Comparing Influenza A H1N1 and H3N2 Genetic Drift 2013-2017, US

Molecular Epidemiolgy Final Report, Spring 2024

Savannah Hammerton

Abstract

Genetic drift leads a need to update influenza vaccine strains most seasons. Distance measures over time can tell us how much drift has occurred. We aimed to determine if H1N1 or H3N2 appeared to drift more over time, and found that there was no significant difference between the subtypes.

# Introduction

Genetic drift leads a regular need to re-assess and update influenza vaccine strains from year to year (1). Differences between influenza vaccine strains and circulating strains have been cited as one of the causes of reduced vaccine efficacy (2–5). “Mismatch” between the strains is often defined somewhat arbitrarily and dichotomously, but using continuous genetic distance measures could lead to a better understanding of the impact of genetic drift and vaccine mismatch on changes in vaccine effectiveness in populations (3,5). However, there is not just one circulating strain per season, as genetic drift happens continuously over time and around the globe. One way we can examine and compare genetic drift is to assess the gradual change and evolution over time and compare genetic distances using phylogenetic tree-based methods. In this analysis, we used such methods to compare genetic drift in H1N1 and H3N2 Influenza A subtypes in the United States from 2013 to 2017.

# Methods

Data were obtained from [BV-VRC.org](https://www.bv-brc.org/view/Taxonomy/11320?and(eq(host_group,Human),eq(isolation_country,%22USA%22),or(eq(collection_year,%222016%22),eq(collection_year,%222015%22),eq(collection_year,%222014%22),eq(collection_year,%222013%22),eq(collection_year,%222017%22)),or(eq(subtype,%22H1N1%22),eq(subtype,%22H3N2%22)),eq(genome_status,%22Complete%22),eq(segment,%224%22))#view_tab=genomes&filter=false) using the filters applied in the attached link (6). The data were loaded and cleaned in R (FASTA file was loaded using the seqinr package) (7,8), and fifty random samples were taken in each subtype of interest in each year to reduce the final sample size from 9,853 to 500 (seed was set to 42 for reproducibility). The msa package was then used perform multiple sequence alignment and prepare data for modeling (9).

Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (10) was used to construct a rooted tree and bootstrapped estimates were calculated for edge lengths using the phangorn R package (11). These estimates were then used to calculate mean total distances and 95% confidence intervals (CIs) for H1N1 and H3N2 over the study period, as well as a contrast between the two (). This contrast was used to test the hypotheses that H3N2 phylogenetic trees would exhibit increased diversity and genetic distance compared to H1N1 trees (based on results from exploratory analysis).

Maximum likelihood methods were then used to select and fit models and generate final phylogenetic tree models using the phangorn R package (11). The JC, K80, and SYM models were tested with all models also testing base, +I, +G, and G+I versions. Final models were selected based on a combination of logLikelihood, AIC, and BIC measures. Collection date was used to root the models in both subtypes. Specific collection dates were missing in five H1N1 and two H3N2 records (only years provided). In these cases, the first of the reported year was used as the date.

# Results

Full bootstrap distributions of distance measures for H1N1 and H3N2 over the study period can be seen in [Figure 1](#fig-dists). Distance and contrast bootstrapped estimates and 95% CIs are shown in [Table 1](#tbl-contrast). Visually, it appears as though H3N2 had a slight trend towards higher distance in the full distributions. However, when examining the final estimates, and in particular, the contrast estimates, we see that the 95% CI crosses zero, and we must thus conclude that there is no statistically significant difference in the total amount of genetic drift over time in our samples over the study period.

|  |
| --- |
| Figure 1: Bootstrap distributions of total distances for H1N1 (red) and H3N2 (blue) subtypes over the study period. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Bootstrapped means and 95% confidence intervals (CIs) for total H1N1 and H3N2 tree distances and contrasts. As the 95% CI for the contrast crosses zero, we conclude there is no statistically significant difference in the distances of the two sub-types, and thus no significant difference in the antigenic drift over the analysis period.   | subtype | mean | lwr | upr | | --- | --- | --- | --- | | H1 | 0.35 | 0.32 | 0.38 | | H3 | 0.38 | 0.34 | 0.41 | | Contrast | −0.03 | −0.07 | 0.02 | |

