Infection, Genetics and Evolution 67 (2019) 67–72

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Anticipating time-dependent antigenic variants of influenza A (H3N2) viruses

Emmanuel S. Adabor

School of Technology, Ghana Institute for Management and Public Administration, P. O. Box AH50, Achimota, Accra, Ghana

ARTICLE INFO

Keywords:
Influenza virus
Evolution
Amino acids
Antigenic sites
Statistical models

ABSTRACT

Frequent variations in influenza vaccines are necessary to match antigenic variants which appear in influenza epidemics. Antigenic variants of influenza viruses result from frequent mutations in amino acid residues located on their hemagglutinin (HA) proteins. Knowledge of specific changes in these amino acids helps to characterize distinct antigenic variants. In this paper, statistical models are developed and used to investigate changes in amino acids which accompany antigenic variants of epidemiological importance. Amino acid sequences of the HA proteins of influenza A (H3N2) strains isolated from 1968 to 2015 were obtained. The sequences were aligned using Clustal Omega and the number of differences in amino acid residues located on annotated positions of antigenic sites of the HA protein between pairs of strains were determined. These were linked in the statistical models and used to assess the relationship between any pair of influenza strains. The results revealed that both antigenic similarity between strains and the amino acid changes are affected by the time of isolation of the strains. Furthermore, the models predicted that rates of changes in amino acids located on the antigenic sites ranged between 5% and 6% per site per year. The findings of the study suggest that time-dependent antigenic variants of influenza A (H3N2) strains may occur as they evolve. The study has the potential to greatly improve influenza surveillance in as much as it supports vaccine designs.

1. Introduction

Influenza vaccinations are carried out to protect against influenza infections in influenza seasons. The vaccines stimulate adaptive immune responses which usually result in the production of antibodies for virus neutralization in future influenza infections. These antibodies (Abs) recognize the hemagglutinin (HA) protein of the influenza viruses. They bind to specific antigenic sites of the HA protein of these viruses. Particularly, five specific regions on the HA protein of influenza A (H3N2) are designated antigenic sites (epitopes) (Wilson and Cox, 1990). However, frequent changes in amino acids in these sites result in new strains (antigenic variants) which favour the viruses to escape Abs (Brown et al., 1990; Laver et al., 1981). It has emerged that at least four amino acid changes located on at least two of the five designated antigenic sites (A-E) appear with antigenic variants of influenza A (H3N2) (Wilson and Cox, 1990).

Consequently, changes in amino acids in those regions affect the interactions between Abs and viruses. These are major concerns in influenza virus studies. The importance of understanding the dynamics of amino acids at these epitopes are crucial to the design of effective influenza vaccines. It has been shown that changes in amino acids at these epitopes affect the neutralization of the mutated viruses (Ndifon et al., 2009a). In particular, Ndifon and co-workers inferred from the study of antibody interference that amino acid changes to epitopes A, B and D could averagely lead to decrease the neutralization of virus while changes to epitopes C and E could lead to increase neutralization of resulting mutated viruses (Ndifon et al., 2009a). For instance, in animal studies, single amino acid substitution at the amino acid position 156 of the antigenic site B resulted in reduced antigenicity and efficacy of vaccines (Katz and Webster, 1989; Kodihalli et al., 1995). Additionally, independent studies of different influenza strains suggest the immunodominance of epitopes A and B over the other designated sites (Popova et al., 2012; Smith et al., 1991; Temoltzin-Palacios and Thomas, 1994; Nobusawa et al., 2012).

Thus, an understanding of the dynamics of changes in amino acids located on the antigenic sites of the HA protein of influenza viruses elucidates the direction of evolution of the influenza virus while informing the design of more effective vaccines through the influenza studies. However, such studies in influenza do not include the effects of differences between years of isolated influenza strains on the changes in the amino acids. With the abundance of data due to the global effort in influenza surveillance, it is possible to systematically ascertain the future of the changes in amino acids in the epitopes of new antigenic variants of the influenza virus from the years of isolation. Inferring

E-mail address: emmanuelsadabor@gimpa.edu.gh.

https://doi.org/10.1016/j.meegid.2018.10.028
Received 22 March 2018; Received in revised form 16 October 2018; Accepted 31 October 2018
Available online 02 November 2018

changes in these amino acids from the years of isolation of influenza strains could lead to the accurate predictions of probable amino acid changes in the epitopes that accompany future influenza variants.

