

Phylogeography of the Highly Pathogenic Avian Influenza H5N6 Virus Determine Transmission from Guangdong Province, China to South Korea and Japan by Wild Bids in 2016

Jung-hoon Kwon¹

¹Kyungpook National University

November 6, 2020

Abstract

The Goose/Guangdong lineage highly pathogenic avian influenza (HPAI) clade 2.3.4.4 viruses have spilled over to wild birds and caused transboundary epizootics. In 2016, the H5N6 subtype of clade 2.3.4.4 HPAI viruses was introduced to Korea and Japan by wild waterfowls. However, the detailed geographic location of the origin of this outbreak was not fully defined yet. In this study, the geographical origin of this outbreak was estimated using phylogeographic analysis. The results indicate that the H5N6 viruses of wild birds originated from a virus in Guangdong province, China. In addition, Guangdong province was an epicenter for H5N6 viruses, and multiple spillovers of viruses from poultry into wild birds have been detected. Thus, to predict and prepare for the next HPAI outbreak in Korea and Japan, viruses in Guangdong province should be carefully monitored and investigated.

Short communication

Title: Phylogeography of the Highly Pathogenic Avian Influenza H5N6 Virus Determine Transmission from Guangdong Province, China to South Korea and Japan by Wild Bids in 2016

Running Title : Phylogeographic Analysis of H5N6 HPAIV

Author: Jung-Hoon Kwon

Affiliation: College of Veterinary Medicine, Kyungpook National University, Daegu, 41566, Republic of Korea.

Corresponding Author

Jung-Hoon Kwon

College of Veterinary Medicine, Kyungpook National University

Daegu, 41566, Republic of Korea.

E-mail: Junghoon.kwon@knu.ac.kr

Phone: +82.53.950.5968

Summary

The Goose/Guangdong lineage highly pathogenic avian influenza (HPAI) clade 2.3.4.4 viruses have spilled over to wild birds and caused transboundary epizootics. In 2016, the H5N6 subtype of clade 2.3.4.4 HPAI viruses was introduced to Korea and Japan by wild waterfowls. However, the detailed geographic location of the origin of this outbreak was not fully defined yet. In this study, the geographical origin of this outbreak

was estimated using phylogeographic analysis. The results indicate that the H5N6 viruses of wild birds originated from a virus in Guangdong province, China. In addition, Guangdong province was an epicenter for H5N6 viruses, and multiple spillovers of viruses from poultry into wild birds have been detected. Thus, to predict and prepare for the next HPAI outbreak in Korea and Japan, viruses in Guangdong province should be carefully monitored and investigated.

Keywords : H5N6, highly pathogenic avian influenza virus, phylogeography, wild birds

Introduction

The Goose/Guangdong (Gs/Gd) lineage highly pathogenic avian influenza viruses (HPAIVs) have been circulating in poultry in endemic countries since it was first detected in 1996 (Swayne, 2016; Swayne, Suarez & Sims, 2020). The HPAIV spillovers from poultry into wild birds have caused long-distance transmission beyond national borders (Lee et al., 2015; Olsen et al., 2006). Transmission of HPAIVs by wild birds has caused significant economic loss in the poultry industry, because most non-endemic countries do not use vaccines and instead have applied a stamping-out policy for the eradication of the virus (Swayne, Pavade, Hamilton, Vallat, & Miyagishima, 2011).

In South Korea and Japan, a total of seven HPAIV outbreaks have been reported: 2003 (H5N1 clade 2.5), 2006–2007 (H5N1 clade 2.2), 2008 (H5N1 clade 2.3.2.), 2010–2011 (H5N1 clade 2.3.2.1), 2014–2016 (H5N8 clade 2.3.4.4), 2016–2017 (H5N6 and H5N8, clade 2.3.4.4) and 2017–2018 (H5N6, clade 2.3.4.4) (Kwon et al., 2018; Kwon et al., 2017; Mo et al., 2016; Woo et al., 2017). In addition, these viruses disseminated to Europe, North America, and Africa by migratory wild waterfowls (Lee, Bertran, Kwon, & Swayne, 2017; Lee et al., 2015; Olsen et al., 2006). Previous studies using phylogenetic analysis indicated that these viruses originated from endemic regions, including China (Kwon et al., 2017; Lee et al., 2017; Olsen et al., 2006).

