

## What These Trends Suggest?

<sup>1</sup>Shaomin Yan and <sup>2</sup>Guang Wu

<sup>1</sup>Guangxi Academy of Sciences, 98 Daling Road,  
Nanning, Guangxi Province, CN-530007, China

<sup>2</sup>DreamSciTech Consulting, 301, Building 12, Nanyou A-zone,  
Jiannan Road, Shenzhen, Guangdong Province CN-518054, China

---

**Abstract: Problem statement:** Global warming is one of most significant factors affecting the biological evolution and the influenza is the disease that threatens humans with possible epidemics or pandemics. It would be important to understand if the global warming would have potential impact on the evolution of influenza virus. For this aim, the first would be to study the trend of evolution of proteins from influenza A virus and compare it with the trend in global warming. **Approach:** The evolution of polymerase acidic proteins of influenza A virus from 1918-2008 was defined using the unpredictable portion of amino-acid pair predictability. Then the trend in this evolution was compared with the trend in the global temperature, the temperature in north and south hemispheres and the temperature in influenza A virus sampling site and species carrying influenza A virus. **Results:** The similar trend was found between global warming and evolution of polymerase acidic proteins although we could not correlate them at this stage of study. **Conclusion:** The study suggested the potential impact of global warming on the evolution of proteins from influenza A virus.

**Key words:** Evolution, global warming, influenza, mutation, polymerase acidic protein

---

### INTRODUCTION

It is said that the climate change has an extinction risk for many species<sup>[1]</sup>. If so, we would expect to see that the protein evolution would be affected to some degree although some proteins could be hidden deeply inside cells. This theme would lead to an important open question.

The protein evolution is a process of mutations within protein family, but definitely we cannot identify all the mutation causes one by one because some cause that led to mutations in the past might not leave any trace due to the environment changes. However, it is certain that all the mutation causes leave their traces in protein, whose amino acids differ over evolutionary process. Thus the open question is how we can present the evolution of different amino-acid proteins along the time course, because the evolution is directly referenced to the time. In other words, how is a protein family represented along the time course?

We perhaps would have many ways for this, just as we have many ways to deal with other things over time. For example, we can follow a man's height, body weight, count of red blood cells from birth to death. Although each measure represents an interesting aspect of life, the

essence is that these measures are numbers. For that, we need to change a 20-letter symbolized protein into a number, more accurate, a scalar datum, because the scalar data are far easy to present in coordinates along the time course as an evolutionary process.

Since 1999, our group has developed three approaches to change a 20-letter symbolized protein into a scalar datum based on the random mechanism<sup>[2-5]</sup>, not only because pure chance is now considered to lie at the very heart of nature<sup>[6]</sup>, but also more importantly the random approach is frequently used when we have no detailed knowledge on a process, in our case, we have no detailed knowledge on all the mutation causes. On the other hand, the data changed by random approach reflects only a single aspect of a protein, as height and weight reflect different aspects of life. Of course, there are many other ways to change a protein into a datum, for example, we can use the physicochemical property of amino acid to change a protein into a numeral sequence<sup>[7]</sup>, however the physicochemical property is not subject to mutations, which engineer the evolution, so we cannot use this approach to plot the evolution of a protein family over time.

If we can effectively represent a protein family along the time course, the pattern of this protein family

---

**Corresponding Author:** Guang Wu, DreamSciTech Consulting, 301, Building 12, Nanyou A-Zone, Jiannan Road, Shenzhen, Guangdong Province CN-518054, China Tel: +867712503930 Fax: +867712503999

would be its evolutionary process from the viewpoint of how the proteins have been changed so far. Furthermore, we can compare the evolutionary process of a protein family with any other time series of interests to see if there is a similar trend between them although similar trends are by no means the correlation.

Influenza viruses replicate and transcribe their segmented negative-sense single-stranded RNA genome in the nucleus of the infected host cell. All RNA synthesizing activities associated with influenza virus are performed by the virally encoded RNA-dependent RNA polymerase that consists of three subunits, polymerase acidic protein (PA), polymerase basic proteins 1 and 2<sup>[8]</sup>. The PA subunit is involved for the conversion of RNA polymerase from transcriptase to replicase<sup>[9]</sup> and contains the endonuclease active site<sup>[10,11]</sup>. A recent study strongly implicates the viral RNA polymerase complex as a major determinant of the pathogenicity of the 1918 pandemic virus<sup>[12]</sup>.

Thus the evolutionary trend of PA would provide more understanding and insight on influenza and how humans can prevent influenza, which is designed as the aim of this study.

