

WestminsterResearch

<http://www.westminster.ac.uk/westminsterresearch>

Complete analysis of the H5 hemagglutinin and N8 neuraminidase phylogenetic trees reveals that the H5N8 subtype has been produced by multiple reassortment events

Dalby, A.R.

This is the publisher version of an article published in *F1000Research* 2016, **5**:2463 (doi: [10.12688/f1000research.9261.1](https://doi.org/10.12688/f1000research.9261.1))

It is available from the publisher at:

<https://doi.org/10.12688/f1000research.9261.1>

© 2016 Dalby AR. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk



RESEARCH ARTICLE

Complete analysis of the H5 hemagglutinin and N8 neuraminidase phylogenetic trees reveals that the H5N8 subtype has been produced by multiple reassortment events [version 1; referees: 1 approved with reservations, 1 not approved]

Andrew R. Dalby

Faculty of Science and Technology, University of Westminster, Westminster, W1W 6UW, UK

v1 First published: 06 Oct 2016, 5:2463 (doi: [10.12688/f1000research.9261.1](https://doi.org/10.12688/f1000research.9261.1))
 Latest published: 06 Oct 2016, 5:2463 (doi: [10.12688/f1000research.9261.1](https://doi.org/10.12688/f1000research.9261.1))

Abstract

The analysis of the complete H5 hemagglutinin and H8 neuraminidase phylogenetic trees presented in this paper shows that the H5N8 avian influenza has been generated by multiple reassortment events. The H5N8 strain does not have a single origin and is produced when the H5 hemagglutinin and N8 neuraminidase re-assort from other H5 and N8 containing strains. While it was known that there had been a re-assortment to incorporate the Guangdong H5 hemagglutinin at the start of the Korean outbreak, the results show that there have also been multiple reassortment events amongst the non-Korean sequences.

Open Peer Review

Referee Status: ? X

	Invited Referees	
	1	2
version 1 published 06 Oct 2016	? report	X report
1	Catherine Macken , University of Auckland New Zealand	
2	Justin Bahl , The University of Texas School of Public Health USA, Joseph T. Hicks , The University of Texas School of Public Health USA	

Discuss this article

Comments (0)

Corresponding author: Andrew R. Dalby (A.Dalby@westminster.ac.uk)

How to cite this article: Dalby AR. Complete analysis of the H5 hemagglutinin and N8 neuraminidase phylogenetic trees reveals that the H5N8 subtype has been produced by multiple reassortment events [version 1; referees: 1 approved with reservations, 1 not approved] *F1000Research* 2016, 5:2463 (doi: [10.12688/f1000research.9261.1](https://doi.org/10.12688/f1000research.9261.1))

Copyright: © 2016 Dalby AR. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: No competing interests were disclosed.

First published: 06 Oct 2016, 5:2463 (doi: [10.12688/f1000research.9261.1](https://doi.org/10.12688/f1000research.9261.1))

Introduction

The H5N8 subtype of influenza A virus was first isolated in Ireland in 1983¹. Until the 2014 outbreak in Korea this was a relatively rare subtype that only occurred sporadically. This Korean outbreak was more sustained and over a much wider geographical region than the earlier outbreaks. The Korean outbreak was also important because it resulted from a viral reassortment which contains the highly pathogenic H5 segment originally isolated from a goose in Guangdong in 1996². The epidemiology of the Guangdong H5 containing H5N8 virus has been investigated extensively³⁻⁷ but there is much less investigation of the subtypes that contain the non-Guangdong H5 which has been the predominant hemagglutinin in North America. The most significant finding from these previous studies is that the H5 Guangdong hemagglutinin has now been introduced to North America via bird migratory pathways and that this H5 has not undergone subsequent reassortment into other avian influenza subtypes including H5N2.

An interesting question is why there are large gaps in the history of the sampling of the H5N8 subtype? This is particularly true in the case of the non-Guangdong H5 containing sequences. In order to create a complete history of the H5N8 subtype a phylogenetic analysis of all of the H5 hemagglutinin and N8 neuraminidase sequences was undertaken.

There are three possible explanations for the breaks in detection. They could result from inadequate sampling of the H5N8 avian influenza, the virus might have been present in wild birds, but because avian influenza is often asymptomatic it might be only cryptically expressed, or H5N8 might occur sporadically because it has been created by reassortment events but the new reassorted virus does not spread widely because it is not competitive with alternative reassortment subtypes and so it does not form a continuous population.

A systematic environmental study of bird diseases in the Delaware Bay, as part of the Southeastern Cooperative Wildlife Disease Study⁸ has only reported the H5N8 subtype sporadically providing evidence against inadequate sampling and cryptic expression. With a systematic collection of biological and environmental samples if a sustained viral population had been present at this location then it is likely that it would have been detected even if infection is asymptomatic.

This leaves the alternative hypothesis that the H5N8 virus occurs sporadically as a result of reassortment events, but that these events do not produce a sustainable H5N8 viral population. This hypothesis can be tested by constructing the complete phylogenetic trees of H5 and N8. The H5N8 subtype samples that fall within a single clade of the H5 hemagglutinin and N8 neuraminidase phylogenetic trees will most likely be the product of a single reassortment. If the H5N8 subtype sequences are scattered widely across the phylogenetic trees, then this would indicate multiple reassortment events that have generated the H5N8 subtype from other subtypes.

This paper shows that the H5N8 subtype is distributed widely across both the H5 and N8 phylogenetic trees and that the sporadic nature of H5N8 is a result of multiple reassortment events that have generated the subtype rather than cryptic expression of the virus.

Materials and methods

All of the available H5 hemagglutinin segments (4007 sequences) and N8 neuraminidase segments (1840 sequences) were downloaded from the NCBI Influenza Virus Resource on the 27th of June 2015⁹. The search was restricted to full-length sequences from any host. Manual inspection and editing of the sequences was carried out using Mega6.06¹⁰. During manual editing the 5' end of the sequence was edited to remove the un-translated region. All sequences were trimmed to the start codon and stop codons. Sequences with missing nucleotides were removed.

The H5 clades for the H5N8 subset of hemagglutinin sequences were assigned using the Highly Pathogenic H5N1 Clade Classification tool available as part of the Influenza Research Database¹¹⁻¹³. While this tool was created for the H5N1 subtype, the recent H5N8 outbreak has been identified as belonging to the new 2.3.4.4 subclade that is part of the classifier.

The H5 hemagglutinin and N8 neuraminidase sequences were aligned using Muscle v 3.8.31. FastTree2.1 was used to create a maximum likelihood tree for all of the sequences using the GTR + gamma evolutionary model¹⁴.

```
Fasttree -boot 10000 -nt -gtr -gamma -quote
filename.fas > filename.tree
```

Given the large number of taxons it is computationally challenging to calculate non-parametric bootstrapped trees, instead FastTree calculates a local support values of each of the splits within the tree using the Shimodaira-Hasegawa (SH) log likelihood test¹⁵. This has been shown to have a high correlation to non-parametric bootstrap values¹⁴.

The resulting trees were edited, visualised and annotated with FigTree 1.4.2¹⁶. The trees were displayed as phylograms in order to examine the effect of sampling. The full name, chronological and geographical information was included in the trees as these are essential for determining the homogeneity of the clades. Nodes were labelled with the support values calculated by FastTree, which are a Log Likelihood Ratio. Trees and sub-trees were all rooted to the earliest chronological sequences within the tree.

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

Results and discussion

The H5N8 hemagglutinin classifications according to the pathogenic H5N1 system

Table 1 shows a summary of the H5N8 sequences that are not classified as being members of the 2.3.4.4 clade by the influenza database pathogenic H5N1 classification tool. The complete table of results is given in supplementary table 1. All of the sequences from the Korean outbreak are classified as part of the 2.3.4.4 clade and this also includes many of the 2014 North American sequences but not Californian quail sequence KP101004, which is part of the American non-Guangdong classification. There are two non-Guangdong clades that can be identified as sources of H5 hemagglutinin in H5N8, an American and Eurasian clade.

Table 1. The hemagglutinin sequences that do not belong to the 2.3.4.4 Guangdong Clade.

Sequence	Clade
GQ923573_A/avian/Colorado/456648/2006	American Non-Guangdong
M18450_A/duck/Ireland/113/1983	American Non-Guangdong
GU052853_A/duck/Ireland/113/1983	American Non-Guangdong
LC053460_A/duck/Quang_Ninh/19c511/2013	2.3.2.1c
KJ161951_A/duck/Thailand/CU-12205C/2012	Eurasian Non-Guangdong
CY134101_A/mallard/California/2559P/2011	American Non-Guangdong
KP101004_A/quail/California/K1400794/2014	American Non-Guangdong
CY144236_A/ruddy_turnstone/New_Jersey/828227/2001	American Non-Guangdong
M18451_A/turkey/Ireland/1378/1983	American Non-Guangdong
CY015089_A/turkey/Ireland/1378/1983	American Non-Guangdong
GU052860_A/turkey/Ireland//1983	American Non-Guangdong

The Quang Ninh sequence belongs to the Guangdong grouping but it is part of a different sub-clade, 2.3.2.1c. This subclade contains H5N1 sequences that were found in long range migratory birds such as Geese, Cranes and Whooper Swans in Mongolia and Japan between 2009 and 2011. This suggests that this sequence occurred from a distinct reassortment in migratory birds to that which produced the Korean outbreak.

These results show that the H5N8 hemagglutinins are widely distributed across the H5 clades and that almost all of the North American sequences fall outside of the clades within the current nomenclature system¹². This demonstrates that there have to have been multiple reassortment events between different H5 and N8 clades which have generated the H5N8 subtype. These need to be explored through a more detailed analysis of the complete phylogenetic trees of the H5 hemagglutinin and N8 neuraminidase. Four distinct reassortment events are already clear. One involving the 2.3.4.4 Guangdong clade, another involving the 2.3.2.1c Guangdong clade and two more involving the Eurasian and American non-Guangdong H5 hemagglutinins.

The H5 hemagglutinin phylogenetic tree

The full H5 tree contains 4007 sequences, and is rooted on the 1959 Scottish H5N1 hemagglutinin sequence CY015081 (Figure 1). The tree has been collapsed into two main clades which correspond to American (clade 1) and Eurasian sequences (clade 2). There is then a small cluster of ancestral sequences to these groups that includes the Irish H5N8 sequences (shown in red) that form a subclade with an H5N2 sequence from Italy in 1980 and the German H5N6 and H5N2 sequences from 1984–1985. This Irish group represents the first recorded reassortment that produced the H5N8 subtype. Given that its sequence neighbours both before 1983 and afterward are from the H5N2 subtype it is plausible that the H5N8 hemagglutinin originated in the H5N2 subtype. Within the highly pathogenic avian influenza H5N1 classification the Irish H5 sequences are attributed to the American non-Guangdong clade

but this more detailed analysis show that there is a European ancestral group that predates the American clade and that the Irish sequences belong there.

