Reconstruction of molecular evolution of human influenza A H1N1/2009 virus in Iran and neighboring countries

Nima Ghalekhani¹, Saied Bokaie², Sana Eybpoosh³, Hesameddin Akbarein², and Hamid Sharifi⁴

 ¹2. HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences
²1. Divisions of Epidemiology & Zoonoses, Department of Food Hygiene and Quality Control, Faculty of Veterinary Medicine, University of Tehran, Iran
³3. Department of Epidemiology and Biostatistics, Research Centre for Emerging and Reemerging Infectious Diseases, Pasteur Institute of Iran, Tehran, Iran
⁴2. HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

April 22, 2022

Abstract

Background In this study, the time and path of transmission of H1N1 serotype influenza A viruses in Iran and neighboring countries have been investigated by using Bayesian phylogeography analysis on the sequences extracted from the gene bank. Methods We obtained all hemagglutinin (HA) and neuraminidase (NA) nucleotide sequences of influenza H1N1 available up to December 25, 2020, from Iran and its neighboring countries (i.e., Pakistan, Afghanistan, Turkmenistan, Armenia, Azerbaijan, Turkey, and Iraq). We also performed a Bayesian Markov chain Monte Carlo method to infer the evolutionary dynamic and the most recent common ancestor for the HA and NA sequences. Results Based on the extracted sequences, the age of emergence of H1N1 influenza virus serotype was older in Iran compared to neighboring countries, and Tehran had a key role and epicenter of transmission to other cities within Iran. The mean time of the most recent common ancestor of H1N1 viruses was 1989 (95% HPD: 1980-1994) for HA and NA as well. Conclusions Along with ordinary measures like resource management, diagnostic approaches, and preparedness to fight against viruses that were in place, continuous monitoring, and screening of H1N1 serotype influenza virus in the country, especially by implementation of feasible, effective, and innovative measures at border line should be initiated and identified gaps and shortage that should be a priority for virus control. It is also important for countries to have a regional monitoring program in addition to internal monitoring programs, as well as to start a virus molecular care program.

Reconstruction of molecular evolution of human influenza A H1N1/2009 virus in Iran and neighboring countries

Running Title: Molecular evolution of A H1N1 virus in Iran and neighbors

Nima Ghalekhani^{1,2}, Saied Bokaie¹, Sana Eybpoosh^{3*}, Hesameddin Akbarein¹, Hamid Sharifi²

- 1. Divisions of Epidemiology & Zoonoses, Department of Food Hygiene and Quality Control, Faculty of Veterinary Medicine, University of Tehran, Iran
- 2. HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

3. Department of Epidemiology and Biostatistics, Research Centre for Emerging and Reemerging Infectious Diseases, Pasteur Institute of Iran, Tehran, Iran

Corresponding Author: Sana Eybpoosh

E-mail: sana.eybpoosh@gmail.com

Acknowledgment

We wish to acknowledge to Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran, especially Reza Sheikhzadeh (IT Manager) for his kind collaboration to access server computer to conduct computational analysis.

Abstract

Background

In this study, the time and path of transmission of H1N1 serotype influenza A viruses in Iran and neighboring countries have been investigated by using Bayesian phylogeography analysis on the sequences extracted from the gene bank.

Methods

We obtained all hemagglutinin (HA) and neuraminidase (NA) nucleotide sequences of influenza H1N1 available up to December 25, 2020, from Iran and its neighboring countries (i.e., Pakistan, Afghanistan, Turkmenistan, Armenia, Azerbaijan, Turkey, and Iraq). We also performed a Bayesian Markov chain Monte Carlo method to infer the evolutionary dynamic and the most recent common ancestor for the HA and NA sequences.

Results

Based on the extracted sequences, the age of emergence of H1N1 influenza virus serotype was older in Iran compared to neighboring countries, and Tehran had a key role and epicenter of transmission to other cities within Iran. The mean time of the most recent common ancestor of H1N1 viruses was 1989 (95% HPD: 1980-1994) for HA and NA as well.