Models were tested as described above, and fit statistics are presented in [Table 2](#tbl-mods). The SYM+I+G models for both H1N1 and H3N2 subtypes appeared to fit generally better when balancing logLikelihood, AIC, and BIC values, so this model was chosen for both subtypes. The same model was intentionally chosen for each subtype to make the resulting trees more comparable. The final trees are presented in [Figure 2](#fig-trees). While initial results revealed that overall distance was similar across the two subtypes, H1N1 appeared to have a more consistent, general evolution, while H3N2 evolution appeared much more sporadic and random.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Fit statistics for H1N1 (top) and H3N2 (bottom) tested phlylogenetic tree models.   | Model | df | logLik | AIC | AICw | AICc | AICcw | BIC | | --- | --- | --- | --- | --- | --- | --- | --- | | H1N1 | | | | | | | | | JC | 497.0 | −8,101.5 | 17,196.9 | 0.0 | 17,581.3 | 0.0 | 19,924.3 | | JC+I | 498.0 | −7,992.6 | 16,981.1 | 0.0 | 17,367.3 | 0.0 | 19,714.0 | | JC+G(4) | 498.0 | −7,985.6 | 16,967.2 | 0.0 | 17,353.4 | 0.0 | 19,700.1 | | JC+G(4)+I | 499.0 | −7,984.2 | 16,966.5 | 0.0 | 17,354.5 | 0.0 | 19,704.9 | | K80 | 498.0 | −7,708.7 | 16,413.4 | 0.0 | 16,799.5 | 0.0 | 19,146.3 | | K80+I | 499.0 | −7,599.3 | 16,196.7 | 0.0 | 16,584.7 | 0.0 | 18,935.0 | | K80+G(4) | 499.0 | −7,592.4 | 16,182.7 | 0.0 | 16,570.7 | 0.0 | 18,921.1 | | K80+G(4)+I | 500.0 | −7,590.9 | 16,181.9 | 0.0 | 16,571.7 | 0.0 | 18,925.7 | | SYM | 502.0 | −7,689.2 | 16,382.4 | 0.0 | 16,776.1 | 0.0 | 19,137.3 | | SYM+I | 503.0 | −7,578.0 | 16,161.9 | 0.0 | 16,557.4 | 0.0 | 18,922.3 | | SYM+G(4) | 503.0 | −7,570.6 | 16,147.2 | 0.4 | 16,542.7 | 0.6 | 18,907.6 | | SYM+G(4)+I | 504.0 | −7,569.0 | 16,146.0 | 0.6 | 16,543.4 | 0.4 | 18,911.8 | | H3N2 | | | | | | | | | JC | 497.0 | −8,325.6 | 17,645.1 | 0.0 | 18,036.7 | 0.0 | 20,365.8 | | JC+I | 498.0 | −8,206.6 | 17,409.2 | 0.0 | 17,802.7 | 0.0 | 20,135.3 | | JC+G(4) | 498.0 | −8,199.6 | 17,395.2 | 0.0 | 17,788.7 | 0.0 | 20,121.3 | | JC+G(4)+I | 499.0 | −8,198.0 | 17,394.0 | 0.0 | 17,789.4 | 0.0 | 20,125.6 | | K80 | 498.0 | −7,965.8 | 16,927.6 | 0.0 | 17,321.1 | 0.0 | 19,653.8 | | K80+I | 499.0 | −7,846.2 | 16,690.5 | 0.0 | 17,085.9 | 0.0 | 19,422.1 | | K80+G(4) | 499.0 | −7,839.2 | 16,676.4 | 0.0 | 17,071.8 | 0.0 | 19,408.0 | | K80+G(4)+I | 500.0 | −7,837.6 | 16,675.1 | 0.0 | 17,072.4 | 0.0 | 19,412.2 | | SYM | 502.0 | −7,938.0 | 16,880.0 | 0.0 | 17,281.2 | 0.0 | 19,628.1 | | SYM+I | 503.0 | −7,814.6 | 16,635.3 | 0.0 | 17,038.3 | 0.0 | 19,388.8 | | SYM+G(4) | 503.0 | −7,807.3 | 16,620.5 | 0.3 | 17,023.6 | 0.5 | 19,374.0 | | SYM+G(4)+I | 504.0 | −7,805.4 | 16,618.8 | 0.7 | 17,023.8 | 0.5 | 19,377.8 | |

|  |
| --- |
| Figure 2: SYM+G+I trees rooted by collection date for H1N1 (left) and H3N2 (right). |

# Conclusions

This analysis aimed to compare genetic drift and evolution of influenza A H1N1 and H3N2 in the United States over the 2013-2017 time period. Both distance-based and maximum likelihood models were used to compare the changes in randomly sampled sequences collected over that period. While we expected to see increased change in H3N2 over H1N1 based on exploratory results, we found no statistically significant difference in the total H1N1 and H3N2 distances. However, final selected models showed a more consistent evolution in H1N1 strains compared to H3N2 strains, which appeared more random. This matches literature that reports H3N2 generally evolves quickly (12). Additionally, it is possible that the study period taking place a few years after the 2009 H1N1 pandemic also impacted the more gradual and consistent evolution of those strains (13).