However, the detailed geographical origin and transmission of these outbreaks had not fully been determined, because only a few nucleotide sequences were available in the database. For instance, only nine hemagglutinin sequences of H5N8 viruses were available in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) before 2104, when it was first detected in Korea (Lee et al., 2014). Recently, however, the implementation of next-generation sequencing systems has allowed for high-throughput sequence analyses of avian influenza viruses, so there are a considerable number of sequences available now (Lee, 2020). In particular, clade 2.3.4.4 H5N6 HPAIVs has been a dominant virus in wide geographic regions in China since 2014, and numerous nucleotide sequences of these viruses have been analyzed (Bi et al., 2016).

Using these sequencing data, in this study, we inferred the geographical origin of the H5N6 outbreak in South Korea and Japan and estimated the time for H5N6 virus spillover from poultry into wild birds for this outbreak.

Materials and Methods

All available hemagglutinin (HA) gene sequences of H5N6 viruses belonging to clade 2.3.4.4 subgroup C ($n = 1388$) were downloaded from the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database (<https://www.gisaid.org/>) on August 13, 2020. After excluding the sequences that did not have an exact collection date or region, we reduced the data set based on the sequence identity at a 99.9% level using the program cd-hit (Li & Godzik, 2006). The reduced data set consisted of 679 HA sequences, and the sequences were aligned by the Multiple Alignment using Fast Fourier Transform (MAFFT) method (Katoh, Misawa, Kuma, & Miyata, 2002) and were manually trimmed to equal lengths (1674 bp) using Geneious Prime (<https://www.geneious.com/>). After phylogenetic analysis using FastTree (Price, Dehal, & Arkin, 2009) (Figure S1), the sequences of a subclade related to the South Korea and Japan outbreak showed >0.9 FastTree support value was extracted ($n = 241$), which was used for the final data set. The taxa were coded into nine discrete groups: Fujian, Guangdong, Hunan, Japan, Jiangxi, Korea, Vietnam, Wild bird, and Yunnan. The Wild bird classification included all wild bird isolates regardless of location.

The phylogeography in discrete space was estimated using BEAST v.1.10.4 (<http://beast.bio.ed.ac.uk>) with an uncorrelated lognormal relaxed molecular clock. The HKY + G nucleotide substitution model (Hasegawa,

Kishino, & Yano, 1985) was used along with a Gaussian Markov random field (GMRF) Bayesian Skyride coalescent tree (Minin, Bloomquist, & Suchard, 2008). The ancestral locations were reconstructed, and the asymmetric viral exchanges between regions were estimated using the non-reversible continuous-time Markov chain (CTMC) model. We applied a Bayesian stochastic search variable selection (BSSVS) procedure to identify the best-supported transitions between discrete states using Bayes factors test, which is a feature of SPREAD software v1.0.7. We identified a transition as significant when the posterior probability was >0.5 and the Bayesian factor (BF) was >6 .

The Markov chain Monte Carlo (MCMC) method was run in parallel for four chains, each with 100 million steps, and the parameters and trees were sampled every 10,000 steps. After burn-in first 10% of each run, samples across the chains combined, yielding a total of 36,000 parameter states and posterior trees. The parameters were analyzed with TRACER v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>), which showed that all parameters had an effective sample size greater than 200. A maximum clade credibility (MCC) tree was generated using TreeAnnotator in BEAST and visualized using FigTree 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). We also estimated the rate and number of transitions between hosts (Markov jumps) using stochastic mapping techniques implemented in the BEAST package (Minin & Suchard, 2008).

Results

The phylogeographic analysis was conducted against viruses belonging to a subgroup of clade 2.3.4.4C viruses, including those of the South Korea and Japan outbreak ($n = 241$; Figure S1). The emergence time of this subgroup was estimated to be February 2014 (95% Bayesian credibility intervals, BCI: October 2013 to March 2014), and the geographic location of the ancestor of this subgroup was estimated to be Guangdong province with 97.87% posterior probability (Figure 1 and Figure S2).

All the poultry isolates of South Korea and Japan were clustered with wild bird isolates. The ancestral reconstruction and the Bayes factor analysis indicated that all HPAIVs of the South Korea and Japan outbreak originated from H5N6 HPAIVs of wild birds (Figure 1, Figure S2, and Table 1). At least thirteen independent introduction events from wild birds into poultry in South Korea and Japan were detected in this study (Figure 1, Table 1). Four Vietnam isolates were also clustered with the South Korean, Japanese, and wild bird viruses, indicating at least two independent introduction events from wild birds to Vietnam (Figure 1 and Figure S2). The transmission from poultry to wild birds or between poultry in different countries was not detected.

