

The multiple origins of the H5N8 avian influenza sub-type

Andrew Dalby

A significant outbreak of H5N8 avian influenza began in early 2014 in Korea. H5N8 is a highly pathogenic avian influenza that is often fatal to chickens and other domestic poultry but which can be asymptomatic in ducks. The virus was also found in migratory birds in Europe and more recently in North America. This is the first time that an H5N8 outbreak has spread so widely and persisted for so long. Previous outbreaks have usually been short and geographically localised. In this study I present a phylogenetic analysis of all of the H5 hemagglutinin and N8 neuraminidase sequences to show that each of the H5N8 outbreaks has resulted from a different re-assortment event and that there have been at least 7 distinct origins of the viral sub-type since it was first characterised in a Turkey in Ireland in 1983.

1 Andrew R. Dalby
2 Faculty of Science and Technology, University of Westminster, 115 New Cavendish Street,
3 Westminster, W1W 6UW, UK.

4

5 A.Dalby@westminster.ac.uk

6

8 **Introduction**

9

10 Low pathogenicity avian influenza (LPAI) such as H9N2 commonly circulates between domestic
11 and wild birds across the world. It is less common for high pathogenicity avian influenza to pass
12 through migratory birds. The most common form of highly pathogenic avian influenza (HPAI)
13 has been H5N1 but recently we have seen the emergences of H5N8 in Korea and then the US
14 and Taiwan as a new form of HPAI. The H5N8 influenza A subtype was first isolated from a
15 turkey in Ireland in 1983 (Murphy 1986). After that initial outbreak the next cases of H5N8 were
16 in 2001 when a case was identified during environmental monitoring in a wild bird in New
17 Jersey. There were then a few sporadic detections in the United States and Thailand until the
18 2013 outbreak in Korea (Lee et al. 2014).

19

20 This outbreak was preceded by cases in Eastern China in 2010 that are distinct from the
21 American and Irish virus. Although the Korean outbreak strains had the same subtype the earlier
22 Asian cases the current outbreak seems to have been the product of another re-assortment of viral
23 segments from other H5 containing subtypes such as H5N1 or H5N5 and another N8 containing
24 subtype, rather than from the evolution of the previous H5N8 lineages (Lee et al. 2014; Zhao et
25 al. 2013). The 2013 Korean outbreak has been subdivided into two lineages one of which is
26 closely related to the Chinese sequences and that has only been identified in two samples in
27 Gochang and a second that contains all the other cases and that was originally identified in Buan
28 (Fan et al. 2014; Jeong et al. 2014). In the winter of 2014 the virus also spread to Europe and the

29 US via migratory birds. There were only a limited number of European and US cases, but there
30 has been widespread infection of domestic geese in Taiwan.

31

32 The H5N8 subtype can be transmitted to ferrets and mice, and antibodies have been detected in
33 domestic dogs, but so far it is not considered a threat to human health as there has not been a
34 case of bird to human transmission (Kim et al. 2014). However the virus does have devastating
35 consequences for the domestic poultry industry and chickens are particularly affected with a very
36 high mortality rate. It has been shown that the current virus has a pathogenicity index of 3 in
37 chickens (Kim et al. 2014). This is significantly higher than that of the original H5N8 from
38 Ireland (Alexander et al. 1986).

39

41 **Materials and Methods**

42 All of the available H5N8 hemagglutinin and neuraminidase nucleotide sequences were
43 downloaded from the NCBI Influenza Virus Resource on the 29th of June 2015. Both sets of
44 sequences, those obtained from the NCBI Influenza Virus Resource were aligned with Muscle
45 v3.8.31 (Edgar 2004) within Mega6.06.

46

47 Manual inspection and editing of the sequences was carried out using Mega6.06 (Tamura et al.
48 2013). During manual editing the 5' end of the sequence was edited to remove the un-translated
49 region. All sequences were trimmed to the start codon and stop codons. Sequences with missing
50 nucleotides were removed (table 1).

51

52 The test for the appropriate evolutionary model was carried out in Mega6.06 and showed that the
53 general time reversible with invariant sites model was optimal (GTR+I) (Posada & Crandall
54 1998; Tavaré 1986). Maximum likelihood phylogenetic trees for the H5N8 hemagglutinin and
55 neuraminidase sequences were calculated using the GTR+I evolutionary model with 1000
56 bootstrap repetitions using Mega6.06. Condensed trees were calculated, displayed and edited in
57 Mega6.06.

58

59 All of the available H5 hemagglutinin subunit and N8 neuraminidase subunits were downloaded
60 from the NCBI Influenza Virus Resource on the 27th of June 2015 (Bao et al. 2008). The search
61 was restricted to full-length sequences from any host. This data was used to carry out a complete

62 phylogenetic analysis for both the H5 hemagglutinin and N8 neuraminidase segments. The
63 sequences were aligned using MAFFT (Katoh & Standley 2013). FastTree2.1 was used to create
64 an approximate maximum likelihood tree for all of the sequences using the GTR + I evolutionary
65 model. The resulting trees were visualised and annotated with FigTree 1.4.2 (Rambaut 2007)

66

67 Supplementary data-files for the phylogenetic analysis of the H5 hemagglutinin are available
68 from <http://dx.doi.org/10.5281/zenodo.20653> and for the N8 neuraminidase from
69 <http://dx.doi.org/10.5281/zenodo.20655>.

71 Results and Discussion

72 The phylogenetic trees for the H5N8 hemagglutinin (figure 1) shows that the US H5N8
73 sequences prior to the current outbreak (2001-2011) form the most distinctive clade. This is then
74 followed by the sequences from the Irish outbreak in 1983. There are then a series of singleton
75 clades before the Gochang clade and the final clade that contains the rest of the sequences from
76 the current outbreak. The current outbreak shows some structure and appears to have broken into
77 3 or four different sub-lineages, one of which contains the 2014 US sequences.

78

79 The phylogenetic tree for the H5N8 neuraminidase dataset (figure 2) contains fewer features
80 when compared to the hemagglutinin trees but the overall topology is the same. For the Irish
81 cluster there is only a single full-length sequence available. There is also less clade structure in
82 the recent outbreak, although there is still a distinct sub-lineage for the US 2014 sequences.

83 Apart from this clade there is very little recent evolutionary change in the neuraminidase
84 sequences.

85

86 The phylogenetic trees for all of the N8 neuraminidase segments and all of the H5 hemagglutinin
87 segments are very large (supplementary data files 1 and 2) and so they have been edited in order
88 to examine the clusters that contain the H5N8 sequences (figures 3 to 10). These clusters show
89 good agreement in the location and sources of the possible rearrangements. These trees show that
90 the H5N8 phylogenetic trees for the two envelope segments (figures 1 and 2) have to be
91 polyphyletic and it is consistent with multiple re-assortment events having occurred that have
92 resulted in the creation of novel H5N8 subtype lineages.

93

94 Out of the 7 clusters 5 of them are singletons, New Jersey 2001, California 2011, Thailand 2012,
95 Quang Ninh 2013 and California 2014 (quail). Of these the most surprising are the US sequences
96 as it was expected that these would be from a single lineage, and that the gaps in time between
97 collected sequences would reflect a lack of sampling. However it is clear that they have arisen
98 from re-assortment events.

99

100 The first cases of H5N8 avian influenza were in Turkeys in Ireland in 1983 (figure 3). From the
101 clusters of sequences for the neuraminidase and hemagglutinin, the neuraminidase is most
102 closely related to those found in H3N8 infected ducks in the Ukraine in 1963 and the
103 hemagglutinin is most closely related to H5N2 found in an Italian turkey in 1982. Considering
104 the geographical spread and the gaps in the timeline it is impossible to state for sure that this is a
105 re-assortment of the H5N2 and H3N8. There are also German sequences from 1984 and 1985
106 with a similar hemagglutinin from the H5N6 and H5N2 subtypes respectively. This lends further
107 support to believing that the H5N2 subtype is likely to be involved in the re-assortment.

108

109 The New Jersey 2001 H5N8 is derived from either an H5N7 or H5N2 hemagglutinin in
110 shorebirds in New Jersey/Delaware Bay (figure 4). The next closest hemagglutinin sequences are
111 for Japanese H5N3 ducks in 2002 but these are geographically very distant and it is difficult to
112 imagine a migratory connection without further evidence of widespread dispersal. This
113 hemagglutinin most likely combined with an H6N8 or possibly an H11N8, which were also
114 circulating in Delaware Bay from 1993. The absence of H6N8 sequence data from Delaware Bay

115 between 1993 and the occurrence of H5N8 in 2001 is of some concern. This location is a focus
116 of the South-eastern Cooperative Wildlife Disease Study that carries out regular sampling. If the
117 virus was present during this period it would be expected that it would be sampled more
118 frequently.

119

120 The California 2011 H5N8 case is less ambiguous and it is clearly a re-assortment of an H5N1
121 viral subtype with and H3N8 subtypes in mallards in California (figure 5). The H5N1 and H3N8
122 sequences seem to have been the dominant subtypes in this location and might the reason for
123 there not being a wider distribution of H5N8.

