

Aggregation of Single Nucleotide Polymorphisms in a Human H5N1 Clade 2.2 Hemagglutinin

Niman HL^{*}, Saad MD[#], Tjaden JA[#], Earhart KC[#], Monteville MR[#], Aly MM[%], Mansour MM[#], Elsayed NM⁺, Nayel AE⁺, Abdelghani AS⁺, Esmat HM^{#+}, Labib EM[#], Ayoub EA[#], Arafa A-SA[%], Raczniak GA^{##,####}, Agyen-Frempong M^{###}, Ampofo WK^{####}, Boynton BR[#]

^{*}Recombinomics, Inc., Pittsburgh, Pennsylvania, USA, [#]U.S. Naval Medical Research Unit 3 (NAMRU-3), Cairo, EGY, ^{##}NAMRU-3 Ghana Detachment, Accra, GHA, [%]Central Laboratory for Veterinary Quality Control, Giza, EGY, ⁺Ministry of Health, Arabic Republic of Egypt, Cairo, EGY, ^{###}Ghana Veterinary Services, Accra, GHA, ^{####}Noguchi Memorial Institute for Medical Research, Accra, GHA

Department of Influenza Recombination, Recombinomics, Inc, 648 Field Club Road, Pittsburgh, PA 15238, USA

Correspondence to Henry L Niman, E-mail: henry_niman@recombinomics.com
617.877.0987, 412.963.1362 FAX

Text word count: 1391
2 Figures 1 Supplemental Table

Running Title: H5N1 Polymorphism Aggregation

The evolution of H5N1 has attracted significant interest ¹⁻⁴ due to linkages with avian ^{5,6} and human infections ^{7,8}. The basic tenets of influenza genetics ⁹ attribute genetic drift to replication errors caused by a polymerase complex that lacks a proof reading function. However, recent analysis ¹⁰ of swine influenza genes identifies regions copied with absolute fidelity for more than 25 years. In addition, polymorphism tracing of clade 2.2 H5N1 single nucleotide polymorphisms identify concurrent acquisition ¹¹ of the same polymorphism onto multiple genetic backgrounds in widely dispersed geographical locations. Here we show the aggregation of regional clade 2.2 polymorphisms from Germany, Egypt, and sub-Sahara Africa onto a human Nigerian H5N1 hemagglutinin (HA), implicating recombination in the dispersal and aggregation of single nucleotide polymorphisms from closely related genomes.

The rapid expansion of the geographical reach of H5N1 clade 2.2 has increased attention on the mechanism of evolution in these rapidly changing genomes. Clade 2.2 was first reported ^{12, 13} at Qinghai Lake in Central China in May, 2005. Infection of long range migratory bird led to a rapid expansion of reported H5N1 infections in wild birds and poultry in Europe, the Middle East, and Africa. Moreover, human H5N1 infections have been subsequently reported in Turkey, Iraq, Azerbaijan, Egypt, Djibouti, and Nigeria ¹⁴. More than 50 countries west of China have reported clade 2.2 infections following the Qinghai Lake outbreak. The expansion into these new regions was associated with the acquisition of regional polymorphisms. Tracking of these newly acquired polymorphisms

demonstrates ¹¹ a non-random appending of the polymorphisms on clade 2.2 genetic backgrounds. One NA polymorphism, G743A, which was a regional polymorphism in Germany in 2006, was appended onto six distinct clade 2.2 genetic backgrounds in Russia, Egypt, and Ghana.

The HA phylogram in Figure 1A is annotated with regional polymorphism on isolates from Egypt or neighboring countries (Israel, Gaza, and Djibouti). We have designated these isolates as clade 2.2.1. The regional markers were present in isolates from early 2006, but were also present in more recent 2006/2007 isolates from Egypt. Two of the polymorphisms, G467A and T937C were present in all of the isolates from the region. Additional polymorphisms, C661T, C727T, A880G, G1019A, were also in these regional isolates, but also extended upstream to isolates in Europe, and downstream to isolates in Nigeria. In addition to these regional markers, the figure identifies some of the sub-regional markers, which will be discussed in more detail elsewhere.

