

New Reassortant Clade 2.3.4.4b Avian Influenza A(H5N6) Virus in Wild Birds, South Korea, 2017–2018

Technical Appendix

Methods

Virus Isolation

Fecal samples were tested for influenza A virus by egg inoculation using 9–1-day-old specific-pathogen-free embryonating chicken eggs. After a 72-h incubation period, allantoic fluids were harvested from an axenic medium, and undiluted allantoic fluids were tested for hemagglutinin (HA) activity. For the definitive diagnosis of influenza A virus, RNA was extracted from HA-positive allantoic fluid using an RNeasy Kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. The extracted RNA was tested for the influenza A virus matrix gene by real-time reverse transcription PCR (RT-PCR) as described previously (1). The hosts of the positive fecal sample were identified as Mandarin ducks (*Aix galericulata*) and mallards (*Anas platyrhynchos*) using DNA barcoding techniques as previously described (2).

Full-genome Sequencing

Full-length genome sequencing was performed for 19 of 40 isolates (5 from Cheongmi-cheon River collected on 22 December 2017, 13 from Gokgyo-cheon River collected on 23 December 2017, and 1 from Gokgyo-cheon River collected on 18 January 2018). For molecular analysis, RNA was extracted using the RNeasy kit (QIAGEN) according to the manufacturer's instructions. The 8 genes of each virus were amplified using 1-step RT-PCR. The RT-PCR amplicons (2 µg) of all 8 gene segments was used to prepare Ion Fragment sequencing libraries (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Briefly, amplicons were loaded onto beads, and emulsion PCR was conducted before sequencing with an Ion 318 chip on an Ion Torrent Personal Genome Machine. De novo and directed assembly of genome sequences were performed using Geneious R9 software (<http://www.geneious.com>). We

provided the genome sequences in the EpiFlu database of Global Initiative on Sharing All Influenza Data (GISAID; <https://www.gisaid.org>), under accession nos. EPI1190336–40, EPI1190342–61, EPI1190381, EPI1190386, EPI1190392, EPI1190410, EPI1190418, EPI1190426, EPI1190426, EPI1190433, EPI1190446, EPI1190498, EPI1190503, EPI1190508, EPI1190514–19, EPI1200474–509, EPI1200511–87.

Phylogenetic Analysis

For phylogenetic analysis, complete coding regions were aligned using MAFFT (<https://mafft.cbrc.jp/alignment/software/>), and manual editing and tree reconstruction were conducted using Geneious 8 software (<https://www.geneious.com/>). The sequence of clade 2.3.4.4 H5 highly pathogenic avian influenza viruses representing each subgroup and low pathogenic avian influenza virus strains related to new reassortment of H5N6 highly pathogenic avian influenza viruses in 2017–18 were selected from GISAID Epiflu and GenBank (<https://www.ncbi.nlm.nih.gov/genomes/FLU>) database as a reference for phylogenetic tree analysis. A maximum-likelihood tree was estimated by the MEGA7 software using the Hasegawa-Kishino-Yano model (3) of nucleotide substitution with gamma-distributed rate variation among sites with 4 rate categories. Statistical analysis of the phylogenetic tree was performed by bootstrap analysis on 1,000 replicates. The percentages of replicate trees (>70%) in which the associated taxa clustered together in the bootstrap test are shown next to the branches.

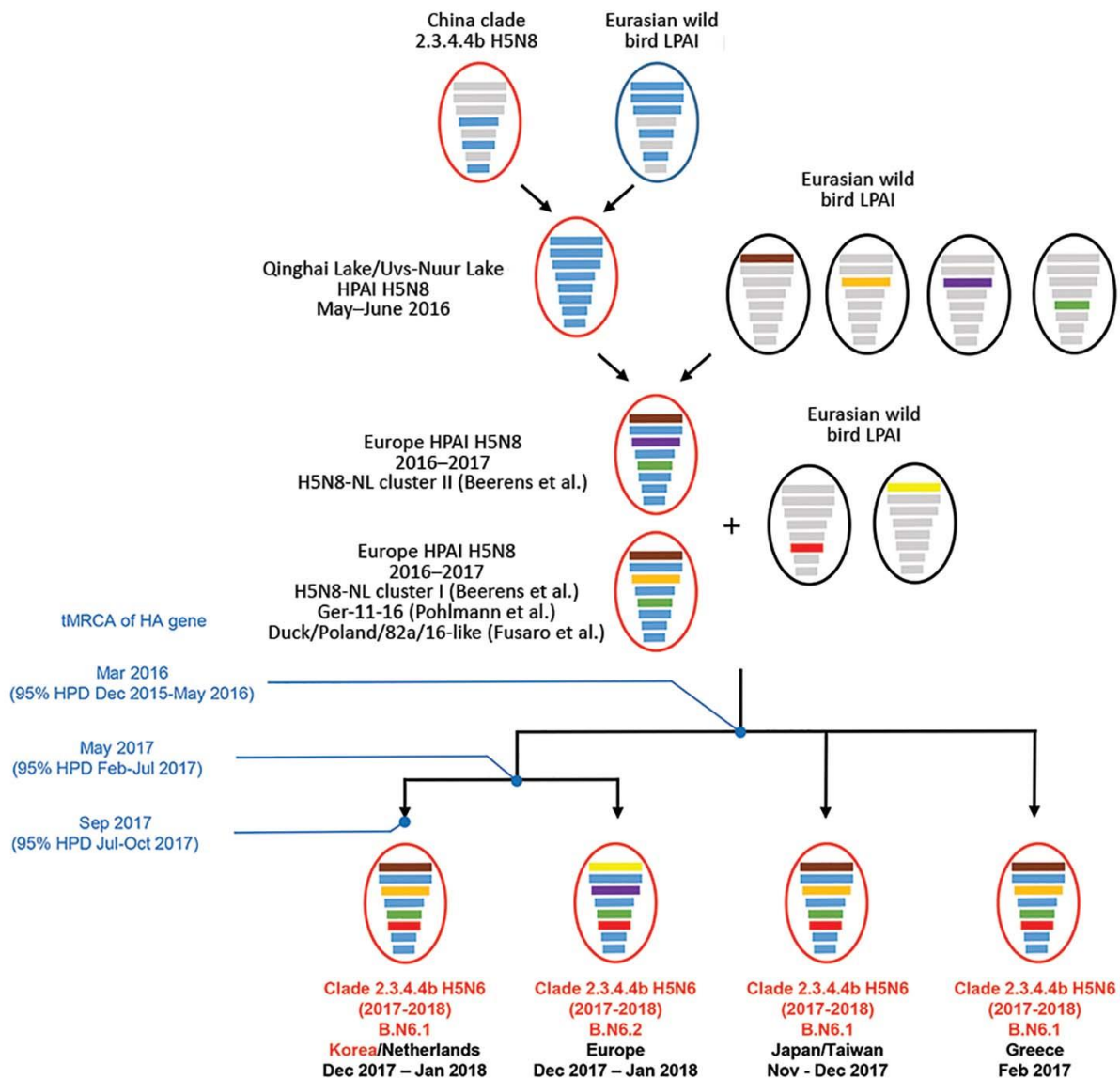
To better visualize the genetic relatedness of viruses, HA gene was analyzed using the median joining method implemented by NETWORK ver. 5.0 with epsilon set to 0. Nucleotide sequences of complete protein coding sequence region of subgroup B H5N6 and H5N8 identified from Asia and Europe in 2016 and 2017 ($n = 250$) were retrieved from GISAID Epiflu database on February 13, 2018. The sequences were pruned by using the online software cd-hit (4) at 97% homology to filter-out 66 representative sequences. The nucleotide sequences of HA segment obtained in this study ($n = 19$) were added to the dataset. A total of 85 nt sequences were aligned using MAFFT, and alignments were manually edited in Geneious 8 software (<https://www.geneious.com/>).

