

1 octoFLU: Automated classification for evolutionary origin of influenza A virus gene  
2 sequences detected in U.S. swine

3

4 Jennifer Chang,<sup>a,\*</sup> Tavis K. Anderson,<sup>a,\*</sup> Michael A. Zeller,<sup>a,b,c,\*</sup> Phillip C. Gauger,<sup>b</sup> Amy  
5 L. Vincent<sup>a,#</sup>

6

7 <sup>a</sup>Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames,  
8 Iowa, USA

9 <sup>b</sup>Department of Veterinary Diagnostic and Production Animal Medicine, College of  
10 Veterinary Medicine, Iowa State University, Ames, Iowa, USA

11 <sup>c</sup>Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa,  
12 USA

13

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

15

16 \*These authors contributed equally.

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

18

19 Abstract: **44/50** words; main text: **678/500** words

20 **ABSTRACT**

21       The diversity of the 8 genes of influenza A viruses (IAV) in swine reflects introductions  
22   from non-swine hosts and subsequent antigenic drift and shift. Here, we curated a dataset and  
23   present a pipeline that assigns evolutionary lineage and genetic clade to query gene segments.

24       Although only H1N1, H1N2, and H3N2 subtypes are endemic in swine around the world,  
25   much diversity can be found in the genes coding for major surface proteins, hemagglutinin (HA)  
26   and neuraminidase (NA), and in the other 6 internal gene segments. The swine IAV that  
27   emerged coincident with the 1918 Spanish flu are classified as classical-swine H1N1 (1). In the  
28   late 1990s, triple-reassortant H3N2 viruses containing gene segments derived from human  
29   seasonal H3N2, avian IAV, and the classical-swine IAV were identified (2, 3). The HA persisted,  
30   evolving into phylogenetic clades (Cluster-IV (C-IV) clades A-F) (4). The triple-reassortant H3N2  
31   viruses reassorted with classical-swine H1N1 viruses, driving new genetic clades of H1N1 and  
32   H1N2 viruses (5), but preserving the triple reassortant internal gene (TRIG) constellation.  
33   Genetically distinct human seasonal H1 spilled into and established in swine in the early 2000s  
34   (6, 7). In 2009, a virus with genes from Eurasian-avian H1N1, TRIG, and classical-swine lineage  
35   genes emerged as a pandemic (H1N1pdm09) and continues to contribute to diversity in swine  
36   (8, 9). More recently, two distinct human H3N2 viruses were transmitted to swine, H3.2010.1  
37   and H3.2010.2 (10, 11). HA genes are paired with N2 genes derived from the 1998 or 2002  
38   human seasonal origin (12), or N1 genes from the classical-swine lineage or pandemic-lineage  
39   (13, 14). In 2018, a live-attenuated influenza virus (LAIV) vaccine became commercially  
40   available in the U.S. (15). The LAIV viruses contain HA, annotated as H3 Cluster-I or H1  
41   gamma2-beta-like, and NA, annotated as N2 LAIV-98 or N1 LAIV-Classical, expressed on a  
42   TRIG internal gene backbone, with all components isolated in the 1990s. Reassorted viruses  
43   with LAIV genes have been detected. Interspecies transmission episodes and the processes of  
44   antigenic shift and drift led to approximately 16 distinct HA clades, 4 NA lineages, and 3 internal  
45   gene lineages (16, 17).

46 We generated reference gene datasets and an analytical pipeline that assigns query HA  
47 to genetic clade and NA and internal IAV genes to evolutionary lineages that are found in IAV  
48 from U.S. swine. Users need the reference dataset and a FASTA with query sequences from  
49 any IAV gene segment. The input data must be of good quality and substantial length  
50 (approximately 50% or greater of the gene of interest). The pipeline (Fig. 1A) processes query  
51 sequences by: (i) identification to one of 8 segments using BLASTn; (ii) alignment to reference  
52 gene segment dataset; (iii) the inference of a maximum likelihood tree; (iv) classification to  
53 evolutionary lineage or genetic clade using patristic distance extracted from the inferred tree;  
54 and (v) generation of a summary classification file and annotated gene trees (Fig. 1B and 1C).  
55 The reference dataset for each gene includes non-swine genes allowing the pipeline to flag  
56 sequences that are not contemporary circulating U.S. swine IAV. Genes derived from  
57 interspecies transmission events are annotated by a non-swine classification, and reassortment  
58 events involving different lineages can be identified in the summary file as disparate lineages  
59 (e.g., a single strain containing a mix of human-seasonal, TRIG, and pandemic genes).  
60 Classification uses patristic distances extracted from gene trees using dendropy in Python (18).  
61 The shortest distance from a query gene to a reference gene is identified, and the reference  
62 gene annotation is assigned to the query. Using swine IAV data collected in the USA from 2014  
63 to present (929 strains, 7432 genes), the pipeline accurately captured classifications assigned  
64 by manual phylogenetic curation (7428 genes classified correct, 99.95% accuracy). Our  
65 approach is reliant upon a relevant reference dataset: the provided reference genes are  
66 adequate for swine IAV in the USA and Canada, but have limited utility for swine IAV in Europe  
67 and Asia. However, this tool maintains utility for international swine IAV researchers if they  
68 generate a custom reference dataset with appropriate clade or lineage annotation. Moreover, if  
69 interspecies transmission events result in the establishment of new lineages, contemporary data  
70 that captures this diversity may be added to the reference files by pipeline users or at the  
71 repository.

