Mutation Rates of Influenza A and Influenza B Over Time

Madison Bowles

Evolution 461

**Introduction**

Although the influenza virus has been around for a thousand or so years, it was only recently isolated from humans and identified as a virus in the 1930s. Since then, we know it has resulted in millions of deaths worldwide, and has sparked a continuation of research into controlling the influenza virus (2019 Apr 18). Currently, there are four known strains of influenza virus: A, B, C, and D. The first two are responsible for what is widely known as the seasonal or winter flu. Influenza A and B can also subdivided into subtypes (for A: H1N1 and H3N2) or lineages (for B: Victoria or Yamagata), then clades, and finally subgroups (CDC 2019 Nov 18). However, this virus is also known for its rapid evolution and mutation rates that has caused the emergence of pandemics within human populations and a vaccine that is comparably one of the least effective within the United States

**Materials and Methods**

*Data and Calculations*

To test the hypothesis, data was obtained from the paper “The Influenza Virus Resource at the National Center for Biotechnology Information” where a link to the NCBI Influenza Virus Resource Database (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi>) contained the genetic sequences, proteins, or coding regions from the different strains of influenza virus sampled and compiled over the years (Information et al.). Each of these individual viruses sampled contained 16 different metadata points, a few of which were date, accession, nucleotide length, host, country, region, virus name, mutation, and lineage. For the purpose of this experiment, the country, date, virus name, lineage, and mutations were kept to run analyses on.

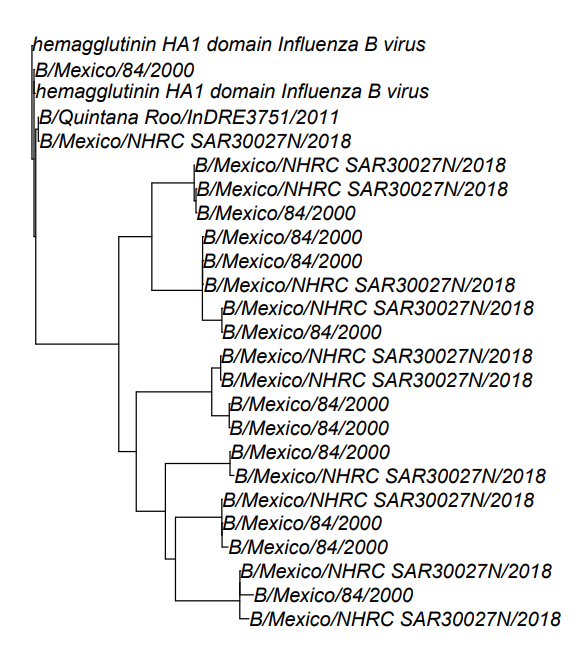
The phylogenies, like Figure 1, were created also using the Influenza Virus Resource Database. Once the specific subset of the data information such as type, host, country/region was selected, the database would then randomly select 1000 sequences to create a phylogeny such as Figure 1. 

Figure . Phylogeny of Influenza B recorded and sequenced in Human hosts located in Mexico between the year 2000 and 2018.

To be able to calculate mutation rate for Influenza A and B in each country, the phylogeny obtained were evaluated using DR metrics in RStudio using the R scripts and data files to estimate speciation rates on phylogenies of different geography using the inverse equal to the splits in the tree (Title and Rabosky 2019, RStudio 2020). The NCBI Influenza Database that built the trees, used the number of mutations as the distance between each node or “edge”, thus speciation rate on the phylogeny is equal to the mutation rate (Information et al.).

*DR Statistics*

The DR statistic is

*Boxplot of Speciation Rate*

**References**

CDC. 2019 Nov 18. Types of Influenza Viruses. Centers for Disease Control and Prevention. [accessed 2021 Mar 26]. https://www.cdc.gov/flu/about/viruses/types.htm.

Influenza Historic Timeline | Pandemic Influenza (Flu) | CDC. 2019 Apr 18. [accessed 2021 Mar 28]. https://www.cdc.gov/flu/pandemic-resources/pandemic-timeline-1930-and-beyond.htm.

Information NC for B, Pike USNL of M 8600 R, MD B, Usa 20894. Influenza virus database - NCBI. [accessed 2021 Mar 29]. https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database.

Title PO, Rabosky DL. 2019. Data from: Tip rates, phylogenies, and diversification: what are we estimating, and how good are the estimates? :3905386069 bytes. doi:10.5061/DRYAD.5HR25NV. [accessed 2021 Mar 31]. http://datadryad.org/stash/dataset/doi:10.5061/dryad.5hr25nv.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.