Computers in Biology and Medicine 52 (2014) 35–40

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

SimFlu: A simulation tool for predicting the variation pattern of influenza A virus

Insung Ahn [a,1,2], Ha-Yeon Kim [b,c,1], Sunghoon Jung [b], Ji-Hae Lee [b,d], Hyeon Seok Son [b,d,*]

[a] National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information, 245 Daehangno, Yuseong-Gu, Daejeon 305-806, Republic of Korea
[b] Laboratory of Computational Biology & Bioinformatics, Institute of Health and Environment, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea
[c] National Institute of Biological Resources, Hwangyeong-ro 42, Seo-gu, Incheon 404-708, Republic of Korea
[d] Graduate Program in Bioinformatics, College of Natural Science, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 18 February 2014
Accepted 5 June 2014

Keywords:
Influenza A virus
Codon variation
Pattern prediction
Simulation
Bioinformatics

ABSTRACT

Since the first pandemic outbreak of avian influenza A virus (H5N1 subtype) in 1997, the National Center for Biotechnology Information (NCBI) has provided a large number of influenza virus sequences with well-organized annotations. Using the time-series sequences of influenza A viruses, we developed a simulation tool for influenza virus, named SimFlu, to predict possible future variants of influenza viruses. SimFlu can create variants from a seed nucleotide sequence of influenza A virus using the codon variation parameters included in the SimFlu package. The SimFlu library provides pre-calculated codon variation parameters for the H1N1, H3N2, and H5N1 subtypes of influenza A virus isolated from 2000 to 2011, allowing the users to simulate their own nucleotide sequences by selecting their preferred parameter options. SimFlu supports three operating systems – Windows, Linux, and Mac OS X. SimFlu is publicly available at http://lcbb.snu.ac.kr/simflu.

1. Introduction

Influenza virus causes respiratory illness and several pandemic outbreaks globally each year. Spanish flu (A/H1N1) in 1918, Asian flu (A/H2N2) in 1957, Hong Kong flu (A/H3N2) in 1968, and swine flu (A/H1N1) in 2009 are representative influenza viruses with pandemic outbreaks that have caused many deaths [1,2]. Influenza virus is a negative-sense RNA virus with haemagglutinin (HA) and neuraminidase (NA) genes, the products of which are present on the viral surface, as well as six internal genes (M1, M2, NP, PA, PB1, PB2) [3]. Influenza virus has a segmented genome, which facilitates genetic reassortment to produce a new subtype [4,5]. The new viral subtype often shows improved evasion of the host immune response, causing more rapid and efficient infections, which can increase viral spread and cause pandemic outbreaks [1,4,6]. Further studies are required to characterize past influenza pandemics, which have occurred periodically.

In this report, we introduce an easy-to-use simulation tool for influenza virus, named SimFlu, which can simulate possible variants of influenza A virus genes using codon variation parameters calculated from available gene sequences. The current version of SimFlu supports variation parameters for the H1N1, H3N2, and H5N1 subtypes of influenza A virus isolated from 2000 to 2011.

The SimFlu tool can be used for computational simulations based on the changing codon patterns over time in the influenza virus nucleotide sequences. Since the pandemic outbreak of the novel influenza A virus (H1N1 subtype) in 2009, many groups have attempted to develop simulation tools to model the large-scale spread of influenza virus [2,7–11]. However, the majority of these tools are designed to simulate the spreading patterns across target cities or countries, instead of considering the changing patterns of viral genomes. For example, FluSurge and FluAid, developed by the U.S. Centers for Disease Control and Prevention (CDC), are pandemic flu preparation tools based on a static modeling approach, while FluTE is an individual-based dynamic modeling platform that creates synthetic populations based on typical American communities, providing information to reduce epidemic peaks, illness, and mortality [7]. Some Asian countries, such as Cambodia, Indonesia, Lao PDR,

* Corresponding author at: Laboratory of Computational Biology & Bioinformatics, Institute of Health and Environment, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. Tel.: +82 2 880 2746; fax: +82 2 745 9104.
E-mail address: hss2003@snu.ac.kr (H.S. Son).
[1] These authors contributed equally to this work.
[2] Co-corresponding author.

http://dx.doi.org/10.1016/j.compbiomed.2014.06.003

Table 1
List of simulation tools using influenza virus-related data.

