# The Peak of a Pandemic? — A Phylogenetic Analysis of the H1N1 Influenza Virus from 2009 to Present

**Anthony Deeter[1,*], Mark Dalman[2,3,*], Gayathri Nimishakavi[1], and Zhong-Hui Duan[1,3,†]**
[1]Department of Computer Science, University of Akron, Akron, OH, USA 44325
[2]Department of Biology, University of Akron, Akron, OH, USA 44325
[3]Integrated Bioscience Program, University of Akron, Akron, OH, USA 44325

**Abstract** – *The swine origin influenza A (S-OIV) virus of 2009 reached pandemic proportions due to its novel sequence identity in human populations of North America and other localities. The S-OIV virus shows subtle change from 2009-2010 in humans, affirming that the HA and NA sequences have been unable to antigenically drift or shift enough to emerge as another pandemic. This study aimed to document the succession of S-OIV from 2009 to current in addition to investigating its relationship among other locations. Based on the phylogenetic analysis, the 2010 H1N1 is similar to other isolates circulating the previous year. Furthermore, the protein sequences with the highest non-synonymous to synonymous ratio were HA and NA thus indicating strong selective pressures for the antigen receptor binding sites to adapt even within human hosts.*

**Keywords**: 2009 influenza pandemic H1N1 influenza virus; antigenetic shift and drift; phylogenetic analysis; non-synonymous to synonymous ratio; neighbor joining method

## 1 Introduction

With the development of antibiotics within the past century, life expectancy has increased despite a prolonged window of susceptibility and transmissibility to viral and bacterial infections during humans' life span. Currently, with over a quarter of a million deaths and upwards of three million cases of influenza globally each year in humans, the emergence of a novel swine-origin influenza virus (S-OIV) in 2009 garnered much attention [1]. The common influenza (flu) is caused by the Orthomyxoviridae family of ssRNA viruses including Influenzavirus A, Influenzavirus B, Influenzavirus C, Isavirus and Thogotovirus. All but Isavirus are detected in vertebrates with Influenzavirus A seemingly the most virulent, diverse, and pathogenetic to humans. The present paper deals with the pandemic H1N1 flu virus, a novel subtype of Influenzavirus A.

Three months after its identification in Mexico in 2009, the S-OIV epidemic had reached alert phase 6, marking the first pandemic in almost forty years to reach that phase [1, 2]. The S-OIV virus, despite its novel sequence and severity, is a triple reassortment from three different "donors" [3, 4]. Phylogenetic analyses conclude that of its eight nucleotide sequences, six of them (HA, PB2, PB1, PA, NP, NS) are highly similar to influenza viruses endemic to pigs in the late 1990's with the other two genes (NA and MP) from a bird lineage isolated in Europe [5-7]. None of the individual genes were previously found in Europe or North America, reaffirming conditions for a pandemic viral outbreak [3, 4, 8, and 9].

Therefore the early detection and continuous monitoring of novel strains in the environment are poised at the interface of molecular biology, viral biology, and, more recently, computer science. Furthermore, the inherent diversity, total number of sequences of Influenzavirus A and the lack of sampling resolution make phylogenetic analysis very complex. The present paper focuses on the antigenic shifting and drifting of the virus from 2009 to present and the post pandemic evolution of the 2009 H1N1 (S-OIV) Influenza virus.

## 2 Materials and Methods

In order to study the evolution of S-OIV since its emergence, phylogenetic trees were constructed using only unique, full length coding sequences, human host H1N1 nucleotide sequences from April 2009 to January 2011 [10]. The trees were constructed via the neighbor-joining method, with distances calculated using the Felsenstein F84 nucleotide method [11]. Non-synonymous (dN) to synonymous (dS) substitution ratio were then calculated using the Nei-Gojobori method [12]. The A/California/04/2009(H1N1) isolate was used as the first identified S-OIV strain (highlighted in green in figures) and 2010 strains were highlighted in red. A/Moscow/ 01/2009(H1N1), A/Boston/ 653/2009(H1N1), A/Korea/ CJ01/2009(H1N1), A/Chile/ 4064/2009(H1N1), A/MexicoCity/ WRAIR1752N/2010(H1N1), A/Newark/ INS429/2010(H1N1), A/Vienna/ INS179/2010(H1N1), and A/Ontario/ 3620/ 2010 were randomly selected in calculating dN/dS ratios. In constructing the phylogenetic trees, only unique isolates were used and a Perl script was written to

[*]both authors contributed equally; [†]duan@uakron.edu

randomly select a sample of 1000 sequences. The A/United Kingdom/1/1933(H1N1) is used as the outgroup. The A/California/04/2009(H1N1) and 2010 sequences were highlighted in each tree for reference and closely related branches were collapsed for tree readability.