To this end, statistical models are developed to determine the changes in amino acids at designated epitopes of the HA of influenza A (H3N2). The principal changes that are reported to affect the neutralization of influenza strains by vaccines are considered. In this way, a more efficient approach to estimate changes in amino acids that lead antigenic variants of influenza viruses will be achieved. The rates of changes in amino acids located on the designated epitopes are inferred from the analysis of differences between the years of isolation of influenza A (H3N2) strains. The results of this study are very significant to the design of vaccines as they form a basis for the choice of vaccines that allow the induction of antibodies for specific epitopes. For instance, knowledge of predicted amino acid changes at immunodominant epitopes could suggest the need to limit probable antibody-induction of those epitopes to enhance the effectiveness of the vaccine (Garrity et al., 1997). A systematic analysis of the results appraises the following assertions: (1) years of isolation of influenza strains are significantly linked to the expected rates of changes in amino acids located on the HA proteins of influenza A (H3N2) strains; (2) changes in amino acids at each epitope vary with years of isolation of influenza strains; and (3) the antigenic relatedness between a pair of influenza strains is linked to the years of isolation of those strains.

2. Materials and methods

2.1. Data sets

In order to obtain data for the study, a list of influenza A (H3N2) strains that circulated from 1968 to 2016 was compiled. The amino acid sequences of the HA proteins of those strains were searched from the National Center for Biotechnology Information (NCBI) protein database (NCBI Resource Coordinators, 2016). However, only 254 complete sequences corresponding to 254 strains were available as at the time of the study. These comprised of isolates characterized from 1968 to 2015. Therefore a set of 254 influenza A (H3N2) viral strains and corresponding amino acid sequences of the HA proteins were considered for the study (see Supplementary Table S1). The choice of application to influenza A (H3N2) is owed to their epidemiological importance and availability of the genetic data. A multiple sequence alignment of all the amino acid sequences of the viruses was performed using Clustal Omega (Sievers et al., 2011).

The HA protein of influenza A (H3N2) has five designated antigenic sites (A - E) (Wilson and Cox, 1990; Wiley et al., 1981). Using published (Ndifon et al., 2009a) annotated amino acid positions on the antigenic sites of an Influenza A (H3N2) HA sequence, the positions in the collection of sequences of viruses were determined after harmonizing to same lengths. More specifically, a multiple sequence alignment of a compendium of the amino acid sequences of the HA of the influenza A (H3N2) viruses and the sequence of influenza A/California/7/2004 (H3N2) was performed using Clustal Omega (Sievers et al., 2011). The sequence of influenza A/California/7/2004 strain was aligned to find the respective designated amino acid positions of interest (Table 1) since the designated positions at the antigenic sites in its sequence were known (Ndifon et al., 2009a). Further, the corresponding amino acid positions found in the antigenic sites in all the influenza A (H3N2) viruses under consideration were identified. Here, the amino acid positions on antigenic sites were used because they better decipher antigenic differences between influenza strains (Lee and Chen, 2004). Furthermore, epitopes A and B were included in the study because they are the most important epitopes in directing antigenic drift of influenza A (H3N2) (Sato et al., 2004; Bush et al., 1999) although epitope D is also important (Ndifon et al., 2009a). This motivated the inclusion of the epitope D. On the other hand, with a recent report (Ndifon et al., 2009a) on the effects of amino acid mutations at epitopes C and E on the neutralization of the mutated viruses, investigating changes in these regions will be relevant to influenza surveillance.

In addition, the numbers of differences in the amino acids at the designated antigenic sites between pairs of the influenza viruses were identified from the sequence data using R codes. Several other works exploit such sequence information for genetic analyses of influenza viruses (Gupta et al., 2006; Liao et al., 2008; Xia et al., 2009). Moreover, the year of isolation of each influenza strain was derived and the pairwise differences between the years of isolation of strains were determined. These formed the data for the differences between years of isolation of strains.

2.2. Modelling the number of differences in amino acids at designated epitopes of distinct strains

For any influenza season, the chances of circulating strains recurring in a subsequent season is uncertain although it is known that frequent changes in amino acids at the designated epitopes results in new strains (Webster et al., 1982; Holmes et al., 2005). With this basis, it suffices to define the number of differences in amino acids at an epitope of HA protein of a strain and its variant as a Poisson process. For any pair of strains, suppose Y_i is the number of differences in amino acids located at the same positions of epitope, i, of two strains. Then the probability of observing any Y_i is defined by the function:

$$P(Y_i = y) = \frac{e^{\mu_i}\mu_i^{\,y}}{y!} \tag{1}$$

where μ_i is the mean of Y_i with variance equivalent to its mean. $P(Y_i = y)$ is the probability that the number of differences in amino acids located at the same positions of an epitope, i, of the two strains is y.

