A Homological Comparison of the COVID-19 Genome with Influenza Strains from the Last 20 Years

Michael Harney*, Jonathan Harney

*mharney1268@yahoo.com

Abstract: A comparative genomic analysis of the recent 2019-nCoV (COVID-19) virus with the known influenza strains of the last 20 years is performed using the current sequence information of 2019-nCoV and the influenza sequences provided by NCBI and FluDB. The analysis uses a ML model selection and shows the most common homology between 2019-nCoV and the influenza-A strain (A/Texas/50/2012(H3N2)) with a pairwise distance of 2.207, with the next two strains that have the most similar homology as A/Fujian/411/2002(H3N2) and A/California/7/2009(H1N1) with a pairwise distance of 2.223 and 2.243 respectively. The analysis of these genetic similarities may allow for the investigation and use of a current treatments or vaccines for 2019-nCoV based on the similarities to these previous strains, although the specific domain analysis of how the anti-viral features work would have to be evaluated. These older treatments may be partially effective while the new trials are ongoing for 2019-nCoV.

I. Introduction

The 2019-nCoV (COVID-19) virus that originated in Wuhan China is 29,903 base pairs long and was sequenced and uploaded to NCBI in February 2020 [1][2]. By comparing the sequence data for influenza A strains for the last 20 years with 2019-nCoV, we look for similarities that may be useful in utilizing current treatments or vaccines for the previous strains that may be partially effective until new treatments are available.

II. Procedure

By performing a multiple sequence analysis of the cDNA sequences of the H1N1 and H3N2 in the northern and southern hemispheres over the past 20 years along with the 2019-nCoV sequence, we look for homology between the strains based on an ML model of viral evolution. The historical data for influenza strains over the last 20 years was tabulated by the WorldHealth

Organization as shown in Table 1 below [3].

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	2007[39]	A/New Caledonia/20/99(H1N1)-like	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/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	B/Brisbane/60/2008-like virus

NH winter season	NH H1N1	SH winter season	SH H1N1	NH H3N2	SH H3N2	NH B-strain	SH B-strain
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[note 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[note 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[note 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[note 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[note 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[note 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[note 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 (B/Victoria lineage) virus

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

We retrieved cDNA sequence data 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 using Mega5 on the Hemagglutimin (HA) gene (for brevity of initial comparison) in each of the influenza strains along with the sequence for 2019-nCoV (Figure 1).

