Annals of Medicine Research and Public Health
Research Article

Epidemiology for the 2009 'Swine flu' Pandemic Novel H1N1 Influenza A Virus using the PB2 Viral Gene

Arjavon Talebzadeh*

California Northstate University School of Medicine, USA

Citation: *Arjavon Talebzadeh. Epidemiology for the 2009 'Swine flu' Pandemic Novel H1N1 Influenza A Virus using the PB2 Viral Gene. Ann Med Res Pub Health. 2022;1(2):1-9.* DOI: https://doi.org/10.5281/zenodo.7850943

Received Date: 18 April, 2023; **Accepted Date:** 20 April, 2023; **Published Date:** 21 April, 2023

***Corresponding author:** Arjavon Talebzadeh. California Northstate University School of Medicine, USA

ABSTRACT

In 2009 Swine flu epidemic claimed significant number of the population globally. In this Article, we investigate the 2009 pH1N1 genetic sequences to track how and where the virus originated and spread from Mexico to USA and internationally. The PB2 gene on segment 1 of 8 segments was chosen as the baseline. Sequence alignment and Phylogeny of the segment was divided to two clades and evaluated. Based on our analysis, the gene sequences did reveal the origin of 2009 pH1N1 to Mexico with subsequent mutations as the virus spread globally.

Keywords: Swine flu; Pandemic; Virus

BACKGROUND

In the first year of the 2009 Swine flu pandemic, the U.S. Center for Disease Control (CDC) estimates that 2009 novel H1N1 swine influenza A virus (pH1N1) claimed between 150,000 to 575,000 lives.[1, 2] The pH1N1 virus is an airborne communicable pathogen that infects the upper respiratory tract and eventually works its way down to the lower portion.[1,2,3] H1N1 swine influenzas are generally harbored in domestic pigs and have been shown to transmit zoonotically to humans. However, pH1N1 is directly transmissible between humans presumably due to mutations that changed the "antigenic epitope regions," or structural elements directly implicated in the function of a protein, of the genes responsible for infection of the host, like hemagglutinin (HA) and neuraminidase (NA).[3,4,5] Influenzas of non-human origin with a hemagglutinin (HA) gene were responsible for three other viral pandemics among humans in the twenty-first century like the 1918 Spanish Flu, another H1N1 swine influenza A virus.[3,6]

In March of 2009, a concerning large number of cases were disregarded as late-season flu illness in Mexico City until further investigation suggested it was novel.[2,6] The source of pH1N1 is suspected to be a pork production plant in the state of Veracruz that overused antibiotics, allowing for an H1N1 virus to evolve resistance.[2] The pH1N1 virus was found to have a novel reassortment of segments from four different influenza A viruses associated with these kinds of hosts: North American bird strains, avian-origin human strains, Eurasian Swine, and North American swine.[3] After spreading within Mexico, pH1N1 is documented to have spread three geographically

separated cities in the United States–New York, NY; Milwaukee, WI; and Boston, MA–in the Spring or Summer of 2009.[5,6,] This occured before cases were documented in the rest of the U.S.A. during Fall 2009 at the start of the new school year, putting school children especially at risk.[6]

In this study, the objective was to compare and relate 2009 pH1N1 viral genetic sequences to track the spread of the outbreak amongst humans. The *PB2* gene on segment 1 of the 8-segment viral RNA genome of pH1N1 was chosen for analysis. The PB2 gene product codes for one segment of the influenza virus RNA-dependent RNA polymerase;[7,8] the PB2 protein complexes with the segment 2 gene products to assist in the process of hijacking the nuclear transcription to make H1N1 viral proteins.[7] We hypothesize that Mexico, the currently accepted source of the pH1N1 outbreak, should be generally the most genetically diverse. With each transmission to a new geographical location, there will effectively be a new founder effect for the location's corresponding pH1N1 viral population. Of all in the U.S.A. to get the virus, New York City is thought to be particularly responsible for the spread of pH1N1 to Europe and Asia due to its role in extensive worldwide commerce and the many visitors it gets every year. We thus suspected that New York City would have strains of pH1N1 closely related to those found outside of the North American continent.[9]

