Published in final edited form as:

Infect Genet Evol. 2014 December ; 28: 351–357. doi:10.1016/j.meegid.2014.07.015.

Toward a Method for Tracking Virus Evolutionary Trajectory Applied to the Pandemic H1N1 2009 Influenza Virus

R. Burke Squires^{1,*}, Brett E. Pickett², Sajal Das³, and Richard H. Scheuermann^{2,4}

¹Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

²Informatics, J. Craig Venter Institute (JCVI), La Jolla, CA, USA

³Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, USA

⁴Department of Pathology, University of California San Diego, San Diego, CA, USA

Abstract

In 2009 a novel pandemic H1N1 influenza virus (H1N1pdm09) emerged as the first official influenza pandemic of the 21st century. Early genomic sequence analysis pointed to the swine origin of the virus. Here we report a novel computational approach to determine the evolutionary trajectory of viral sequences that uses data-driven estimations of nucleotide substitution rates to track the gradual accumulation of observed sequence alterations over time. Phylogenetic analysis and multiple sequence alignments show that sequences belonging to the resulting evolutionary trajectory of the H1N1pdm09 lineage exhibit a gradual accumulation of sequence variations and tight temporal correlations in the topological structure of the phylogenetic trees. These results suggest that our evolutionary trajectory analysis (ETA) can more effectively pinpoint the evolutionary history of viruses, including the host and geographical location traversed by each segment, when compared against either BLAST or traditional phylogenetic analysis alone.

Keywords

influenza; virus; evolution; metadata; swine flu; pandemic

1. Introduction

In late March 2009, the first influenza pandemic of the 21st century arrived in the form of the 2009 influenza A H1N1 pandemic (H1N1pdm09) virus. The virus originated in Mexico though the earliest samples were isolated from patients in California and Texas (Garten et

© 2014 Elsevier B.V. All rights reserved.

*Corresponding Author: R. Burke Squires, richard.squires@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

al., 2009; Smith et al., 2009). According to the World Health Organization, as of August 31, 2010 there were ~19,000 deaths due to the pandemic worldwide; though recent analysis by Simonsen et al. (Simonsen et al., 2013) casts doubt on that number, instead estimating 10 times the number of deaths. Although it is difficult to determine the precise number of deaths from any pandemic, including the one declared for the 2009 influenza virus; it is nevertheless understood that influenza virus pandemics can be very deadly. Such was the case when a separate H1N1 influenza A virus caused the pandemic of 1918, which was responsible for an estimated 50 to 100 million human deaths. Additional pandemics in the twentieth century caused an estimated 2 million deaths in 1957 (H2N2) and 1 million deaths in 1968 (H3N2) (World Health Organization, 2005).

The influenza virus genome consists of eight negative-sense single-stranded RNA segments. Each of these segments encodes at least one protein — yielding at least 11 proteins in total. Elucidating the evolutionary trajectories of individual segments enables us to: determine how frequently segments jump between different host species, identify key reassortment events, and track down the geographic origin of a pandemic virus. Ultimately, understanding how influenza viruses evolve and reassort is crucial to designing novel vaccines and antiviral drugs as well as predicting the efficacy of existing treatments.

2. Material and methods

2.1. Identification of Similar Sequences

The A/California/04/2009 (H1N1) pandemic strain was chosen as the reference strain for our analysis because it was the first pandemic isolate for which a complete genome became publically available through GenBank (Garten et al., 2009; Smith et al., 2009). All genome sequences from strains circulating before the pandemic began were downloaded from the Influenza Research Database (IRD, www.fludb.org) (Squires et al., 2012). Sequence similarity searches for all A/California/04/2009 segments were performed using local BLASTn (Camacho et al., 2009) against all previously-isolated influenza sequences (data downloaded from GenBank on Jan. 2, 2014). The highest-scoring 1,000 BLAST hits for each segment were used for further analysis.