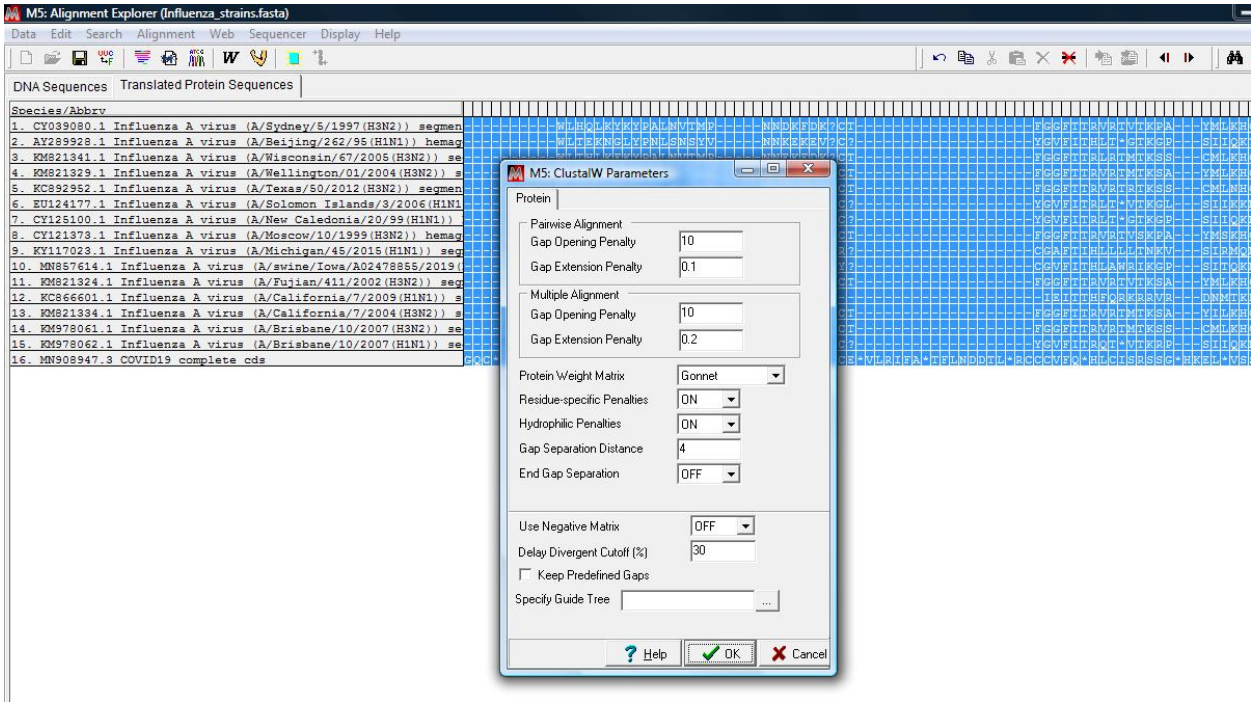