Figure 2 shows clade 1, the American non-Guangdong clade, rooted on the 1966 H5N9 sequence from a Turkey in Ontario (AB558456). This clade contains the H5N8 sequences from Colorado 2006, a quail in California in 2014, a mallard in California in 2011 and a ruddy turnstone in New Jersey in 2001. Each of these appears as a single sequence in a subclade made up of other non-H5N8 viral subtypes. This suggests that each of these occurrences of H5N8 is the result of a different reassortment event.

The 2001 New Jersey sequence is in a subclade with H5N2 and H5N7 virus sequences collected in the same location in the same year. However, it is not clear which subtype is the source of the H5 hemagglutinin. The 2006 Colorado sequence is part of a subclade with the H5N2 subtype and there are a number of other H5N2 viruses from Arkansas, Minnesota and Wisconsin that are closely related and that were also detected in 2006. It is therefore most likely that a reassortment took place between H5N8 and H5N2. The 2011 California sequence forms a distinct subclade where the other members are all H5N1 sequences also from California and also from 2011. This suggests that the H5N8 is the result of a reassortment between H5N8 and H5N1. The 2011 Californian sequence, 2006 Colorado and California 2014 sequences are the most similar to each other but the clade is dominated by H5 hemagglutinins from the H5N5 subtype.

The quail sequence from California in 2014 is particularly important because this could easily be mistaken as being part of the main outbreak of the Guangdong H5 containing virus which is found in the other American H5N8 sequences from that year¹⁷. However, sequence analysis makes it clear that this is not part of that group and that it is part of the American non-Guangdong clade and that this most likely originated in another reassortment of H5N5 or H5N6 with an N8 containing subtype.

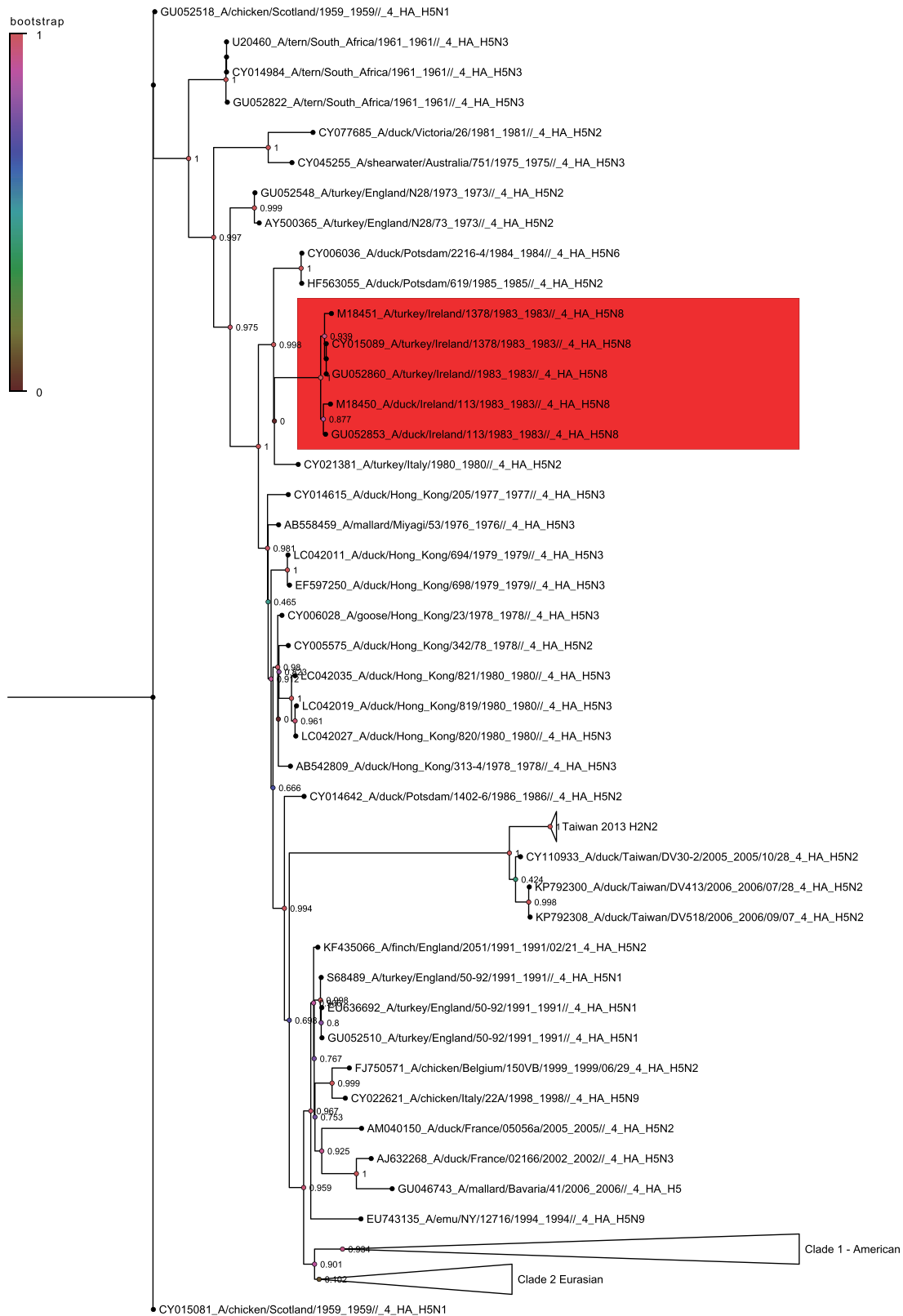


Figure 1. The complete H5 hemagglutinin tree rooted on CY01581 a chicken sequences from Scotland in 1959. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in red.

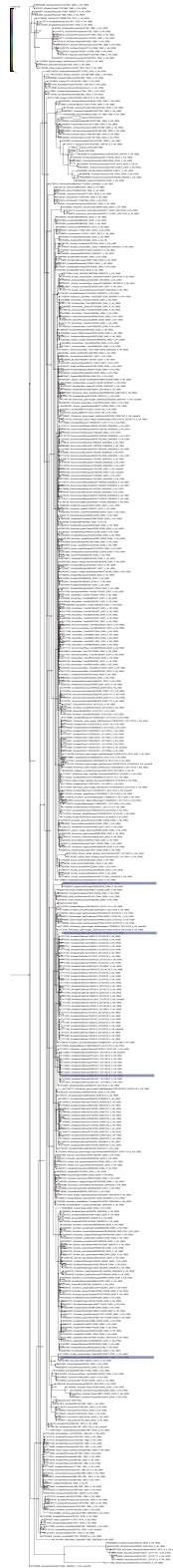


Figure 2. The expanded Clade 1, non-Guangdong American clade of the H5 hemagglutinin tree rooted on CY087808 a turkey sequences from Ontario in 1966. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

Confusion about the sources of outbreaks can affect the measures taken to prevent the spread of the disease. This is especially true in the case of highly pathogenic avian influenza where there is likely to be a significant economic impact if the outbreak cannot be managed successfully. Both the Japanese and European outbreaks were contained and while the highly pathogenic Guangdong H5 was not allowed to spread widely in domestic flocks it was present in wild birds¹⁸⁻²⁰.

Clade 2 is predominantly Eurasian was rooted on the 1991 H5N3 Altai sequence and can be divided into two subclades (Figure 3). One that contains non-Guangdong Eurasian sequences and a second that contains the Guangdong sequences. The non-Guangdong Eurasian sequences include a single H5N8 virus from a duck in Thailand in 2002, which is classified as Eurasian Non-Guangdong under the existing nomenclature¹² (Figure 4). There is a considerable distance between this sequence and any other H5N8 subtype sequences. This is a clear indication that there must have been a reassortment in Thailand to produce the H5N8 and that it was most likely between an H5N2 subtype hemagglutinin and an N8 containing subtype. This new tree shows that the existing nomenclature for H5 sequences outside of the Guangdong lineage does not adequately cover the diversity of this group of sequences.

The Guangdong containing subclade is the lineage that has been studied most extensively and this is taken as the prevailing form of pathogenic H5 containing highly pathogenic avian influenza^{19,21}. Within the Guangdong sequences there are three distinct sets of H5N8 sequences. The large bulk correspond to the Buan H5N8 sequences (Figure 5) that have been previously described¹⁹. This grouping also includes the North American sequences of H5N8 and H5N2 spread by long distance bird migration²². However the Gochang sequences form a distinct grouping along with Chinese sequences from Zhejiang, Shandong and Jiangsu (Figure 6). This strongly suggests that the Gochang and Buan groups are two different reassortment events even though they are both within the 2.3.4.4 highly pathogenic H5 clade¹². The other distinct group contains a single H5N8 sequence from Quang Ninh and is in clade 2.3.2.1c. (Figure 7) again this must have been the product of another reassortment event.

These results from the hemagglutinin trees suggest that there have been a minimum of five reassortment events within the American non-Guangdong sequences, another within the Eurasian non-Guangdong sequences and at least two more in the Guangdong clade. This makes a total of at least eight separate reassortment events that have produced H5N8 from other subtypes. Examining the N8 neuraminidase tree can be used to confirm these reassortment events and to show that absences of H5N8 in the chronological record do not result from poor sampling.

