Infection, Genetics and Evolution 32 (2015) 107–112

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Similarity of currently circulating H1N1 virus with the 2009 pandemic clone: Viability of an imminent pandemic

Rachana Banerjee [a], Ayan Roy [b], Santasabuj Das [c], Surajit Basak [d,e,*]

[a] Department of Bio-Physics, Molecular Biology and Bioinformatics, University of Calcutta, 92, Acharya Prafulla Chandra Road, Kolkata 700009, India
[b] NBU Bioinformatics Facility, Department of Botany, University of North Bengal, Raja Rammohanpur, Siliguri 734013, India
[c] Biomedical Informatics Center, National Institute of Cholera and Enteric Diseases, P-33, C.I.T Road, Scheme-XM, Beliaghata, Kolkata 700010, India
[d] Department of Molecular Biology & Bioinformatics, Tripura University, Suryamaninagar, Tripura 799 022, India
[e] Bioinformatics Centre, Tripura University, Suryamaninagar, Tripura 799 022, India

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form 22 February 2015
Accepted 23 February 2015
Available online 28 February 2015

Keywords:
H1N1
Hemagglutinin
Pandemic
Pathogenicity
Evolution

ABSTRACT

The first influenza pandemic in the 21st century commenced in March, 2009 causing nearly 300,000 deaths globally within the first year of the pandemic. In late 2013 and in early 2014, there was gradual increase in the reported case of H1N1 infection and according to World Health Organization (WHO) report, influenza activity increased in several areas of the Southern Hemisphere and was dominated by the H1N1 pandemic strain of 2009. In the present study, a comprehensive comparison of the global amino acid composition and the structural features of all HA gene sequences of H1N1, available in the Flu Database (NCBI), from 1918 to December, 2014 has been performed to trace out the possibility of a further H1N1 pandemic in near future. The results suggest that the increased potential to enhance pathogenicity for the H1N1 samples of 2013 (latter part) and 2014 could lead to a more severe outbreak in the near future.

1. Introduction

The first influenza pandemic of the present century, commonly known as "Mexican flu", commenced in March 2009 with the emergence of a new swine-origin (triple-reassorted) H1N1 influenza A virus that has been remarkably different from the previously circulating H1N1 strains (Dawood et al., 2009). Clinical symptoms associated with the 2009 pandemic H1N1 strain (H1N1 pdm09) include mild respiratory irritations that extend up to severe pneumonia, associated with acute respiratory distress syndrome (ARDS) with the gradual progress of infection (Chowell et al., 2009; Perez-Padilla et al., 2009; Go et al., 2012). The H1N1 pandemic that initiated in 2009 resulted in nearly 300,000 deaths globally within the first year of its outbreak (Dawood et al., 2012). A distinctive feature of influenza virus infection is a successful viral entry into the host system followed by subsequent destruction of the host immune responses and inception of disease (Go et al., 2012; de Jong et al., 2006; Kobasa et al., 2007). Continuous antigenic variations occur in influenza A virus that ensure its competent replication inside the human host (Schmolke and Garcia-Sastre, 2010). H1N1 subtype of influenza A virus has been found to be display variations at the amino acid level which allow them to evade the host immune signals and establish the disease with absolute efficacy. H1N1 pdm09 influenza A virus is a classic example supporting the fact that the antigenic variations correlate strongly with host immune elusion and crucially govern the efficacy of the virus in pronouncing the disease. The origin of human-isolated H1N1 virus has been well explored from the various phylogenetic analyses however, proper know-how of the determinants of cross-species transmission still remain somewhat obscure (Liu et al., 2014; Lin et al., 2009; Landolt and Olsen, 2007; Banerjee et al., 2012).

Hemagglutinin (HA) has been confirmed to be one of the major glycoproteins present on the surface of influenza virus that facilitate proper viral attachment to the host cellular receptors (Guarnaccia et al., 2013; Caton et al., 1982; Gerhard et al., 1981). Successful attachment of HA with host sialic-acid receptors plays a significant role to pave the way for the onset of infection of the respiratory epithelial cells. HA serves to be the primary antigen of the influenza virus and is also efficient in evading the host immune response. Interestingly, mutations causing antigenic drift are typically restricted to the antigenic sites adjacent to the receptor binding site on the globular head of HA (Yewdell and Gerhard, 1981; Both et al., 1983; Skehel et al., 1984). Therefore, it is always

* Corresponding author at: Department of Molecular Biology & Bioinformatics, Tripura University, Suryamaninagar, Tripura 799 022, India. Tel.: +91 9862924152.
E-mail address: basaksurajit@gmail.com (S. Basak).

http://dx.doi.org/10.1016/j.meegid.2015.02.023
1567-1348/

essential to thoroughly scrutinize the properties of the HA protein in H1N1 virus, as it can provide good evidence for elucidating the complex infective mechanisms of the virus. Mutations that involve amino acid changes in antigenic regions of influenza proteins assist proficient immune escape and are frequently termed as antigenic drifts. Such mutations commonly occurs with the genes encoding the HA surface glycoprotein. Thus, HA proteins appear to be the major targets of neutralizing antibodies that circulate as a result of vaccination (Guarnaccia et al., 2013). Several reports have been published in the past one year, highlighting the unique mutations of the HA gene and their impact on the structure and function of the protein, especially the human receptor binding affinity (Wu et al., 2014; Łepek et al., 2014; Linderman et al., 2014).

