

Predicting Influenza Strains Based on Correlations Between Northern and Southern Hemispheres

Jonathan Harney, Michael Harney*

*mharney1268@yahoo.com

Abstract: There is research showing that influenza strains are correlated between the Northern and Southern hemispheres. If this data is corroborated, this may lead to a prediction algorithm that relies on the predominant influenza strain in one hemisphere that is coming out of its flu season in order to predict the strain in the opposing hemisphere that is going into its flu season. This may provide enough time to develop a vaccine that is more accurately targeting the correct strain than in previous years.

I. Introduction

Predicting the predominant influenza strain each season is a challenge. In order to provide the millions of vaccines necessary for the beginning of each flu season, the CDC and World Health Organization need to predict the influenza strain that is expected from 6 to 9 months in advance to verify a vaccine in the laboratory and manufacture the many samples required. There is research showing that influenza strains are correlated between the Northern and Southern hemispheres [1]. If this data is corroborated, this may lead to a prediction algorithm that relies on the predominant influenza strain in one hemisphere that is coming out of its flu season in order to predict the strain in the opposing hemisphere that is going into its flu season. This may provide enough time to develop a vaccine that is more accurately targeting the correct strain than in previous years.

II. Procedure

By analyzing the cDNA sequences of the H1N1 and H3N2 influenza strains in the northern and southern hemispheres over the past 20 years, we work to obtain a correlation between the mutated strains that arise in one hemisphere based on the previous hemisphere's influenza season. The historical data for influenza strains over the last 20 years was