A Poisson regression model provides a fit to the mean of the number of differences in amino acids at the same epitope of two distinct strains. More formally, a simple linear regression of the mean, μ_i, is expressed as:

$$\mu_i = \beta_0 + \beta_1 X_i \tag{2}$$

where X_i is the difference between the years of isolation of the two strains, β_1 is the average number of differences in amino acids at the epitope for a unit difference between the years of isolation for the two distinct strains under consideration, and β_0 is a constant which account for other (biological and physical) factors that contribute to the differences in amino acids between strains. In the current data, differences between the years of isolation of strains significantly correlate with the number of differences in amino acids located on the annotated positions of antigenic sites and the antigenic relatedness between strains (See Results). These provide further evidence to support Eq. (2).

However, μ_i is positive because it represents the mean number of differences in amino acid residues between two strains. Therefore the natural way to satisfy the positivity constraint is to take the logarithm of the mean number of differences in amino acids between the two strains. Thus, it suffices to reformulate the model in order to explain the mean number of differences in the amino acid residues between two strains as:

$$\log(\mu_i) = \beta_0 + \beta_1 X_i \tag{3}$$

Hence, the alternative expression is used:

$$\mu_i = \alpha e^{\beta_1 X_i} \tag{4}$$

where α is a constant.

Thus, Eq. (4) represents the expected number of changes in amino acids at an epitope, i, that associates new antigenic variants. This implies that a unit difference between years of isolation of strains has a multiplicative effect on the mean number of differences in amino acid residues at the antigenic sites for any two strains. More specifically, it suggests that the marginal effect of a difference between the years of

Table 1
Amino acid positions found in designated antigenic sites of hemagglutinin of influenza A (H3N2).

Antigenic site	Amino acid positions
A	122, 124, 126, 130, 131, 132, 133, 135, 137, 138, 140, 142, 143, 144, 145, 146, 150, 152, 168
B	128, 129, 155, 156, 157, 158, 159, 160, 163, 164, 165, 186, 187, 188, 189, 190, 192, 193, 194, 196, 197, 198
C	44, 45, 46, 47, 48, 50, 51, 53, 54, 273, 275, 276, 278, 279, 280, 294, 297, 299, 300, 304, 305, 307, 308, 309, 310, 311, 312
D	96, 102, 103, 117, 121, 167, 170, 171, 172, 173, 174, 175, 176, 177, 179, 182, 201, 203, 207, 208, 209, 212, 213, 214, 215, 216, 217, 218, 219, 226, 227, 228, 229, 230, 238, 240, 242, 244, 246, 247, 248
E	57, 59, 62, 63, 67, 75, 78, 80, 81, 82, 83, 86, 87, 88, 91, 92, 94, 109, 260, 261, 262, 265

A, B, C, D and E are the five designated antigenic sites (regions) of the influenza A (H3N2) hemagglutinin. The annotated amino acid positions were obtained from the literature (Ndifon et al., 2009a).

isolation of two strains on the expected number of differences in amino acids (located at the annotated positions of the epitope) is a change in the amino acids by an exponential factor. Furthermore, Eq. (4) permits the direct prediction of the chances of observing specific number of changes in amino acids at any of the epitopes that will follow future antigenic variants of influenza virus for any influenza season. Similarly, it is possible to infer the time it will take to observe specific changes in the amino acids that will accompany new antigenic variants.

2.3. Robust models

Under normal circumstances, the mean of a Poisson model should be equal to the variance. However, for a non-constant mean of observations, it is possible for the variance to be greater than the mean (overdispersion) in some instances (Ver Hoef and Boveng, 2007; Zelterman, 2006). In other cases, the variance may be less than the mean (underdispersion). On such occasions, robust regression approaches, based on exponential family distribution, are applied to obtain robust model estimates in a way that controls the errors associated with the models and provide naturally interpretable parameters.

In general, for a distribution of Y_i belonging to the exponential family, the mean of Y_i is defined as $E(Y_i) = \mu_i$ and the variance is defined as $var(Y) = \theta\mu_i$, while θ is a dispersion parameter. The comparative relationships between the mean and variance of the Poisson and the exponential family distributions provide the basis for the name of the robust approach called *quasi-poisson* model. Thus, a *quasi-poisson* model permits reformulation of a Poisson model into a function of a mean and a dispersion parameter that account for the difference that exists between mean and variance of the Poisson model. In addition, the negative binomial distribution is applied to modelling overdispersion in a Poisson model. However, it is distinguished from the *quasi-poisson* model because while the variance-mean relationship is linear in the *quasi-poisson* model, it is quadratic in a negative binomial regression model. More specifically, for negative binomial regression, the mean and the variance are defined by $E(Y_i) = \mu_i$ and $var(Y_i) = \mu + k\mu^2$ respectively, where $\mu > 0$ and $k > 0$. The expression, $1 + k\mu$, defines an overdispersion parameter. Estimates of regression coefficients of both the *quasi-poisson* regression and the negative binomial regression are obtained by the iterative weighted least squares method (Lee and Nelder, 2000; Wedderburn, 1974). The estimates of regression coefficients in both the *quasi-poisson* and the negative binomial models differ due to the difference in the variance-mean relationships in both models. Statistical tests to confirm the significance of overdispersion or underdispersion are regression-based as suggested in (Cameron and Trivedi, 1990).