124

125 The Thailand 2012 re-assortment is more complex (figure 6). The sequence is part of a H5N2
126 cluster in wild birds in Xianghai but there are also some H5N3 cases. The neuraminidase clusters
127 with a group of H3N8 sequences mostly from ducks in Eastern China. This group also includes a
128 Vietnamese duck sequence. From 2013 onwards this cluster is dominated by H10N8 sequences
129 and these could have replaced H5N8. This would fit with the re-assortment event having taken
130 place in the Xianghai region.

131

132 Quang Ninh is a coastal region of North Vietnam that border with China. The Quang Ninh
133 cluster for hemagglutinin is mostly H5N1 subtype sequences from Vietnamese Muscovy ducks
134 (figure 7). The neuraminidase cluster is from a mixed group that also includes H3

135 hemagglutinins in ducks from Jiangxi and Vietnam. This is consistent with the re-assortment
136 having occurred in Vietnamese wild ducks between the H5N1 and H3N8 subtypes.

137

138 The California 2014 re-assortment is also unambiguous. Almost all of the sequences in the
139 hemagglutinin cluster are of the H5N5 subtype from mallards in California (figure 8). The
140 neuraminidase cluster is similarly almost homogeneous for H3N8 sequences also from mallard in
141 California and so it seems clear that the re-assortment took place in Californian wild ducks that
142 then spread the virus to quail.

143

144 In the past H5N8 outbreaks have been short lived and localised, but in the recent outbreak the
145 subtype has persisted through two breeding seasons and spread over three continents (figures 9
146 and 10). This new outbreak contains the Guangdong H5 hemagglutinin that has become the
147 predominant form of the H5 hemagglutinin in China. While the current outbreak came to
148 prominence in Korea there were earlier cases in China. The tree structure shows the presence of
149 at least two distinct lineages. One lineage probably originated around Jiangsu in 2010 from a
150 possible reassortment of H5N1 and H3N8 (Kang et al. 2015). This has been named the Gochang
151 lineage after the region in Korea where the most recent sequence was detected (Jeong et al.
152 2014). This lineage forms a distinctive cluster at the base of both the neuraminidase and
153 hemagglutinin trees containing the current outbreak. The Shandong 2013 neuraminidase
154 sequence is particularly distinct to other members of this lineage and it is also breaks off the
155 hemagglutinin tree for the Gochang lineage close to the origin.

156

157 A second Korean lineage has been named the Buan lineage (Jeong et al. 2014) from the Buan
158 region of Korea where it was isolated. From the hemagglutinin tree this lineage appears to be
159 splitting into two sub-lineages in Korea and a third in North America. However from the
160 neuraminidase tree the picture is more complex and it is not clear that there are sub-lineages as it
161 has a nested structure. The presence of sub-clades or sub-lineages has also been proposed in
162 recent work by Hill *et al.* in a study that combines ecological data with the phylogenetic data
163 (Hill et al. 2015). They propose that the lineages are dependent on geographical location, but
164 they only used the H5 hemagglutinin in the analysis.

165

166 The North American sequences do form a distinct sub-clade in both of the trees, although there
167 are only a small number of N8 neuraminidase sequences. There is also some evidence that
168 branch lengths are getting longer within this group, indicating more sequence variation. This is
169 interesting as there has been a reassortment in North America to produce a new H5N2 virus,
170 which contains the Jiangsu H5 lineage hemagglutinin. It is possible that this hemagglutinin might
171 undergo further re-assortment allowing the Guangdong H5 hemagglutinin to displace the
172 existing US H5 hemagglutinin in other influenza subtypes (Verhagen et al. 2015).

173

174 **Conclusions**

175 The H5N8 subtype is made up of at least 7 distinct lineages that have each been produced by a
176 distinct re-assortment event. The presence of these recombination events affects the phylogenetic
177 analysis and has to be accounted for in the H4N8 phylogenetic trees. This is likely to be a more
178 general observation when constructing phylogenetic trees of influenza sub-types where re-

179 assortment will mean that segments can have a very different evolutionary history. These effects
180 need to be accounted for before any phylogenetic analysis can be carried out, as the methods
181 need to account for reassortment events. Analysis that does not explicitly account for re-
182 assortments is likely to be unreliable, especially if it is used for calculating varying mutation
183 rates along different branches.

184

185 The results here have shown an unexpected degree of re-assortment, especially amongst the
186 sequences from the United States. These events happen very rapidly. Mostly re-assortment does
187 not produce a persistent new subtype and this explains the presence of a high proportion of
188 singleton sequences. For H5N8 it is possible that the most recent re-assortment has finally
189 generated a viable subtype that will continue to circulate but it is still possible that it might die
190 out once again and that it will only return sporadically.

191

192 **Acknowledgments**

193

194 I would like to thank Dr Edward Wright for his helpful discussions on viral re-assortment.

195

196

197

198 **References**

- 199 Alexander D, Parsons G, and Manvell R. 1986. Experimental assessment of the pathogenicity of eight
200 avian influenza A viruses of H5 subtype for chickens, turkeys, ducks and quail. *Avian pathology*
201 15:647-662.
- 202 Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, and Lipman D. 2008. The
203 influenza virus resource at the National Center for Biotechnology Information. *Journal of*
204 *virology* 82:596-601.
- 205 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
206 *acids research* 32:1792-1797.
- 207 Fan S, Zhou L, Wu D, Gao X, Pei E, Wang T, Gao Y, and Xia X. 2014. A novel highly pathogenic H5N8 avian
208 influenza virus isolated from a wild duck in China. *Influenza Other Respir Viruses* 8:646-653.
- 209 Hill SC, Lee Y-J, Song B-M, Kang H-M, Lee E-K, Hanna A, Gilbert M, Brown IH, and Pybus OG. 2015. Wild
210 waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic
211 H5N8 influenza in the Republic of Korea. *Infection, Genetics and Evolution*.
- 212 Jeong J, Kang H-M, Lee E-K, Song B-M, Kwon Y-K, Kim H-R, Choi K-S, Kim J-Y, Lee H-J, and Moon O-K.
213 2014. Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its relationship
214 with migratory birds in South Korea during 2014. *Veterinary microbiology* 173:249-257.
- 215 Kang H-M, Lee E-K, Song B-M, Jeong J, Choi J-G, Jeong J, Moon O-K, Yoon H, Cho Y, and Kang Y-M. 2015.
216 Novel reassortant influenza A (H5N8) viruses among inoculated domestic and wild ducks, South
217 Korea, 2014. *Emerging infectious diseases* 21:298.
- 218 Katoh K, and Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
219 improvements in performance and usability. *Molecular biology and evolution* 30:772-780.
- 220 Kim Y-I, Pascua PNQ, Kwon H-I, Lim G-J, Kim E-H, Yoon S-W, Park S-J, Kim SM, Choi E-J, and Si Y-J. 2014.
221 Pathobiological features of a novel, highly pathogenic avian influenza A (H5N8) virus. *Emerging*
222 *Microbes & Infections* 3:e75.
- 223 Lee Y-J, Kang H-M, Lee E-K, Song B-M, Jeong J, Kwon Y-K, Kim H-R, Lee K-J, Hong M-S, and Jang I. 2014.
224 Novel reassortant influenza A (H5N8) viruses, South Korea, 2014. *Emerging infectious diseases*
225 20:1087.
- 226 Murphy T. 1986. The control and epidemiology of an influenza A outbreak in Ireland. *Acute Virus*
227 *Infections of Poultry*: Springer, 23-28.
- 228 Posada D, and Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*
229 14:817-818.
- 230 Rambaut A. 2007. FigTree, a graphical viewer of phylogenetic trees. See <http://tree.bio.ed.ac.uk/software/figtree>.
- 232 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: molecular evolutionary genetics
233 analysis version 6.0. *Molecular biology and evolution* 30:2725-2729.
- 234 Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on*
235 *mathematics in the life sciences* 17:57-86.
- 236 Verhagen JH, Herfst S, and Fouchier RA. 2015. How a virus travels the world. *science* 347:616-617.
- 237 Zhao K, Gu M, Zhong L, Duan Z, Zhang Y, Zhu Y, Zhao G, Zhao M, Chen Z, and Hu S. 2013.
238 Characterization of three H5N5 and one H5N8 highly pathogenic avian influenza viruses in China.
239 *Veterinary microbiology* 163:351-357.

