



octoFLU: Automated Classification for the Evolutionary Origin of Influenza A Virus Gene Sequences Detected in U.S. Swine

 Jennifer Chang,^a  Tavis K. Anderson,^a Michael A. Zeller,^{a,b,c}  Phillip C. Gauger,^b  Amy L. Vincent^a

^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

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

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

We generated reference gene data sets and an analytical pipeline that assigns queried HA to genetic clade and queried NA and internal IAV genes to evolutionary lineages that are found in IAV from U.S. swine. Users need the reference data set and a FASTA file with query sequences from any IAV gene segment. The input data must be of good quality and substantial length (approximately 50% or greater of the gene of interest). The pipeline (Fig. 1A) processes query sequences by (i) identification to one of 8 segments using BLASTn, (ii) alignment to the reference gene segment data set, (iii) the inference of a maximum likelihood tree, (iv) classification to evolutionary lineage or genetic clade using patristic distance extracted from the inferred tree, and (v) gener-

Citation Chang J, Anderson TK, Zeller MA, Gauger PC, Vincent AL. 2019. octoFLU: automated classification for the evolutionary origin of influenza A virus gene sequences detected in U.S. swine. *Microbiol Resour Announc* 8:e00673-19. <https://doi.org/10.1128/MRA.00673-19>.

Editor Irene L. G. Newton, Indiana University, Bloomington

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to Amy L. Vincent, amy.vincent@ars.usda.gov.

J.C., T.K.A., and M.A.Z. contributed equally to this work.

Received 4 June 2019

Accepted 18 July 2019

Published 8 August 2019

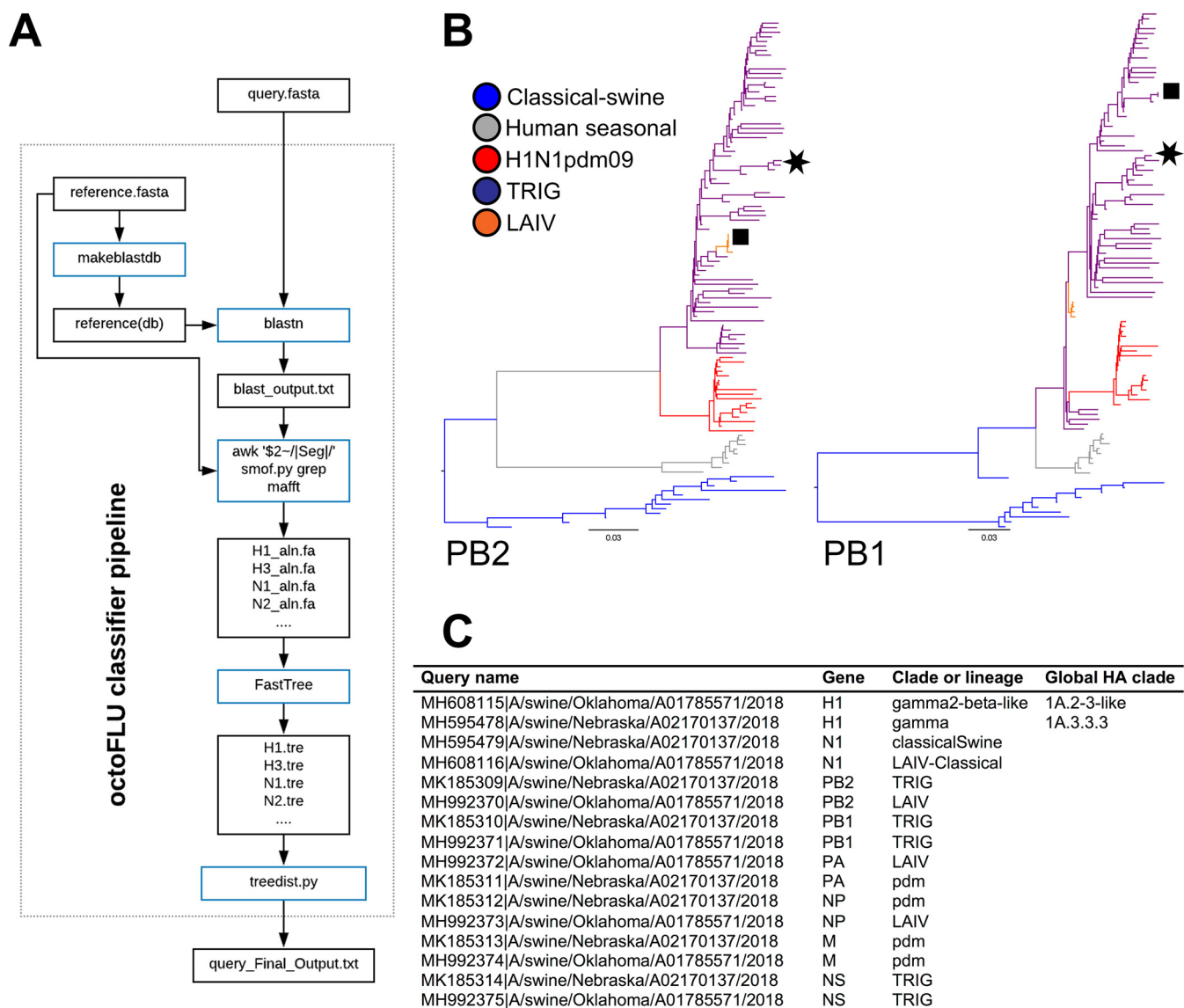


FIG 1 The octoFLU classifier pipeline (A), PB2 and PB1 inferred maximum likelihood trees generated with 2 query strains, including the reference gene sequences (B), and example summary output generated for two contemporary U.S. swine influenza A genomes (C). The PB1 and PB2 gene tree examples demonstrate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