tabulated by the WorldHealth Organization as shown in Table 1

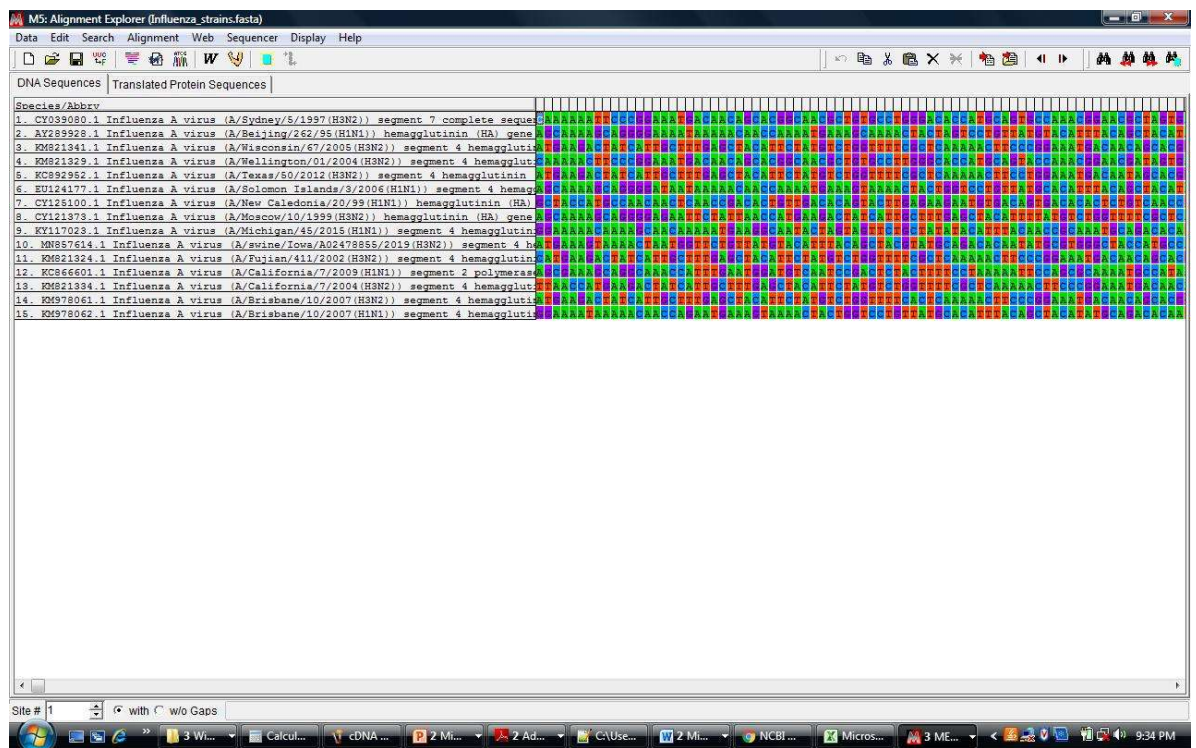
NH winter season	NH H1N1	SH winter season	SH H1N1	NH H3N2	SH H3N2	NH B-strain	SH B-strain
November 1998–April 1999 [2]	A/Beijing/262/95 (H1N1)-like virus	1999 [31]	A/Beijing/262/95 (H1N1)-like virus	A/Sydney/5/97 (H3N2)-like virus	A/Sydney/5/97 (H3N2)-like virus	B/Beijing/184/93-like virus	B/Beijing/184/93-like virus
November 1999–April 2000 [3]	A/Beijing/262/95 (H1N1)-like virus	May–October 2000 [32]	A/New Caledonia/20/99 (H1N1)-like virus	A/Sydney/5/97 (H3N2)-like virus	A/Moscow/10/99 (H3N2)-like virus	B/Beijing/184/93-like virus or B/Shangdong/7/97-like virus	B/Beijing/184/93-like virus or B/Shangdong/7/97-like virus
2000–2001 [4]	A/New Caledonia/20/99 (H1N1)-like virus	May–October 2001 [33]	A/New Caledonia/20/99 (H1N1)-like virus	A/Moscow/10/99 (H3N2)-like virus	A/Moscow/10/99 (H3N2)-like virus	B/Beijing/184/93-like virus	B/Sichuan/379/99-like virus
2001–2002 [5]	A/New Caledonia/20/99 (H1N1)-like virus	2002 [34]	A/New Caledonia/20/99 (H1N1)-like virus	A/Moscow/10/99 (H3N2)-like virus	A/Moscow/10/99 (H3N2)-like virus	B/Sichuan/379/99-like virus	B/Sichuan/379/99-like virus
2002–2003 [6]	A/New Caledonia/20/99 (H1N1)-like virus	2003 [35]	A/New Caledonia/20/99 (H1N1)-like virus	A/Moscow/10/99 (H3N2)-like virus	A/Moscow/10/99 (H3N2)-like virus	B/Hong Kong/330/2001-like virus	B/Hong Kong/330/2001-like virus
2003–2004 [7]	A/New Caledonia/20/99 (H1N1)-like virus	2004 [36]	A/New Caledonia/20/99 (H1N1)-like virus	A/Moscow/10/99 (H3N2)-like virus	A/Fujian/411/2002 (H3N2)-like virus	B/Hong Kong/330/2001-like virus	B/Hong Kong/330/2001-like virus