Tools	Description	URL
FluSurge 2.0	For estimating the number of hospitalizations and deaths of an influenza pandemic	http://www.cdc.gov/flu/pandemic-resources/tools/flusurge.htm
FluAid 2.0	For preparing for the next influenza pandemic by providing estimates of potential impact specific to their locality	http://www.cdc.gov/flu/pandemic-resources/tools/fluaid.htm
Community Flu 2.0	For simulating the spread of influenza through a model community using the impact of a variety of potential interventions	http://www.cdc.gov/flu/pandemic-resources/tools/communityflu.htm
FluTE	For simulating the spread of influenza across a large population	https://www.epimodels.org/midas/flute.do
AsiaFluCap simulator	For assessing health system capacity for responding to various pandemic influenza scenarios	http://www.cdprg.org/asiaflucap-simulator.php
SEARUMS	For predicting the routes of avian influenza infection spread	http://www.searums.org

Taiwan, Thailand, and Vietnam, undertook the AsiaFluCap project from 2008 to 2011, and developed a simulation tool named the AsiaFluCap simulator to estimate the impact on health care resource capacity during pandemics [11] (Table 1). The feature that most distinguishes SimFlu from other simulation tools is that it can simulate possible variants from the user's seed sequence. With this program, users can generate future variant sequences on a case-by-case basis. Because the SimFlu program provides the pre-calculated codon variation parameters as a library, it is not necessary to perform complicated and time-consuming analyses.

2. Methods

2.1. Development of the SimFlu program

We compared the codon usage patterns among various subtypes of influenza A virus, and observed directional changes over time. In our previous study, we analyzed the codon usage patterns of the HA genes of the human-origin influenza A virus subtype H3N2 (A/H3N2) isolated between 1999 and 2005, and found that the relative synonymous codon usage (RSCU) values of the HA gene tended to be grouped into three different types [12]. To confirm this time-dependent change, we conducted a more intensive study by analyzing the RSCU values and G+C contents of eight major genes of A/H3N2. We observed that the directional changes in RSCU values that occurred in the HA, NA, polymerase basic protein 1 (PB1), and polymerase basic protein 2 (PB2) genes were significantly correlated with variations in the G+C contents of codons [13]. Moreover, the proportion of codons substituted for the same codon within each synonymous codon group, namely the exactly matched codons (EMC), in the internal genes NP, PA, PB1, and PB2 gradually decreased over time, indicating that another rule of variation may influence changes in the synonymous codon usage patterns over time. Typically, influenza A viruses are known to randomly change their antigenic codes, making it impossible to predict future variants. Nonetheless, in our previous study, we observed directional changes of A/H3N2s based on the codon unit, and those changes were observed not only in the surface proteins, HA and NA, but also in the replication-related proteins, PB1 and PB2 [14]. Therefore, we hypothesized that there could be a rule of variation over the long-term evolutionary process of influenza viruses, although their genetic codes change randomly. The majority of pandemic outbreaks are known to occur accidentally. However, we could predict symptoms of the 2009 pandemic outbreak 1 year before the pandemic in our follow-up study that simulated variants of H1N1 influenza A viruses from the seed sequence of A/Taiwan/929/99 (DQ249252) using the positively accumulated variation patterns from 1999 to 2008 [13]. Thus, we developed the simulation tool SimFlu, which can predict variation patterns of the influenza A virus. The SimFlu program is freely available under the PRODUCT page on the SimFlu homepage (http://lcbb.snu.ac.kr/simflu/). Since SimFlu was developed for the three major computer operating systems (OSs) – *Windows*, *Linux*, and *Macintosh OS X* – users can download the suitable SimFlu package for their OS of choice. Three major modes of the SimFlu interface are available: interactive mode; command file mode; and option mode. Users can select the mode that suits their requirements, while the manner of uses for each mode is explained on the SimFlu homepage.