The wild bird HPAI isolates in South Korea, Japan, and Shanghai were clustered with Korean, Japanese, and Vietnamese viruses. HPAIV infections in wild birds were also detected from Hong Kong, Guangdong, Hunan, and Vietnam, but these were phylogenetically distinct from the South Korea and Japan isolates (Figure 1 and Figure S2). However, in this phylogeographic analysis, the ancestral locations of all the wild bird isolates were traced to Guangdong province, showing >0.9 posterior probability (Figure 1 and Figure S2). The Markov jump count from Guangdong to the wild birds was estimated with a mean of 6.89 (95% BCI: 6–8) in this study (Table 1).

The time period of the most recent common ancestor (tMRCA) of the South Korea and Japan outbreak and related wild bird viruses (Figure 1) was estimated to be March 2016 (95% BCI: January to April 2016). The tMRCA of the Guangdong viruses and wild bird viruses that caused the South Korea and Japan outbreak was estimated to be October 2015 (95% BCI: September to November 2015), which indicates that the spillover of HPAIV from poultry to wild birds occurred during this period. The first real-world detection of wild bird infection with these viruses was October 28, 2016 (J. H. Kwon et al., 2017), which shows a one year interval from the estimated spillover time.

Discussion

Despite repeated transboundary transmission of HPAIVs, the detailed geographic origins and transmission routes of the evolution and epidemiology of the viruses have not fully been determined yet. The migration

of wild birds, which is a major contributor to long-distance HPAIV transmission, is highly complex, unpredictable, and impacted by climate change (Fusaro et al., 2019). In this study, we estimated the origin of the 2016–2017 South Korea and Japan H5N6 HPAI outbreak using Bayesian phylogenetic analysis.

Guangdong province, China was estimated to be the site for wild bird infection of H5N6 HPAIV by poultry in this study. Guangdong province is known as one of the major regions for duck and geese production in China, and a high incidence of H5N6 HPAIV infection was detected in a previous study (Bi et al., 2016). In addition, a previous phylogeographic study showed that Guangdong province and southern China were the epicenters for H5N6 HPAIV in China (Zhang et al., 2020). These results of the previous studies support the finding in this study that wild bird H5N6 viruses originated from Guangdong province, China.

The tMRCA analysis result indicated that H5N6 HPAIV spilled over to wild birds around October 2015, and then circulated in the wild bird population for approximately one year. However, the virus was first detected in October 2016. A previous experimental infection study suggests possible subclinical infection and replication of H5N6 HPAIV in Mandarin ducks (*Aix galericulata*) (Son et al., 2018). In addition, simultaneous detection of multiple reassortants of H5N6 HPAIVs in South Korea and Japan also supports the possibility of long-term silent circulation of H5N6 HPAIVs in wild birds (Lee et al., 2018; Takemae et al., 2017). The results suggest that spring migration of wild birds in 2016 caused the dissemination of viruses to other population of wild birds, and fall migration in 2016 introduced the virus into South Korea, Japan, and Shanghai.

The Bayesian phylogenetic inference and ancestral reconstruction methods used in this study can be affected by sampling bias (Kwon et al., 2020). To minimize the effect of sampling bias, the sequences were equally subsampled by sequence identity using all available sequences in the database. However, each province in China has a different surveillance system and sampling frequency. Thus, the potential role of un-sampled or less-sampled viruses cannot be completely excluded. Nevertheless, this study presents the most reliable transmission route using all available data, and the results strongly indicate the significant role of poultry in Guangdong province, China, for the H5N6 outbreak in wild birds. Continuous, constant, and regular surveillance and a nucleotide sequencing system will improve the quality of epidemiological analysis of HPAIV, including the molecular epidemiology.

Clade 2.3.4.4 H5N6 HPAIVs are causing continuous outbreaks and emergence of novel reassortants in endemic countries. This study recommends careful monitoring and control of HPAIVs in Guangdong province to prevent wild bird infection and transboundary transmission of HPAIVs. Although the emergence and spread of novel viruses are unpredictable, the accumulation of epidemiologic data, such as that reported in this study, will allow prediction and preparation that can be taken to minimize the economic loss caused by epidemics in the event of future HPAIV outbreaks.

Acknowledgments

We gratefully acknowledge the authors and submitting laboratories of the sequences from GISAID's EpiFlu (<https://www.gisaid.org/>) database. The GISAID acknowledgement tables for laboratory contributions are shown in Table S1.

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1G1A1101032).

