Comparing antigenic distance metrics for influenza against genetic and antigenic history

EPID 8200 project proposal

W. Zane Billings

Amanda Skarlupka

Lambodhar Damodaran

Justin Bahl

Andreas Handel

2024-03-13

# Project summary

Influenza A evolves rapidly and causes seasonal epidemics in temperate climates. Subsequent epidemics are often induced by immune escape and the spread of a novel variant. The antigenic distance between variants of influenza can be calculated using multiple different methods, and is not always equivalent to the genetic or evolutionary distance between to strains. There is no consensus on which distance calculation methods most accurately map to immune phenotypes and immune escape variants. We propose to calculate many different distance metrics on the same database, including antigenic cartography. We can then use the sequences to reconstruct a maximum likelihood evolutionary history, and compare distance-based evolutionary trees within a likelihood framework to determine which metrics most accurately reflect the evolutionary history of the immune escape variant strains.

# Significance

The natural history of influenza is shaped predominantly by rapid evolution due to host selective pressures, along with periodic recombination events which lead to the proliferation of novel variants. We call these evolutionary processes antigenic drift and antigenic shift respectively. Effectively vaccinating against influenza requires careful consideration of these evolutionary processes, and current vaccines rely on predicting variants of concern which are expected to circulate in the upcoming season.

Current influenza vaccines rarely elicit strong responses to many different strains of influenza, and so are less effective when predictions for circulating strains are incorrect, or when novel viruses emerge due to recombination as in 2009. Developing a broadly protective influenza vaccine which responds to many different strains of influenza is, therefore, a crucial goal for reduction of the burden of current and future influenza epidemics.

However, measuring the breadth of a vaccine response is difficult, and many proposed measures involve using the year of isolation of influenza viruses as a proxy for sequence dissimilarity due to antigenic drift. Especially for H1N1, which has two distinct historical lineages, year of isolation is an adequate measure of antigenic distance. However, there are many different measures of genetic and antigenic distance, with no clear consensus on which is best for influenza. Comparing the similarity of antigenic distance measurements and their ability to reconstruct the evolutionary history of influenza is necessary for improving methods for quantifying the breadth of vaccine response.

# Approach

We will combine data from an influenza cohort study along with publicly avaiable influenza sequence data in order to compare multiple measurements of genetic and antigenic distance. The cohort data we will use is comprised of volunteers enrolled at three different study sites from September 2013 ongoing. The study design is a prospective, open cohort study with one treatment arm. Recruited individuals completed a questionnaire at enrollment to provide demographic information (including sex, race, age, comorbidities, and prior vaccination history), gave a pre-vaccination serum sample, and received a seasonal influenza vaccine. Individuals returned for a followup visit with a target timeframe of 21 or 28 days post-vaccination (depending on the year) and gave a second serum sample. Individuals who were under the age of 65 received the standard dose (SD) FluZone (Sanofi Pasteur) vaccine, while individuals who were age 65 or older were allowed to choose either FluZone or FluZone High Dose (HD). Individuals were allowed to return in subsequent years.

For our study, we selected records that had both a pre and post-vaccination serum time point. We also only used records from study years 2013 up to March 2020, as the historical panel was reduced in subsequent years due to concerns arising from the COVID-19 pandemic. [Table 1](#tbl-demographics) shows demographics and counts for participants who were selected for inclusion in our study. We will also only consider responses to influenza A, and we will treat H1N1 and H3N2 as independent lineages for the purpose of our analyses.

Researchers conducted hemagglutination inhibition (HAI) assays to a panel of historical strains, which varied slightly over the duration of the study due to funding and virus availibilities. The strains they used, along with the UniProt and/or GISAID acession numbers for the protein sequences are given in [Table 3](#tbl-sequence-numbers). The years for which each strain was used are shown in [Table 2](#tbl-strain-years).

In order to compare genetic distance measures, we will first build a maximum likelihood (ML) phylogenetic tree using the RNA sequences for each strain. Then, we will calculate the pairwise distances between strains using several different methods: Hamming distance, Levenshtein distance, Grantham’s distance, -epitope distance, and normalized Euclidean distance from antigenic cartography (see following section). Then, we can use each of the computed distance matrices to build a distance-based tree via neighbor joining. The likelihood of the resulting distance-based tree can be calculated under the same framework as the ML tree, and we can use a likelihood ratio test to determine whether the distance-based tree is significantly worse than the ML tree. Note that we can easily extend our proposed methods to include any number of different genetic or antigenic distance metrics, but as this is a proof of concept study, we limited our concern to these metrics.

We will also check whether the ML tree reflects what we know about evolutionary history. Since the strains we will use emerged in a specific and known order, we can compare the inferred phylogeny based on sequence distance to the true phylodynamic history. In particular, comparing the likelihoods of the ML and cartography-based tree should give us insight into how similar the genetic and antigenic patterns of evolution are for both H1N1 and H3N2 vaccine strains. We can also compare the trees based on genetic distance measurements to the tree based on the cartographic distance using a likelihood ratio test, so we can determine which measures of genetic distance are similar to the genetic and antigenic evolution patterns we detect in the data.