2.2. SimFlu library

The SimFlu library consists of pre-calculated variation parameters of influenza A virus genes between two different year-of-isolation groups. The source nucleotides of these library files were collected from the Influenza Virus Resource of the National Center for Biotechnology Information (NCBI) [15]. We collected messenger RNAs (mRNAs) of 10 genes (HA, NA, NP, PA, PB1, PB2, M1, M2, NS1, and NS2) for three influenza A virus subtypes (H1N1, H3N2, and H5N1) in the library of the current version of SimFlu (version 1.0). Target years-of-isolation were from 2000 to 2011, and target host species were human and swine for H1N1 and H3N2, and human and avian for H5N1. The source sequences used in this study included 13,607 H1N1 sequences (10,743 of human-origin and 2864 of swine-origin), 12,225 H3N2 sequences (11,195 of human-origin and 1030 of swine-origin), and 10,536 H5N1 sequences (1274 of human-origin and 9262 of avian-origin). The first step of the calculation was to perform multiple sequence alignments (MSAs) among all possible pairs of years-of-isolation between 2000 and 2011. SimFlu provides two types of selection options in the "Library Settings" step: type A and B. If the user chooses 'type A', the time intervals between the initial (time t) and final target (time t') years would be 1 year, while in 'type B' the initial target year is fixed and the final target year is increased. Each pair of target years should be aligned using the MSA tools, and the aligned output sequences are divided into two files according to their years-of-isolation for further processing. All possible codon variations between a pair of MSAs results in 'gap' information, and codon variations are counted in the next step. First, each sequence (Sequence #1, #2, …, #n) in the "Initial year-of-isolation (t)" is compared with the sequences (Sequence #1', #2',…, #p) in the "Final year-of-isolation (t')". Each codon of group t is compared with that of the group t', and the total of $n \times p$ comparisons are conducted. All counted codon variations are saved in a 61×61 matrix named the "codon variation matrix (CVM)", which is converted into the codon transition matrix (CTM) using a Markov model [13,16]. The names of the calculated CTMs are encoded and saved in the lib folder of SimFlu (Fig. 1).