METHODS

37 viral sequences of the complete PB2 gene were collected on the NCBI virus database. Efforts were concentrating to ensure a representative sample of sequences was collected as early as possible for each location we were interested in. The earlier the collection date of the sequence, the less likely it was that the given viral strain had undergone random mutations–making it easier to determine phylogenetic relationships. We collected 12 viral H1N1 influenza A sequences from Mexico between February 6, 2009 and the beginning of April 2009. We collected 7 sequences from Boston and 2 sequences from New York. While a thorough search was made to find additional PB2 sequences from New York and Milwaukee, there were only complete "segment 1 sequences" rather than "complete PB2 gene sequences" before the end June 2009. Additional *PB2* gene sequences from China, South Korea, Argentina, Chile, Brazil, India, Italy, England, and France were also collected to get some view of the pH1N1 gene pool in Asia, South America, and Europe respectively. All 37 sequences were formatted as ".FASTA" files, carefully aligned, and imported as one ".meg" file into MEGA 7 to create a "Maximum likelihood" phylogenetic tree using the default parameters (i.e. Tamura-Nei model, Nearest-Neighbor Interchange (NNI)). The aligned sequences were subsequently inspected for polymorphisms, and additional smaller phylogenetic trees were made for clades observable on the original generated phylogenetic tree for all 37 viral gene sequences.