For estimation of the time to the most recent common ancestor, Bayesian analysis was performed for all 8 gene segments using BEAST version 1.8.4 (<http://beast.bio.ed.ac.uk>). We used an Hasegawa-Kishino-Yano substitution model with 4 gamma categories and specified an

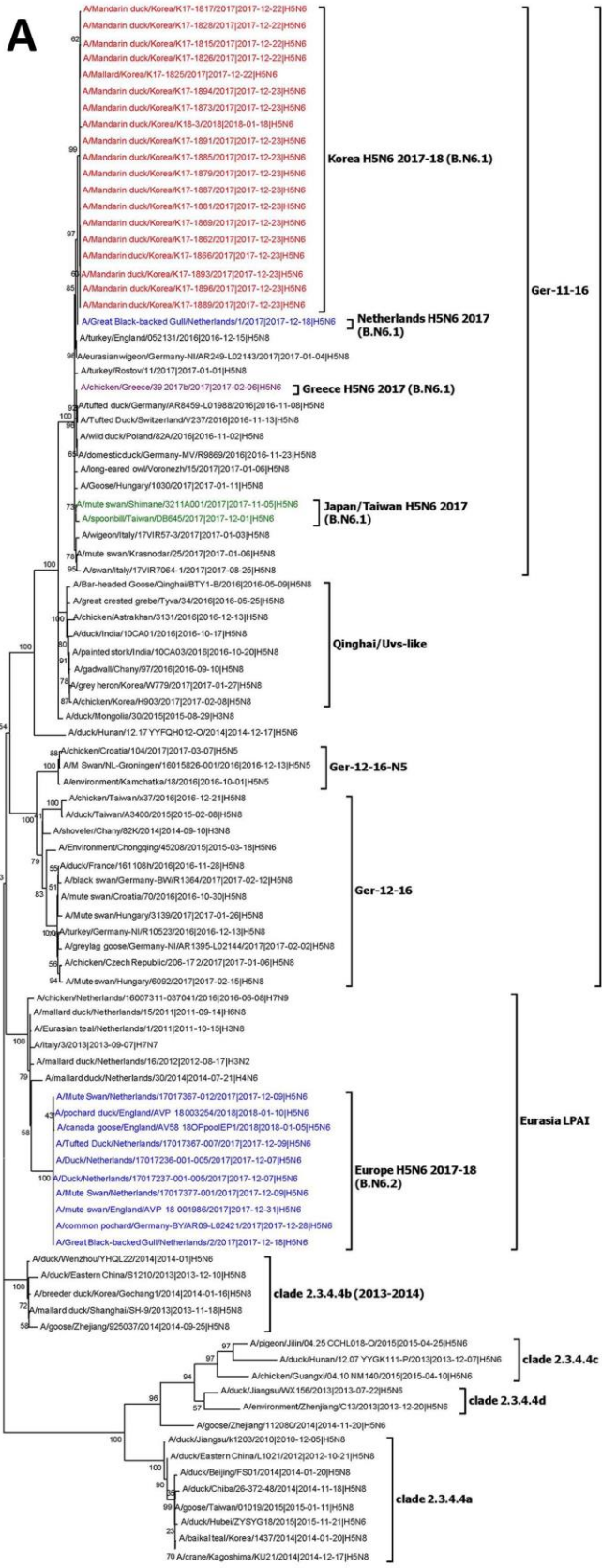
uncorrelated lognormal relaxed clock and GMRF Bayesian skyride tree prior for each segment. A Markov chain Monte Carlo method was used with 50 million chain lengths to draw inference under this model. A maximum clade credibility tree with common ancestor height was generated for each dataset using TreeAnnotator version 1.8.4 after 10% burn-in. The maximum clade credibility trees were visualized using the program FigTree v1.4.2.

References

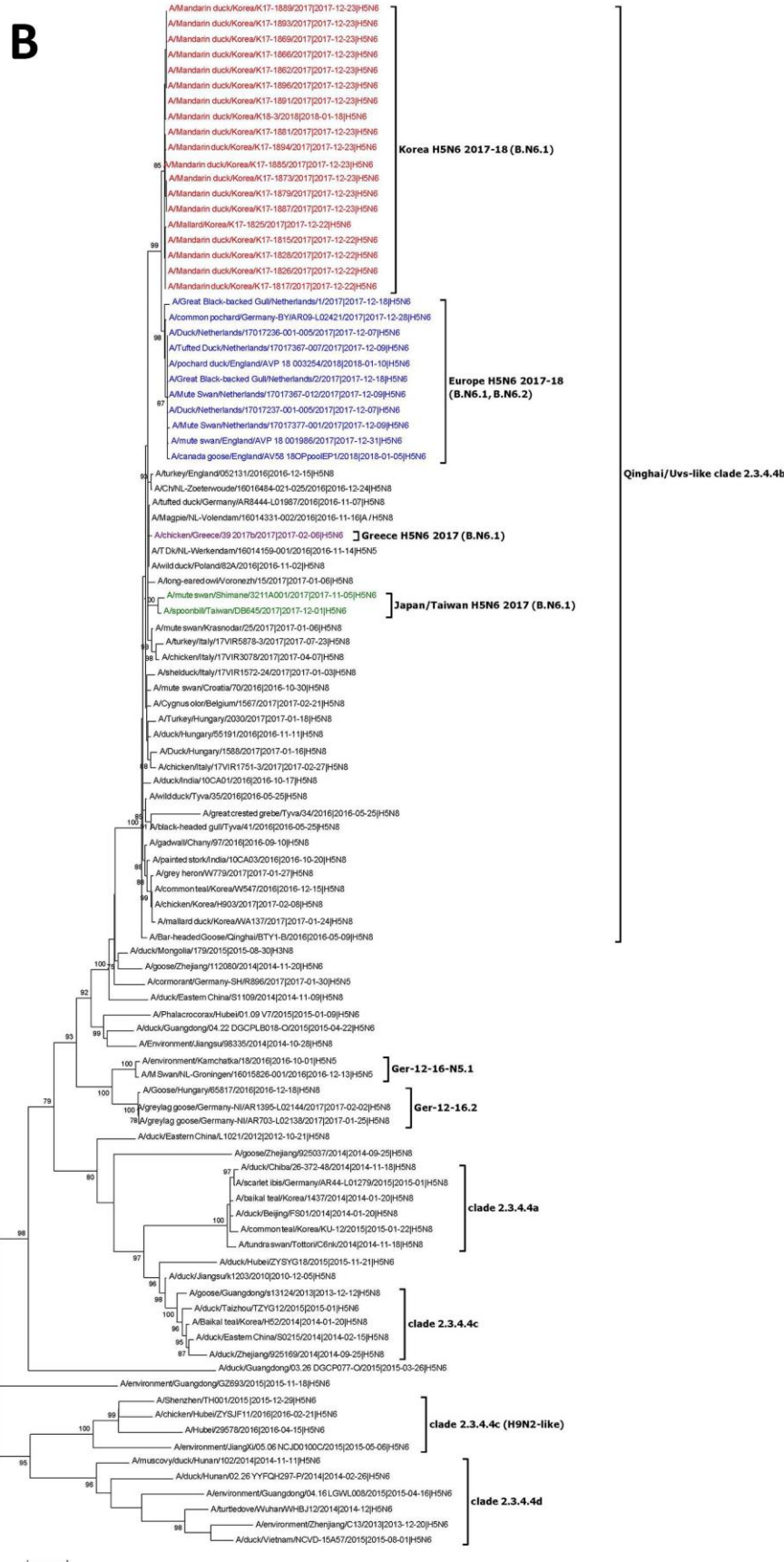
1. Spackman E, Senne DA, Bulaga LL, Myers TJ, Perdue ML, Garber LP, et al. Development of real-time RT-PCR for the detection of avian influenza virus. *Avian Dis.* 2003;47(Suppl):1079–82. [PubMed](https://pubmed.ncbi.nlm.nih.gov/15111111/)
<http://dx.doi.org/10.1637/0005-2086-47.s3.1079>
2. Lee DH, Lee HJ, Lee YJ, Kang HM, Jeong OM, Kim MC, et al. DNA barcoding techniques for avian influenza virus surveillance in migratory bird habitats. *J Wildl Dis.* 2010;46:649–54. [PubMed](https://pubmed.ncbi.nlm.nih.gov/20444444/)
<http://dx.doi.org/10.7589/0090-3558-46.2.649>
3. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985;22:160–74. [PubMed](https://pubmed.ncbi.nlm.nih.gov/27222222/)
<http://dx.doi.org/10.1007/BF02101694>
4. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26:680–2. [PubMed](https://pubmed.ncbi.nlm.nih.gov/20444444/)
<http://dx.doi.org/10.1093/bioinformatics/btq003>