240

241

242 **Tables**

243

244

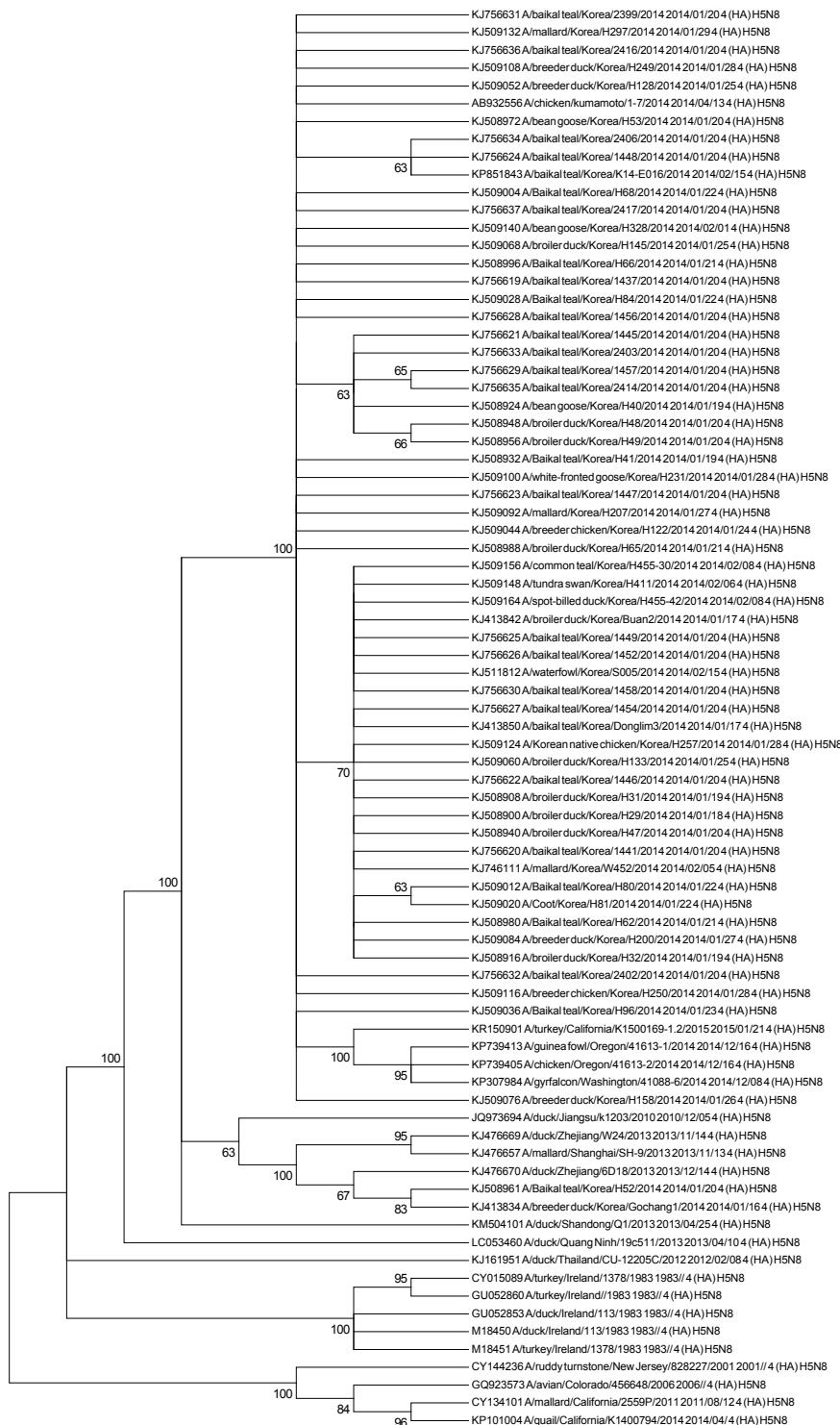
Segment	Deleted Sequence Identifiers.
HA	KR233690, AY296086, KR233674, KR232364, KR233682, KJ413855, KJ413856, EF607893.
NA	CY054463, GU052855, GU052862, KP101006, KR233692, JQ973696, KR233676, KR232366, KR233684,