Finally, we will also conduct correlation and reliability analyses on the different measures of antigenic distance to assess how similar they are to each other. Our tree-based analysis can tell us which simple measures are best at reconstructing the genetic or antigenic history of the virus. Susequent correlation and reliability analyses can provide additional information about how similar different measurements are and may help highlight why certain measurements are more similar to the genetic or cartographic phylogenies.

We conducted antigenic cartography using the post-vaccination cohort data described previously. We used the Racmacs R package to construct the cartographic maps from the HAI data, which employs an iterative multidimensional scaling (MDS) process to reduce the matrix of measurements into fewer dimensions, from which we can easily calculate Euclidean distance measurements. We computed cartographic maps in 1 – 5 dimensions, and based on the relative error for each, we determined that 2 dimensions were suitable, which matches previous findings.

# Data tables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Age, sex, race, dose, and study site distributions for the selected study years. Notice that the study was conducted at the FL and PA study sites from fall 2013 through spring 2017, and at the UGA study site beginning in January 2017 onwards.   |  | **2013 - 2014**, N = 1331 | **2014 - 2015**, N = 2381 | **2015 - 2016**, N = 2121 | **2016 - 2017**, N = 3351 | **2017 - 2018**, N = 2551 | **2018 - 2019**, N = 2421 | **2019 - 2020**, N = 3921 | | --- | --- | --- | --- | --- | --- | --- | --- | | Study |  |  |  |  |  |  |  | | FL | 60 (45%) | 150 (63%) | 130 (61%) | 125 (37%) | 0 (0%) | 0 (0%) | 0 (0%) | | PA | 73 (55%) | 88 (37%) | 82 (39%) | 65 (19%) | 0 (0%) | 0 (0%) | 0 (0%) | | UGA | 0 (0%) | 0 (0%) | 0 (0%) | 145 (43%) | 255 (100%) | 242 (100%) | 392 (100%) | | Age | 58 (35, 66) | 57 (39, 64) | 56 (35, 63) | 44 (27, 60) | 25 (17, 45) | 16 (14, 27) | 37 (17, 52) | | Sex |  |  |  |  |  |  |  | | Female | 100 (75%) | 172 (72%) | 159 (75%) | 226 (67%) | 142 (56%) | 137 (57%) | 232 (59%) | | Male | 33 (25%) | 66 (28%) | 53 (25%) | 109 (33%) | 113 (44%) | 105 (43%) | 160 (41%) | | Race |  |  |  |  |  |  |  | | White | 100 (76%) | 158 (67%) | 149 (71%) | 241 (72%) | 205 (80%) | 203 (84%) | 324 (83%) | | Black | 27 (21%) | 44 (19%) | 39 (19%) | 46 (14%) | 19 (7.5%) | 11 (4.5%) | 33 (8.4%) | | Hispanic | 2 (1.5%) | 27 (11%) | 17 (8.1%) | 24 (7.2%) | 16 (6.3%) | 11 (4.5%) | 13 (3.3%) | | Other | 2 (1.5%) | 6 (2.6%) | 5 (2.4%) | 23 (6.9%) | 15 (5.9%) | 17 (7.0%) | 22 (5.6%) | | Unknown | 2 | 3 | 2 | 1 | 0 | 0 | 0 | | 1n (%); Median (IQR) | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Strains used for HAI assays on cohort study samples, and the seasons in which those strains were used.   | Strain | 14/15 | 15/16 | 16/17 | 17/18 | 18/19 | 19/20 | | --- | --- | --- | --- | --- | --- | --- | | SC/18 | x | x | x | x |  |  | | Wei/43 | x | x | x | x |  |  | | FM/47 | x | x | x | x |  |  | | Den/57 | x | x | x | x |  |  | | NJ/76 | x | x | x | x |  |  | | USSR/77 | x | x | x | x |  |  | | Bra/78 |  |  | x | x |  |  | | CA/78 | x | x |  |  |  |  | | Chi/83 | x | x | x | x | x | x | | Sing/86 | x | x | x | x | x |  | | TX/91 | x | x | x | x | x |  | | Bei/95 | x | x | x | x | x |  | | NC/99 | x | x | x | x | x |  | | SI/06 | x | x | x | x | x |  | | Bris/07 | x | x | x | x | x | x | | CA/09 | x | x | x | x | x | x | | MI/15 |  |  | x | x | x | x | | Bris/18 |  |  |  |  |  | x | | HK/68 | x | x | x | x |  |  | | PC/73 | x | x | x | x |  |  | | TX/77 | x | x | x | x |  |  | | MI/85 | x | x | x | x |  |  | | Sich/87 | x | x | x | x |  |  | | Shan/93 | x | x | x | x |  |  | | Nan/95 | x | x | x | x |  |  | | Syd/97 | x | x | x | x |  |  | | Pan/99 | x | x | x | x | x | x | | Fuj/02 | x | x |  |  |  |  | | NY/04 | x | x | x | x | x |  | | WI/05 | x | x | x | x | x |  | | Uru/07 | x | x | x | x | x |  | | Per/09 | x | x | x | x | x |  | | Vic/11 | x | x | x | x | x |  | | TX/12 | x | x | x | x | x | x | | Switz/13 | x | x | x | x | x | x | | HK/14 | x | x | x | x | x | x | | Sing/16 |  |  |  | x | x | x | | KS/17 |  |  |  |  |  | x | | SA/19 |  |  |  |  |  | x | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: The strains used in the study along with the GISAID or UniProt accession numbers for the protein sequence of the virus stock used in the lab. For one sequence, the full length is not available. The nucleic acid sequences are also available for each of the strains, from the provided sources.   | Strain Name | Abbreviation | Full Length? | HA Sequence Source | | --- | --- | --- | --- | | A/H1N1/South Carolina/1/1918 | SC/18 | Yes | UniProt: Q9WFX3 | | A/H1N1/Weiss/43 | Wei/43 | Yes | UniProt: Q20N27 | | A/H1N1/Fort Monmouth/1/1947 | FM/47 | Yes | UniProt: Q84110 | | A/H1N1/Denver/1957 | Den/57 | Yes | UniProt: Q2IBI1 | | A/H1N1/New Jersey/8/1976 | NJ/76 | Yes | UniProt: L0L141 | | A/H1N1/Ussr/90/1977 | USSR/77 | Yes | UniProt: P03453 | | A/H1N1/Brazil/11/1978 | Bra/78 | Yes | UniProt: A4GBX7 | | A/H1N1/California/10/1978 | CA/78 | Yes | UniProt: A4U6W3 | | A/H1N1/Chile/1/1983 | Chi/83 | Yes | UniProt: A4GCH5 | | A/H1N1/Singapore/6/1986 | Sing/86 | Yes | UniProt: A4GCN0 | | A/H1N1/Texas/36/1991 | TX/91 | Yes | UniProt: B4UPL3 | | A/H1N1/Beijing/262/1995 | Bei/95 | Yes | UniProt: B4UPF7 | | A/H1N1/New Caledonia/20/1999 | NC/99 | Yes | UniProt: Q6WG00 | | A/H1N1/Solomon Islands/3/2006 | SI/06 | Yes | UniProt: A7Y8I1 | | A/H1N1/Brisbane/59/2007 | Bris/07 | Yes | UniProt: D5F1Q8 | | A/H1N1/California/07/2009 | CA/09 | Yes | UniProt: C3W5X2 | | A/H1N1/Michigan 45/2015 | MI/15 | Yes | UniProt: A0A144YDV8 | | A/H1N1/Brisbane/02/2018 | Bris/18 | Yes | Gisaid: EPI1415369 | | A/H3N2/Hong Kong/8/1968 | HK/68 | Yes | UniProt: A6YBG1 | | A/H3N2/Port Chalmers/1/1973 | PC/73 | Yes | UniProt: Q1PUD9 | | A/H3N2/Texas/1/1977 | TX/77 | Yes | UniProt: I6RX51 | | A/H3N2/Mississippi/1/1985 | MI/85 | No (17-345AA) | Gisaid: EPI129066; Uniprot: Q67178 | | A/H3N2/Sichuan/2/1987 | Sich/87 | Yes | UniProt: H9XCU1 | | A/H3N2/Shangdong/9/1993 | Shan/93 | Yes | UniProt: H9XM74 | | A/H3N2/Nanchang/933/1995 | Nan/95 | Yes | UniProt: H9XED1 | | A/H3N2/Sydney/5/1997 | Syd/97 | Yes | UniProt: C3PR59 | | A/H3N2/Panama/2007/1999 | Pan/99 | Yes | UniProt: Q1K9M3 | | A/H3N2/Fujian/411/2002 | Fuj/02 | Yes | UniProt: H9XEX9 | | A/H3N2/New York/55/2004 | NY/04 | Yes | UniProt: B4UPJ0 | | A/H3N2/Wisconsin/67/2005 | WI/05 | Yes | UniProt: W0RXT2 | | A/H3N2/Uruguay/716/2007 | Uru/07 | Yes | UniProt: B2ZV32 | | A/H3N2/Perth/16/2009 | Per/09 | Yes | UniProt: C6KNH7 | | A/H3N2/Victoria/361/2011 | Vic/11 | Yes | UniProt: A0A097PF39 | | A/H3N2/Texas/50/2012 | TX/12 | Yes | UniProt: R4L4F3 | | A/H3N2/Switzerland/9715293/2013 | Switz/13 | Yes | Gisaid: EPI530687 | | A/H3N2/Hong Kong/4801/2014 | HK/14 | Yes | Gisaid: EPI834581 | | A/H3N2/Singapore/infimh-16-0019/2016 | Sing/16 | Yes | Gisaid: EPI780183 | | A/H3N2/Kansas/14/2017 | KS/17 | Yes | UniProt: A0A2L2FM43 | | A/H3N2/South Australia/34/2019 | SA/19 | Yes | Gisaid: EPI1387331 | |

# References