Published online 8 May 2009 *Nucleic Acids Research, 2009, Vol. 37, Web Server issue* ***W643–W646***
doi:10.1093/nar/gkp321

ATIVS: analytical tool for influenza virus surveillance

Yu-Chieh Liao[1], Chin-Yu Ko[1], Ming-Hsin Tsai[1], Min-Shi Lee[2,*] and Chao A. Hsiung[1,*]

[1]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences and [2]Vaccine R&D Center, National Health Research Institutes, Zhunan 350, Miaoli County, Taiwan

Received January 30, 2009; Revised April 9, 2009; Accepted April 20, 2009

ABSTRACT

The WHO Global Influenza Surveillance Network has routinely performed genetic and antigenic analyses of human influenza viruses to monitor influenza activity. Although these analyses provide supporting data for the selection of vaccine strains, it seems desirable to have user-friendly tools to visualize the antigenic evolution of influenza viruses for the purpose of surveillance. To meet this need, we have developed a web server, ATIVS (Analytical Tool for Influenza Virus Surveillance), for analyzing serological data of all influenza viruses and hemagglutinin sequence data of human influenza A/H3N2 viruses so as to generate antigenic maps for influenza surveillance and vaccine strain selection. Functionalities are described and examples are provided to illustrate its usefulness and performance. The ATIVS web server is available at http://influenza.nhri.org.tw/ATIVS/.

INTRODUCTION

Influenza viruses cause substantial medical and social burdens worldwide. Vaccination is the primary method to prevent influenza and its complications. Hemagglutinin (HA) of influenza viruses is the main surface protein inducing protective antibody responses. The HA protein is synthesized as a single polypeptide (HA0), which is subsequently cleaved into two polypeptides, HA1 and HA2, and forms into homotrimers. The HA1 mutates more frequently than the HA2 and plays a major role in the process of natural selection (1,2). Accumulation of point mutations on the HA result in antigenic drift, so that antibody raised in response to one virus may have reduced effectiveness against a drifted virus. Since 1977, influenza A/H1N1, A/H3N2 and B viruses have been circulating globally, and thus current vaccines are usually trivalent, containing these three strains.

In order to tackle the seasonal epidemics of influenza, the WHO Global Influenza Surveillance Network was established in 1952 (http://www.who.int/csr/disease/influenza/surveillance/). The collaborative centres in the network perform antigenic and genetic analyses of viral isolates regularly. Antigenic characterization of influenza viruses is based on hemagglutinin-inhibition (HI) tests using ferret antisera. Cross-reactive HI titers among reference antisera and circulating viruses are summarized into tables. To utilize the information in these tables sensibly, Smith *et al*. (2) presented a quantitative relationship for mapping antigenic evolution of influenza viruses (cartography), but the cartography software does not seem to be available for public use. Another effort to monitor the changes in the surface antigens of influenza viruses starts with analyzing sequence data in viral RNA genes containing the antigenic regions (the HA1 domain). Examining the relationship between the sequence changes and the antigenic differences of these viruses, Lee *et al*. (3) and Liao *et al*. (4) identified some potential immunodominant positions and proposed statistical models for predicting antigentic variants of influenza A/H3N2 viruses on the basis of sequence information.

In this study, we present a web server ATIVS (Analytical Tool for Influenza Virus Surveillance) to analyze serological data of influenza viruses and provide easily interpretive summaries; to compare the HA1 sequences of viruses with those of reference vaccine strains to predict influenza A/H3N2 antigenic distances; based on the serological data and the predicted antigenic distances, to generate antigenic maps, which further demonstrate the antigenic evolution of influenza viruses and facilitate the selection of vaccine strains.

OVERVIEW

ATIVS has functionalities including analyzing serology data for all influenza subtypes and HA1 sequence data for influenza A/H3N2 viruses. After inputting a serological table, antigenicity analysis is performed to show two supporting figures along with an easily interpretive table. Subsequently, an antigenic map can be generated after parameter setting. When the serological data are not available and only amino-acid sequence data are available, antigenic prediction models are implemented to predict antigenic distances based on the differences of

*To whom correspondence should be addressed. Tel: +886 37 246 166. Ext. 36100; Fax: +886 37 586 467; Email: hsiung@nhri.org.tw
Correspondence may also be addressed to Min-Shi Lee. Tel: +886 37 246 166. Ext. 37742; Fax: +886 37 583 009; Email: minshi.lee@hotmail.com

