Selecting vaccine strains for H3N2 human influenza A virus

CrossMark

Yoshiyuki Suzuki *

Graduate School of Natural Sciences, Nagoya City University, Nagoya, Japan

ARTICLE INFO

ABSTRACT

H3N2 human influenza A virus causes epidemics of influenza mainly in the winter season in temperate regions. Since the antigenicity of this virus evolves rapidly, several attempts have been made to predict the major amino acid sequence of hemagglutinin 1 (HA1) in the target season of vaccination. However, the usefulness of predicted sequence was unclear because its relationship to the antigenicity was unknown. Here the antigenic model for estimating the degree of antigenic difference (antigenic distance) between amino acid sequences of HA1 was integrated into the process of selecting vaccine strains for H3N2 human influenza A virus. When the effectiveness of a potential vaccine strain for a target season was evaluated retrospectively using the average antigenic distance between the strain and the epidemic viruses sampled in the target season, the most effective vaccine strain was identified mostly in the season one year before the target season (pre-target season). Effectiveness of actual vaccines appeared to be lower than that of the strains randomly chosen in the pre-target season on average. It was recommended to replace the vaccine strain for every target season with the strain having the smallest average antigenic distance to the others in the pre-target season. The procedure of selecting vaccine strains for future epidemic seasons described in the present study was implemented in the influenza virus forecasting system (INFLUCAST) (http://www.nsc.nagoya-cu.ac.jp/~yossuzuk/influcast.html).

  * Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya-shi, Aichi-ken 467-8501, Japan. Tel./fax: +81 52 872 5821.
    E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp.

Introduction

H3N2 influenza A virus appeared and caused a pandemic in the human population in 1968 (Coleman et al., 1968). Subsequently, this virus has been circulating in humans causing epidemics every winter season in temperate regions (Nelson and Holmes, 2007; World Health Organization, 2009). Vaccines are available for prophylaxis against human influenza viruses (Henle et al., 1943; Francis et al., 1944). The main target of vaccination is hemagglutinin (HA), which is the envelope protein most abundantly exposed on the virion surface (Mitnaul et al., 1996). HA is a homotrimeric type I transmembrane protein glycosylated at the *N*-linked glycosylation sites (NGS), and the HA1 domain of HA contains the sialic-acid receptor binding sites (RBS) and the neutralization epitopes (Skehel and Wiley, 2000).

The antigenicity of human influenza viruses evolves rapidly due to the high mutation rate, the short generation time, the host immune responses, and the relatively weak functional constraint operating on surface amino acids of HA1 (Suzuki and Nei, 2002; Smith et al., 2004; Suzuki, 2006, 2013a; Bedford et al., 2012). Consequently, vaccines are re-formulated biannually by the Global Influenza Surveillance and Response System of the World Health Organization under the assumption that the dominant strain in a year may dominate in the next year (Treanor, 2004; Russell et al., 2008a). However, this assumption does not always hold and the antigenicity of vaccines sometimes fails to match that of epidemic viruses (Centers for Disease Control and Prevention, 2004, 2010).

To improve the effectiveness of vaccines, several attempts have been made to predict the major amino acid sequence of HA1 in the target season of vaccination (Bush et al., 1999; Xia et al., 2009; He and Deem, 2010; Ito et al., 2011; Suzuki, 2013b; Luksza and Lassig, 2014). However, the precision of prediction was limited by the random genetic drift, the incomplete sampling of epidemic viruses, the occurrence of *de novo* mutations, and the lack of knowledge on the mechanism of sequence evolution. In addition, the usefulness of predicted sequence was unclear because its relationship to the antigenicity was unknown.

The antigenicity of influenza viruses may be determined by the physical and chemical properties of surface amino acids in HA1 (Hensley et al., 2009; Lees et al., 2010), as well as their relative positions to RBS (Whittle et al., 2011) and NGS (Kobayashi and Suzuki, 2012b). Several methods have been developed for estimating the degree of antigenic difference (antigenic distance) between amino acid sequences of HA1 (Lee and Chen, 2004; Gupta et al., 2006; Lee et al., 2007; Liao et al., 2008; Huang et al., 2009; Lees et al., 2010; Suzuki, 2013b). The purpose of the present study was to integrate the information on antigenicity into the process of selecting vaccine strains for H3N2 human influenza A virus.

Materials and methods

Influenza seasons

H3N2 human influenza A virus causes epidemics of influenza mainly in the winter season in temperate regions. Therefore, in the present study, the influenza seasons are defined such that season 2000.0 indicates the period between October of year 1999 and March of year 2000, corresponding to the winter season in the northern hemisphere, and season 2000.5 indicates the period between April and September of year 2000, corresponding to the winter season in the southern hemisphere (Luksza and Lassig, 2014). The recommended composition of vaccine strains for human influenza viruses has been publicized such that the vaccine strains for season 2001.0 were announced in February of year 2000, and those for season 2001.5 in September of year 2000 (World Health Organization, 2014a,b). In the present study, the vaccine strain for target season 2001.0 was attempted to be identified from the strains isolated up to the end of season 2000.0, and the vaccine strain for target season 2001.5 from the strains isolated up to the end of season 2000.5, and so forth. Here, seasons 2000.0 and 2000.5 are called the pre-target season for target seasons 2001.0 and 2001.5, respectively.