The new swine-origin H1N1pdm09 (the pandemic strain of H1N1 of 2009) virus was first isolated in April, 2009, from patients with febrile respiratory illnesses in the United States and Mexico and it spread rapidly across the world by human-to-human transmission. Later, the World Health Organization declared the 2009 H1N1 infection a global pandemic. The rapid spread of this swine influenza virus, mainly among young healthy adults and outside the classical influenza season added to the unpredictability of this virus. Thus, the virus and its molecular evolution raised a number of questions, which are of prime international public health concern. After the first reported case of infection with the H1N1pdm09 virus, the virus was found to be transmitted in different countries. The epidemiology of H1N1pdm09 virus in the United Kingdom during 2009–2011 was characterized by 3 distinct waves: first wave, April–August 2009; second wave, September 2009–April 2010; and third wave, August 2010–April 2011 (Sridhar et al., 2013). In the post pandemic period, the virus remained in circulation and showed unusually increased activity and severity in March and April, for three consecutive years, which is certainly unseasonal (Dakhave et al., 2013). Since November 2013, the Public Health Agency of Canada has received a number of reports of illness caused by the influenza A H1N1 flu virus among young and middle-aged adults (http://www.phac-aspc.gc.ca/influenza/ah1n1-eng.php). The most alarming scenario was reported for the 2013–2014 season when CDC (Centers for Disease Control and Prevention) received reports of severe flu illness among young and middle-aged adults, many of whom were infected with the H1N1pdm09 virus (http://www.cdc.gov/washington/fluBrief/CDCInfluenza_E-Brief.pdf). It has also been detected by CDC's influenza surveillance system that most of the patients affected with flu-infection were aged between 18 and 64 years. Notably, similar records were evident among non-elderly adults during the 2009 H1N1 pandemic (http://www.cdc.gov/flu/pastseasons/1314season.htm). In the present study, a comprehensive comparison of the global amino acid composition of all HA gene sequences of H1N1 available in the Flu Database (NCBI) from 1918 to December, 2014 has been performed to trace out significant features pointing towards a further probable H1N1 pandemic in the near future.

2. Methods and calculations

All the gene sequences of haemagglutinin (HA) of the influenza A (H1N1) viruses isolated until October, 2014 were downloaded from the NCBI Influenza Virus Resource database (http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html). Single linkage cluster analysis on amino acid usage of all HA genes was performed using STATISTICA (version 6.0, published by Statsoft Inc., USA). Since amino acid usage is multivariate in nature, therefore, it is necessary to analyze this data with multivariate statistical techniques i.e., Correspondence analysis (Greenacre, 1984; Sabbía et al., 2007; Basak et al., 2004; Banerjee et al., 2004; Basak and Ghosh, 2006). Major trend in amino acid usage variation among the HA genes was predicted by Correspondence analysis available in CodonW 1.4.2 (Peden, 2000; http://www.molbiol.ox.ac.uk/cu/).

In order to assess the extent of divergence between different groups or clusters, we used Mahalanobis distance, which is well known in multivariate statistical analysis. Mahalanobis distance (D^2) between two clusters is calculated as:

$$D^2 = (X - \bar{Y})^T S^{-1} (X - \bar{Y})$$

where X is a vector of amino acid usage values for the data points in the Cluster 2, Y is a mean vector of amino acid usage values calculated from the data points in the Cluster 1, S is the variance–covariance matrix of the amino acid usage calculated from the data points in the Cluster 1 (S^{-1} is the inverse matrix of S), and the superscript T is the transposition operator. To test if the group means significantly differ between the clusters, we used Hotelling's T-square distribution that is applied in multivariate statistics in undertaking tests of differences between the multivariate means of different populations. Let vector d be the difference between sample means, n_1 and n_2 the sample sizes, p the number of variables and S the variance–covariance matrix. Then, the test statistic is

$$T^2 = \frac{(n_1 + n_2 - p - 1) n_1 n_2}{p(n_1 + n_2 - 2)(n_1 + n_2)} d' S^{-1} d$$

A transformation of T^2 yields an exact F distribution, so that:

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 \sim F_{p, n_1 + n_2 - p - 1}$$

This can be evaluated on p and $(N - p - 1)$ degrees of freedom, where p is the number of dependent variables and $N = n_1 + n_2$. Therefore, F can be evaluated in terms of statistical significance by computing the p-values when F and the two degrees of freedom are given.

We have constructed homology based structure of seven HA proteins taking one representative from each of the six groups and one HA protein of the year 2013 from group 6, using the crystal structure of HA protein of 2009 H1N1 influenza virus (PDB ID: 3LZG) as template. We also created the heavy chain of the antibody 2D1 computationally using the 2D1 from crystal structure of Fab 2D1 in complex with the 1918 Influenza Virus Hemagglutinin (PDB ID: 3LZF) as template. The interaction between the HA protein and the heavy chain of the antibody 2D1 and energy minimization of the docked structures were carried out using Accelrys Discovery Studio software (version 2.0).

3. Results and discussion

3.1. Similarity in RAAU of HA genes from 2009 and 2014

All HA gene sequences from 1918 to December, 2014 were retrieved from GenBank that counted a total of 4671 and Relative Amino Acid Usage (RAAU) was calculated. Single linkage clustering on RAAU identified six major clusters (Fig. 1). It is noteworthy that all the HA gene sequences isolated from various seasonal outbreaks from the year 1933 to 2009 have been found to be present under Cluster 1, which has been revealed to be the root of the six clusters as denoted by the single linkage distances. Fig. 1 also indicates the year wise variation of amino acid composition of HA genes starting from 1918 and successfully detects very similar patterns of amino acid composition as reflected from the close association between Cluster 3 and Cluster 6. In other words, very similar patterns of amino acid composition of HA genes exists for 2009 (April to December) and June, 2013–2014. When the actual values of RAAU were compared between HApdm09 (HA gene of pandemic H1N1 strain of 2009) and HA genes collectively of 2013 (June to