Conclusions

Along with ordinary measures like resource management, diagnostic approaches, and preparedness to fight against viruses that were in place, continuous monitoring, and screening of H1N1 serotype influenza virus in the country, especially by implementation of feasible, effective, and innovative measures at border line should be initiated and identified gaps and shortage that should be a priority for virus control. It is also important for countries to have a regional monitoring program in addition to internal monitoring programs, as well as to start a virus molecular care program.

Keywords: Influenza virus; H1N1; Bayesian phylogeographic; Iran

Introduction

Influenza is still a major cause of health disorders and mortality worldwide and affects a large portion of the human population each year, this potential could cause epidemics in humans, such as H1N1 virus pandemic in 2009. In addition, many species of animals, including birds, pigs, horses, and dogs, can be infected with influenza viruses.¹²

Influenza viruses belong to the Orthomyxoviridae family defined by viruses that have a negative-sense, single-stranded, and fragmented RNA genome. There are currently 18 different hemagglutinin subtypes (H1 to H18) and 11 different neuraminidase subtypes (N1 to N11) for influenza A viruses.³ Antigenic changes occur only in the proteins of influenza virus neuraminidase and hemagglutinin, and hemagglutinin is more important than neuraminidase because hemagglutinin is more susceptible to such changes.

The history of the first influenza A (H1N1) pandemic in the 21st century (March 2009) occurred 33 years after the last pandemic of the 20th century (1976), followed by the detection of cases in the United States and many other countries. With the spread of the disease to several continents in April, the sixth phase of the pandemic was announced by the World Health Organization.⁴ On the other hand, the emergence and pandemic of influenza A (H1N1) virus with swine-origin occurred so suddenly that apparently the pathogenic stages in environment (first phase), pathogenicity in pigs (second phase), transmission from pig to human (third phase), limited human-to-human transmission in Mexico (fourth Phase), limited human outbreaks in the Americas (fifth Phase), and widespread epidemics across the continent (sixth Phase) occurred in less than three months.

With respect to viral epidemics that may emerge, molecular epidemiological techniques have been useful tools in detecting the source of epidemics and tracking the spread of regional or global viruses. In this regard, phylogenetic methods, a kind of mathematical modeling, provide new opportunities to study the evolutionary history of pathogens and to reconstruct the spatio-temporal and demographic dynamics of viral epidemics. Many viruses, such as the influenza virus, evolve faster than normal transmission, and phylogenetic techniques as an ideal and objective tool can be used for reconstructing spatio-temporal dynamics and transmission history.

Phylogeography is the study of historical processes that may have shaped the current geographical distribution of living species, such as viruses. Nowadays, these studies are performed according to the information obtained from the genealogy. Recent advances have made it possible to integrate the theory of coalescent into phylogeography in such a way that both demographic and spatial aspects of the epidemic are inferred simultaneously. This information is used to track the patterns of viral spread, which in turn can become effective intervention and prevention strategies and can shed light on the main factors responsible for the spread and transmission of the virus.⁵

One of the limitations of the traditional epidemiological approaches is their inability to estimate the path of the virus and identify the estimated time and place of virus entry into a specific geographical area based on available data. Bayesian phylogeographical approaches using nucleotide sequences have made it possible to study the path of the virus and the approximate time and place of virus entry into a population along with other characteristics such as molecular dynamics and so on. In this study, we investigate the above gaps using the Bayesian phylogeographic approach for influenza virus H1N1 serotypes in Iran and neighboring countries.

