Acta Scientific Microbiology (ISSN: 2581-3226)

Volume 1 Issue 8 August 2018

Research Article

Phylogenetic Study and Principal Component Analysis of the Matrix Gene of the Influenza A Virus Subtype H1N1 from 1918 - 1999

Raúl Isea*

Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela

***Corresponding Author:** Raúl Isea, Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela.

Received: June 26, 2018; **Published:** July 13, 2018

Abstract

The goal of this work is to review the evolution of the 81 matrix genes of influenza A Virus subtype H1N1 from 1918 until 1999 with phylogenetic tools and compare the results with the principal component analysis (PCA). The results obtained from the phylogenetic analysis are consistent with the evolution of the virus according to the scientific literature while the results of the PCA reveals that the sequences can be grouped in 7 clusters where the strain that was isolated in Tientsin, China in 1977, is not included in these clusters.

Keywords: Influenza; H1N1; Phylogenetic; Matrix Gene; Principal Component Analysis; PCA

Introduction

The human population has always been affected by the influenza virus. We recall the pandemics that occurred in 1781, 1789, 1830 and 1847 where the last two pandemics also coincided with the second and third cholera pandemics in the world [1]. It was later followed by the pandemic of 1889 that began in the spring and expanded worldwide for several months, reaching its maximum in northern Europe and the US at the beginning of 1890. Immediately thereafter, there was a recurrence that peaked near the end of 1891 and another one that peaked at the beginning of 1892 [2].

In 1918, the worst pandemic in the history of mankind occurred and is now called the "Spanish influenza" and it affected one third of the world's population. Between 20 and 50 million people died worldwide [3]. Most of the autopsies performed at the time revealed that the cause of death was pneumonia and respiratory failure.

It was in 2006 when Taubenberger and Morens published their work that the contagion process occurred in three pandemic waves [4]. The first wave began in March of 1918 and it affected various states in the US, Europe and probably Asia. Later there was another wave that occurred from September to November in 1918 which registered the highest number of deaths. The final wave was at the beginning of 1919. Its propagation was very fast compared with the 1889 event. In parallel, they also indicated that the eight genes of the H1N1 virus are very narrow relative to the avian influenza viruses so that an avian virus must have infected humans and subsequently adapted to them in order to spread from person to person.

A new variant of the H1N1 influenza was identified in 1933 in Puerto Rico and several studies revealed that it was serologically distinct from the seasonal variant that was previously circulating around the world. The virus then spread to several continents and replaced the prior seasonal influenza strain. In 1947, an antigenic change of the H1N1 virus product of the 1943 virus was detected. This new H1N1 virus was called A-Prime and it was mild but widespread. Nelson., et al. [5] suggested that the 1947 strain was a combination of the two previous H1N1 viruses that combined with the 1943 virus along with another virus of an unknown type. Moreover, they concluded that the HA1 region of the 1947 HA is significantly different from the 1940 strain.

For 20 years, there were no more cases of influenza A caused by the H1N1 virus. However, it reappeared in the Soviet Union in November, 1977 and has been given the name "Russian influenza" that affects mainly the younger population under 25. It is very similar to the 1947/1948 H1N1 strain. Hence, this particular strain has been suggested in the scientific literature that there was an accidental release of a virus from a laboratory [6,7].

In the fall of 1986, there was another outbreak of influenza at the Naval Air Station in Key West, Florida which has been suggested to be caused by strains that had been circulating in Asia during the spring of that year. The Taiwanese/1986 strain was isolated and after a phylogenetic study, it was concluded that it had evolved from a virus that had circulated in the early 1980s in Hong Kong [8]. Also, it was suggested that this virus was the product of a recombination event that occurred between two H1N1 viruses.

Figure 1: Radial representation of consensus tree obtained from 10,000 replicates using the GTR+G model.

Materials and Methods

The nucleotide sequences of the matrix gene was obtained from the Influenza Research Database (available at fludb.org). The search criteria was to select only those sequences of type A, subtype H1N1, of any country dated between 1918 and 1999. On the other hand, the genome must be complete and we only select the sequences in humans. The abbreviation of each sequence was defined with 8 numbers as follows: the first four represent the year of the sequenced strain and the last four correspond to the last four of the GenBank Id. Using this procedure, it is easy to identify the evolution of the virus over the years.