# 3 Results and Discussion

Phylogenetic trees were constructed using neighbor-joining method to understand how the novel S-OIV (H1N1) influenzavirus A strain has changed through its pandemic period (April 2009 - January 2011) in human hosts. The phylogenies for the unique protein coding sequences HA, NA, M1, M2 and PB2 are shown in Figures 1-5 (phylogenies for other coding sequences are not shown here).These trees show the genotypic variation of encoded proteins in H1N1 influenzavirus A. Our results indicate that the A/California/04/2009(H1N1) strain is genetically similar to 2010 isolated strains and is always present within every tree. Furthermore A/United Kingdom/1/1933 (H1N1) is the outgroup for all trees. Interestingly, 2010 isolates are genetically similar to the previous 2009 isolates once again reiterating that most epidemic H1N1 stem from circulating viral reservoirs. Surprisingly, a large polytomy occurred within the 2009 pandemic and 2010 isolates are much more diverse from one another, yet still similar to ones from 2009 than any other sequences.


Figure 1. Phylogenetic relationships among human H1N1 viruses (HA)

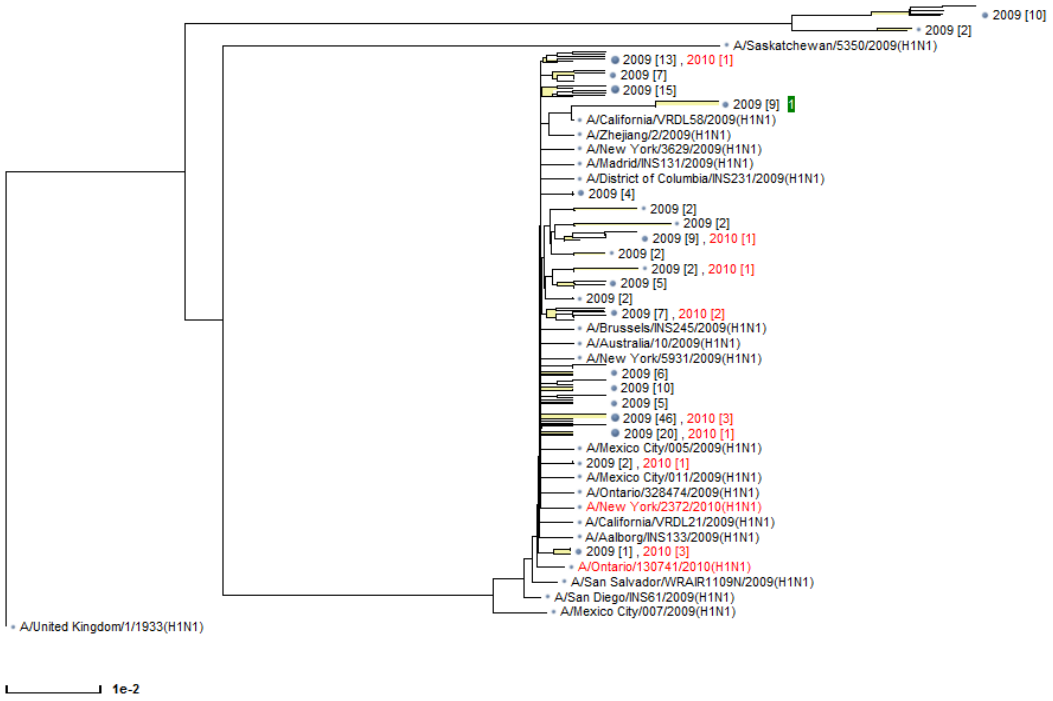


Figure 2. Phylogenetic relationship among human H1N1 viruses (M2)