72       **Data availability.** Gene segment sequences were extracted from the Influenza  
73   Research Database (19). The pipeline and reference gene sets are provided on GitHub  
74   (<https://github.com/flu-crew/octoflu>), and DockerHub ([flucrew/octoflu](#)).

75   **ACKNOWLEDGMENTS**

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

90   **REFERENCES**

- 91   1. **Shope RE.** 1931. SWINE INFLUENZA : III. FILTRATION EXPERIMENTS AND  
92   ETIOLOGY. *J Exp Med* **54**:373–385.
- 93   2. **Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L, Yoon  
94   KJ, Krauss S, Webster RG.** 1999. Genetic reassortment of avian, swine, and human  
95   influenza A viruses in American pigs. *J Virol* **73**:8851–8856.
- 96   3. **Karasin AI, Schutten MM, Cooper LA, Smith CB, Subbarao K, Anderson GA,  
97   Carman S, Olsen CW.** 2000. Genetic characterization of H3N2 influenza viruses isolated  
98   from pigs in North America, 1977-1999: evidence for wholly human and reassortant virus  
99   genotypes. *Virus Res* **68**:71–85.

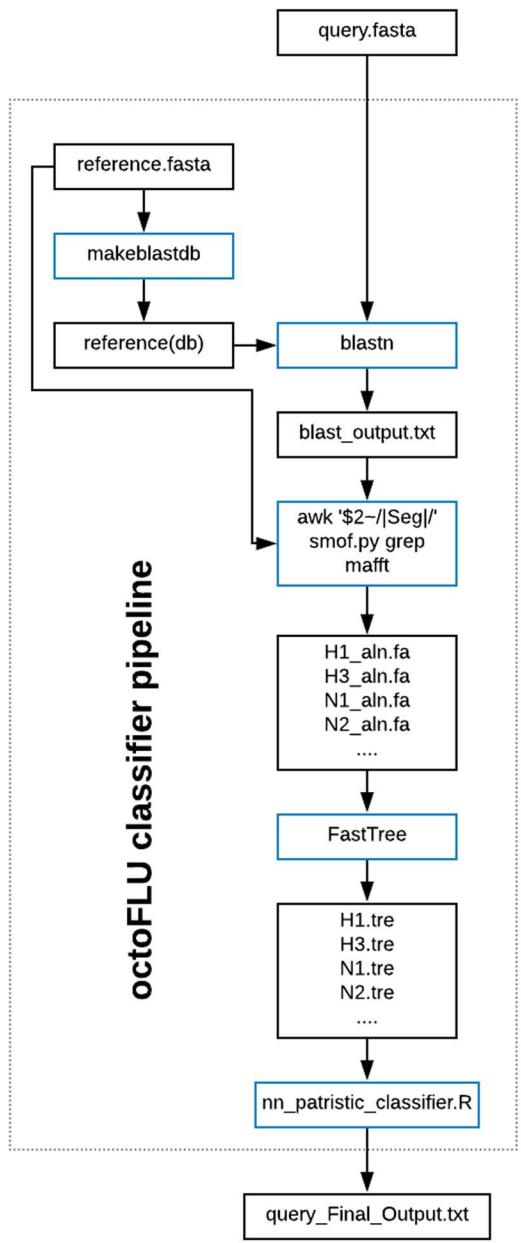
- 100 4. **Kitikoon P, Nelson MI, Killian ML, Anderson TK, Koster L, Culhane MR, Vincent AL.**  
101 2013. Genotype patterns of contemporary reassorted H3N2 virus in US swine. *J Gen*  
102 *Virol* **94**:1236–1241.
- 103 5. **Walia RR, Anderson TK, Vincent AL.** 2019. Regional patterns of genetic diversity in  
104 swine influenza A viruses in the United States from 2010 to 2016. *Influenza Other Respir*  
105 *Viruses* **2**:65–12.
- 106 6. **Karasin AI, Landgraf J, Swenson S, Erickson G, Goyal S, Woodruff M, Scherba G,**  
107 **Anderson G, Olsen CW.** 2002. Genetic characterization of H1N2 influenza A viruses  
108 isolated from pigs throughout the United States. *J Clin Microbiol* **40**:1073–1079.
- 109 7. **Vincent AL, Ma W, Lager KM, Gramer MR, Richt JA, Janke BH.** 2009.  
110 Characterization of a newly emerged genetic cluster of H1N1 and H1N2 swine influenza  
111 virus in the United States. *Virus Genes* **39**:176–185.
- 112 8. **Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu**  
113 **X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB,**  
114 **Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier**  
115 **RAM, Pappas C, Alpuche-Aranda CM, López-Gatell H, Olivera H, López I, Myers CA,**  
116 **Faix D, Blair PJ, Yu C, Keene KM, Dotson PD, Boxrud D, Sambol AR, Abid SH, St**  
117 **George K, Bannerman T, Moore AL, Stringer DJ, Blevins P, Demmler-Harrison GJ,**  
118 **Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA,**  
119 **Beatrice ST, Donis R, Katz J, Finelli L, Bridges CB, Shaw M, Jernigan DB, Uyeki TM,**  
120 **Smith DJ, Klimov AI, Cox NJ.** 2009. Antigenic and genetic characteristics of swine-  
121 origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* **325**:197–201.
- 122 9. **Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK,**  
123 **Cheung CL, Raghwanji J, Bhatt S, Peiris JSM, Guan Y, Rambaut A.** 2009. Origins and  
124 evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*  
125 **459**:1122–1125.
- 126 10. **Zeller MA, Li G, Harmon KM, Zhang J, Vincent AL, Anderson TK, Gauger PC.** 2018.  
127 Complete Genome Sequences of Two Novel Human-Like H3N2 Influenza A Viruses,  
128 A/swine/Oklahoma/65980/2017 (H3N2) and A/Swine/Oklahoma/65260/2017 (H3N2),  
129 Detected in Swine in the United States. *Microbiol Resour Announc* **7**:41–2.
- 130 11. **Rajão DS, Gauger PC, Anderson TK, Lewis NS, Abente EJ, Killian ML, Perez DR,**  
131 **Sutton TC, Zhang J, Vincent AL.** 2015. Novel Reassortant Human-Like H3N2 and  
132 H3N1 Influenza A Viruses Detected in Pigs Are Virulent and Antigenically Distinct from  
133 Swine Viruses Endemic to the United States. *J Virol* **89**:11213–11222.
- 134 12. **Nelson MI, Lemey P, Tan Y, Vincent A, Lam TT-Y, Detmer S, Viboud C, Suchard MA,**  
135 **Rambaut A, Holmes EC, Gramer M.** 2011. Spatial dynamics of human-origin H1  
136 influenza A virus in North American swine. *PLoS Pathog* **7**:e1002077.
- 137 13. **Anderson TK, Campbell BA, Nelson MI, Lewis NS, Janas-Martindale A, Killian ML,**  
138 **Vincent AL.** 2015. Characterization of co-circulating swine influenza A viruses in North  
139 America and the identification of a novel H1 genetic clade with antigenic significance.  
140 *Virus Res* **201**:24–31.