The N8 neuraminidase trees

The neuraminidase trees are harder to summarise even though they contain far fewer sequences. This is because they have a less clear clade structure. Figure 8 shows the N8 phylogenetic tree rooted on the 1963 Ukrainian duck sequences. Once again, the H5N8 subtype is close to the root of the tree in the case of the initial Irish outbreak in 1983. There is however, only a single Irish H5N8

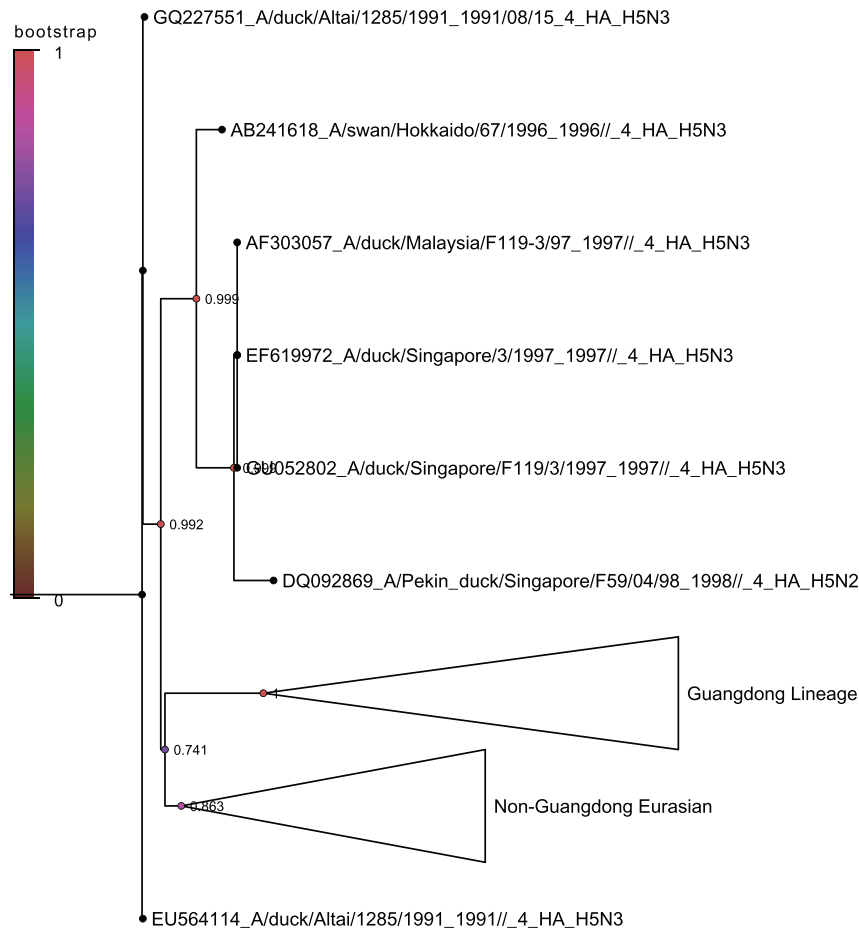


Figure 3. The expanded Clade 2, Eurasian clade of the H5 hemagglutinin tree that includes the Guangdong lineage. This tree is rooted on the EU564116 sequence from a duck in Altai in 1991. Nodes are labelled and coloured with the local bootstrap likelihood values. There are both non-Guangdong and Guangdong subclades.

sequence amongst a group of H3N8 sequences. Unlike the H5 hemagglutinin there is no existing clade nomenclature for the N8 neuraminidase sequences.

From the neuraminidase tree there are four distinct clades but these are much more heterogeneous than in the case of the H5 trees. The four clades correspond roughly to long range migratory birds, far eastern migratory birds, Gochang and Buan and finally an American clade. Only the Gochang and Buan and the American clade contain the H5N8 subtype and will be considered here.

The simplest of these clades to view is the Gochang and Buan clade where the H5N8 subtypes are clustered tightly together in one subclade with only a few non H5N8 sequences (Figure 9). This highly homogeneous clade is very different to the mixtures of subtypes found in the other clades. The structure of this clade suggests that the Gochang and Buan H5N8 viruses originate from a single source of the N8 neuraminidase but that they divided from one another before the Korean outbreak. This is in agreement with

the hemagglutinin trees which show that the H5 hemagglutinins for the Buan and Gochang groups have two more distinct origins. This shows that reassortment can occur on a local level between closely related sequences and produce multiple lineages in the same geographical location.

Clade 4 (Figure 10) contains the seven distinct H5N8 sequences that are distributed widely across the clade as singletons. None of the sequences are adjacent to each other in the tree and most are in distantly related subclades. Clade 4.1 (Figure 11) contains the 2001 New Jersey and 2011 Californian H5N8 sequences along with the 2006 Colorado and 2013 California and 2013 Quang Ninh sequences within sub-sub-clade 4.1.1. This agrees with the results of the H5 hemagglutinin sequences phylogenetic tree and is strong evidence that each of the H5N8 viruses corresponds to a reassortment event. What is more significant is the presence of the Quang Ninh sequence amongst the North American sequences as this shows that the N8 neuraminidase circulates more widely than the different H5 hemagglutinin lineages. Clade 4.2 contains the 2002

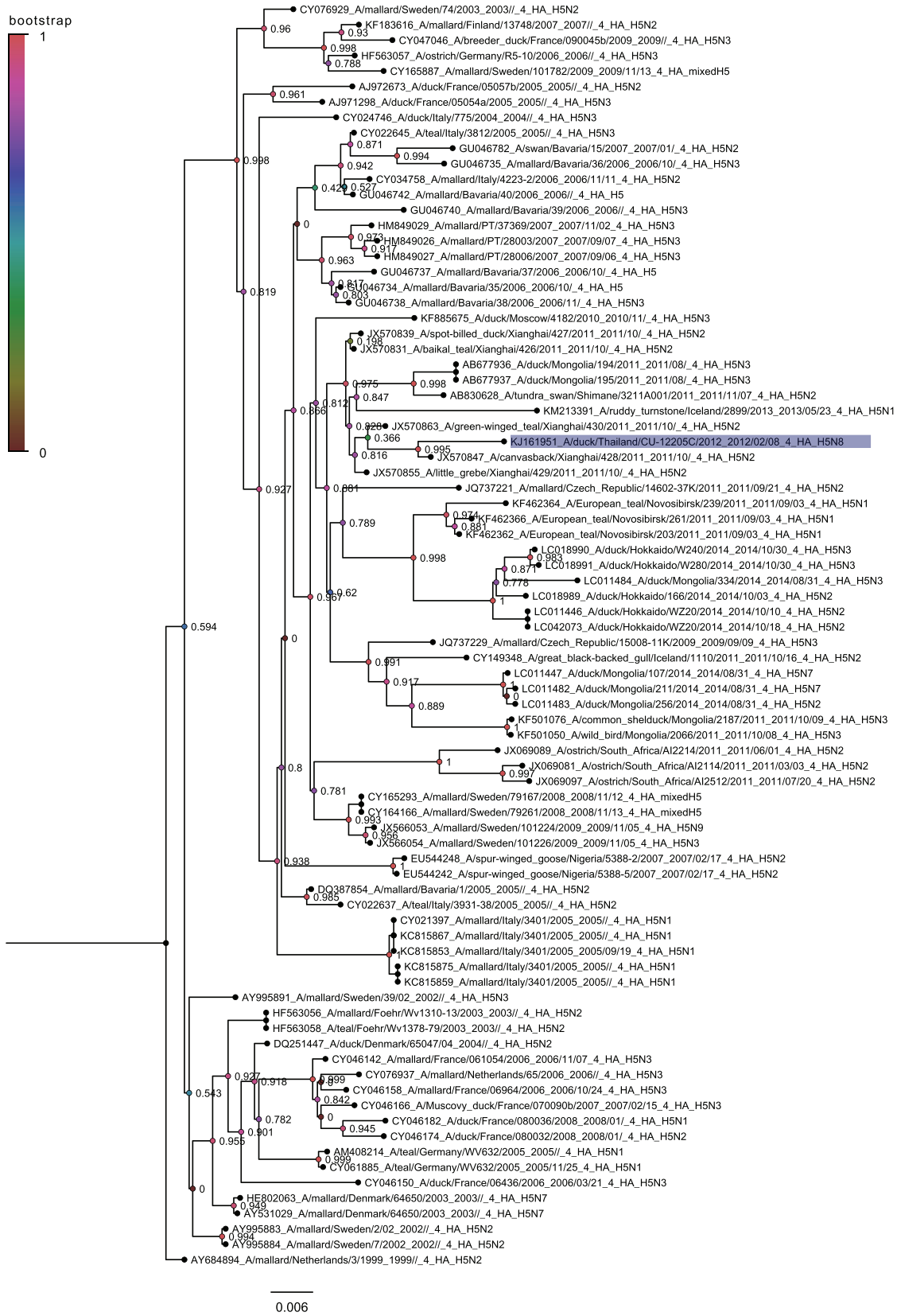


Figure 4. The non-Guangdong Eurasian clade rooted on the AY684894 Dutch mallard sequences from 1999. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

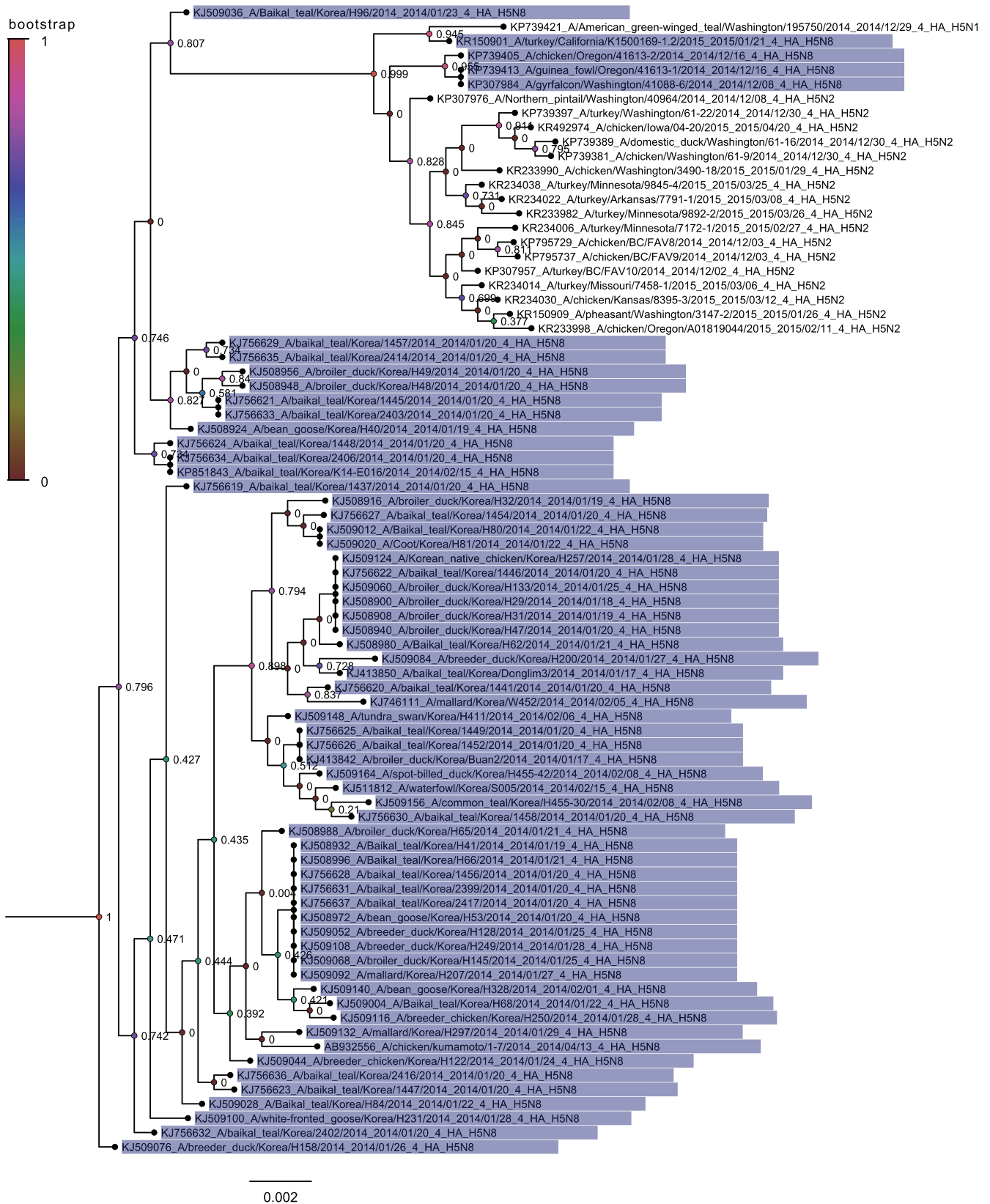


Figure 5. The Buan sequences within the Guangdong subclade. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