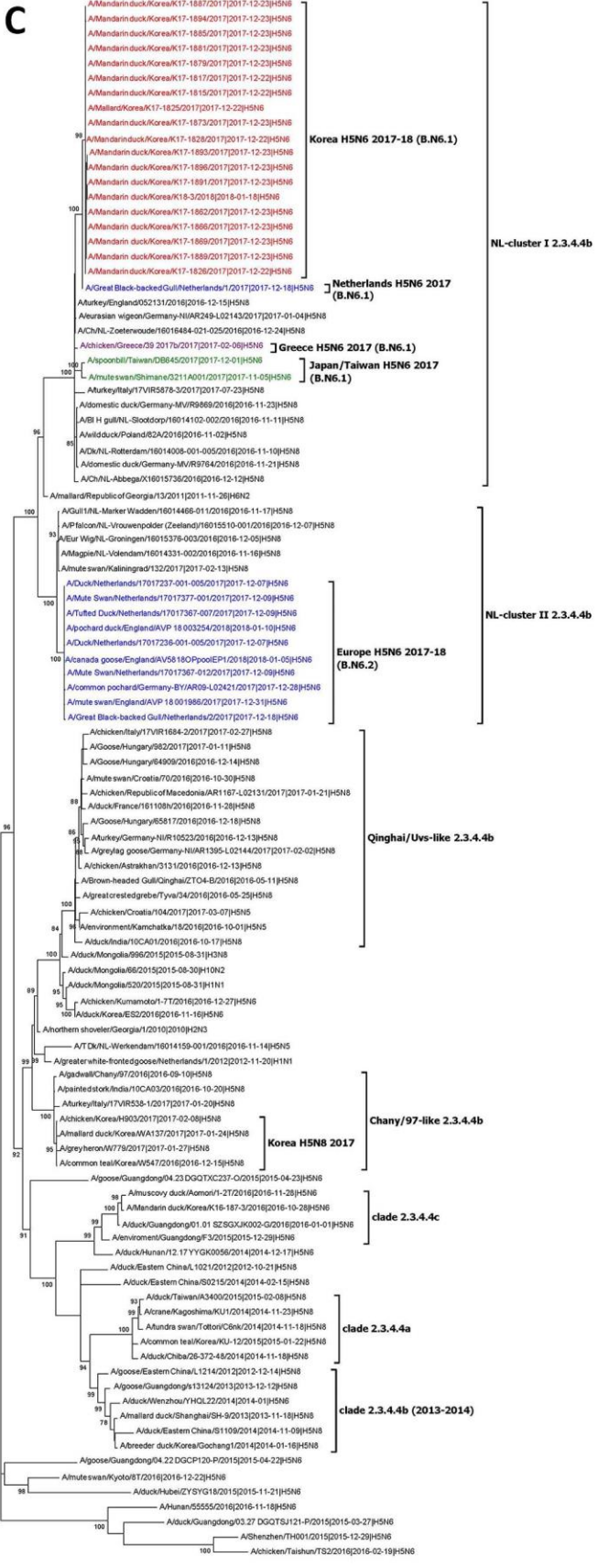
Technical Appendix 1 Figure 1. Proposed reassortment events leading to the new subgroup B influenza A(H5N6) viruses. Eight gene segments in each of the schematic virus particles are arranged from top to bottom to represent the PB2, PB1, PA, HA, NP, NA, M, and NS genes. The circled HPAI viruses are colored in red, and unknown LPAI viruses are colored in black. For each gene segment, different phylogenetic groupings are in different colors. The mean time to the most recent common ancestor of HA gene are shown at the node. HA, hemagglutinin; HPAI, highly pathogenic avian influenza; HPD, highest posterior density; LPAI, low pathogen avian influenza; M, matrix; NA, neuraminidase; NP, nucleoprotein; NS, nonstructural; PA, polymerase acidic; PB, polymerase basic; tMCRA, time to the most recent common ancestor.