Antigenic distance

Difference in antigenicity between influenza virus strains can be quantified using the hemagglutination inhibition (HI) titer. HI titer t_{ij} is defined as the magnitude of the maximum dilution of anti-serum raised against strain i that inhibits hemagglutination of strain j. It is considered that vaccination by strain i effectively

prevents infection by strain j if $t_{ij}/t_{ii} \leq 4$ (antigenic escape threshold). The antigenic distance between strains i and j (d_{ij}) is defined as $d_{ij} = \frac{1}{2} \log_2 \left(\frac{t_{ij}}{t_{ii}} \cdot \frac{t_{ji}}{t_{jj}} \right)$ (Lees et al., 2010). Models have been developed for estimating the antigenic distance from the difference in amino acid sequences of HA1 (Lee and Chen, 2004; Gupta et al., 2006; Lee et al., 2007; Liao et al., 2008; Huang et al., 2009; Lees et al., 2010; Suzuki, 2013b).

In the model of Suzuki (2013b), d_{ij} is estimated by

$$\hat{d}_{ij} = \sum_s \left\{ \left(\left| vol_{s(i)} - vol_{s(j)} \right| w_{\text{vol}} + \left| pI_{s(i)} - pI_{s(j)} \right| w_{\text{pI}} \right) \cdot f\left(d_{RBS_s} \right) \cdot g\left(\frac{d_{NGS_{s(i)}} + d_{NGS_{s(j)}}}{2} \right) \cdot \delta_s \right\},$$

where s denotes the amino acid site, vol and pI denote the volume (Grantham, 1974) and the isoelectric point (Nelson and Cox, 2008) of amino acid representing the physical and chemical properties, respectively, and w_{vol} and w_{pI} denote the weights for them. The d_{RBS} and d_{NGS} indicate the Euclidean distances of C_α atom in amino acid from the closest C_α atom in RBS and NGS, respectively, in the three-dimensional structure of HA trimer for A/Hong Kong/19/1968 (Protein Data Bank [PDB] ID: 2HMG). The functions $f(x)$ and $g(y)$ are given by $f(x) = e^{-kx}$ $(k > 0)$ and $g(y) = 1 - e^{-ly}$ $(l > 0)$. The δ is 1 or 0 according to whether the relative solvent accessibility of amino acid is > 0 or 0, respectively, which is obtained by computer program ASAVIEW (Ahmad et al., 2004) using 2HMG.

A total of 1673 antigenic distances experimentally determined between the strains of H3N2 human influenza A virus, for which the nucleotide sequences encoding HA1 were available in the Influenza Virus Resource at the National Center for Biotechnology Information (Bao et al., 2008), were retrieved from the literature (Supplementary Table S1). After eliminating the distances associated with the sequences containing minor gaps, ambiguous nucleotides, or premature termination codons and averaging the distances for the same pairs of strains, 382 distances between 88 strains were available for optimizing model parameters w_{vol}, w_{pI}, k, and l (Supplementary Table S2).

In the present study, the antigenic model was integrated into the process of selecting vaccine strains for H3N2 human influenza A virus. To imitate the real situation of selecting vaccine strains, the model parameters were re-optimized for each target season using only the antigenic distances between the strains that were isolated up to the end of pre-target season. Optimization was performed by minimizing the square sum of residual errors in the estimates of antigenic distances using the genetic algorithm (Tomita et al., 2000). Three sets of random real numbers were assigned as the initial parameter values to confirm the convergence of the results (Suzuki, 2013b).

Sequence data

Effectiveness of a potential vaccine strain for a target season of H3N2 human influenza A virus may be evaluated retrospectively using the average antigenic distance between the strain and the epidemic viruses circulating in the target season; strains with smaller average antigenic distances may be considered as more effective (Luksza and Lassig, 2014). To obtain reliable estimates for the average antigenic distance, it is required to sample a relatively large number of epidemic viruses in the target season, so that they could reflect the composition of epidemic viruses in nature.

Nucleotide sequences encoding HA1 of H3N2 human influenza A virus without minor gaps, ambiguous nucleotides, or premature termination codons, for which the information on the time (year and month) and the place (country) of isolation was available, were retrieved from the Influenza Virus Resource (Bao et al., 2008) on July 25, 2014. The sequences were classified according to the season and the continent (Africa, Asia, Europe, North America, Oceania, or South America) of isolation, and the identical sequences classified in the same category were collapsed into one (Luksza and Lassig, 2014). Since the sample size was relatively large (≥ 5) for the seasons of 1992.0 and later, 6297 sequences were regarded as representatives of epidemic viruses in seasons 1992.0 to 2014.5 (Supplementary Table S3).