The HA phylogram in Figure 1B is annotated with regional markers on isolates from Germany and neighboring countries. The German isolates fell into three major sub-clades. Isolates from northern Germany were similar to isolates from Denmark (designated clade 2.2.2.1), and had sub-regional markers G1235A, T1510C, and T1615C. A second sub-group (clade 2.2.2.2) has the broader Egyptian markers, as well as G142A and A658B. A third sub-group (clade 2.2.2.3), has another series of markers (G41A, G295A, C689T, C1012T,

C1177T, C1402T, C1480T). Clade 2.2.2.3 has additional markers in NA, including G743A, which was appended onto four different Egyptian backgrounds in 2007¹¹. Additional polymorphisms shared between German and Egyptian isolates will be discussed in detail elsewhere.

The HA phylogram in Figure 1C is annotated with regional markers on isolates from sub-Saharan Africa. These isolates also fall into three major sub-clades. One group (clade 2.2.3.1) has the broader Egyptian markers plus G496A, C627A, G1672A. A second group (clade 2.2.3.2) has G209A and T1415C. A larger sub-Saharan group has A433G, G643A, and A1708G. The isolates from Ghana had a number of additional polymorphisms appended onto this background. The polymorphisms on the 2007 isolates from Ghana will be discussed in detail elsewhere.

The first confirmed human clade 2.2 infection in Nigeria was in February, 2007. The HA phylogram with this isolate and isolates which share polymorphisms are listed in the phylogram in figure 2. The Nigerian isolate has aggregated regional and sub-regional clade 2.2 polymorphisms from Egypt (clade 2.2.1) Germany (clade 2.2.2.3), and sub-Saharan Africa. The isolate has the three sub-Saharan polymorphisms, plus a sub-regional polymorphism, A1006G from this sub-clade. This isolate also has a sub-set of the German clade 2.2.2.3 markers (G295A and C1480T) as well as a sub-regional marker from Egyptian clade 2.2.2.3, C1614T. In addition, the isolate has T937C, which is one of the clade 2.2.1 regional

markers. The human sequence also has two clade 2.2.1 sub-regional markers (T610C and G643A). In addition, the sequence has additional sub-regional markers from Mongolia and Siberia. Thus, 13 of the 14 newly acquired polymorphisms are present in sequences from clade 2.2 isolates, and six of the polymorphisms are regional markers.

Earlier reports have used phylogenetic analysis to conclude that H5N1 infections in Nigeria involved multiple introductions¹⁵⁻¹⁸. Similar observations have been made for Germany¹⁹⁻²¹. The similarities between isolates from Egypt and Israel and Gaza have also been noted²². The data presented here support those conclusions, but the tracing of polymorphisms identifies exchanges of polymorphisms between the sub-clades identified by this analysis. These exchanges are not easily explained by random mutations due to copy errors.

Currently, genetic drift in influenza is explained by random mutations that became dominant because of selection / adaptation pressures. However, analysis of recent swine influenza sequences identified large regions of nucleotide identity with sequences for isolates collected over 25 years earlier. These isolate also have clear examples of homologous recombination, based on matches with multiple parental sequences¹⁰.

Recombination also offers an explanation for the aggregation data, which are not easily explained by random mutations. The number of regional markers is small,

and the aggregation of subsets of polymorphisms into a single sequence does not appear to be random. The aggregation data compliments the polymorphism dispersal data ¹¹ noted for NA G743A. The polymorphisms are appended onto regional genetic backgrounds, but the new acquisitions have linkages to earlier sequences that lie along migration pathways.

The acquisitions are most easily explained by recombination between closely related sequences. A series of closely related sequences has been found in the H5N1 reported in countries west of China. All of the reported cases have been clade 2.2, and isolates are closely related to each other. However, the sequences, as seen in the Egyptian isolates, are becoming more genetically complex. As the sequence database grows, the ability to find matching polymorphisms increases.

