Forecasting influenza A pandemic outbreak using protein dynamical network biomarkers

Jie Gao¹², Kang Wang¹, Tao Ding¹ and Shanshan Zhu¹

From The 10th International Conference on Systems Biology (ISB 2016)
Weihai, China. 19-22 August 2016

Abstract

Background: Influenza A virus is prone to mutation and susceptible to human beings and spread in the crowds when affected by the external environment or other factors. It is very necessary to forecast influenza A pandemic outbreak.

Methods: This paper studies the different states of influenza A in the method of dynamical network biomarkers. Through establishing protein dynamical network biomarkers of influenza A virus protein, a composite index is ultimately obtained to forecast influenza A pandemic outbreak.

Results: The composite index varies along with the state of pandemic influenza virus from a relatively steady state to critical state before outbreak and then to the outbreak state. When the composite index continuous decreases for 2 years and increases of more than 0.1 suddenly, it means the next year is normally in the outbreak state. Therefore, we can predict and identify whether a certain year is in the critical state before influenza A outbreak or outbreak state by observing the variation of index value. Meanwhile, through data analysis for different countries influenza A pandemic outbreak in different countries can also be forecasted.

Conclusions: This indicates the composite index can provide significant warning information to detect the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic.

Keywords: Influenza A pandemic, Protein dynamical network biomarker (PDNB), The critical state, The outbreak state

Background

It is proved that there is a kind of common critical phenomenon in lots of complex biological process, i.e. a relative stable state quickly enters into another state after a critical point in a very short period of time [1, 2]. There is the kind of critical phenomenon for influenza A, because it needs only a very short period of time quickly from a relative stable state to outbreak state after a critical point. Thus in order to timely and effectively prevent and control the outbreak of influenza A pandemic, the key lies in predicting the critical point before the outbreak.

At present, influenza A is studied from all aspects. Pan et al. found that the spatio-temporal network that connects the cities with human cases along the order of outbreak timing emerges two-section-power-law edge-length distribution, using the empirical analysis and modeling studies [3]. Chang et al. studied the vaccine for influenza, so as to achieve the effect of prevention of influenza [4]. Banerjee et al. made full comparisons for the structural features of all H1N1 HA gene sequences and the composition of global amino acid to make it possible to depict the developing trend of influenza A [5]. He et al. also made in-depth studies to identify HA protein epitopes of avian influenza virus [6].

This paper studies the different states of influenza A using DNB. Through establishing PDNB of influenza A virus protein and using the nature of DNB, a composite index is ultimately obtained to forecast influenza A pandemic outbreak. The composite index varies along with the state of pandemic influenza virus from a relatively steady state to critical state before outbreak and then to the outbreak state. Therefore, we can predict and identify whether a certain year is in the critical state before influenza A pandemic outbreak using protein dynamical network biomarkers.
influenza A outbreak or outbreak state by observing the variation of index value. This indicates the composite index can provide significant warning information to detect the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic. Meanwhile, through data analysis for different countries influenza A pandemic outbreak in different countries can also be forecasted.

Methods
DNB analysis
The concept of network biomarkers is set up with the development of high-throughput genomic technologies and the systematic and multidimensional study of molecular expression profiling [7, 8]. This concept refers to a series of markers as well as their mutual relations and has been proposed as a new marker type [9]. Compared with traditional biomarkers, these markers can accurately distinguish disease states for taking the links between the molecules into consideration [10, 11]. However, it is used to diagnose the states of diseases, not for the detecting the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic. The method of dynamic network biomarkers focuses on the detection and assessment of different stages of the disease in the development of disease and shows it is a time-dependent method [12]. It studies the location changes of the markers over time and the relationship among network markers over time changing and then constructs three-dimensional images showing the interaction relationship between the markers. Therefore the study of Network markers focuses on the molecular interactions and distinguishes normal and disease states, and the study of dynamic network markers focuses on dynamic changes, which is helpful to discover the marker accurately and comprehensively and further to distinguish the state of disease before outbreak. It not only does not depend on the method of small sample excavation mode markers, but also make it easier for clinical application. At the same time it can be used in future studies to find early warning signals in any biological process, such as differentiation, senescence and cell cycle of each phase as well as key change.