Method

We screened the NCBI Influenza Virus Resource and downloaded all HA and NA nucleotide sequences of influenza H1N1 available up to December 25, 2020 from Iran and its neighboring countries (i.e., Pakistan, Afghanistan, Armenia, Turkmenistan, Azerbaijan, Turkey and Iraq). We captured sampling date and location of sequences by extracting information deposited in the NCBI Influenza Virus Resource. In our final genome sets, a total of 398 HA and 339 NA sequences remained after the exclusion of identical sequences and sequences with unknown sampling time and location. By using the ClustalW algorithm in MEGA v7.0, multiple sequence alignment was performed and manually inspected. The dataset was trimmed for phylogenetic analysis.⁶ The initial maximum likelihood (ML) phylogenetic analysis was conducted using Hasegawa-Kishino-Yano (HKY) and General Time Reversible (GTR) nucleotide substitution models for HA and NA, respectively. The best-fitting nucleotide substitution models were identified using an online execution on the ATGC bioinformatics platform (http://www.atgc-montpelier. fr/). Analysis was performed with 1000 bootstraps in MEGA v7.0.⁶

To infer the evolutionary dynamic and time to the most recent common ancestor (tMRCA) for the HA and NA sequences, Bayesian Markov chain Monte Carlo (MCMC) method was performed using BEAST v2.5.241 package.⁷ In these analyses, HKY+G (for the HA dataset) and GTR+G (for the NA dataset) substitution model, uncorrelated lognormal relaxed clock model, and Bayesian skyline coalescent tree priors were used.⁸ Four independent MCMC chains were run for 25 million generations (sampling every 2,500 steps)

and were combined using the LogCombiner program v1.54.⁸ Convergence was assessed based on the effective sampling size (ESS) after a 10% burn-in, using Tracer software v1.5. ESSs of 200 and above were accepted. The maximum clade credibility tree was generated in the TreeAnnotator program, while the initial 10% of trees were discarded as burn-in.^{7,9}The maximum clade credibility tree (MCC) was visualized in the FigTree program v1.2.3.¹⁰

Inferences about spatial dynamic and potential viral migration patterns of H1N1 within Iran and between Iran and its neighboring countries were made based on a discrete Bayesian phylogeographic model developed in BEAST software v2.5.241.⁸ The symmetric substitution model with the Bayesian Stochastic Search Variable Selection (BSSVS) approach was performed in the SPREAD program v1.0.7.¹¹

To determine selection pressure acting on HA and NA lineages, we estimated the ratio of non-synonymous (dN) to synonymous (dS) substitutions per site (ratio dN/dS) for each lineage, using all the sequences included in this study. Positively selected codons were detected using the single likelihood ancestor counting (SLAC) and fixed effects likelihood (FEL) methods with a significance level of 0.1, all procedures are available in the HyPhy package and accessed through the Datamonkey web server.^{12,13}

Result

The H1 phylogeographic tree showed two large independent clusters within Iran, suggesting the initiation of two independent effective outbreaks with different tMRCAs, one in the past years (1997) and the other in recent years (2005). The first outbreak was initiated from the north region of the country and most of time was circulated between the north and northwest region within the country and established some small introduction in the southwest region in the late years (2008). The phylogeographic tree showed that the second initiation was introduced in the north region and had concentrated circulation in the cities within this region by 2013 and then, circulation of virus expanded to the west region and continued to 2015, therefore based on our result inferred cautiously there were small and concentrated epidemics which had occurred in most parts of the country which came along until recently, despite the rapid expansion and decline in the region, it has remained stable (Figure 1). Along with the circulation of virus within the country, the phylogeographic tree revealed that intercountry transmission occurred in circulation between southwest region, Turkey and Afghanistan and again back transmission to the north region in Iran from 2007 to 2010, in addition, same back transmission was observed in cities were located in the west region and close to the Turkey border during 2013 to 2015. So, we could infer cautiously that, the virus has entered the country many times in the form of multiple introductions and spread, and according to our data, it did not have a single independent source and the age of presence of the virus in Iran is older than neighboring countries. Based on our genome set, lineages from different parts of the country branched within the lineages from Tehran, highlighting the role of Tehran as the epicenter of the outbreaks in the country. After 2010, small and concentrated epidemics in other cities have their origins in Tehran.