## MATERIALS AND METHODS

**Temperature data:** The global, north and south hemispheric temperature anomalies from 1850-2007, whose anomaly is based on the period 1961-1990, were obtained from HadCRUT3v<sup>[13,14]</sup>. The local temperature from 1918 to 1998 based on 0.5 by 0.5° latitude and longitude grid-box basis cross globe was obtained from New *et al.*<sup>[15]</sup>.

**PA data:** A total of 5165 full-length PA sequences of influenza A virus sampled from 1918-2008 was obtained from the influenza virus resources<sup>[16]</sup>. After excluded identical sequences, 2433 PAs are used in this study.

**Changing of PAs into scalar data:** We need using a scalar datum to represent a PA, which must differ for different PAs. Among our three random approaches<sup>[5]</sup>, the simplest one is the amino-acid pair predictability, which we have used in many studies<sup>[5,17-21]</sup>.

For a PA sequence, we count the first and second amino acids as a pair, the second and third amino acids as another pair, until the next to terminal and the terminal amino acids as the last pair. Then, we determine whether an amino-acid pair can be explained by permutation, or predicted by random mechanism. Finally we determine the percentage of how many amino-acid pairs in a PA are predictable and unpredictable.

For example, there are 77 glutamic acids “E” and 48 isoleucines “I” in the PA isolated from 1918 influenza A virus, accession number ABA55040 A/Brevig Mission/1/1918(H1N1), which is composed of 716 amino acids. If the appearance of amino-acid pair EI can be explained by permutation, it must appear five times in this PA ( $77/716 \times 48/715 \times 715 = 5.162$ ). Actually we do find five EIs in this PA. Thus, the appearance of EI can be explained by permutation or predicted by random mechanism. By clear contrast, there are 36 glycines “G” in 1918 PA. If the appearance of GE can be explained by permutation, it must appear four times ( $36/716 \times 77/715 \times 715 = 3.8715$ ). However, it appears only once in reality, which cannot be explained by permutation or randomly unpredictable. In this way, we classify all of the amino-acid pairs in this PA as predictable and unpredictable.

It is absolutely necessary that the predictable/unpredictable portion is subject to a tiny difference between two PAs, thus different PAs should have different values to be distinguishable. In the past, we have tested many proteins to verify this request and got the positive answer<sup>[2-5]</sup>.

For example, there are two H7N2 influenza A viruses isolated from New York in 2003 (accession number ACJ69195 and ACJ69220) and their PAs have only one amino acid different at position 235. However, the predictable and unpredictable portions are 26.15 and 73.85% for ACJ69195 PA, while they are 26.71 and 73.29% for ACJ69220 PA.

In this manner, we change 2433 letter-symbolized PAs into 2433 scalar data<sup>[22]</sup>. As each PA has its sampling year, we would have two scalar datasets, one is temperature and the other is the 2433 scalar data, thus we can plot both along the time course.

## RESULTS

Figure 1 shows the trends in both global warming and PA evolution, where both trends are quite similar as indicated by the regressed lines. This means that the PA evolution in sense of our measure has a similar trend as the trend in global temperature.

Moreover, the global temperature is generally divided as temperatures in north and south hemisphere, thus we can therefore group PAs accordingly to see if the trend still holds on in such circumstance. As shown in Fig. 2, the similar trends are much clearer in north hemisphere than in south hemisphere, which can be explained by the fact that there is more area of ocean in south hemisphere than in north hemisphere<sup>[15]</sup>. This explanation once again supports the trends because most of PAs were sampled in north hemisphere, whose trends go along the same direction.

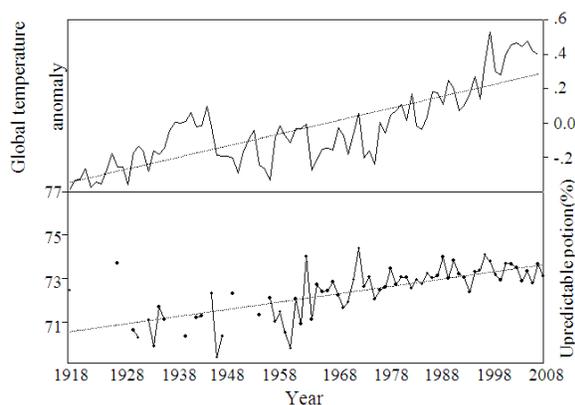


Fig. 1: Global temperature anomaly (°C) and PA evolution of influenza A viruses. The dotted lines and points are regressed lines and the mean of all PAs in a given year

Actually the PA data in Fig. 1 and 2 were averaged in each year, for example, there is only a single sample in 1918, but 22 PAs were sampled in 2008. Another way to analyze the trends is to apply the point-to-point method, that is, we correspond the temperate according the place and year a particular PA was sampled. In other word, we take the temperature measured at each geographical latitude and longitude of place where a PA was sampled at the same year to make the comparison.