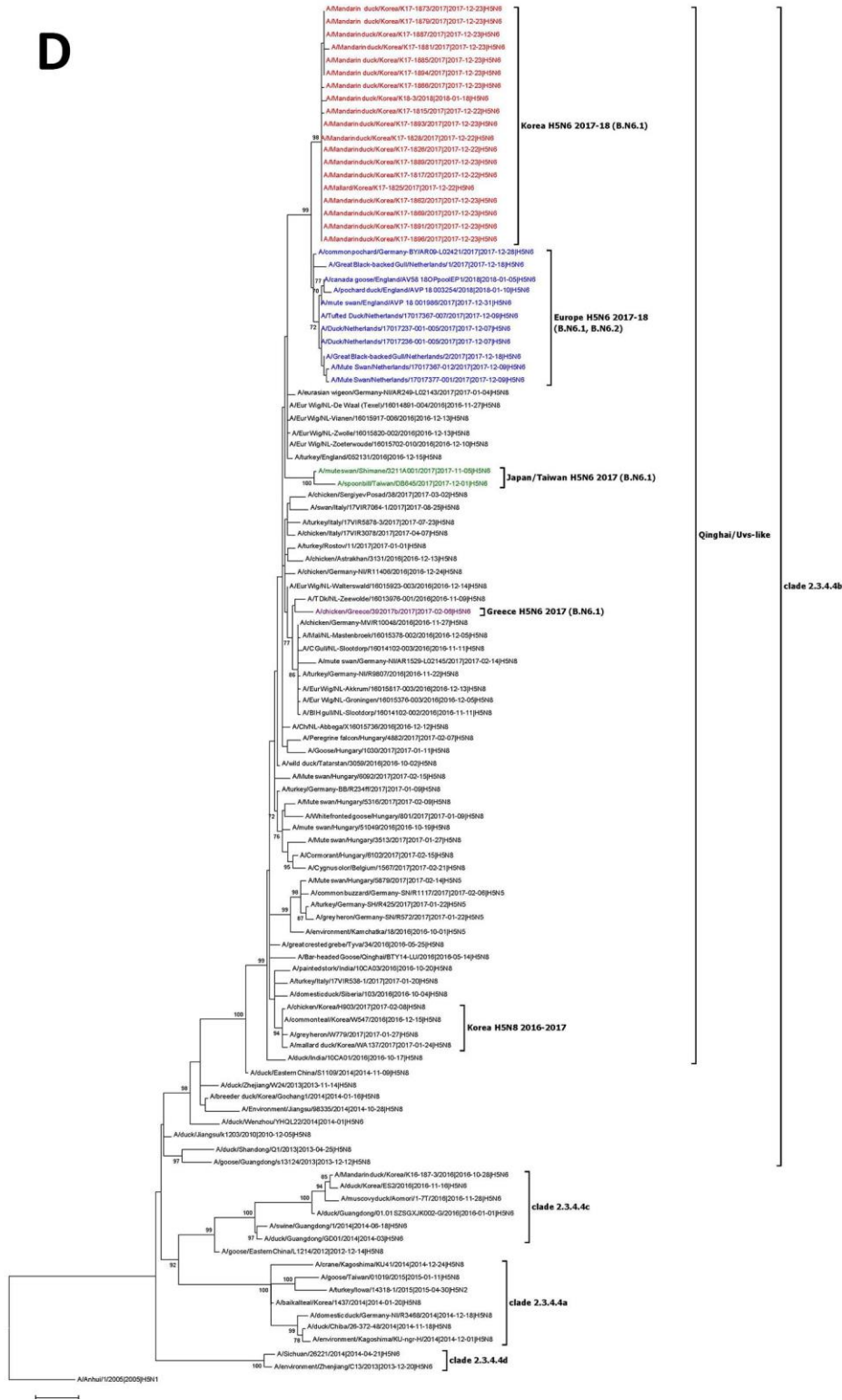
B



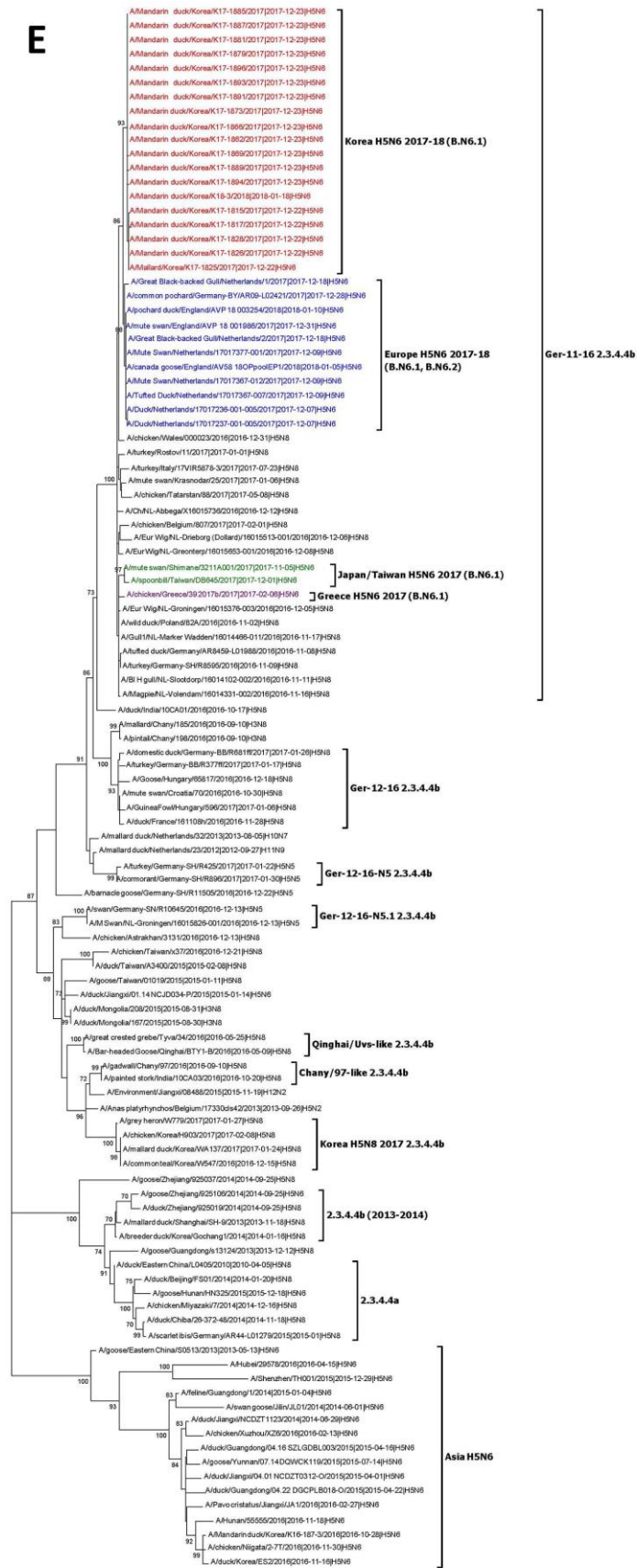
C

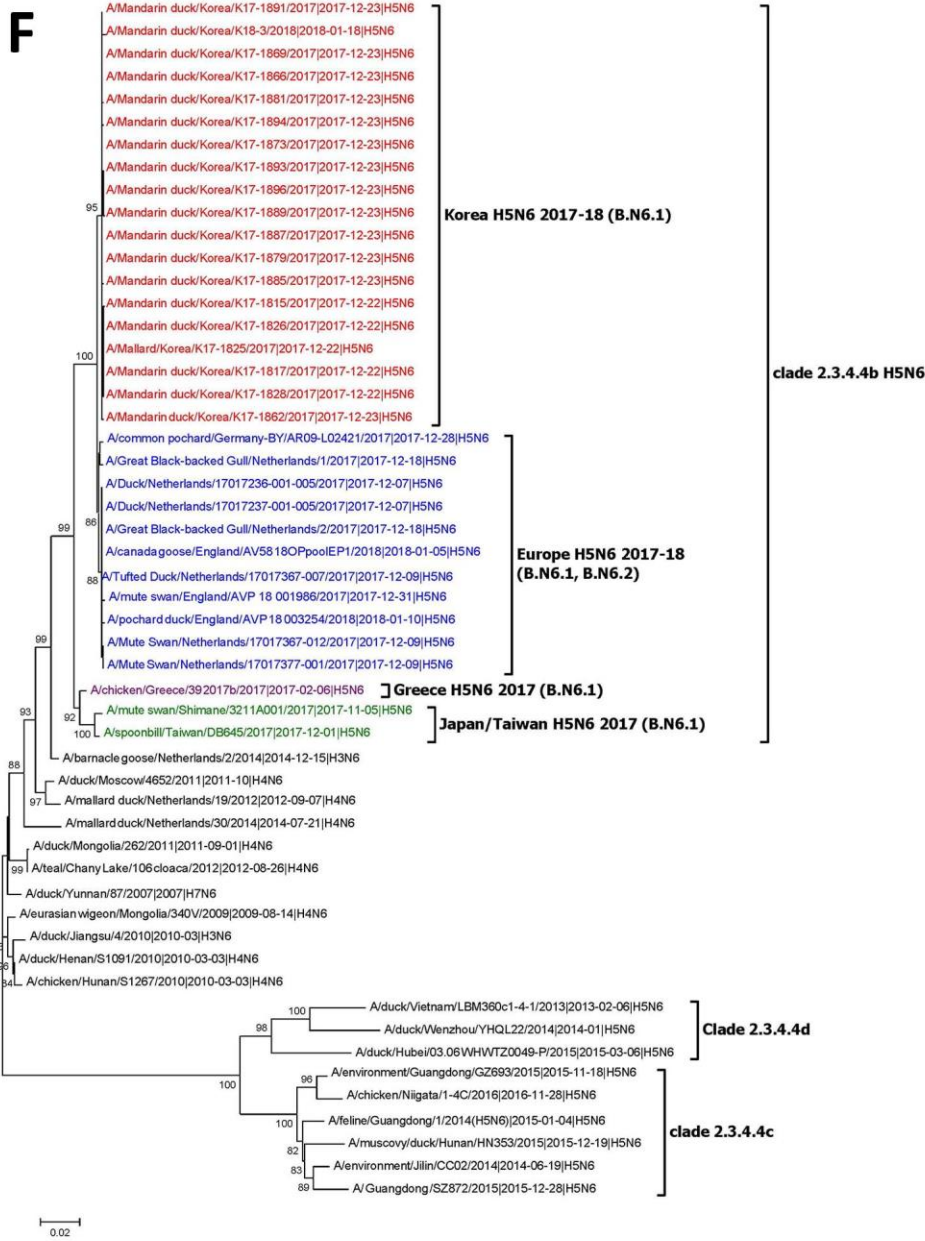