According to the above-mentioned, generally we concluded that probably the virus has been circulating in Tehran with low frequency for a little period and has gradually spread throughout the region and neighboring countries in the following years in a single and sporadic cases. The small genetic distance between the different branches of these clusters in this period suggests rapid spread of the virus in this specific geographical area during the narrow time (2007-2009) in Iran and neighboring countries. Based on our genome set the virus was still present in small numbers in some countries, such as Turkey, where the virus was mainly circulating in comparison to other neighboring countries.

The H1 phylogeographic tree based on different geographical regions was constructed separately and showed that two independent clusters same as the above results were created for our total sequences and geographically suggested that the virus circulation started from north and northwest to east and southwest of Iran over time, and then moved to east and southeast in recent years (Figure 2), and this indicates the correct goodness of fit of the model.

Results of the N1 phylogeographic tree were in line with the H1 tree, which two independent clusters among sequences from Iran, Turkey, Iraq, and Turkmenistan were observed, with the sequences related to Turkey

occupying the root of the tree (Figure 3). The phylogeographic tree showed that the transmission and circulation pattern was like H1 tree, and the northern region is the epicenter of this transmission which had taken place bilaterally and had occurred between the north of the country and Turkey. During a short time, the circulation of virus involved the whole country (Figure 4).

It should be noted that, if there were more sequences from other countries in our genome set, different results could be expected from the model outputs.

The evolutionary rate for the HA and NA genes of the H1N1 viruses were estimated at 4.01×10^{-3} [95% HPD: (2.82-5.40) $\times 10^{-3}$] and 3.30×10^{-3} [95% HPD: (1.33-5.43) $\times 10^{-3}$] substitutions/site/year, respectively. The mean time of the most recent common ancestor of

H1N1 viruses was 1989 (95% HPD: 1980-1994) for HA and NA as well.

The estimation of dN/dS ratio per site for each lineage was conducted. After excluding the duplicated sequences, 210 sequences with 443 sites for HA and 85 sequences with 91 sites for NA lineages were analyzed. The SLAC analysis of HA lineages showed that 17 sites were under negative/purifying selection while no positive/diversifying selection were observed. The overall dN/dS for HA lineages was 0.403. For the NA lineages, 14 negative/purifying selection sites and no positive/diversifying selection were observed. The overall dN/dS for NA lineages was 0.286. Based on the FEL analysis of HA lineages, 48 and 2 sites were under negative/purifying and positive/diversifying selections, respectively. This was 24 and 1 selection sites for the NA lineages, respectively. The results of Generic Algorithm Recombination Detection (GARD) showed no evidence of recombination.

Discussion

In this study, the time and path of transmission of H1N1 serotype influenza A viruses in Iran and neighboring countries have been investigated by using Bayesian phylogeography analysis on the sequences extracted from the gene bank. As of now several studies on geographical phylodynamic of different influenza virus serotypes with different goals and details have been developed and published around the world.^{14,15}Based on our knowledge, this study is the first study for Influenza in the field of Bayesian phylogeographical analysis in Iran and the region.

The results of our study showed that based on the extracted sequences, the age of emergence of H1N1 influenza virus serotype was older in Iran compared to neighboring countries, and among the cities within the country, Tehran had a key role and epicenter of transmission to other cities. In the meanwhile, it should be noted that there were not enough evidence diverse sequences from Iran and other possible countries of virus transmission in our genome set.