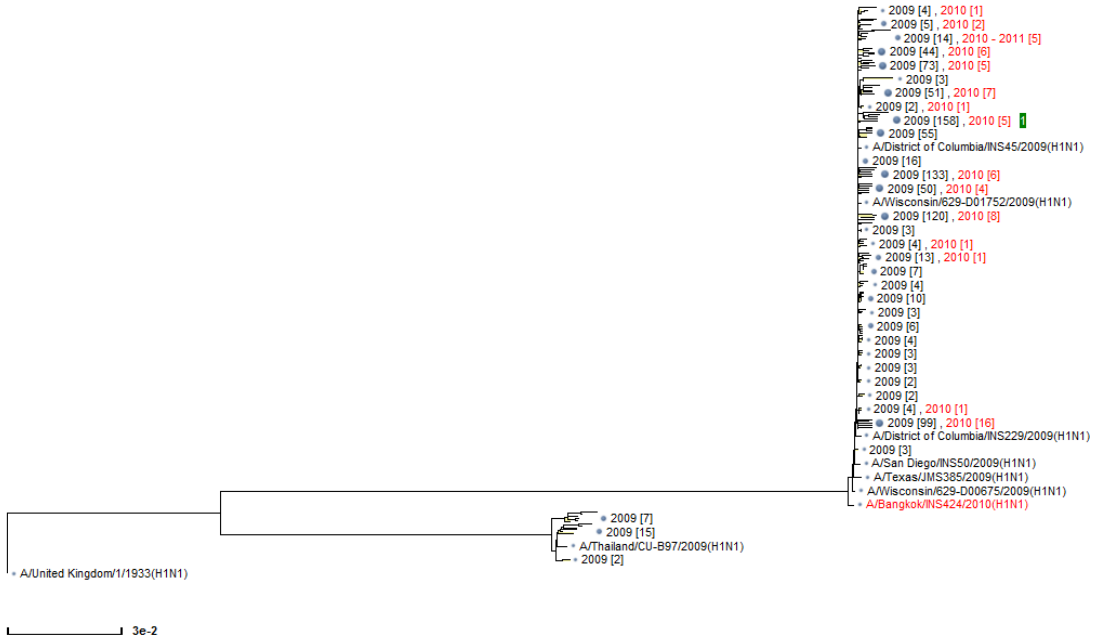


Figure 3. Phylogenetic relationship among human H1N1 viruses (NA)

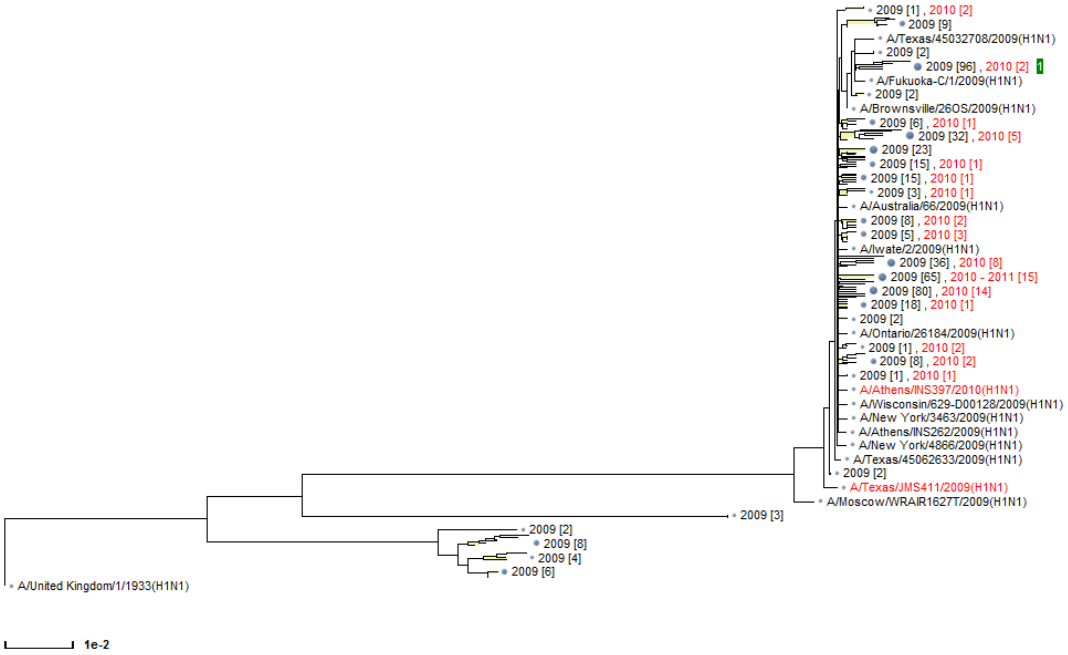


Figure 4. Phylogenetic relationship among human H1N1 viruses (M1)