2004–2005 [8]	A/New Caledonia/20/99 (H1N1)-like virus	2005 [37]	A/New Caledonia/20/99 (H1N1)-like virus	A/Fujian/411/2002 (H3N2)-like virus	A/Wellington/1/2004 (H3N2)-like virus	B/Shanghai/361/2002-like virus	B/Shanghai/361/2002-like virus
2005–2006 [9]	A/New Caledonia/20/99 (H1N1)-like virus	2006 [38]	A/New Caledonia/20/99 (H1N1)-like virus	A/California/7/2004 (H3N2)-like virus	A/California/7/2004 (H3N2)-like virus	B/Shanghai/361/2002-like virus	B/Malaysia/2506/2004-like virus
2006–2007 [10]	A/New Caledonia/20/99 (H1N1)-like virus	2007 [39]	A/New Caledonia/20/99 (H1N1)-like virus	A/Wisconsin/67/2005 (H3N2)-like virus	A/Wisconsin/67/2005 (H3N2)-like virus	B/Malaysia/2506/2004-like virus	B/Malaysia/2506/2004-like virus
2007–2008 [11]	A/Solomon Islands/3/2006 (H1N1)-like virus	2008 [40]	A/Solomon Islands/3/2006 (H1N1)-like virus	A/Wisconsin/67/2005 (H3N2)-like virus	A/Brisbane/10/2007 (H3N2)-like virus	B/Malaysia/2506/2004-like virus	B/Florida/4/2006-like virus
2008–2009 [12]	A/Brisbane/59/2007 (H1N1)-like virus	2009 [41]	A/Brisbane/59/2007 (H1N1)-like virus	A/Brisbane/10/2007 (H3N2)-like virus	A/Brisbane/10/2007 (H3N2)-like virus	B/Florida/4/2006-like virus	B/Florida/4/2006-like virus
2009–2010 [13]	A/Brisbane/59/2007 (H1N1)-like virus	2010 [42]	A/Brisbane/59/2007 (H1N1)-like virus	A/California/7/2009 (H1N1)-like virus	A/Brisbane/10/2007 (H3N2)-like virus	A/Perth/16/2009 (H3N2)-like virus	B/Brisbane/60/2008-like virus

