A pipeline and dataset facilitating the classification of influenza A virus gene sequences detected in U.S. swine to evolutionary origin

Jennifer Chang,a,\* Tavis K. Anderson,a,\* Michael A. Zeller,a,b,c,\* Phillip C. Gauger,b Amy L. Vincenta,#

aVirus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, Iowa, USA

bDepartment of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA

cBioinformatics and Computational Biol­ogy Program, Iowa State University, Ames, Iowa, USA

Running Head: Classification of IAV genes detected in U.S. swine

\*Author order was determined by a random selection script.

#Address correspondence to: Dr. Amy L. Vincent, amy.vincent@ars.usda.gov

Abstract: **48/50** words; main text: **562/500** words

**ABSTRACT**

Influenza A viruses (IAV) in swine are classified as H1N1, H1N2, or H3N2 subtype. The 8 IAV genes are named by host source (human- or avian-origin) or evolutionary lineage. Here, we curated a dataset and pipeline that reproducibly assigns evolutionary lineage and genetic clade to query gene segments.

Influenza A virus (IAV) is a negative-sense, single-stranded, enveloped RNA virus of the *Orthomyxoviridae* family. Although only H1N1, H1N2, and H3N2 subtypes are endemic in swine around the world, much diversity can be found in the genes coding for major surface proteins, hemagglutinin (HA) and neuraminidase (NA), and in the other 6 internal gene segments. This diversity is the result of frequent bidirectional transmission between swine and humans, occasional transmission of an avian virus into swine, followed by periods of antigenic drift and shift.

Early IAV detected in swine emerged coincident with the 1918 Spanish flu, and are classified as classical-swine H1N1 (1). In the late 1990s, a triple-reassortant H3N2 virus was identified containing gene segments derived from seasonal human H3N2 (HA, NA, and PB1), avian IAV (PB2 and PA), and the classical H1N1 swine IAV (NP, M, and NS) (2, 3). These H3N2 persisted and evolved into defined phylogenetic clades, with Cluster-IV (C-IV) H3 sustaining and expanding in diversity to the present day (4). The triple-reassortant H3N2 viruses then reassorted with classical-swine H1N1 viruses, resulting in new lineages of H1N1 and H1N2 viruses. The major reassortment events preserved the triple reassortant internal gene (TRIG) constellation with swine (M, NP and NS), avian (PB2 and PA), and human (PB1) influenza virus origins. Genetically and antigenically distinct human seasonal IAV H1 and N2 also spilled into and established in swine in the early 2000s (5, 6). In 2009, a virus with NA and M genes from Eurasian avian H1N1 swine in addition to TRIG and classical-swine lineage genes emerged in swine, infected humans as a pandemic (H1N1pdm09), and now via reverse zoonoses contributes to genetic diversity in swine (7, 8). Hemagglutinin genes are paired with N2 genes derived from the 1998 or 2002 human seasonal origin (9), or an N1 gene from the classical-swine lineage or pandemic-lineage (10, 11). More recently, a distinct human H3N2 virus was transmitted to swine, termed H3.2010.1, this virus is distinct from the 1998 H3N2 lineage C-IV viruses (12). The processes of antigenic shift and drift in North America have led to approximately ten genetically distinct hemagglutinin (HA) genes, 4 neuraminidase (NA) genes, and 3 internal gene lineages (13, 14).

Here we introduce reference gene sequence datasets and an analytical pipeline that allow the assignment of evolutionary lineage to IAV gene segments. This workflow and reference data may play a role in the detection of reassorted viruses containing gene segments derived from avian or human sources. Users need (i) the reference gene datasets, and (ii) a FASTA with query sequences. Input FASTA sequences may be aligned or unaligned, but must have unique names for each strain and for each gene. The default pipeline (Fig. 1) takes query gene sequences and processes by: (i) identification to one of 8 genes using BLASTn; (ii) alignment of queries to appropriate reference gene dataset; (iii) the inference of a maximum likelihood tree; (iv) classification of query sequences using patristic distance extracted from the inferred tree and assignment to evolutionary lineage (TRIG, H1N1pdm2009, Classical-swine, Human-seasonal); and (v) generation of a summary classification file. The reference dataset for each gene includes outgroup sequences allowing the pipeline to flag sequences that are not contemporary circulating U.S. swine IAV.