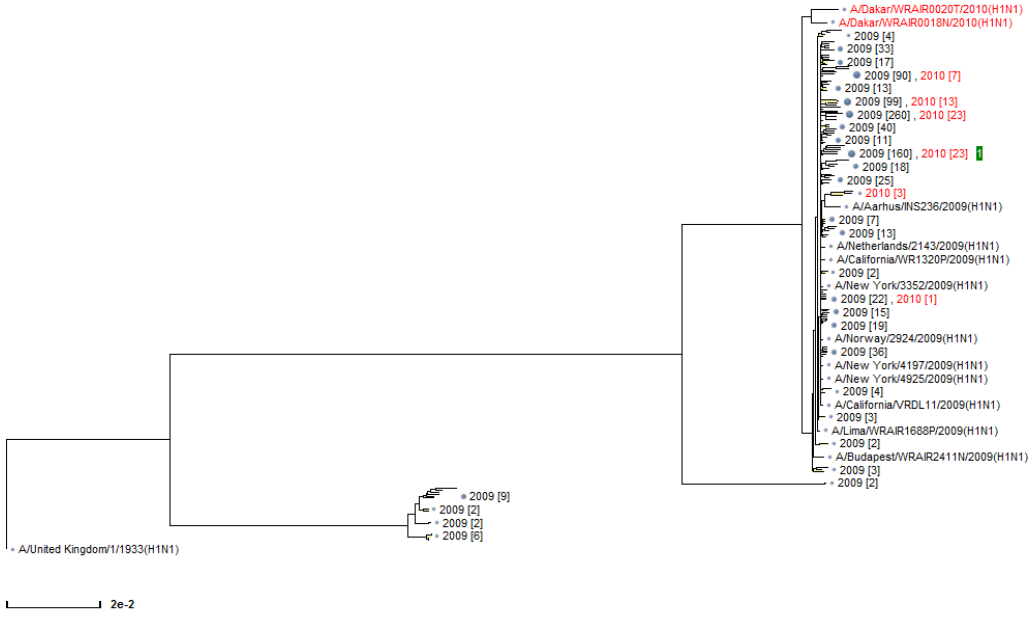


Figure 5. Phylogenetic relationship among human H1H1 viruses (PB2)

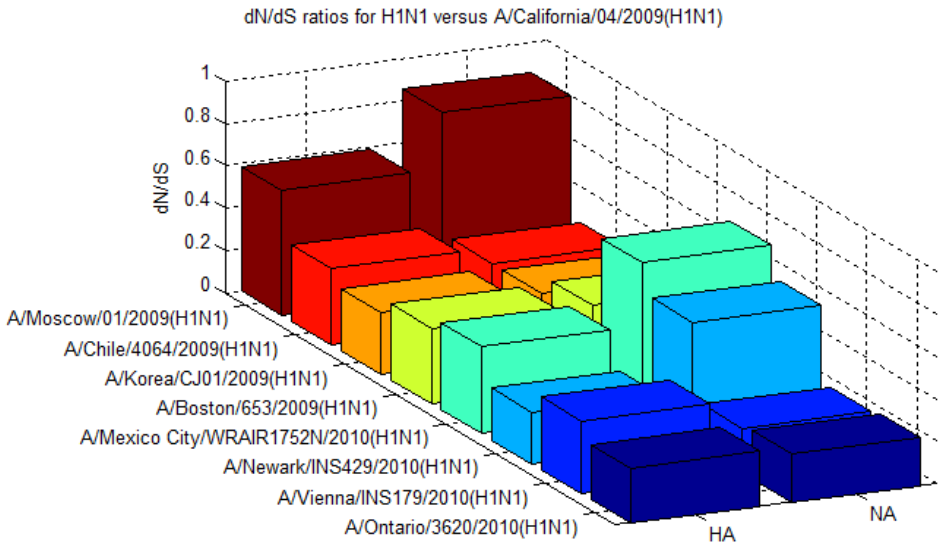


Figure 6. Non-synonymous to synonymous (dN/dS) ratios for selected human H1N1 isolates from 2009 and 2010.