245

246 Table 1: The sequences that were removed from the phylogenetic analysis because of truncation.

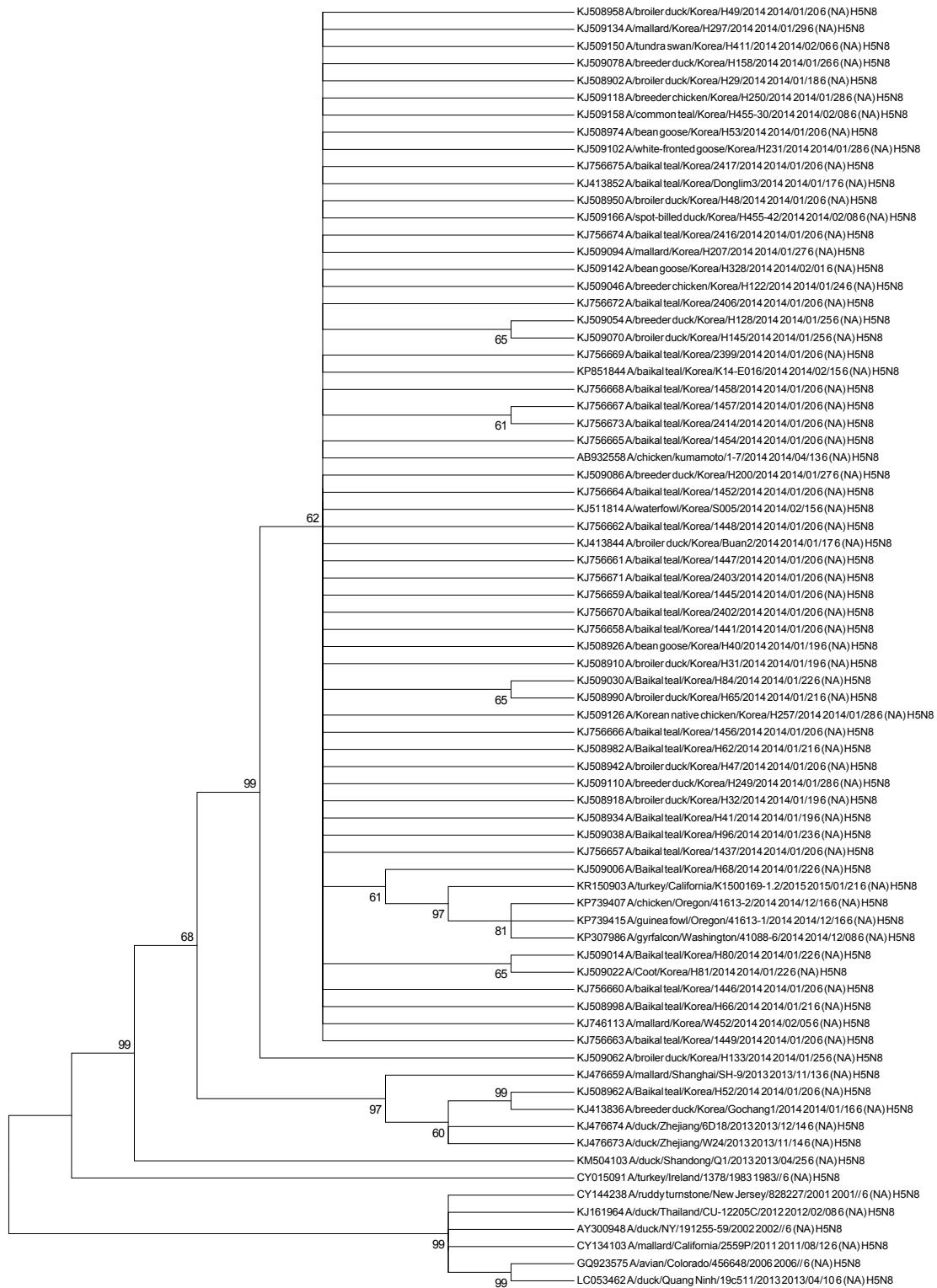
247

248

249 **Figures**

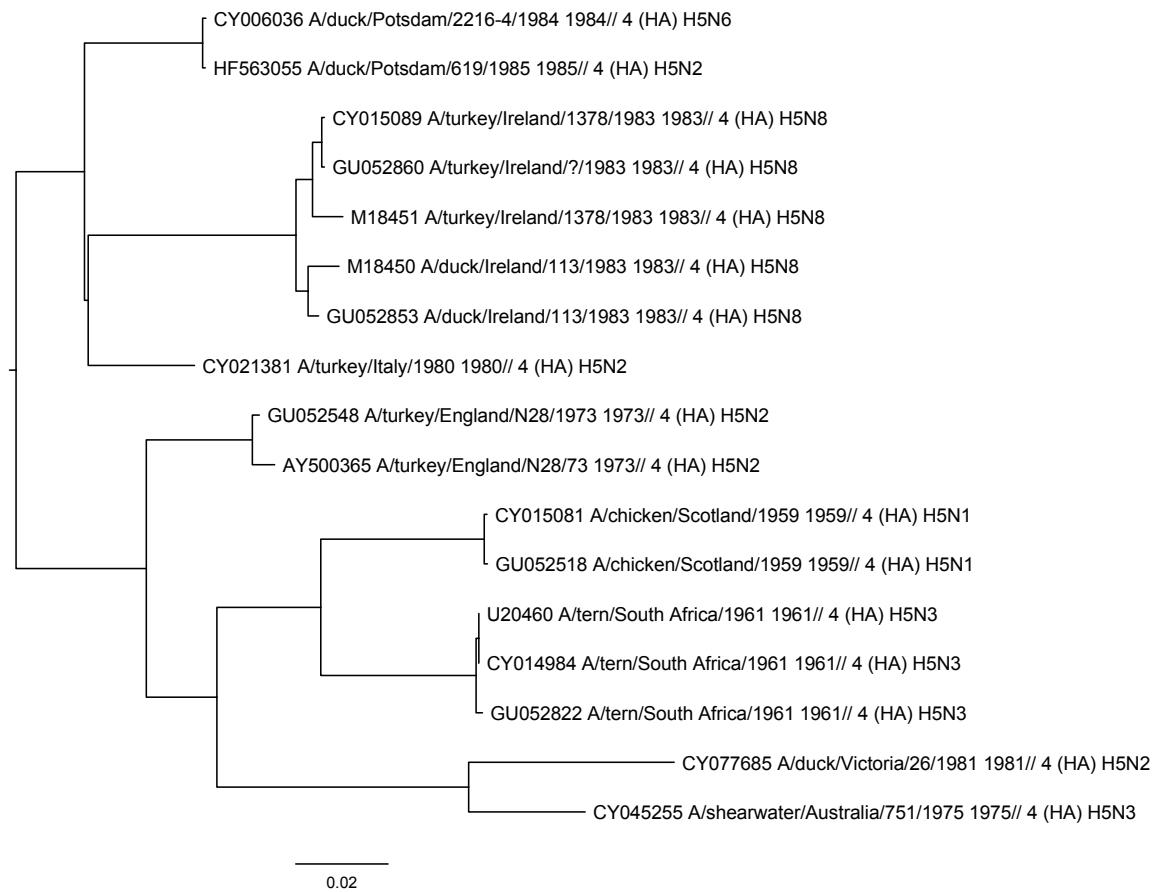
250

251 Figure 1: The condensed H5N8 hemagglutinin phylogenetic tree. Internal nodes are labelled with
252 the bootstrap values that have a cut-off of 60%.

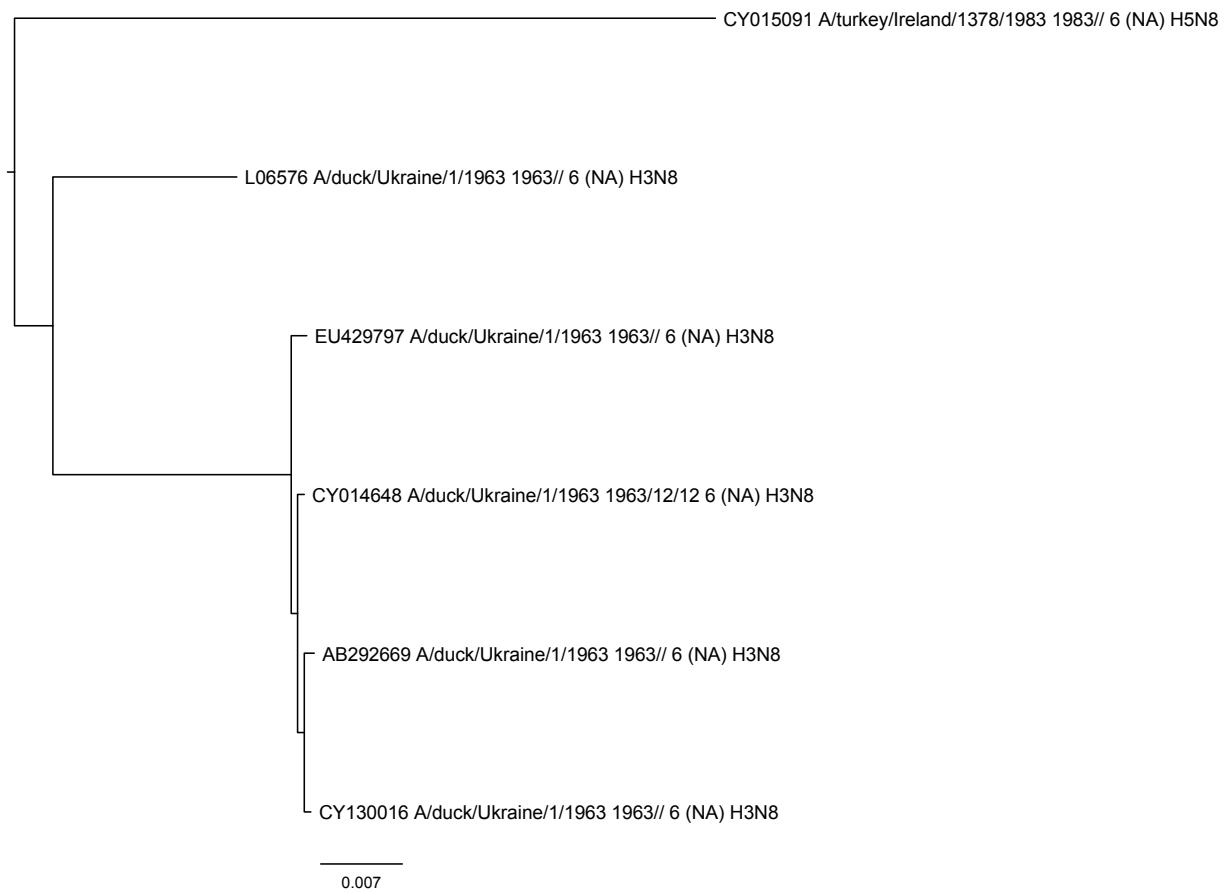


253

254 Figure 2: The condensed H5N8 neuraminidase phylogenetic tree. Internal nodes are labelled with
 255 the bootstrap values that have a cut-off of 60%.