Conflict of Interest: None

References

- Bi, Y., Chen, Q., Wang, Q., Chen, J., Jin, T., Wong, G., . . . Gao, G. F. (2016). Genesis, Evolution and Prevalence of H5N6 Avian Influenza Viruses in China. *Cell Host Microbe*, 20 (6), 810-821. doi:10.1016/j.chom.2016.10.022
- Fusaro, A., Zecchin, B., Vrancken, B., Abolnik, C., Ademun, R., Alassane, A., . . . Monne, I. (2019). Disentangling the role of Africa in the global spread of H5 highly pathogenic avian influenza. *Nat Commun*,

10 (1), 5310. doi:10.1038/s41467-019-13287-y

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*, 22 (2), 160-174. doi:10.1007/BF02101694

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, 30 (14), 3059-3066. doi:10.1093/nar/gkf436

Kwon, J.-H., Lee, D.-H., Criado, M. F., Killmaster, L., Ali, M. Z., Giasuddin, M., & Swayne, D. E. (2020). Genetic evolution and transmission dynamics of clade 2.3.2.1a highly pathogenic avian influenza A/H5N1 viruses in Bangladesh. *Virus Evol*, 6 (2). doi:10.1093/ve/veaa046

Kwon, J. H., Jeong, S., Lee, D. H., Swayne, D. E., Kim, Y. J., Lee, S. H., . . . Song, C. S. (2018). New Reassortant Clade 2.3.4.4b Avian Influenza A(H5N6) Virus in Wild Birds, South Korea, 2017-18. *Emerg Infect Dis*, 24 (10), 1953-1955. doi:10.3201/eid2410.180461

Kwon, J. H., Lee, D. H., Swayne, D. E., Noh, J. Y., Yuk, S. S., Erdene-Ochir, T. O., . . . Song, C. S. (2017). Reassortant Clade 2.3.4.4 Avian Influenza A(H5N6) Virus in a Wild Mandarin Duck, South Korea, 2016. *Emerg Infect Dis*, 23 (5), 822-826. doi:10.3201/eid2305.161905

Lee, D. H. (2020). Complete Genome Sequencing of Influenza A Viruses Using Next-Generation Sequencing. *Methods Mol Biol*, 2123 , 69-79. doi:10.1007/978-1-0716-0346-8_6

Lee, D. H., Bertran, K., Kwon, J. H., & Swayne, D. E. (2017). Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. *J Vet Sci*, 18 (S1), 269-280. doi:10.4142/jvs.2017.18.S1.269

Lee, D. H., Torchetti, M. K., Winker, K., Ip, H. S., Song, C. S., & Swayne, D. E. (2015). Intercontinental Spread of Asian-Origin H5N8 to North America through Beringia by Migratory Birds. *J Virol*, 89 (12), 6521-6524. doi:10.1128/JVI.00728-15

Lee, E. K., Lee, Y. N., Kye, S. J., Lewis, N. S., Brown, I. H., Sagong, M., . . . Lee, Y. J. (2018). Characterization of a novel reassortant H5N6 highly pathogenic avian influenza virus clade 2.3.4.4 in Korea, 2017. *Emerg Microbes Infect*, 7 (1), 103. doi:10.1038/s41426-018-0104-3

Lee, Y. J., Kang, H. M., Lee, E. K., Song, B. M., Jeong, J., Kwon, Y. K., . . . Lee, H. S. (2014). Novel reassortant influenza A(H5N8) viruses, South Korea, 2014. *Emerg Infect Dis*, 20 (6), 1087-1089. doi:10.3201/eid2006.140233

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22 (13), 1658-1659. doi:10.1093/bioinformatics/btl158

Minin, V. N., Bloomquist, E. W., & Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*, 25 (7), 1459-1471. doi:10.1093/molbev/msn090

Minin, V. N., & Suchard, M. A. (2008). Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci*, 363 (1512), 3985-3995. doi:10.1098/rstb.2008.0176

Mo, I. P., Bae, Y. J., Lee, S. B., Mo, J. S., Oh, K. H., Shin, J. H., . . . Lee, Y. J. (2016). Review of Avian Influenza Outbreaks in South Korea from 1996 to 2014. *Avian Dis*, 60 (1 Suppl), 172-177. doi:10.1637/11095-041715-Review

Olsen, B., Munster, V. J., Wallensten, A., Waldenstrom, J., Osterhaus, A. D., & Fouchier, R. A. (2006). Global patterns of influenza a virus in wild birds. *Science*, 312 (5772), 384-388. doi:10.1126/science.1122438

