1   OctoFLU: Automated classification to evolutionary origin of influenza A virus gene

2   sequences detected in U.S. swine

3

4   Jennifer Chang,[a,*] Tavis K. Anderson,[a,*] Michael A. Zeller,[a,b,c,*] Phillip C. Gauger,[b] Amy

5   L. Vincent[a,#]

6

7   [a]Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames,

8   Iowa, USA

9   [b]Department of Veterinary Diagnostic and Production Animal Medicine, College of

10 Veterinary Medicine, Iowa State University, Ames, Iowa, USA

11 [c]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: **51/50** words; main text: **599/500** words

20   **ABSTRACT**

21   Influenza A viruses (IAV) in swine are classified as H1N1, H1N2, or H3N2 subtype, with

22   diversity of the 8 genes reflecting introductions from non-swine hosts and subsequent antigenic

23   drift and shift. Here, we curated a dataset and present a pipeline that assigns evolutionary

24   lineage and genetic clade to query gene segments.

25   Influenza A virus (IAV) is a negative-sense, single-stranded, enveloped RNA virus of the

26   *Orthomyxoviridae* family. Although only H1N1, H1N2, and H3N2 subtypes are endemic in swine

27   around the world, much diversity can be found in the genes coding for major surface proteins,

28   hemagglutinin (HA) and neuraminidase (NA), and in the other 6 internal gene segments. This

29   diversity is the result of bidirectional transmission between swine and humans, the occasional

30   transmission of an avian virus into swine, followed by periods of antigenic drift and shift.

31   Swine IAV emerged coincident with the 1918 Spanish flu, and genes derived from this

32   lineage are classified as classical-swine H1N1 (1). In the late 1990s, triple-reassortant H3N2

33   viruses were identified containing gene segments derived from seasonal human H3N2 (HA, NA,

34   and PB1), avian IAV (PB2 and PA), and the classical H1N1 swine IAV (NP, M, and NS) (2, 3).

35   The HA persisted, evolving into phylogenetic clades that are detected to present day (Cluster-IV

36   (C-IV) clades A-F) (4). The triple-reassortant H3N2 viruses also reassorted with classical-swine

37   H1N1 viruses, driving diversification and new genetic clades of H1N1 and H1N2 viruses (5), but

38   preserving the triple reassortant internal gene (TRIG) constellation. Genetically distinct human

39   seasonal H1 also spilled into and established in swine in the early 2000s (6, 7). In 2009, a virus

40   with NA and M genes from Eurasian-avian H1N1 swine in addition to TRIG and classical-swine

41   lineage genes emerged in swine, and infected humans as a pandemic (H1N1pdm09). Although

42   sharing common ancestors, the human H1N1pdm09 genes were phylogenetically distinct from

43   contemporary swine IAV in the U.S.. Via reverse zoonoses, the H1N1pdm09 continues to

44   contribute to genetic diversity in swine, particularly the internal gene segments (8, 9). More

45   recently, a human H3N2 virus was transmitted to swine, H3.2010.1, this virus is distinct from the

2

46   H3N2 lineage C-IV viruses (10). In the U.S., HA genes are paired with N2 genes derived from

47   the 1998 or 2002 human seasonal origin (11), or an N1 gene from the classical-swine lineage or

48   pandemic-lineage (12, 13). In 2018, a live-attenuated influenza virus (LAIV) vaccine became

49   commercially available in the U.S. (14). The LAIV uses HA (H1 and H3) and NA (N1 and N2)

50   expressed on a TRIG internal gene backbone, with all components isolated from swine in the

51   1990s. These LAIV genes are distinct from contemporary IAV and reassorted viruses have been

52   detected with vaccine-derived internal genes (annotated as LAIV), and surface genes

53   (annotated as H3 Cluster-I, H1 gamma2-beta-like, N2 LAIV-98, N1 LAIV-Classical). Thus,

54   interspecies transmission episodes and the processes of antigenic shift and drift have led to

55   approximately 16 distinct HA clades, 4 NA lineages, and 3 internal gene lineages (15, 16).