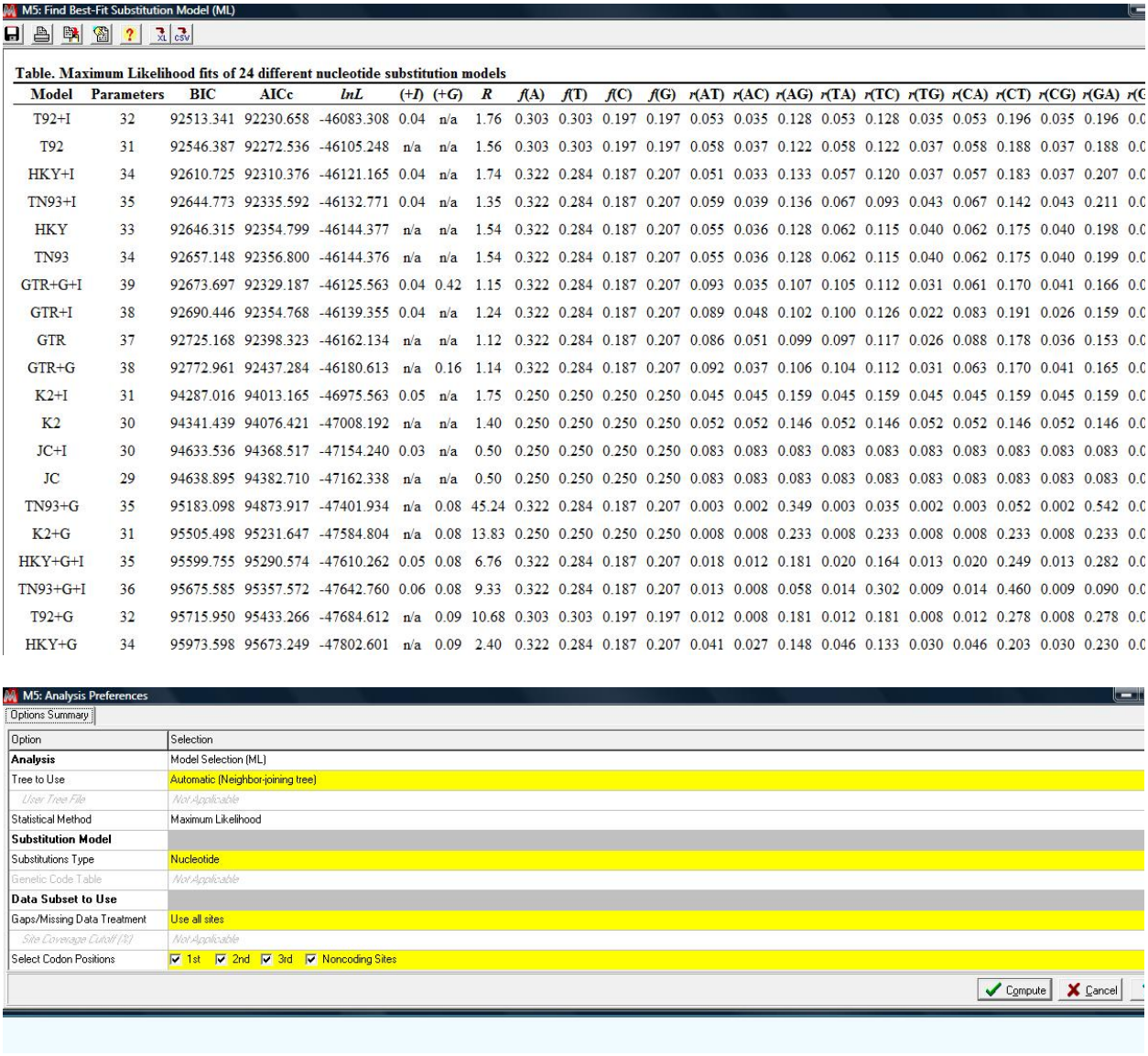