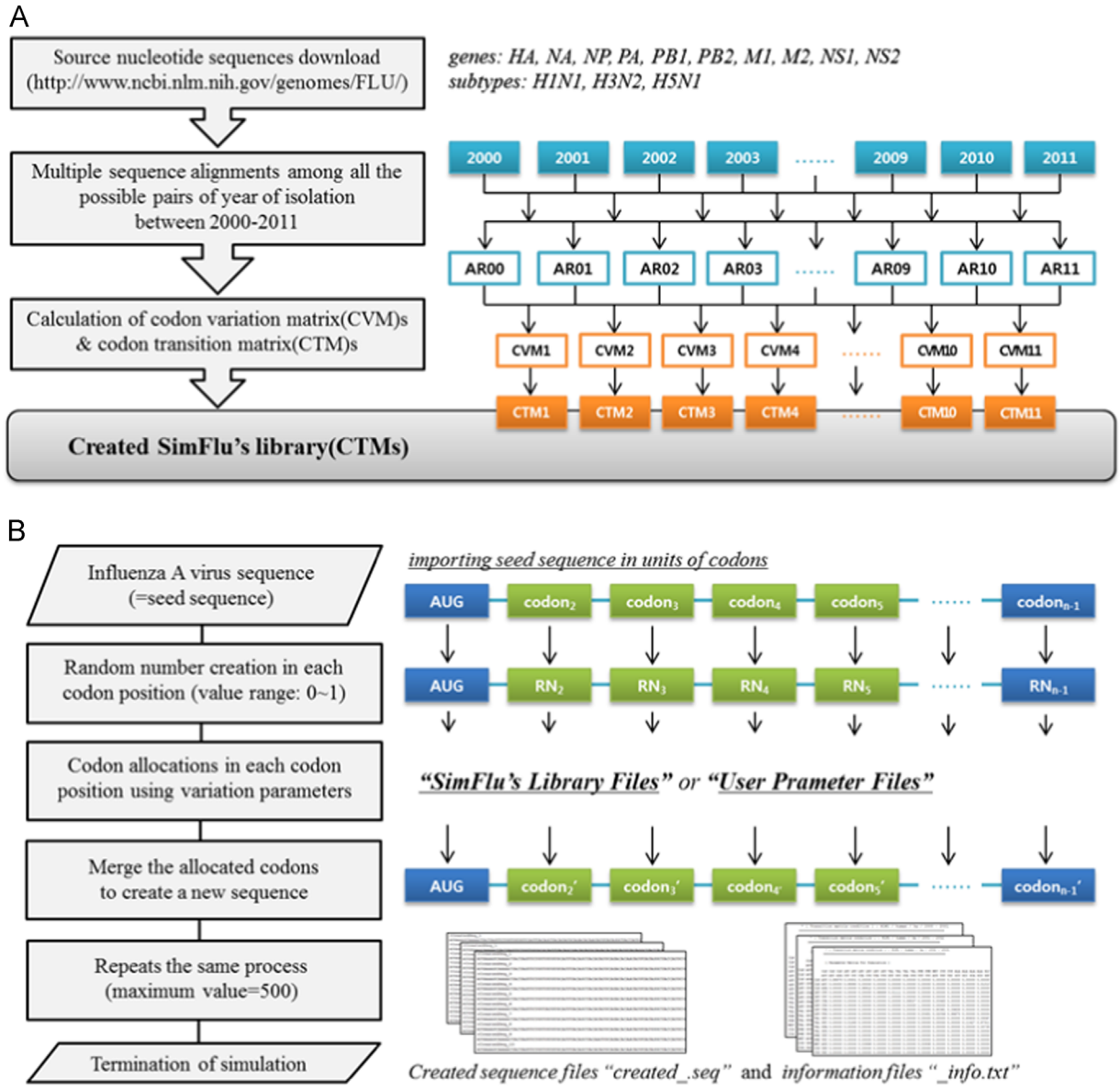

Fig. 1. Overview of SimFlu simulation. (A) SimFlu program provides pre-calculated variation parameters of influenza A virus genes between two different year-of-isolation groups through the library. The source nucleotide sequences of these library files were collected from the Influenza Virus Resource of NCBI (http://www.ncbi.nlm.nih.gov/genomes/FLU/). The first step in calculating SimFlu library files is to perform multiple sequence alignments (MSAs) among all possible pairs of years-of-isolation between 2000 and 2011. Each pair of target years should be aligned, and the output files are divided into two files according to their years-of-isolation. In the second step, all possible codon variations between two MSA result files that contain the 'gap' information are counted. Each codon in each sequence region along the aligned result of the first year of isolation is compared with that in the second year of isolation result; the counted variation are saved in the 61×61 matrix termed the 'codon variation matrix' (CVM). In the final step, all CVMs are converted into the codon transition matrix (CTM) for use in the SimFlu library. (B) The SimFlu simulation process begins with importing the seed sequence according to the units of codons. In the next step, SimFlu generates random numbers between 0 and 1, excluding the start and termination codons. SimFlu converts each random number into a new codon, which is assigned based on the probability of the variation parameters in the SimFlu library or from user input information. This process is repeated as many times as desired by the user (maximum, 500).

2.3. SimFlu simulation algorithm

SimFlu is a simulation tool that predicts future nucleotide sequences from a real influenza A virus sequence (seed sequence) using the codon variation parameters provided in the SimFlu library files. SimFlu first reads all codon units in each codon position from a seed sequence. SimFlu then generates random numbers, excluding the start and termination codons. SimFlu then converts each random number into a new codon, which is assigned based on the probability of the variation parameters. This process is repeated until SimFlu finishes creating the sequences, up to a maximum of 500 times. The output files of SimFlu can be divided into two categories: created sequences and information files. To minimize user workload, SimFlu automatically assigns the names of sequence and information files. All simulated sequences are saved in FASTA format, and each title is automatically assigned with the number of the generated sequence with the prefix "CreatedSeq_". The information file (_info.txt) is composed of three parts: description, simulation conditions, and parameters. In the simulation conditions section, SimFlu provides details including the date, file paths to the simulated result files and seed sequence files, and the setup information for the SimFlu library file. SimFlu provides these simulation details in the output files, so that users are not required to remember this information. At the end of the information file, the user can review the 61×61 transition matrix of codon

variations obtained from the parameters specified by user input or from the SimFlu library files.