D

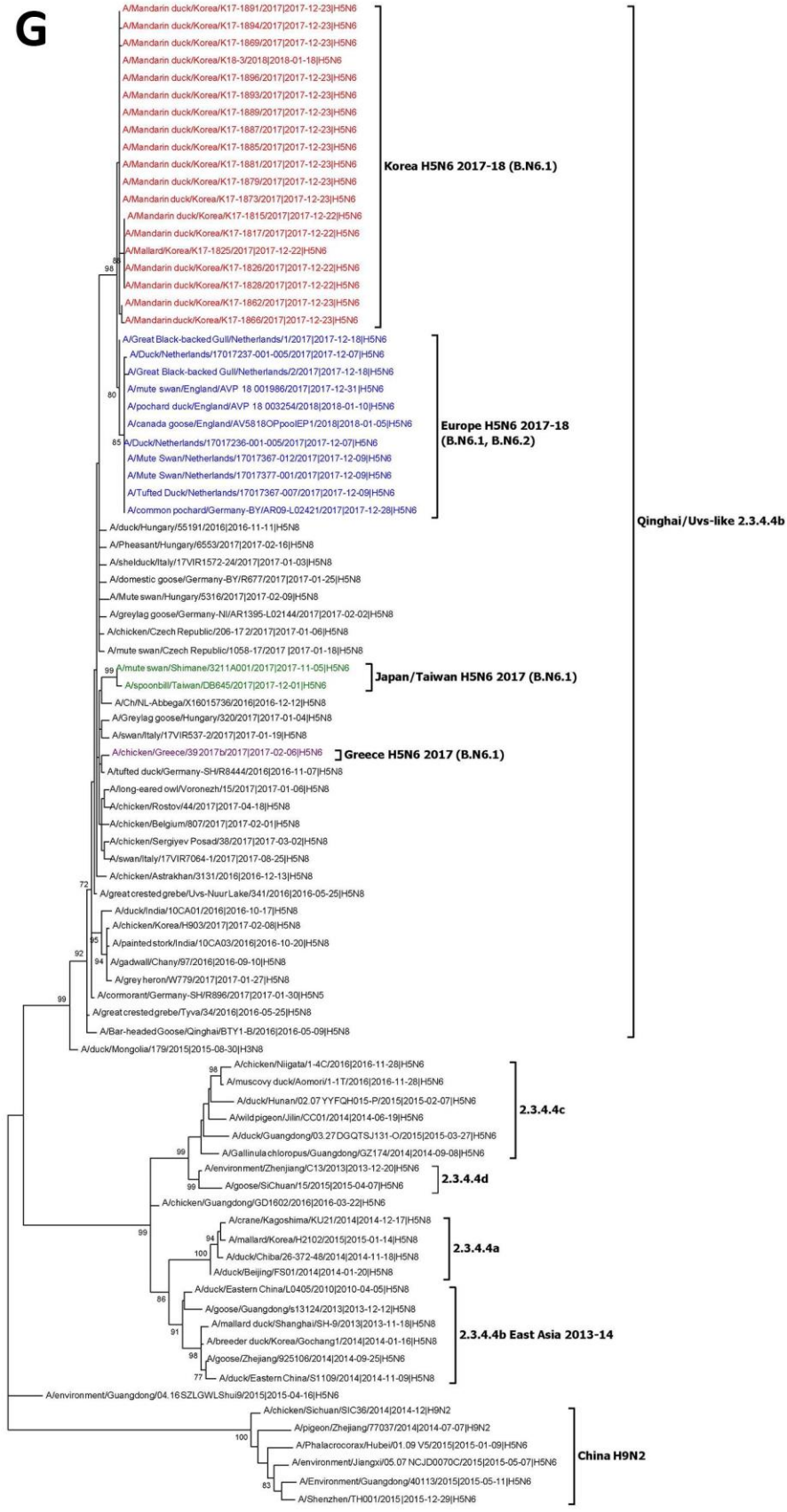


E

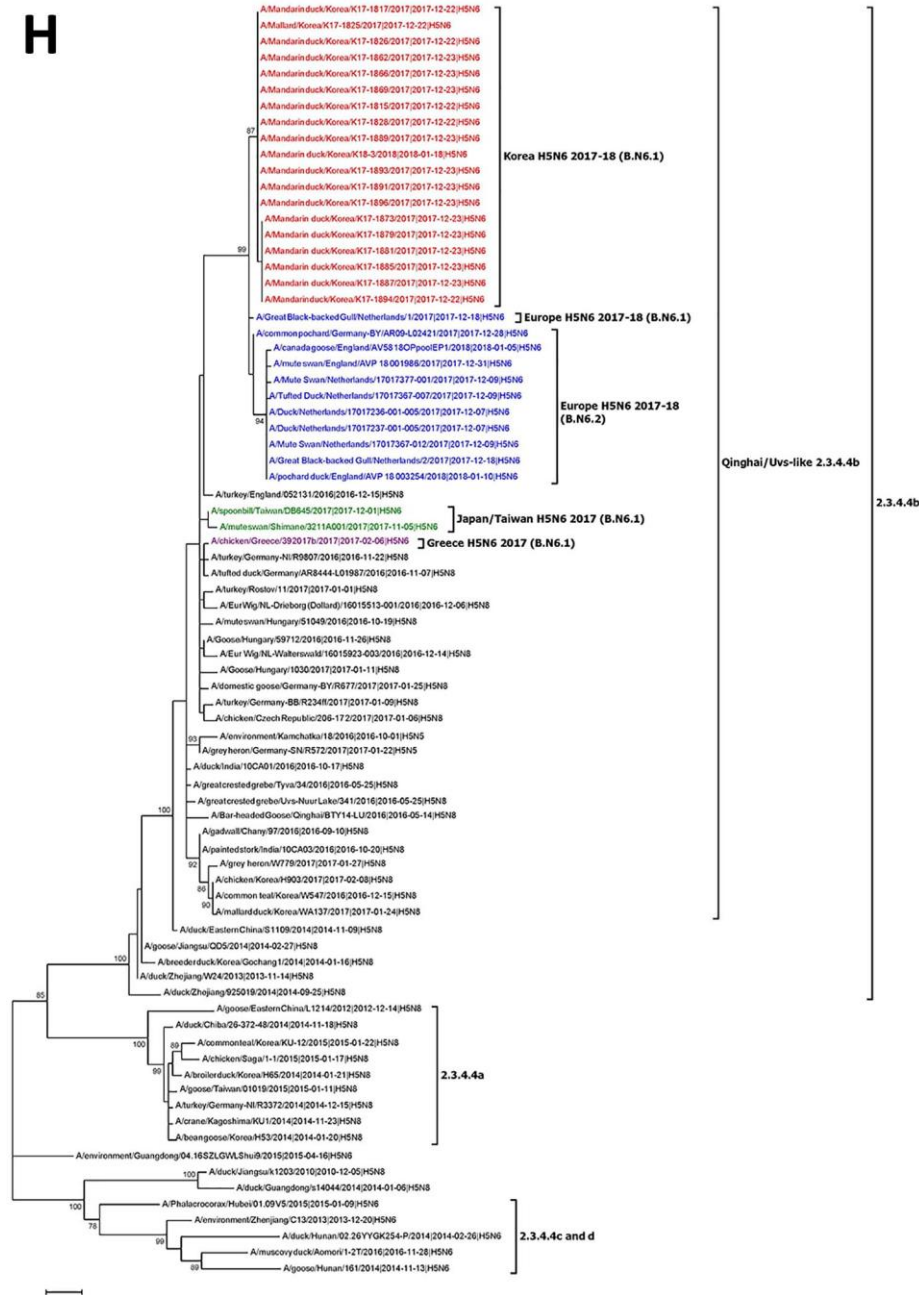