Figure 1. Multiple Sequence Alignment of all A-Strains From 1999-2020 and 2019-nCoV

After alignment, several evolutionary models were investigated with Maximum Likelihood methods as shown in Figure 2.



M5: Find Best-Fit Substitution Model (ML)

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(G
T92+I	32	92513.341	92230.658	-46083.308	0.04	n/a	1.76	0.303	0.303	0.197	0.197	0.053	0.035	0.128	0.053	0.128	0.035	0.053	0.196	0.035	0.196	0.0
T92	31	92546.387	92272.536	-46105.248	n/a	n/a	1.56	0.303	0.303	0.197	0.197	0.058	0.037	0.122	0.058	0.122	0.037	0.058	0.188	0.037	0.188	0.0
HKY+I	34	92610.725	92310.376	-46121.165	0.04	n/a	1.74	0.322	0.284	0.187	0.207	0.051	0.033	0.133	0.057	0.120	0.037	0.057	0.183	0.037	0.207	0.0
TN93+I	35	92644.773	92335.592	-46132.771	0.04	n/a	1.35	0.322	0.284	0.187	0.207	0.059	0.039	0.136	0.067	0.093	0.043	0.067	0.142	0.043	0.211	0.0
HKY	33	92646.315	92354.799	-46144.377	n/a	n/a	1.54	0.322	0.284	0.187	0.207	0.055	0.036	0.128	0.062	0.115	0.040	0.062	0.175	0.040	0.198	0.0
TN93	34	92657.148	92356.800	-46144.376	n/a	n/a	1.54	0.322	0.284	0.187	0.207	0.055	0.036	0.128	0.062	0.115	0.040	0.062	0.175	0.040	0.199	0.0
GTR+G+I	39	92673.697	92329.187	-46125.563	0.04	0.42	1.15	0.322	0.284	0.187	0.207	0.093	0.035	0.107	0.105	0.112	0.031	0.061	0.170	0.041	0.166	0.0
GTR+I	38	92690.446	92354.768	-46139.355	0.04	n/a	1.24	0.322	0.284	0.187	0.207	0.089	0.048	0.102	0.100	0.126	0.022	0.083	0.191	0.026	0.159	0.0
GTR	37	92725.168	92398.323	-46162.134	n/a	n/a	1.12	0.322	0.284	0.187	0.207	0.086	0.051	0.099	0.097	0.117	0.026	0.088	0.178	0.036	0.153	0.0
GTR+G	38	92772.961	92437.284	-46180.613	n/a	0.16	1.14	0.322	0.284	0.187	0.207	0.092	0.037	0.106	0.104	0.112	0.031	0.063	0.170	0.041	0.165	0.0
K2+I	31	94287.016	94013.165	-46975.563	0.05	n/a	1.75	0.250	0.250	0.250	0.250	0.045	0.045	0.159	0.045	0.159	0.045	0.045	0.159	0.045	0.159	0.0
K2	30	94341.439	94076.421	-47008.192	n/a	n/a	1.40	0.250	0.250	0.250	0.250	0.052	0.052	0.146	0.052	0.146	0.052	0.052	0.146	0.052	0.146	0.0
JC+I	30	94633.536	94368.517	-47154.240	0.03	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.0
JC	29	94638.895	94382.710	-47162.338	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.0
TN93+G	35	95183.098	94873.917	-47401.934	n/a	0.08	45.24	0.322	0.284	0.187	0.207	0.003	0.002	0.349	0.003	0.035	0.002	0.003	0.052	0.002	0.542	0.0
K2+G	31	95505.498	95231.647	-47584.804	n/a	0.08	13.83	0.250	0.250	0.250	0.250	0.008	0.008	0.233	0.008	0.233	0.008	0.008	0.233	0.008	0.233	0.0
HKY+G+I	35	95599.755	95290.574	-47610.262	0.05	0.08	6.76	0.322	0.284	0.187	0.207	0.018	0.012	0.181	0.020	0.164	0.013	0.020	0.249	0.013	0.282	0.0
TN93+G+I	36	95675.585	95357.572	-47642.760	0.06	0.08	9.33	0.322	0.284	0.187	0.207	0.013	0.008	0.058	0.014	0.302	0.009	0.014	0.460	0.009	0.090	0.0
T92+G	32	95715.950	95433.266	-47684.612	n/a	0.09	10.68	0.303	0.303	0.197	0.197	0.012	0.008	0.181	0.012	0.181	0.008	0.012	0.278	0.008	0.278	0.0
HKY+G	34	95973.598	95673.249	-47802.601	n/a	0.09	2.40	0.322	0.284	0.187	0.207	0.041	0.027	0.148	0.046	0.133	0.030	0.046	0.203	0.030	0.230	0.0

