BRIEF COMMUNICATION

Using combinatorial bioinformatics methods to analyze annual perspective changes of influenza viruses and to accelerate development of effective vaccines

Yu-Jen Hu, Kuan-Chih Chow, Ching-Chuan Liu, Li-Jen Lin, Sheng-Cheng Wang, Shulhn-Der Wang

a Department of Mathematical Sciences, National Chung Hsing University, Taichung, Taiwan
b Graduate Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan
c Department of Pediatrics, College of Medicine, National Cheng Kung University and Hospital, Tainan, Taiwan
d School of Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung, Taiwan
e National Taichung Wen Hua Senior High, Taichung, Taiwan
f School of Post-Baccalaureate Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung, Taiwan

Received 23 September 2014; received in revised form 24 March 2015; accepted 25 March 2015

KEYWORDS
biological sequences; phylogenetic tree; vaccine development

The standard World Health Organization procedure for vaccine development has provided a guideline for influenza viruses, but no systematic operational model. We recently designed a systemic analysis method to evaluate annual perspective sequence changes of influenza virus strains. We applied dnaml of PHYLIP 3.69, developed by Joseph Felsenstein of Washington University, and ClustalX2, developed by Larkin et al, for calculating, comparing, and localizing the most plausible vaccine epitopes. This study identified the changes in biological sequences and associated alignment alterations, which would ultimately affect epitope structures, as well as the plausible hidden features to search for the most conserved and effective epitopes for...
Introduction

Influenza virus contains hemagglutinin (HA) and neuraminidase (NA), which were coded as H and N strands, respectively. These H and N strands represent different strains of the virus. Generally, immune response loci of HA and NA were used for designing vaccines. The standard World Health Organization procedure for vaccine development has provided a guideline for influenza viruses, but no systematic operational model. Our design can help scientists shorten the time of blind tests and be a better method for producing vaccines against unknown strains of influenza virus based on previously known flu vaccines.

Our methods could rapidly identify similarity among strains and localize the closest strains. When a new strain of virus is confirmed for disease manifestation, we can instantaneously find the closest pre-made vaccine by whole virus based on previously known flu vaccines. Our methods could rapidly identify similarity among strains and localize the closest strains. When a new strain of virus is confirmed for disease manifestation, we can instantaneously find the closest pre-made vaccine by whole virus based on previously known flu vaccines. However, the error in prediction of influenza virus mutant is inevitable. For example, the lack of a vaccine development. Addition our newly designed systemic analysis method to supplement the WHO guidelines could accelerate the development of urgently needed vaccines that might concurrently combat several strains of viruses within a shorter period.

Materials and methods

Analysis programs

In this study, we applied the unweighted pair group method with arithmetic mean (UPGMA), maximum likelihood estimate (MLE),11 and N–W dynamic programming12 to evaluate evolutionary changes and relationship between viruses in polymerase basic 2 (PB2) and NA, respectively. Two evolution analysis software, dnaml of PHYLIP 3.69 and ClustalX2, were used to show the phylogenetic pedigrees among viruses based on the results of MLE and UPGMA, respectively.

Sequences of influenza viruses

The genetic sequences of H1N1 (GenBank: GQ132185.1), H1N2 (GenBank: AY129157.1), H2N2 (GenBank: DQ508843.1), H3N2 (GenBank: U71142.1), H3N8 (GenBank: KF790582.1), H5N1 (GenBank: AF509095.1), H5N2 (GenBank: AY300929.1), H5N3 (GenBank: AY207510.1), H7N1 (GenBank: AY207538.1), H7N2 (GenBank: CY037097.2), H7N3 (GenBank: CY076858.1), H7N7 (GenBank: CY133723.1), H7N9 (GenBank: KF420298.1), and H10N7 (GenBank: CY157240.1) had been collected from the databases of the National Center for Biotechnology Information.

Analysis of immune loci

We applied ClustalX2 and dynamic programming to analyze the different bases (immune loci) between those 14 NA sequences.

Results

To analyze the relationships among various influenza viruses, we collected genetic sequences of H1N1, H1N2, H2N2, H3N2, H3N8, H5N1, H5N2, H5N3, H7N1, H7N2, H7N3, H7N7, and H10N7 from the databases of National Center for Biotechnology Information. Moreover, an influenza virus contains eight nucleic acid fragments that encode 11 proteins, namely PB2, PB2-F1, PB1, PA, HA, NP, NA, M1, M2, NS1, and NS2. The sequence of PB2 reveals the evolutionary origin of a virus and has often been used in phylogenetic analysis. Therefore, the MLE method was used to determine phylogenetic pedigrees of PB2 sequences among viruses. Results showed that H5N3 had the most evolutionary relationship with H7N3, and there was also high similarity between H7N7, H7N9, and H7N2 (Fig. 1A).

