Journal of Microbiology, Immunology and Infection (2016) 49, 477–486

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jmii.com

ORIGINAL ARTICLE

Analysis of codon usage preference in hemagglutinin genes of the swine-origin influenza A (H1N1) virus

Sheng-Fan Wang [a,b,c,d,j], Ming-Wei Su [c,e,j], Sung-Pin Tseng [a], Ming-Chun Li [f], Ching-Han Tsao [c,g], Szu-Wei Huang [d,h], Woei-Chyn Chu [e], Wu-Tse Liu [g], Yi-Ming Arthur Chen [d,i,**], Jason C. Huang [b,g,*]

[a] *Department of Medical Laboratory Science and Biotechnology, Kaohsiung Medical University, Kaohsiung, Taiwan*
[b] *AIDS Prevention and Research Center, National Yang-Ming University, Taipei, Taiwan*
[c] *Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan*
[d] *Center for Infectious Disease and Cancer Research, Kaohsiung Medical University, Kaohsiung, Taiwan*
[e] *Institute of Biomedical Engineering, National Yang-Ming University, Taipei, Taiwan*
[f] *Department of Pediatrics, Taipei City Hospital, Yang-Ming Branch, Taipei, Taiwan*
[g] *Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, Taipei, Taiwan*
[h] *Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan*
[i] *Department of Microbiology, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan*

Received 16 April 2014; received in revised form 14 August 2014; accepted 23 August 2014
Available online 31 October 2014

KEYWORDS
antigenic sites;
phylogenetic tree;
S-OIV;
synonymous codon usage

Background: The swine-origin influenza A (H1N1) virus (S-OIV) has come to the forefront since 2009 and was identified as a new reassortant strain. The hemagglutinin (HA) glycoprotein mediates virus binding, contains antigenic regions recognized by neutralizing antibodies, and is associated with viral cross-species infection and adaption. The comparison study of codon usage preferences in influenza viral genomes was less extensive. In this study, we used codon usage pattern analyses to validate the adaption and origins of S-OIV.

** Corresponding author. Yi-Ming Arthur Chen, Department of Microbiology, College of Medicine, Kaohsiung Medical University, Number 100, Shih-Chuan 1st Road, San Ming District, 80708 Kaohsiung, Taiwan.
* Corresponding author. Jason C. Huang, Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, Number 155, Li-Nong Street, Section 2, Taipei, Taiwan.
E-mail addresses: arthur@kmu.edu.tw (Y.-M.A. Chen), jchuang2@ym.edu.tw (J.C. Huang).

[j] These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jmii.2014.08.011

Methods: Codon usage pattern was used to estimate the host adaption of S-OIVs. Phylogenetic analysis of the *HA* gene was conducted to understand the phylogeny of H1N1 viruses isolated from different hosts. Amino acid signature pattern on antigenic sites of HA was analyzed to understand the antigenic characteristics.
Results: Results of phylogenetic analyses of *HA* gene indicate that S-OIVs group in identical clusters. The synonymous codon usage pattern analyses indicate that the effective number of codons versus GC content at the third codon position in the *HA1* gene slightly differ from those in swine H1N1 and gradually adapted to human. Our data indicate that S-OIV evolution occurred according to positive selection within these antigenic regions. A comparison of antigenic site amino acids reveals similar signature patterns between S-OIV and 1918 human influenza strains.
Conclusion: This study proposes a new and effective way to gain a better understanding of the features of the S-OIV genome and evolutionary processes based on the codon usage pattern. It is useful to trace influenza viral origins and cross-species virus transmission.

Introduction

Influenza A viruses have the potential to trigger epidemics or pandemics.[1] Glycosylated oligosaccharides, the major cell receptors of influenza viruses, terminate with α-2,3 or α-2,6-linked sialic acid (SA) residues.[2] Host range determination is not completely clear, although the receptor specificity of hemagglutinin (HA) is considered an important determinant.[3] The HA is one of the determinates for cross-species virus transmission.[4] The receptor binding specificity of influenza viruses can be changed by HA protein mutations. Previous reports indicate that: (1) the general receptor specificity of different HA subtypes of influenza A viruses is maintained by conserved amino acids in positions 190, 194, 225, 226, and 228 (H3 numbering); and (2) only two mutations (D190 E and D225 G) are required to convert the HA of 1918 strains from classic α-2,6 receptor specificity to avian α-2,3 specificity.[3,5]

Pigs with both receptors in their respiratory tract cells have been called "mixing vessels" for generating new influenza viruses.[2,6] In mid-April 2009, a novel swine-origin influenza A H1N1 virus (S-OIV) was identified in the United States and Mexico, and swine flu cases were soon reported throughout North America and in Europe. S-OIV was identified as a reassortant strain with six gene segments (*PB2*, *PB1*, *PA*, *HA*, *NP*, and *NS*) from triple-reassortant influenza A viruses circulating in North American pigs. Two other gene segments (*NA* and *M*) have their origin in Eurasian swine influenza viruses.[7,8] Previous studies indicate that S-OIVs show higher virulence and cause systemic infections in nonhuman primates.[9] The S-OIV HA and NA proteins are primary targets for neutralizing antibodies during influenza infections. The major immunogenic domain, HA1, contains five major regions involved in antigenic drift: Sa, Sb, Ca1, Ca2, and Cb in H1N1, and A, B, C, D, and E in H3N2. It is thought that more severe epidemics are caused by antigenic variants.[10–13]

Codon usage has been focused on understanding the general cause of codon choices, and the preference of codon usage correlated to genome evolution.[14] Codon usage preference patterns indicate the bias shown in different organisms and genes in the codon choices among a synonymous group of codons that all code the same amino acid.[15] Previous studies indicate a consistency in bacteria between codon choice and highly expressed genes with stronger selective preferences.[16,17] Specific codon usage preferences in different species were subsequently identified in many other organisms, and closely related organisms have more similar patterns of codon usage. For example, the codon usage preferences in *Salmonella typhimurium* closely resemble those in *Escherichia coli* when these two bacteria infect a human.[18] The diverse codon usage preferences may be the result of translation selection as the populations of isoaccepting transfer RNA (tRNA) contents vary in different organisms and tissues.[19] To date, the comparison of codon usage preferences in viral genomes has been investigated less extensively. Viruses complete their life cycle intracellularly, so they have to exploit and coevolve with host molecular mechanisms. Reports have indicated that the expression level of papillomavirus capsid protein depends on the match between the codon usage and tRNA availability in the host cells.[20] A study of human immunodeficiency virus (HIV)-1 codon usage pattern indicated that the HIV-1 within a host changes its codon usage patterns to more closely resemble human codon usage patterns during a period of infection.[21]

However, analysis of codon usage preferences among a number of species is complicated. Because there are 64 codons for 20 amino acids, there are a vast number of genes in a single species. Analyzing synonymous codon usage patterns is a common method for examining species origin.[22] Compositional constraints (mutation and natural selection) are thought to be two primary factors triggering codon usage variation in different organisms.[21,23] However, codon usage patterns are not randomly selected; some codons are used more frequently than others in different species or reservoirs, or in different genes within the same genome.[22] Relative synonymous codon usage (RSCU) values may be virus specific and unaffected by translational selection or gene length.[21,24] Codon usage pattern analyses are therefore helpful for comparing differences among large numbers of virus strains and for determining predominant driving forces during virus evolution.