3. Validation of SimFlu simulated sequences

Validation of SimFlu was performed in two ways, using simulated sequences. Influenza A virus (H1N1 subtype) sequences were downloaded from NCBI and validated for the simulation method using our SimFlu application (Fig. 2). HA gene sequences from two influenza A virus (H1N1 subtype) strains isolated in 2000, A/Bangkok/163/2000 and A/South Australia/25/2000 (GenBank accession nos. CY125108 and ABK40028, respectively), were downloaded and used as seed sequences. First, we generated 500 variant sequences using the HA A/Bangkok/163/2000 gene as a seed sequence. The variation parameters between 2000 and 2009 (type B option) were used in SimFlu's library. To validate the actual effect of the simulation, we performed similarity searches against the complete genomic database of influenza virus A genes using the BLASTN program [17]. According to the results of our pre-analysis using the BLASTN program, the seed sequence displayed high similarities with the target sequences within ± 2 years-of-isolation, so we conducted the sequence validation using an enlarged span of the target year ± 2 years. Sequences were randomly collected from available sequences and BLASTN was performed. A total of 1000 sequences were obtained from this BLASTN analysis. Target years were defined as each year during the 10 years after the seed sequence's year-of-isolation, from 2000 to 2009, and each group was defined with a span of ± 2 years. Sequence identity percentages between the seed sequence and the target group sequences were determined and averaged for the BLAST results (Fig. 2A). According to the results, sequence similarity between the seed sequence and target sequences decreased only slightly over time as the distance from the seed sequence's year-of-isolation increased. However, target sequences gradually evolved into new nucleotide sequences that were distantly related to the seed sequence. These results suggest that the SimFlu program is effective in simulating viral gene evolution by searching for variation patterns in the influenza virus and predicting the evolutional direction.

Next, a validation test was performed using the HA A/South Australia/25/2000 gene as the seed sequence. Three simulations were performed and for each simulation, 100 sequences were generated based on the variation parameters for the years of the simulation: from 2000 to 2003 (P2000–2003), 2000 to 2006 (P2000–2006), and 2000 to 2009 (P2000–2009). We blasted the generated sequences against the H1N1 subtype sequences isolated from 2000 to 2013 in the NCBI Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html); the results are shown in Fig. 2B. The total numbers of BLAST hits were 31 for the seed sequence, 41 for