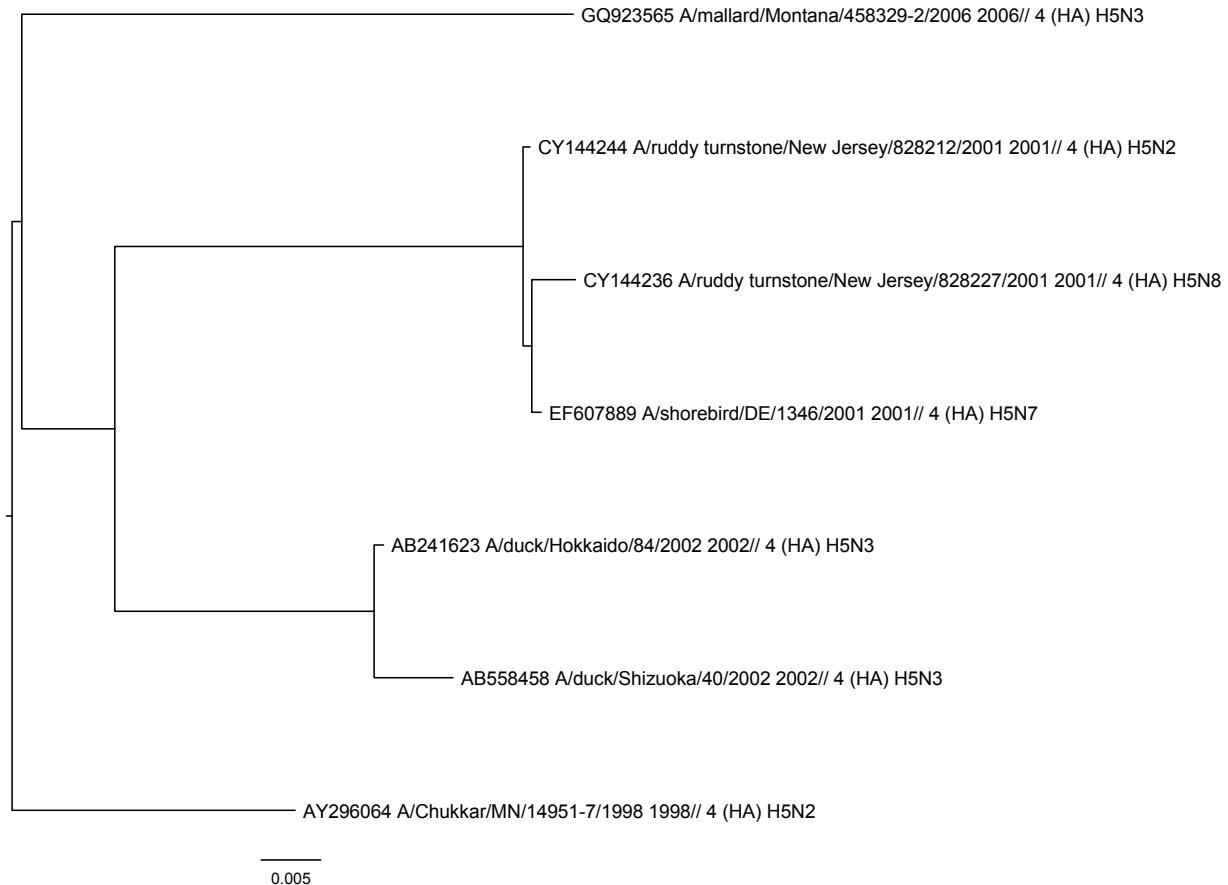


257 Figure 3A: The hemagglutinin clade for Ireland in 1983.



260 Figure 3B: The neuraminidase clade for Ireland in 1983.

261



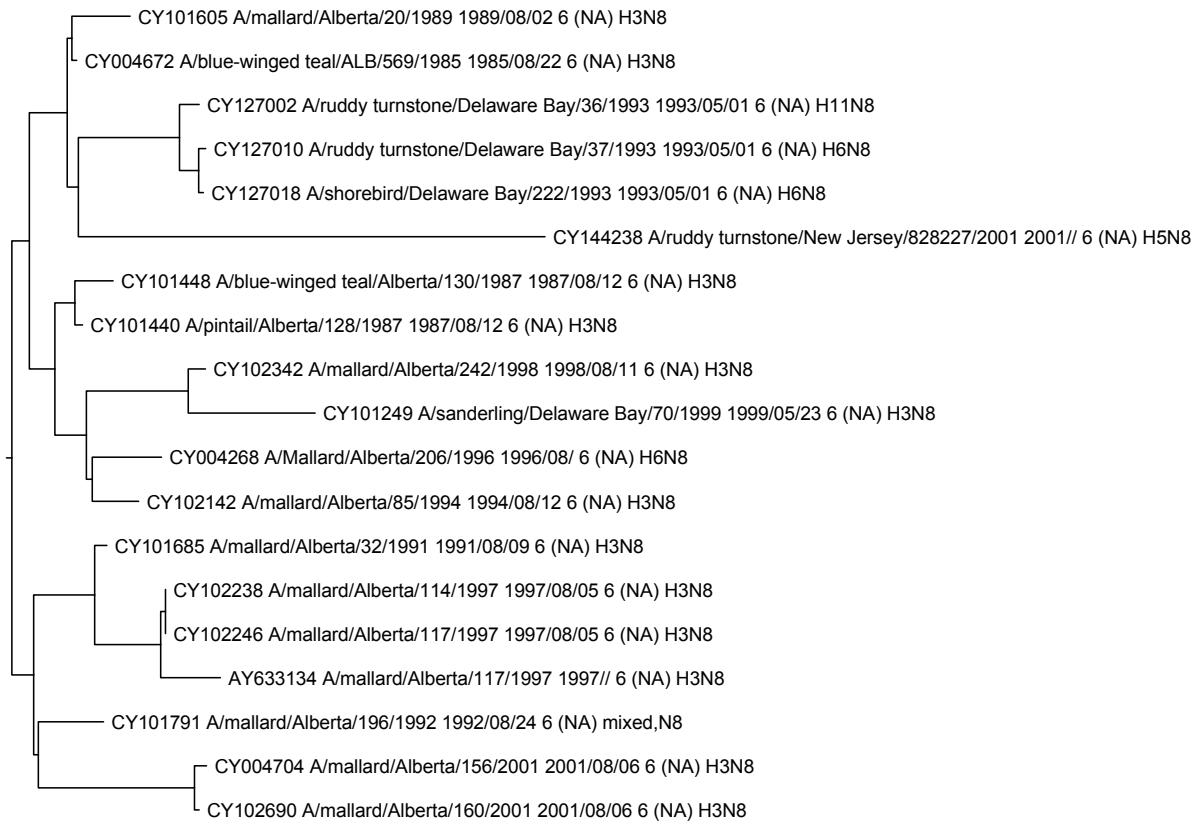
262

0.005

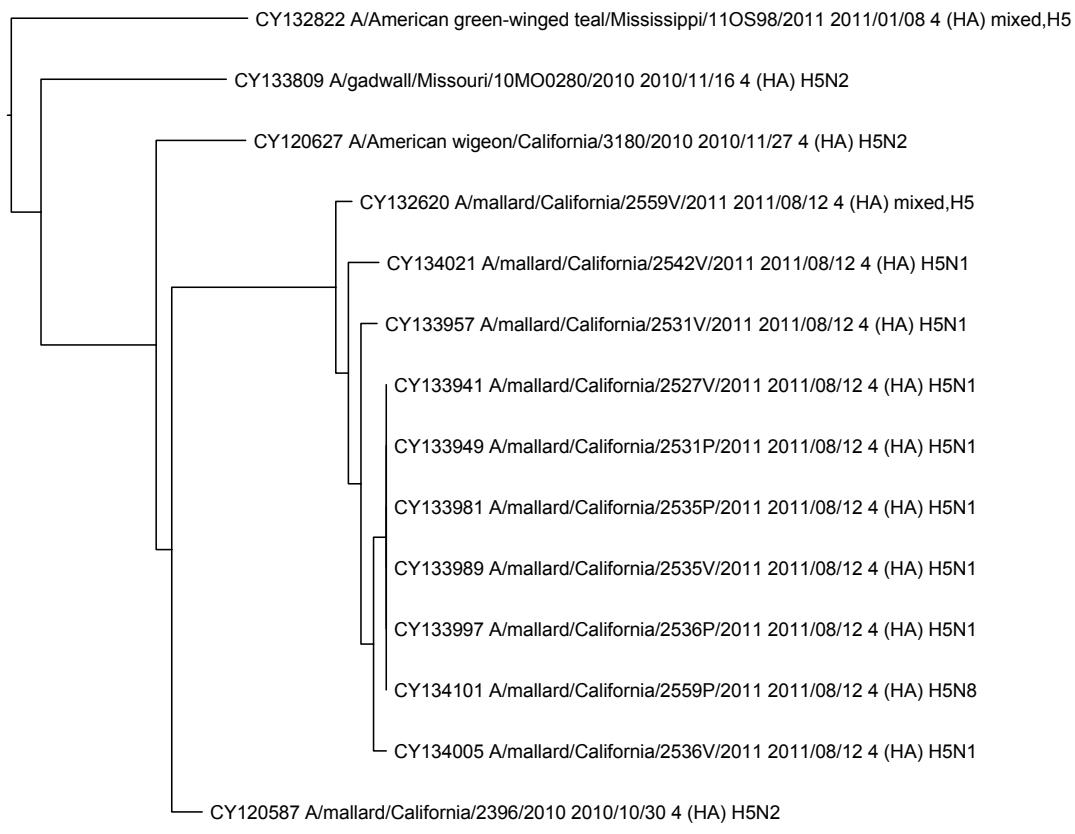
263 Figure 4a: The hemagglutinin clade for New Jersey in 2001.