Sequence Alignment and Phylogeny

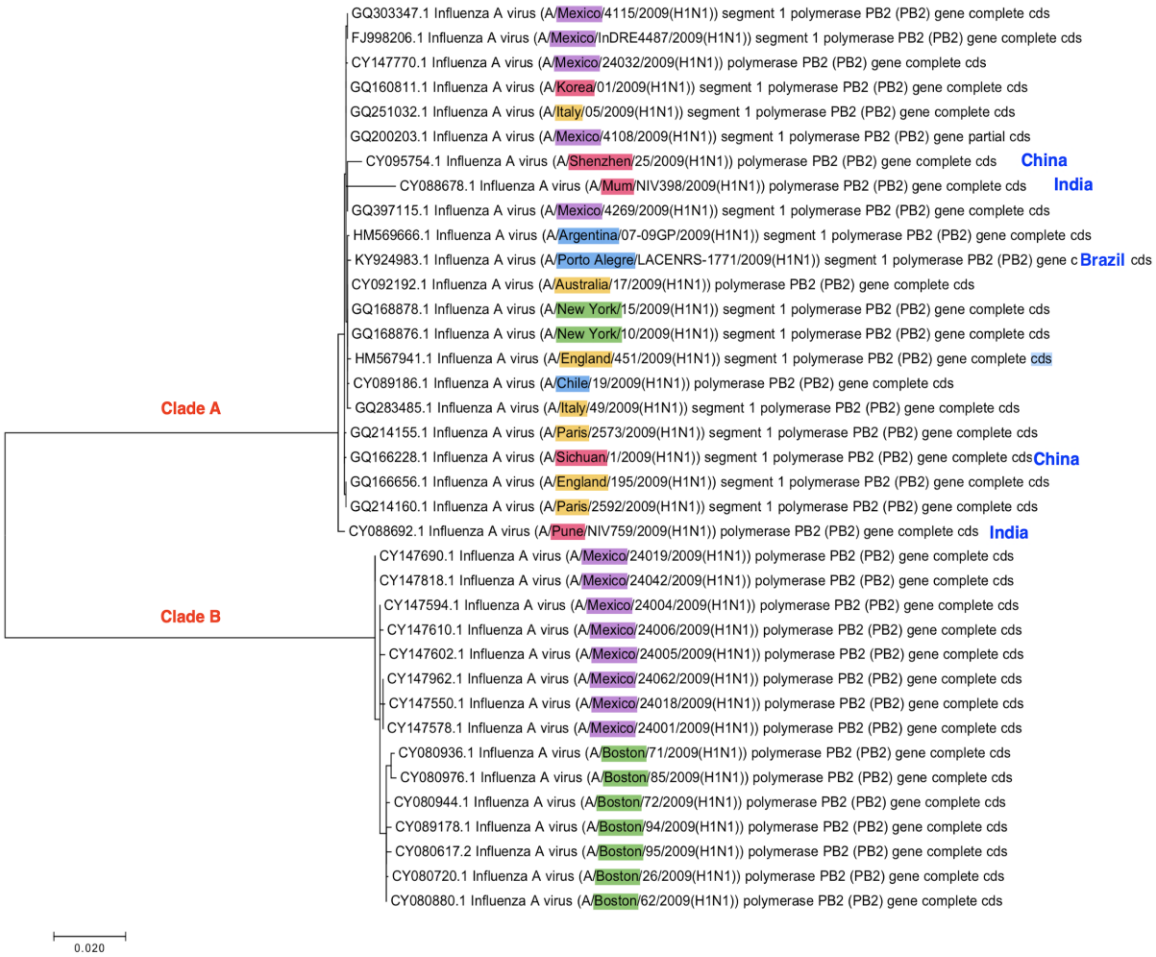

Figure 1: Complete phylogeny generated for all 37 sequences

A complete phylogeny was generated for all 37 aligned *PB2* sequences in MEGA 7. Overall, we observed two distinct large clades with comparatively large branch lengths, which are labeled in Figure 1 as “Clade A” and “Clade B” respectively. Clade A is definitely more diverse than Clade B as it contains every single sequence we obtained from outside North America and three sequences from Mexico. Meanwhile Clade B contains the majority of sequences from Mexico and all of the sequences collected from Boston. If our computational analysis is broadly trustworthy, it is reasonable to presume that there were at least two separate transmission events to hit the United States in Boston and New York respectively. Additionally, Mexico must be quite genetically diverse as it is the only region (even if we consider the regions outside of North America as entire continents) that is represented in both clades. The representation of Mexico in both clades is supporting evidence of our hypothesis that Mexico is the source of pH1N1 due to its genetic diversity.