- 141 14. **Anderson TK, Nelson MI, Kitikoon P, Swenson SL, Korslund JA, Vincent AL.** 2013.  
142 Population dynamics of cocirculating swine influenza A viruses in the United States from  
143 2009 to 2012. *Influenza Other Respir Viruses* **7 Suppl 4**:42–51.
- 144 15. **Richt JA, Lekcharoensuk P, Lager KM, Vincent AL, Loiacono CM, Janke BH, Wu W-H, Yoon K-J, Webby RJ, Solórzano A, García-Sastre A.** 2006. Vaccination of pigs  
145 against swine influenza viruses by using an NS1-truncated modified live-virus vaccine. *J Virol* **80**:11009–11018.
- 146 148 16. **Vincent AL, Perez DR, Rajao D, Anderson TK, Abente EJ, Walia RR, Lewis NS.**  
149 2017. Influenza A virus vaccines for swine. *Vet Microbiol* **206**:35–44.
- 150 17. **Gao S, Anderson TK, Walia RR, Dorman KS, Janas-Martindale A, Vincent AL.** 2017.  
151 The genomic evolution of H1 influenza a viruses from swine detected in the United States  
152 between 2009 and 2016. *J Gen Virol* **98**:2001–2010.
- 153 18. **Sukumaran J, Holder MT.** 2010. DendroPy: a Python library for phylogenetic computing.  
154 *Bioinformatics* **26**:1569–1571.
- 155 19. **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.