For example, the PA of 1918 H1N1 influenza virus was sampled at Brevig Mission (accession number AAF77036)<sup>[23-25]</sup>, whose latitude and longitude are 65.34 and 166.49 west according to Get Lat Lon<sup>[26]</sup>. So we can find that the average yearly temperature was -6.26°C in 1918 according to the 0.5° by 0.5° latitude and longitude grid-box basis cross globe obtained from New *et al.*<sup>[15]</sup>.

Figure 3 shows 823 point-to-point relationships between temperature and PA and the regression indicates the similar trends between temperature and PA evolution. The results in Fig. 3 are in consistent with what we found in Fig. 1 and 2, that is, there are similar trends between global warming and PA evolution.

Until now, our analysis concentrates on the PAs as whole. However, we can advance our analysis along the thought in Fig. 3 to analyze the point-to-point relationship between temperature and species, from which the PAs were sampled, as we know influenza viruses are hosted in different species.

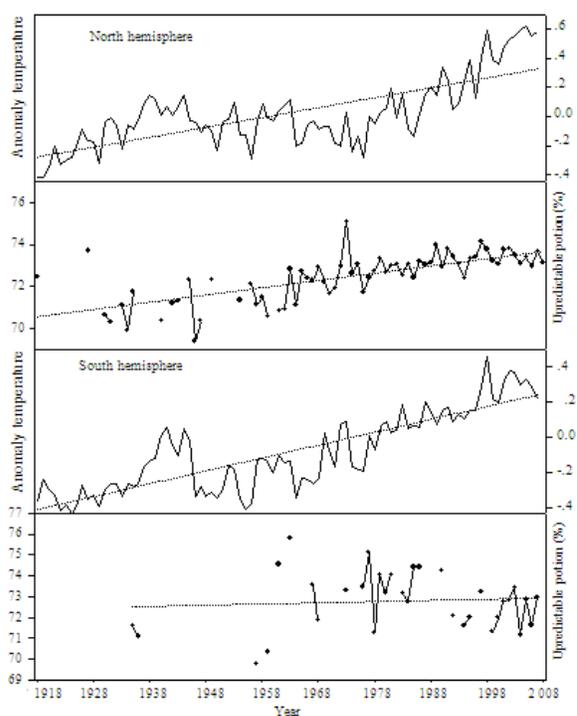


Fig. 2: Trend in temperature and PA evolution grouped according to north (n = 2241) and south (n = 192) hemispheres. The dotted lines are regressed lines

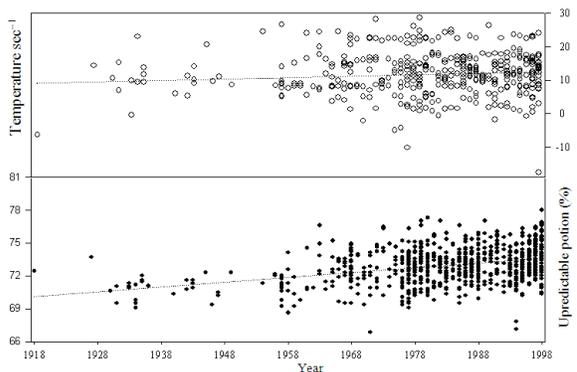


Fig. 3: Point-to-point temperature versus PA (n = 823). Each point presents a local temperature (°C) at the given year (upper panel), corresponding to the place where a PA was sampled (lower panel). The dotted lines are regressed lines

Figure 4 shows the trends of PA evolution with respect to temperature in three major host species, avian, human and swine. Once again, the trends are quite clear.

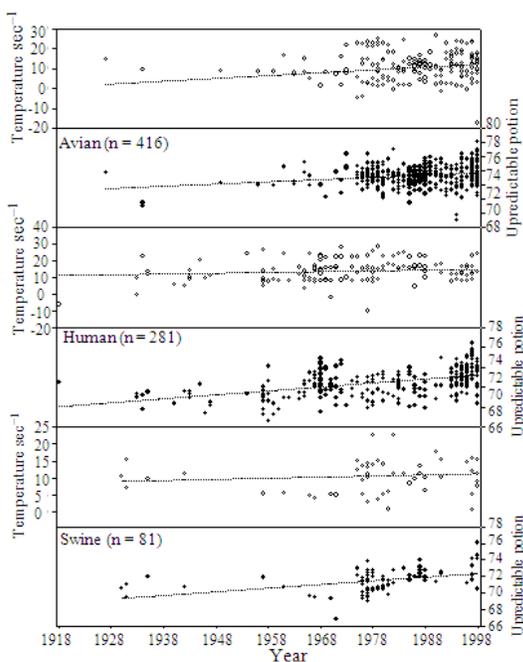


Fig. 4: Global temperature anomaly (°C) and PA evolution sampled from different species

### DISCUSSION

Although the trends in global warming and PA evolution are similar in this study, they only indicate the direction for studies in future, because we would not expect to determine such an important issue within a few studies and we still have much to discuss.