The global transmission of virus during the pandemic can be explained by the fact that and based on the existence agreements on previous reports, started in April 2009 in Mexico, followed by North America. Expanding transmission of the virus to Europe and its spread in the population of European countries, in connection with the travel of people living in European countries to Mexico or the United States one week before the onset of clinical symptoms.^{4,16-18}Based on the results of our study, the virus has been circulating in Tehran with low frequency for some time and gradually it has spread throughout the region and neighboring countries as sporadic event in the recent years. Our results showed that after rapid inhibition, the virus was still present in small numbers in some countries, such as Turkey. The results of our phylodynamic reconstruction were in accordance with other evidences that stated disease was initiated by the arrival of air travelers from non-neighboring European countries, first by an Iranian teenager living abroad in June and then by other travelers, gradually coinciding with other foreign trips and return of pilgrims from Hajj Umrah, the disease spread and was observed almost all over the country, although the prevalence of the disease in some provinces, especially the northwestern provinces and Tehran due to more population density was higher than other centers in the country.¹⁹

In general, the influenza epidemic caused by type A virus usually occurred in a certain state, for example, the pandemic of influenza, starts almost suddenly and reaches its maximum within 2-3 weeks, and lasted for 5-6

weeks.⁴The small genetic distance between the different branches of these clusters in this period indicated the high rate of spread of the virus, which reflected the sudden onset and rapid inhibition of the epidemic in this period (2007-2009) both in Iran and in other neighbor countries. The 2009 pandemic influenza virus is like seasonal influenza viruses in many epidemiological features and when the pandemic is fully subsided and at least after one year has passed, it is expected to follow the same trend as other seasonal influenza viruses.

Therefore, due to the significant threat to public health caused by such viruses, regular monitoring of H1N1 viruses in the country, especially in entering point to the country, strengthening the social measures and screening of newcomers and tourists from neighboring countries, should be considered as impactful tools to fight against future infections.²⁰ It is also necessary for countries to have a regional monitoring program in addition to internal monitoring programs, because control of this disease would not be possible without such a program. In addition, given the potential threat to the public health of the H1N1 virus, it is also important to develop effective measures like establishing a molecular surveillance program to prevent epidemics and pandemics.

Our study showed that the estimated evolutionary rate for HA and NA genomes for H1N1 serotype in our genome set from Iran and neighboring countries is similar to that previously found in different countries in Scotland, Argentina, Italy, China, India, Japan, South Korea, and Israel.^{14,21,22}This finding validates the estimates of the recent common ancestor in our study, and supports the validity of our tree estimates and topology. Also presence of similar mutation rates for HA and NA genes seems they were coevolving, and leads to minimizing the possibility of emergence of new reassortments.²³

Based on our results we did not find any positive selection sites, but other studies identified different number of positive selection sites and these differences can be attributed to the differences in significance levels and data sets used in the other studies.²⁴

This study had some limitations such as, we could not study the epidemiological parameters of the H1N1 virus in neighboring countries such as Iraq, Armenia, Afghanistan, and the Persian Gulf countries in depth due to the existence of limited sequences of these countries in the gene bank. In addition, we could not directly estimate the full extent of the evolution of the genome set because the time frame of this data set was not large enough. Our models and time estimates need to be further explored in future studies as new sequences from other countries and wider time periods become available, allowing for a more comprehensive assessment of the virus's evolutionary history.

In summary, our results suggested that the H1N1 serotype influenza virus epidemic in Iran may have started much earlier than in other countries surveyed in the region, and it was estimated that the date is earlier than the identified years by molecular detection. Therefore, along with such measures like resource management, diagnostic approaches, and preparedness to fight against viruses that already were in place, we need to initiate continuous monitoring, and screening of H1N1 serotype influenza virus in the country, especially by implementation of feasible, effective, and innovative measures at borderline for virus control. It is also important for countries to have a regional monitoring program in addition to internal monitoring programs, as well as to start a virus molecular care program. In the future, with the availability of new sequences from other countries and more open time intervals, our model and timeline estimates need to be further explored.

References

1. Nguyen-Van-Tam JS, Hampson AW. The epidemiology and clinical impact of pandemic influenza. *Vaccine* . 2003;21(16):1762-1768.

2. Rimmelzwaan G, Kuiken T, Van Amerongen G, Bestebroer T, Fouchier RAM, Osterhaus ADM. A primate model to study the pathogenesis of influenza A (H5N1) virus infection. *Avian diseases* . 2003;47(s3):931-933.

