Comparison of the Four Representative Strains of Influenza A Virus

**Abstract**

Influenza A viral strains are studied for mutation patterns. Geographical or temporal significance is not found for the genomes included in this research, and there is a slower mutation rate found for the two most recent strains of the virus. Polymerase PB1-F2 protein is truncated in the 2009 strain due to a transversion, but it is functional in the following viral genome.

**Introduction**

The Influenza A virus (IAV) genome consists of eight single-stranded negative-sense RNA chromosomes.(Moreira et al., 2016) The segments are enveloped in copies of the nucleocapsid protein (NP) which is encoded in the fifth viral segment. IAV was first isolated in the laboratory in 1932.(Potter, 2001) The IAV genome has 12 genes coding for 12 proteins, with the exception of the Californian strain which emerged in 2009. This strain, termed H1N1, produced only 11 functional proteins – its PB1-F2 protein, derived from a +1 alternative reading frame of the PB1 polymerase subunit gene, was truncated. Nearly half of all IAV variants isolated between 2009 and 2011 had the mutation which produced a non-functional PB1-F2 gene. The prevalence of the mutated gene peaked in 2011, when it reached 55%.(McAuley et al., 2017)

Influenza A viruses are grouped into subtypes based on the two surface proteins expressed by the virus: hemagglutinin (HA) and neuraminidase (NA), encoded in the viral segments 4 and 6, respectively. There are 18 hemagglutinin subtypes and 11 neuraminidase subtypes, giving rise to IAV viral segment combinations ranging from H1 to H18 and N1 to N11, indicated in the name of each strain.(CDC, 2021) The two glycoproteins are subject to rapid mutation because the immune response targets both hemagglutinin and neuraminidase.(Potter, 2001)

**Materials and Methods**

Sequences used in this paper have been retrieved from the National Centre for Biotechnology Information database. Out of the available genomes, four representative strains have been selected: A/Korea/426/1968(H2N2), A/New York/392/2004(H3N2), A/California/07/2009(H1N1), and A/Shanghai/02/2013(H7N9).(*Influenza A Virus (ID 10290) - Genome - NCBI*, n.d.) In the paper, the strains will be referred to by their isolation year.

Multiple Sequence Alignment

The sequences were aligned using the Clustal Omega program. All the genomes were aligned against the 1968 strain in chronological order. This is reflected in the results presented below, with the 1968 strain bearing 100% similarity and 0 mutations – it is included in the charts as a reference point but the data has no significance when taken out of context.

Sequence Analysis

The data generated by our program was analysed and visually represented in MS Excel. Special attention was paid on comparing the mutations in the coding and non-coding sequences, as well as contrasting the amount of indels and gaps that occur between the two.

Protein Sequence Analysis

The non-functional PB1-F2 sequence from 2009 was translated and aligned with the three functional protein sequences expressed by the other three strains. Other protein sequences resulted in consistent products with identical functions.

**Results**

Overall similarity between sequences decreased over time, amounting to 75.34% in 2013. However, coding sequences were relatively well-preserved, with an average similarity of 86.03%. Figure 1 compares the similarities between the four strains.

To get a clearer representation of the coding-versus-non-coding sequence conservation, Figure 2 contrasts the two while taking into account the number of nucleotides that underwent mutation. The IAV genomes from 2009 and 2013 are notably close in terms of their non-coding sequence similarity, while their coding sequences differ more prominently.

In Figure 3, we can see the total number of gaps, insertions, and deletions between the strains. Again, the 2009 and 2013 strains have more in common with each other than with the 1968 sequence. Both have 28 gaps and insertions in their coding sequences, while the 2004 strain has none.

Transversions have more impact on the protein structure and function seeing as they are more likely to alter the amino acid sequence than transitions.(Guo et al., 2017) A higher Ts:Tv ratio indicates a more closely similar protein product expressed by the sequence, while a lower ratio may point to a more significant change in protein conformation and function. Figure 4 depicts the transition-to-transversion ratios of the three subsequent strains compared against the 1968 strain.

The non-functional amino acid sequence of PB1-F2 in 2009 genome was translated and compared to the functional 90-residue sequences of that protein [Figure 5]. The truncated version produces a protein composed of 11 amino acid residues in the virus due to a transversion that occurs in the second base of the twelfth codon of the protein sequence. TCA, which codes for Serine, is replaced by TAA, a stop codon.

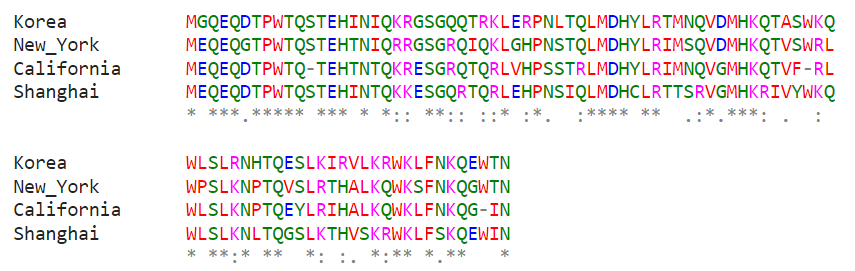


Figure 5

**Discussion**

Similarity of the studied sequences is difficult to attribute to either temporal or geographical influences. Geographically and temporally close strains (from 2004 and 2009) displayed a significant distinction, most notably for the dramatic increase in the mutation number. It could be argued that the Influenza A virus had experienced the greatest change in sequence prior to the 2009 outbreak, simply due to the resulting truncation of PB1-F2. Given that the protein was once again functional in the 2013 outbreak, the mutation must not have been beneficial for the virus.

**Conclusion**

Influenza A viral strains have caused large-scale pandemics across the globe and will likely continue to evolve in the future, seeing as there is a strong mutation pressure on the hemagglutinin and neuraminidase glycoproteins. Immunity against one HA or NA subtype does not indicate immunity against other subtypes, so any newly emerging subtype of IAV has a potential to result in a pandemic.

**References**

CDC. (2021, November 2). *Types of Influenza Viruses*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/about/viruses/types.htm

Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., & Reddy, T. E. (2017). Transversions have larger regulatory effects than transitions. *BMC Genomics*, *18*, 394. https://doi.org/10.1186/s12864-017-3785-4

*Influenza A virus (ID 10290)—Genome—NCBI*. (n.d.). Retrieved January 10, 2022, from https://www.ncbi.nlm.nih.gov/genome/10290

McAuley, J., Deng, Y.-M., Gilbertson, B., Mackenzie-Kludas, C., Barr, I., & Brown, L. (2017). Rapid evolution of the PB1-F2 virulence protein expressed by human seasonal H3N2 influenza viruses reduces inflammatory responses to infection. *Virology Journal*, *14*(1), 162. https://doi.org/10.1186/s12985-017-0827-0

Moreira, É. A., Weber, A., Bolte, H., Kolesnikova, L., Giese, S., Lakdawala, S., Beer, M., Zimmer, G., García-Sastre, A., Schwemmle, M., & Juozapaitis, M. (2016). A conserved influenza A virus nucleoprotein code controls specific viral genome packaging. *Nature Communications*, *7*(1), 12861. https://doi.org/10.1038/ncomms12861

Potter, C. w. (2001). A history of influenza. *Journal of Applied Microbiology*, *91*(4), 572–579. https://doi.org/10.1046/j.1365-2672.2001.01492.x