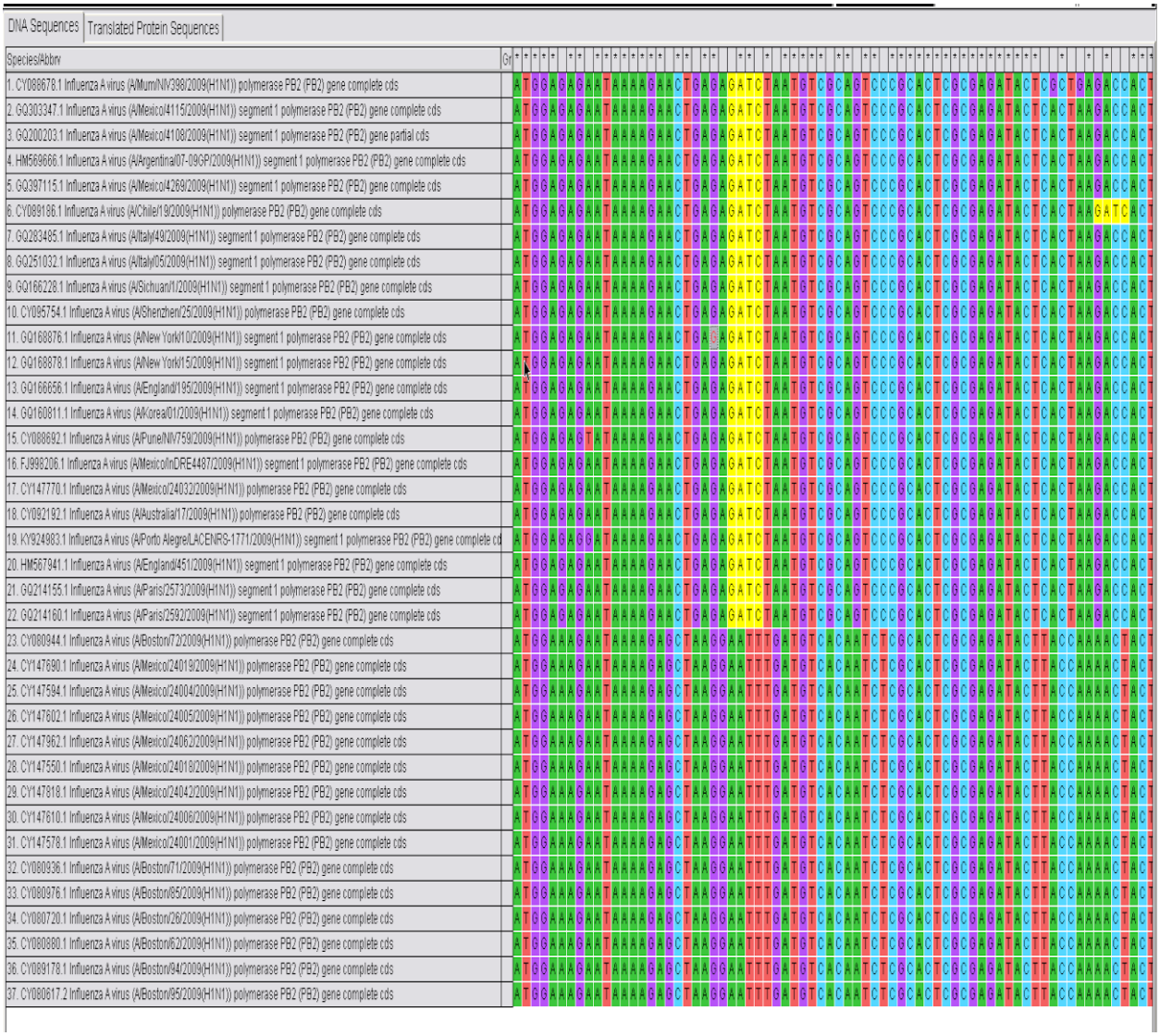

Figure 2: Aligning *PB2* sequences in MEGA 7: Display of the first ~50 aligned bases

With the general phylogeny in Figure 1 alone, there may be concerns regarding the validity of our tree's quality. With how short the branch lengths were within each clade, we made sure to go the extra mile to ensure that our observable evolutionary relationships were valid. Figure 2 contains a screenshot from MEGA 7 of the aligned sequences for all 37 *PB2* sequences. For this alignment to have been facilitated properly, between 1 and 12 bases were removed amongst the 37 genes to line up with the sequence with the latest base in segment 1 of the H1N1 genome. In general, an excessive amount of SNPs between the two clades were observable at the beginning of the aligned sequences. A consistent single nucleotide difference was observed at bases 6, 25, 27, 34, 40, 60, 69, 75, 81, 96, 99, 130, 114, 138, 161, 177, 180, 186, 191,195, 204, and 207 between members of Clade A and Clade B. As highlighted in the figure above, the first 22 sequences (all from Clade A) had a "GATC" motif at bases 25 to 28 different from the "AATT" nucleotide motif found in the bottom 15 sequences (all from Clade B). This consistent

pattern of binary base substitutions–like the difference of a guanine or adenine at base 6 for Clade A or B respectively–strongly supports our assertion that there are two different clades of pH1N1 virus.