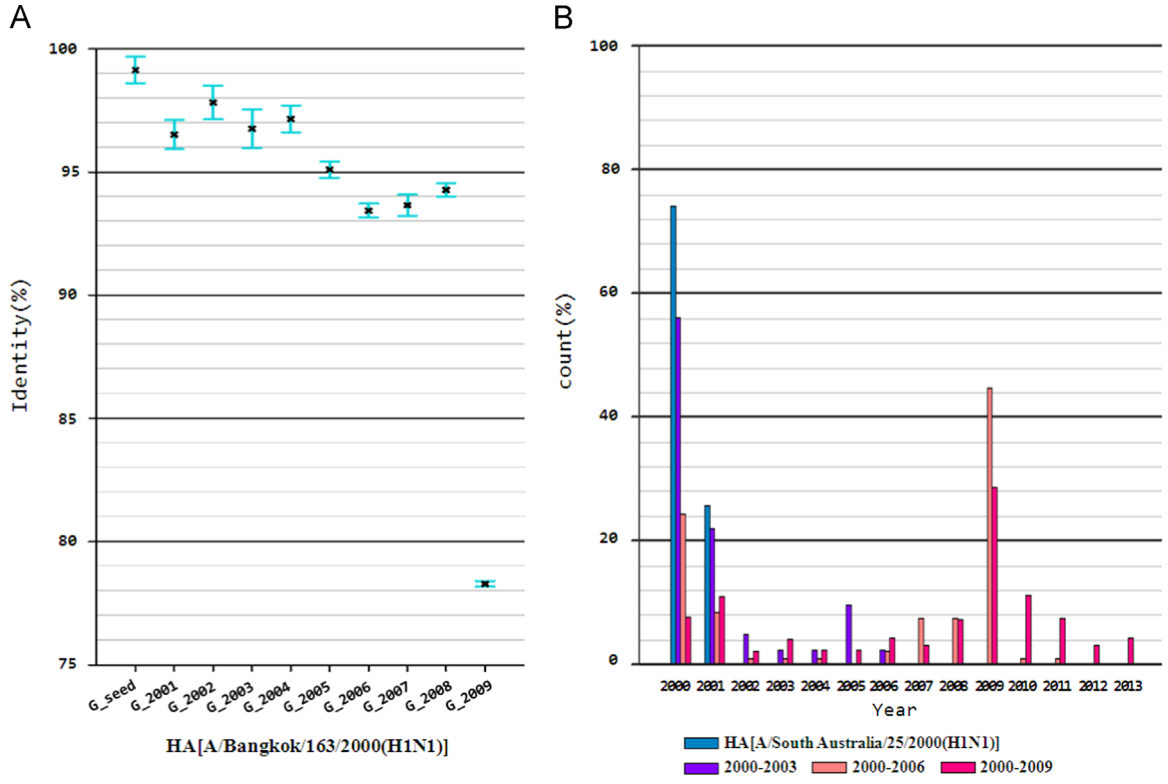

Fig. 2. Validation of the SimFlu simulated sequence by BLAST searching. HA seed sequences from influenza A virus (H1N1 subtype) strains A/Bangkok/163/2000 and A/South Australia/25/2000 downloaded from NCBI were used for the simulation to validate the SimFlu application. (A) The simulated sequences were generated for each year and the results of the simulation were tested using BLAST. Groups (named G_groupyear) were generated for the following target years: G_seed (2000 ± 2); G_2001 (2001 ± 2); G_2002 (2002 ± 2); G_2003 (2003 ± 2); G_2004 (2004 ± 2); G_2005 (2005 ± 2); G_2006 (2006 ± 2); G_2007 (2007 ± 2); G_2008 (2008 ± 2); and G_2009 (2009 ± 2). The graph shows the simulation results as percentage identity distributions derived from comparing the initial seed sequence to the sequences in the corresponding group. As shown in the figure, the sequences simulated for years closer to the seed year shared greater identity ($\geq 90\%$) with the seed sequence, while sequences with lower shared identity (75–80%) appeared over time. (B) Three sets of simulations were performed (S1, S2 and S3) and for each simulation, 100 sequences were generated based on the variation parameters of the simulation years: from 2000 to 2003 (S1), 2000 to 2006 (S2), and 2000 to 2009 (S3). The simulation results were tested using BLAST with the NCBI Influenza Virus Resources (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html); H1N1 subtype sequences from 2000 to 2013 were extracted from the BLAST results and the percentages of count (defined below) for each year and each group were calculated; BLAST searches were conducted to verify the direction of simulation results and BLAST hits for each simulation (C_{S1}, C_{S2} and C_{S3}) were counted, using the default e-value of 10. Thus for each year, the percentage count (PC) of simulation *Sn* was calculated with the formula: $PC_{yearSn} = (C_{yearSn}/C_{Sn}) \times 100$, where C_{Sn} is the total number of hits for simulation *Sn* ($n = 1, 2, 3$) and C_{yearSn} is the number of hits for simulation *Sn*, for a specific year (e.g., for S1 simulation, the percent count for the year 2000 is $PC_{2000S1} = (C_{2000S1}/C_{S1}) \times 100 = (23/41) \times 100 = 56.09$). As shown in the figure, simulation results varied among the variation parameters. The variation parameter for years from 2000 to 2009 better predicted the sequences for years later than 2009 than did the variation parameters for years from 2000 to 2003. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

P2000–2003, 94 for P2000–2006, and 652 for P2000–2009. The calculation method for the percentages is described in the figure caption, along with an example. Based on the results of the simulations, the variant sequences that were artificially generated by the SimFlu program showed directional changes according to their parameters. The variant sequences generated by the P2000–2009 (red color) showed the highest similarities with the influenza A virus sequences isolated in 2009, and also showed similar patterns with the sequences isolated from 2010 to 2013, while the variants generated by P2000–2003 (blue color) showed patterns most similar to the sequences isolated ca. 2000. The variants generated by P2000–2006 demonstrated intermediate characteristics. In conclusion, we observed that the sequences artificially generated by SimFlu were successfully evolved according to their parameter groups. SimFlu can generate possible variants based on a user's seed sequences with pre-calculated codon variation parameters; this reduces complicated simulation calculations and may be one of the reasons for the plausible results generated by SimFlu, as presented in Fig. 2B.

4. Conclusion

Studies designed to predict new emergences of influenza viruses are important, as they may provide insights on potentially hazardous strains. A previous study found that a rapidly evolving subset of codons appeared through positive selection in the HA1 domain of the hemagglutinin gene of human influenza A subtype H3 [18]. Such monitoring studies, detecting changes in positively selected codons of influenza viruses, can provide useful information about evolution of influenza viruses [18]. In addition, with the development of advanced sequencing techniques, massive amounts of sequence data have accumulated for influenza viruses; with various analysis methods, these data can be utilized for new research studies. In another study based on bioinformatics, the authors performed analyses of mono-, di-, tri-, and tetra-nucleotide compositions with all genome sequences of influenza A and B viruses using a Batch-Learning Self-Organizing Map (BLSOM) method and found host-dependent clustering of influenza viruses [19]. Our study has developed a simulation method based on codon variation patterns though analysis of vast quantities of influenza virus sequence. In a previous study, we examined the evolutionary changes of influenza A viruses using codon sequences, and observed that not only the envelope proteins but also the polymerase basic proteins showed the directional changes over time [12]. We therefore focused on all possible variations of codons between target groups of influenza A viruses to calculate the simulation parameters for the SimFlu program.