**Data availability.** Gene sequences are in NCBI GenBank and reference datasets are hosted at the Influenza Research Database (XXX) (15). The pipeline is provided on GitHub (flu-crew/nn\_patristic\_classifier), and DockerHub (j23414/nn\_patristic\_classifier).

**ACKNOWLEDGMENTS**

The authors were supported by USDA-ARS, USDA-APHIS, and by an NIH-National Institute of Allergy and Infectious Diseases (NIAID) interagency agreement associated with CRIP (Center of Research in Influenza Pathogenesis), an NIAID funded Center of Excellence in Influenza Research and Surveillance (CEIRS, HHSN272201400008C) to ALV. JC and TKA were supported by an appointment to the USDA-ARS Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and USDA under contract number DE-AC05-06OR23100. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. PCG and MAZ were supported by the Iowa State University Veterinary Diagnostic Laboratory and a Presidential Interdisciplinary Research Initiative Award. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture, DOE, or ORISE. USDA is an equal opportunity provider and employer.

**REFERENCES**

1. **Shope RE**. 1931. SWINE INFLUENZA : III. FILTRATION EXPERIMENTS AND ETIOLOGY. J Exp Med **54**:373–385.

2. **Zhou NN**, **Senne DA**, **Landgraf JS**, **Swenson SL**, **Erickson G**, **Rossow K**, **Liu L**, **Yoon KJ**, **Krauss S**, **Webster RG**. 1999. Genetic reassortment of avian, swine, and human influenza A viruses in American pigs. J Virol **73**:8851–8856.

3. **Karasin AI**, **Schutten MM**, **Cooper LA**, **Smith CB**, **Subbarao K**, **Anderson GA**, **Carman S**, **Olsen CW**. 2000. Genetic characterization of H3N2 influenza viruses isolated from pigs in North America, 1977-1999: evidence for wholly human and reassortant virus genotypes. Virus Research **68**:71–85.

4. **Kitikoon P**, **Nelson MI**, **Killian ML**, **Anderson TK**, **Koster L**, **Culhane MR**, **Vincent AL**. 2013. Genotype patterns of contemporary reassorted H3N2 virus in US swine. J Gen Virol **94**:1236–1241.

5. **Karasin AI**, **Landgraf J**, **Swenson S**, **Erickson G**, **Goyal S**, **Woodruff M**, **Scherba G**, **Anderson G**, **Olsen CW**. 2002. Genetic characterization of H1N2 influenza A viruses isolated from pigs throughout the United States. J Clin Microbiol **40**:1073–1079.

6. **Vincent AL**, **Ma W**, **Lager KM**, **Gramer MR**, **Richt JA**, **Janke BH**. 2009. Characterization of a newly emerged genetic cluster of H1N1 and H1N2 swine influenza virus in the United States. Virus Genes **39**:176–185.

7. **Garten RJ**, **Davis CT**, **Russell CA**, **Shu B**, **Lindstrom S**, **Balish A**, **Sessions WM**, **Xu X**, **Skepner E**, **Deyde V**, **Okomo-Adhiambo M**, **Gubareva L**, **Barnes J**, **Smith CB**, **Emery SL**, **Hillman MJ**, **Rivailler P**, **Smagala J**, **de Graaf M**, **Burke DF**, **Fouchier RAM**, **Pappas C**, **Alpuche-Aranda CM**, **López-Gatell H**, **Olivera H**, **López I**, **Myers CA**, **Faix D**, **Blair PJ**, **Yu C**, **Keene KM**, **Dotson PD**, **Boxrud D**, **Sambol AR**, **Abid SH**, **St George K**, **Bannerman T**, **Moore AL**, **Stringer DJ**, **Blevins P**, **Demmler-Harrison GJ**, **Ginsberg M**, **Kriner P**, **Waterman S**, **Smole S**, **Guevara HF**, **Belongia EA**, **Clark PA**, **Beatrice ST**, **Donis R**, **Katz J**, **Finelli L**, **Bridges CB**, **Shaw M**, **Jernigan DB**, **Uyeki TM**, **Smith DJ**, **Klimov AI**, **Cox NJ**. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science **325**:197–201.