3. Results

3.1. Models for investigating changes in amino acids and years of isolation of strains

In order to establish a link between the years of isolation of strains and the observed changes in amino acids in the epitopes, statistical models of the amino acid changes and time are developed. On the lines of review of the epitopes considered in this study, eight different models were developed. Each model described changes in amino acids at the antigenic sites which support antigenic drifts of the influenza A (H3N2) relevant to epidemiological studies. In particular, for each model, the *quasi-poisson* regression fit and negative binomial regression fit were compared to establish the better of the two models. The two models were tested in an 11-fold cross validation after performing dispersion tests of the Poisson models (Table 2). All the Poisson fits to the current data sets were found to be overdispersed. Therefore the comparisons of models were made between the *quasi-poisson* regression and the negative binomial regression. The models are compared on the basis of root-mean-square errors and the coefficients of determination (R-squared) associated with the models. Note that both Poisson regression and the *quasi-poisson* regression result in the same estimates of model coefficients although the *quasi-poisson* regression controls standard errors of estimates.

In all, the *quasi-poisson* regression models were found to provide better fits to the expected number of changes in amino acids located on the designated antigenic sites compared to the negative binomial regression (Table 2). In addition, models of difference between years of isolations of influenza strains explained more variations in the changes in amino acids at the antigenic sites A and B compared to the other designated antigenic sites. These were observed because changes in amino acids at sites A and B are more advantageous to viruses compared to antigenic sites C, D and E as indicated in a previous study

Table 2
Evaluating models of the changes in amino acids.

	Quasi-poisson		Negative binomial	
Model[a]	R-squared	RMSE[b]	R-squared	RMSE
A	0.37	4.45	0.35	4.81
B	0.37	4.63	0.35	5.22
C	0.34	4.79	0.29	6.28
D	0.35	5.60	0.31	7.85
E	0.35	4.74	0.31	5.57
A + B[c]	0.37	10.61	0.34	13.01
A + B + D[c]	0.37	21.13	0.33	26.54
C + E	0.34	8.68	0.29	13.82

[a] Model of expected changes in amino acids on the epitopes (A - E) due to year of isolation.
[b] RMSE is the root-mean-square error associated with model in an 11-fold cross validation.
[c] Model involving sum of epitopes.

Table 3
Rates of changes in amino acids due to a difference between years of isolation of strains.

Model[a]	[b]Expected rates of changes in amino acids (%)	Significance (*p*-value)
A	5.16	< 2e-16
B	4.99	< 2e-16
C	6.49	< 2e-16
D	6.02	< 2e-16
E	6.23	< 2e-16
A + B	5.07	< 2e-16
A + B + D	5.42	< 2e-16
C + E	6.36	< 2e-16

[a] Model of expected changes in amino acids on the epitopes (A - E) due to year of isolation alone.
[b] This is the variable coefficient in model.

(Ndifon et al., 2009a). However, on average, both *quasi-poisson* regression and the negative binomial regression could explain 36% and 32% of the variations in the number of amino acid changes on the antigenic sites of the HA protein of influenza A (H3N2) strains respectively. These results are remarkable since mutations of amino acids are mostly affected by biochemical factors and exposure to certain environmental factors (Lodish et al., 2000). Moreover, these results suggest that the models are statistically significant (Table 2). These provide further support for analysis of the rates of changes in amino acids located on the antigenic sites based on the differences between the years of isolation of influenza strains.

3.2. Years of isolation of influenza strains affect changes in amino acids on each epitope

Furthermore, investigations of the actual rates of changes of the amino acids located on the designated antigenic sites were conducted using the *quasi-poisson* modelling since it was shown to be more accurate (Table 3). In addition, the contributions of the difference between the years of isolation of strains to the changes in amino acid residues on the designated antigenic sites of the influenza strains were examined. All models indicated that the difference between years of isolations of influenza strains had statistically significant effect on the changes in amino acids at the same positions on the antigenic sites of the strains (Table 3). These results suggest that for any influenza strain isolated in a particular year, it has a non-zero chance to mutate some number of amino acids on the antigenic sites in the absence of factors other than change in season. This is consistent with expectations since the frequent changes in amino acids at the antigenic sites leading to new strains are known (Layne, 2006; Carrat and Flahault, 2007; Fitch et al., 1997). This is the reason for which influenza vaccines are examined annually.