We are currently developing a message-passing interface (MPI) version of SimFlu to conduct simulations using single codon variation parameters and multiple matrices calculated from each genomic region along the target nucleotide. The simulation parameters of the current SimFlu version (1.0) were calculated by comparing the codon substitution patterns between two target groups, and all variations were accumulated in one matrix (61×61). However, the majority of influenza A virus genes are known to have highly conserved regions along the sequence; these regions are used as antigenic epitopes in vaccine production [20–22]. Therefore, prior to conducting multi-level simulations, SimFlu must generate codon variation parameters for each genomic region in its library. Hence, it would be valuable to develop an analytical tool to calculate multi-level codon variation parameters along the genomic regions. This MPI-version of SimFlu and the related analytical tools may be installed and tested on the high-performance computing infrastructure at the National Institute of Supercomputing and Networking, and may later be made available to the public.

5. Summary

In our previous study, we compared the codon usage patterns among various subtypes of influenza A virus, and observed directional changes over time. Using the time-series sequences of influenza A viruses, we developed a simulation tool for influenza virus, named SimFlu, to predict possible future variants of influenza viruses. SimFlu can create variants from a seed nucleotide sequence of influenza A virus using the codon variation parameters included in the SimFlu package. The SimFlu library provides pre-calculated codon variation parameters for the H1N1, H3N2, and H5N1 subtypes of influenza A viruses isolated from 2000 to 2011, allowing the users to simulate their own nucleotide sequences by selecting their preferred parameter options.

Validation of SimFlu was performed in two ways using simulated sequences. HA gene segments from influenza A virus (H1N1 subtype) strains A/Bangkok/163/2000 and A/South Australia/25/2000 were used as seed sequences. According to the results, sequence similarity between the seed sequence and target sequences decreased only slightly over time as the distance from the seed sequence's year-of-isolation increased. However, target sequences gradually evolved into new nucleotide sequences that were distantly related to the seed sequence. SimFlu supports three operating systems: *Windows*, *Linux*, and *Mac OS X*. We are currently developing a message-passing interface (MPI) version of SimFlu to conduct simulations using single codon variation parameters and multiple matrices calculated from each genomic region along the target nucleotide. SimFlu is publicly available at http://lcbb.snu.ac.kr/simflu.

Conflict of interest statement

None declared.

Acknowledgments

The authors thank Prof. Moon Ahn at Korea University for his support on this article. This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MEST 2012008344). This research was also supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF), funded by the Ministry of Science, ICT & Future Planning (2012M3A9D1054622). Support from the National Institute of Supercomputing and Networking at the Korea Institute of Science and Technology Information (KISTI K-14-L01-C02-S04) is gratefully acknowledged.

References

[1] J.C. de Jong, E.C. Claas, A.D. Osterhaus, R.G. Webster, W.L. Lim, A pandemic warning? Nature 389 (1997) 554.
[2] F.S. Dawood, S. Jain, L. Finelli, M.W. Shaw, S. Lindstrom, R.J. Garten, L.V. Gubareva, X. Xu, C.B. Bridges, T.M. Uyeki, Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team, Emergence of a novel swine-origin influenza A (H1N1) virus in humans, N. Engl. J. Med. 360 (2009) 2605–2615.
[3] N. Marshall, L. Priyamvada, Z. Ende, J. Steel, A.C. Lowen, Influenza virus reassortment occurs with high frequency in the absence of segment mismatch, PLoS Pathog. 9 (2013) e1003421.
[4] R.G. Webster, W.J. Bean, O.T. Gorman, T.M. Chambers, Y. Kawaoka, Evolution and ecology of influenza A viruses, Microbiol. Rev. 56 (1992) 152–179.
[5] R.A. Fouchier, V. Munster, A. Wallensten, T.M. Bestebroer, S. Herfst, D. Smith, G.F. Rimmelzwaan, B. Olsen, A.D. Osterhaus, Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls, J. Virol. 79 (2005) 2814–2822.
[6] T. Horimoto, Y. Kawaoka, Pandemic threat posed by avian influenza A viruses, Clin. Microbiol. Rev. 14 (2001) 129–149.