This analysis was limited in computational resources, requiring smaller sample sizes and less complex models. Future analyses could benefit from further examination using larger samples over longer study periods to assess if one subtype appears to drift more rapidly than the other. Sub-analyses could also stratify by different, shorter time periods to examine how the evolution rates change over time and interact with each other.

# References

1. CDC. How Flu Viruses Can Change [Internet]. Centers for Disease Control; Prevention; 2022 [cited 2024 Mar 18]. Available from: <https://www.cdc.gov/flu/about/viruses/change.htm>

2. Belongia EA, Simpson MD, King JP, Sundaram ME, Kelley NS, Osterholm MT, et al. Variable influenza vaccine effectiveness by subtype: A systematic review and meta-analysis of test-negative design studies. The Lancet Infectious Diseases [Internet]. 2016 Aug 1 [cited 2023 Sep 28];16(8):942–51. Available from: <https://www.sciencedirect.com/science/article/pii/S1473309916001298>

3. da Costa JC, Siqueira MM, Brown D, Lopes JO, da Costa BC, Gama EL, et al. Vaccine Mismatches, Viral Circulation, and Clinical Severity Patterns of Influenza B Victoria and Yamagata Infections in Brazil over the Decade 2010–2020: A Statistical and Phylogeny–Trait Analyses. Viruses [Internet]. 2022 Jul 5 [cited 2024 Feb 29];14(7):1477. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9321334/>

4. Tenforde MW, Kondor RJG, Chung JR, Zimmerman RK, Nowalk MP, Jackson ML, et al. Effect of Antigenic Drift on Influenza Vaccine Effectiveness in the United States—2019–2020. Clinical Infectious Diseases [Internet]. 2021 Dec 1 [cited 2024 Mar 13];73(11):e4244–50. Available from: <https://doi.org/10.1093/cid/ciaa1884>

5. Tricco AC, Chit A, Soobiah C, Hallett D, Meier G, Chen MH, et al. Comparing influenza vaccine efficacy against mismatched and matched strains: A systematic review and meta-analysis. BMC Med [Internet]. 2013 Jun 25 [cited 2023 Jul 14];11(1):153. Available from: <https://doi.org/10.1186/1741-7015-11-153>

6. BRC B. BV-BRC Viruses::Taxonomy Overview [Internet]. [cited 2024 Mar 18]. Available from: <https://www.bv-brc.org/view/Virus/10239>

7. Charif D, Lobry JR. SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural approaches to sequence evolution: Molecules, networks, populations. New York: Springer Verlag; 2007. p. 207–32. (Biological and medical physics, biomedical engineering).

8. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: <https://www.R-project.org/>

9. Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S. [Msa: An R package for multiple sequence alignment](https://doi.org/10.1093/bioinformatics/btv494). Bioinformatics. 2015;31(24):3997–9.

10. Moulton V, Spillner A, Wu T. UPGMA and the normalized equidistant minimum evolution problem. Theoretical Computer Science [Internet]. 2018 Apr 18 [cited 2024 May 6];721:1–15. Available from: <https://www.sciencedirect.com/science/article/pii/S0304397518300690>

11. Schliep KP. [Phangorn: Phylogenetic analysis in R](https://doi.org/10.1093/bioinformatics/btq706). Bioinformatics. 2011;27(4):592–3.

12. Potter BI, Kondor R, Hadfield J, Huddleston J, Barnes J, Rowe T, et al. Evolution and rapid spread of a reassortant A(H3N2) virus that predominated the 2017–2018 influenza season. Virus Evolution [Internet]. 2019 Jul 1 [cited 2024 May 6];5(2):vez046. Available from: <https://doi.org/10.1093/ve/vez046>

13. Wilasang C, Suttirat P, Chadsuthi S, Wiratsudakul A, Modchang C. Competitive evolution of H1N1 and H3N2 influenza viruses in the United States: A mathematical modeling study. Journal of Theoretical Biology [Internet]. 2022 Dec 21 [cited 2024 May 6];555:111292. Available from: <https://www.sciencedirect.com/science/article/pii/S0022519322002831>