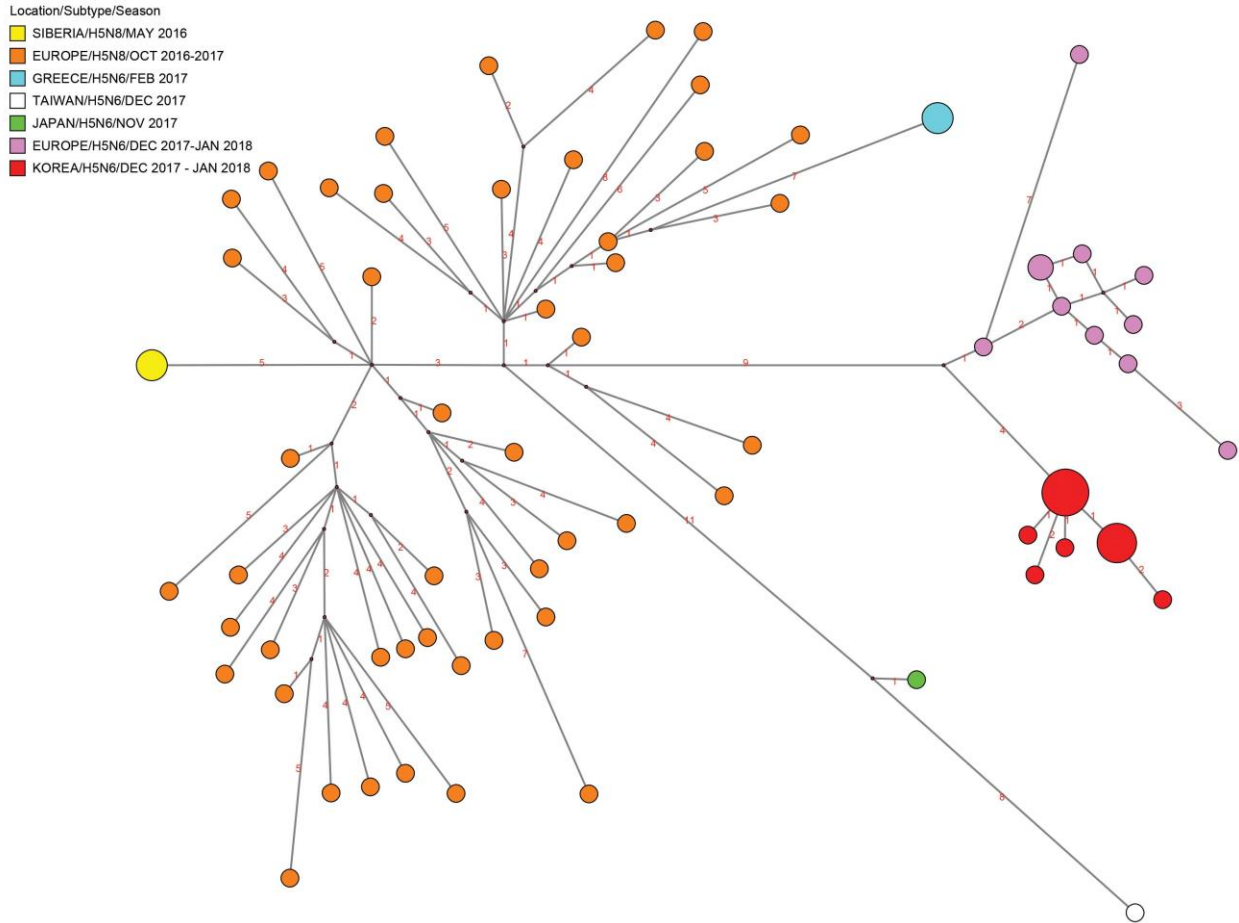
G



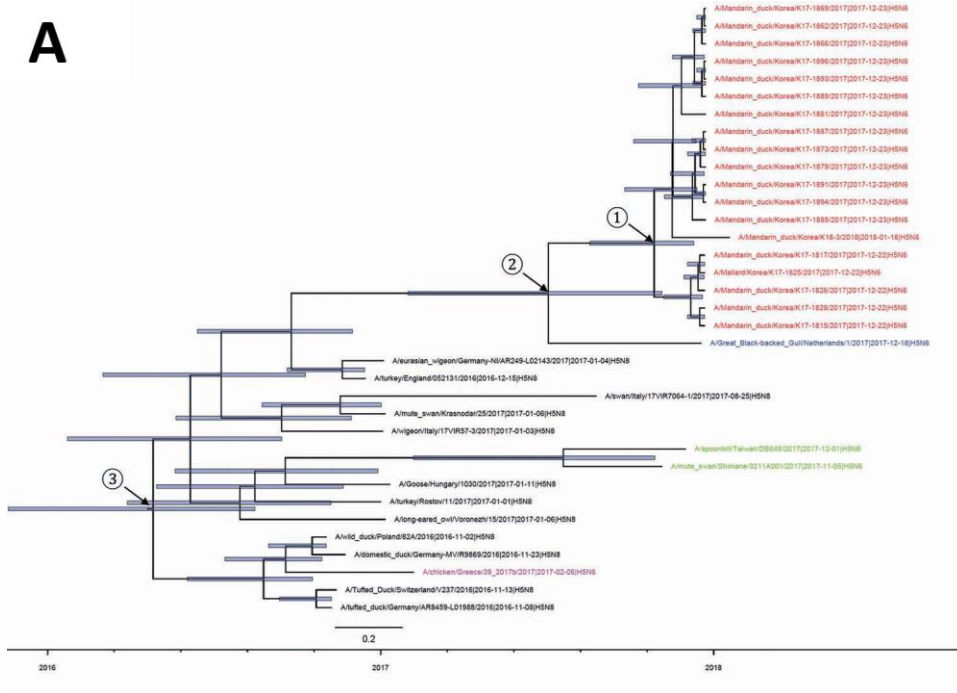
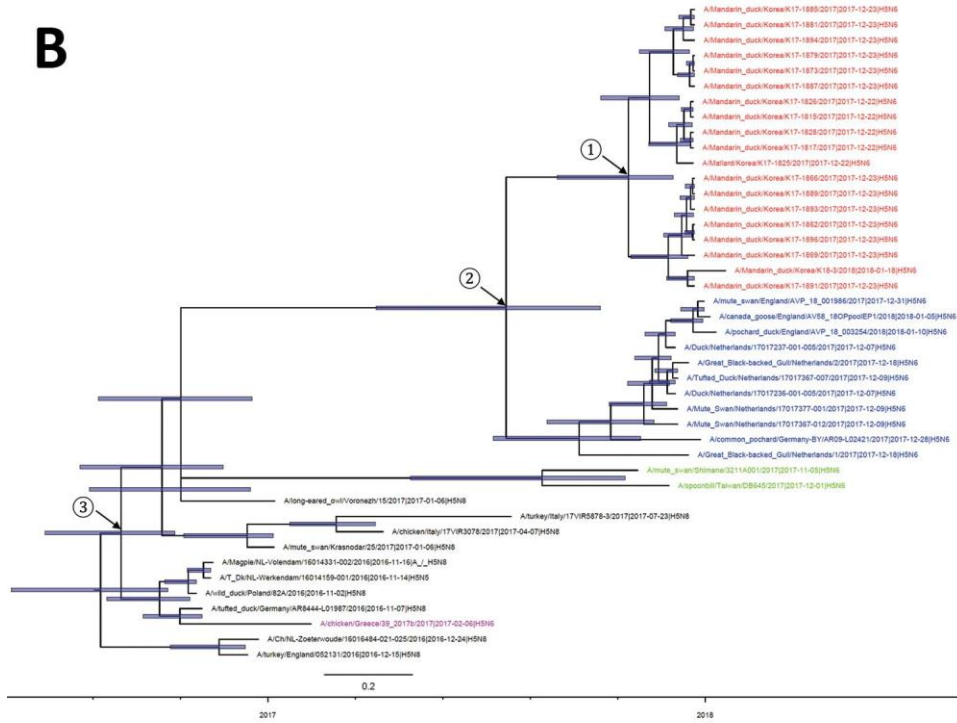
H