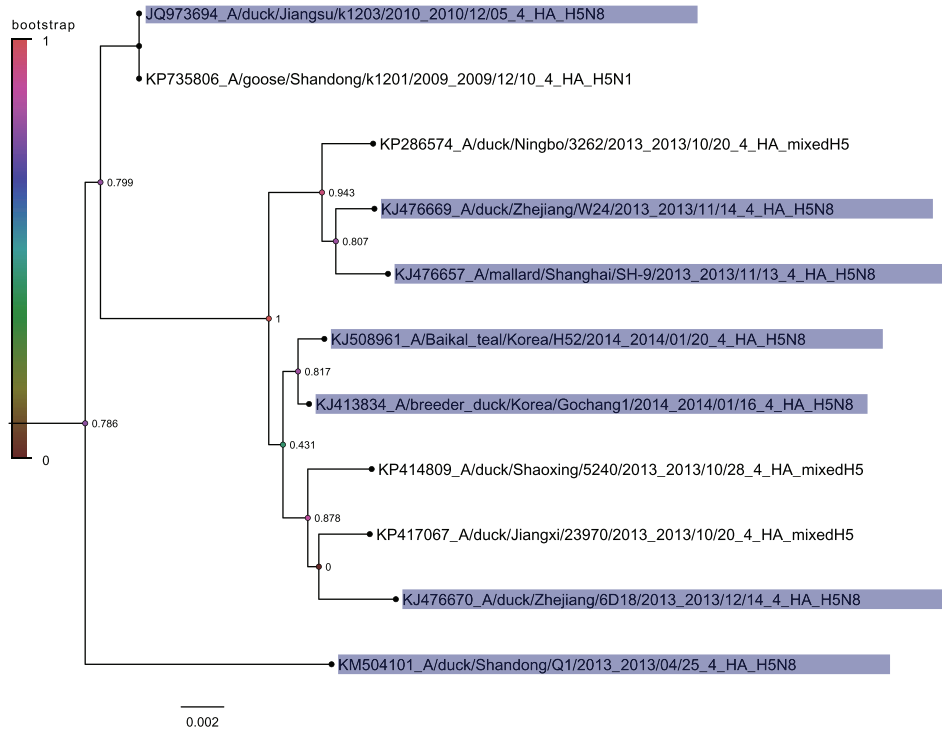


Figure 6. The Gochang sequences within the Guangdong subclade. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

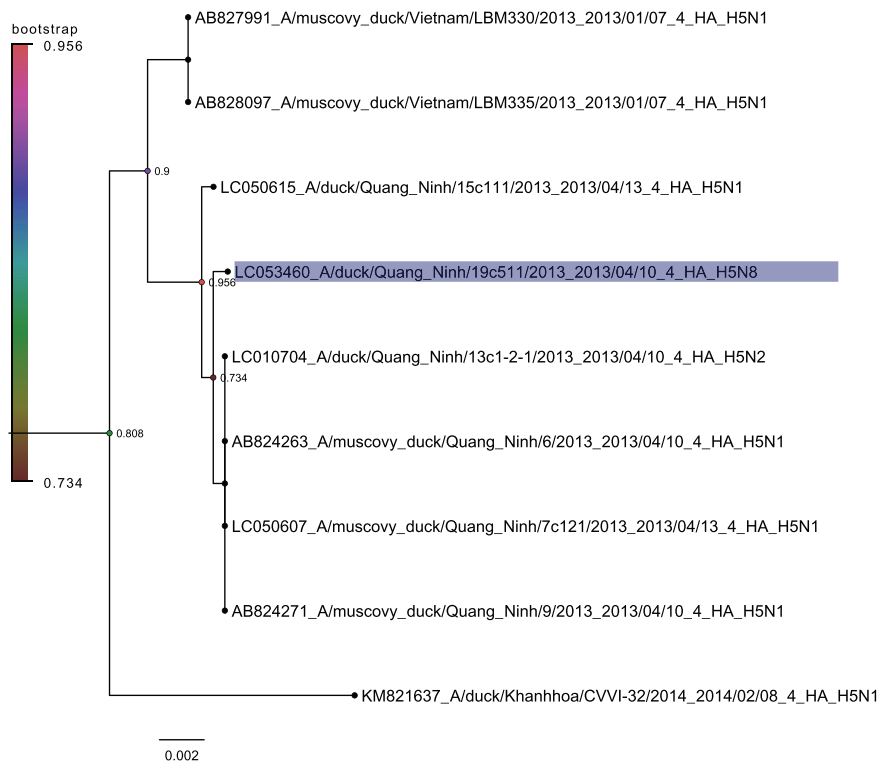


Figure 7. The Quang Ninh sequence within the Guangdong subclade. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequence is highlighted in blue.

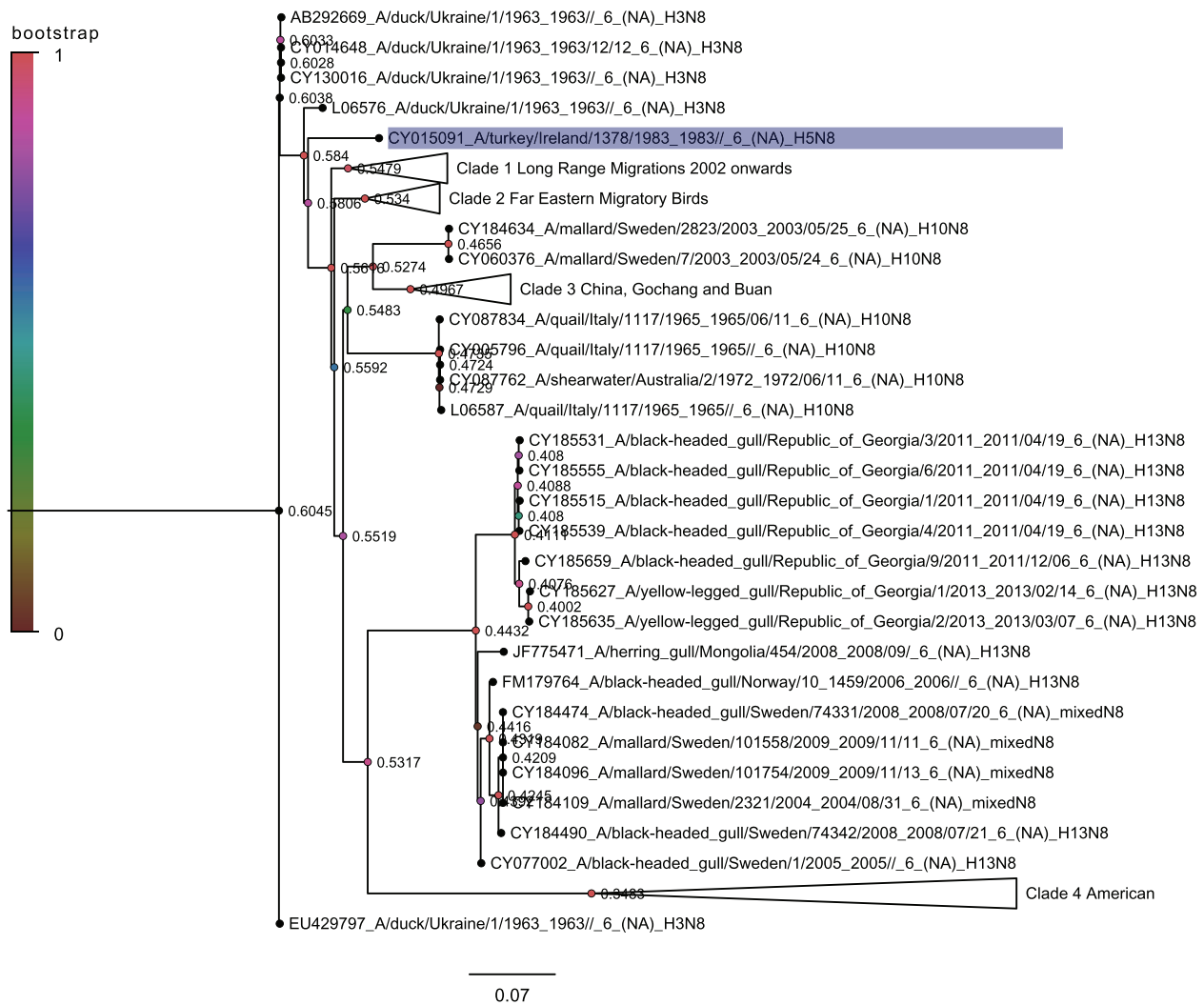


Figure 8. The complete N8 neuraminidase tree rooted on EU429797 a duck from the Ukraine in 1963. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequence is highlighted in blue.

New York sequence (Figure 13) and clade 4.3 contains the 2012 Thailand sequence (Figure 14). Again this demonstrates how there is a global dispersion of the N8 neuraminidase clade 4, compared to the Guangdong H5 which until 2014 was not present in North America.

Conclusions

The results presented here show that there is a high degree of reassortment that generates new influenza subtypes. The presence of a high proportion of singleton sequences shows that most of H5N8 is often not the preferred subtype as it did not produce a wide-spread outbreak. The Guangdong H5 containing reassortment has produced an H5N8 capable of wider circulation. However it is still possible that once again H5N8 will die out only to return sporadically given that the virus has already undergone further reassortment in North America to produce an H5N2 Guangdong H5 containing subtype²³.

In a wider context, reassortment events that create new subtypes need to be accounted for before reliable phylogenetic analysis can be carried out. Sampling for tree construction based on viral subtype, without any consideration of reassortment will be misleading. If we ignore these reassortment events, then we will introduce sampling bias to the trees. Sampling bias is introduced because you selectively sample sequences within a clade that share the same subtype whereas many of the neighbouring hemagglutinin and neuraminidase sequences may actually be from other subtypes.