3.3. Ranges of rates of changes in amino acids in antigenic sites

The rates of changes in amino acids located on the antigenic sites ranged between 5% and 6% (5e-2 and 6e-2) per site per year (Fig. 1). This is in the order of the results of a very recent study which found the rates to be 1.29e-2 changes per site per year (Kirkpatrick et al., 2018). However, while all amino acids located on the annotated positions of the antigenic sites were considered in this study, Kirkpatrick et al. (2018) analysed only some portions of the sites (Kirkpatrick et al., 2018). Perhaps, this could partly account for the difference which appears in the two empirical estimates of the rates of changes in amino acids. The expected number of differences in amino acids located on the antigenic sites A and B of the HA proteins of strains changes at a rate of about 5% whenever the years of isolation differ by one. This is, however, lower compared to the rates of changes in the amino acids on the antigenic sites C to E which change at a rate of about 6% for every passing year (Fig. 1). These suggest that the difference between years of isolation of strains has higher effect on the expected number of amino acid changes at antigenic sites C to E compared to sites A and B. In addition, these results indicate that the other factors explaining > 36% of the variations in amino acid changes in the antigenic sites may have higher effect on the changes in antigenic sites A and B compared to sites C to E.

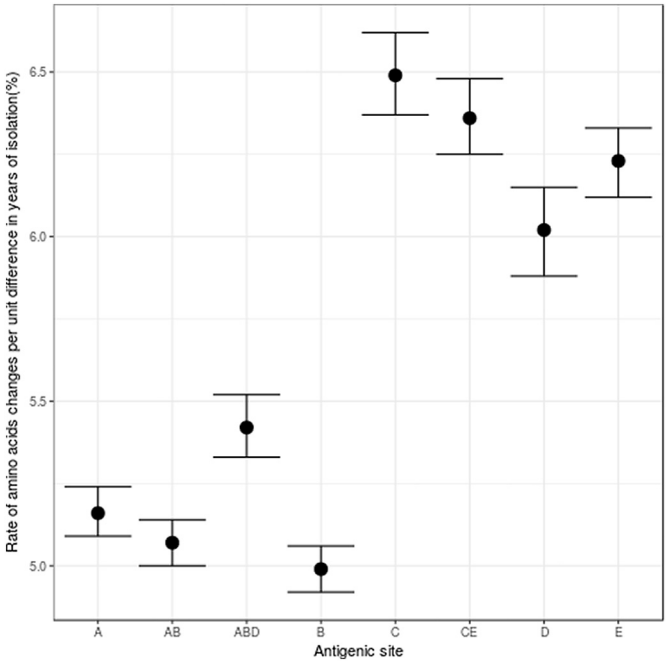

Fig. 1. 95% confidence intervals of incidence of amino acid changes on antigenic sites of the HA protein of influenza A (H3N2). The rates of changes are changes at each antigenic site per year.

3.4. Antigenic relatedness is linked to the years of isolation of strains

Furthermore, the effect of years of isolation of influenza strains on the measures of antigenic similarity between strains was examined. This result provides an indication of the existence of a relationship between antigenic relatedness and the years of isolation of strains. To achieve this, hemagglutination-inhibition (HI) titres corresponding to influenza A (H3N2) strain-serum pairs were required. These were obtained for 809 pairs of distinct strains isolated from 1968 to 2015 (Supplementary Table S2). The HI titres would enable the estimation of antigenic relatedness between strains. Antigenic relatedness (difference) between pairs of strains were measured by the reciprocal of the normalized HI titres (Ndifon et al., 2009b; Webster et al., 2002). For any pair of strains, say *V1* and *V2*, the strain *V2* is said to be antigenically drifted relative to *V1* when the reciprocal of the normalized HI titre of *V2* relative to antisera raised against *V1* is greater than or equal to four. Besides its many usage in measuring antigenic similarity in such studies in the past (Ndifon et al., 2009b; Webster et al., 2002), the reciprocal of the normalized HI titres permitted the measurements of antigenic similarities among most of the strains collected for the study. The correlation coefficient between the logarithm (base 2) of the reciprocal of the normalized HI titres and the difference between years of isolation of strains was determined.