Fitness model

A model for predicting the composition of epidemic viruses in season $t + 1$ from that observed in the sample obtained in season t has been proposed for H3N2 human influenza A virus (Luksza and Lassig,

Table 1
Average antigenic distances between actual vaccines or potential vaccine strains and epidemic viruses sampled in target seasons.

Target season	Actual vaccine	Antigenic distance	Most effective vaccine strain[a]	Antigenic distance	Center-of-mass strain	Antigenic distance	Strain selected by fitness model	Antigenic distance
2001.0	A/Moscow/10/1999	1.487	A/Malaysia/12974/1999	0.687	A/Nelson Marlborough/1/2000	1.221	N.A.	N.A.
2001.5	A/Moscow/10/1999	1.600	A/Malaysia/12974/1999	0.767				
			A/Hong Kong/CUHK28038/2000	0.822	A/Western Australia/5/2000	0.824	N.A.	N.A.
2002.0	A/Moscow/10/1999	1.613	A/Netherlands/118/2001	0.621	A/Western Australia/9/2000	1.147	N.A.	N.A.
2002.5	A/Moscow/10/1999	1.726	A/Queensland/20/2001	0.658	A/West Coast/55/2001	1.144	N.A.	N.A.
2003.0	A/Moscow/10/1999	2.491	A/Perth/201/2002	1.112	A/New York/22/2002	1.790	N.A.	N.A.
2003.5	A/Moscow/10/1999	2.943	A/Hunan/407/2002	0.565	A/Western Australia/31/2002	2.355	N.A.	N.A.
2004.0	A/Moscow/10/1999	3.170	A/Sydney/7/2003	0.411	A/Sydney/7/2003	0.411	N.A.	N.A.
2004.5	A/Fujian/411/2002	0.892	A/Chanthaburi/219/2003	0.374	A/Uruguay/UY-PA03-4/2003	0.629	N.A.	N.A.
2005.0	A/Fujian/411/2002	1.022	A/Taiwan/296/2004	0.387	A/Texas/TX-AC24395/2003	0.783	N.A.	N.A.
2005.5	A/Wellington/1/2004	0.894	A/Jeonnam/336/2004	0.502				
			A/California/CA-S71051/2004	0.588	A/Whanganui/386/2004	0.594	N.A.	N.A.
2006.0	A/California/7/2004	1.196	A/Tianjin/275/2005	0.434	A/Western Australia/63/2004	0.730	N.A.	N.A.
2006.5	A/California/7/2004	1.372	A/Argentina/AG-R570-05/2005	0.496	A/Argentina/AG-R570-05/2005	0.496	N.A.	N.A.
2007.0	A/Wisconsin/67/2005	1.350	A/Western Australia/71/2005	0.688	A/Western Australia/71/2005	0.688	N.A.	N.A.
2007.5	A/Wisconsin/67/2005	1.396	A/Connecticut/CT-5204-7547/2006	0.355	A/Victoria/503/2006	0.733	N.A.	N.A.
2008.0	A/Wisconsin/67/2005	1.511	A/Wisconsin/03/2007	0.351	A/New York/UR06-0510/2007	0.862	N.A.	N.A.
2008.5	A/Brisbane/10/2007	0.994	A/LangSon/LS218/2007	0.181	A/Perth/142/2007	0.253	N.A.	N.A.
2009.0	A/Brisbane/10/2007	1.100	A/Wisconsin/09/2008	0.258	A/Wisconsin/09/2008	0.258	N.A.	N.A.
2009.5	A/Brisbane/10/2007	2.259	A/Wisconsin/13/2008	1.400	A/Wisconsin/13/2008	1.400	N.A.	N.A.
2010.0	A/Brisbane/10/2007	2.038	A/Texas/WRAIR1239P/2009	0.961	A/Wisconsin/05/2009	1.179	N.A.	N.A.
2010.5	A/Perth/16/2009	1.166	A/Australia/30/2009	0.512	A/Hawaii/14/2009	1.002	N.A.	N.A.
2011.0	A/Perth/16/2009	1.317	A/Pennsylvania/02/2010	0.636	A/Florida/26/2009	0.697	A/Singapore/C2009.803bV/2009	1.235
2011.5	A/Perth/16/2009	1.485	A/Peru/PER379/2010	0.787	A/Peru/PER379/2010	0.787	A/Wyoming/04/2010	0.819
2012.0	A/Perth/16/2009	1.356	A/Peru/PER315/2010	0.576	A/Hyogo/10 K304/2011	0.898	A/Sao Paulo/14805/2011	0.905
2012.5		1.482		0.702		0.702		0.967

Table 1 (continued)