below [2].

NH winter season	SH		NH H3N2	SH H3N2	NH B-strain	SH B-strain	
	NH H1N1	winter season					SH H1N1
2010–2011[14]	A/California/7/2009 (H1N1)-like virus	2011[43]	A/California/7/2009 (H1N1)-like virus	A/Perth/16/2009 (H3N2)-like virus	A/Perth/16/2009 (H3N2)-like virus	B/Brisbane/60/2008-like virus	B/Brisbane/60/2008-like virus
2011–2012[15]	A/California/7/2009 (H1N1)-like virus	2012[44]	A/California/7/2009 (H1N1)pdm09 ^{not} ^{e 1} -like virus ^[17]	A/Perth/16/2009 (H3N2)-like virus	A/Perth/16/2009 (H3N2)-like virus	B/Brisbane/60/2008-like virus	B/Brisbane/60/2008-like virus
2012–2013[16]	A/California/7/2009 (H1N1)pdm09 ^{note 1} -like virus ^[17]	2013[45]	A/California/7/2009 (H1N1)pdm09-like virus	A/Victoria/361/2011 (H3N2)-like virus	A/Victoria/361/2011 (H3N2)-like virus	B/Wisconsin/1/2010-like virus	B/Wisconsin/1/2010-like virus
2013–2014[18]	A/California/7/2009 (H1N1)pdm09[not e 1]-like virus	2014[46]	A/California/7/2009 (H1N1)pdm09-like virus	A(H3N2) virus antigenically like the cell-propagated prototype virus A/Victoria/361/2011[note 2]	A/Texas/50/2012 (H3N2)-like virus[note 3]	B/Massachusetts/2/2012-like virus	B/Massachusetts/2/2012-like virus
2014–2015[19]	A/California/7/2009 (H1N1)pdm09[not e 1]-like virus	2015[47]	A/California/7/2009 (H1N1)pdm09-like virus	A/Texas/50/2012 (H3N2)-like virus[note 3]	A/Switzerland/9715293/2013 (H3N2)-like virus	B/Massachusetts/2/2012-like virus	B/Phuket/3073/2013-like virus
2015–2016[20]	A/California/7/2009 (H1N1)pdm09[not e 1]-like virus	2016[48]	A/California/7/2009 (H1N1)pdm09-like virus	A/Switzerland/9715293/2013 (H3N2)-like virus	A/Hong Kong/4801/2014 (H3N2)-like virus	B/Phuket/3073/2013-like virus	B/Brisbane/60/2008-like virus
2016–2017^{[21][22]}	A/California/7/2009 (H1N1)pdm09[not e 1]-like virus	2017^{[49][50]}	A/Michigan/45/2015 (H1N1)pdm09-like virus[25]	A/Hong Kong/4801/2014 (H3N2)-like virus	A/Hong Kong/4801/2014 (H3N2)-like virus	B/Brisbane/60/2008-like virus	B/Brisbane/60/2008-like virus
2017–2018^{[23][24]}	A/Michigan/45/2015 (H1N1)pdm09 ^{note 1} -like virus ^[25]	2018^{[51][52]}	A/Michigan/45/2015 (H1N1)pdm09-like virus	A/Hong Kong/4801/2014 (H3N2)-like virus	A/Singapore/INFIMH-16-0019/2016 (H3N2)-like virus	B/Brisbane/60/2008-like virus	B/Phuket/3073/2013-like virus
2018–2019^{[26][27]}	A/Michigan/45/2015 (H1N1)pdm09[not e 1]-like virus	2019^{[53][54]}	A/Michigan/45/2015 (H1N1)pdm09-like virus	A/Singapore/INFIMH-16-0019/2016 (H3N2)-like virus	A/Switzerland/8060/2017 (H3N2)-like virus	B/Colorado/06/2017-like virus (B/Victoria/2/87 lineage)	B/Colorado/06/2017-like virus (B/Victoria/2/87 lineage)
2019–2020^{[28][29][30]}	A/Brisbane/02/2018 (H1N1)pdm09[not e 1]-like virus	2020^{[55][56]}	A/Brisbane/02/2018 (H1N1)pdm09-like virus	A/Kansas/14/2017 (H3N2)-like virus	A/South Australia/34/2019 (H3N2)-like virus	B/Colorado/06/2017-like virus (B/Victoria/2/87 lineage)	B/Washington/02/2019-like virus (B/Victoria/2/87 lineage)