M5: Analysis Preferences

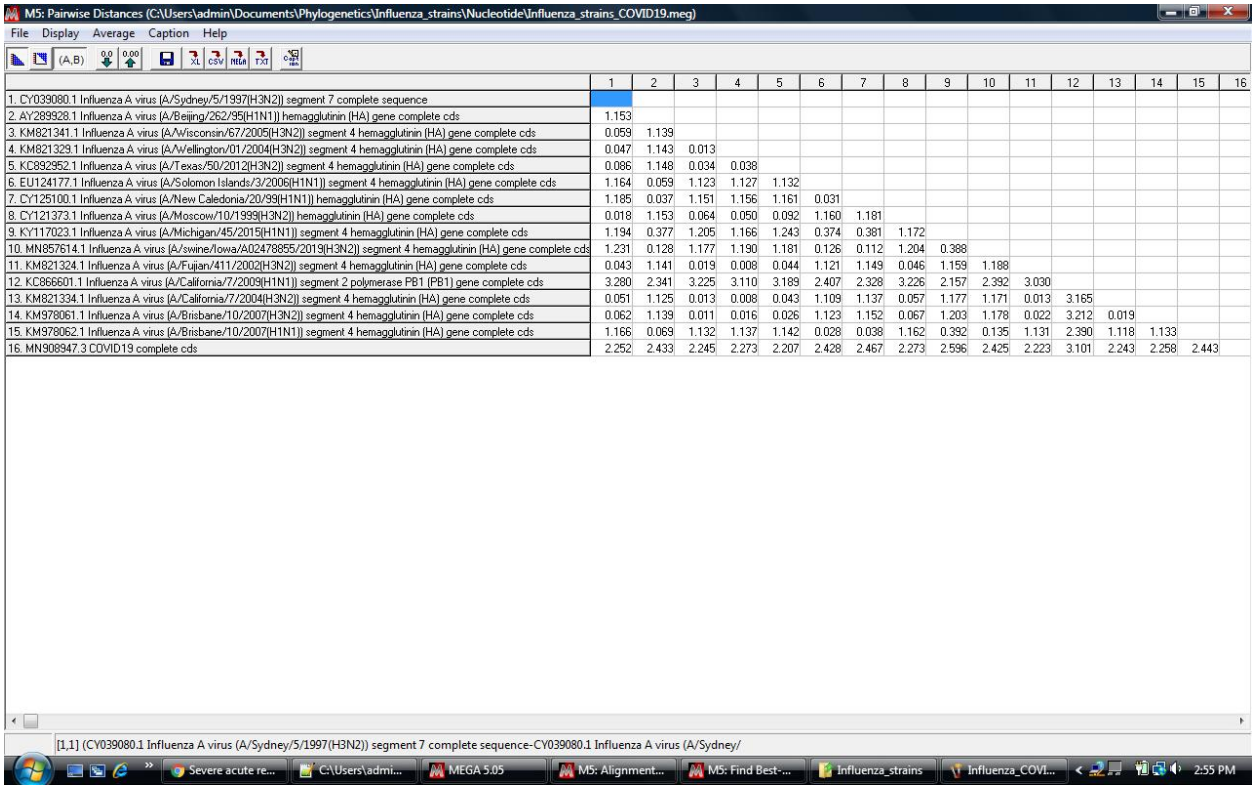
Options Summary

Option	Selection
Analysis	Model Selection (ML)
Tree to Use	Automatic (Neighbor-joining tree)
User Tree File	*Not Applicable*
Statistical Method	Maximum Likelihood
Substitution Model	
Substitutions Type	Nucleotide
Genetic Code Table	*Not Applicable*
Data Subset to Use	
Gaps/Missing Data Treatment	Use all sites
Site Coverage Cutoff (%)	*Not Applicable*
Select Codon Positions	☑ 1st ☑ 2nd ☑ 3rd ☑ Noncoding Sites