Target season	Actual vaccine	Antigenic distance	Most effective vaccine strain[a]	Antigenic distance	Center-of-mass strain	Antigenic distance	Strain selected by fitness model	Antigenic distance
	A/Perth/16/2009		A/Peru/PER337/2011		A/Peru/PER337/2011		A/Peru/PER397/2011	
2013.0	A/Victoria/361/2011	1.404	A/Wisconsin/07/2012	0.553	A/Peru/PER358/2012	0.892	A/California/2972/2012	1.098
2013.5	A/Victoria/361/2011	1.486	A/Alabama/17/2012	0.359	A/Peru/PER331/2012	0.982	A/Peru/PER353/2012	1.732
2014.0	A/Victoria/361/2011	1.710	A/Wyoming/03/2013	0.455	A/Sao Paulo/11079/2013	0.615	A/Colorado/20/2012	2.342
2014.5	A/Texas/50/2012	1.158	A/Texas/91/2012	0.635				
			A/Oklahoma/3414/2013	0.689	A/Santiago/p36d5/2013	0.765	A/Oklahoma/3414/2013	0.689
2015.0	A/Texas/50/2012	N.A.[b]	N.A.	N.A.	A/Guam/3712/2013	N.A.	A/Guam/3712/2013	N.A.
2015.5	A/Switzerland/9715293/2013	N.A.	N.A.	N.A.	A/Virginia/10/2014	N.A.	A/Washington/3810/2014	N.A.

[a] Two strains are indicated when the most effective vaccine strain identified in seasons 1992.0 to the pre-target season (upper) was different from that identified in the pre-target season (lower).

[b] Not available.

2014). Given the relative frequency of strain i observed in season t ($x_{i(t)}$), its relative frequency in season $t + 1$ ($x_{i(t + 1)}$) was estimated by

$$\hat{x}_{i(t+1)} = x_{i(t)} e^{f_i},$$

where f_i denotes the fitness of strain i determined by the cross-immunity to the former epidemic viruses and the thermodynamic stability of HA1 (Luksza and Lassig, 2014). In the present study, f_i was defined as

$$f_i = f_0 - \left(c_i w_c + |g_i - g_o| w_g \right),$$

where c_i denotes the sum of cross-immunity to strain j isolated in season τ of $1992.0 \leq \tau \leq t$; that is $c_i = \sum_{j,\tau} \frac{1}{2^{d_{ij}}} x_{j(\tau)}$. The g_i indicates the thermodynamic stability of HA1 measured as the difference in the ΔG value from that for A/Hong Kong/19/1968 (PDB ID: 2HMG); that is $g_i = \sum_s \Delta\Delta G_s$, where $\Delta\Delta G_s$ is the
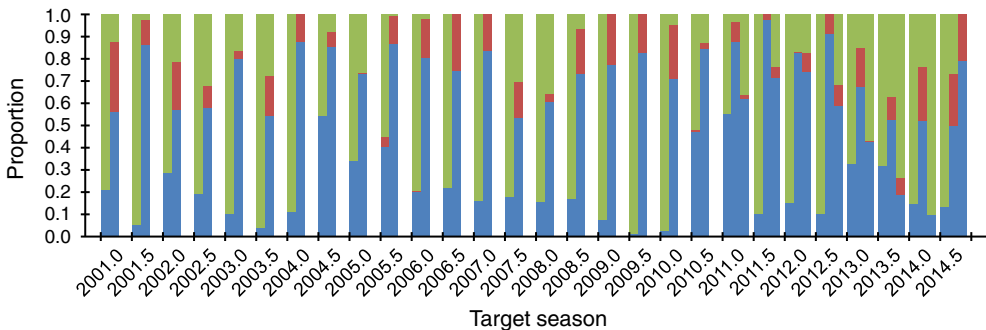

Fig. 1. Proportions of epidemic viruses sampled in the pre-target seasons with greater (blue), same (red), and smaller (green) average antigenic distances to those sampled in the target seasons of 2001.0 to 2014.5, in comparison to the actual vaccines (left), the center-of-mass strains (middle), and the strains identified by the fitness model (right). Note that the results of the fitness model could be obtained only for the target seasons of 2011.0 to 2014.5.

difference at amino acid site s estimated by FOLDX (version 3.0 beta 5.1) (Guerois et al., 2002; Schymkowitz et al., 2005). It was assumed that the stability of HA1 is optimum at g_o and the fitness declines linearly to the difference from it (Bloom and Glassman, 2009). The w_c and w_g are the weights for cross-immunity and thermodynamic stability, respectively, and f_0 is the constant to set $\sum_i \hat{x}_{i(t+1)} = 1$.

Similar to the case for the antigenic model, model parameters w_c, w_g, and g_o were re-optimized for each target season t as follows. First, in each of season τ, the strain with the smallest average antigenic distance to those sampled in season $\tau + 1$ was identified using the antigenic model. Then, w_c, w_g, and g_o were optimized by minimizing the square sum of the estimates of average antigenic distances between the identified strains in season τ and the predicted compositions of epidemic viruses in season $\tau + 1$, using the genetic algorithm (Tomita et al., 2000) with three sets of random real numbers assigned as the initial parameter values to confirm the convergence of the results (Suzuki, 2013b).