Where phylogenetic analysis focuses on a viral segment of a specific subtype a complete analysis of all of the sequences for from all subtypes for that segment, as has been performed here, is rarely carried out. The criteria for including sequences in these analyses are usually based on chronological or geographical limits, but these limits reduce the generalizability of the hypothesis being

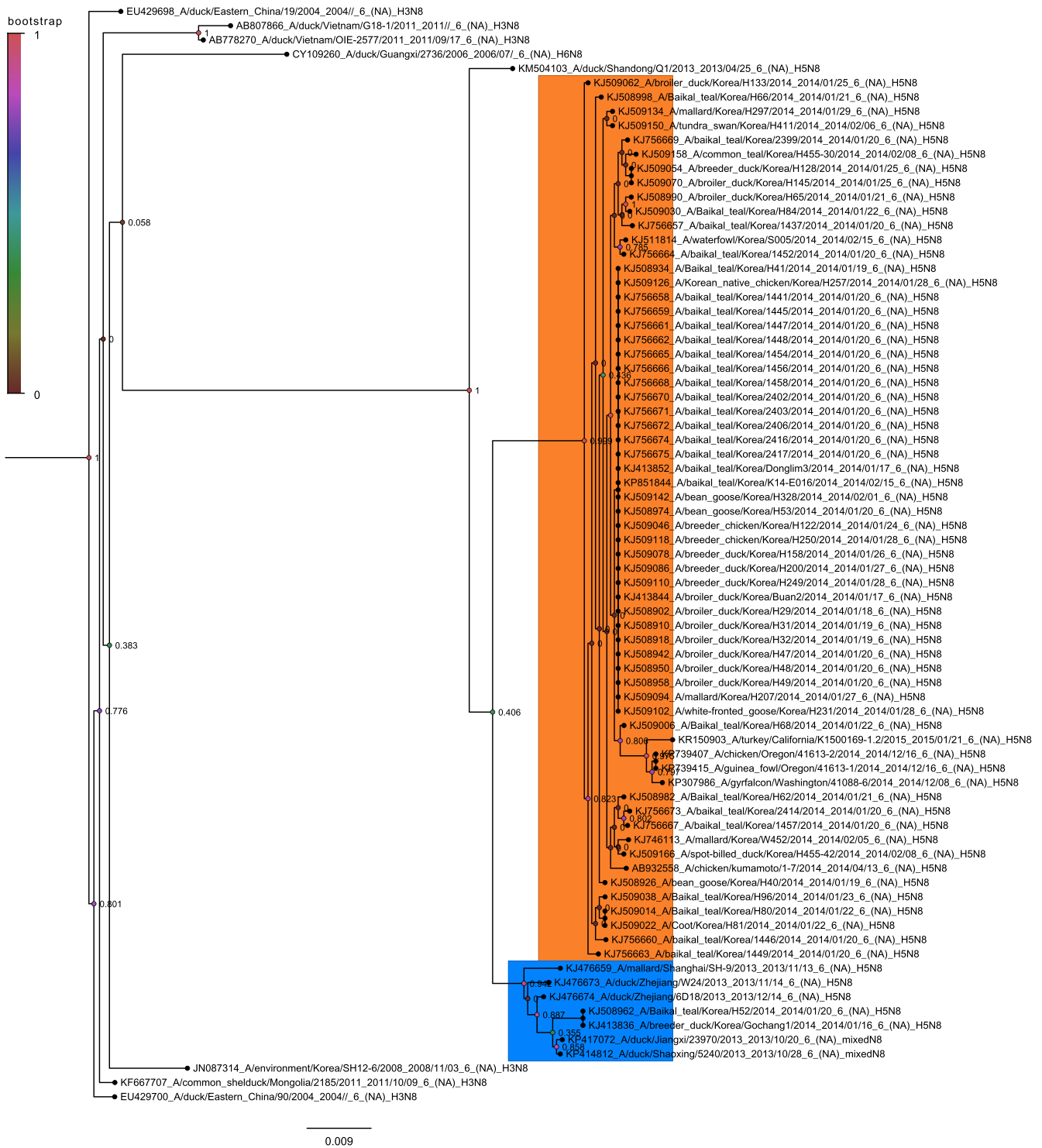


Figure 9. The Gochang and Buan containing clade of the N8 neuraminidase tree. The tree is rooted on the EU429700 sequence from a duck in Eastern China in 2004. The Buan sequences are highlighted in orange and the Gochang sequences are highlighted in blue. Nodes are labelled and coloured with the local bootstrap likelihood values.

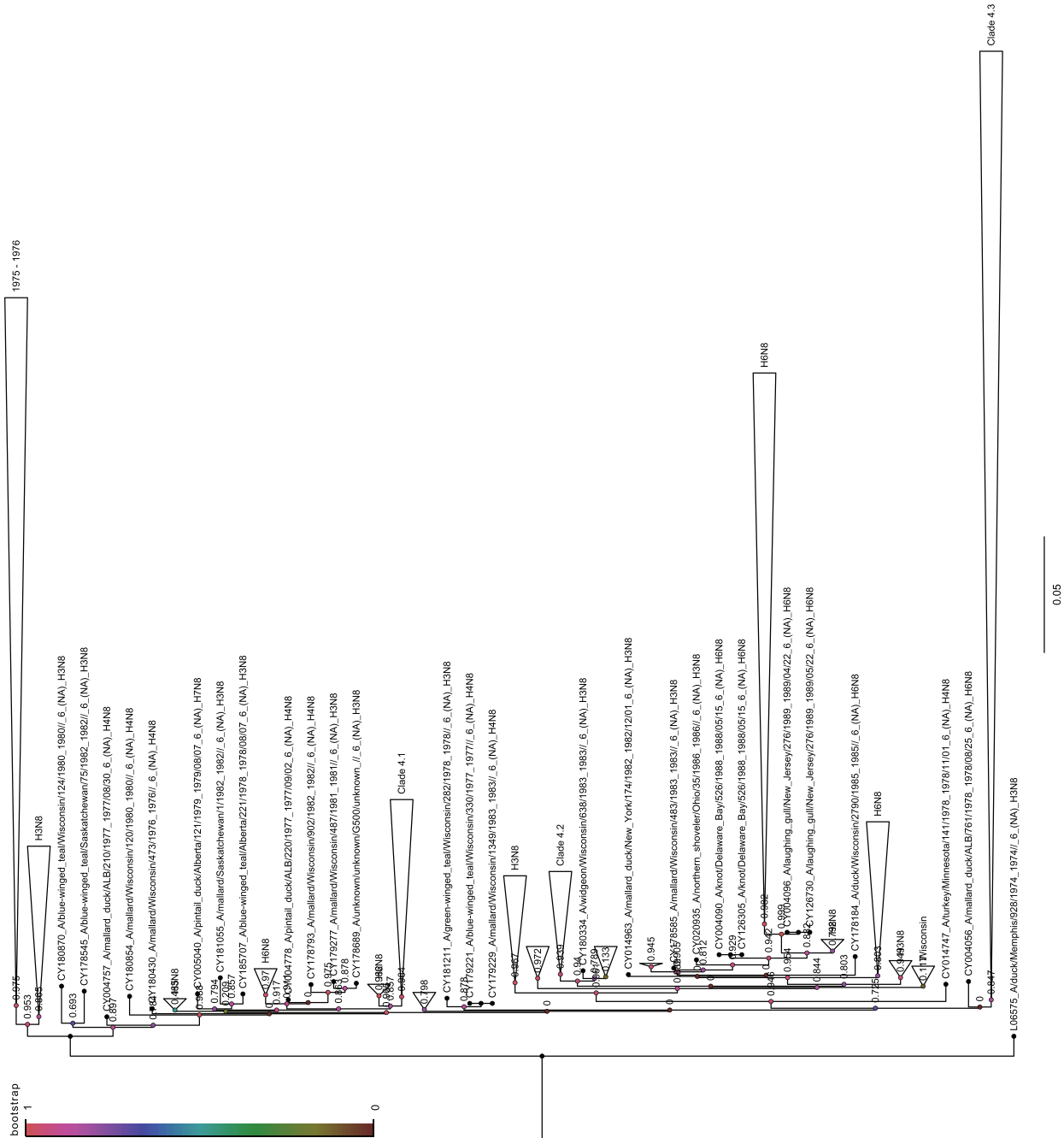


Figure 10. The American N8 neuraminidase clade, clade 4 from the main tree. This is a collapsed view, nodes are labelled and coloured with the local bootstrap likelihood values.

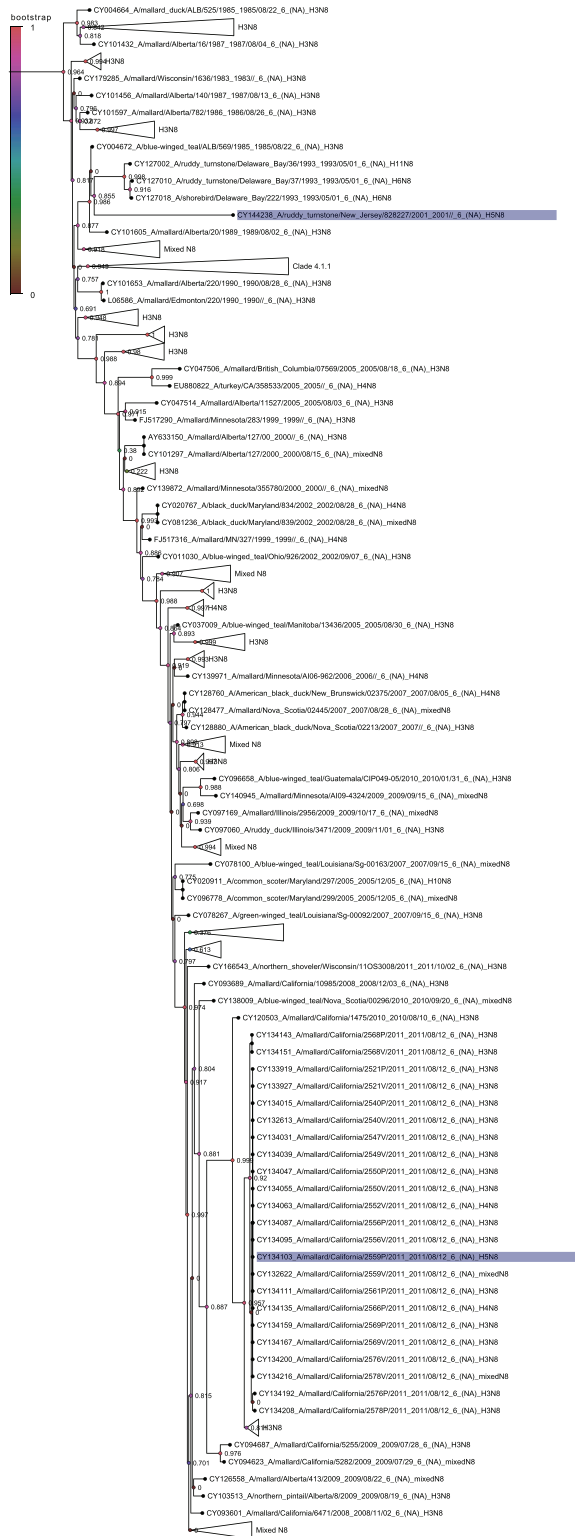


Figure 11. Sub-clade 4.1 of the N8 neuraminidase tree. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

