Saudi Journal of Biological Sciences (2017) **24**, 724–728

King Saud University

Saudi Journal of Biological Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Detecting early-warning signals for influenza A pandemic based on protein dynamical network biomarkers

Jie Gao [a,b,*], **Kang Wang** [a], **Tao Ding** [a]

[a] *School of Science, Jiangnan University, Wuxi 214122, China*
[b] *Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China*

Received 3 November 2016; revised 25 December 2016; accepted 6 January 2017
Available online 25 January 2017

KEYWORDS

Influenza A virus;
Early-warning signal;
Dynamical network biomarker (DNB);
The critical state;
The outbreak state

Abstract The outbreak of influenza A comes from a relatively stable state is a critical phenomenon on epidemic. In this paper, influenza A varying from different states is studied in the method of dynamical network biomarkers (DNB). Through studying DNB of influenza A virus protein, we can detect the warning signals of outbreak for influenza A and obtain a composite index. The composite index varies along with the state of pandemic influenza, which gives a clue showing the turn point of outbreak. The low value (<1) steady state of the composite index means influenza A is normally in the relatively steady stage. Meanwhile, if the composite index of a certain year increases by more than 0.8 relative to the previous year and it is less than 1 and it increases sharply and reaches a peak being larger than 1 in next year, it means the year is normal in the critical state before outbreak and the next year is normally in the outbreak state. Therefore, we can predict the outbreak of influenza A and identify the critical state before influenza A outbreak or outbreak state by observing the variation of index value.

1. Introduction

It is proved that there exists a kind of common critical phenomenon in many complex biological processes, i.e. a relative stable state enters into another state quickly after a soon critical point (Chen et al., 2012; Liu et al., 2012) 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 prevent and control the outbreak of influenza A pandemic timely and effectively, the key solution lies in predicting the critical point before the outbreak.

At present, influenza A is studied from many aspects. Ya-Nan 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

* Corresponding author at: School of Science, Jiangnan University, Wuxi 214122, China.
E-mail address: gaojie@jiangnan.edu.cn (J. Gao).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.sjbs.2017.01.048

(Pan et al., 2014; Zhang, 2016). Chang et al. (2009) studied the vaccine for influenza, so as to achieve the effect of prevention of influenza. Banerjee et al. (2015) 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. He et al. (2014) also made indepth studies to identify HA protein epitopes of avian influenza virus.

This paper studies the different states of influenza A using dynamical network biomarkers (DNB). Through studying DNB of influenza A virus protein, we can detect the warning signals of outbreak for influenza A and obtain a composite index. The composite index varies along with the state of pandemic influenza, which gives a clue showing the turn point of outbreak. Therefore, we can predict the outbreak of influenza A and identify the critical state before influenza A outbreak or outbreak state by observing the variation of index value. This indicates the composite index can provide reliable and significant warning information to detect the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic.

2. Method

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 (Liu et al., 2014; Wu et al., 2012). This concept refers to a series of markers as well as their mutual relations and has been proposed as a new marker type (Jin et al., 2008; Yao et al., 2015). Compared with traditional biomarkers, these markers can accurately distinguish disease states for taking the links between the molecules into consideration (Simon, 2005; Ludwig and Weinstein, 2005). However, it is used to diagnose the states of diseases, not for the detecting the critical point before the outbreak of diseases.

The method of dynamic network biomarkers focuses on the detection and assessment of different stages of the disease in the development of disease. This is a time-dependent method (Sun et al., 2014). It studies the location changes of the markers over time and the relationship among network markers over time changing. Meanwhile, this method can construct 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. The study of dynamic network markers focusing on dynamic changes, is helpful to discover the marker accurately, comprehensively, and further to distinguish the state of disease before outbreak. It does not only 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 wide 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.

3. Results

3.1. Data

Here are ten of proteins for influenza A virus hemagglutinin (HA), matrix protein, matrix protein 2, neuraminidase, nonstructural 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 1933 to 2015 from the NCBI website (www.ncbi.nlm.nih.gov/), whereas some data in 1937, 1938, 1939, 1940, 1941, 1942, 1944, 1951, 1952, 1953, 1954 and 1955 years are absent.

3.2. Model

3.2.1. Defining dynamic network biomarker

Taking HA protein as an example firstly, we suppose that a HA protein marked y is linked sequentially by t numbers of amino acids. Its amino acid sequence is represented by $y = x_1x_2\cdots x_t$, in which $x_i \in \{A, V, L, I, P, F, W, M, D, E, G, S, T, C, Y, N, Q, K, R, H\}$; $i = 1,2,\ldots,t$. 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 by $y_{s-1,1}$, $y_{s-1,2},\ldots,y_{s-1,m}$. 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 by $y_{s,1}$, $y_{s,2},\ldots,y_{s,n}$. The amino acid number of the $y_{i,j}$ is marked $c_{i,j}$, where $i = s-1,s$; $j = 1,2,\ldots,q$; $q = \max\{m,n\}$. Sequentially selecting the i-th amino acid for $y_{s-1,1},y_{s-1,2},\ldots,y_{s-1,m}$ to form a new amino acid sequence is defined $Z_{s-1,i}$, and then take out the largest one of amino acids number. 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, these amino acids are marked x_i, where $i = 1,2,\ldots,k$; $k = \max\{c_{s-1,1},c_{s-1,2},\ldots,c_{s-1,m}\}$. We individually connect them in order to form a new amino acid sequence ($U_{S-1} = x_1x_2\ldots x_k$) and then separately compare with corresponding amino acids of $y_{s,1},y_{s,2},\ldots,y_{s,n}$ one by one. If they are different, the assignment is one, on the contrary the assignment is zero. Therefore, n new sequences are represented by $E_{s,1},E_{s,2},\ldots,E_{s,n}$ are 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:

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

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

$$CV_s = \frac{SD_S}{M_s} \tag{3}$$

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 values of CV_s are the top three are defined as core protein (CP), and the others are no-core protein (NP). CP is a set of high confidence interactions of proteins, which forms a sub-network called influenza A virus proteins of the protein dynamical network biomarkers.

3.2.2. The early warning model for influenza A

We calculate the frequencies of the 20 kinds of amino acids, and the computation formulas are as follows:

$$f_{x_i}(s) = \frac{\sum_{j=1}^{n} f_{x_i}(s,j)}{n} \tag{4}$$