56   We generated reference gene datasets and an analytical pipeline that assigns query HA

57   to genetic clade and NA and internal IAV genes to evolutionary lineages that are commonly

58   found in IAV from North American swine. The pipeline can aid in the detection of interspecies

59   transmission or reassorted viruses containing gene segments derived from avian, human, or

60   other sources. Users need (i) the reference datasets; and (ii) a FASTA with query sequences

61   from any IAV gene segment. The pipeline (Fig. 1) processes query sequences by: (i)

62   identification to one of 8 segments using BLASTn; (ii) alignment of queries to reference gene

63   segment dataset; (iii) the inference of a maximum likelihood tree; (iv) classification to

64   evolutionary lineage or genetic clade of queries using patristic distance extracted from the

65   inferred tree; and (v) generation of a summary classification file. The reference dataset for each

66   gene includes outgroup sequences allowing the pipeline to flag sequences that are not

67   contemporary circulating U.S. swine IAV.

68   **Data availability.** Gene segment sequences were extracted from the Influenza

69   Research Database (17). The pipeline and reference gene sets are provided on GitHub

70   (https://github.com/flu-crew/octoflu), and DockerHub (flucrew/octoflu).

71   **ACKNOWLEDGMENTS**

86   **REFERENCES**

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

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

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

96   4.   **Kitikoon P**, **Nelson MI**, **Killian ML**, **Anderson TK**, **Koster L**, **Culhane MR**, **Vincent AL**.
97   2013. Genotype patterns of contemporary reassorted H3N2 virus in US swine. J Gen
98   Virol **94**:1236–1241.

99   5.   **Walia RR**, **Anderson TK**, **Vincent AL**. 2019. Regional patterns of genetic diversity in
100   swine influenza A viruses in the United States from 2010 to 2016. Influenza Other Resp
101   Viruses **2**:65–12.

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

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

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

118 9.   **Smith GJD**, **Vijaykrishna D**, **Bahl J**, **Lycett SJ**, **Worobey M**, **Pybus OG**, **Ma SK**,
119   **Cheung CL**, **Raghwani J**, **Bhatt S**, **Peiris JSM**, **Guan Y**, **Rambaut A**. 2009. Origins and
120   evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature
121   **459**:1122–1125.

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

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

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

133 13.   **Anderson TK**, **Nelson MI**, **Kitikoon P**, **Swenson SL**, **Korslund JA**, **Vincent AL**. 2013.
134   Population dynamics of cocirculating swine influenza A viruses in the United States from
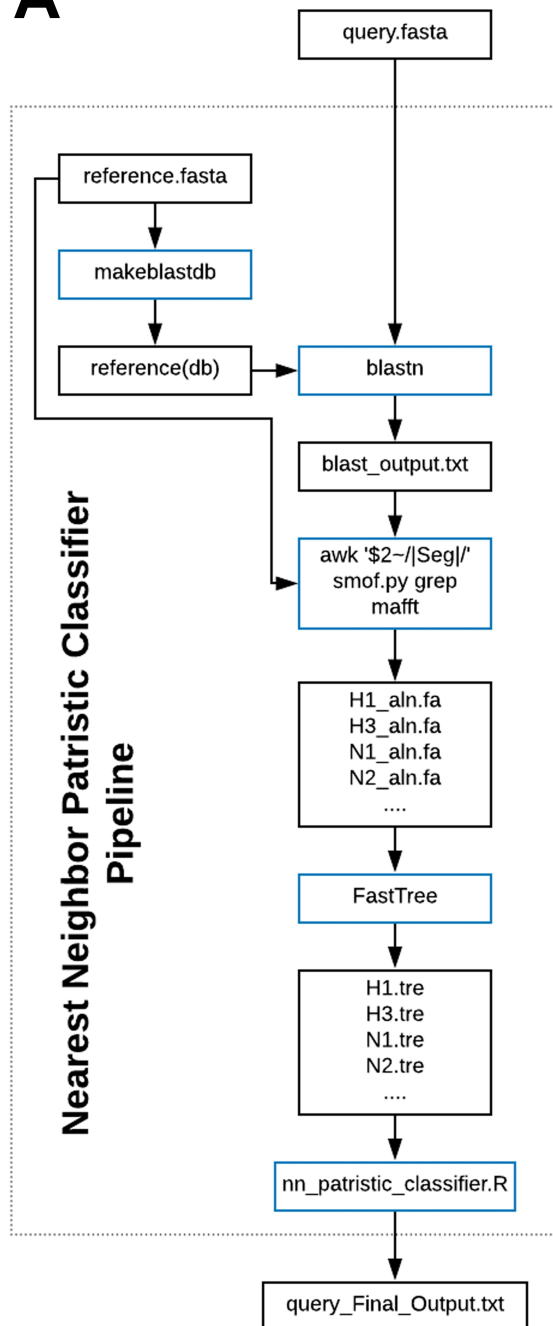135   2009 to 2012. Influenza Other Resp Viruses **7 Suppl 4**:42–51.