After finding convincing evidence that there are two clades, "sub-phylogenies" for each clade of pH1N1 *PB2* sequences were made to better clarify the relationships within each clade. New phylogenetic trees were regenerated for each tree and aligned sequences were reanalyzed in the context of only members of the same given clade. Certain strains exhibited greater numbers of polymorphisms or substitutions than others, many of which were documented in **Appendices A** and **B** for Clades A and B respectively. For Clade A, the first 1240 bases in the ~2340-base sequence were analyzed. There were about 25 documented substitution positions–half of which were solely used by the A/Mum/NIV 398/2009(H1N1) strain in India. While the first phylogeny simply put this strain on a longer strand in the middle of the tree, it was at least indicated that the Mum strain was significantly different from the others. The actual Mum strain had a total of 19 substitutions relative to the rest of clade A along with some strange triplet substitutions at positions 614–616 ACA motif (TAC) and 637–639 CGT motif (GTG). While a frameshift mutation could have been possible around these base positions, everything remained aligned. Because the *PB2* gene is a coding segment for an important protein implicated in the pH1N1 viral replication cycle, it should be expected that the gene is largely conserved. Analogous to how exons of genes are less likely to be mutated in a general population, significant mutation of a pH1N1 viral genome segment would more likely occur in a noncoding region since there would likely be fewer functional implications. With so many substitutions for Mum, it makes sense that its new placement within the Clade A sub-phylogeny is as an outgroup presumably part of a smaller clade exclusive from all of the others. Within Clade A, there was also a double-substituted motif for A/Paris/2592/2009(H1N1) (AA)at 546–547 GC similar to a singly-substituted motif at the same position for another Parisian strain A/Paris/2573/2009(H1N1) (AC). It may be plausible to suggest that the 2592 Parisian strain was formed after the 2573 Parisian strain–possibly even from something closely related to the other strain–as posed by its higher number being indicative of the order in which it was sequenced by a given research group or organization. Meanwhile, Clade B was relatively uneventful likely due to the nature of our sampling method. In general, the sequences in Clade B were selected based on both location and time period, which thus limited the potential for diversification of strains. Not until after 980th nucleotide is there finally some observable markers that distinguishes the Mexican Clade B pH1N1 virus from the Bostonian counterpart. At sites 981 (G–>A), 1032 (T–>G), and 1357 (T–>C), there are nucleotide substitutions consistent across all samples within both groups where Clade B Bostonian pH1N1 has a single given nucleotide while Clade B Mexican pH1N1 has an entirely different base. With this information in mind, it can be inferred that some Clade B Mexican virus likely acquired these three mutations before one member of its clone was transmitted from Mexico to Boston.