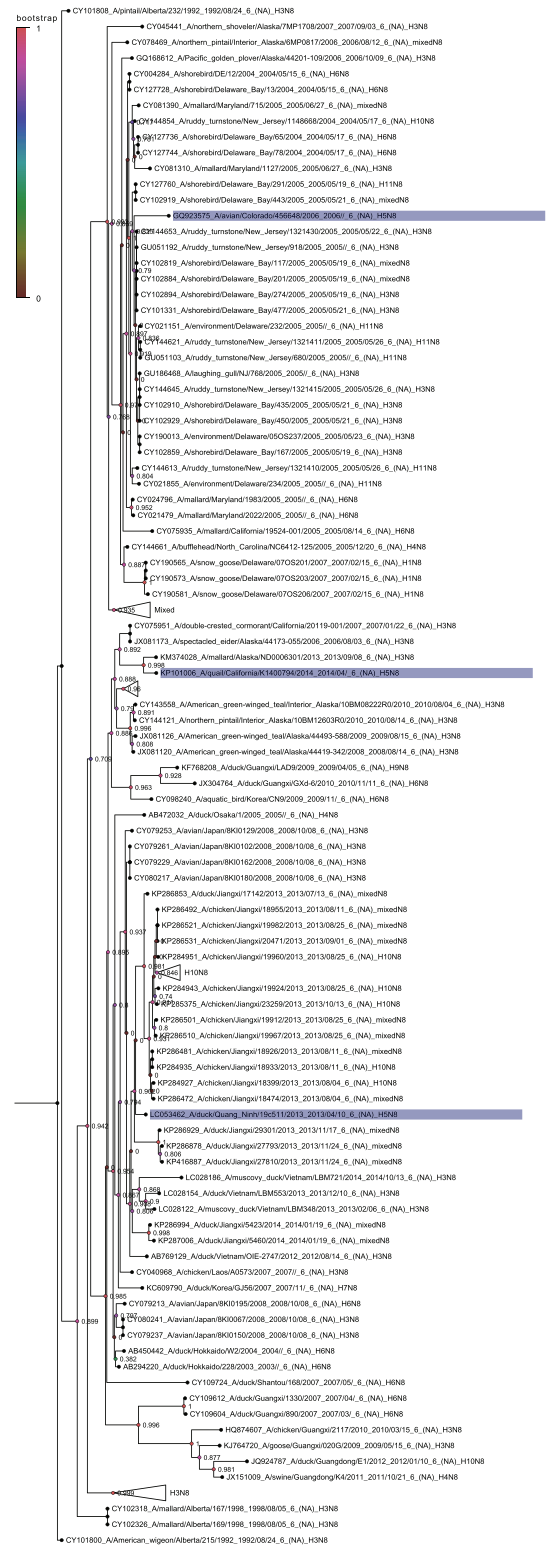


Figure 12. Sub-sub-clade 4.1.1 of the N8 neuraminidase tree. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequences are highlighted in blue.

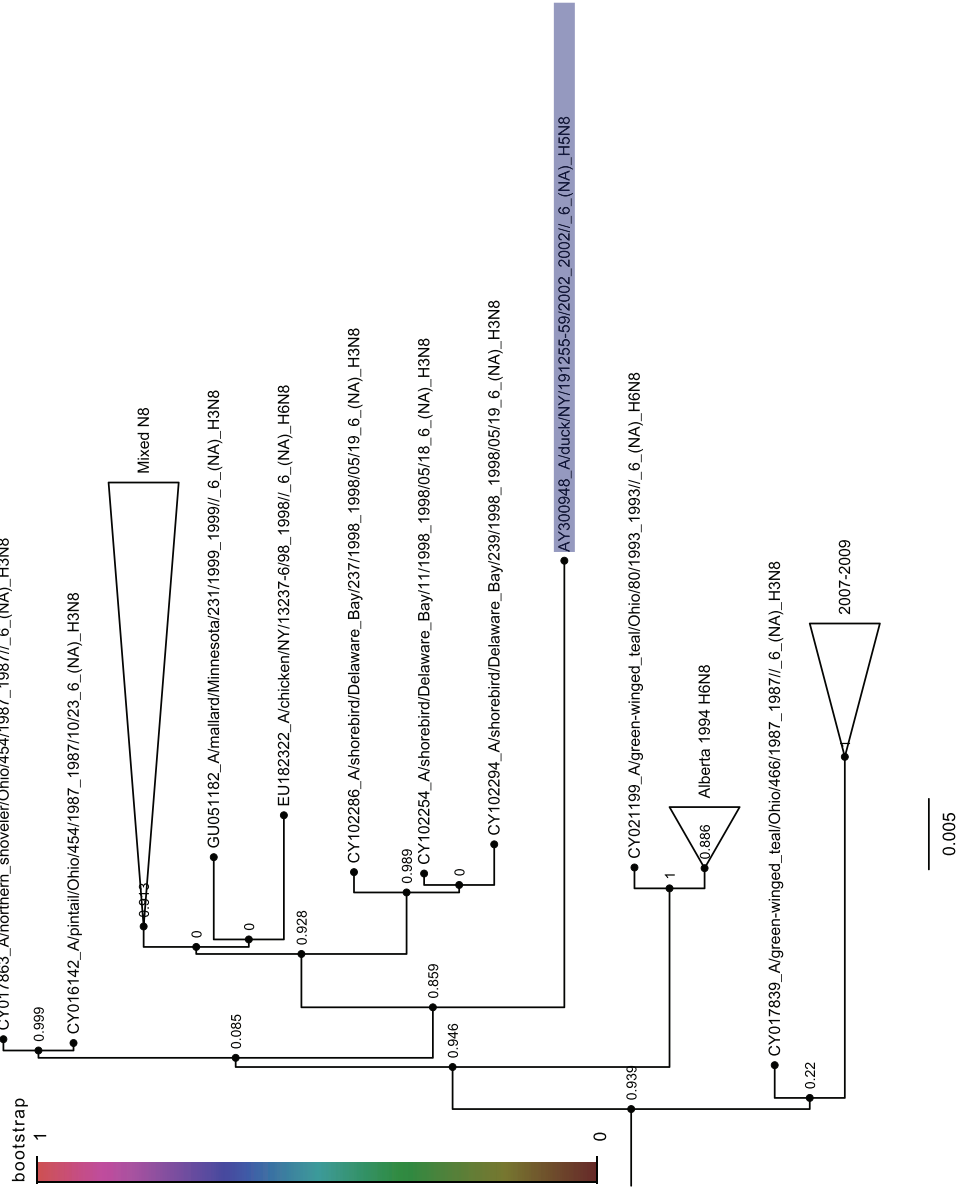


Figure 13. Sub-clade 4.2 of the N8 neuraminidase tree. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequence is highlighted in blue.

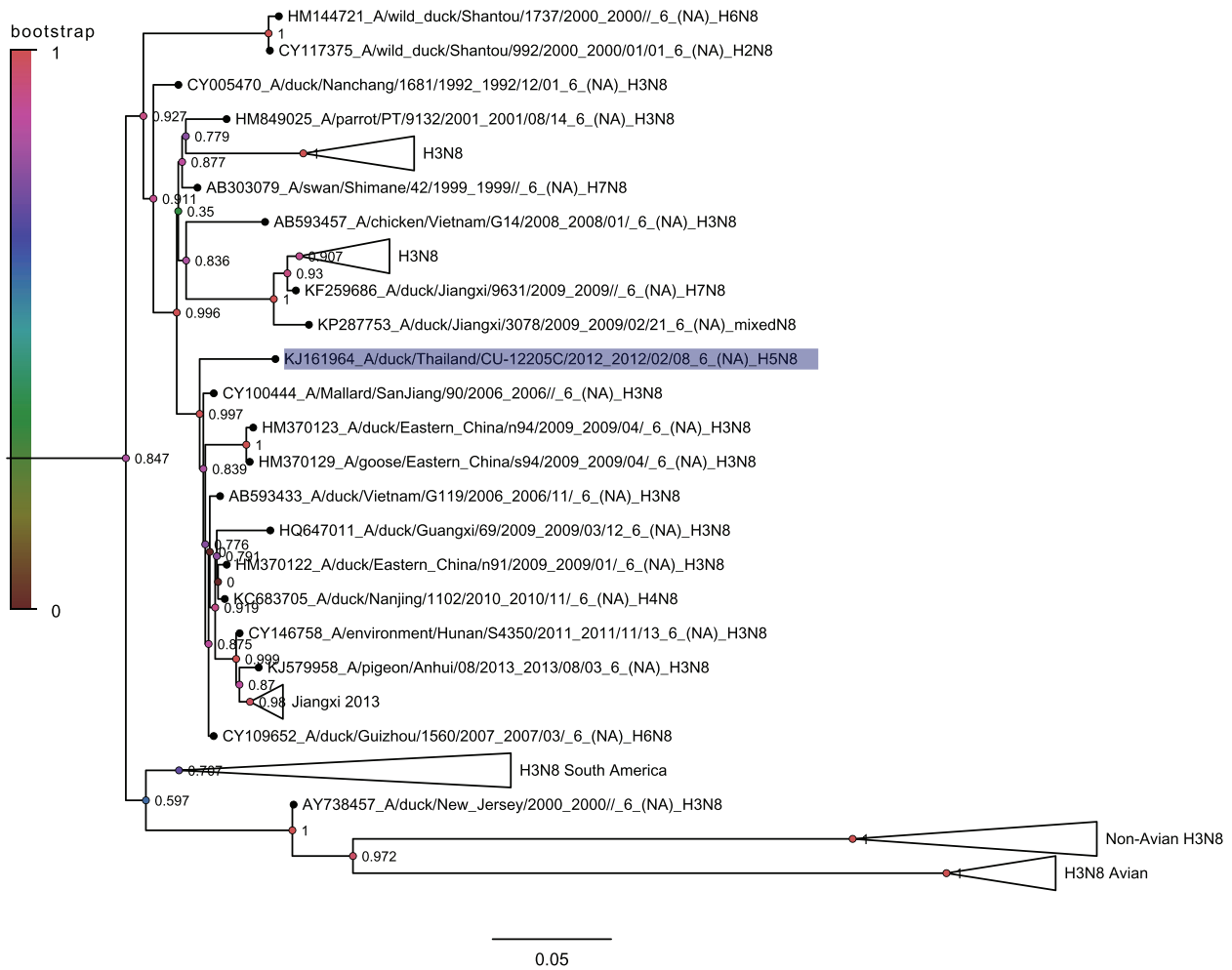


Figure 14. Sub-clade 4.3 of the N8 neuraminidase tree. Nodes are labelled and coloured with the local bootstrap likelihood values. The H5N8 sequence is highlighted in blue.

considered. To make sure that sampling is effective a complete phylogenetic analysis of that segment is required. After this then clades and subclades can be selected for further analysis using geographical or chronological criteria. In this way the only bias introduced is that from sequence collection and availability.