2.2. Multiple Sequence Alignment and Phylogenetic Tree Construction

Selected groups of sequences were aligned and ordered by year of isolation using the “multiple sequence alignment based on fast Fourier transform” (MAFFT (Katoh and Standley, 2013) software. Phylogenetic trees were constructed using the PhyML algorithm (Guindon et al., 2010) with HKY + 5 class model to build trees for each group of sequences. Each tree was manually rooted at the strain with the earliest isolation date within each group. Subtrees were ordered using FigTree viewer (<http://tree.bio.ed.ac.uk/software/figtree>) to mimic the multiple sequence alignment organization.

2.3 Calculation of substitution rates

Best-fit trend lines were determined by first computing the slopes of each hit from the origin (0, 0) in the isolation year difference versus nucleotide difference plots and binning the sequences according to the nearest slope. Using a data-driven peak finding approach we

counted the frequencies of slopes through each hit in our graph and the origin (0, 0). We then analyzed the frequencies with a peak detection algorithm and selected the first valley slope to differentiate between Groups 1 and 2. A best-fit linear trend line passing through the origin was then calculated for each group separately along with a 95% confidence level.

3. Results

3.1 Temporal Relationships of Sequence Similarity

In order to determine how measures of sequence similarity correlate with the date of strain isolation, the normalized nucleotide difference between each A/California/04/2009 segment and the top 1000 BLASTn hits was plotted against the time since the strain was isolated (e.g. 21 years for a strain isolated in 1988 in comparison with 2009). These plots revealed several interesting consistent patterns of sequence similarity-temporal relationships (Figure 1). Most dramatically, plots for segments 4, 5, 7 and 8 show a triangular pattern suggesting that three different groups of sequences existed corresponding to the three sides of the triangle (Figure 2). Note that the triangular pattern remains when the isolation year differences are plotted against pairwise distances, based upon a codon model of evolution (Fig. S1) as described by Criscuolo and colleagues (Criscuolo and Michel, 2009) (see supplemental materials). Group 1 showed a gradual linear relationship between sequence similarity and time since isolation; Group 2 showed much larger nucleotide differences in shorter periods of time; Group 3 showed similar levels of nucleotide differences regardless of the time of isolation. We focused our analysis on Groups 1 and 2 since they showed the closest sequence similarity to the A/California/04/2009 segments of the three groups.

A nucleotide substitution rate could be estimated for Group 1 and Group 2 sequences from a best-fit line going through the origin of the plot. For most segments, the nucleotide substitution rate estimated from Group 1 sequences was closely similar to the segment-specific mutation rates reported in the literature (Table 1). In contrast, the nucleotide substitution rate estimated from Group 2 sequences was 2 – 11 times faster than would be expected for normal influenza virus evolution.

Based on these findings, we hypothesized that the Group 1 sequences represented the true evolutionary trajectory (ET) of the H1N1pdm09 lineage, whereas the Group 2 sequences represented isolates that were similar but distantly related (SDR).

3.2 Persistence of Substitutions Over Time in the ET Group

To further investigate the temporal relationships of sequence similarity in these groups, multiple sequence alignments for Group 1 (ET) and Group 2 (SDR) sequences were generated and compared. When the sequences in the multiple sequence alignment were ordered by date of isolation, distinct patterns were observed in the two groups. The alignment of Group 1 (ET) sequences showed a pattern in which nucleotide changes would occur and then persist in subsequent strains generating “vertical tracking” patterns in the alignments (Figure 3A). In contrast, Group 2 (SDR) sequences showed a “starry sky” pattern in the temporally ordered alignments (Figure 3B). Indeed, the persistence of nucleotide changes is consistent with Group 1 (ET) corresponding to the true evolutionary trajectory.

To quantify the nucleotide change persistence, the number of nucleotide changes that occurred as one proceeded from the top to the bottom of each column in the alignment (i.e. across time) was calculated. Smaller numbers of changes were observed for Group 1 (ET) versus Group 2 (SDR) viruses, especially for segments 4, 5 and 7. This observation is consistent with the higher degree of persistence of nucleotide substitutions in the Group 1 alignment (Table 2).