136 14.   **Richt JA**, **Lekcharoensuk P**, **Lager KM**, **Vincent AL**, **Loiacono CM**, **Janke BH**, **Wu W-**
137   **H**, **Yoon K-J**, **Webby RJ**, **Solórzano A**, **García-Sastre A**. 2006. Vaccination of pigs
138   against swine influenza viruses by using an NS1-truncated modified live-virus vaccine. J
139   Virol **80**:11009–11018.

140 15.   **Vincent AL**, **Perez DR**, **Rajao D**, **Anderson TK**, **Abente EJ**, **Walia RR**, **Lewis NS**.
141   2017. Influenza A virus vaccines for swine. Vet Microbiol **206**:35–44.

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

145 17. **Zhang Y**, **Aevermann BD**, **Anderson TK**, **Burke DF**, **Dauphin G**, **Gu Z**, **He S**, **Kumar**
146   **S**, **Larsen CN**, **Lee AJ**, **Li X**, **Macken C**, **Mahaffey C**, **Pickett BE**, **Reardon B**, **Smith T**,
147   **Stewart L**, **Suloway C**, **Sun G**, **Tong L**, **Vincent AL**, **Walters B**, **Zaremba S**, **Zhao H**,
148   **Zhou L**, **Zmasek C**, **Klem EB**, **Scheuermann RH**. 2017. Influenza Research Database:
149   An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res
150   **45**:D466–D474.

151

152 **FIGURE LEGEND**

153 **Figure 1.** The nearest neighbor patristic classifier pipeline (A) and an inferred maximum

154 likelihood tree generated on query and reference PB2 sequences (B). The PB2 gene example

155 demonstrates the genetic lineages of contemporary influenza A virus circulating in United States

156 swine populations: the H1N1 pandemic 2009 (red) and LAIV PB2 genes (orange) are

157 monophyletic clades nested within the TRIG lineage (purple), human seasonal (grey), and

158 classical-swine (blue) lineage PB2 genes are separate monophyletic clades. The tree is

159 midpoint rooted for clarity; branch lengths are drawn to scale; and the scale bar indicates the

160 number of nucleotide substitutions per site.

A

query.fasta

Nearest Neighbor Patristic Classifier Pipeline

reference.fasta → makeblastdb → reference(db) → blastn → blast_output.txt → awk '$2~/|Seg|/' smof.py grep mafft → H1_aln.fa H3_aln.fa N1_aln.fa N2_aln.fa .... → FastTree → H1.tre H3.tre N1.tre N2.tre .... → nn_patristic_classifier.R → query_Final_Output.txt

B

TRIG lineage

LAIV

H1N1pdm09

Classical-swine lineage

0.03