Journal of the Formosan Medical Association (2015) 114, 456–463

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

ORIGINAL ARTICLE

Influenza A virus plasticity—A temporal analysis of species-associated genomic signatures

Guang-Wu Chen [a,b,*], Yu-Nong Gong [a], Shin-Ru Shih [b,c]

[a] *Department of Computer Science and Information Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan*
[b] *Research Center for Emerging Viral Infections, College of Medicine, Chang Gung University, Taoyuan, Taiwan*
[c] *Department of Medical Biotechnology and Laboratory Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan*

Received 2 October 2014; received in revised form 29 January 2015; accepted 30 January 2015

KEYWORDS
amino acid residue;
genome analysis;
host adaptation;
influenza A virus;
species-specific signatures

Background/purpose: An influenza A pandemic occurred in 2009–2010. A novel H1N1 virus (hereafter H1N1pdm) was responsible for this outbreak. H1N1pdm viruses have been largely seen in recent human influenza A viruses. This virus was descended from a triple-reassorted swine virus consisting of human, avian, and swine origins. As a result, the previously established species-associated signatures could be in jeopardy.
Methods: We analyzed all influenza A sequences in the past 5 years after the inclusion of H1N1pdm into human viruses since 2009, and examined how human signatures may lose their distinctness by mixing with avian residues that H1N1pdm have brought in. In particular, we compared how those signatures were changed/shifted in the past 5 years for human-isolated avian influenza A viruses and discussed their implications.
Results: Only eight out of 47 signatures remained human-like for human influenza A viruses in the past 5 years. They are PB2 271A; PB1 336I; PA 356R and 409N; NP 33I, 305K, and 357K; and NS1 227R. Although most avian-like residues were preserved in human-isolated avian influenza A viruses, a number of them were found to have become or on the verge of becoming human-like, including PB2 627, PA 100, 356, 404, 409, NP 33, 61, 305, 357, M2 20, and NS1 81.
Conclusion: Analyzing how species-associated signatures are becoming human-like in human-isolated avian influenza A viruses helps in assessing their potential to go pandemic as well as providing insights into host adaptation.

Conflicts of interest: The authors have no conflicts of interest relevant to this article.

* Corresponding author. Department of Computer Science and Information Engineering, Chang Gung University, 259 Wen-Hua 1st Road, Guishan District, Taoyuan City 333, Taiwan.
E-mail address: gwchen@mail.cgu.edu.tw (G.-W. Chen).

http://dx.doi.org/10.1016/j.jfma.2015.01.015

Introduction

Influenza A virus can infect various hosts including birds, swine, and some mammals. A number of pandemics occurred in humans, including the 1918 H1N1 (Spanish flu), 1957 H2N2 (Asian flu), and 1968 H3N2 (Hong Kong flu). Casualties were catastrophic in these global infections. Recent endemics of highly pathogenic avian influenza A viruses also caused human deaths. Such seemingly sporadic and regional infections constantly raise concern that an avian influenza A virus could mutate to become easily transmissible between humans, raising the possibility for another influenza pandemic. Notable examples include the H5N1 virus since 2003, and the H7N9 virus since 2013.[1–3] Other reported human cases of avian influenza A virus also include H7N7 and H9N2.[4,5]

Recently, a single substitution from glutamic acid (E) to lysine (K) at PB2 627 of avian influenza viruses was found to enhance their replications in mammalian cells.[6,7] Observing that the distribution of PB2 K versus E was each dominant in human and avian influenza A viruses, we developed an entropy-based method that outlined a number of amino acid positions, wherein each contains a distinct residue in either avian or human viruses. Fifty-two signatures were revealed based on large-scale sequence analysis of 306 human and 95 avian virus genomes in 2006.[8] We further revalidated them into 47 signatures in 2009 based on more than 3000 genomes each for human and avian influenza A viruses.[9] These signatures provide not only markers of diagnosing if a novel influenza A virus can be attributed to either a human or avian virus, but also molecular targets for scientists to probe host adaptability.[10,11] Note that an entropy-based algorithm heavily depends on a good multiple sequence alignment. As a result, the two surface proteins, HA and NA, have been excluded from the discussion, specifically their species-specific signatures, because too many gaps would have been introduced from aligning these diverse sequences among subtypes.[8]