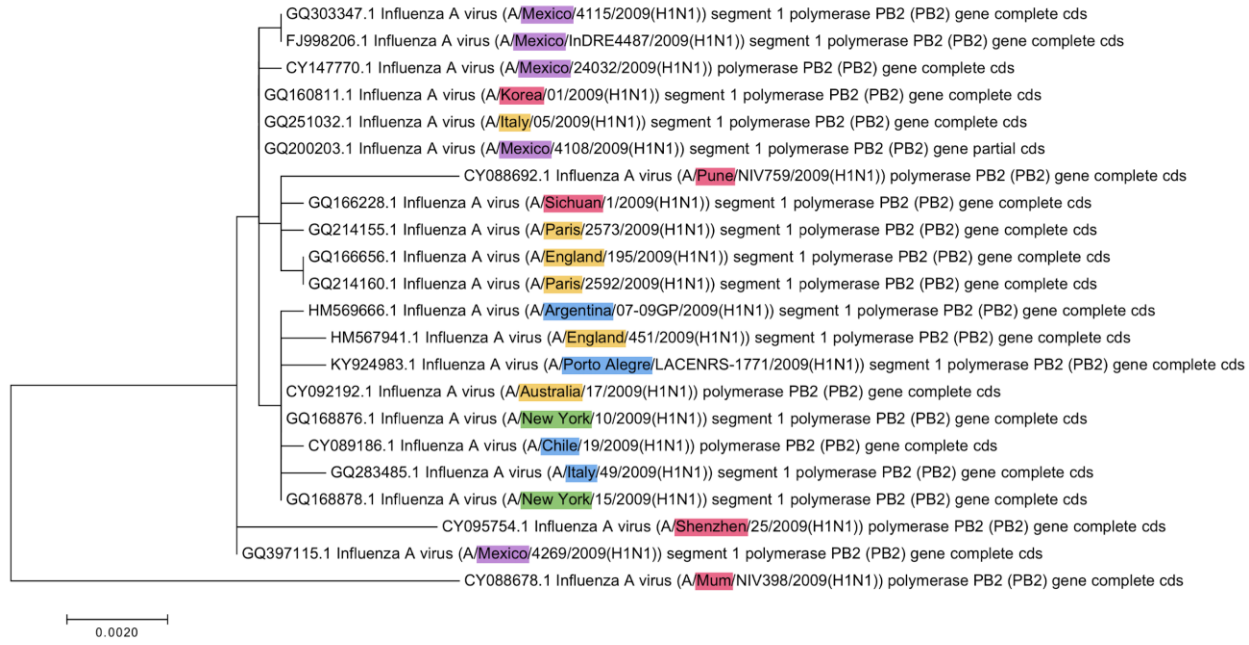

Figure 3: Clade A: New York, Mexico, South America, East Asia, and Europe

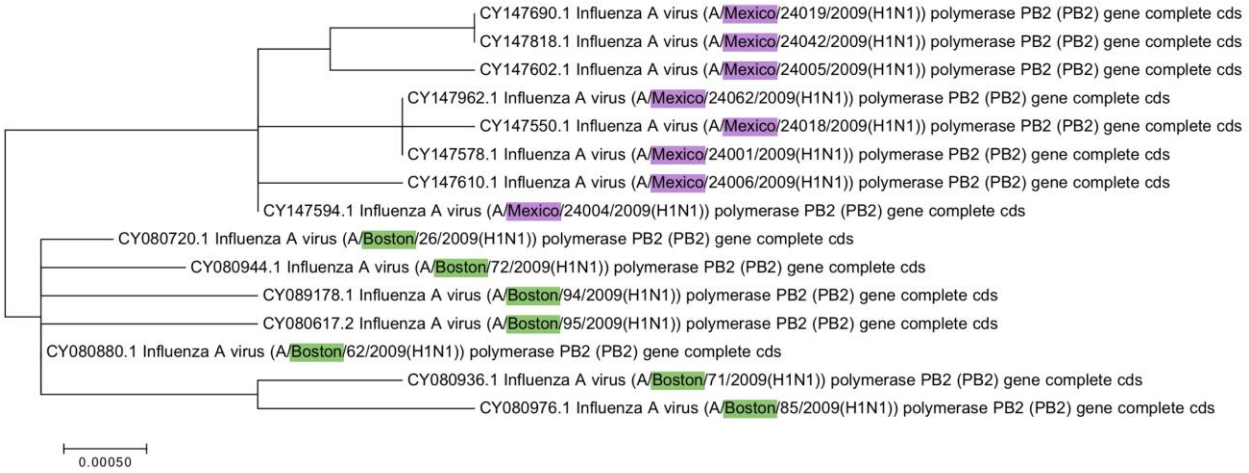

Figure 4: Clade B: Boston and Mexico

Appendix A.

All polymorphisms in the first 1240 bases of the aligned *PB2* sequences of Clade A:

1. Site 9A substitution for Brazil: Porto Alegre (G) and Pune (T)

2. Site 61A substitution for Mum (G)
3. Site 68C substitution for Chile (T)
4. Site 94A substitution for Mum (G)
5. Site 135C substitution for Mum (T)
6. Site 245A substitutions for Mum (G) and lnDRE4487 Mexico (G)
7. Site 348A substitution for Mum (C)
8. Site 355T substitution for Mum (G)
9. Site 367G substitution for Mum (C)
10. Site 480G substitution for Mum (C)
11. Single and double substitutions for site 546–547 GC motif

Site 546–547 Motif	Sequences with motif
GC	Everything else
AC	Paris 2573 (first), Sichuan(China); Pune(India)
AA	Paris 2592 (second)