Table 1. Historical Influenza Strains for the Last 20 Years In Each Hemisphere

We retrieved sequence data (cDNA data derived from RNA sequencing) from NCBI and the Influenza Research Database (www.fludb.org) of the H1N1 and H3N2 influenza strains in both the northern and southern hemispheres over the last 20 years. We then performed a multiple sequence alignment using the ClustalW algorithm on the Hemagglutinin (HA) gene in each of the strains using Mega 5 (Figure 1).



**Figure 1. Multiple Sequence Alignment of all Strains
From 1999- 2020**

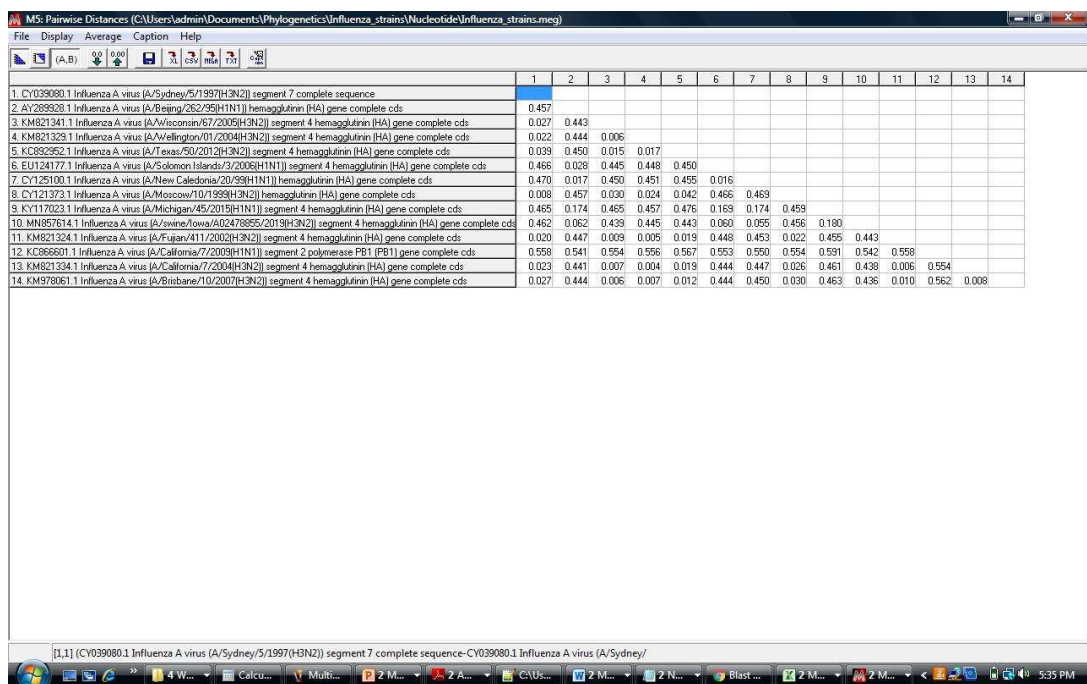


Figure 2. PairWise Distance Matrix From PAM30 Analysis

Based on the PAM30 substitution matrix, a pairwise distance estimate between the strains was constructed (Figure 2) along with a phylogenetic tree (Figure 3).

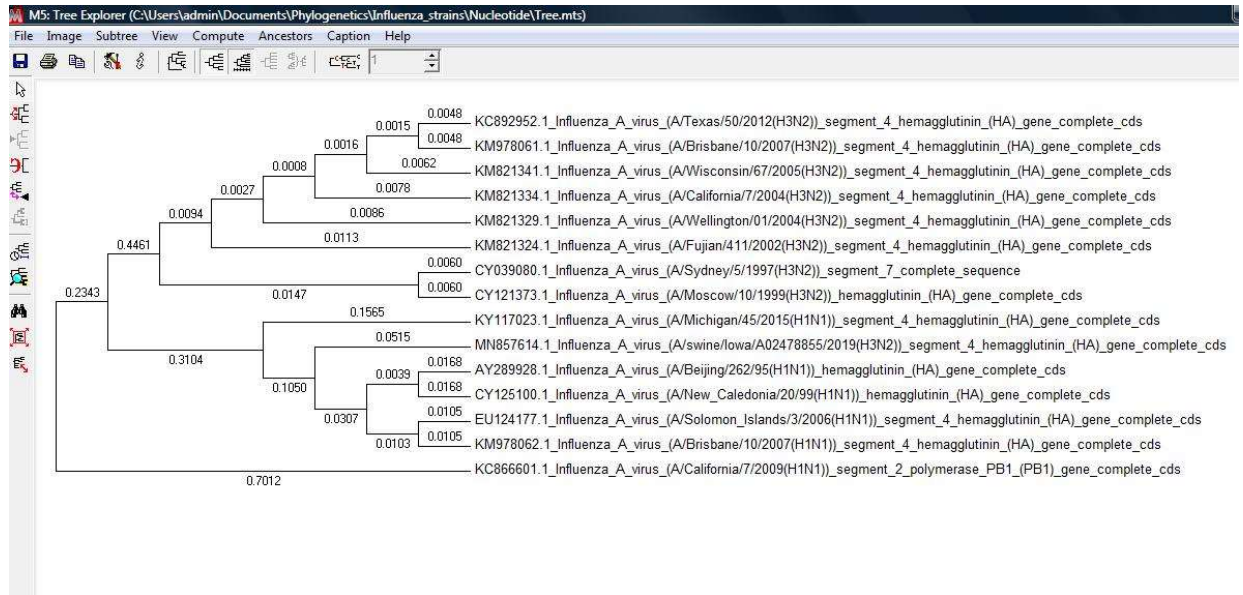


Figure 3. Phylogenetic Tree Showing Relationships Between Strains

The phylogenetic tree shows how closely the strains are related based on the pairwise distance matrix. It is theorized that most mutations between strains will be less than one branch change in the tree. After examining the changes in the 20 year historical data, this assumption appears to be validated.