264

0.007



265 Figure 4b: The neuraminidase clade for New Jersey in 2001.



266

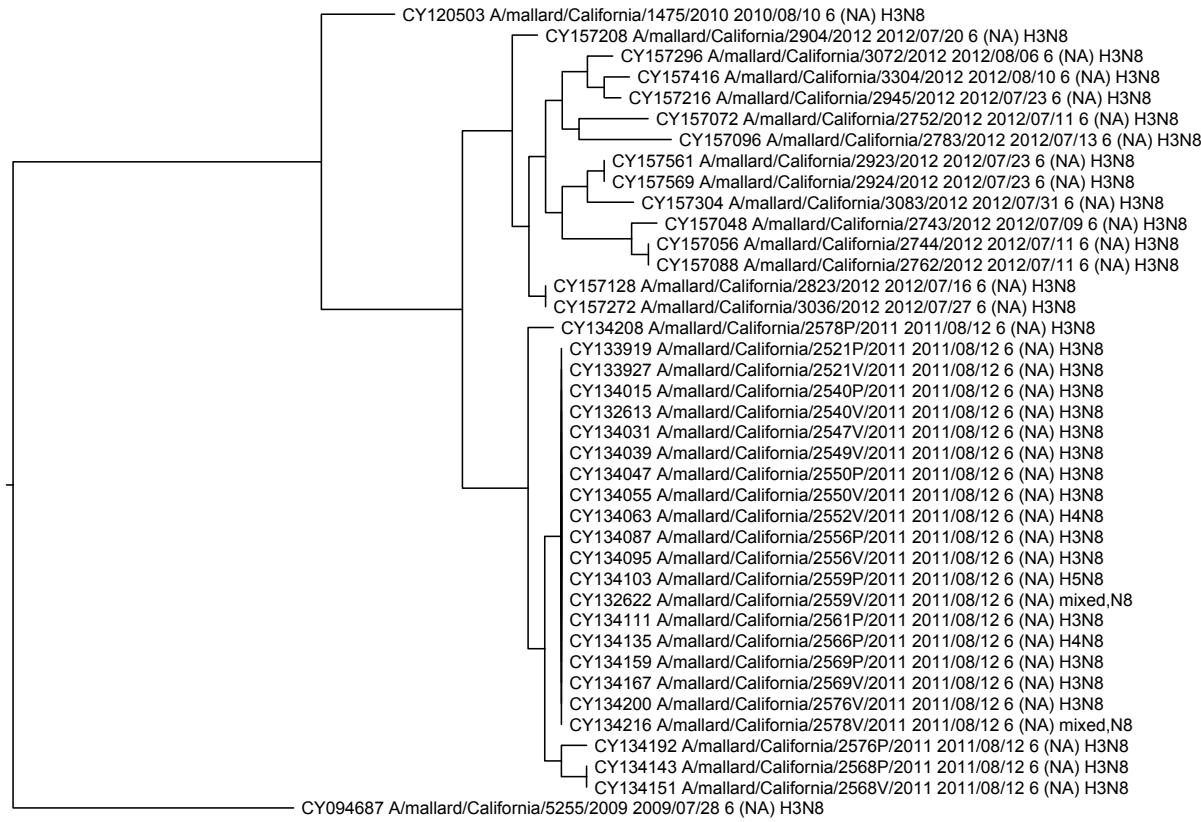
0.002

267 Figure 5a: The hemagglutinin clade for California in 2011.

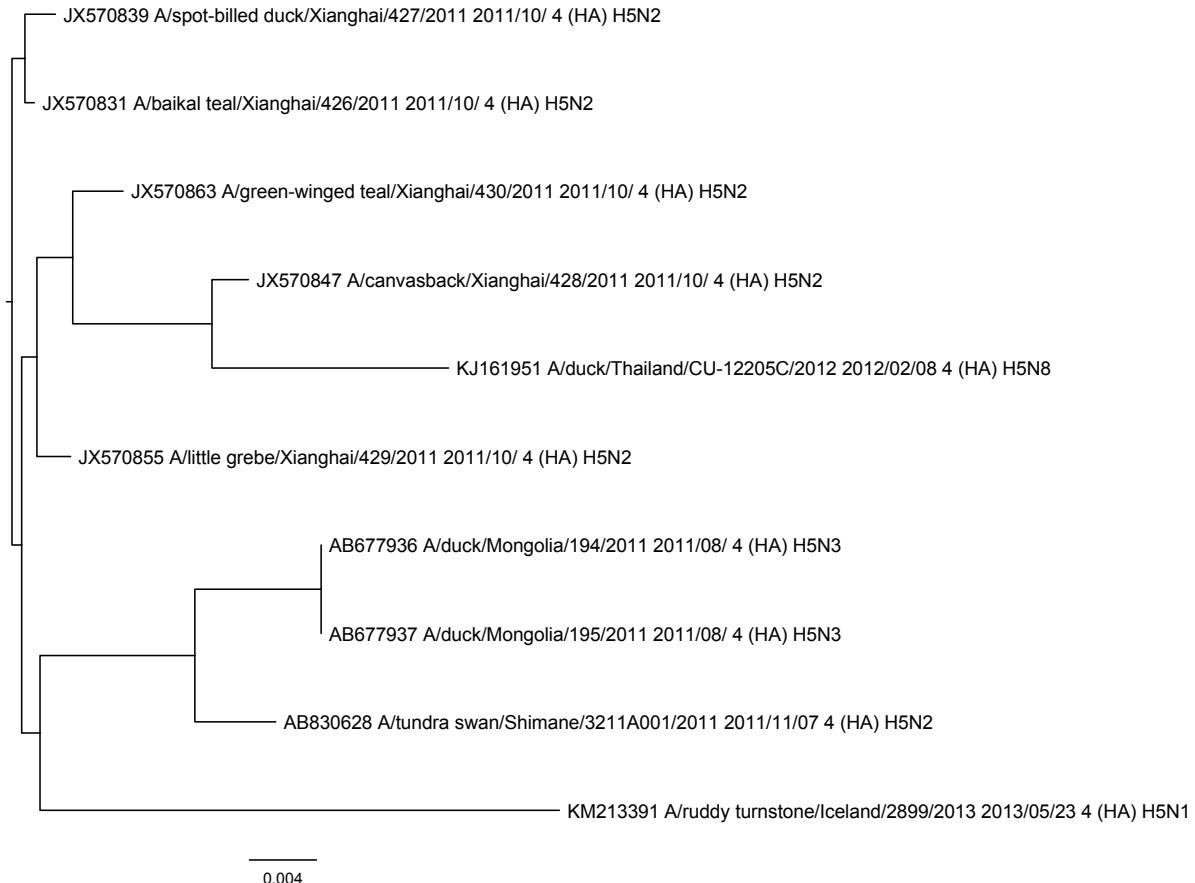
268

269

0.003



270 Figure 5b: The neuraminidase clade for California in 2011.



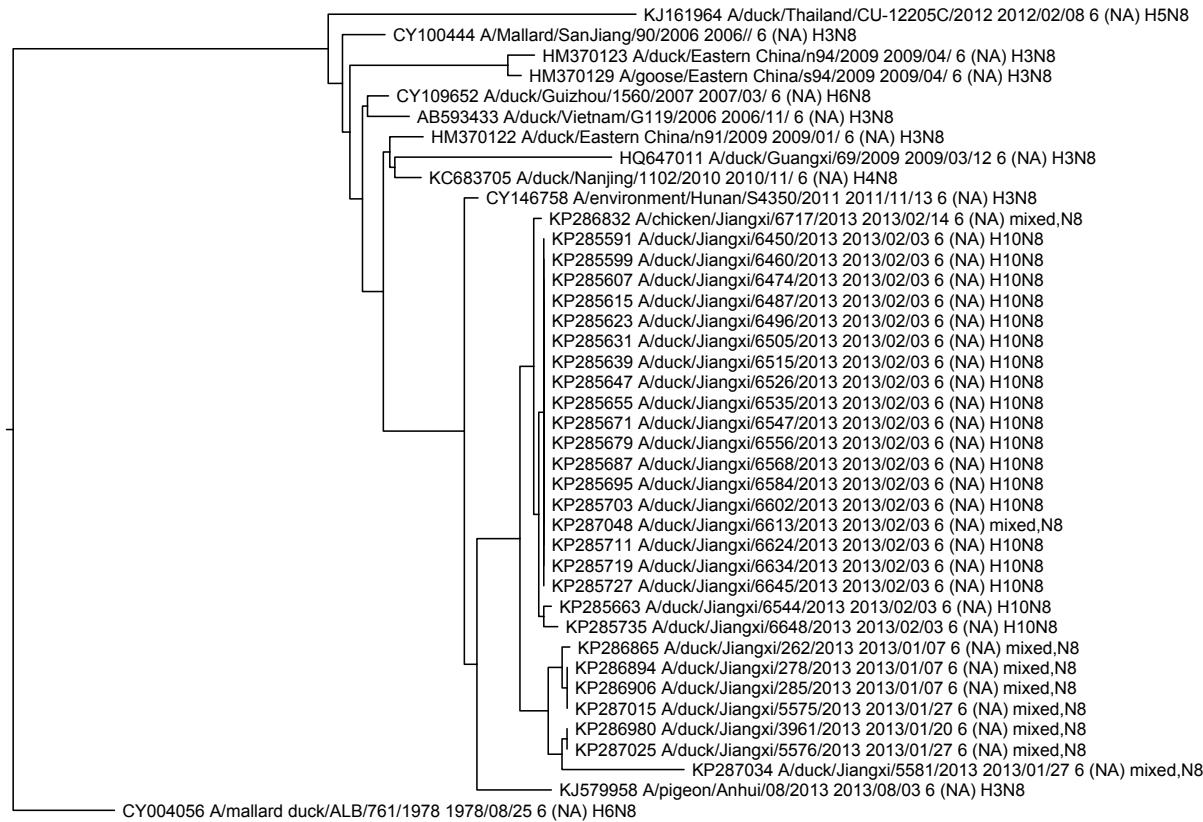
271

0.004

272

273 Figure 6a: The hemagglutinin clade for Thailand in 2012.

