the RT-PCR product. The complete genomic sequence of the virus has been deposited in GenBank under the accession number MT094142. At the nucleotide level, the genome of AVS-US consisting of 8,350 nt excluding a poly(A) tail was 95.2 and 92.3% identical to AVS and AVS-C (MN864567), respectively, as predicted by BLASTn analysis. Amino acid identities of the six putative proteins encoded by the genome of the virus with the corresponding proteins of AVS and AVS-C varied from 96.6 to 98.9% (AVS) and from 92.8 to 96.8% (AVS-C), implying a closer relationship between the U.S. and Sudanese isolates of the virus. Phylogenetic analysis performed using Qiagen CLC Main Workbench software (version 20.0) with the predicted amino acid sequence of AVS-US RNA-dependent RNA polymerase and that of other known allexiviruses placed the three AVS isolates together in the same subcluster. As revealed by HTS, the alfalfa sample in which AVS was identified was coinfecting with other pathogens including alfalfa mosaic virus, bean leafroll virus, pea streak virus, *Medicago sativa* alphapartitivirus, *Medicago sativa* amalgavirus, and *Pseudomonas* bacterial species. This once again demonstrated that viral coinfections are common in alfalfa and may constitute a substantial part of the disease complexes. Thus far, AVS has been found in three global geographic locations: Sudan, China, and the United States. Pinpointing the source of AVS in alfalfa is difficult. However, virus genome diversity could be indicative of independent geographic origins for all three isolates.

References:

- Nemchinov, L. G., and Grinstead, S. C. 2020. Plant Dis. <https://doi.org/10.1094/PDIS-04-20-0906-SC>
Nemchinov, L. G., et al. 2017. PLoS One 12:e0178222.

The author(s) declare no conflict of interest.

e-Xtra

Keywords: alfalfa virus S, U.S. isolate, complete genome sequence, specific diagnosis

[†]Indicates the corresponding author.

E-mail: lev.nemchinov@usda.gov (L.G.N.)