Figure 2: Results of the Principal Components Analysis (see text for more details).

The alignment of the sequences was performed with the Clustal Omega program which is available on the EBI web server (https://www.ebi.ac.uk/Tools/msa/clustalo/). The phylogenetic analyses were carried out using the PhyML software version [9] including the automatic model selection with the best-fit substitution model with PhyMl-SMS (http://www.atgc-montpellier.fr/phyml-sms/). All analyses were performed with 10,000 iterations.

The Principal Components Analysis (PCA) was calculated with JalView software [10]. This software yielded the PCA from eigenvector decomposition of the matrix that was obtained from the sum of the substitution matrix scores at each aligned position. Finally, the clustered sequences were grouped together manually.

Results and Discussion

In this study, the 1994 sequences were obtained according to the conditions indicated above in the Materials and Method's section. Only 81 sequences were selected that were characterized as having up to a pair of sequences for each country, thereby eliminating duplicate sequences (Table 1).

Abbreviation	Year	Strain name	GenBank ID
19403272	1940	A/Hickox/1940	CY013272
19779293	1977	A/Hong Kong/117/1977	CY009293
19988523	1998	A/Hong Kong/427/1998	AF258523
19978522	1997	A/Hong Kong/470/1997	AF258522
19800454	1980	A/India/6263/1980	CY020454
19975069	1997	A/Johannesburg/159/1997	CY125069
19575863	1957	A/Kw/1/1957	CY125863
19517892	1951	A/Liverpool/1951	CY077892
19957692	1995	A/Malaysia/06768/1995	CY117692
19977745	1997	A/Malaysia/12530/1997	CY117745
19989123	1998	A/Malaysia/15042/1998	CY119123
19549341	1954	A/Malaysia/1954	CY009341
19357399	1935	A/Melbourne/JY2/1935	CY147399
19799740	1979	A/Memphis/1/1979	CY019740
19841726	1984	A/Memphis/1/1984	CY021726
19869102	1986	A/Memphis/12/1986	CY019102
19969110	1996	A/Memphis/2/1996	CY019110
19879772	1987	A/Memphis/3/1987	CY019772
19830925	1983	A/Memphis/4/1983	CY010925
19985085	1998	A/Moscow/13/1998	CY125085
19335814	1933	A/NWS/1933	L25814
19340985	1934	A/NWS/1934	CY120985
19963814	1996	A/Nanchang/13/1996	CY013814
19995013	1999	A/Nanchang/16A/1999	CY125013
19567714	1956	A/Netherlands/001B1/1956	CY077714
19547728	1954	A/Netherlands/001H1/1954	CY077728
19537751	1953	A/Netherlands/001R1/1953	CY077751
19487759	1948	A/Netherlands/001S1/1948	CY077759
19497766	1949	A/Netherlands/002K1/1949	CY077766
19985093	1998	A/Ostrava/801/1998	CY125093
19350470	1935	A/Phila/1935	CY020470
19341099	1934	A/Puerto Rico/8/1934	V01099
19498872	1949	A/Roma/1949	CY019972
19975077	1997	A/Shanghai/2/1997	CY125077
19965061	1996	A/Shanghai/8/1996	CY125061
19955053	1995	A/Shengzhen/227/1995	CY125053
19953615	1995	A/Beijing/262/1995	CY033615
19896824	1989	A/Siena/10/1989	CY036824
19956840	1995	A/Siena/14/1995	CY036840
19875853	1987	A/Siena/4/1987	CY045853
19975350	1997	A/TW/3355/1997	DQ415350
19868876	1986	A/Taiwan/01/1986	DQ508876
19969265	1996	A/Taiwan/117/1996	DQ249265
19995351	1999	A/Taiwan/4845/1999	DQ415351
19860566	1986	A/Texas/2922-3/1986	CY020566
19919317	1991	A/Texas/36/1991	CY009317
19770574	1977	A/Tientsin/78/1977	CY020574
19791910	1979	A/USSR/46/1979	CY021910
19773029	1977	A/USSR/90/1977	X53029
19335757	1933	A/United Kingdom/1/1933	CY045757
19439453	1943	A/Weiss/1943	CY009453
19925037	1992	A/Wellington/47/1992	CY125037

Table 1: Matrix gene sequences employed in this study.