3.3 Gradual Phylogenetic Transitions in the ET Group

In order to further investigate the evolutionary relationships between these two groups, phylogenetic trees were generated for each segment belonging to each group and rooted using the oldest sequence in each group. In the trees generated from sequences assigned to Group 1 (ET), a gradual “staircase” pattern of progression was observed along the tree, with relatively short branch lengths (Figure 4). In addition, the isolation date ranges paralleled the phylogenetic progression. In contrast, the Group 2 (SDR) trees showed relatively flat structures and appeared to have longer branch lengths. Branch length quantification, as shown in Table 3, demonstrated that virtually all Group 1 (ET) segments showed significantly shorter branch lengths on average than Group 2 (SDR) segments.

4 Discussion

Traditional phylogenetic tree generation algorithms focus on reconstructing the underlying evolutionary relationships between sequences without explicitly incorporating the temporal information that is associated with the sequence data. Here we report an approach that segregates and tracks groups of sequences by combining sequence similarity data with temporal information and is independent from traditional phylogenetic approaches. We used this technique to analyze the H1N1pdm09 lineage and identified a subset of similar sequences that appear to represent the true evolutionary trajectory (ET) of each of the pandemic segments. Phylogenetic trees constructed with these ET subsets validated this method by showing gradual and temporal transitions consistent with genetic drift of the viral genome over time. In contrast, construction of phylogenetic trees of the similar but distantly related (SDR) subset also validated this method by showing shorter branch lengths and many fewer monophyletic groups, which indicates the occurrence of random nucleotide substitutions over time. Phylogenetic trees that analyze 2009 H1N1 pandemic sequences without regard of their evolutionary relationships can give rise to incorrect interpretations about the origins of an outbreak strain. Indeed, in the initial H1N1pdm09 lineage publication (Garten et al., 2009), sequences from both our Group 1 (ET) and Group 2 (SDR) groups were included in their phylogenetic analysis, with the Group 1 sequences coming from North America (e.g. A/Swine/Minnesota/593/99_H3N2) and the Group 2 sequences coming from Asia (e.g. A/Swine/Guangxi/13/2006_H1N2). Consequently, the likely geographic source of the pandemic outbreak was more uncertain than it would have been if only sequences belonging to the true evolutionary trajectory had been included in the analysis.

In the current analysis, distinct groups of sequences were identified through an approach designed to capture sequence similarity-temporal relationships between related sequences identified by BLAST. Based on the results obtained for segment 5, the nucleotide difference distributions for both Groups 1 and 2 ranged from approximately 25 to 175 nucleotides.

Thus, the groups do not differ in their ranges of nucleotide differences; instead it is the distribution of *isolation years* that primarily differentiates the groups. Group 1 (ET) has isolation years from as recent as 2002 all the way back to the 1918 pandemic. In stark contrast, the date distribution of Group 2 (SDR) ranges from 1993 to 2008. Group 2 sequences overall have a similar number of substitutions in comparison with the 2009 pandemic strain as the Group 1 sequences. However, if the Group 2 sequences were truly direct ancestors of the 2009 outbreak sequence, they would have had to mutate at a much faster-than-normal rate between when they first emerged in the mid-1990's and the start of the 2009 H1N1 pandemic in order to support these observations. In other words, bioinformatics techniques that look exclusively at sequence differences would not distinguish between sequences belonging to either Group 1 or Group 2.

In an effort to further compare Groups 1 and 2, we utilized a linear best-fit algorithm to estimate the overall mutation rate of each group. While a power best-fit line was also examined, in the end we selected a linear best-fit line to enable inclusion of the origin. This approach enables the most direct comparison with experimentally determined mutation rates. The use of a best-fit algorithm will smooth out over all evolutionary processes in a weighted average of mutation rates somewhere in between the times of rapid evolution interspersed by periods of relative genetic stability that might occur during punctuated equilibrium. The rapid evolution periods will be smoothed to a greater extent than the periods of relative genetic stability. That being said we think the evidence in the influenza field is that during genetic drift, periods of rapid evolution and genomic stability are relatively short, frequently interspersed, and relatively continuous.