Theoretically, homologous recombination between closely related sequences would be more common because of extensive regions of sequence identity. Moreover, such recombination would result in acquisitions of single nucleotide polymorphisms because most of the acquired sequences would be identical in both parental and progeny sequences. Examples of closely related sequences in human H5N1 have been reported previously in isolates with different receptor binding domains ²³, as well as susceptibility to anti-viral treatment with oseltamivir ²⁴. Plaques purified clones from the same patient had different combinations of receptor binding domain polymorphisms. Similarly, plaque purified clones with

two different oseltamivir resistance changes, H274Y and N294S, were isolated, in addition to clones with wild type polymorphisms. However, both of these changes have been detected in poultry or wild bird sequences that have not been linked to oseltamivir sequences, raising the possibility that the resistance markers that emerged were already present in low abundance prior to treatment.

One of the markers, N294S, was also in clade 2.2 sequences present in family members from the Egyptian governorate of Gharbiya. This change was present in isolates collected prior to treatment, as well as isolates from samples collected two days following the start of treatment. Sequences from two of the patients were similar but distinct, and these distinctions were present in both sets of sequences, suggesting that the source of infection for the patients also contained distinct but closely related sequences.

Similarly, plaque purified clones of H5N1 from a chicken in Gharbiya identified two distinct populations. One was closely related to the sequences from the patients in Gharbiya, while the other was identical to sequences from other chickens in Gharbiya. Moreover, the clones had evidence of recombination (data will be reported elsewhere). These examples of dual infections involving closely related sequences provide the genetic basis for acquisition of single nucleotide polymorphisms by recombination.

Tracing of these polymorphisms defines distribution routes. The distribution of additional polymorphisms described above will be detailed elsewhere. The mapping of these polymorphisms has the potential of defining predictable acquisitions. The predicted acquisitions can then be used to create vaccine targets representing emerging genomes.

Figure legends

Figure 1 HA Phylograms of Egyptian, German, and Sub-Sahara Isolates

A. Phylogram of Egyptian and regional isolates with clade 2.2.1 Egyptian regional markers G467A and T937C. Extended markers include C661T, C727T, A880G, G1019T.

B. Phylogram of German and regional isolates. Clade 2.2.2.1 German regional markers G1235A and T1510C, and extended marker T1615C. Clade 2.2.2.2 with German regional markers G142A and A658G. Extended markers include C661T, C727T, A880G, G1019T. Clade 2.2.2.3 German regional markers G41A, G295A, C689T, C1012T, C1177T, C1402T, C1480T.

C. Phylogram of sub-Sahara isolates. Sub-Sahara markers are A433G, G643A, A1708G. Clade 2.2.3.1 has Nigeria regional markers G496A, C627A, G1672A and extended markers C661T, C727T, A880G, G1019T. Clade 2.2.3.2 has Nigeria regional markers G209A and T1415C.

Phylograms represent positions 93-1688. Isolates and accession numbers are in Table S1. Trees were generated using neighbor joining with 100 bootstrap repetitions. Sequences generated as described previously ²⁵

Figure 2 Aggregated Polymorphisms on Human Nigerian Hemagglutinin Phylogram as described in Figure 1. 13 of the 14 polymorphisms are found in other clade 2.2 including Egyptian regional polymorphism T937C, German clade 2.2.2.3 regional polymorphisms G295A and C1480T, and sub-Sahara regional polymorphisms A433G, G643A, G1708A.

Acknowledgements

We thank Evelyn Yayra Bonney, Ken-Edwin Aryee, Kofi Bonney, and Professor Alexander Nyarko from Naguchi Memorial Institute for Medical Research, and Drs Joseph GAARI-KWEKU, Joseph ADONGO AWUNI, John NIENDOW KARIMU, Bashiru KIKIMOTO and Jerry ODOI from the Ghana Veterinary Services for the collection of avian samples.

References

1. Peiris, J.S. *et al.* Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* **363**, 617-619 (2004).

2. Fouchier,R., Kuiken, T., Rimmelzwaan, G., & Osterhaus, A. Global task force for influenza. *Nature* **435**, 419-420 (2005).
3. Li, K.S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*. **430**, 209-213 (2004).
4. Guan, Y. *et al.* H5N1 influenza: A protean pandemic threat. *Proc. Natl. Acad. Sci. U.S.A.***101**, 8156 – 8161 (2004).
5. Osterholm, M.T. Preparing for the next pandemic. *N. Engl. J. Med.***352**, 1839 - 1842 (2005).
6. Monto, A.S. The Threat of an Avian Influenza Pandemic. *N. Engl. J. Med.* **352**, 323 - 325 (2005).
7. Subbarao, K. *et al.* Characterization of an Avian Influenza A (H5N1) Virus Isolated from a Child with a Fatal Respiratory Illness *Science* **279**, 393-396 (1998).
8. Claas, E.C. *et al.* Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**, 472-477 (1998).

9. Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M. & Kawaoka, Y. Evolution and ecology of influenza A viruses *Microbiol. Rev.***56**, 152 - 179 (1992).
10. Niman, H.L. Swine influenza A evolution via recombination – genetic drift reservoir. Available from Nature Precedings <http://precedings.nature.com/documents/385/version/1> (2007).
11. Niman, H.L. *et al.* Concurrent acquisition of a single nucleotide polymorphism in diverse influenza H5N1 clade 2.2 sub-clades. Available from Nature Precedings <http://precedings.nature.com/documents/459/version/1> (2007).
12. Chen, H. *et al.* Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* **436**, 191-192 (2005).
13. Liu, J. *et al.* Highly Pathogenic H5N1 Influenza Virus Infection in Migratory Birds *Science*, **309**, 1266-1267 (2005).
14. Cumulative number of confirmed cases of human influenza A/(H5N1) reported to WHO. Available from http://www.who.int/csr/disease/avian_influenza/country/cases_table_2007_05_16/en/index.html

15. Ducatez, M.F. *et al.* Avian flu: multiple introductions of H5N1 in Nigeria. *Nature* **442**, 37 (2006).
16. Ducatez, M.F. *et al.* Molecular and antigenic evolution and geographical spread of H5N1 highly pathogenic avian influenza viruses in western Africa. *J. Gen. Virol.* **88**, 2297-2306 (2007).
17. Salzberg, S.L. *et al.* Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerg. Infect. Dis.* **13**, May (2007). Available from <http://www.cdc.gov/EID/content/13/5/713.htm>
18. Ducatez, M.F. *et al.* Genetic characterization of HPAI (H5N1) viruses from poultry and wild vultures, Burkina Faso. *Emerg. Infect. Dis.* **13**, 611-613 (2007).
19. Rinder, H. *et al.* Genetic evidence for multi-event imports of avian influenza virus A (H5N1) into Bavaria, Germany. *J. Vet. Diagn. Invest* **19**, 279-282 (2006).
20. Weber, S. *et al.* Molecular analysis of highly pathogenic avian influenza virus of subtype H5N1 isolated from wild birds and mammals in northern Germany. *J. Gen Virol* **88**, 554-558 (2007).

21. Bragstad, K. *et al.* First introduction of highly pathogenic H5N1 avian influenza A viruses in wild and domestic birds in Denmark, Northern Europe. *Virology* **4**, 43 (2007).
22. Perk, s. *et al.* Molecular characterization of the glycoprotein genes of H5N1 influenza A viruses isolates in Israel and the Gaza Strip during 2006 outbreaks. *Virus Genes* July 7 [Epub ahead of print (2007)].
23. Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378-382 (2006).
24. Le, Q.M. *et al.* Avian flu: isolation of drug-resistant H5N1 virus, *Nature* **437**, 1108 (2005).
25. Yingst, S.L., Saad, M.D., & Felt, S.A. Qinghai-like H5N1 from domestic cats, Northern Iraq. *Emerg. Infect. Dis.* **12**, 1295-1297 (2006).