The model that was obtained from PhyML-SMS was GTR+G where the value of the Gamma shape parameter was estimated to 0.19. The phylogenetic tree is shown in figure 1 and is visualized with the PRESTO software (Phylogenetic tReEviSualisaTiOn available un atgc-montpellier, fr/presto). The radial presentation of the consensus maximum likelihood tree was obtained from 10,000 replicates. In figure 1, we can see that the evolution of the virus is according to the years as indicated in the introduction of the present work. The result of the Principal Components Analysis (PCA) study is shown in figure 2 where seven clusters are identified

and are highlighted with a yellow circle. In table 2, the sequences in each cluster are identified. From this result, it can be seen that the particular strain (A/NWS/1933) belonged to the1918-1936 cluster but the strain (A/Tientsin/78/1977) did not belong to this cluster. This means that it is not in the cluster with the 1947/1948 sequences.

1	19180766, 19335757, 19340985, 19341099, 19359956, 19357327, 19357399, 19350470, 19360446
2	19403272, 19426770, 19430286, 19439453, 19451710, 19467351, 19467766, 19472084, 19487884, 19487759
3	19489948, 19497766, 19498872, 19508091, 19509333, 19511822, 19517375, 19517892, 19537751, 19549341, 19547728, 19567714, 19576794, 19575863
4	19781798, 19789964, 19780294, 19796412, 19799740, 19791910
5	19800454, 19811030, 19829621, 19821038, 19838947, 19830925, 19841726, 19869102, 19868876, 19860566, 19879772, 19875853
6	19896824, 19919317, 19925037, 19935045, 19958798, 19958876, 19957692, 19955053, 19956840, 19969110, 19965061, 19969265, 19977745, 19985085, 19988523
7	19953615, 19975077, 19978522, 19975069, 19975350, 19989123, 19985093, 19995013, 19995351

Table 2: Clusters obtained by PCA from the matrix gene of Influenza A Virus.

Conclusion

The present work calculates the evolution of the influenza A virus with the phylogeny program using the matrix gene subtype H1N1. This result is in accordance with the scientific literature. However, it is surprising that when forming a cluster based on the PCA results, the strain (A/Tientsin/78/1977) appears to be isolated from the rest of the clusters and it does not belong in cluster 3 or 4. This last result suggests that the Tientsin strain must have been altered from the 1947/1948 H1N1 strain.

Acknowledgement

The author wishes to express his sincere thanks to Prof. Karl E. Lonngren for his unconditional help and comments concerning this manuscript.

Bibliography

1. Taubenberger JK *et al*. "Discovery and characterization of the 1918 pandemic influenza virus in historical context". *Antiviral Therapy* 12.4B (2007): 581-591.

2. Trilla A., *et al*. "The 1918 Spanish Flu in Spain". *Clinical Infectious Diseases* 47.5 (2008): 668-673.

3. Reid AH., *et al*. "The 1918 Spanish influenza: integrating history and biology". *Microbes and Infection* 3.1 (2001): 81-87.

4. Taubenberger JK., *et al*. "1918 Influenza: the Mother of All Pandemics". *Emerging Infectious Diseases* 12.1 (2006): 15-22.

5. Kilbourne ED. "Influenza Pandemics of the 20th Century". *Emerging Infectious Diseases* 12.1 (2006): 9-14.

6. Rozo M., *et al*. "The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate". *mBio* 6 (2015): e01013-e01015.

7. Zimmer SM., *et al*. "Historical perspective--Emergence of influenza A (H1N1) viruses". *New England Journal of Medicine* 361.3 (2009): 279-285.

8. Klontz KC., *et al*. "An outbreak of influenza A/Taiwan/1/86 (H1N1) infections at a naval base and its association with airplane travel". *American Journal of Epidemiology* 129.2 (1989): 341-348.

9. Lefort V., *et al*. "SMS: Smart Model Selection in PhyML". *Molecular Biology and Evolution* 34.9 (2017): 2422-2424.

10. Waterhouse AM., *et al*. "Jalview Version 2-a multiple sequence alignment editor and analysis workbench". *Bioinformatics* 25.9 (2009) 1189-1191.

Volume 1 Issue 8 August 2018