First, can we correlate both trends in this study statistically? At this stage, it would be difficult to determine such correlation because (i) To the best of our knowledge there is no statistical method available to determine the correlation between two lines including a discontinued one in this study and (ii) Moreover, many statistical books tell that the correlation does not mean the cause-consequence relationship, that is, even we would find the so-called correlation between two trends, we still need to determine if there is any direct or indirect cause-consequence relationship.

Second, can we ignore these trends as if they would not exist at any rate? At this stage, we can only admit the trends existed, because we cannot create another earth without global warming but with active influenza virus for comparison over the same time scale. As the validation of global warming is done through the comparison along the time course, we would argue that the validation of PA evolution should also be done along the time course, say, the comparison between any two different time points.

Third, can the scientific reasoning support the results? This is possible and is also the first step for all the initial observations in scientific history: (i) Currently the only well-known and profound factor existed for last 100 years is global warming. Although one may suggest the increase in global population, for instance, would have the similar trend along the time course, we would argue that the increase in global population is a contributing factor to the global warming. (ii) If it is widely considered that the climate change endangers many species<sup>[1]</sup>, then why have proteins in species not been affected?

Fourth, can we build a direct connection, which goes from global warming to the protein within a host? This could be possible although the interrelationships would make the connection very complex, which would be the objective of network biology.

Fifth, can we explain the PA evolution in context that two trends go along the time course with the similar direction? In other word, what does our measurement, unpredictable portion, mean? Our method divides a protein as randomly predictable and unpredictable portions. In daily-life, we can use certain equations to predict the range of bodyweight of certain man, if the prediction is correct, then this man's bodyweight is predictable, if not, the unprofitable means either overeating or over-fasting. Similar, the unpredictable portion of protein suggests that the construction of this portion should require more time and energy as we know that the predictable event based on permutation would have the biggest chance of occurrence<sup>[2-5]</sup>. Hence, the nature should have a sufficient reason to deliberately spend more time and energy to construct a larger unpredictable portion through a protein evolution, which is what we have shown in this study. Therefore as near as we can determine, the sufficient reason for the nature to deliberately to do so would be the global warming.

### CONCLUSION

This study suggested that the global warming could have a potential impact on the evolution of proteins from influenza A virus.

### ACKNOWLEDGMENT

The researchers wish to thank Tammy W. Beaty, ESD GIS facilities manager, computer analysis, design, systems support staff and Les A Hook, documentation coordinator, ORNL NASA-DAAC, Oak Ridge National Laboratory, Bethel Valley Road, MS. 6407, Oak Ridge, TN 37831-6407, USA for suggesting the filezilla

(<http://filezilla-project.org>) to download the datasets and related information on data format.

This study is supported in part by Guangxi Sci Key 0630003A2.