Defining PDNB
Firstly, taking hemagglutinin (HA) protein as an example, we suppose that a HA protein marked y is linked sequentially by t numbers of amino acids. Its amino acid sequence is represented as \( y = x_1, x_2 \cdots x_t \), in which \( x_i \in \{A, V, L, I, P, F, W, M, D, E, G, S, T, C, Y, N, Q, K, R, H\}; i = 1, 2, \cdots , t \).
We suppose s-1-th year have m numbers of influenza virus HA proteins all over the world and its amino acid sequence is represented as \( y_{s-1,1}, y_{s-1,2}, \cdots , y_{s-1,m} \). Meanwhile, We suppose s-th year have n numbers of influenza virus HA proteins all over the world and its amino acid sequence is represented as \( y_{s,1}, y_{s,2}, \cdots , y_{s,n} \).
The amino acid number of the \( y_{i,j} \), is marked \( c_{i,j} \)
where \( i=s-1,s; j = 1, 2, \cdots , q \) and \( q = \max \{m, n\} \). Sequentially selecting the i-th amino acid for \( y_{s-1,1}, y_{s-1,2}, \cdots , y_{s-1,m} \) to form a new amino acid sequence is defined as \( Z_{s-1,i} \) and then take out the one of the largest number of amino acids. If the maximum number of amino acids has two or more than two, we take the first amino acid without loss of generality. At the same time, it is marked \( x_0 \) where \( i = 1, 2, \cdots , k \) and \( k = \max \{c_{s-1,1}, c_{s-1,2}, \cdots , c_{s-1,m}\} \). We individually connect them in order to form a new amino acid sequences (\( U_{s-1} = x_1 x_2 \cdots x_k \)) and then separately compare with corresponding amino acids of \( y_{s,1}, y_{s,2}, \cdots , y_{s,n} \) one by one. If they are different, the assignment is 1, on the contrary the assignment is 0. Therefore, n new sequences are represented by \( E_{s,1}, E_{s,2}, \cdots , E_{s,n} \) obtained in s-th year. Then we calculate their mean (M), standard deviation (SD) and coefficient of variation (CV). Their computation formulas are as follows:

\[
M_s = \frac{\sum_{i=1}^{n} f(s,i)}{n} 
\]

\[
SD_s = \sqrt{\frac{\sum_{i=1}^{n} (f(s,i) - M_s)^2}{n}} 
\]

\[
CV_s = \frac{SD_s}{M_s} 
\]

where \( f(s,i) \) represents the frequency of occurrence of one in sequence \( E_{s,i} \). Similarly, we calculate M, SD and CV of the other nine proteins. The protein that the top three values of \( CV_s \) are defined as core protein (CP), and the others are no-core protein (NP). CP is a set of high confident interactions of proteins, which forms a sub-network called influenza A virus proteins of protein dynamical network biomarkers (see Fig. 1).

Defining forecasting index
The frequencies of the 20 kinds of amino acids can be calculated in turn, so we can get a 23 dimensional characteristic value vector of each phase as well as key change.
where \( V_t(s) \) represents the characteristic value vector of the \( t \)-th influenza A protein, \( t = 1, 2, \ldots, 10 \). Defining the characteristic distance between proteins:

\[
d_{vw} = \sqrt{(M_v - M_w)^2 + (\sigma_v - \sigma_w)^2 + (C_v - C_w)^2 + \sum_{i=1}^{20} (f_{vx}(s) - f_{wx}(s))^2)
\]

where \( v \) and \( w \) respectively represents the \( v \)-th and the \( w \)-th protein.

The core proteins are not only the universal indicators to detect the complex outbreak signal of influenza A, but also the dominant or driving network of the whole protein system in the development, mutation and outbreak of the critical stages. In fact, the dominant network breaks through the limits of variation in the first time, first enters to the state of variation, and then affects other proteins and lead to the transfer of the entire system. Therefore, the determination of the dominant network can not only detect system in the critical state before break out, also help to reveal the underlying mechanism of influenza A virus proteins from the dimension of dynamic network.

By combining the above properties of the core proteins, we can get a composite index:

\[
I = \frac{CV_k CP_{cd}}{NP_{cd}}
\]

Where \( \overline{CV_k} \) represents the average value of the core proteins’ \( CV_v \), \( CP_{cd} \) is the average value of the characteristic distance between the core proteins, \( NP_{cd} \) is the average value of the characteristic distance between the core and non-core proteins.