274

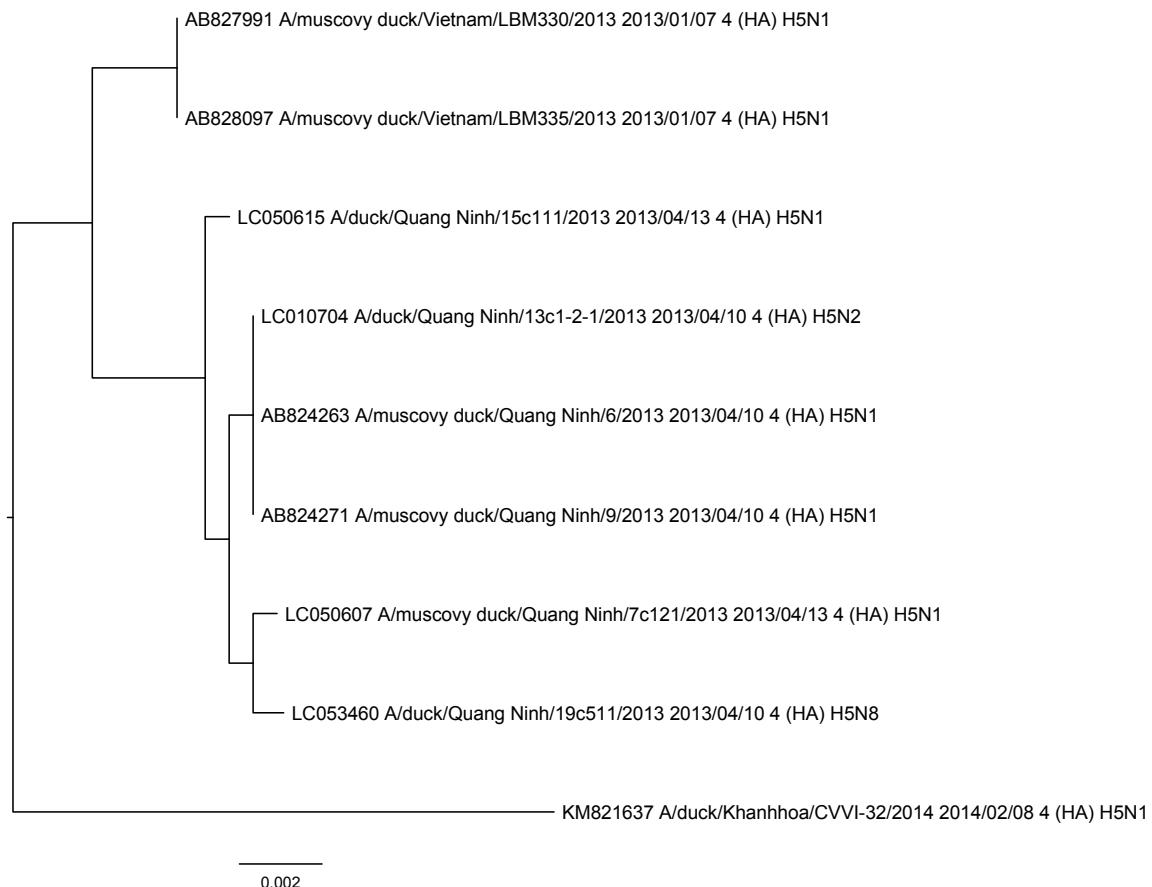


275

0.008

276 Figure 6b: The neuraminidase clade for Thailand in 2012.

277

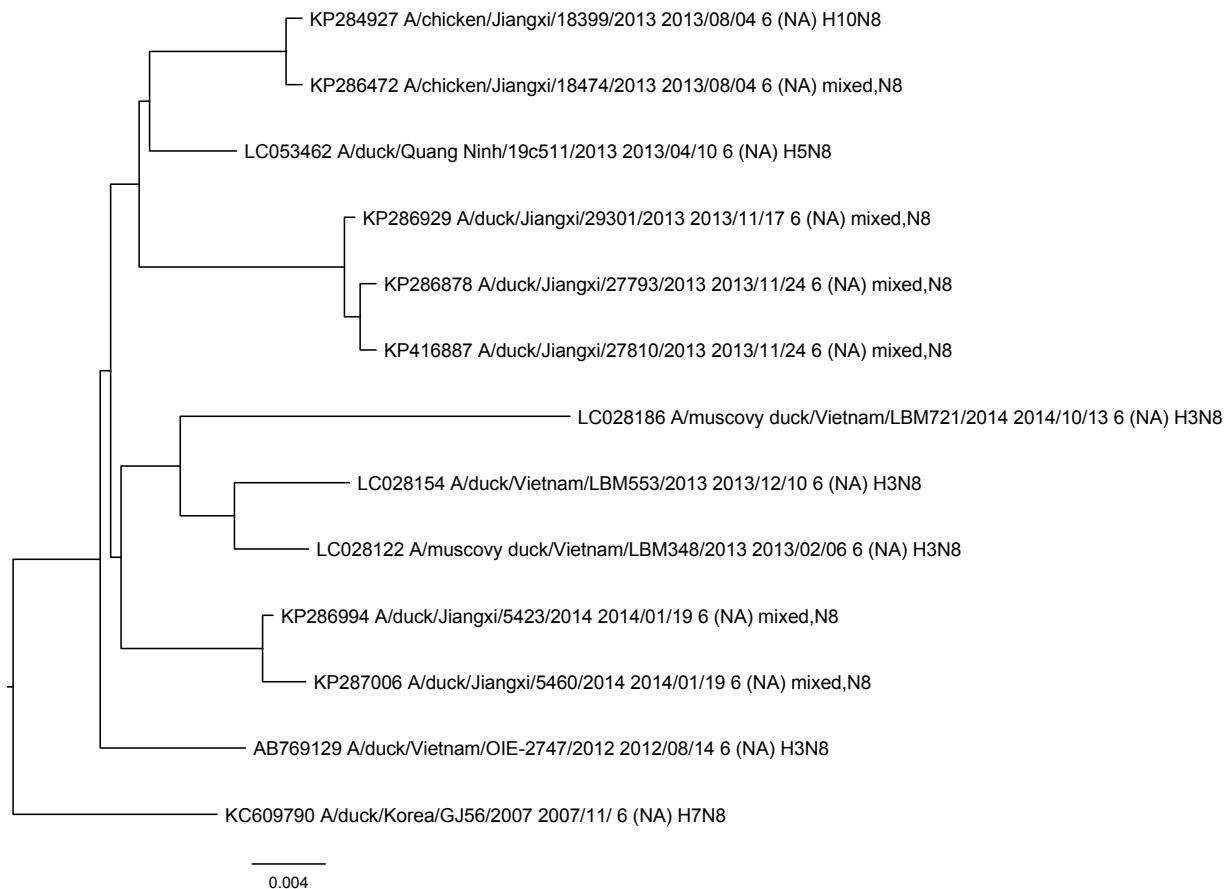


278

0.002

279 Figure 7a: The hemagglutinin clade for Quang Ninh in 2013.

280

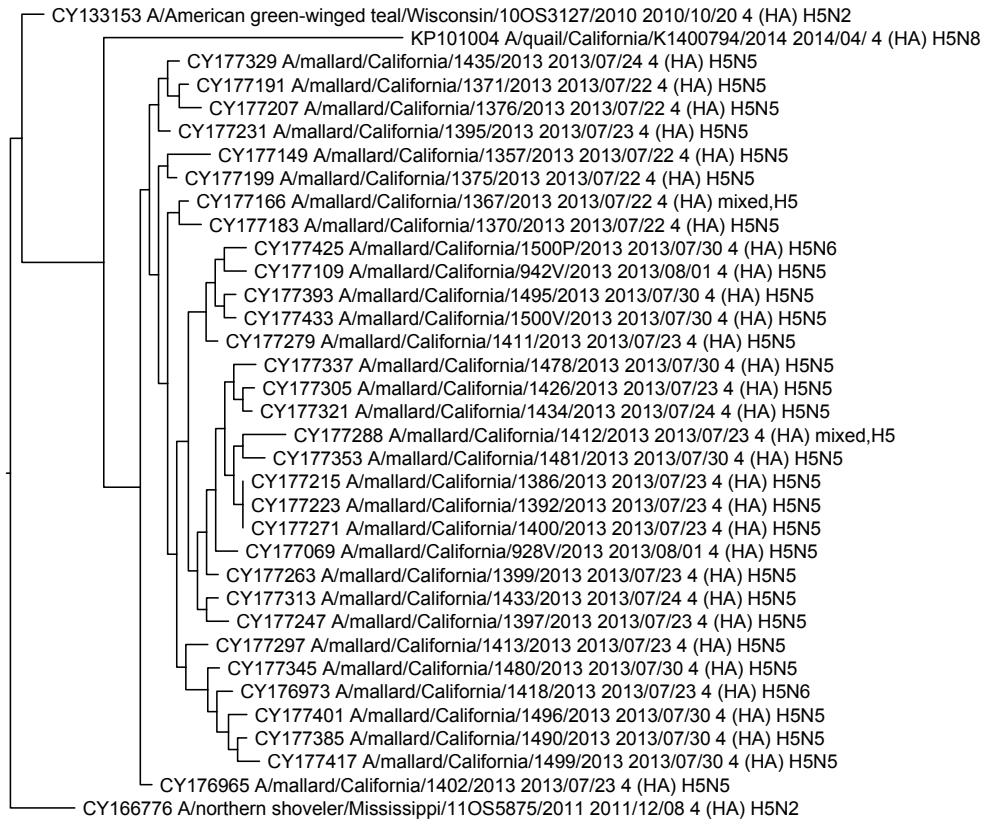


281

0.004

282 Figure 7b: The neuraminidase clade for Quang Ninh in 2013.

283



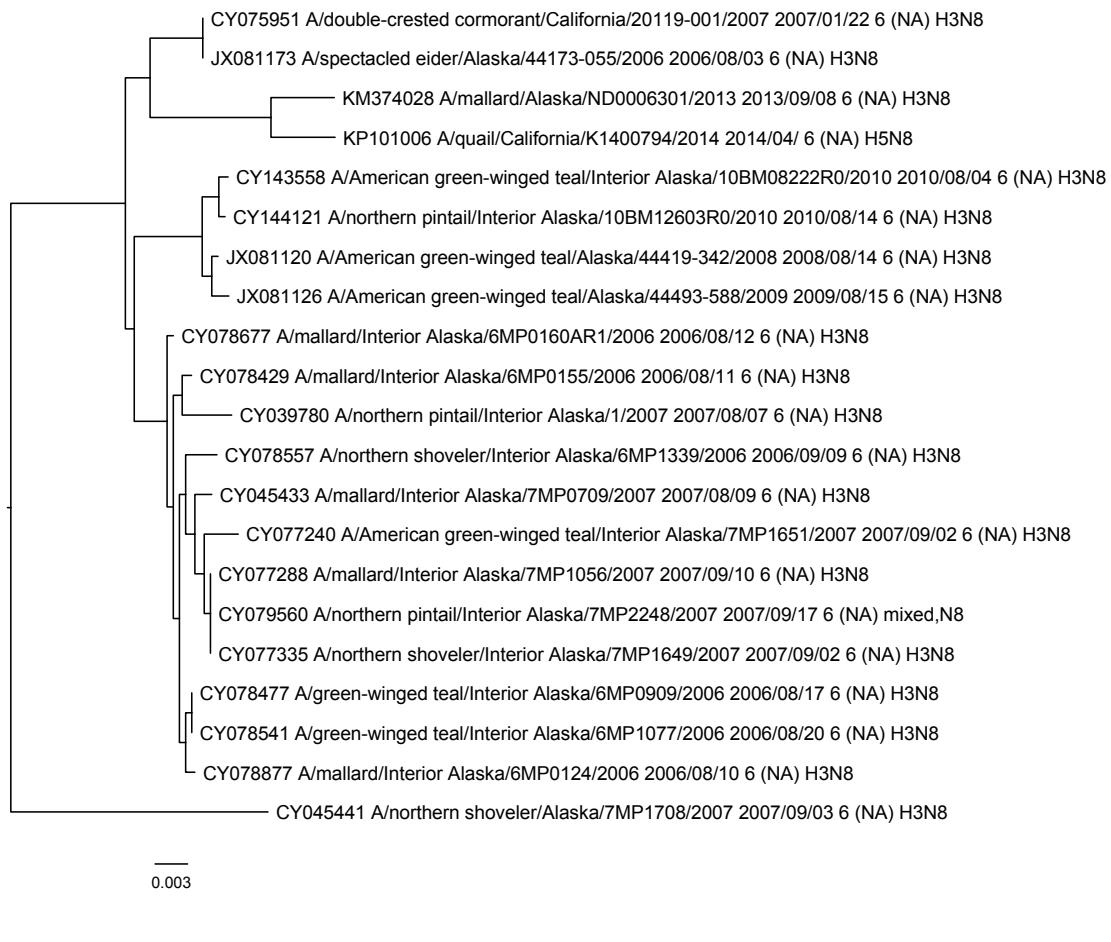