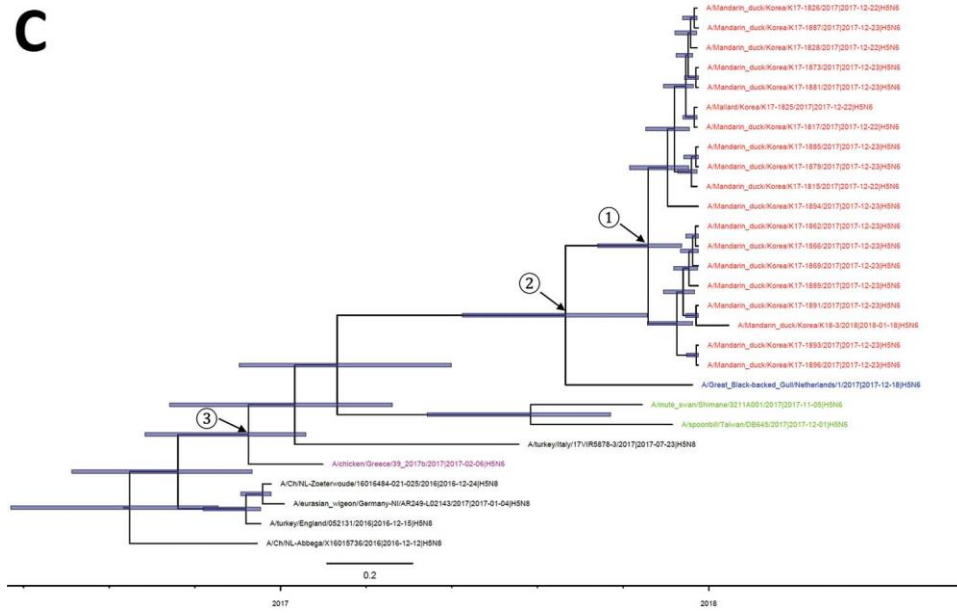
Technical Appendix 1 Figure 2. Maximum-likelihood phylogenetic tree of the PB2, PB1, PA, HA, NP, NA, M, and NS genes. The new subgroup B H5N6 isolates are colored according to the isolated location (South Korea, red; Europe, blue; Japan/Taiwan, green; Greece, purple). The percentages of replicate trees (>70%) in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. Each cluster is labeled according to classification of Pohlmann et al. (5) and Beerens et al. (6). HA, hemagglutinin; M, matrix; NA, neuraminidase; NP, nucleoprotein; NS, nonstructural; PA, polymerase acidic; PB, polymerase basic. Scale bars indicate nucleotide substitutions per site.



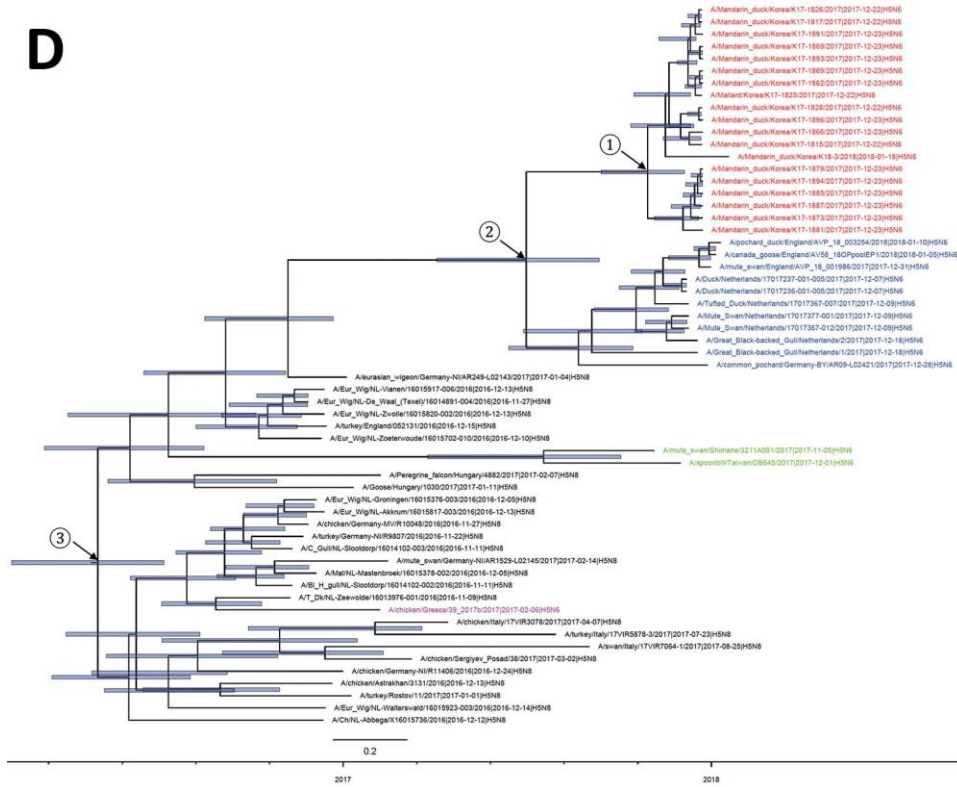
Technical Appendix 1 Figure 3. Median-joining phylogenetic network of subgroup B H5 highly pathogenic avian influenza viruses. The median-joining network was constructed from the HA gene. It includes all the most parsimonious trees linking the sequences. Each unique sequence is represented by a circle sized relative to its frequency in the dataset. Branch length is proportional to the number of mutations. Isolates are colored according to the location. HA, hemagglutinin.