When \( I_{s,3} > I_{s,2} > I_{s,1} \) and \( I_{s-1} - I_{s-2} > 0.1 \), it can be concluded that \( s + 1 \) year is in the outbreak state.

Although the amino acid sequence of each protein will fluctuate randomly, the composite index can provide significant early warning information when the influenza A virus is close to the critical state before the outbreak or the outbreak state.

**Results**

**Forecasting influenza A pandemic outbreak**

Ten of proteins for influenza A virus are hemagglutinin (HA), matrix protein, matrix protein 2, neuraminidase, non-structural protein 1, non-structural protein 2, nucleocapsid protein, PA RNA polymerase, PB1 RNA polymerase and PB2 RNA polymerase. They are composed of 20 different amino acids link to form polymers. This paper selects influenza A virus protein sequences from 1934 to September 2016 from the NCBI website (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database), lots of data before 1934 are absent.
As shown in Table 1, by using the above methods to calculate the composite index of the 1934 to September 2016. However, we can’t figure out the composite index of some years, because some data in 1937–1942, 1944–1945, 1952–1956 years are absent.

Forecasting influenza A pandemic outbreak in pandemic occurrence place

Through influenza A virus protein data analysis for different countries influenza A pandemic outbreak in different countries can also be forecasted. Take China as an example, this paper selects influenza A virus protein sequences occurred in China from the NCBI website to forecast influenza A pandemic outbreak in China. Whereas lots of data in 1954–1956, 1958–1963, 1965, 1967 years are absent, all data before 1954 and in 2016 year are absent.

As shown in Table 2, by using the above methods to calculate the composite index of the 1954 to 2015. However, we can’t figure out the composite index of some years, because lots of data in 1954–1956, 1958–1963, 1965, 1967 years are absent, all data before 1954 and in 2016 year are absent.

Discussion

Forecasting influenza A pandemic outbreak

The dynamic network markers of Pandemic influenza virus vary in the whole process from a relatively stable state to the critical state before outbreak as well as the outbreak state, which results in the status transfer of the entire network and finally results in fluctuations in the composite index. Therefore, by observing the transformation of the composite index, we can predict the critical state before the outbreak of pandemic influenza and the outbreak state.

The flu broke out in Hong Kong in 1968 and continued until 1969, of which 7.5 million people died. In 1972, influenza broke out in Henan Province and quickly spread to the entire province. As shown in Fig. 2, in 1964, the composite index value is 0.650854, 1965 is 0.527201; 1966 is 0.500452; 1967 is 0.666783; 1968 is 2.31271; 1969 is 1.081257; 1970 is 0.405805; 1971 is 0.728516. Because I_{1964} > I_{1965} > I_{1966} and I_{1967} - I_{1966} > 0.1, 1968 is in the outbreak state. Similarly, I_{1969} > I_{1966} > I_{1967} and I_{1971} - I_{1970} > 0.1, so 1972 is in the outbreak state.

The influenza A broke out in The United States, Russia and Japan in 1976 and 1977. Although the prevalence of this flu was typical of the outbreak, adults were slightly infected, and the incidence rate

| Year | Year | Year | Year |
|------|------|------|------|
| 1957 | 0.835938 | 1982 | 0.566947 | 1999 | 0.702834 |
| 1964 | 0.464712 | 1983 | 0.489632 | 2000 | 0.718456 |
| 1966 | 0.571425 | 1984 | 0.467346 | 2001 | 0.645986 |
| 1968 | 2.198217 | 1985 | 0.896537 | 2002 | 0.673956 |
| 1969 | 1.217563 | 1986 | 1.025471 | 2003 | 0.653958 |
| 1970 | 0.645643 | 1987 | 0.637485 | 2004 | 0.543824 |
| 1971 | 0.765451 | 1988 | 0.653873 | 2005 | 0.854385 |
| 1972 | 2.189534 | 1989 | 0.712854 | 2006 | 1.632657 |
| 1973 | 0.815746 | 1990 | 0.762368 | 2007 | 0.549367 |
| 1974 | 0.776357 | 1991 | 0.772543 | 2008 | 0.924375 |
| 1975 | 0.780346 | 1992 | 0.805734 | 2009 | 2.184658 |
| 1976 | 0.856282 | 1993 | 0.784756 | 2010 | 0.843569 |
| 1977 | 0.902765 | 1994 | 0.812367 | 2011 | 0.483967 |
| 1978 | 0.536847 | 1995 | 0.765347 | 2012 | 0.513975 |
| 1979 | 0.422576 | 1996 | 0.714358 | 2013 | 0.563954 |
| 1980 | 0.468257 | 1997 | 0.726856 | 2014 | 0.606837 |
| 1981 | 0.496629 | 1998 | 0.683546 | 2015 | 0.663854 |