Results and discussion

Effectiveness of actual vaccines and the most effective vaccine strains

The average antigenic distances were first computed between the actual vaccines for H3N2 human influenza A virus that have been administered in practice (Table 1) and the epidemic viruses sampled in their target seasons of 2001.0 to 2014.5, to evaluate their effectiveness retrospectively. Here it was assumed that vaccination did not affect the composition of epidemic viruses to any large extent because worldwide coverage of vaccination is small (Osterholm et al., 2012). The average antigenic distances for actual vaccines were found to be relatively large, ranging from 0.892 to 3.170 with the overall average of 1.558 (Table 1). In particular, for 5 out of 28 seasons examined, the average antigenic distance exceeded the antigenic escape threshold of 2. Indeed, the occurrence of antigenic drift and the discrepancy in antigenicity between actual vaccines and epidemic viruses have been reported in some of these seasons (Centers for Disease Control and Prevention, 2004, 2010).

In human influenza viruses, the vaccine strain for target season t is identified at the end of season $t - 1$ (pre-target season) from the epidemic viruses isolated by then (Treanor, 2004; Russell et al., 2008a). To identify, retrospectively, the most effective vaccine strain for each target season of $2001.0 \leq t \leq 2014.5$, the average antigenic distance was computed between each of the strains isolated in seasons 1992.0 to $t - 1$ and the epidemic viruses sampled in season t. The strain with the smallest average antigenic distance was considered as the most effective vaccine strain (Table 1). It was observed that the average antigenic distance for the most effective vaccine strain was always smaller than 2, ranging from 0.181 to 1.400 with the overall average of 0.587 (Table 1). These values were also always smaller than those for actual vaccines obtained above. Notably, the most effective vaccine strain was identified in the pre-target season for 25 out of 28 seasons examined (Table 1).

When the average antigenic distance for the actual vaccine was compared with those for the strains isolated in the pre-target season, the former was greater than the latter in 79.1% of the cases on average (same: 0.2%; smaller: 20.7%) (Fig. 1). These results suggest that a strain randomly chosen in the pre-target season was expected to be more effective than the actual vaccine. This is probably because H3N2 human influenza A virus has been continuously escaping from the host immune responses through antigenic evolution (Shih et al., 2007; Suzuki, 2008). It is therefore recommended to replace the vaccine strain for every target season by a strain isolated in the pre-target season.

Selecting vaccine strains

In replacing the vaccine strain for every target season of H3N2 human influenza A virus, it is preferable to identify the most effective vaccine strain in the pre-target season. For this purpose, it may be helpful to understand the evolutionary dynamics of the virus. It has been proposed that H3N2 human influenza A virus circulates continuously in East and Southeast (E-SE) Asia, and epidemics in temperate regions are seeded from E-SE Asia every winter season (Rambaut et al., 2008; Russell et al., 2008b). Indeed, the phylogenetic tree of HA is known to comprise trunk branches and side clusters, which apparently represent evolutionary pathways of this virus in E-SE Asia and temperate regions, respectively. Furthermore, positive and negative

selection has been detected mainly on trunk branches and side clusters, respectively, suggesting that antigenic evolution occurs mainly in E-SE Asia (Fitch et al., 1991; Suzuki, 2008). According to this evolutionary scenario, it would be expected that the viruses circulating in E-SE Asia are not only positioned near the antigenic center of epidemic viruses in each season, but also most closely related to those in the following seasons. Then, the center-of-mass strain, which is defined as the strain with the smallest average antigenic distance to the others, in the pre-target season, may correspond to the most effective vaccine strain for the target season.

The center-of-mass strains identified for the target seasons of 2001.0 to 2014.5 are listed in Table 1. The average antigenic distances of these strains to the epidemic viruses sampled in their target seasons ranged from 0.253 to 2.355 with the overall average of 0.887, which were always smaller than those for actual vaccines (Table 1). The center-of-mass strain turned out to be the most effective vaccine strain for 7 out of 28 seasons examined (Table 1). In each pre-target season, the average antigenic distance for the center-of-mass strain was greater than those for the other strains only in 13.0% of the cases on average (same: 13.9%; smaller: 73.1%), suggesting that the center-of-mass strain was expected to be more effective than a strain randomly chosen in the pre-target season (Fig. 1). Vaccine strains for the future epidemic seasons of 2015.0 and 2015.5 could also be identified as the center-of-mass strains, which were A/Guam/3712/2013 and A/Virginia/10/2014, respectively (Table 1).