The addition of temporal data to help elucidate evolutionary trajectory is by no means the only possible metadata enhancement. The temporal component of samples is but one of many pieces of metadata; additional metadata such as geographic location, host (and vector), and passage history can all be used to improve the estimation of the evolutionary trajectory of a sequence. For example, geographic relationships that depend on known avian migratory patterns could be used to exclude trajectory sequences that lie outside of these defined migratory patterns. The extent to which each of these metadata components influences the evolutionary trajectory will require additional study.

The findings presented in this work can be used to enhance the accuracy of future sequence analyses by including only strains that appear to lie along the true evolutionary trajectory of the influenza virus subtype. The results from such analyses can then be applied to the development of novel vaccines as well as effective prophylactic and therapeutic antiviral drugs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Megan Coakley for critical reading of the manuscript. We also acknowledge the generous support received from the NIAID – N01AI40041.

References

- Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, Mallal S, Mullins JI, Nickle DC, Herbeck J, Rousseau C, Learn GH, Miura T, Brander C, Walker B, Korber B. Founder Effects in the Assessment of HIV Polymorphisms and HLA Allele Associations. *Science*. 2007; 315:1583–1586. [PubMed: 17363674]
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. [PubMed: 20003500]
- Criscuolo A, Michel CJ. Phylogenetic Inference with Weighted Codon Evolutionary Distances. *Journal of Molecular Evolution*. 2009; 68:377–392. [PubMed: 19308635]
- Fitch WM, Bush RM, Bender CA, Cox NJ. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proceedings of the National Academy of Sciences*. 1997; 94:7712–7718.
- Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier RAM, Pappas C, Alpuche-Aranda CM, Lopez-Gatell H, Olivera H, Lopez I, Myers CA, Faix D, Blair PJ, Yu C, Keene KM, Dotson PD, Boxrud D, Sambol AR, Abid SH, St George K, Bannerman T, Moore AL, Stringer DJ, Blevins P, Demmler-Harrison GJ, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA, Beatrice ST, Donis R, Katz J, Finelli L, Bridges CB, Shaw M, Jernigan DB, Uyeki TM, Smith DJ, Klimov AI, Cox NJ. Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans. *Science*. 2009; 325:197–201. [PubMed: 19465683]
- Gorman OT, Donis RO, Kawaoka Y, Webster RG. Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. *J Virol*. 1990; 64:4893–4902. [PubMed: 2398532]
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*. 2010; 59:307–321. [PubMed: 20525638]
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. 2013; 30:772–780. [PubMed: 23329690]
- Maurer-Stroh S, Lee RTC, Eisenhaber F, Cui L, Phuap SP, Lin RT. A new common mutation in the hemagglutinin of the 2009 (H1N1) influenza A virus. *PLoS Curr*. 2010; 2:RRN1162. [PubMed: 20535229]
- Nelson M, Spiro D, Wentworth D, Beck E, Fan J, Ghedin E, Halpin R, Bera J, Hine E, Proudfoot K, Stockwell T, Lin X, Griesemer S, Kumar S, Bose M, Viboud C, Holmes E, Henrickson K. The early diversification of influenza A/H1N1pdm. *PLoS Curr*. 2009; 1:RRN1126. [PubMed: 20029664]
- Nelson MI, Viboud C, Simonsen L, Bennett RT, Griesemer SB, St George K, Taylor J, Spiro DJ, Sengamalai NA, Ghedin E, Taubenberger JK, Holmes EC. Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus Since 1918. *PLoS Pathog*. 2008; 4:e1000012. [PubMed: 18463694]
- Nobusawa E, Sato K. Comparison of the Mutation Rates of Human Influenza A and B Viruses. *J Virol*. 2006; 80:3675–3678. [PubMed: 16537638]
- Parvin JD, Moscona A, Pan WT, Leider JM, Palese P. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J Virol*. 1986; 59:377–383. [PubMed: 3016304]
- Romero-Tejeda A, Capua I. Virus-specific factors associated with zoonotic and pandemic potential. *Influenza Other Respi Viruses*. 2013; 7(Suppl 2):4–14.
- Saitou N, Nei M. Polymorphism and evolution of influenza A virus genes. *Molecular Biology and Evolution*. 1986; 3:57–74. [PubMed: 3444396]
- Shu LL, Bean WJ, Webster RG. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J Virol*. 1993; 67:2723–2729. [PubMed: 8474171]