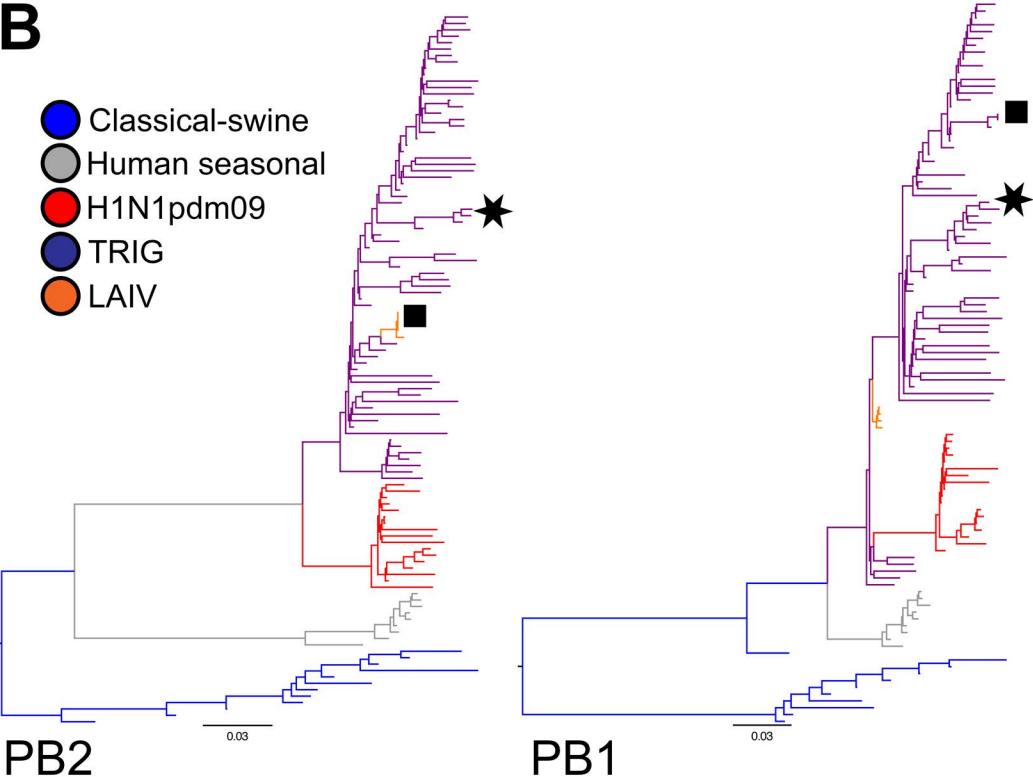
161

162 **FIGURE LEGEND**

163 **Figure 1.** The octoFLU classifier pipeline (A), PB2 and PB1 inferred maximum likelihood trees  
164 generated with 2 query strains including the reference gene sequences (B), and example  
165 summary output generated for two contemporary US swine influenza A genomes (C). The PB1  
166 and PB2 gene tree examples demonstrate the genetic lineages of contemporary influenza A  
167 virus circulating in United States swine populations: the H1N1 pandemic 2009 (red) and LAIV  
168 genes (orange) are monophyletic clades nested within the TRIG lineage (purple), human  
169 seasonal (grey), and classical-swine (blue) lineage genes are separate monophyletic clades.  
170 The query genes for A/swine/Nebraska/A02170137/2018 are labeled with a black star: the  
171 query genes for A/swine/Oklahoma/A01785571/2018 are labeled by a black square. The trees  
172 are midpoint rooted for clarity; branch lengths are drawn to scale; and the scale bar indicates  
173 the number of nucleotide substitutions per site.

**A****B**

- Classical-swine
- Human seasonal
- H1N1pdm09
- TRIG
- LAIV

**C**

Query name	Gene	Clade or lineage	Global HA clade
MH608115 A/swine/Oklahoma/A01785571/2018	H1	gamma2-beta-like	1A.2-3-like
MH595478 A/swine/Nebraska/A02170137/2018	H1	gamma	1A.3.3.3
MH595479 A/swine/Nebraska/A02170137/2018	N1	classicalSwine	
MH608116 A/swine/Oklahoma/A01785571/2018	N1	LAIV-Classical	
MK185309 A/swine/Nebraska/A02170137/2018	PB2	TRIG	
MH992370 A/swine/Oklahoma/A01785571/2018	PB2	LAIV	
MK185310 A/swine/Nebraska/A02170137/2018	PB1	TRIG	
MH992371 A/swine/Oklahoma/A01785571/2018	PB1	TRIG	
MH992372 A/swine/Oklahoma/A01785571/2018	PA	LAIV	
MK185311 A/swine/Nebraska/A02170137/2018	PA	pdm	
MK185312 A/swine/Nebraska/A02170137/2018	NP	pdm	
MH992373 A/swine/Oklahoma/A01785571/2018	NP	LAIV	
MK185313 A/swine/Nebraska/A02170137/2018	M	pdm	
MH992374 A/swine/Oklahoma/A01785571/2018	M	pdm	
MK185314 A/swine/Nebraska/A02170137/2018	NS	TRIG	
MH992375 A/swine/Oklahoma/A01785571/2018	NS	TRIG	