

Analysis of HA nucleotide sequence of influenza A virus, and its predictive simulation

T14M002 Kazuya Shibata

Adviser Qing Ma

Graduate Course of Applied Mathematics and Informatics

Graduate School of Science and Technology

Ryukoku University

Abstract

Influenza virus is a zoonotic pathogen of the influenza diseases. It has caused great damage to humans and animals. Especially, new influenza virus appeared every several decades causes the flu pandemic. Seasonal influenza is also caused by the influenza virus which changes own antigenicity to escape human immunity.

In this study, we have analyzed the hemagglutinin(HA) nucleotide sequence of influenza A/H3N2 virus that causes seasonal influenza infectious disease in humans. We have investigated the HA nucleotide sequences of influenza virus strains collected from 1968 to 2015 using multivariate analysis techniques. We have also tried to predict the interannual variation of the virus strains by the analytical results.

At first, we have described the phylogenetic trees from genetic relationship of HA nucleotide sequence of A/H3N2 viruses, and confirmed a specific structure of A/H3N2 virus has a long stem and many short branches. Secondly, we have calculated p-distance between the sequence of virus strains, and obtained a representation of the strains in low-dimensional Euclidean space by multidimensional scaling method. The distribution of virus strains in the low-dimensional space is similar to the structure of its phylogenetic tree. Additionally, we have applied Lam's Binary encoding scheme to the A/H3N2 virus strains, and obtained a similar result in another low-dimensional representation.

Furthermore, we have proposed an interannual variation model and a prediction algorithm of the virus strains. For interannual prediction of A/H3N2 virus strains, we have calculated the number of minimum nucleotide substitutions of HA sequence and the variation frequency on each site from the annual virus strains dataset. We have tried to predict the annual virus variation of A/H3N2 by our prediction algorithm through the use of calculated parameters. Our prediction results have been similar in distribution to the actual isolated strains.