To quantitatively identify the changes in primary sequence from 2009 to 2010, non-synonymous to synonymous (dN/dS) substitution ratio were calculated to identify whether changes in nucleotide sequence actually resulted in changes in primary sequence. Unusually high number of non-synonymous substitutions is widely accepted as a result of positive selection [13]. Within our case study, the non-synonymous substitutions between A/California/04/2009 (H1N1) strain and the 2010 sequences of HA and NA are considerably smaller than their respective synonymous ones. Ratios of less than one across all 2010 isolates suggest concrete evidence as to why no genetically dissimilar S-OIV isolates have arisen (Figure 6). Furthermore, in comparing randomly selected 2009 and 2010 sequences to A/California/04/2009(H1N1), there was less variation between A/California/04/2009(H1N1) and more recent 2010 sequences than amongst 2009 sequences (0.424 vs 0.371, respectively). This suggests there was more genetic variation among the initial outbreak than subsequently documented in 2010 and even more so, this signifies that the strains identified are likely small antigenic drifts from other viruses circulating in 2009.

Genetically dissimilar and novel isolates to a population are the cruxes of a pandemic. Additionally, the dN/dS ratios of HA and NA are considerably larger than those of other six proteins (not shown in paper); suggesting that as these ratios increase, the prevalence of new coded amino acids will increase. A new amino acid can be the difference between viral detection and infection [14-16]. However, the direction of selection is not well articulated within dN/dS ratios and begs the question of whether neutral theory is the evolutionary process underlying epidemic viral outbreaks and the "perfect storm" reassortments in pigs and birds causing pandemic outbreaks [17].

By crossing the positive selection threshold, the possibility of novel strains of influenza virus A increase, requiring new vaccines. In other words, as the dN/dS ratio goes above one, the novelty of its structure begins to be advantageous to the virus's transmission, ultimately increasing its fitness. Despite research indicating that selective pressures will increase non-synonymous substitutions, the lack of biochemical and evolutionary data is not in accordance. For all proteins, there are both essential and nonessential amino acids, those which are responsible for function and those that are not. HA and NA are two proteins on the viral coat which, by being genetically different through selective pressures from innate and adaptive immunity, can cause a pandemic. Yet proteins within the virion that are not as plastic show little variation from one host or year to the next (Figure 2). The later example is similar to most proteins in the human body in which there are areas capable of nonsynonomous substition and areas that have conserved sequences. Consequently, in viral biology producing a genetically different coat is advantageous as oppossed to maintaining the status quo and being erradicated. An inability to infect (highly detected) or novelize (highly virulent) may both result in the erradication of the strains from the gene pool.

# 4 Conclusions

The swine origin influenzavirus A S-OIV pandemic of 2009 has a unique genetic composition as suggested by

almost a century of viral data. Our study reveals that despite being phylogenetically similar to 2010 influenza viruses in human, the dN/dS ratios indicate that the surface proteins HA and NA do antigenetically drift fastest amongst human hosts. Furthermore, the dN/dS ratios suggest that sequences during 2009 are significantly more dissimilar than recent 2010 isolates, suggesting that the 2009 S-OIV pandemic might have peaked during the summer of 2009.

Future studies should comparatively measure the substitution rates amongst host types and by locations to further elucidate whether avian and swine lineages are the most capable and dominating viral incubators or whether attention should be focused at a more macroscopic regional or continental understanding of viral transmission. Thus further research of immunoinformatics will increase the interdisciplinary understanding of viral transmission, vaccination, documentation, and retrieval.

## 5 References


[1] WHO Pandemic (H1N1) 2009. http://www.who.int/csr/disease/swineflu/en/. Retrieved on 2/12/2011.