## REFERENCES

1. Thomas, C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F. Erasmus, M.F. De Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A.S. Van Jaarsveld, G.F. Midgley, L. Miles, M.A. Ortega-Huerta, A.T. Peterson, O.L. Phillips and S.E. Williams, 2004. Extinction risk from climate change. *Nature*, 427: 145-148. DOI: 10.1038/nature02121
2. Wu, G. and S. Yan, 2002. Randomness in the primary structure of protein: Methods and implications. *Mol. Biol. Today*, 3: 55-69. <http://www.horizonpress.com/mbt/v/v3/08.pdf>
3. Wu, G. and S. Yan, 2006. Mutation trend of hemagglutinin of influenza a virus: A review from computational mutation viewpoint. *Acta Pharmacol. Sinica*, 27: 513-526. DOI: 10.1111/j.1745-7254.2006.00329.x
4. Wu, G. and S. Yan, 2006. Fate of influenza A virus proteins. *Protein Peptide Lett.*, 13: 399-406. DOI: 10.2174/092986606775974474
5. Wu, G. and S. Yan, 2008. *Lecture Notes on Computational Mutation*. 1st Edn., Nova Science Publishers, New York, ISBN: 10: 1604565160, pp: 230.
6. Everitt, B.S., 1999. *Chance Rules: An Informal Guide to Probability, Risk and Statistics*. 1st Edn., Springer, New York, ISBN: 10: 0387987762, pp: 202.
7. Chou, K.C., 2004. Structural bioinformatics and its impact to biomedical science. *Curr. Med. Chem.*, 11: 2105-2134. DOI: 10.2174/0929867043364667
8. Engelhardt, O.G. and E. Fodor, 2006. Functional association between viral and cellular transcription during influenza virus infection. *Rev. Med. Virol.*, 16: 329-345. DOI: 10.1002/rmv.512
9. Honda, A. and A. Ishihama, 1997. The molecular anatomy of influenza virus RNA polymerase. *Biol. Chem.*, 378: 483-488. <http://www.ncbi.nlm.nih.gov/pubmed/9224927>
10. Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S. and R.W. Ruigrok, 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, DOI: 10.1038/nature07745
11. Yuan, P., M. Bartlam, Z. Lou, S. Chen, J Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao and Y. Liu, 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*, DOI: 10.1038/nature07720
12. Watanabe, T., S. Watanabe, K. Shinya, J.H. Kim, M. Hatta and Y. Kawaoka, 2009. Viral RNA polymerase complex promotes optimal growth of 1918 virus in the lower respiratory tract of ferrets. *Proc. Natl. Acad. Sci. USA.*, 106: 588-592. DOI: 10.1073/pnas.0806959106
13. Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett, 2006. Improved analyses of changes and uncertainties in marine temperature measured *in situ* since the mid nineteenth century: The HadSST2 dataset. *J. Clim.*, 19: 446-469. [http://hadobs.metoffice.com/hadsst2/rayner\\_et\\_al\\_2005.pdf](http://hadobs.metoffice.com/hadsst2/rayner_et_al_2005.pdf)
14. Climatic Research Unit, 2008. <http://www.cru.uea.ac.uk/cru/data/temperature/>
15. New, M., M. Hulme and P. Jones, 2000. Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *J. Clim.*, 13: 2217-2238. DOI: 10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2
16. Influenza virus resources, 2008. <http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/multiple.cgi>
17. Wu, G. and S. Yan, 2008. Prediction of mutations engineered by randomness in H5N1 neuraminidases from influenza a virus. *Amino Acids*, 34: 81-90. DOI: 10.1007/s00726-007-0579-z
18. Wu, G. and S. Yan, 2005. Timing of mutation in hemagglutinins from influenza A virus by means of unpredictable portion of amino-acid pair and fast Fourier transform. *Biochem. Biophys. Res. Commun.*, 333: 70-78. DOI: 10.1016/j.bbrc.2005.05.094
19. Wu, G. and S. Yan, 2008. Prediction of mutations initiated by internal power in H3N2 hemagglutinins of influenza a virus from North America. *Int. J. Peptide Res. Ther.*, 14: 41-51. DOI: 10.1007/s10989-007-9104-1
20. Wu, G. and S. Yan, 2008. Prediction of mutations engineered by randomness in H5N1 hemagglutinins of influenza a virus. *Amino Acid*, 35: 365-373. DOI: 10.1007/s00726-007-0579-z
21. Wu, G. and S. Yan, 2007. Improvement of prediction of mutation positions in H5N1 hemagglutinins of influenza A virus using neural network with distinguishing of arginine, leucine and serine. *Protein Peptide Lett.*, 14: 465-470. DOI: 10.2174/092986607780782713
22. Amino-acid pair predictability. 2008. <http://www.dreamscitech.com/Service/rationale.htm>

23. Reid, A.H., T.G. Fanning, J.V. Hultin and J.K. Taubenberger, 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. USA.*, 96: 1651-1656. <http://www.pnas.org/content/96/4/1651.full>
24. Reid, A.H., T.G. Fanning, T.A. Janczewski and J.K. Taubenberger, 2000. Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. USA.*, 97: 6785-6790. <http://www.pnas.org/content/97/12/6785.full>
25. Basler, C.F., A.H. Reid, J.K. Dybing, T.A. Janczewski, T.G. Fanning, H. Zheng, M. Salvatore, M.L. Perdue, D.E. Swayne, A. García-Sastre, P. Palese and J.K. Taubenberger, 2001. Sequence of the 1918 pandemic influenza virus Nonstructural Gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc. Natl. Acad. Sci. USA.*, 98: 2746-2751. <http://www.pnas.org/content/98/5/2746.full.pdf>
26. Get Lat Lon. 2008. <http://www.getlatlon.com/>