Fig. 2 Trend Chart of composite index values from 1964 ~ 1972. Horizontal axis represents the year from 1964 ~ 1972, vertical axis represents the composite index value.
was very high in young people. As shown in Fig. 3, in 1972, the composite index value is 2.379322; 1973 is 0.79888; 1974 is 0.527835; 1975 is 0.801294. \( I_{1972} > I_{1973} > I_{1974} \) and \( I_{1975} - I_{1974} > 0.1 \), so 1976 is in the outbreak state.

The influenza A broke out in The United States and Japan in 1986. Meanwhile, many countries in Asia and Europe had the outbreak of influenza A. As shown in Fig. 4, in 1982, the composite index value is 0.650454; 1983 is 0.449789; 1984 is 0.354632; 1985 is 0.939702. \( I_{1982} > I_{1983} > I_{1984} \) and \( I_{1985} - I_{1984} > 0.1 \), so 1986 is in the outbreak state.

The influenza A broke out in China in 2006. Global influenza pandemic caused by the new influenza A virus in 2009, of which 0.3 million people died [13, 14]. As shown in Fig. 5, in 2002, the composite index value is 0.660193; 2003 is 0.465805; 2004 is 0.45421; 2005 is 0.772595; 2006 is 1.595902; 2007 is 0.476057; 2008 is 0.798138. \( I_{2002} > I_{2003} > I_{2004} \) and \( I_{2005} - I_{2004} > 0.1 \), \( I_{2006} > I_{2007} \) and \( I_{2008} - I_{2007} > 0.1 \), so 2006 is in the outbreak state. Although \( I_{2005} \) is not larger than \( I_{2006} \), 2006 is outbreak year and other conditions are in line, so there is still the outbreak state in 2009.

The influenza A broke out in India in 2015, of which 1.5 thousand people died [15]. As shown in Fig. 6, in 2011, the composite index value is 0.740067; 2012 is 0.63573; 2013 is 0.606009; 2014 is 0.806321. \( I_{2011} > I_{2012} > I_{2013} \) and \( I_{2014} - I_{2013} > 0.1 \), so 2015 is in the outbreak state.

In general, the composite index varies along with the state of pandemic influenza virus from a relatively steady state to critical state before outbreak and then to the outbreak state. When the composite index continuous decreases for 2 years and increases of more than 0.1 suddenly, it means the next year is normally in the outbreak state. Therefore, we can predict and identify whether a certain year is in the critical state before influenza A outbreak or outbreak state by observing the variation of index value.
Forecasting influenza A pandemic outbreak in pandemic occurrence place

Take China as an example. The flu broke out in Hong Kong in 1968 and continued until 1969, of which 7.5 million people died. In 1972, influenza broke out in Henan Province and quickly spread to the entire province. As shown in Table 2, the data in 1965 and 1967 are absent, so we cannot forecast. 1968 is 2.198217; 1969 is 1.217563; 1970 is 0.645643; 1971 is 0.756451. I1968 > I1969 > I1970 and I1971-I1970 > 0.1, so 1972 is in the outbreak state.

Many countries in Asia including China had the outbreak of influenza A in 1986. As shown in Table 2, 1982 is 0.566947; 1983 is 0.489632; 1984 is 0.467346; 1985 is 0.896537. I1982 > I1983 > I1984 and I1985-I1984 > 0.1, so 1986 is in the outbreak state.

The influenza A broke out in China in 2006 and 2009. As shown in Table 2, 2002 is 0.673956; 2003 is 0.653958; 2004 is 0.543824; 2005 is 0.854835; 2006 is 1.632657; 2007 is 0.549367; 2008 is 0.924375. I2002 > I2003 > I2004 and I2005-I2004 > 0.1, I2006 > I2007 and I2008-I2007 > 0.1, so 2006 and 2009 are in the outbreak state.