- Simonsen L, Spreeuwenberg P, Lustig R, Taylor RJ, Fleming DM, Kroneman M, Van Kerkhove MD, Mounts AW, Paget WJ, Teams TGC. Global Mortality Estimates for the 2009 Influenza Pandemic from the GLaMOR Project: A Modeling Study. *Plos Med.* 2013; 10:e1001558. [PubMed: 24302890]
- Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, Peiris JSM, Guan Y, Rambaut A. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature.* 2009; 459:1122–1125. [PubMed: 19516283]
- Squires RB, Noronha J, Hunt V, García-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, Ramsey A, Zhou L, Zaremba S, Kumar S, Deitrich J, Klem E, Scheuermann RH. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respi Viruses.* 2012; 6:404–416.
- Subbarao K, Murphy BR, Fauci AS. Development of Effective Vaccines against Pandemic Influenza. *Immunity.* 2006; 24:5–9. [PubMed: 16413916]
- Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, O'Connor DH, Hughes AL, Neumann G, Kawaoka Y, Friedrich TC. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nature Communications.* 2013; 4
- World Health Organization. Ten things you need to know about pandemic influenza (update of 14 October 2005). *Wkly Epidemiol Rec.* 2005; 80:428–431. [PubMed: 16372665]
- Xu X, Cox NJ, Bender CA, Regnery HL, Shaw MW. Genetic variation in neuraminidase genes of influenza A (H3N2) viruses. *Virology.* 1996; 224:175–183. [PubMed: 8862412]

Highlights

- Quantifying sequence similarity-temporal relationships revealed added structure in most similar to H1N1pdm09 by Blast.
- One of the subgroups exhibiting characteristics of the true evolutionary trajectory of the H1N1pdm09 lineage.
- Suggests sequence similarity in absence of temporal component not enough to correctly determine evolutionary relationships.
- Utilizing isolation year metadata enables a more accurate tracking of the true evolutionary trajectory.

Figure 1. Sequence similarity-temporal patterns in top BLAST hits

Plots of the isolation year differences versus normalized nucleotide difference between the reference A/California/04/2009 segments 1 – 8 and the top 1,000 BLAST hits for each segment are shown. Isolation year differences were computed by subtracting the isolation year of each hit from the isolation year of the reference strain (2009). Nucleotide differences were computed from the BLAST results and normalized to the full-length sequence.

Figure 2. Groups of sequences revealed in the sequence similarity-temporal plots

A triangular pattern is overlaid on the previous plot of segment 5 with red lines highlighting Groups “1”, “2”, and “3”.

Figure 3. Multiple sequence alignments for each group

Alignments were colored to highlight sequence variations when compared against the consensus sequence with sequence difference highlighted in white. Representative strains were chosen to best distribute year, host and geographical distributions. (A) Alignment of sequences from strains representing the Group 1 (ET) for segment 5, sorted by isolation year, which shows the persistence of nucleotide changes over time. (B) Alignment of Group 2 (SDR) sequences, sorted by isolation year, which shows a lack of persistence of nucleotide changes (lighter background) but rather a “starry sky” pattern.

Figure 4. Distinct phylogenetic tree shapes inferred for Group 1 (ET) versus Group 2 (SDR) sequences

(Left) Phylogenetic tree inferred for Group 1 (ET) nucleotide sequences of the segment 5 (NP) was rooted at A/Brevig Mission/1918 in the lower left, and the subtrees ordered.