It was anticipated that the antigenic relatedness between the influenza strains should correlate with differences between years of isolation because antigenically distinct strains isolated in different influenza seasons usually evolved from previous strains (Fitch et al., 2000). As expected, a very significant positive correlation coefficient was found between the antigenic relatedness and the differences between years of isolation of strains ($r = 0.46$, $p < 2.2e\text{-}16$). This result

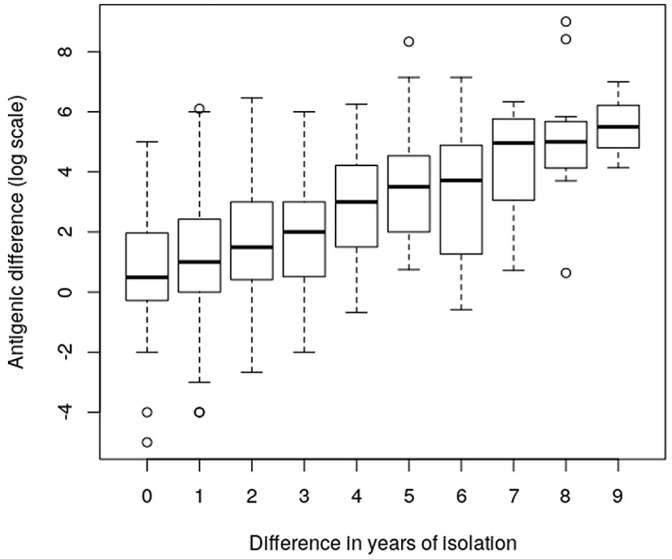

Fig. 2. Antigenic similarity depends on difference between years of isolation of strains. Antigenic difference is logarithm (base 2) of the reciprocal of the normalized hemagglutination-inhibition titres.

suggests that antigenic difference between strains increases when the difference between the years of isolation of the strains increases.

In addition, the variations in antigenic similarities across the differences among the years of isolations of pairs of strains were examined. The current data allowed the grouping of pairs of strains isolated in the same years and those isolated in years that differed up to nine years. Consistent with the earlier results, antigenic difference between strains was found to increase as the difference between years of isolation of strains increases (Fig. 2). Furthermore, it was found that for any particular difference between years of isolation of pairs of strains, the antigenic differences among strains in that group were more similar compared to other groups. Nevertheless, the variations in antigenic difference within the same groups are similar across the groupings except for pairs of strains with eight (8) and nine (9) differences in years of isolation (Fig. 2). These are attributed to the smaller amount of strains that were available for those groups for this study. Nonetheless, these common variations within each group demonstrate that influenza strains isolated in the same years are more related compared to those isolated in different years. Furthermore, this result connotes that any current influenza A (H3N2) isolate is closely linked to recently isolated strains compared to formerly isolated strains. This is important and consistent with the process of evolution of the influenza virus because the virus mutates quickly in efforts to escape neutralizing antibodies as noted in the human immune system. They evolve from formally recognized particles into mutant strains with changes in amino acids on the HA protein which are recognized by Abs (Fitch et al., 2000).

4. Discussion

The frequent transformations of influenza viruses require effective surveillance in order to monitor the antigenic changes in new variants of influenza strains. In this regard, studies in influenza have involved analysing the changes in amino acids. These are very important since alterations in the antigenicity of the influenza virus are correlated with key amino acid changes (Sleigh et al., 1981). In this study, statistical models were developed and used to investigate the changes in amino acids located on the antigenic sites of the influenza A (H3N2) HA protein. More specifically, *quasi-poisson* models found that the differences in amino acids on the designated antigenic sites of the HA protein of influenza A (H3N2) strains have statistically significant connection with the years of isolation of the strains. The number of differences in amino acids at the epitopes of pairs of influenza strains increases as the difference between the years of isolation of strains increases. These results affirm that antigenic variants causing seasonal epidemics have partly resulted from changes in key amino acids on the antigenic sites (Both et al., 1983).

Further, with the results that years of isolation of strains have significant relationship with the amino acid substitutions, they suggest that regular substitutions of amino acids are to be expected with each passing year. Particularly, this is consistent with reports that indicate that antigenic variants emerge from previous strains in approximately two to three years in a process known as antigenic drift (World Health Organization, 1999). Thus, the study supports the significance of time as influenza viruses evolve into new strains of distinct antigenicity.