It has been proposed that the relative frequency of a strain in season $t + 1$ may be predicted from that observed in season t using the fitness model, where the fitness of a strain was assumed to be determined by the cross-immunity to the former epidemic viruses and the thermodynamic stability of HA1 (Luksza and Lassig, 2014). Using this model, the vaccine strain may be identified in the pre-target season as the strain with the smallest average antigenic distance to the predicted composition of epidemic viruses in the target season (Luksza and Lassig, 2014). Note that identifying the center-of-mass strain in the pre-target season is equivalent to assuming the same composition of epidemic viruses in the target and pre-target seasons for selecting vaccine strains by the fitness model (Luksza and Lassig, 2014).

In the retrospective analysis, vaccine strains could be identified by the fitness model only for the target seasons of 2011.0 to 2014.5, because of a failure in optimizing the model parameters due to relatively small numbers of data points available for computation (Table 1). In addition, for seven out of eight seasons examined, the average antigenic distance for the vaccine strain identified by the fitness model was greater than that for the center-of-mass strain obtained above (Fig. 1). The vaccine strain for the future epidemic season of 2015.0 identified by the fitness model (A/Guam/3712/2013) was the same as the center-of-mass strain, supporting the reliability. For season 2015.5, however, a different strain (A/Washington/3810/2014) was identified by the fitness model, which was likely to be less reliable than the center-of-mass strain as suggested from the results of the retrospective analysis (Table 1). Similar results were observed when either the cross-immunity or the thermodynamic stability was allowed to be included in the fitness model (data not shown).

The influenza virus forecasting system (INFLUCAST)

In the present study, the effectiveness of a potential vaccine strain for a target season of H3N2 human influenza A virus was evaluated retrospectively using the average antigenic distance between the strain and the epidemic viruses sampled in the target season. The most effective vaccine strain was identified mostly in the pre-target season and the effectiveness of actual vaccines appeared to be lower than that of the strains randomly chosen in the pre-target season on average. It was recommended to replace the vaccine strain for every target season with the center-of-mass strain in the pre-target season. Recently, complicated algorithms based on the phylogenetic analysis have been devised to identify the strains that were likely to be more effective than actual vaccines in the pre-target seasons (Luksza and Lassig, 2014; Steinbruck et al., 2014). However, the strains isolated in the pre-target seasons were expected to be more effective than actual vaccines as indicated above, and thus it may be more meaningful to identify the most effective vaccine strains in the pre-target seasons. The procedure described in the present study provides a simple way directed toward this purpose.

The procedure of selecting vaccine strains for future epidemic seasons described in the present study was implemented in the influenza virus forecasting system (INFLUCAST) (http://www.nsc.nagoya-cu.ac.jp/~yossuzuk/influcast.html). The system should be continuously updated with accumulation of HI and sequence data and upgraded by improving the antigenic model (Wilson et al., 1981; Xia et al., 2012), the measure of thermodynamic stability (Wiley et al., 1981; Han and Marasco, 2011; Kryazhimskiy et al., 2011), and the fitness model (Soundararajan et al., 2011; Ekiert et al., 2012; Kobayashi and Suzuki, 2012a; Omori and Sasaki,

2013). It is also important to incorporate missing variations due to incomplete sampling of epidemic viruses and *de novo* mutations into the system. The last problem arises because inactivated vaccines are administered to induce systemic humoral immunity against variable proteins for human influenza viruses. This problem may be avoided by targeting conserved proteins as in the case of poliovirus (Suzuki, 2004), or by inducing heterologous immunity with live-attenuated vaccines as in the case of rotavirus (Soares-Weiser et al., 2010; Cortese et al., 2013). It should be noted, however, that the evolutionary mechanism of amino acid sequence may be integrated with that of antigenicity (Suzuki, 2013b), suggesting that the system may be extended to predict the antigenic evolution.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mgene.2015.03.003.

Acknowledgments

References

Ahmad, S., Gromiha, M.M., Fawareh, H., Sarai, A., 2004. ASAView: database and tool for solvent accessibility representation in proteins. BMC Bioinf. 5, 51.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., Lipman, D., 2008. The Influenza Virus Resource at the National Center for Biotechnology Information. J. Virol. 82, 596–601.

Bedford, T., Rambaut, A., Pascual, M., 2012. Canalization of the evolutionary trajectory of the human influenza virus. BMC Biol. 10, 38.

Bloom, J.D., Glassman, M.J., 2009. Inferring stabilizing mutations from protein phylogenies: application to influenza hemagglutinin. PLoS Comput. Biol. 5, e1000349.

Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., Fitch, W.M., 1999. Predicting the evolution of human influenza A. Science 286, 1921–1925.

Centers for Disease Control and Prevention, 2004. Update: influenza activity—United States, 2003–2004 season 53, pp. 284–287.

Centers for Disease Control and Prevention, 2010. Update: influenza activity—United States, 2009–2010 season 59, pp. 901–908.

Coleman, M.T., Dowdle, W.R., Pereira, H.G., Schild, G.C., Chang, W.K., 1968. The Hong Kong/68 influenza A2 variant. Lancet 292, 1384–1386.

