

Research of H5N6 Treatment by Comparing with H6N1 and H10N8 by Using Decision Tree and Apriori Algorithm

Sunghyun Kim¹, Hojoon Lee¹ and Taeseon Yoon²

¹Hankuk Academy of Foreign Studies, Natural Science Department, Yong-in, Republic of Korea

²Korea University, Computer Science and Engineering, Seoul, Republic of Korea

Abstract. Since 2003, 608 people in 15 countries have infected with human-infectious AI viruses and 359 of them died. Especially, in China, H6N1 and H10N8 viruses were wide-spread and a lot of people were infected and died. Recently, H5N6 virus emerged in China and the number of patients has been increasing gradually. Therefore, this research compared amino acid strain of Matrix Protein, Hemagglutinin, Neuraminidase and Nucleoprotein of H5N6, H6N1 and H10N8, by using Decision tree and Apriori Algorithm, to figure out their similarity and devise the treatment. In result, Matrix protein and Nucleoprotein sequences of H5N6 were similar with those of H6N1 and H10N8. Therefore, this research concluded that the treatment targeting those proteins of H6N1 and H10N8 will be also effective to H5N6.

1 Introduction

Since 1878, when the first AI(Avian Influenza) was reported in northern Italy, a great number of chickens and ducks have infected with various AI viruses and been buried. [1] At the beginning of the outbreak, it was considered that only avian features could come down with the flu like its name, but now even humans are the host of AI viruses.

Since 2003, 608 people in 15 countries have infected with human-infectious AI viruses and 359 of them died. Especially, in China, the number of people who died due to viruses including H10N8 and H6N1, is overwhelming compared to other East Northern Asian nations. [2] As a considerable number of AI viruses in East Asia were from China, there is a strong possibility that human-infectious AI viruses arisen in China will spread to near countries. [3] In this situation, examining of H5N6, a new human-infectious AI virus in China, can contribute to the treatment of not only the virus but future-emerging AI viruses.

Thus, we are going to compare H5N6 with H10N8 and H6N1 by using Decision tree, Apriori Algorithm and discuss the treatment of it according to the result. [4]

2 Materials and methods

2.1 Avian viruses

Since 1990s, many influenza A viruses have circulated in poultry populations in China, which offer abundant gene pools for mutations. All of three viruses we use in this

experiment are subtypes of influenza A viruses from China. [5]

2.1.1 H5N6

The first case of human H5N6 infection is a 49-year-old man from southwest China's Sichuan Province in April, 2014. Since the initial case, nine human infections were reported, including the recent cases in 4 January 2016 and 11 January 2016 in China. [6]

2.1.2 H6N1

H6N1 viruses had circulated persistently in poultry, and continued to evolve and accumulate changes, increasing the risk of cross-species infection. Although H6N1 virus is one of the most common viruses in wild avian species, it has only infected one person, a woman in China, who even recovered. [7]

2.1.3 H10N8

Human infection with H10N8 was reported in China in December 2013 and January 2014. Those two patients showed signs of pneumonia, and died soon despite the administration of antibiotics and antiviral agents. Thanks to the later experiments, there is no more fatal infection case. [8]

2.2 Algorithms

2.2.1 Decision tree

The A decision tree is a way of algorithm specialized in classifying data and sequential decision problems. A decision tree uses explicit tree-shape graph so that many can easily read and understand.

A decision tree splits the remaining data and make common attribute to node. Then set groups into subsets that usually branches down with particular attribute. When expressed, the data is described mathematically for its generalization and categorization. It follows the form of $(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$. Vector x is composed of input variables and a dependent variable Y is the purpose should be understood, classified, or generalized. [9]

2.2.2 Apriori algorithm

Apriori is one of the usual algorithms of association rule. Apriori is a method for clarify correlation among data on the basis of frequency and general rules of data. Apriori has two steps to start processes. First, it finds groups of high frequency following the minimum support previously set. Second, figure out the association rules which satisfy value of confidence among the groups of high frequency. Then, all the groups of high frequency's subsets have high frequency also, and certain group can not have higher support than before regardless of a new added item. Therefore, Apriori works based on following sentence, for a start, it finds groups of high frequency among the groups whose size is 1. Then among the groups whose size is 2, 3, k, until it can't make the group, and get the answer. called bottom up approach. [10]

2.3 Experiment method

First of all, we collected nucleotide sequences of H5N6, H10N8, H6N1 viruses from NCBI, "National Center for Biotechnology Information". We designed experiment using Decision tree and Apriori algorithm in window5, window7, and window9 for each virus. By using those algorithms, we track out certain frequencies in Matrix Protein(M), Hemagglutinin(HA), Neuraminidase(NA), and Nucleoprotein(NP).