[7] D.L. Chao, M.E. Halloran, V.J. Obenchain, I.M. Longini Jr., FluTE, a publicly available stochastic influenza epidemic simulation model, PLoS Comput. Biol. 6 (2010) e1000656.
[8] B. Bonte, J.D. Mathias, R. Duboz, Moment approximation of infection dynamics in a population of moving hosts, PLoS One 7 (2012) e51760.
[9] K.C. Chong, B.C. Ying Zee, Modeling the impact of air, sea, and land travel restrictions supplemented by other interventions on the emergence of a new influenza pandemic virus, BMC Infect. Dis. 12 (2012) 309.
[10] E.O. Nsoesie, R.J. Beckman, M.V. Marathe, Sensitivity analysis of an individual-based model for simulation of influenza epidemics, PLoS One 7 (2012) e45414.
[11] M.L. Stein, J.W. Rudge, R. Coker, C. van der Weijden, R. Krumkamp, P. Hanvoravongchai, I. Chavez, W. Putthasri, B. Phommasack, W. Adisasmito, S. Touch, M. Sat le, Y.C. Hsu, M. Kretzschmar, A. Timen, Development of a resource modelling tool to support decision makers in pandemic influenza preparedness: the AsiaFluCap Simulator, BMC Public Health 12 (2012) 870.
[12] I. Ahn, H.S. Son, Comparative study of the hemagglutinin and neuraminidase genes of influenza A virus H3N2, H9N2, and H5N1 subtypes using bioinformatics techniques, Can. J. Microbiol. 53 (2007) 830–839.
[13] I. Ahn, H.S. Son, Computational model for analyzing the evolutionary patterns of the neuraminidase gene of influenza A/H1N1, Comput. Biol. Chem. 36 (2012) 23–30.
[14] I. Ahn, H.S. Son, Evolutionary analysis of human-origin influenza A virus (H3N2) genes associated with the codon usage patterns since 1993, Virus Genes 44 (2012) 198–206.
[15] L. Zaslavsky, Y. Bao, T.A. Tatusova, Visualization of large influenza virus sequence datasets using adaptively aggregated trees with sampling-based subscale representation, BMC Bioinform. 9 (2008) 237.
[16] W.J. Stewart, Probability, Markov Chains, Queues, and Simulation: the Mathematical Basis of Performance Modeling, first ed., Princeton University Press, New Jersey, 2009.
[17] L. Cummings, L. Riley, L. Black, A. Souvorov, S. Resenchuk, I. Dondoshansky, T. Tatusova, Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes, FEMS Microbiol. Lett. 216 (2002) 133–138.
[18] R.M. Bush, C.A. Bender, K. Subbarao, N.J. Cox, W.M. Fitch, Predicting the evolution of human influenza A, Science 286 (1999) 1921–1925.
[19] Y. Iwasaki, T. Abe, Y. Wada, K. Wada, T. Ikemura, Novel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains, BMC Infect. Dis. 13 (2013) 386.
[20] P. Huang, S. Yu, C. Wu, L. Liang, Highly conserved antigenic epitope regions of hemagglutinin and neuraminidase genes between 2009 H1N1 and seasonal H1N1 influenza: vaccine considerations, J. Transl. Med. 11 (2013) 47.
[21] A. Vitelli, M.R. Quirion, C.Y. Lo, J.A. Misplon, A.K. Grabowska, A. Pierantoni, V. Ammendola, G.E. Price, M.R. Soboleski, R. Cortese, S. Colloca, A. Nicosia, S. L. Epstein, Vaccination to conserved influenza antigens in mice using a novel simian adenovirus vector, PanAd3, derived from the bonobo *Pan paniscus*, PLoS One 8 (2013) e55435.
[22] J. Yang, E. James, T.J. Gates, J.H. Delong, R.E. Lafond, U. Malhotra, W.W. Kwok, CD4+ T cells recognize unique and conserved 2009 H1N1 influenza hemagglutinin epitopes after natural infection and vaccination, Int. Immunol. 25 (2013) 447–457.