The NA gene of an influenza virus encodes a surface protein that is essential for releasing progeny virions from the infected cell surface and prevent self-aggregation. Thus, the NA has been targeted to design antiviral drugs. In fact, the NA inhibitor is one of the two major classes of antivirals licensed and effectively limits the severity of viral infections. Therefore, we further investigated the NA section of the 14 viruses by UPGMA. In Figs. 1B and 1C, we found that PB2 analysis could be extrapolated to those for NA.

Annual production of influenza virus vaccines was based on historical records used to predict the future prevalence of the disease. However, the error in prediction of influenza virus mutant is inevitable. For example, the lack of a
Figure 1  (A) Rootless tree diagram analysis of PB2 of influenza viruses. NA (UPGMA) (B) rooted (B) or (C) rootless tree diagram of 14 viruses prior to using N–W dynamic programming to all-sequence comparison. NA (UPGMA) (D) rooted or (E) rootless tree diagram of 14 viruses after using N–W dynamic programming to all-sequence comparison. NA = neuraminidase; N–W = Needleman–Wunsch; UPGMA = unweighted pair group method with arithmetic mean.
Figure 2  NA (MLE) (A) rooted or (B) rootless tree diagram of 14 viruses after using N–W dynamic programming to all-sequence comparison. (C and D) NA all-sequence comparison and immune loci analysis of 14 viruses. MLE = maximum likelihood estimate; NA = neuraminidase; N–W = Needleman–Wunsch.
framework for modeling nucleotide insertions/deletions together with nucleotide substitutions would also lead to an error in phylogenetic analysis. To avoid such conditions, multiple alignments of NA were constructed based on the complete nucleotide sequences, using the N–W dynamic programming and then UPGMA for precise calculation and cross-examination (Figs. 1D and 1E).

To reduce an arithmetic error in the calculation of phylogenetic diversity, we also applied N–W dynamic programming and MLE to evaluate the relationships between 14 viruses in NA sequences (Figs. 2A and 2B). The result showed that there was no difference between UPGMA and MLE calculations (Figs. 1D and 2A, and Figs. 1E and 2B, respectively).

To find the basic difference between those 14 influenza viruses for the determination of immune loci, we further conducted a continuous analysis of NA sequences by N–W dynamic programming and MLE. We used H7N9 and H5N3 sequences. The total length of H7N9 sequence was 1426 base pairs and that of H5N3 was 1422 base pairs. After calculation, H7N9 was found to have 623 loci of immune response, and 43.32% of blind tests could be reduced. The number of loci of immune response of H5N3 was 191, and reduction in blind tests was 13.55% (Figs. 2C and 2D).

Discussion

Previously available biomedical information is used for vaccine development; however, in this case, the major problems would be to predict the correct future viral strains that can cause widespread infection and the possible viral mutations that may influence the presentation of effective epitopes to provoke protective immune responses. Projection of the best epitopes is particularly important for influenza viruses, which from time to time become lethal to human beings. Moreover, because of the distinctive features of RNA viruses, persistent gene mutations as well as possible genome recombination and reassortments are inevitable, which by contrast, could generate more deadly strains to cross species barrier, leading to more widespread manifestation.

In this study, we have developed a novel method by combining UPGMA, MLE, and N–W dynamic programming to define the annually detected mutational changes. We used these annual results to determine mutational ranges and the most probable mutation sites in the gene, which could then be used to project by using computer analysis the coming unknown virus which would be the major target for production of the vaccine.

In this way, there are many uncertainty products of vaccines which could be eliminated. The derivation of statistical outcomes is shown in Figs. 1D–2D. We could not only identify the relationship as well as subtle differences between known viruses, but also pinpoint the most plausible immune loci for biochemical experiments to raise effective antibodies (Figs. 2C and 2D). In addition, our results could be further extrapolated to the drug design; for instance, the current design of antiviral drugs was focused primarily on NA, such as oseltamivir and zanamivir, both of which are NA inhibitors. However, our results have shown that HAs can be the next potential target, similar to NAs.

Moreover, our previous study had been awarded by the 2013 International Biochemical Technology Invention. The patent rights of this new methods for vaccine development are under application via Intellectual Property Rights (IPR) Center of National Chung Hsing University, Taichung, Taiwan (patent numbers: 102134134 and 103117788) and the USA (patent number: 14/482,506). This study could effectively improve the current reversed genetics procedure of vaccine development (Figs. 2C and 2D). As the best epitopes to provoke appropriate immunity could be identified faster, 3–6 weeks of developing time could be saved. The cost of vaccine development can also be reduced markedly. It also helps medical policy making and provides more protections for peoples.

References

2. Shors T. Understanding viruses. 2nd ed. Massachusetts, USA: Jones and Bartlett Learning; 2013.