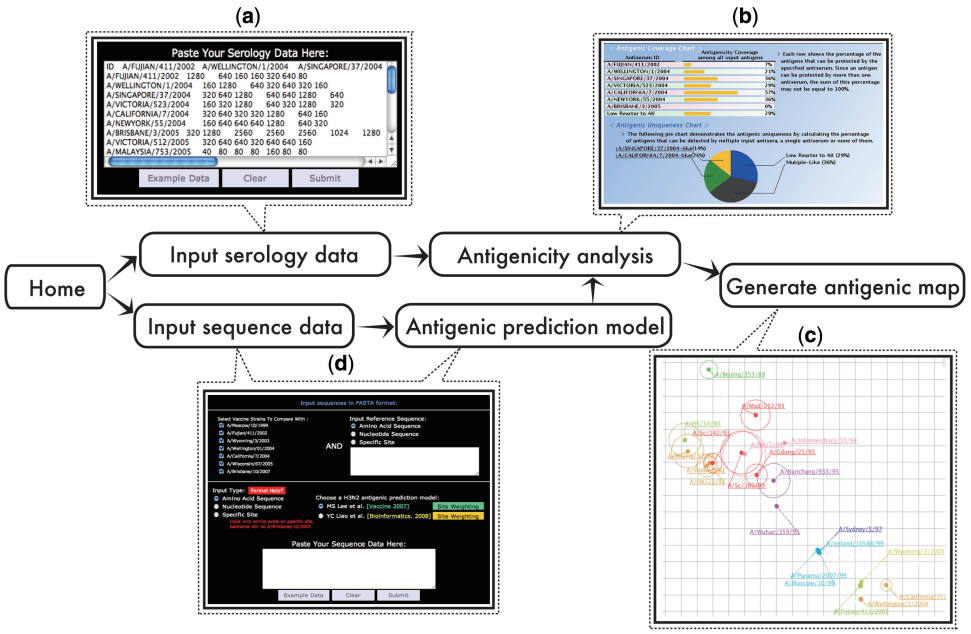

Figure 1. Overall flowchart of ATIVS. (**a**) Example Data in the Serology Data. (**b**) Two supporting figures are performed by antigenicity analysis based on the example in (a). (**c**) Antigenic map generated by the Example 3. (**d**) Screenshot of sequence data input.

amino-acid sequences; after that, the antigenicity analysis is performed and an antigenic map is constructed as previously described. The flowchart of ATIVS is shown in Figure 1. Notably, the subfigures presented in the Figure 1 were generated from different example data; details can be found in the figure legend.

FUNCTIONALITIES

Serological data analysis

Given a serological table that describes cross-reactive titers between antigens of circulating and reference strains and antisera of reference strains, we calculate ratios between homologous and heterologous antibody titers (Figure 1a); ratios of antibody titers <4 are defined as similar antigenicity between the reference and circulating strains and ratios ≥ 4 are defined as low-reactors or antigen variants to the reference strains (3,4). Based on the ratios of antibody titers in the serological table, we would know if any reference strains are similar to the circulating strains (3,4). From these, an easily interpretive table is summarized and downloadable. Besides, two supporting figures are provided as bar and pie charts to summarize the antigenic relationship (Figure 1b). The bar chart is to determine which reference strains provide optimal antigenic coverage rate against the circulating strains. Alternatively, the pie chart demonstrates the antigenicity distribution of the circulating strains that can be covered by the reference strains (single, multiple or low-reactors to all). Take the example in the serology data input (as shown in Figure 1a) for illustration. The antigenic coverage rate of the antiserum of A/California/7/2004 against the circulating viruses is about 57% (Figure 1b), with 21% solely reacted to the antiserum of A/California/7/2004 and 36% reacted to multiple antisera (Figure 1b). Based on this kind of serological table including recent circulating viruses versus contemporary vaccine strains and potential reference strains, the two supporting figures provided by antigenicity analysis help us to determine which reference strains can be selected as vaccine candidates. The viruses with both high antigenic coverage rate and antigenic uniqueness are selected as vaccine strain candidates.