A novel influenza A H1N1 pandemic (hereafter H1N1pdm) occurred in early 2009 and caused severe morbidity and fatality.[12,13] Not only was this the first human pandemic in recent years, the new virus also featured multiple genes from different origins, and some of these genes were previously reassorted from their more distant ancestors, including birds.[14,15] For example, the two polymerase genes PB2 and PA were traced back to the avian influenza A viruses of North American lineage in the late 1970s. They were cointroduced together with human-origin genes of PB1, HA, and NA (from human H3N2 dated back to as early as 1968) into a swine population in around the early 1990s. This virus population was named a triple-reassorted swine influenza A virus, with its third origin consisting of swine H1N1 genes of NP, M, and NS from the classical swine virus of Eurasia and North America. It later acquired a new HA and NA gene also from the classical swine virus in Eurasia and North America and was detected as a novel swine reassortant in late 1990s. Avian HA and M genes from the North American H1N1 avian virus then replaced the two genes in this swine virus, and the resulting swine virus was detected with sporadic human infections in the late 1990s into early the 2000s. Finally, an outbreak of new human H1N1commenced in 2009 based on this reassorted, multiple-origin genome.

Regardless of the multiple origins of this new influenza A virus, H1N1pdm has been considered a human virus since its debut in 2009, primarily because of its extensive infection and transmission capability in humans. After the 2009 outbreak, the H1N1pdm literally replaced the old H1N1 (previously known as the seasonal H1N1) viruses and became a human seasonal influenza virus just like H3N2. Viewing that many H1N1pdm virus genes were descended from avian viruses (some directly, and some indirectly via swine species), we wondered if some of the 47 human—avian genomic signatures that we established earlier would be in jeopardy. In this study, we analyzed all influenza A sequences in the past 5 years after the inclusion of H1N1pdm into human viruses, and examined how human signatures may lose their distinctness by mixing with avian residues that H1N1pdm has brought in. In particular, we compared how those signatures were changed/shifted in the past 5 years for human-isolated avian influenza A viruses and discuss their implications.

Materials and methods

Influenza virus sequences

Influenza A virus sequences were downloaded from Influenza Virus Resource as of April 2014.[16] Only full-length amino acid sequences were gathered, and identical sequences in each dataset were collapsed by keeping only the oldest one in the group. All sequences were grouped into two temporal phases—Phase 1 of viruses isolated before December 2008 and Phase 2 of viruses isolated in the most recent five seasons from January 2009 to April 2014. In each phase, we further divided the viruses into three categories: avian-isolated avian viruses, human-isolated avian viruses, and human-isolated human viruses.

Template-based multiple sequence alignments

Sequences were multiply aligned using an in-house-developed algorithm called template-based multiple sequence alignment (TMSA). It takes advantage of the fact that the sequences to be aligned are genetically close, so that no gaps (or very limited ones) are inserted. In short, TMSA outlines a reference sequence called R; based on it, every sequence to be aligned (T_i, where $i = 1, n$) is compared using a pairwise alignment program *needle* in the European Molecular Biology Open Software Suite package.[17] All resulting pairwise alignments are then assembled into the multiple sequence alignment (MSA) and curated for any anomalies that are not compatible with the coordinate systems in documenting the 47 signature positions. The time complexity is reduced from $O(n^2)$ to $O(n)$ by using TMSA rather than some traditional MSA tools such as ClustalW.[18]

Sequence logos for signature positions

We used WebLogo (version 3.4, http://weblogo.threeplusone.com/) to graph the sequence logos of