284

0.003

285

286 Figure 8a: The hemagglutinin clade for California in 2014.

287

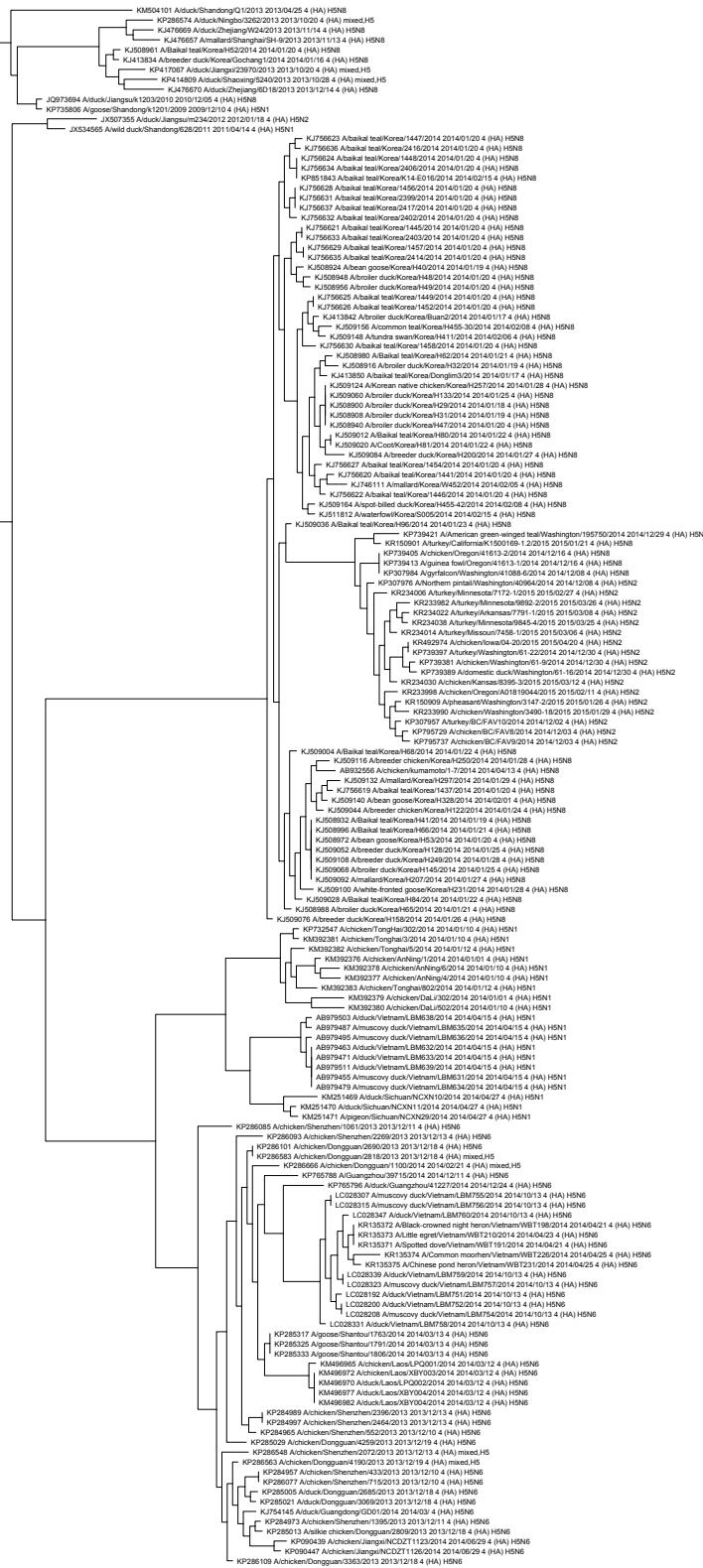


288

0.003

289

290 Figure 8b: The neuraminidase clade for California in 2014.

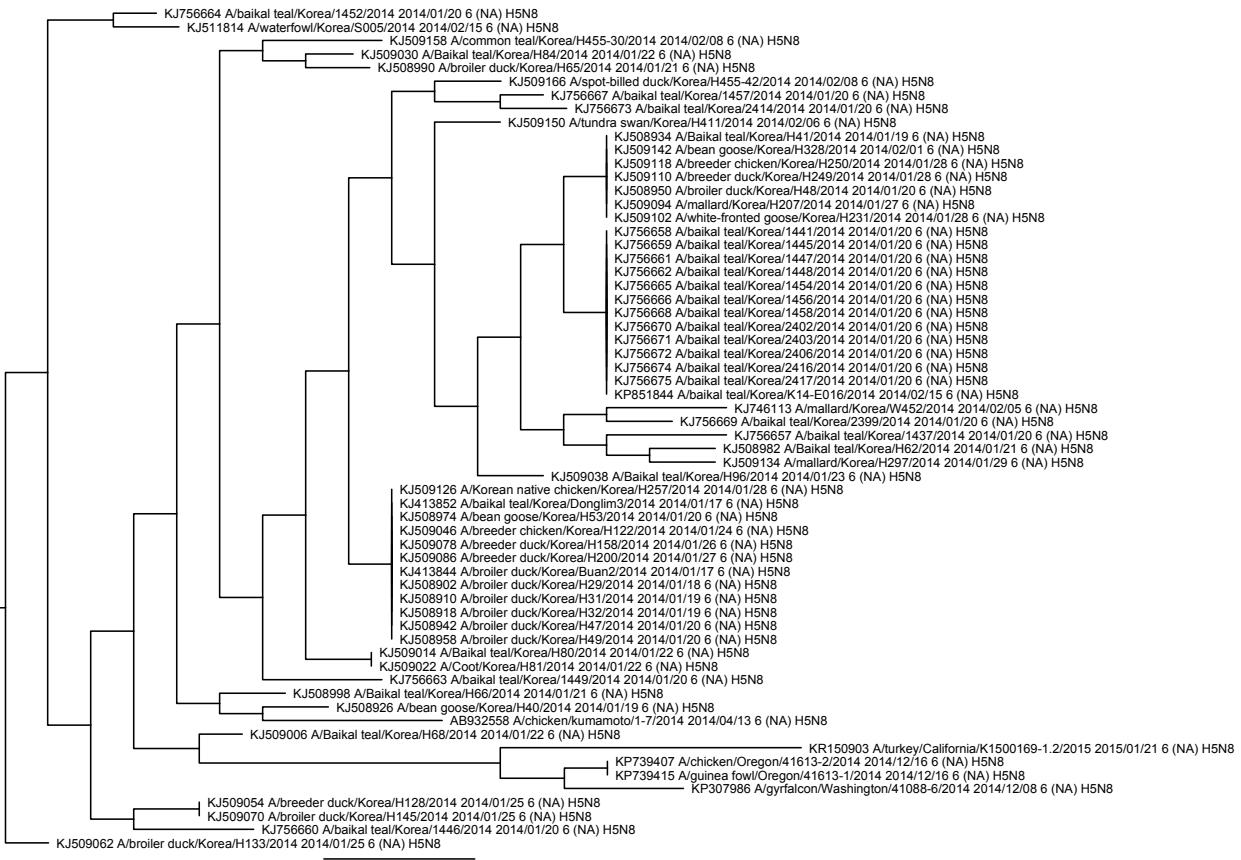


291

0.005

292 Figure 9: The hemagglutinin clade for the current outbreak of H5N8.

293



294

295

296 Figure 10: The neuraminidase clade for the current outbreak of H5N8.

297