Subsequently, the serological data can be used to construct a 2D antigenic map (2), as shown in Figure 1c. The construction procedure is detailed in the following subsection on sequence data analysis. One unit of antigenic distance on an antigenic map corresponds to a 2-fold difference in the serological assay. Smith *et al.* (2) use the periphery of each shape to denote an increase in the error; here we use an outer circle to express the error and an antigen point with larger circle indicates larger uncertainty of its location. We note that the map can present antigenic evolution only when the reference antisera points are widely spread. In the case of human influenza A/H3N2 viruses, Russell *et al.* (5) mentioned that 'periods of relative stasis lasting from 3 to 8 years were followed by rapid antigenic changes', which suggests that the antisera data in a serological table should cover a period of more than a decade in order to present antigenic evolution. Since the serological data covering broad antigenicity may not be always available from a single laboratory, one of the illustrations on antigenic evolution in the following Example section uses combined serological data. One of the main purposes of the Example section is to show how antigenicity evolves based on serological data. Note that the statement 'serology data should cover more than a decade to present antigenic evolution' is only for human influenza A/H3N2 viruses, not for other viruses with higher evolutionary rates.

Although all the example data we used in the web server are human influenza A/H3N2 viruses, the serological data analysis, in fact, can be applied to other influenza viruses and animal viruses that can drift antigenically.

Sequence data analysis

In case antigenic distances can not be reliably provided by serological data alone, we use sequence data to obtain predicted antigenic distances. Based on the relationship between the genetic differences and the antigenic distances, we proposed antigenic prediction models, which have good agreement rates, for predicting antigenic variants (3,4). These antigenic prediction models are now incorporated into ATIVS. Sequence data analysis in ATIVS allows one to input nucleotide or amino-acid sequences of influenza A/H3N2 HA1 and to compare these sequences against WHO-recommended vaccine strains or other reference strains (as shown in Figure 1d). In addition to upload nucleotide or amino-acid sequences in FASTA format, ATIVS provides a specific site mode where sequence sites and their corresponding amino-acid residues are required for inputting data. Furthermore, user-defined colors encoded in RGB color model can also be specified in this mode. An additional column with the title of COLOR is used to set up the user-defined colors. Using the amino-acid differences between input strains and reference strains obtained by ClustalW and the above model, ATIVS automatically provides the predicted antigenic distances, which are shown in a downloadable table. Then, the antigenicity analysis can be performed as described before.

For antigen i and antiserum j, let D_{ij} be the log2 of the predicted antigenic distances of the input virus strain i (antigen) to reference strain j (antiserum), and d_{ij} denote the Euclidean distance between them in the map, which are obtained by minimizing the error function $E = \sum_{ij} e(D_{ij},d_{ij})$, where $e(D_{ij},d_{ij}) = (D_{ij} - d_{ij})^2$ when $D_{ij} <$ threshold (default is 8); $e(D_{ij},d_{ij}) = (D_{ij} - d_{ij})^2 \times g(D_{ij} - d_{ij})$, $g(x) = 1/(1 + e^{-10x})$ when $D_{ij} \geq$ threshold.

The equations are modified from the error function in Smith *et al.* (2). The antigenic map is then obtained by applying the multidimensional scaling algorithm to d_{ij}. Since sequence data of viruses are widely available, this approach to generate antigenic map may facilitate efficient surveillance of influenza viruses.

In sequence data analysis, ATIVS allows users to upload the maximum of 500 sequences and the results can be obtained in a couple of minutes. However, since the antigenic prediction models are only for human influenza A/H3N2 viruses, at least 50% sequence similarity to HA1 is required for sequence data input.

EXAMPLES

We have created examples that demonstrate the functionalities of the web server. Three examples are included for the illustration of the use of serological data, and one example is for the use of sequence data. Example 1 exhibits an ordinary HI table, in which the contemporary vaccine strains and potential reference strains are compared with the recent circulating viruses. Using the two supporting figures, we know immediately which reference strains are suitable vaccine candidates. Example 2 uses an HI table of selected successive strains over 25 years, which shows how antigenicity evolves over a long period from 1968 to 1997. In Example 3, we combine five datasets, obtained at different times, to form the HI table (the detailed sources are shown in the website). The antigenic map obtained from the combined data has a high consistency with the map that was shown in the study of Berkhoff *et al.* except a spinning of 45° (6). As shown in Figure 1c, the antigenic map demonstrates that the viruses drifted from A/Beijing/353/89 to A/Beijing/32/92, to A/Wuhan/353/95 and A/Sydney/5/97, to A/Fujian/411/2002 then to A/California/7/2004, which is concordant with the influenza epidemics (the high resolution figure can be found in ATIVS). The two HI tables in Example 2 and Example 3, including over-decade reference strains, can be used to generate antigenic maps which show continuous antigenic drift over a long period.