Moreover, the antigenic similarity between any pair of influenza strains coordinates with the time (year) in which each strain was isolated. This further supports the evidence that years of isolation of each new antigenic variant partly account for antigenic difference between vaccine strain and newly isolated strain. Consequently, with the progress in time, antigenic difference between a strain and its variants will increase. This is because the antigenic similarity between influenza strains has positively significant correlation with the years in which the strains were first characterized. In addition, new antigenically distinct strains which are observed in the same year are more similar compared to strains which were characterized in different years. The results of the modelling are consistent with the evolution pattern of influenza A (H3N2) which clusters into more homogeneous groups based on the antigenic similarity. More specifically, Smith and colleagues showed that the influenza A (H3N2) strains isolated from 1968 to 2002 form chronological clusters based on the years of characterization of the strains (Smith et al., 2004). This is supported by the evidence of antigenic relatedness produced among strains having the same years of isolation and those isolated in different years (Fig. 2).

Furthermore, the variations in antigenic relatedness (which have been linked to the years of isolation) are the results of the differences in amino acids located on the antigenic sites of the HA protein of the viruses. These changes in amino acids at the antigenic sites were found to be significantly linked to the years of isolations of the strains. Here, the rates of changes in the amino acids in the various sites differ. However, the results suggest that changes in antigenic sites A and B are less affected by the differences between the years in which the strains appeared compared to the other designated antigenic sites (C-E).

Interestingly, the statistical models make it possible to infer amino acid changes at the designated epitopes from the years of isolations of the influenza strains. Particularly, they reveal the expected rates of changes in amino acids located on the antigenic sites as the influenza virus evolves with time. Hence, this work further serves to potentially inform the design of vaccines with differential specificities for epitopes with respect to varying antigenic changes due to time alone. This is very relevant since it was suggested earlier that approaches to improve the effectiveness of vaccines could be achieved from limiting antibody-induction potential of certain epitopes (Garrity et al., 1997). In this study, these are the epitopes that are less affected when there is transition in years or influenza seasons.

5. Conclusion

This study has revealed a significant relationship between the years of isolation and the antigenic relatedness of influenza A (H3N2) strains. Thus, time-dependent antigenic variants may occur with the evolution of the virus. These findings were obtained from the analysis of statistical models used to investigate changes in amino acids of the HA protein of influenza A (H3N2) virus. In addition, the statistical models predicted important amino acid changes which may accompany future antigenic variants at rates consistent with the literature. Furthermore, time for observing specific antigenic changes in the evolution of the virus can be predicted from the models. These results have the potential

to complement vaccine strain recommendation and to support the global surveillance of the influenza virus.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.10.028.

Acknowledgement

The author would like to thank Wilfred Ndifon for the meaningful discussions on the evolution influenza virus.

Competing interests

There are no competing financial or other interests in relation to this work.

References

Both, G.W., Sleigh, M.J., Cox, N.J., Kendal, A.P., 1983. Antigenic Drift in Influenza Virus H3 Hemagglutinin from 1968 to 1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J. Virol. 48 (1), 52–60.

Brown, L.E., Murray, J.M., White, D.O., Jackson, D.C., 1990. An analysis of the properties of monoclonal antibodies directed to epitopes of influenza virus hemagglutinin. Arch. Virol. 114, 1–26.

Bush, R.M., Fitch, W.M., Bender, C.A., Cox, N.J., 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol. Biol. Evol. 16, 1457–1465.

Cameron, A.C., Trivedi, P.K., 1990. Regression-based tests for overdispersion in the Poisson model. J. Econ. 46, 347–364.

Carrat, F., Flahault, A., 2007. Influenza vaccine: the challenge of antigenic drift. Vaccine 25, 6852–6862.

Fitch, W.M., Bush, R.M., Bender, C.A., Cox, N.J., 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. Proc. Natl. Acad. Sci. U. S. A. 94 (15), 7712–7718.

Fitch, W.M., Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., 2000. Predicting the evolution of human influenza A. The Journal of Heredity 91 (3), 183–185.

Garrity, R.R., Rimmelzwaan, G., Minassian, A., Tsai, W.P., Lin, G., de Jong, J.J., Goudsmit, J., Nara, P.L., 1997. Refocusing neutralizing antibody response by targeted dampening of an immunodominant epitope. J. Immunol. 159, 279–289.

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

Holmes, E.C., et al., 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol. 3, e300.

Katz, J.M., Webster, R.G., 1989. Efficacy of inactivated influenza A virus (H3N2) vaccines grown in mammalian cells or embryonated eggs. J. Infect. Dis. 160, 191–199.

Kirkpatrick, E., Qiu, X., Wilson, P.C., Bahl, J., Krammer, F., 2018. The influenza virus hemagglutinin head evolves faster than the stalk domain. Sci. Rep. 8 (1), 10432.

Kodihalli, S., Justewicz, D.M., Gubareva, L.V., Webster, R.G., 1995. Selection of a single amino acid substitution in the hemagglutinin molecule by chicken eggs can render influenza A virus (H3) candidate vaccine ineffective. J. Virol. 69, 4888–4897.