Figure 1A. Phylogram of Egyptian and regional isolates

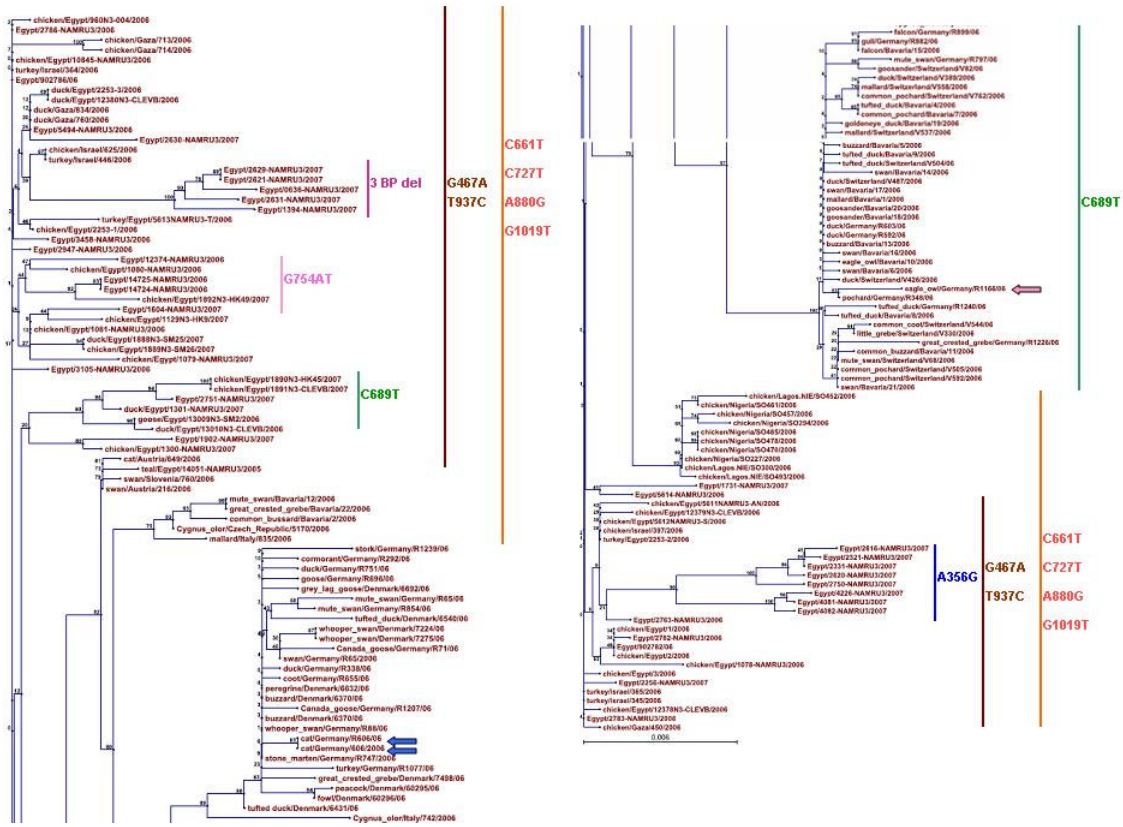


Figure 1B. Phylogram of German and regional isolates

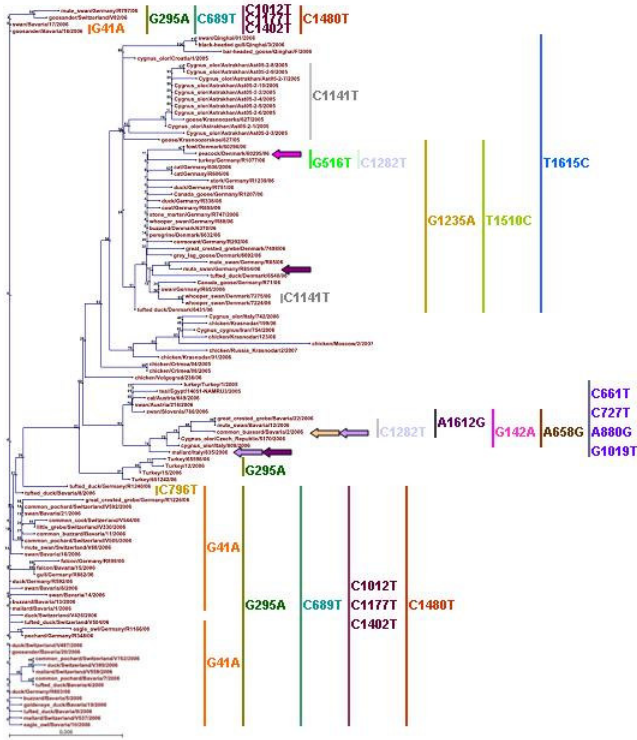


Figure 1C. Phylogram of sub-Saharan isolates