Cortese, M.M., Immergluck, L.C., Held, M., Jain, S., Chan, T., Grizas, A.P., Khizer, S., Barrett, C., Quaye, O., Mijatovic-Rustemapasic, S., Gautam, R., Bowen, M.D., Moore, J., Tate, J.E., Parashar, U.D., Vazquez, M., 2013. Effectiveness of monovalent and pentavalent rotavirus vaccine. Pediatrics 132, e25–e33.

Ekiert, D.C., Kashyap, A.K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J.H., Dillon, M.A., O'Neil, R.E., Faynboym, A.M., Horowitz, M., Horowitz, L., Ward, A.B., Palese, P., Webby, R., Lerner, R.A., Bhatt, R.R., Ian, A., Wilson, I.A., 2012. Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature 489, 526–532.

Fitch, W.M., Leiter, J.M., Li, X.Q., Palese, P., 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. U. S. A. 88, 4270–4274.

Francis Jr., T., Salk, J., Pearson, H.E., Brown, P.N., 1944. Protective effect of vaccination against induced influenza A. Proc. Soc. Exp. Biol. Med. 55, 104–105.

Grantham, R., 1974. Amino acid difference formula to help explain protein evolution. Science 185, 862–864.

Guerois, R., Nielsen, J.E., Serrano, L., 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. 320, 369–387.

Gupta, V., Earl, D.J., Deem, M.W., 2006. Quantifying influenza vaccine efficacy and antigenic distance. Vaccine 24, 3881–3888.

Han, T., Marasco, W.A., 2011. Structural basis of influenza virus neutralization. Ann. N. Y. Acad. Sci. 1217, 178–190.

He, J., Deem, M.W., 2010. Low-dimensional clustering detects incipient dominant influenza strain clusters. Protein Eng. Des. Sel. 23, 935–946.

Henle, W., Henle, G., Stokes Jr., J., 1943. Demonstration of the efficacy of vaccination against influenza type A by experimental infection of human beings. J. Immunol. 46, 163–175.

Hensley, S.E., Das, S.R., Bailey, A.L., Schmidt, L.M., Hickman, H.D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J.R., Yewdell, J.W., 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science 326, 734–736.

Huang, J.-W., King, C.-C., Yang, J.-M., 2009. Co-evolution positions and rules for antigenic variants of human influenza A/H3N2 viruses. BMC Bioinf. 10, S41.

Ito, K., Igarashi, M., Miyazaki, Y., Murakami, T., Iida, S., Kida, H., Takada, A., 2011. Gnarled-trunk evolutionary model of influenza A virus hemagglutinin. PLoS One 6, e25953.

Kobayashi, Y., Suzuki, Y., 2012a. Compensatory evolution of net-charge in influenza A virus hemagglutinin. PLoS One 7, e40422.

Kobayashi, Y., Suzuki, Y., 2012b. Evidence for N-glycan shielding of antigenic sites during evolution of human influenza A virus hemagglutinin. J. Virol. 86, 3445–3451.

Kryazhimskiy, S., Dushoff, J., Bazykin, G.A., Plotkin, J.B., 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 7, e1001301.

Lee, M.-S., Chen, J.S.-E., 2004. Predicting antigenic variants of influenza A/H3N2 viruses. Emerg. Infect. Dis. 10, 1385–1390.

Lee, M.-S., Chen, M.-C., Liao, Y.-C., Hsiung, C.A., 2007. Identifying potential immunodominant positions and predicting antigenic variants of influenza A/H3N2 viruses. Vaccine 25, 8133–8139.

Lees, W.D., Moss, D.S., Shepherd, A.J., 2010. A computational analysis of the antigenic properties of haemagglutinin in influenza A H3N2. Bioinformatics 26, 1403–1408.

Liao, Y.-C., Lee, M.-S., Ko, C.-Y., Hsiung, C.A., 2008. Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. Bioinformatics 24, 505–512.

Luksza, M., Lassig, M., 2014. A predictive fitness model for influenza. Nature 507, 57–61.

Mitnaul, L.J., Castrucci, M.R., Murti, K.G., Kawaoka, Y., 1996. The cytoplasmic tail of influenza A virus neuraminidase (NA) affects NA incorporation into virions, virion morphology, and virulence in mice but is not essential for virus replication. J. Virol. 70, 873–879.

Nelson, D.L., Cox, M.M., 2008. Lehninger Principles of Biochemistry. 5th edition. W. H. Freeman and Company, New York.

Nelson, M.I., Holmes, E.C., 2007. The evolution of epidemic influenza. Nat. Rev. Genet. 8, 196–205.

Omori, R., Sasaki, A., 2013. Timing of the emergence of new successful viral strains in seasonal influenza. J. Theor. Biol. 329, 32–38.

Osterholm, M.T., Kelley, M.S., Sommer, A., Belongia, E.A., 2012. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. Lancet Infect. Dis. 12, 36–44.

Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453, 615–619.

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., de Jong, J.C., Kelso, A., Klimov, A.I., Kageyama, T., Komadina, N., Lapedes, A.S., Lin, Y.P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A.D.M.E., Rimmelzwaan, G.F., Shaw, M.W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R.A.M., Smith, D.J., 2008a. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. Vaccine 26, D31–D34.

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., de Jong, J.C., Kelso, A., Klimov, A.I., Kageyama, T., Komadina, N., Lapedes, A.S., Lin, Y.P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A.D.M.E., Rimmelzwaan, G.F., Shaw, M.W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R.A.M., Smith, D.J., 2008b. The global circulation of seasonal influenza A (H3N2) viruses. Science 320, 340–346.

Schymkowitz, J.W.H., Rousseau, F., Martins, I.C., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., 2005. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc. Natl. Acad. Sci. U. S. A. 102, 10147–10152.

Shih, A.C.-C., Hsiao, T.-C., Ho, M.-S., Li, W.-H., 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc. Natl. Acad. Sci. U. S. A. 104, 6283–6288.

Skehel, J.J., Wiley, D.C., 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. 69, 531–569.

Smith, D.J., Lapedes, A.S., de Jong, J.C., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305, 371–376.

Soares-Weiser, K., Maclehose, H., Ben-Aharon, I., Goldberg, E., Pitan, F., Cunliffe, N., 2010. Vaccines for preventing rotavirus diarrhoea: vaccines in use. Cochrane Database Syst. Rev. CD008521.

Soundararajan, V., Zheng, S., Patel, N., Warnock, R., Wilson, I.A., Raguram, S., Sasisekharan, V., Sasisekharan, R., 2011. Networks link antigenic and receptor-binding sites of influenza hemagglutinin: mechanistic insight into fitter strain propagation. Sci. Rep. 1, 200.

Steinbruck, L., Klingen, T.R., McHardy, A.C., 2014. Computational prediction of vaccine strains for human influenza A (H3N2) viruses. J. Virol. 88, 12123–12132.

Suzuki, Y., 2004. Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene 328, 127–133.

Suzuki, Y., 2006. Natural selection on the influenza virus genome. Mol. Biol. Evol. 23, 1902–1911.

Suzuki, Y., 2008. Positive selection operates continuously on hemagglutinin during evolution H3N2 human influenza A virus. Gene 427, 111–116.

Suzuki, Y., 2013a. Detection of positive selection eliminating effects of structural constraints in hemagglutinin of H3N2 human influenza A virus. Infect. Genet. Evol. 16, 93–98.

Suzuki, Y., 2013b. Predictability of antigenic evolution for H3N2 human influenza A virus. Genes Genet. Syst. 88, 225–232.

Suzuki, Y., Nei, M., 2002. Origin and evolution of influenza virus hemagglutinin genes. Mol. Biol. Evol. 19, 501–509.

Tomita, M., Hashimoto, K., Takahashi, K., Matsuzaki, Y., Matsushima, R., Saito, K., Yugi, K., Miyoshi, F., Nakano, H., Tanida, S., Saito, Y., Kawase, A., Watanabe, N., Shimizu, T., Nakayama, Y., 2000. The E-CELL project: towards integrative simulation of cellular processes. N. Gener. Comput. 18, 1–12.

Treanor, J., 2004. Weathering the influenza vaccine crisis. N. Engl. J. Med. 351, 2037–2040.

Whittle, J.R.R., Zhang, R., Khurana, S., King, L.R., Manischewitz, J., Golding, H., Dormitzer, P.R., Haynes, B.F., Walter, E.B., Moody, M.A., Kepler, T.B., Liao, H.-X., Harrison, S.C., 2011. Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc. Natl. Acad. Sci. U. S. A. 108, 14216–14221.

Wiley, D.C., Wilson, I.A., Skehel, J.J., 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289, 373–378.

Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289, 366–373.

World Health Organization, 2009. Influenza (Seasonal). http://www.who.int/mediacentre/factsheets/fs211/en/index.html.

World Health Organization, 2014a. Recommended composition of influenza virus vaccines for use in the 2014–2015 northern hemisphere influenza season. Wkly. Epidemiol. Rec. 89, 93–104.

World Health Organization, 2014b. Recommended composition of influenza virus vaccines for use in the 2015 southern hemisphere influenza season. Wkly. Epidemiol. Rec. 89, 441–452.

Xia, Z., Jin, G., Zhu, J., Zhou, R., 2009. Using a mutual information-based site transition network to map the genetic evolution of influenza A/H3N2 virus. Bioinformatics 25, 2309–2317.

Xia, Z., Huynh, T., Kang, S.-g., Zhou, R., 2012. Free-energy simulations reveal that both hydrophobic and polar interactions are important for influenza hemagglutinin antibody binding. Biophys. J. 102, 1453–1461.