Laver, W.G., Air, G.M., Webster, R.G., 1981. Mechanism of antigenic drift in influenza virus. Amino acid sequence changes in an antigenically active region of Hong Kong (H3N2) influenza virus hemagglutinin. J. Mol. Biol. 145, 339–361.

Layne, S.T., 2006. Human Influenza Surveillance: the demand to Expand. Emerg. Infect. Dis. 12 (4), 562–568.

Lee, M.S., Chen, J.S.E., 2004. Predicting antigenic variants of influenza A/H3N2 viruses. Emerg. Infect. Dis. 10 (8), 1385–1389.

Lee, Y., Nelder, J.A., 2000. The relationship between double-exponential families and extended quasi-likelihood families, with application to modeling Geissler's human sex ratio data. Appl. Stat. 49, 413–419.

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 (4), 505–512.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J., 2000. Molecular Cell Biology, 4th edition. W. H. Freeman, New York.

NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7–D19.

Ndifon, W., Wingreen, N.S., Levin, S.A., 2009a. Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines. Proc. Natl. Acad. Sci. U. S. A. 106 (21), 8701–8706.

Ndifon, W., Dushoff, J., Levin, S.A., 2009b. On the use of hemagglutination-inhibition for influenza surveillance: surveillance data are predictive of influenza vaccine effectiveness. Vaccine 27, 2447–2452.

Nobusawa, E., Omagari, K., Nakajima, S., Nakajima, K., 2012. Reactivity of human convalescent sera with influenza virus hemagglutinin protein mutants at antigenic site A. Microbiol. Immunol. 56, 99–106.

Popova, L., Smith, K., West, A.H., Wilson, P.C., James, J.A., Thompson, L.F., Air, G.M., 2012. Immunodominance of Antigenic Site B over Site A of Hemagglutinin of recent H3N2 Influenza Viruses. PLoS One 7 (7), e41895. https://doi.org/10.1371/journal.pone.0041895.

Sato, K., Morishita, T., Nobusawa, E., Tonegawa, K., Sakae, K., et al., 2004. Amino acid change on the antigenic region B1 of H3 haemagglutinin may be a trigger for the emergence of drift strain of influenza A virus. Epidemiol. Infect. 132, 399–406.

Sievers, F., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Sleigh, M.J., Both, G.W., Underwood, P.A., Bender, V.J., 1981. Antigenic Drift in Hemagglutinin of the Hong Kong Influneza Subtype: Correlation of Amino Acid changes with Alterations in Viral Antigenicity. J. Virol. 37 (3), 845–853.

Smith, C.A., Barnett, B.C., Thomas, D.B., Temoltzin-Palacios, F., 1991. Structural assignment of novel and immunodominant antigenic sites in the neutralizing antibody response of CBA/Ca mice to influenza hemagglutinin. J. Exp. Med. 173, 953–959.

Smith, D.J., Lapedes, A.S., de Jong, J.C., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D., Fouchier, R.A., 2004. Mapping the Antigenic and Genetic Evolution of Influenza Virus. Science 305, 371–376. https://doi.org/10.1126/science.1097211.

Temoltzin-Palacios, F., Thomas, D.B., 1994. Modulation of immunodominant sites in influenza hemagglutinin compromise antigenic variation and select receptor-binding variant viruses. J. Exp. Med. 179, 1719–1724.

Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-poisson vs negative binomial regression: how should we model overdispersed count data? Ecology 88 (11), 2766–2772.

Webster, R.G., Laver, W.G., Air, G.M., Schild, G.C., 1982. Molecular mechanisms of variation in influenza viruses. Nature 296, 115–121.

Webster, R., Cox, N., Stohr, K., 2002. WHO Manual on Animal Influenza Diagnosis and Surveillance. In: WHO/CDS/CSR/NCS/2002.5, Rev.1, . http://www.who.int/csr/resources/publications/influenza/en/whocdscsrncs20025rev.pdf (Last date accessed: February 2, 2018).

Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61 (3), 439–447.

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., Cox, N.J., 1990. Structural basis of immune recognition of influenza virus hemagglutinin. Annu. Rev. Immunol. 8, 737–771.

World Health Organization, 1999. Influenza Pandemic Plan: the role of WHO and guidelines for national and regional Planning World Health Organization. In: WHO/CDS/CSR/EDC/99.1, . http://apps.who.int/iris/bitstream/10665/66155/1/WHO_CDS_CSR_EDC_99.1.pdf (Last date accessed: February 2, 2018).

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 (18), 2309–2317.

Zelterman, D., 2006. Models for Discrete Data. Oxford University Press Inc, New York, USA.