This analysis only considered reassortment from the perspective of the glyco-proteins as the reassortment of these protein produces a novel influenza sub-type. Further analysis needs to also include the other viral segments in order to provide a more complete picture of reassortment in avian influenza.

Data availability

ZENODO: Phylogenetic Analysis of the Influenza H5 hemagglutinins, doi: [10.5281/zenodo.206534](https://doi.org/10.5281/zenodo.206534)

ZENODO: Phylogenetic analysis of the influenza N8 neuraminidases, doi: [10.5281/zenodo.206525](https://doi.org/10.5281/zenodo.206525)

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no supporting grants were involved in this work.

Acknowledgments

I would like to thank an anonymous referee from a preceding paper for pointing out the importance of identifying reassortment events in sampling for phylogenetic analysis.

References

1. Alexander DJ, Parsons G, Manvell RJ: **Experimental assessment of the pathogenicity of eight avian influenza A viruses of H5 subtype for chickens, turkeys, ducks and quail.** *Avian Pathol.* 1986; **15**(4): 647–62.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Xu X, Subbarao K, Cox NJ, *et al.*: **Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong.** *Virology.* 1999; **261**(1): 15–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Lee YJ, Kang HM, Lee EK, *et al.*: **Novel reassortant influenza A(H5N8) viruses, South Korea, 2014.** *Emerg Infect Dis.* 2014; **20**(6): 1087–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Wu H, Peng X, Xu L, *et al.*: **Novel reassortant influenza A(H5N8) viruses in domestic ducks, eastern China.** *Emerg Infect Dis.* 2014; **20**(8): 1315–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Jeong J, Kang HM, Lee EK, *et al.*: **Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its relationship with migratory birds in South Korea during 2014.** *Vet Microbiol.* 2014; **173**(3–4): 249–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Jhung MA, Nelson DI; Centers for Disease Control and Prevention (CDC): **Outbreaks of avian influenza A (H5N2), (H5N8), and (H5N1) among birds--United States, December 2014-January 2015.** *MMWR Morb Mortal Wkly Rep.* 2015; **64**(4): 111.
[PubMed Abstract](#)
7. Dalby AR, Iqbal M: **The European and Japanese outbreaks of H5N8 derive from a single source population providing evidence for the dispersal along the long distance bird migratory flyways.** *PeerJ.* 2015; **3**: e934.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Brown JD, Swayne DE, Cooper RJ, *et al.*: **Persistence of H5 and H7 avian influenza viruses in water.** *Avian Dis.* 2007; **51**(1 Suppl): 285–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Bao Y, Bolotov P, Demovoy D, *et al.*: **The influenza virus resource at the National Center for Biotechnology Information.** *J Virol.* 2008; **82**(2): 596–601.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Tamura K, Stecher G, Peterson D, *et al.*: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.** *Mol Biol Evol.* 2013; **30**(12): 2725–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Squires RB, Noronha J, Hunt V, *et al.*: **Influenza research database: an integrated bioinformatics resource for influenza research and surveillance.** *Influenza Other Respir Viruses.* 2012; **6**(6): 404–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Smith GJ, Donis RO; World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization (WHO/OIE/FAO) H5 Evolution Working Group: **Nomenclature updates resulting from the evolution of avian influenza A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013–2014.** *Influenza Other Respir Viruses.* 2015; **9**(5): 271–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Matsen FA, Kodner RB, Armbrust EV: **pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree.** *BMC Bioinformatics.* 2010; **11**(1): 538.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Price MN, Dehal PS, Arkin AP: **FastTree 2--approximately maximum-likelihood trees for large alignments.** *PLoS One.* 2010; **5**(3): e9490.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Shimodaira H, Hasegawa M: **Multiple comparisons of log-likelihoods with applications to phylogenetic inference.** *Mol Biol Evol.* 1999; **16**(8): 1114–6.
[Reference Source](#)
16. Rambaut A: **FigTree, a graphical viewer of phylogenetic trees.** 2007.
[Reference Source](#)
17. Lee DH, Torchetti MK, Winker K, *et al.*: **Intercontinental Spread of Asian-Origin H5N8 to North America through Beringia by Migratory Birds.** *J Virol.* 2015; **89**(12): 6521–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Adlhoch C, Gossner C, Koch G, *et al.*: **Comparing introduction to Europe of highly pathogenic avian influenza viruses A(H5N8) in 2014 and A(H5N1) in 2005.** *Euro Surveill.* 2014; **19**(50): 20996.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Ku KB, Park EH, Yum J, *et al.*: **Highly pathogenic avian influenza A(H5N8) virus from waterfowl, South Korea, 2014.** *Emerg Infect Dis.* 2014; **20**(9): 1587–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Verhagen JH, Herfst S, Fouchier RA: **Infectious disease. How a virus travels the world.** *Science.* 2015; **347**(6222): 616–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Yoon H, Moon OK, Jeong W, *et al.*: **H5N8 Highly Pathogenic Avian Influenza in the Republic of Korea: Epidemiology During the First wave, from January Through July 2014.** *Osong Public Health Res Perspect.* 2015; **6**(2): 106–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Ip HS, Dusek RJ, Bodenstein B, *et al.*: **High Rates of Detection of Clade 2.3.4.4 Highly Pathogenic Avian Influenza H5 Viruses in Wild Birds in the Pacific Northwest During the Winter of 2014–15.** *Avian Dis.* 2016; **60**(1 Suppl): 354–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Pasick J, Berhane Y, Joseph T, *et al.*: **Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014.** *Sci Rep.* 2015; **5**: 9484.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Dalby A: **Phylogenetic Analysis of the Influenza H5 hemagglutinins.** *Zenodo.* 2015.
[Data Source](#)
25. Dalby A: **Phylogenetic analysis of the influenza N8 neuraminidases.** *Zenodo.* 2015.
[Data Source](#)

Open Peer Review

Current Referee Status: ? X

Version 1

Referee Report 25 January 2017

doi:10.5256/f1000research.9969.r18901



Justin Bahl, Joseph T. Hicks

The University of Texas School of Public Health, Houston, TX, USA

The author describes the history of the H5N8 avian influenza subtype by performing phylogenetic analysis of all H5 hemagglutinin and N8 neuraminidase gene sequences regardless of subtype. The objective of this analysis is to demonstrate reassortment events between the H5 and N8 gene segments to explain gaps in collection history of the H5N8 subtype. However, the quality of this analysis is doubtful for several reasons: foundational research on this subject appears to be ignored, assumptions are unclear and/or incorrect, the methods are subjective, strong assertions and conclusions are made with weak evidence, and the presented figures are often uninterpretable. In comparing this study to the published literature, the methods used are qualitative and highly sensitive to bias. I recommend the author uses other, more quantitative methods rather than *reading* the tree.

Influenza A virus has been an important model organism to study the mechanisms of evolutionary change. In addition to mutation, migration, natural selection and genetic drift, genomic reassortment of influenza gene segments is a well-known, well-established mechanism of influenza evolution and diversity¹⁻⁸. Scientists have been reporting reassortment events since the 1970's, including several that have resulted in pandemic disease. While the author no doubt knows this, results and conclusions are posed in such a way as to lead the reader to believe reassortment is a novel discovery. Reframing these findings in the context of previous studies on reassortment would be a helpful means to avoid this. Alternatively, the author may wish to approach the subject with a question that delves deeper into the phenomenon of reassortment itself or further investigate the methods of its analysis, rather than stopping at the conclusion that reassortment exists. The author has also overlooked some important systematic studies demonstrating that persistence of LPAI subtypes through time is rare (if ever). The author asks whether the H5 and N8 associations are randomly distributed across the tree or do they cluster assuming adequate sampling (a dubious assertion). With the understanding that LPAI subtypes rarely persist in wild birds, the signal of persistence in domestic birds might be an artifact of systematic surveillance vs. outbreak investigation of an HPAI subtype.

The methods in this analysis are also unclear and subjective. The author hopes to estimate whether or not H5 and N8 subtype genes associate persistently through time and space, immediately ruling out sampling bias and "cryptic expression" as causes for these associations. But what is meant by cryptic expression? And how was it assessed? Maybe branch lengths? To evaluate reassortment of the H5N8 subtype, the author compares phylogenetic analyses of the H5 and N8 genes. This is a valid method to investigate reassortment, especially if the goal of the manuscript is to report circulating genotypes. But, the author is making inferences or suggestions when the strength of those assertions has not been assessed. For example, the author, referring to the early H5N8, states, "Given that its sequence neighbours both before

1983 and afterward are from the H5N2 subtype it is plausible that the H5N8 hemagglutinin originated in the H5N2 subtype." Two immediate issues with this statement become clear: 1) there is no scale bar provided to allow the readers to assess how much evolution occurred on the branch leading to the H5N8 and 2) the ancestral state being H5N8 is equally likely. The author also makes inferences about the Gochang and Buan groups, but again, no scale bar is provided, the relationship of these two groups is not depicted, and the NA analysis shows them to be monophyletic (ie. single introduction into HPAI H5). It's unclear how robust this assessment of multiple reassortment histories is.

Bootstrapping, as presented here, only provides support for the estimated bifurcation of a particular node, but provides no information on the strength of evidence for a reassortment event. For instance, a single sequence of H5N8 within a clade of sequences from a different subtype may indicate a reassortment event (assuming systematic surveillance, as the author does). However, the evidence for reassortment becomes less clear as in the case of the author's claim that the Gochang and Buan sequences developed from a local reassortment. The author provides no evidence to test this claim except for the heterogeneity/homogeneity of the clade based on bootstrap values, which can be influenced by phylogenetic error. The author should consider using BATS to assess the probability of association given the phylogenies generated. Furthermore, the analysis provides no information on the extent of the circulation, context of surveillance, or number of infected or durations of circulation. All of these can affect branch lengths that might lead to incorrect inference of ancestral subtype. Additionally, the author might want to look at all H5N8 genomes to assess the relative diversity resulting from reassortment using for example tree-to-tree comparison methods. Within the past several years a variety of computational methods have been developed that can help test these claims of reassortment, both dependent^{5, 9, 10} on and independent¹¹⁻¹³ of phylogeny. These listed methods would provide statistical support for reassortment events that would go beyond the subjective claims made by this author.