Compute Cancel

Figure 2. Model Selection based on Alignment and ML

A PAM30 model was selected and based on the data, a pairwise distance estimate between the influenza-A strains and 2019-nCoV was constructed (Figure 3) along with a phylogeny test tree (Figure 4).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. CY039080.1 Influenza A virus (A/Sydney/5/1997(H3N2)) segment 7 complete sequence																
2. AY289928.1 Influenza A virus (A/Beijing/262/95(H1N1)) hemagglutinin (HA) gene complete cds	1.153															
3. KM821341.1 Influenza A virus (A/Wisconsin/67/2005(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.059	1.139														
4. KM821329.1 Influenza A virus (A/Wellington/01/2004(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.047	1.143	0.013													
5. KC892952.1 Influenza A virus (A/Texas/50/2012(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.086	1.148	0.034	0.038												
6. EU124177.1 Influenza A virus (A/Solomon Islands/3/2006(H1N1)) segment 4 hemagglutinin (HA) gene complete cds	1.164	0.059	1.123	1.127	1.132											
7. CY125100.1 Influenza A virus (A/New Caledonia/20/99(H1N1)) hemagglutinin (HA) gene complete cds	1.185	0.037	1.151	1.156	1.161	0.031										
8. CY121373.1 Influenza A virus (A/Moscow/10/1999(H3N2)) hemagglutinin (HA) gene complete cds	0.018	1.153	0.064	0.050	0.092	1.160	1.181									
9. KY117023.1 Influenza A virus (A/Michigan/45/2015(H1N1)) segment 4 hemagglutinin (HA) gene complete cds	1.194	0.377	1.205	1.166	1.243	0.374	0.381	1.172								
10. MN857614.1 Influenza A virus (A/swine/Iowa/A02478855/2019(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	1.231	0.128	1.177	1.190	1.181	0.126	0.112	1.204	0.388							
11. KM821324.1 Influenza A virus (A/Fujian/411/2002(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.043	1.141	0.019	0.008	0.044	1.121	1.149	0.046	1.159	1.188						
12. KC866601.1 Influenza A virus (A/California/7/2009(H1N1)) segment 2 polymerase PB1 (PB1) gene complete cds	3.280	2.341	3.225	3.110	3.189	2.407	2.328	3.226	2.157	2.392	3.030					
13. KM821334.1 Influenza A virus (A/California/7/2004(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.051	1.125	0.013	0.008	0.043	1.109	1.137	0.057	1.177	1.171	0.013	3.165				
14. KM978061.1 Influenza A virus (A/Brisbane/10/2007(H3N2)) segment 4 hemagglutinin (HA) gene complete cds	0.062	1.139	0.011	0.016	0.026	1.123	1.152	0.067	1.203	1.178	0.022	3.212	0.019			
15. KM978062.1 Influenza A virus (A/Brisbane/10/2007(H1N1)) segment 4 hemagglutinin (HA) gene complete cds	1.166	0.069	1.132	1.137	1.142	0.028	0.038	1.162	0.392	0.135	1.131	2.390	1.118	1.133		
16. MN908947.3 COVID19 complete cds	2.252	2.433	2.245	2.273	2.207	2.428	2.467	2.273	2.596	2.425	2.223	3.101	2.243	2.258	2.443	