Conclusions

We select the data of protein amino acid sequence of pandemic influenza virus between 1934 and September 2016, and the different countries’ data such as China’s data between 1957 and 2015 in which only some data in a very few years are absent, and obtain a composite index by using PDNB. Although the amino acid sequence of each protein will randomly fluctuate, the composite index can still provide reliable, significant early warning information when influenza pandemic is close to the critical state or outbreak state. The network markers and other traditional markers cannot provide an early warning signal of the critical state before pandemic outbreak in comparison with dynamic network biomarker. This fully shows the dynamic network biomarker is more stable and accurate to determine the state in which the pandemic influenza virus, particularly the critical state of pandemic influenza. This will achieve the aim of early warning and then strengthen preventive measures in advance. This is of great significance for the research and warning of pandemic influenza virus.
Abbreviations
DNB: Dynamical network biomarkers; PDNB: Protein dynamical network biomarkers

Acknowledgments
We would like to thank Prof. Luonan Chen from SBS, China for his helpful guidance and comments to an earlier version of this manuscript. We would also like to sincerely thank the reviewers for their extremely helpful comments to this manuscript.

Funding
The work was supported by the National Natural Science Foundations of China (Grant No. 11271163 and No. 11371174). The publication costs were funded by the National Natural Science Foundations of China (Grant No. 11271163).

Availability of data and materials
The datasets analyzed during the current study are available in the [NCBI website] repository, [http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database].

About this supplement
This article has been published as part of BMC Systems Biology Volume 11 Supplement 4, 2017: Selected papers from the 10th International Conference on Systems Biology (ISB 2016). The full contents of the supplement are available online at https://bmcsysbiol.biomedcentral.com/articles-supplements/volume-11-supplement-4.

Authors’ contributions
JG conceived and coordinated the study, designed research methods, developed mathematical models and drafted the manuscript. KW carried out the statistical analysis, evaluated the mathematical models and participated in drafting the manuscript. TD and SZ helped to calculate and process data. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 21 September 2017

References
1. Chen LN, Liu R, Liu ZP, Li MY, Alhara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep. 2012;2:342–9.
2. Liu R, Li MY, Liu ZP, Wu JR, Chen LN, Alhara K. Identifying critical transitions and their leading biomolecular networks in complex diseases. Sci Rep. 2012;2:2813–21.
3. Pan YN, Lou JJ, Han XP. Outbreak patterns of the novel avian influenza (H7N9). Physica A: Stat Mech Appl. 2014;401:265–70.
4. Chang LY, Shih SR, Shao PL, et al. Novel swine-origin influenza virus a (H1N1) the first pandemic of the 21st century. J Formos Med Assoc. 2009;108(7):526–32.
5. Banerjee R, Roy A, Das S, et al. Similarity of currently circulating H1N1 virus with the 2009 pandemic clone: viability of an imminent pandemic. Infect Genet Evol. 2015;32:107–12.
6. He J, Hsieh MS, Juang RH, et al. A monoclonal antibody recognizes a highly conserved neutralizing epitope on hemagglutinin of H6N1 avian influenza virus. Vet Microbiol. 2014;174(3-4):333–41.
7. Liu R, Wang XD, Alhara K, Chen LN. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers and dynamical network biomarkers. Med Res Rev. 2014;34(3):455–78.
8. Wu DJ, Rice CM, Wang XD. Cancer bioinformatics: a new approach to systems clinical medicine. BMC Bioinformatics. 2012;13:71–80.
9. Jin GX, Zhou XB, Wang HH, Zhao H, Cui K, Zhang XS, Chen LN, Hazen SL, Li K, Wong STC. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. J Proteome Res. 2008;7(9):4013–21.
10. Simon R. Development and validation of therapeutically relevant multi-gene biomarker classifiers. J Natl Cancer Inst. 2005;97(12):866–7.
11. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer. 2005;5(11):845–56.
12. Li S, Depuy GW, Evanc GW. Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. Comput Oper Res. 2014;51:350–9.
13. Di R, Gao J. Early-warning signals for an outbreak of the influenza pandemic. Chin Phys B. 2011;20:1287011–4.
14. Girard MP, Tam JS, Assossou OM, Kieny MP. The 2009 a (H1N1) influenza virus pandemic: a review. Vaccine. 2010;28(31):4895–902.
15. Pardini M, Dash PK, Kumar JS, Joshi G, Tandel K, Sharma S, Srivastava A, Agarwall A, Saha A, Sarawat S. Emergence of influenza a(h1n1)pdm09 genogroup 6b and drug resistant virus. Eur Secur. 2016;21(5):6–11.