[2] Christophe Fraser, Christl A. Donnelly, Simon Cauchemez, William P. Hanage, Maria D. Van Kerkhove, T. Déirdre Hollingsworth, Jamie Griffin, Rebecca F. Baggaley, Helen E. Jenkins, Emily J. Lyons, Thibaut Jombart, Wes R. Hinsley, Nicholas C. Grassly, Francois Balloux, Azra C. Ghani, Neil M. Ferguson, Andrew Rambaut, Oliver G. Pybus, Hugo Lopez-Gatell, Celia M. Alpuche-Aranda, Ietza Bojorquez Chapela, Ethel Palacios Zavala, Dulce Ma. Espejo Guevara, Francesco Checchi, Erika Garcia, Stephane Hugonnet, Cathy Roth, and The WHO Rapid Pandemic Assessment Collaboration. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings; Science, 324:1557-1561, June 2009.

[3] F. S. Dawood, S. Jain, L. Finelli, M. W. Shaw, S. Lindstrom, R. J. Garten, L. V. Gubareva, X. Xu, C. B. Bridges, T. M. Uyeki, and Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. Emergence of a Novel Swine-origin Influenza A (H1N1) Virus in Humans; New England Journal of Medicine, 360:2605-2615, June 2009.

[4] Gavin J. D. Smith, Dhanasekaran Vijaykrishna, Justin Bahl, Samantha J. Lycett, Michael Worobey, Oliver G. Pybus, Siu Kit Ma, Chung Lam Cheung, Jayna Raghwani, Samir Bhatt, J. S. Malik Peiris, Yi Guan, and Andrew Rambaut. Origins and Evolutionary Genomics of the 2009 Swine-origin H1N1 Influenza A Epidemic; Nature, 459:1122-1125, June 2009.

[5] Mikhail Matrosovich, Alexander Tuzikov, Nikolai Bovin, Alexandra Gambaryan, Alexander Klimov, Maria R. Castrucci, Isabella Donatelli, and Yoshihiro Kawaoka. Early Alterations of the Receptor-Binding Properties of H1, H2, and H3 Avian Influenza Virus Hemagglutinins after Their Introduction into Mammals; Journal of Virology, 74: 8502-8512, September 2000.

[6] Gabriele Neumann, Takeshi Noda, and Yoshihiro Kawaoka. Emergence and Pandemic Potential of Swine-Origin H1N1 Influenza Virus; Nature, 459(7249): 931-939, June 2009.

[7] Adrian J Gibbs, John S Armstrong, and Jean C Downie. From Where Did The 2009 'Swine-Origin' Influenza A Virus (H1N1) Emerge? Virology Journal, 6(207): 1-11, November 2009.

[8] Alan J. Hay, Victoria Gregory, Alan R. Douglas and Yi Pu Lin. The Evolution of Human Influenza Viruses; Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 356:1861-1870, December 2001.

[9] Troy Day, Jean-Baptiste André, and Andrew Park. The Evolutionary Emergence of Pandemic Influenza; Proceedings: Biological Sciences, 273(1604): 2945-2953, December 2006.

[10] NCBI Influenza Virus Sequence Database. http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/select. cgi?go=1. Data retrieved: March, 15, 2011.

[11] Joseph Felsenstein and Gary A. Churchill. A Hidden Markov Model Approach to Variation among Sites In Rate of Evolution. Molecular Biology and Evolution, 13:93–104, January 1996.

[12] Masatoshi Nei and Takashi Gojoborit. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions; Molecular Biology and Evolution, 3:418-426, September 1986.

[13] Sergei L. Kosakovsky Pond, Art F.Y. Poon, Andrew J. Leigh Brown, and Simon D.W. Frost. A Maximum Likelihood Method for Detecting Directional Evolution In Protein Sequences And Its Application To Influenza A Virus; Molecular Biology Evolution, 25(9):1809-1824 September 2008.

[14] Christoph Scholtissek. Source for Influenza Pandemics; European Journal of Epidemiology, 10(4): 455-458, August 1994.

[15] G. G. Brownlee and E. Fodor. The Predicted Antigenicity of the Haemagglutinin of the 1918 Spanish Influenza Pandemic Suggests an Avian Origin; Philosophical Transactions: Biological Sciences, 356(1416): 1871-1876, December 2001.

[16] D. R. Perez, R. J. Webby, E. Hoffmann, R. G. Webster. Land-Based Birds As Potential Disseminators of Avian Mammalian Reassortant Influenza A Viruses. Avian Diseases, 47: 1114–1117, 2003.

[17] Tomoko Ohta. The Nearly Neutral Theory of Molecular Evolution; Annual Review of Ecology and Systematics, 23: 263–286, 1992.