A**B**

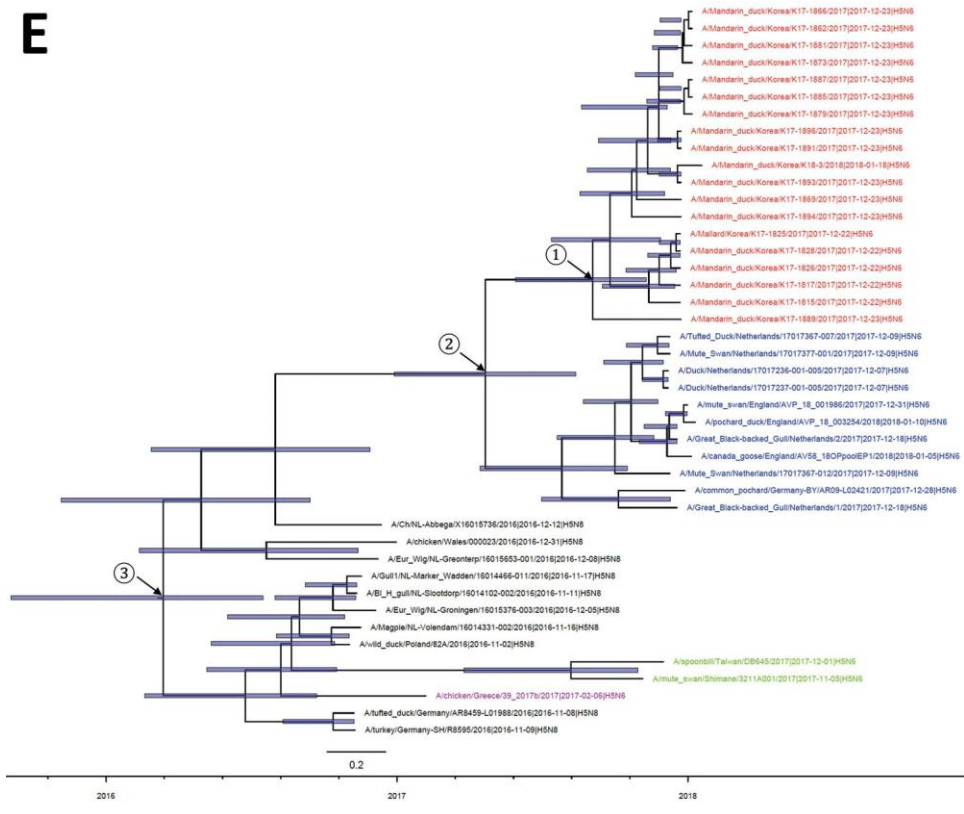
C



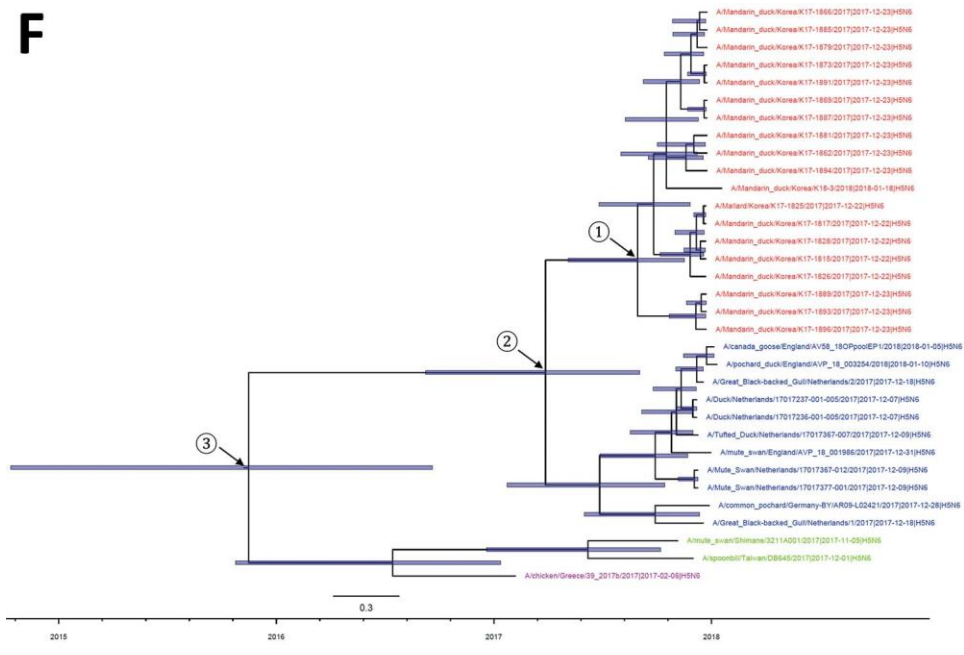
D

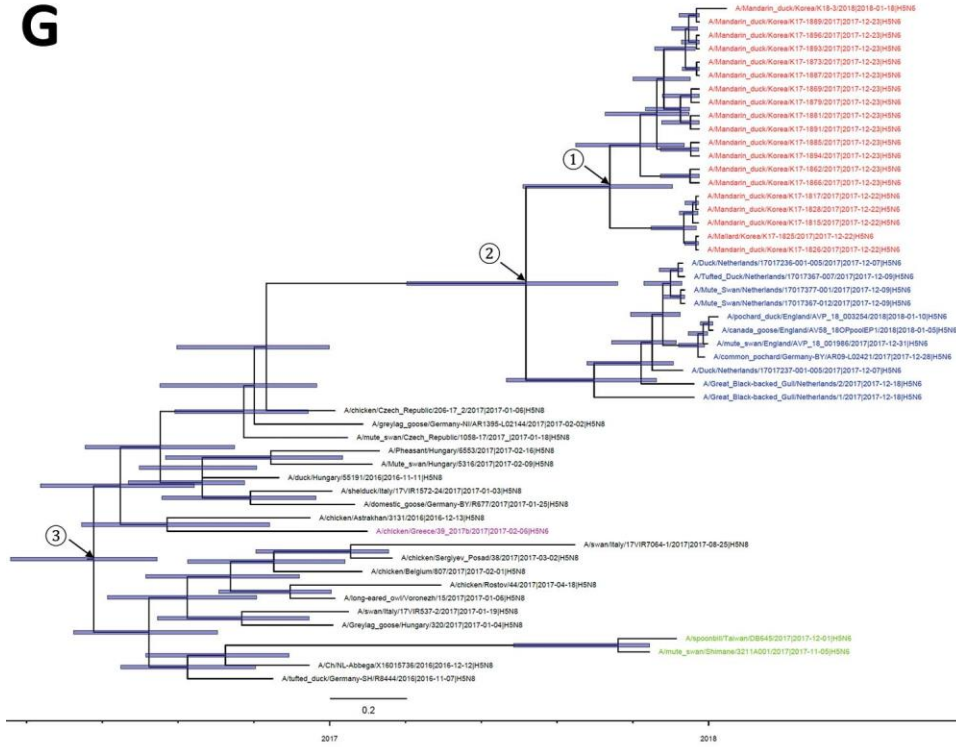
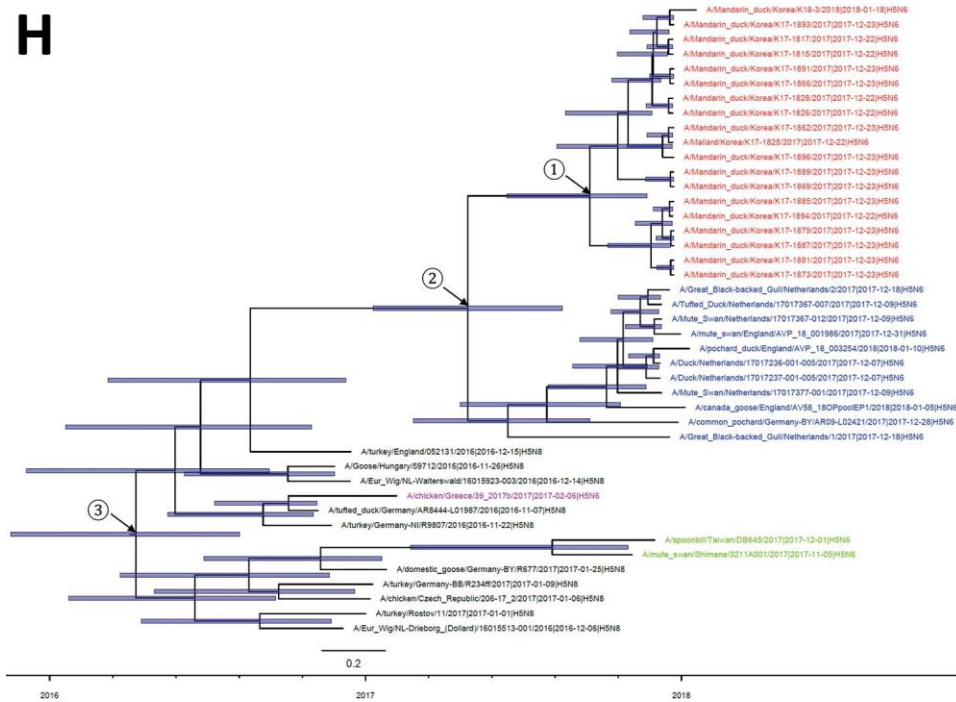


E



F



G**H**

Technical Appendix 1 Figure 4. Temporally structured maximum clade credibility phylogenetic tree (years on the horizontal axis) of the PB2, PB1, PA, HA, NP, NA, M, and NS genes of subgroup B HPAIV. The new subgroup B H5N6 Isolates are colored according to the isolated location (South Korea, red;

Europe, blue; Japan/Taiwan, green; and Greece, purple). The horizontal bars indicate the 95% highest posterior density (HPD) intervals of the most recent common ancestor. HA, hemagglutinin; HPD, highest posterior density; M, matrix; NA, neuraminidase; NP, nucleoprotein; NS, nonstructural; PA, polymerase acidic; PB, polymerase basic. Scale bars indicate nucleotide substitutions per site.