8. **Smith GJD**, **Vijaykrishna D**, **Bahl J**, **Lycett SJ**, **Worobey M**, **Pybus OG**, **Ma SK**, **Cheung CL**, **Raghwani J**, **Bhatt S**, **Peiris JSM**, **Guan Y**, **Rambaut A**. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature **459**:1122–1125.

9. **Nelson MI**, **Lemey P**, **Tan Y**, **Vincent A**, **Lam TT-Y**, **Detmer S**, **Viboud C**, **Suchard MA**, **Rambaut A**, **Holmes EC**, **Gramer M**. 2011. Spatial dynamics of human-origin H1 influenza A virus in North American swine. PLoS Pathog **7**:e1002077.

10. **Anderson TK**, **Campbell BA**, **Nelson MI**, **Lewis NS**, **Janas-Martindale A**, **Killian ML**, **Vincent AL**. 2015. Characterization of co-circulating swine influenza A viruses in North America and the identification of a novel H1 genetic clade with antigenic significance. Virus Research **201**:24–31.

11. **Anderson TK**, **Nelson MI**, **Kitikoon P**, **Swenson SL**, **Korslund JA**, **Vincent AL**. 2013. Population dynamics of cocirculating swine influenza A viruses in the United States from 2009 to 2012. Influenza Other Resp Viruses **7 Suppl 4**:42–51.

12. **Rajão DS**, **Gauger PC**, **Anderson TK**, **Lewis NS**, **Abente EJ**, **Killian ML**, **Perez DR**, **Sutton TC**, **Zhang J**, **Vincent AL**. 2015. Novel Reassortant Human-Like H3N2 and H3N1 Influenza A Viruses Detected in Pigs Are Virulent and Antigenically Distinct from Swine Viruses Endemic to the United States. J Virol **89**:11213–11222.

13. **Vincent AL**, **Perez DR**, **Rajao D**, **Anderson TK**, **Abente EJ**, **Walia RR**, **Lewis NS**. 2017. Influenza A virus vaccines for swine. Vet Microbiol **206**:35–44.

14. **Gao S**, **Anderson TK**, **Walia RR**, **Dorman KS**, **Janas-Martindale A**, **Vincent AL**. 2017. The genomic evolution of H1 influenza a viruses from swine detected in the United States between 2009 and 2016. J Gen Virol **98**:2001–2010.

15. **Zhang Y**, **Aevermann BD**, **Anderson TK**, **Burke DF**, **Dauphin G**, **Gu Z**, **He S**, **Kumar S**, **Larsen CN**, **Lee AJ**, **Li X**, **Macken C**, **Mahaffey C**, **Pickett BE**, **Reardon B**, **Smith T**, **Stewart L**, **Suloway C**, **Sun G**, **Tong L**, **Vincent AL**, **Walters B**, **Zaremba S**, **Zhao H**, **Zhou L**, **Zmasek C**, **Klem EB**, **Scheuermann RH**. 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res **45**:D466–D474.

**FIGURE LEGEND**

**Figure 1.** The nearest neighbor patristic classifier pipeline (A) and an inferred maximum likelihood tree generated on query and reference PB2 sequences (B). The PB2 gene example demonstrates the genetic lineages of contemporary influenza A virus circulating in United States swine populations: the H1N1 pandemic 2009 (red) and vaccine PB2 genes (orange) are monophyletic clades nested within the TRIG lineage (purple), human seasonal (grey), and classical-swine (blue) lineage PB2 genes are separate monophyletic clades. The tree is midpoint rooted for clarity; branch lengths are drawn to scale; and the scale bar indicates the number of nucleotide substitutions per site.