(Right) Phylogenetic tree inferred for Group 2 (SDR) was also rooted and ordered but the earliest isolation data of 2001. The red lines show the approximate isolation year ranges of the sequences, which clearly show a much shorter isolation year range for Group 2 (SDR). (Full trees can be seen in supplementary material.)

Table 1

Comparison of substitution rates inferred from best-fit trend lines for each sequence group in comparison with estimated segment-specific mutation rates reported in the literature.

Segment	Group 1 - ET (95% Conf.) ^a	Group 2 - SDR (95% Conf.) ^a	Mutation Rate ^b	Reference(s)
1 (PB2)	7.74 (7.48-7.99)	21.65 (20.47-22.82)	4.30	(Gorman et al., 1990)
2 (PB1)	7.20 (7.10-7.31)	15.05 (14.13-15.96)		
3 (PA)	5.36 (5.20-5.53)	21.65 (20.55-22.72)		
4 (HA)	5.82 (5.73-5.91)	12.99 (12.32-13.67)	5.70	(Fitch et al., 1997)
5 (NP)	3.16 (3.06-3.26)	15.26 (14.48-16.04)	3.60	(Shu et al., 1993)
6 (NA)	6.15 (5.69-6.55)	24.81 (23.85-25.78)	3.20	(Xu et al., 1996)
7 (M)	2.08 (1.98-2.17)	10.05 (9.57-10.53)	1.50	(Saitou and Nei, 1986)
8 (NS)	2.50 (2.40-2.60)	18.00 (17.07-18.86)	1.60	(Nobusawa and Sato, 2006; Parvin et al., 1986)

^aSlopes of the best-fit trend lines were calculated for each group separately as described in the Methods section along with a 95% confidence level (shown in parenthesis).

^bEstimated segment-specific mutation rate reported in the indicated reference.

Table 2

Average number of nucleic acid residue changes per strain in alignments of Groups 1 and 2.

Segment	Protein(s)	Group 1 -ET ^a	Group 2 - SDR ^a
1	PB2	61.0	70.8
2	PB1, PB1-F2	21.9	53.8
3	PA, PA-X	52.0	90.1
4	HA	30.2	90.8
5	NP	23.6	61.5
6	NA	75.6	110.5
7	M1, M2	32.8	51.7
8	NS1, NS2	21.5	33.0

^a Sequences were aligned and ordered from oldest isolation year to most recent isolation year, and the number of residue changes at a specific residue or column among the sequences was counted as one progressed down the residue column in the alignment. For quantitative purposes, counts were normalized to the full-length sequences.

Table 3

Analysis of branch length of the phylogenetic trees for the two groups.

Group 1 - ET ^a					Group 2 - SDR				
Segment	Min. Iso. Yr.	Max. Iso. Yr.	Average	Avg. / Year	Min. Iso. Yr.	Max. Iso. Yr.	Average	Avg. / Year	P-value
1	1975	2002	3.33E-03	1.23E-04	1995	2008	2.48E-03	1.91E-04	7.77E-32
2	1988	2000	6.57E-04	5.48E-05	1997	2008	2.29E-03	2.08E-04	3.32E-36
3	1972	2001	3.15E-03	1.09E-04	1996	2008	2.10E-03	1.75E-04	1.99E-26
4	1935	2000	2.29E-03	3.53E-05	1981	2008	3.84E-03	1.42E-04	9.82E-04
5	1918	2002	2.80E-05	3.34E-07	1993	2008	3.79E-03	2.53E-04	2.65E-36
6	1959	2003	5.49E-03	1.25E-04	1998	2008	4.17E-03	4.17E-04	1.81E-02
7	1902	2003	3.39E-03	3.35E-05	1997	2008	3.28E-03	2.98E-04	6.47E-01
8	1930	2006	3.17E-03	4.17E-05	2000	2008	4.64E-03	5.80E-04	4.02E-05

^a Branch_lengths were extracted from the phylogenetic Newick files, divided by the isolation year range to compute an average branch length per year and a Student t-test was performed on the branch lengths from groups 1 and 2 with the averages and p-values shown