3 Result

3.1. Decision tree

Table 1. Result of hemagglutinin(window5)

	Pos1	Pos2	Pos3	Pos4	Pos5
H5N6	H		K E V	A	
H6N1			V		
H10N8	A N		E		

Table 2. Result of neuraminidase(window5)

	Pos1	Pos2	Pos3	Pos4	Pos5
H5N6	I		N S L T	L N I	V
H6N1	S	P K	I N Y V W	I G N Q	T Y K
H10N8	V		H S Q N V	Q P D G W N	T A G

Table 3. Result of hemagglutinin(window7)

	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7
H5N6				V	S		
H6N1			L	S D R T			S N
H10N8	S	T N		F	N S K E H	R T	

Table 4. Result of neuraminidase(window7)

	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7
H5N6				K			A G
H6N1			I	S V	S		D A E
H10N8	I		T	L G			E D

Table 5. Result of hemagglutinin(window9)

	P1	P2	P3	P4	P5	P6	P7	P8	P9
H5N6				N		I		V	
H6N1	R			T	G	E		V	
H10N8	A W I	G E	G	A E M	E H		L	D T I	

Table 6. Result of neuraminidase(window9)

	P1	P2	P3	P4	P5	P6	P7	P8	P9
H5N6	K				M	E S I			I S H
H6N1	N					D G		S	I F G H
H10N8	R E C						I	Q	S V Q G

By using decision tree, we track out rules of patterns with certain frequencies in M, HA, NA, and NP. A, C, D, E, F,

G, H, I, K, L, M, N, P, Q, R, S, T, V are names of amino acids. We use three viruses as experiment variables in every fold, and the results show both similarities and differences. We used the sequences for a 10-fold cross validation experiment, that is to say, we test 10 experiments Table1-6. When we gather extraction results, we first selected the rules of patterns of three viruses whose frequency rates higher than 0.750 and extracted at least two times. This is because degree under 0.750 is considered as almost unfixed rule patterns that are too unclear to be sure.

The decision tree reveals whole kinds of amino acids of HA and NA, so we can find lots of similar and different amino-acids. Also, there were proportional relationship between the lengths of the rule(window number) and the amount of particular rules of high frequency. Usually, the rules have frequency of 0.75, but window9 pos8=M and pos1=W have frequency of 0.833, for instance.

Furthermore, not as we expect, there was no rule in the experiment results in NP and M. Considering the characteristic of decision tree, we can see this result as different amino acids among these viruses did not have differences that are enough to be a key factor. Therefore we can see these viruses as similar species.

3.2 Apriori algorithm

A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V are names of amino acids and the number means frequency. Star marks (*) beside amino acids in H6N1 and H10N8 mean that the rule is same with that of H5N6.

Table 7. Result of matrix protein(window5)

	H5N6	H6N1	H10N8
Rules	amino4=L 8 amino2=K 7 amino1=E 6 amino2=L 6 amino2=N 6 amino3=A 6 amino5=R 6 amino1=A 5 amino1=Q 5 amino1=S 5 amino2=A 5 amino3=G 5 amino3=M 5 amino3=Q 5 amino4=I 5 amino5=A 5	amino4=L 8 * amino1=A 6 amino1=E 6 * amino2=K 6 amino2=L 6 * amino3=A 6 * amino1=Q 5 amino2=S 5 amino3=G 5 * amino3=M 5 * amino3=Q 5 * amino4=I 5 * amino5=A 5 amino5=R 5 amino5=T 5 * amino5=V 5	amino4=L 8 * amino1=A 6 amino1=E 6 * amino2=A 6 amino2=L 6 * amino3=A 6 * amino5=A 6 amino1=Q 5 * amino2=K 5 amino2=T 5 amino3=G 5 * amino3=L 5 amino3=Q 5 * amino4=I 5 *

In M experiment Table 7, 9 of 17 rules of H5N6 were same with H6N1 and 8 were same with H10N8. Key factors are L(Leucine) and E(Glutamic acid). We discovered similar rate and feature in the experiment by window7 and window9.