- Site 580C substitution for Italy 49 (A)
- 2 triply-changed motif of Mum:
    - Mum has 614–616 base TAC motif in place of ACA motif that every other sequence has
    - Mum has a 637–639 baseGTG motif in place of the CGT motif everything else has
- Site 682T substitution for Mum (G)
- Site 687T substitution for Mum (C)
- Site 695T substitution for Mum (C)
- Site 746A substitution for Mum (T)
- Site 749A substitution for Mum (T)
- Double substitution for Mum (CA) at 749–750 TG
- Site 750A substitution for Mum (G)
- Site 776A substitution for Mum (G)
- Site 780C substitution for Mum (T)
- Site 852C substitution for Mexico 24032 (T)
- Site 858G substitution for Shenzhen (A)
- Site 957T substitution for Pune (C)
- Site 963A substitution for Pune (G)
- Site 966T substitution for Pune (C)
- Site 972 substitution for Pune (C)
- Site 1056G substitution for Brazil: Porto Alegre (A)
- Site 1056G substitution for Brazil: Porto Alegre (A)

- Site 1080T substitution for Shenzhen (A)
- Site 1088T substitution for Shenzhen (C)
- Site 1201G substitution for Mum (A)

Appendix B.

All Clade B polymorphisms at first about ~1500 bases

1. First substitution at 192G for Boston 85 (T)
2. Substitution at 204T for Boston 72 (G)
3. Substitution at 258A for Boston 24005 (G)
4. **SAME** substitutions at 303A for Mexico 24019 (G) and Mexico 24042 (G)
5. Substitution at 322A for Boston 95 (G)
6. Substitution at 387C for Boston 95 (T)
7. **SAME** substitution at 417C for Boston 71 (T) and Boston 85 (T)
    - Supporting evidence for their proximity on sub-phologeny tree
8. Substitution at 469A for Boston 26 (G)
9. Substitution at 550A for Boston 85 (G)
10. No mutations from 550 to 745
11. Substitution at 746A for Mexico 24018 (G)
12. Substitution 896G for Mexico 24006 (A)
13. **SAME** substitution at 900A for Mexico 24019 (G), Mexico 24042 (G), Mexico 24005 (G)
14. Substitution at 948G for Mexico 24005 (A)
15. **BIG DIFFERENCE AT SITE 981G – a consistent substitution between Boston and Mexico**
    - Boston has A only
    - Mexico only has G
    - A virus that reached Boston from Mexico of the general Clade B lineage must have acquired a mutation at the 981 position before reaching Boston
16. **A SECOND BIG DIFFERENCE at 1032T**
    - Mexico has T
    - Boston has G
    - No mutations from 1031 to 1356
17. **A THIRD MARKER AT 1357T**
    - Mexico has T
    - Boston has C

REFERENCES

1. Al-Muharrmi Z. Understanding the Influenza A H1N1 2009 Pandemic. Sultan Qaboos Univ Med J. 2010;(10):187-195.
2. Dotis J, Roilides E. H1N1 Influenza A Infection. Hippokratia. 2009;13(3):135-138.
3. Talha JN, Jamil RT, Siddiqui AH. H1N1 Influenza (Swine Flu). StatsPearls [Internet] [Online]; StatPearls Publishing LLC. Treasure Island, FL., 2019.
4. Huang P, Yu S, Wu C, Liang L. Highly Conserved Antigenic Epitope Regions of Hemagglutinin and Neuraminidase Genes between 2009 H1N1 and Seasonal HN1 Influenza: Vaccine Considerations. J Trans Med. 2013;11:47.
5. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez, Lara-Puente JH, et al. Origins of the 2009 H1N1 Influenza Pandemic in Swine in Mexico. Elife. 2016;5(16777):1-21.
6. Gog JR, Ballesteros S, Viboud C, Simonsen L, Bjornstad ON, Shaman J, et al. Spatial transmission of 2009 pandemic influenza in the US. PLoS Comp Bio. 2014;10:6.
7. Graef KM, Vreede FT, Lau Y, Mccall AW, Carr SM, Subbarao K, et al. The PB2 subunit of the Influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. J Virol. 2010;84(17):8433-8445.
8. Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JR, Nagata K, et al. Structural Insight into the Essential PB1-PB2 Subunit Contact of the Influenza Virus RNA Polymerase. EMBO J. 2009; 28(12):1803-1811.
9. Heneghan C, Jefferson T. COVID-19 Deaths Compared with “Swine Flu.” The Centre for Evidence-Based Medicine, University of Oxford. 2020.