Figure 3. PairWise Distance Matrix

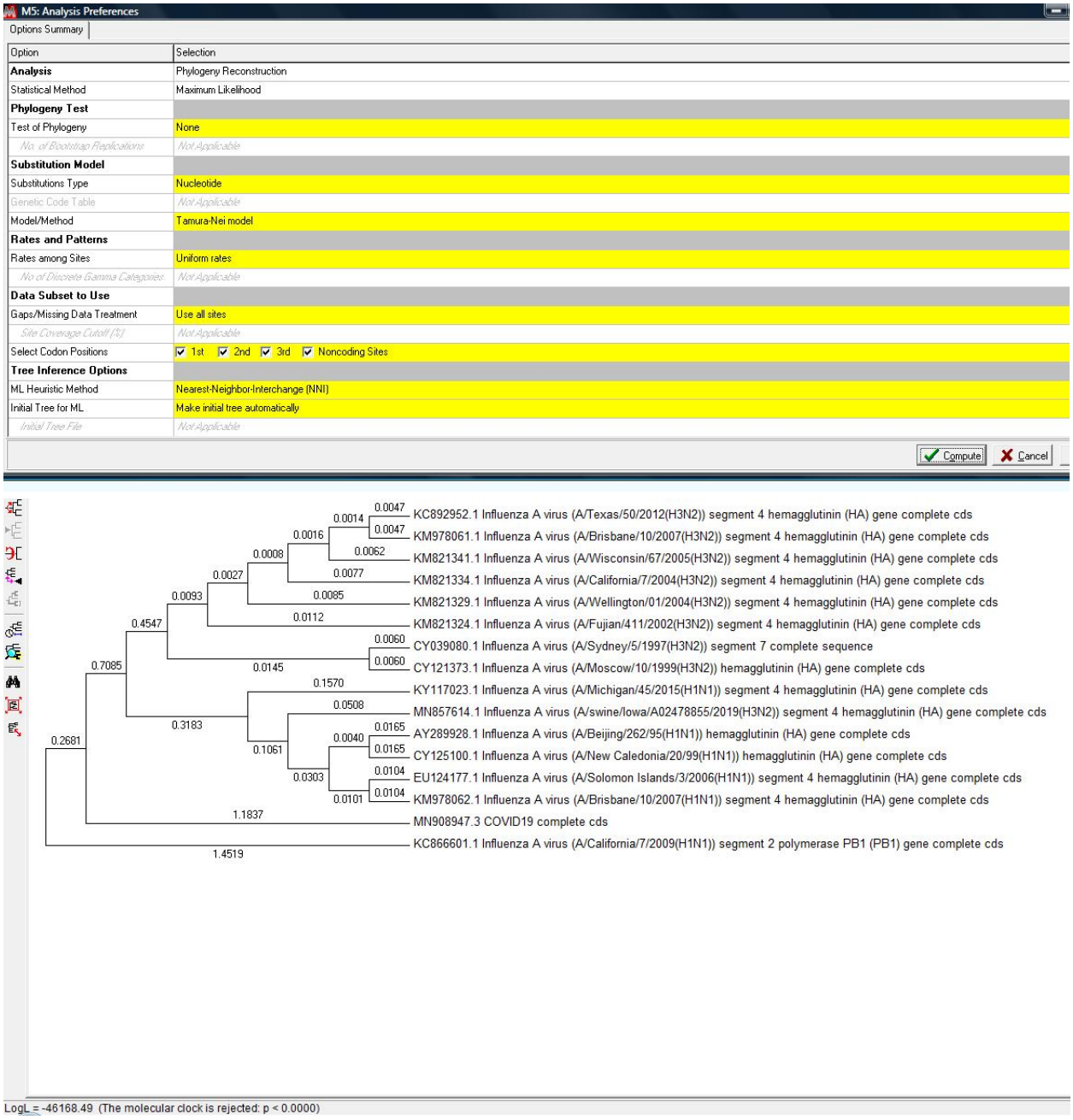


Figure 4. Phylogeny Test Tree

The phylogenetic tree shows how closely the strains are related based on the pairwise distance matrix. If it is the case that most

single-step mutations between strains will be less than one branch change in the tree, then this relationship is mostly validated for H1N1 and H3N2 strains which with one exception are in the same sub-branches.

The pairwise distance matrix in Figure 3 shows the shortest distance of 2.207 from 2019-nCoV to Influenza-A_virus_(A/Texas/50/2012(H3N2)) with the next two shortest distances being Influenza_A_virus_(A/Fujian/411/2002(H3N2) of 2.223 and Influenza_A_virus_(A/California/7/2009(H1N1)) with a distance of 2.243, then closely followed by Influenza_A_virus_(A/Wisconsin/67/2005(H3N2) with a distance of 2.245. These are the most likely candidates for further investigation of potential analogs to COVID-19.

III. Conclusions

The genetic sequences of the various influenza A strains of the last 20 years are compared against the recently sequenced 2019-nCoV with closest homology to the strain Influenza_A_virus_(A/Texas/50/2012(H3N2)), followed by Influenza_A_virus_(A/Fujian/411/2002(H3N2) and Influenza_A_virus_(A/California/7/2009(H1N1)). Further research is required to analyze the particular domains that are in common with these sequences and what differences in confirmation that the glycoprotein surfaces undergo in order to determine if the anti-viral treatments for older influenza strains

would truly be effective for 2019-nCoV. The shelf-life of any existing vaccine or anti-viral treatment would also have to be evaluated, although the possibility of some type of existing protection against 2019-nCoV is worth considering until a specific version of treatment for this new virus is available.

References

1. Zhang and Yakob, “DNA sleuths read the coronavirus genome, tracing its origins and looking for dangerous mutations”, StatNews, January 24, 2020
   https://www.statnews.com/2020/01/24/dna-sleuths-read-coronavirus-genome-tracing-origins-and-mutations

2. “Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome”, NCBI
   https://www.ncbi.nlm.nih.gov/nuccore/MN908947

3. “Historical annual reformulations of the influenza Vaccine”, Wikipedia
   https://en.m.wikipedia.org/wiki/Historical_annual_reformulations_of_the_influenza_vaccine