In HA experiment Table 8, only 2 of 9 rules of H5N6 were same with H6N1, and none were same with H10N8. Even in the experiment by window7 and window9, we couldn't find any certain key factor..

Table 8. Result of hemagglutinin(window5)

	H5N6	H6N1	H10N8
Rules	amino3=L 16 amino5=N 14 amino1=L 13 amino5=A 13 amino3=K 12 amino2=E 11 amino3=S 11 amino4=N 11 amino5=I 11	amino2=N 14 amino1=S 13 amino3=K 13 amino3=L 13 amino4=G 12 amino5=E 12 amino1=L 11 amino4=N 11 * amino5=I 11 *	amino3=I 13 amino1=G 12 amino2=L 12 amino3=G 12 amino4=G 12 amino4=N 12 amino3=T 11

Table 9. Result of neuraminidase(window5)

	H5N6	H6N1	H10N8
Rules	amino1=I 13 amino3=T 11 amino1=N 10 amino1=S 10 amino3=S 10 amino2=I 9 amino2=T 9 amino4=C 9 amino4=G 9 amino5=G 9 amino5=S 9	amino5=S 17 amino3=I 13 amino4=I 12 amino1=G 11 amino2=S 11 amino2=S 10 amino4=G 10 amino4=S 10 amino5=G 10 amino2=G 9 amino3=N 9 amino4=N 9 amino5=T 9	amino5=G 12 amino4=S 11 amino4=V 11 amino2=V 10 amino3=G 10 amino3=S 10 * amino2=E 9 amino3=V 9 amino4=G 9 *

In NA experiment Table 9, none of 11 rules of H5N6 were same with H6N1, and 2 were same with H10N8. We got the similar results in the experiment by window7 and window 9. In the whole, S(Serine) and G(Glycine) appeared in large numbers, but we didn't regard amino acid in different location as a standard of choosing key factor.

Table 10. Result of nucleoprotein(window5)

	H5N6	H6N1	H10N8
Rules	amino1=R 18 amino4=E 15 amino3=A 12 amino5=G 12 amino1=A 11 amino1=L 11 amino2=G 10 amino3=L 10 amino5=R 10	amino1=R 18 * amino4=E 15 * amino3=A 12* amino5=G 12* amino1=A 11* amino2=S 11 amino1=L 10 amino2=G 10* amino2=R 10 amino3=L 10 * amino5=R 10 *	amino1=R 18 * amino4=E 15 * amino3=A 13 amino5=G 12* amino1=A 11* amino1=I 11 * amino1=L 10 amino2=G 10* amino2=R 10 amino2=S 10 amino3=L 10 * amino5=R 10 *

In NP experiment Table10, 8 of 9 rules of H5N6 were same with H6N1 and H10N8. We got the similar results in the experiment by window7 and window 9. We got the highest sameness rate(88.9% in both) in this experiment. Key factors are R(Arginine) and E(Glutamic acid).

4 Discussion and conclusion

According to Decision tree and Apriori algorithms, we found new-emerging H5N6 has significant similar Matrix protein(M) and Nucleoprotein(NP) sequences compared with H6N1 and H10N8. Therefore, we concluded that H6N1 and H10N8 virus treatment targeting M and NP will be effective to cure H5N6 virus. M2 protein is a channel protein that protons selectively pass through and plays an important role in replication cycle of influenza virus and NP provides important information in viral NA replication [11], [12].

This study provides efficient way to treat avian influenza virus by discovering similar sequences between three kinds of viruses and targeting those sequences with one specific treatment such as M and NP inhibitors. We could find similarities between three influenza viruses. This gives us relevant possibility to make a treatment targeting the inhibitors. If human inhibitors also have relevant similarities, scientists could make an efficient cure for the disease. But we should be cautious to this because if we target the wrong inhibitor due to misunderstanding, it could damage normal genes, which could lead to other serious disorder. For example, a non-replicating cell such as neurons or brain cells, a single-strand break or other type of damage in the transcribed strand of DNA can block RNA polymerase II to transcript gene sequence. Consequently, the body might not be able to make essential proteins, which regulate gene revelation. [13]

Even if the treatment for influenza is devised, there is still remaining concern as mentioned above. Researchers of this disease should notice that making the treatment of the influenza disease is not the only problem, but checking the possibility of malfunction is also important, and that taking care of all is the ultimate goal of public welfare.

