

SWINE HEALTH

Title: United States Swine Pathogen Database – **NPB #16-222**

Investigator: Dr. Kay Faaberg, Ph.D.

Institution: USDA-ARS National Animal Disease Center

Date Submitted: February 20, 2018

Scientific Abstract

Veterinary diagnostic laboratories derive partial nucleotide sequences of thousands of isolates of PRRSV, SVA, PEDV and other coronaviruses annually, and with the advent of next generation sequencing, near full-length genomes are also rapidly produced. Presently, the sequence data are only released to the client, as the samples are associated with sensitive information. At the same time, however, this information is critical and can provide objective criteria for: 1) vaccine design; 2) determining when and how fast pathogens are spreading across the landscape; and 3) identifying transmission hotspots. In tandem with the USDA Agriculture Research Service Big Data initiative, we have generated a centralized nucleotide sequence relational database housed at the National Animal Disease Center. We have implemented the Tripal toolkit, using Drupal for Content Management, and the Chado relational database schema. Hosting is via a BlueHost cloud service with resource scaling, dedicated support for the prevention of data theft and control of database vulnerabilities, and the service is ~2X faster for general database query tasks and ~6X faster for more complex analyses. Each genetic sequence housed in the database contains at a minimum four core data items: genomic information; the date of collection; the US State the collection was made; and a unique identifier. Additionally, custom curation and annotation pipelines have determined PRRSV genotype (Type 1 or 2), the location of open reading frames and non-structural proteins, generated amino acid sequences, and identified putative frame shifts. This repository is currently private but will be publicly accessible in the future. The resource will ultimately provide researchers timely access to sequences discovered by highly qualified veterinary diagnosticians, allowing for biological data mining and epidemiological studies. The result of this effort will be a better understanding concerning the appearance of novel viruses in the United States, how these novel isolates are moving through the US and abroad, and discovering new patterns of biological consequence.

These research results were submitted in fulfillment of checkoff-funded research projects. This report is published directly as submitted by the project's principal investigator. This report has not been peer-reviewed.

For more information contact:

National Pork Board • PO Box 9114 • Des Moines, IA 50306 USA • 800-456-7675 • Fax: 515-223-2646 • pork.org