One last recommended improvement for this article would be to focus on making the figures more readable and easier to interpret. Bootstrap values are often difficult to read as they are overlapped by branches and nodes. Coloring nodes by bootstrap value is probably not necessary since the values are listed, but if it is felt that the color coding must be included, the nodes should be enlarged, otherwise it is difficult to tell colors apart. In several cases these figures are uninterpretable. It also seems slightly misleading to keep the Gochang and Buan clades together for N8 phylogeny, but separate them into two figures when analyzing H5, especially without providing a tree to put these clades in context. There are also better ways to label taxa names than highlighting them in FigTree and taking a screen capture.

While the reassortment history of the H5N8 subtype is an interesting subject, the analysis presented qualifies as an initial step in understanding the global dynamics of H5N8 reassortment and re-emergence. Improved analytical methods that provide quantitative support would greatly strengthen this paper and provide an empirical framework to test the robustness of the inferences made. Finally, those inferences must be presented in context of previously published works.

References

1. Dugan VG, Chen R, Spiro DJ, Sengamalay N, Zaborsky J, Ghedin E, Nolting J, Swayne DE, Runstadler JA, Happ GM, Senne DA, Wang R, Slemons RD, Holmes EC, Taubenberger JK: The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* 2008; **4** (5): e1000076 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Pinsent A, Fraser C, Ferguson NM, Riley S: A systematic review of reported reassortant viral lineages of influenza A. *BMC Infect Dis.* 2016; **16**: 3 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y: Evolution and ecology of influenza A viruses. *Microbiol Rev.* 1992; **56** (1): 152-79 [PubMed Abstract](#)

4. Desselberger U, Nakajima K, Alfino P, Pedersen F, Haseltine W, Hannoun C, Palese P: Biochemical evidence that. *Proceedings of the National Academy of Sciences*. 1978; **75** (7): 3341-3345 [Publisher Full Text](#)
5. Dudas G, Bedford T, Lycett S, Rambaut A: Reassortment between influenza B lineages and the emergence of a coadapted PB1-PB2-HA gene complex. *Mol Biol Evol*. 2015; **32** (1): 162-72 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Holmes EC, Ghedin E, Miller N, Taylor J, Bao Y, St George K, Grenfell BT, Salzberg SL, Fraser CM, Lipman DJ, Taubenberger JK: Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol*. 2005; **3** (9): e300 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Karasin A, Schutten M, Cooper L, Smith C, Subbarao K, Anderson G, Carman S, Olsen C: Genetic characterization of H3N2 influenza viruses isolated from pigs in North America, 1977–1999: evidence for wholly human and reassortant virus genotypes. *Virus Research*. 2000; **68** (1): 71-85 [Publisher Full Text](#)
8. Lindstrom SE, Cox NJ, Klimov A: Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. *Virology*. 2004; **328** (1): 101-19 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Nagarajan N, Kingsford C: GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res*. 2011; **39** (6): e34 [PubMed Abstract](#) | [Publisher Full Text](#)
10. Svinti V, Cotton JA, McInerney JO: New approaches for unravelling reassortment pathways. *BMC Evol Biol*. 2013; **13**: 1 [PubMed Abstract](#) | [Publisher Full Text](#)
11. Dong C, Ying L, Yuan D: Detecting transmission and reassortment events for influenza A viruses with genotype profile method. *Virology*. 2011; **8**: 395 [PubMed Abstract](#) | [Publisher Full Text](#)
12. de Silva UC, Tanaka H, Nakamura S, Goto N, Yasunaga T: A comprehensive analysis of reassortment in influenza A virus. *Biol Open*. 2012; **1** (4): 385-90 [PubMed Abstract](#) | [Publisher Full Text](#)
13. Rabadan R, Levine AJ, Krasnitz M: Non-random reassortment in human influenza A viruses. *Influenza Other Respir Viruses*. 2008; **2** (1): 9-22 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Referee Report 02 November 2016

doi:10.5256/f1000research.9969.r17349



Catherine Macken

Bioinformatics Institute, University of Auckland, Auckland, New Zealand

The author examines the emergence of avian influenza A viruses (AIAV) having the H5N8 subtype, using genomic sequences from a publicly available influenza sequence database. He points out, correctly, that isolates of H5N8 AIAV has been reported sporadically, but infrequently, since 1983, with the majority of detections occurring during the recent outbreak of H5N8 viruses in Korea. These H5N8 viruses carry a hemagglutinin (HA) segment from the so-called “Guangdong lineage”, the lineage associated with the highly pathogenic AIAV H5N1 viruses that were first detected in Asia in 1996.

In order to demonstrate the sporadic emergence of H5N8 AIAV over 30+ years of history, the author conducts a phylogenetic analysis all full-length HA from all influenza A viruses having H5Nx subtype, and the complementary analysis of all full-length neuraminidase (NA) from all influenza A viruses having HxN8

subtype.

The analytical approach is appropriate. Perhaps other phylogenetic inference methods could give stronger results, but I doubt that alternatives will change the basic conclusions. It is clear that AIAV having H5N8 subtype have been seen sporadically in the past, but not until 2010 were these found in conjunction with HA from the Guangdong lineage.

Using his extensive HA(H5) and NA(N8) phylogenies, the author attempts to infer the subtype of the donor viruses of the sporadically occurring H5N8 viruses. In some instances, the evidence is reasonably strong. However, circumstantial evidence based on collocation of subtypes with high bootstrap support in a phylogeny does not constitute proof of the donor subtypes, a point that the author recognizes. (Such proof is extremely difficult to obtain, requiring a field sample with a mixture of the relevant donors and reassortant viruses.)

I would like the author to have provided some context for his results. How “unusual” is this pattern of sporadic emergence of newly formed reassortants having a particular mixture of HA and NA subtypes? For example, the database contains four sequences from H5N4 viruses, one collected from each of 2006 and 2009, and two collected in 2010. Given the time period separating these detections, it is highly likely that at least three separate reassortment events gave rise to these four viruses. The author could also describe the various other NA subtypes that have reassorted with HA from the Guangdong lineage. The fact that the Guangdong HA's have been circulating for two decades and only recently have they reassorted with another subtype to produce a “successful” novel H5Nx subtype is of interest.

Frequent reassortment among avian influenza viruses has been well documented. For the most part, novel AIAV reassortants do not persist long before they reassort again, as is seen in this paper. An important question is why a particular combination of segments becomes “successful”, i.e. able to spread widely in its host species. The answer to this question is beyond the scope of the author's work. This paper does show, for one subtype, the history of emergence, disappearance and re-emergence, including a recent outbreak.

I have a number of minor comments.

1. I found the tip labels of phylogenetic trees confusing. I would prefer the format “strain name (subtype).”
2. Please collapse subtrees in Fig. 2.
3. Numerical support for branches: I would prefer these values to be given only when the branch is relevant to the thesis of the paper. Reporting support values that are very low or even 0, such as in Figures 5 and 8, is not helpful.
4. I would like Figs 5-7 to be combined, so that I can see how these subtrees relate to the overall evolution of the Guangdong HA/H5.
5. A number of the sequences from old isolates are duplicates. These should be removed before analysis.
6. The author criticizes phylogenetic analysis based on a specific subtype. I disagree with this criticism. I believe that the choice of dataset depends on the hypothesis of interest. It may be most appropriate to focus on A/H3N2 viruses from humans or A/H5N1 viruses from chickens when, for

example, considering antigenic drift in the respective hemagglutinin proteins.

7. Surveillance of avian influenza is a major focus of the CEIRS program. It extends beyond the study in the Delaware Bay. Subtypes not represented in the public sequence database are likely to be rare in the host population. Except in restricted situations such as outbreaks in a domestic poultry flock, it is unlikely that the subtype of an avian virus is known before sampling and laboratory identification. Therefore, sampling bias is unlikely to be a source of lack of sequences from a particular subtype.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Nov 2016

Andrew Dalby, University of Westminster

I would like to thank Dr Macken for her comments and review.

While it is not possible to know for sure the ancestors of a reassortment event collocation in a specific time-frame is strongly suggestive especially if reassortment can be placed within the same time frame.

The bootstrap is often treated as a measure of the accuracy of a phylogenetic tree when it is actually a measure of the precision of that sample of data producing that tree topology with that tree construction method. Tree reconstruction is limited by the quality of the sampling of the data, and no number of bootstraps can create the correct tree from a poor sample. I do not consider bootstraps as a measure of likely ancestry but more as a measure of clustering of sequences and their position within clades. If there is a clade with different subtypes then these are more likely to have a shared ancestry.

The H5 trees contain the sporadic occurrence of many other H5 containing subtypes and show that even H5N2 and H5N1 have multiple reassortment events. The reasons for the focus on H5N8 is because there are such large breaks in the historical record, because of the global distribution of the subtype and most importantly because there was finally a sustained outbreak in Korea and then later Taiwan. It was this different behaviour between the Korean outbreak and previous cases of H5N8 that drew my attention. This does point to the actual question of interest, which is why some reassortments thrive but many do not. Even for the Korean outbreak when the Guangdong containing sequence reached North America it rapidly reassorted to produce a H5N2 subtype. This implies that while in North America the N2 is preferred, there is no equivalent N2 replacement for N8 in Korea or Taiwan. This shows a specific preference for neuraminidase lineages for hemagglutinin lineages that goes beyond the level of subtype.

The reason for criticising studies that use just sequences from a single subtype is that if you do this then you do not get a contiguous sample of the hemagglutinin or neuraminidase sequence changes. If you consider the H5 hemagglutinin in H5N8 for example, if H5N8 becomes H5N2 and then returns to H5N8 you are missing the sequence changes that occur in the hemagglutinin in the H5N2 subtype. This will affect antigenic studies as much as phylogenetic studies (and possibly

even more severely depending on the number of changes that occur in the H5N2 hemagglutinin). This is because you will be sampling irregularly in time. For the best possible sampling you need to see all of the individual nucleotide sequence changes. Where you have steps that involve more than a single change then there are several different orders for these changes which cannot be resolved unambiguously.

By sampling bias I am considering the relatively small number and narrow geographical focus of current sampling efforts. I was not questioning the experimental validity of studies such as CEIRS and implying that the experiments are biased. My meaning was that most sampling is convenience sampling and this is inherently biased. For example Africa and South America have very few sequences available and only now is there increased sampling in Russia where the migratory breeding grounds for many long range bird migrants are found. While a subtype might be rare locally this might not be true globally. For example the Guangdong H5 lineage was not known in North America until the H5N8 outbreak and localisation of lineages seems to occur in many subtypes including H5 and H9.

Competing Interests: No competing interests.