ation of a summary classification file and annotated gene trees (Fig. 1B and C). The reference data set for each gene includes nonswine genes, allowing the pipeline to flag sequences that are not contemporary circulating U.S. swine IAV. Genes derived from interspecies transmission events are annotated by a nonswine classification, and reassortment events involving different lineages can be identified in the summary file as disparate lineages (e.g., a single strain containing a mix of human seasonal, TRIG, and pandemic genes). Classification uses patristic distances extracted from gene trees using DendroPy in Python (18) and smof for processing FASTA files (19). The shortest distance from a query gene to a reference gene is identified, and the reference gene annotation is assigned to the query. Using swine IAV data collected in the United States from 2014 to present (929 strains and 7,432 genes), the pipeline accurately captured classifications assigned by manual phylogenetic curation (7,428 genes classified correctly; 99.95%

accuracy). Our approach is reliant upon a relevant reference data set; the provided reference genes are adequate for swine IAV in the United States and Canada but have limited utility for swine IAV in Europe and Asia. However, this tool maintains utility for international swine IAV researchers if they generate a custom reference data set with appropriate clade or lineage annotation. Moreover, if interspecies transmission events result in the establishment of new lineages, contemporary data that capture this diversity may be added to the reference files by pipeline users or at the repository.

Data availability. Gene segment sequences were extracted from the Influenza Research Database (20). The pipeline and reference gene sets are provided on GitHub (<https://github.com/flu-crew/octoflu>) and DockerHub (<https://hub.docker.com/r/flucrew/octoflu>).

ACKNOWLEDGMENTS

We 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). J.C. and T.K.A. 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. P.C.G. and M.A.Z. were supported by the Iowa State University Veterinary Diagnostic Laboratory and by an Iowa State University 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.

REFERENCES

- Shope RE. 1931. Swine influenza: III. Filtration experiments and etiology. *J Exp Med* 54:373–385. <https://doi.org/10.1084/jem.54.3.373>.
- 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.
- 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 Res* 68:71–85. [https://doi.org/10.1016/S0168-1702\(00\)00154-4](https://doi.org/10.1016/S0168-1702(00)00154-4).
- 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. <https://doi.org/10.1099/vir.0.51839-0>.
- Walia RR, Anderson TK, Vincent AL. 2019. Regional patterns of genetic diversity in swine influenza A viruses in the United States from 2010 to 2016. *Influenza Other Respir Viruses* 2:65–72. <https://doi.org/10.1111/irv.12559>.
- 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. <https://doi.org/10.1128/jcm.40.3.1073-1079.2002>.
- 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. <https://doi.org/10.1007/s11262-009-0386-6>.
- 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, Rivaller 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. <https://doi.org/10.1126/science.1176225>.
- 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. <https://doi.org/10.1038/nature08182>.
- Zeller MA, Li G, Harmon KM, Zhang J, Vincent AL, Anderson TK, Gauger PC. 2018. Complete genome sequences of two novel human-like H3N2 influenza A viruses, A/swine/Oklahoma/65980/2017 (H3N2) and A/swine/Oklahoma/65260/2017 (H3N2), detected in swine in the United States. *Microbiol Resour Announc* 7:41–42. <https://doi.org/10.1128/MRA.01203-18>.
- 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. <https://doi.org/10.1128/JVI.01675-15>.
- 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. <https://doi.org/10.1371/journal.ppat.1002077>.
- 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 Res* 201:24–31. <https://doi.org/10.1016/j.virusres.2015.02.009>.

14. 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 Respi Viruses* 7(Suppl 4):42–51. <https://doi.org/10.1111/irv.12193>.
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 against swine influenza viruses by using an NS1-truncated modified live-virus vaccine. *J Virol* 80:11009–11018. <https://doi.org/10.1128/JVI.00787-06>.
16. 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. <https://doi.org/10.1016/j.vetmic.2016.11.026>.
17. 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. <https://doi.org/10.1099/jgv.0.000885>.
18. Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26:1569–1571. <https://doi.org/10.1093/bioinformatics/btq228>.
19. Arendsee Z, Chang J, Leung E. 2018. *incertae-sedis/smf: first release* (version 2.13.1). <https://doi.org/10.5281/zenodo.1434656>.
20. 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. <https://doi.org/10.1093/nar/gkw857>.