3. Fouchier RA, Munster V, Wallensten A, et al. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *Journal of virology* . 2005;79(5):2814-2822.

4. Organization WH. Influenza-like illness in the United Staites and Mexico. http://www who int/csr/don/2009_04_24/en/index html . 2009;

5. Holmes EC. The phylogeography of human viruses. *Molecular ecology* . 2004;13(4):745-756.

6. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011;28(10):2731-2739.

7. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* . 2007;7(1):1-8.

8. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution*. 2005;22(5):1185-1192.

9. Jin Y, Yu D, Ren H, et al. Phylogeography of Avian influenza A H9N2 in China. $BMC\ genomics$. 2014;15(1):1-10.

10. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution* . 2018;4(1):vey016.

11. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* . 2011;27(20):2910-2912.

12. Pond SLK, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* . 2005;21(10):2531-2533.

13. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* . 2005;22(5):1208-1222.

14. Lycett S, McLeish NJ, Robertson C, et al. Origin and fate of A/H1N1 influenza in Scotland during 2009. *The Journal of general virology*. 2012;93(Pt 6):1253.

15. Ma Y, Liu K, Yin Y, et al. The phylodynamics of seasonal influenza A/H1N1pdm virus in China between 2009 and 2019. *Frontiers in microbiology* . 2020;11:735.

16. Abdel-Ghafar A, Chotpitayasunondh T, Gao Z, et al. Writing committee of the second world health organization consultation on clinical aspects of human infection with avian influenza A (H5N1) virus. Update on avian influenza A (H5N1) virus infection in humans. *N Engl J Med* . 2008;358(3):261-273.

17. Lemey P, Suchard M, Rambaut A. Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm.*PLoS currents* . 2009;1

18. Jones S, Nelson-Sathi S, Wang Y, et al. Evolutionary, genetic, structural characterization and its functional implications for the influenza A (H1N1) infection outbreak in India from 2009 to 2017. *Scientific reports* . 2019;9(1):1-10.

19. BAGHERI LK. Hajj and Swine Flu Pandemic (H1N1, 2009): What Is Expected and What Should Be Done? 2010;

20. Naguib MM, Arafa A-SA, El-Kady MF, et al. Evolutionary trajectories and diagnostic challenges of potentially zoonotic avian influenza viruses H5N1 and H9N2 co-circulating in Egypt. *Infection, Genetics and Evolution*. 2015;34:278-291.

21. Zehender G, Pariani E, Piralla A, et al. Reconstruction of the evolutionary dynamics of the A (H1N1) pdm09 influenza virus in Italy during the pandemic and post-pandemic phases. *PLoS One* . 2012;7(11):e47517.

22. Barrero P, Viegas M, Valinotto L, Mistchenko A. Genetic and phylogenetic analyses of influenza A H1N1pdm virus in Buenos Aires, Argentina. *Journal of virology* . 2011;85(2):1058-1066.

23. Smith GJ, Vijaykrishna D, Bahl J, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* . 2009;459(7250):1122-1125.

24. Xu J, Christman MC, Donis RO, Lu G. Evolutionary dynamics of influenza A nucleoprotein (NP) lineages revealed by large-scale sequence analyses. *Infection, Genetics and Evolution*. 2011;11(8):2125-2132.



Figure 1. Maximum clade credibility phylogenies for the hemagglutinin gene of influenza A H1N1. The branches are colored according to the location of their nodes. The scale bar at the bottom indicates the years before the most recent sampling time (2019).



Figure 2. Significant locations of transmission of influenza H1N1 based on HA nucleotide sequences



Figure 3. Maximum clade credibility phylogenies for the neuraminidase gene of influenza A H1N1. The branches are colored according to the location of their nodes. The scale bar at the bottom indicates the years before the most recent sampling time (2019).



Figure 4. Significant locations of transmission of influenza H1N1 based on NA nucleotide sequences







