Dear Editor

We analyze proteins of virus H1N1 a virus responsible for swine flu; it was observed that year of evolution of H1N1 is predictable. Following is description of our work.

Virus H1N1, a subtype of Influenza A virus has put the world on high alert leading to the "2009 H1N1 pandemic". According to the World Health Organization (www.who.int/csr/don/2009_06_15), published on June 15, 2009, the virus had spread to 76 countries, with 35, 928 reported cases including 163 deaths (Huang et al., 2009)**.** The H1N1 strain of influenza is a single stranded RNA virus composed of a segmented genome originated from various influenza viruses (Solovyov et al., 2009). The existing vaccines for Influenza A might not provide effective cross-protection due to difference in Hemagglutinin (HA) and Neuraminidase (NA) protein sequences of current swine H1N1 and conventional influenza vaccine strains (CDC 2009). There are two surface proteins, HA and NA of the virus against which majority of the cellular immune response mounted upon infection. There are total 16 serotypes of the HA and 9 for NA, based on which Influenza A broadly classified. HA in particular has a high amino acid substitution rate in its epitope regions leading to the evasion from the host immune system, termed as "antiogen drift" in literature (Hay et al., 2001). Given earlier experience with SARS and avian flu pandemic (Vogel 2006), global health surveillance mechanism provided sequences of new flu strains in public sequence databases.

There are 11 proteins in the pandemic H1N1 virus and sequences from more than 65 different locations are available with information on type, sampling year etc. Based on the availability of the sequences from 1933 to 2009 for all 11 proteins make it interesting and feasible to analyze the variations in the sequences. There have been many significant efforts in the past by the scientific community in this direction to aid in the design of better vaccine for current and future H1N1 strains. However these sequence analysis was limited to a particular year (De Groot et al., 2009) or particular protein NA (Maurer-Stroh et al., 2009). Heiny et al in 2007 did a comprehensive study on all proteins of Influenza A virus covered major subtypes including H1N1 leading to the selection of conserved sequences for the synthesis of a supertype epitope-based genetic vaccine. Moreover in the current situation the data for the H1N1 for all 11 proteins has increased tremendously from 7301 in 2007 to 11968 in 2009 due to recent pandemic and synergistic global efforts.

Based on the availability of the sequence data for H1N1 since 1933, in the present study we asked a simple question as how amino acid composition of sequences varies due to course of time. Is there any

correlation between amino acid composition of sequences and the corresponding year of sampling? We need to know the variation behavior of the virus before the next possible pandemic.

The question arises whether mutations in H1N1 proteins are random or foreseeable. In order to address this question, we analyzed 1168 neuraminidase (NA) proteins (Tamiflu acts as NA inhibitor) of H1N1 sampled between 1933 to 2009. It was observed that composition of certain type of residues (like Asp, Gly, Phe, Met, Asn, Arg) shows high correlation with year of sampling of NA protein. This means overall composition of residues is not random instead it is increasing or decreasing with time. We developed a Support Vector Machine (SVM) based model for predicting year of evolution using amino acid composition of protein NA and achieve very high correlation 0.99 between predicted and actual year of sampling. This is surprising that year of evolution of H1N1 is predictable from amino acid composition of NA protein. We extend our study to other proteins of H1N1 and observed similar trend. It was observed that year of evolution is also predictable for strain H3N2 using amino acid composition of its proteins. We also extend our study to another virus HIV and develop SVM based model for predicting year of evolution of HIV from amino acid composition of Gag protein; a low correlation 0.61 was achieved between predicted and actual year of sampling of Gag protein of HIV. Our study demonstrates that year of evolution is predictable in case of strains of influenza A but not not applicable for HIV virus. Influenza A virus has unique property, in which composition of a protein evolved in a given time period have unique composition as well as composition is unidirectional (either increasing or decreasing) over the year. This is interesting observation and can be used to predict composition of proteins to be evolved in future. Based on our study, we developed a web server Sfepred for predicting year of evolution virus H1N1, which is available from url http://www.imtech.res.in/raghava/sfepred/.

REFERENCES

CDC; Centers for Disease Control and Prevention (2009) Serum cross-reactive antibody response to a novel influenza A (H1N1) virus after vaccination with seasonal influenza vaccine. MMWR Morb Mortal Wkly Rep 58: 521-524.

De Groot AS, Ardito M, McClaine EM, Moise L, Martin WD (2009) Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008-2009 conventional influenza vaccine. Vaccine 27: 5740-5747.

Hay AJ, Gregory V, Douglas AR, Lin YP (2001) The evolution of human influenza viruses. Philos Trans R Soc Lond B Biol Sci 356: 1861-1870.

Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, et al. (2007) Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets. PLoS One 2: e1190.

Huang T, Cui W, He ZS, Hu L, Liu F, et al. (2009) Functional association between influenza A (H1N1) virus and human. Biochem Biophys Res Commun. 390(4):1111-3.

Maurer-Stroh S, Ma J, Lee RT, Sirota FL, Eisenhaber F (2009) Mapping the sequence mutations of the 2009 H1N1 influenza A virus neuraminidase relative to drug and antibody binding sites. Biol Direct 4: 18; discussion 18.

Solovyov A, Palacios G, Briese T, Lipkin WI, Rabadan R (2009) Cluster analysis of the origins of the new influenza A(H1N1) virus. Euro Surveill 14.

Vogel G (2006) Public health. SARS and bird flu veteran to take WHO helm. Science 314: 1067.