One newer strain of H3N2 (A\swine\Iowa\H3N2\2019) emerged in the sub-branch of H1N1 next to (A\Michigan\H1N1\2015), we suspect that the new 2019 sequences of A\swine\Iowa\H3N2\2019 may be unverified as the rest of the strains collapse into their respective H1N1 or H3N2 sub-branches as expected.

III. Prediction Algorithm

The proposal is to apply supervised machine learning in the form of neural net which uses the historical data (Table 1) as training data and the pairwise distance matrix (Figure 2) as the node weights that are used in the computational stage of the neural net. This will keep consistency between single sub-branch mutations as shown in the pairwise distance matrix (and associated tree) but also incorporate historical data to account for the mutations that are more than one sub-branch away. The proposed architecture is shown in Figure 4.

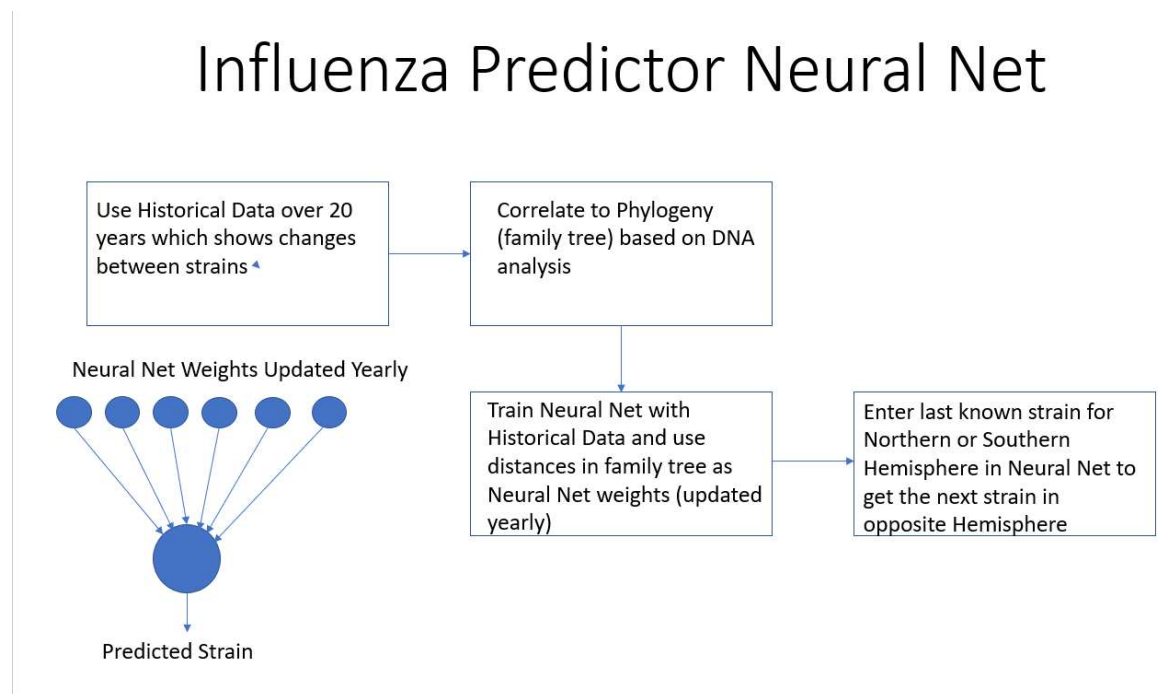


Figure 4. Influenza Predictor

IV. Conclusions

Using the research performed on correlation between influenza strains in the northern and southern hemispheres to predict upcoming influenza epidemics, we believe there is evidence that the strain for each hemisphere can be predicted with some reliability. The use of historical data and genomic analysis may provide a path for machine learning to determine future predominant strains in each hemisphere based on the flu season results in the opposing hemisphere.

References

1. Zhang and Yakob, “Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local internet query data”, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6397245/>
2. Historical annual reformulations of the influenza vaccine, Wikipedia
https://en.m.wikipedia.org/wiki/Historical_annual_reformulations_of_the_influenza_vaccine