References

1. Alexander, Dennis J. "A Review of Avian Influenza in Different Bird Species." *Veterinary Microbiology* 74.1-2 (2000): 3-13. Web
2. *Cumulative Number of Confirmed Human Cases for Avian Influenza A(H5N1) Reported to WHO, 2003-2013* (n.d.): n. pag. Web.
3. Hoffmann, E., J. Stech, I. Leneva, S. Krauss, C. Scholtissek, P. S. Chin, M. Peiris, K. F. Shortridge, and R. G. Webster. "Characterization of the Influenza A Virus Gene Pool in Avian Species in Southern China: Was H6N1 a Derivative or a Precursor of H5N1?" *Journal of Virology* 74.14 (2000): 6309-315. Web.
4. Somsri Wiwanitkit and Viroj Wiwanitkit. "The Three Newest Emerging Influenza H10N8, H5N6 and H6N1: Comparison of Its Morbidity and Mortality." *Asian Pacific Journal of Tropical Biomedicine* 5.5 (2015): 360. Web.
5. Gao, W., A. C. Soloff, X. Lu, A. Montecalvo, D. C. Nguyen, Y. Matsuoka, P. D. Robbins, D. E. Swayne, R. O. Donis, J. M. Katz, S. M. Barratt-Boyes, and A. Gambotto. "Protection of Mice and Poultry from Lethal H5N1 Avian Influenza Virus through Adenovirus-Based Immunization." *Journal of Virology* 80.4 (2006): 1959-964. Web.
6. Pan, Ming, Rongbao Gao, Qiang Lv, Shunhe Huang, Zhonghui Zhou, Lei Yang, Xiaodan Li, Xiang Zhao, Xiaohui Zou, Wenbin Tong, Suling Mao, Shumei Zou, Hong Bo, Xiaoping Zhu, Lei Liu, Heng Yuan, Minghong Zhang, Daqing Wang, Zumao Li, Wei Zhao, Maoli Ma, Yaqiang Li, Tianshu Li, Huiping Yang, Jianan Xu, Lijun Zhou, Xingyu Zhou, Wei Tang, Ying Song, Tao Chen, Tian Bai, Jianfang Zhou, Dayan Wang, Guizhen Wu, Dexin Li, Zijian Feng, George F. Gao, Yu Wang, Shusen He, and Yuelong Shu. "Human Infection with a Novel, Highly Pathogenic Avian Influenza A (H5N6) Virus: Virological and Clinical Findings." *Journal of Infection* 72.1 (2016): 52-59. Web.
7. "Human Infection with Avian Influenza A H6N1 Virus: An Epidemiological Analysis." *National Center for Biotechnology Information*. U.S. National Library of Medicine, n.d. Web. 30 Jan. 2016.
8. Zhang, Tao, Yuhai Bi, Huaiyu Tian, Xiaowen Li, Di Liu, Ying Wu, Tao Jin, Yong Wang, Qianjiao Chen, Ze Chen, Jianyu Chang, George F. Gao, and Bing Xu. "Human Infection with Influenza Virus A(H10N8) from Live Poultry Markets, China, 2014." *Emerg. Infect. Dis. Emerging Infectious Diseases* 20.12 (2014): n. pag. Web.
9. https://en.wikipedia.org/wiki/Decision_tree
10. https://en.wikipedia.org/wiki/Apriori_algorithm
11. http://en.wikipedia.org/wiki/M2_proton_channel
12. <http://drugdiscoveryopinion.com/22010/06/targeting-influenza-a-nucleoprotein/>
13. Schinecker, T. M. "Human RNA Polymerase II Is Partially Blocked by DNA Adducts Derived from Tumorigenic Benzo[c]phenanthrene Diol Epoxides: Relating Biological Consequences to Conformational Preferences." *Nucleic Acids Research* 31.20 (2003): 6004-015. Web.