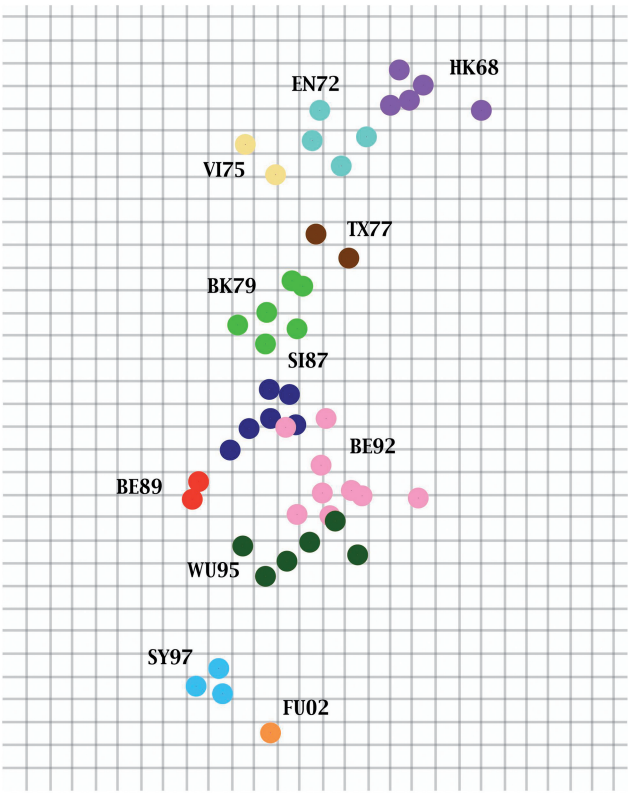

Figure 2. Antigenic map of influenza A/H3N2 viruses from 1968 to 2003 generated by ATIVS. The abbreviations and colors of the clusters follow the definitions of Smith *et al.* (2).

Example 4 uses sequence data to generate antigenic map. Two hundred and fifty-three sequences of specific sites were extracted from the Supplementary Data of Smith *et al.* (2). For each of the 11 clusters identified by Smith *et al.*, we randomly select one to three sequences based on the size of the cluster. Twenty-five sequences were accordingly selected and used as reference strains. These sequence data are available in our website. After sequence alignment with A/Brisbane/10/2007,

antigenic distances between each of the 253 strains and each of the 25 reference strains are predicted and an antigenic map is subsequently generated by the server. This antigenic map (Figure 2) is highly consistent with the Smith's map, which shows the robustness of our method.

CONCLUSIONS

ATIVS is a java-based web server built on Linux. Both serology data of all influenza viruses and HA1 sequence data of human influenza A/H3N2 viruses can be utilized to generate antigenic maps that are useful in influenza virus surveillance and vaccine strain selection.

ACKNOWLEDGEMENTS

The authors thank Julie Chih-yu Chen and Dr I-Shou Chang for help revising the web server, and Guan-Yuan Liou, Dr Feng-Chi Chen and Dr Ben-Yang Liao for stimulating discussion.

FUNDING

Centers for Disease Control, Taiwan (DOH95-DC-1401); National Research Program for Genomic Medicine, National Science Council, Taiwan (NSC95-3112-B-400-005). Funding for open access charge: National Health Research Institutes, Taiwan.

Conflict of interest statement. None declared.

REFERENCES

1. Lee,M.S. and Chen,J.S. (2004) Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg. Infect. Dis.*, **10**, 1385–1390.
2. Smith,D.J., Lapedes,A.S., de Jong,J.C., Bestebroer,T.M., Rimmelzwaan,G.F., Osterhaus,A.D. and Fouchier,R.A. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science*, **305**, 371–376.
3. Lee,M.S., Chen,M.C., Liao,Y.C. and Hsiung,C.A. (2007) Identifying potential immunodominant positions and predicting antigenic variants of influenza A/H3N2 viruses. *Vaccine*, **25**, 8133–8139.
4. Liao,Y.C., Lee,M.S., Ko,C.Y. and Hsiung,C.A. (2008) Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. *Bioinformatics*, **24**, 505–512.
5. Russell,C.A., Jones,T.C., Barr,I.G., Cox,N.J., Garten,R.J., Gregory,V., Gust,I.D., Hampson,A.W., Hay,A.J., Hurt,A.C. *et al.* (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science*, **320**, 340–346.
6. Berkhoff,E.G., Geelhoed-Mieras,M.M., Jonges,M., Smith,D.J., Fouchier,R.A., Osterhaus,A.D. and Rimmelzwaan,G.F. (2007) An amino-acid substitution in the influenza A virus hemagglutinin associated with escape from recognition by human virus-specific CD4 + T-cells. *Virus Res.*, **126**, 282–287.