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, 26 (7), 1641-1650. doi:10.1093/molbev/msp077

Son, K., Kim, Y. K., Oem, J. K., Jheong, W. H., Sleeman, J. M., & Jeong, J. (2018). Experimental infection of highly pathogenic avian influenza viruses, Clade 2.3.4.4 H5N6 and H5N8, in Mandarin ducks from South Korea. *Transbound Emerg Dis*, 65 (3), 899-903. doi:10.1111/tbed.12790

Swayne, D. E. (2016). *Animal influenza* (2nd ed.). Ames, IA: John Wiley & Sons

Swayne, D. E., Suarez, D. L., Sims, L. (2020). *Influenza* In: Swayne, D. E., Boulianne, M., Logue, C. M., McDougald, L. R., Nair, V., Suarez, D. L., editors. *Diseases of poultry* . (14th ed.) Ames, IA: John Wiley & Sons

Swayne, D. E., Pavade, G., Hamilton, K., Vallat, B., & Miyagishima, K. (2011). Assessment of national strategies for control of high-pathogenicity avian influenza and low-pathogenicity notifiable avian influenza in poultry, with emphasis on vaccines and vaccination. *Rev Sci Tech*, 30 (3), 839-870. doi:10.20506/rst.30.3.2081

Takemae, N., Tsunekuni, R., Sharshov, K., Tanikawa, T., Uchida, Y., Ito, H., . . . Saito, T. (2017). Five distinct reassortants of H5N6 highly pathogenic avian influenza A viruses affected Japan during the winter of 2016-2017. *Virology*, 512 , 8-20. doi:10.1016/j.virol.2017.08.035

Woo, C., Kwon, J. H., Lee, D. H., Kim, Y., Lee, K., Jo, S. D., . . . Jeong, J. (2017). Novel reassortant clade 2.3.4.4 avian influenza A (H5N8) virus in a grey heron in South Korea in 2017. *Arch Virol*, 162 (12), 3887-3891. doi:10.1007/s00705-017-3547-2

Zhang, J., Chen, Y., Shan, N., Wang, X., Lin, S., Ma, K., . . . Qi, W. (2020). Genetic diversity, phylogeography, and evolutionary dynamics of highly pathogenic avian influenza A(H5N6) viruses. *Virus Evol* . doi:10.1093/ve/veaa079

Table 1. Well-supported viral transitions of clade 2.3.4.4C H5N6 HPAIVs among geographic regions and wild birds.

From	to	Mean actual migration rate ^a (95% BCI ^b)	Mean number of Markov jump (95% BCI)	BF
Guangdong	Fujian	1.02 (0.15, 2.20)	3.16 (3, 4)	>100
Guangdong	Hunan	2.37 (0.75, 4.37)	8.77 (7, 11)	>100
Guangdong	Jiangxi	0.31 (0.00, 1.08)	0.65 (0, 1)	12.60
Guangdong	Wild_Bird	1.89 (0.53, 3.58)	6.89 (6, 8)	>100
Guangdong	Yunnan	0.43 (0.00, 1.42)	1.13 (0, 3)	12.30
Hunan	Guangdong	0.85 (0.00, 2.14)	1.32 (0, 3)	89.57
Wild_Bird	Japan	2.67 (0.89, 4.72)	11.02 (10, 12)	>100
Wild_Bird	Korea	0.89 (0.12, 1.95)	3.00 (3, 3)	>100
Wild_Bird	Vietnam	1.04 (0.00, 2.16)	3.64 (1, 4)	>100

^a Actual transition rates were calculated as rate * indicator.

^b BCI, Bayesian credibility intervals.

Figure Legends

Figure 1. Bayesian phylogenetic MCC tree of the HA gene of a subgroup of clade 2.3.4.4C H5 HPAIVs. Branches are colored according to discrete traits of geographic regions or wild bird isolates, and the thickness of branches indicates posterior probabilities of the ancestral host type. The scale axis indicates the branch lengths by year.

Supporting Information

Figure S1. FastTree Phylogenetic tree of HA gene of clade 2.3.4.4 H5N6 HPAIVs. Node labels indicate FastTree support values. The scale bar indicates the number of substitutions per site. The sequences of the

highlighted subclade were used for the final dataset.

Figure S2. Bayesian phylogenetic MCC tree of the HA gene of a subgroup of clade 2.3.4.4C H5 HPAIVs with full taxa names and node bars. Horizontal node bars indicate 95% Bayesian credibility intervals for the divergence in time estimates.

