

Article

Flow of Information during an Evolutionary Process: The Case of Influenza A Viruses

Víctor Serrano-Solís and Marco V. José *

Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Apdo. Postal 70228, México D.F. 04510, Mexico;

E-Mail: vserrano@biomedicas.unam.mx

* Author to whom correspondence should be addressed; E-Mail: marcojose@biomedicas.unam.mx; Tel.: +52-55-5622-3894; Fax: +52-55-5622-3894.

Received: 14 March 2013; in revised form: 18 July 2013 / Accepted: 23 July 2013 /

Published: 29 July 2013

Abstract: The hypothesis that Mutual Information (MI) dendrograms of influenza A viruses reflect informational groups generated during viral evolutionary processes is put forward. Phylogenetic reconstructions are used for guidance and validation of MI dendrograms. It is found that MI profiles display an oscillatory behavior for each of the eight RNA segments of influenza A. It is shown that dendrograms of MI values of geographically and historically different segments coming from strains of RNA virus influenza A turned out to be unexpectedly similar to the clusters, but not with the topology of the phylogenetic trees. No matter how diverse the RNA sequences are, MI dendrograms crisply discern actual viral subtypes together with gain and/or losses of information that occur during viral evolution. The amount of information during a century of evolution of RNA segments of influenza A is measured in terms of bits of information for both human and avian strains. Overall the amount of information of segments of pandemic strains oscillates during viral evolution. To our knowledge this is the first description of clades of information of the viral subtypes and the estimation of the flow content of information, measured in bits, during an evolutionary process of a virus.

Keywords: mutual information function; mutual information dendrograms; viral evolution; flow of information; Influenza A; phylogenetic reconstruction

1. Introduction

Viruses have been considered as important players in the history of life on this planet [1]. Living beings cannot be conceived without viruses. They impinge on a wide variety of biological processes ranging from the abiotic and fundamental global carbon cycles [2] to such an exquisite process as placentation [3]. As a rule, they act as information transmitting vectors in processes such as horizontal gene transfer, or as infectious agents. Viral evolution has been studied thoroughly in many research fields [4]. Viruses are ubiquitous in the biosphere and they show an astonishing genetic diversity so that viral genomes do not resemble to their hosts' genomes [5,6], and consequently they have not been considered, so far, as part of any tree of life. RNA viruses evolve extremely rapidly, often with mutation rates one million times greater than those of vertebrate species [7]. This rate of mutation allows viral populations to rapidly cope with selective pressures. Viral pathogens, such as influenza virus, HIV, hepatitis C virus, and dengue virus, place a substantial burden on global human health. Despite the fact RNA viruses display high mutation and recombination rates, they remain as discrete recognizable evolutionary units. In fact, the error rates of RNA viruses usually approach the Eigen's error threshold [8,9]. Current methodologies for determining the type of evolutionary paths of a taxon consist in the comparison of sequences in order to infer kinship relationships as displayed by a phylogenetic reconstruction [10]. Those *in silico* methods have been shown to be useful both for DNA and RNA viruses [11]. Viruses are suitable models for evolutionary studies, e.g., the HIV-1 virus [12], influenza A virus [13] and for studying even the origin of genetic information [8]. Influenza A virus is the result of several evolutionary processes such as genetic drift, random mutations, purifying and positive selection, plus genetic recombination and the generation of genomic reassortants [14–16]. There are two major evolutionary approaches for describing viral evolution: the quasispecies theory [17–19] and standard population genetics theory [20,21]. However, none of these theories contemplate the flow of information through the viral evolution.

The extensive flu databases offer a great opportunity to uncover and to examine relevant mutational information associated to influenza A pandemics for almost a century of registries. Based on the antigenic specificities of 16 hemagglutinin (HA) (H1-H16) and nine neuraminidase (NA) (N1-N9) subtypes, there are 144 possible subtypes of influenza A viruses [22]. Unlike most pathogens, where persistent immunity arises after recovery of illness, influenza A virus presents a moving antigenic target, evading any specific immunity triggered by previous infections. The virological basis for annual recurrent epidemics is a continual process of small changes in influenza surface antigens to escape host immunity. This process, called antigenic drift, is the result of the selective fixation of mutations in the gene encoding the HA protein, the major target for the host immune response [23]. HA variants that best escape the host immune response are thought to have a significant reproductive advantage [23]. Influenza A viruses host range comprise avian as well as mammal species, and their genomic segments are subject to occasional reassortment (shift) giving rise to new viral strains consisting of novel HA and/or NA genes [24]. Although less common than antigenic drift, antigenic shift is considered another major force in the evolution of influenza viruses [23–25]. By definition antigenic shift occurs when the virus acquires an HA and/or NA of a different influenza subtype via reassortment of one or more gene segments and is thought to be the basis for the more devastating influenza pandemics that occurred several times in the last century [26]. Inter-pandemic evolution of influenza A virus involves a

complex interplay between neutral evolution during periods of antigenic stasis, positive selection during relatively short intervals of rapid change in fitness, and multiple effects of reassortment [15]. A living entity can be described as a complex adaptive system which differs from any, however complex, chemical structure by its capability of functional self-organization based on the processing of information [8]. Evolutionary processes not only entail loss and gain of relevant information for survival, but its regulation is certainly critical. The mutual information (MI) measures the differences between the average uncertainty in the input of an information channel before and after the output are received [27].

In this work, we propose an approach that considers the MI from Shannon's theory of information perspective [27] in order to measure the amount of flow of information throughout ~100 years of evolution of influenza A virus. The use of the MI has already been applied to study the genetic drift of influenza A/H3N2 virus [28]. The MI concept is useful when analyzing symbolic RNA or DNA sequences. The main goal of this work was to determine if the MI could capture relevant viral evolutionary information associated to antigenic shifts. To this end, we hypothesized that similarity hierarchizations (dendrograms) of MI as derived from actual RNA sequences of influenza A subtypes could display biologically informational coherent groups. Therefore, these dendrograms should also be, in principle, in agreement with the hierarchical clusterization and topology of phylogenetic trees as obtained from *in silico* methodologies of molecular evolution bioinformatics. The article is organized as follows. First, we calculated the MI profiles for each of the eight influenza A genomic segments from 39 antigenic shift subtypes (20 human and 19 avian) that have been clearly identified during the period of 1918 to 2012 for human subtypes and from 1902 to 2011 for avian subtypes. Secondly, dendrograms were constructed from the MI values for each of the eight influenza A segments. Next, phylogenetic reconstructions were obtained and, remarkably, they resemble the hierarchical clusters of the MI dendrograms. In particular, we illustrate results for the MI dendrograms of segments 1, 4, 6, and 8, here denoted by S1, S4, S6, and S8, which encode a highly conserved RNA polymerase (PB2), HA, NA, and a bi-cistronic sequence coding for both the only Non-Structural protein of influenza A (NS1) and the Nuclear Export Protein (NEP), respectively. Finally, a discussion of the present results in terms of viral evolutionary theory and practical implications is presented.

2. Viral Sequences

A set of 312 sequences comprising 39 antigenic shift influenza A whole genomes, (20 human and 19 avian), were selected from the Influenza Virus Resource [29]. Accession numbers for each sequence as well as relevant information are provided in Table 1. The strains were selected according to its relevance in antigenic shift history [30] associated to different pandemics. An important criterion for selection was that all of them were complete sequences. We remark a peculiarity in the analysis of the alignment of S4: for the subtype A/Brevig_Mission/1918 genome [31], the HA sequence (S4) at the 3' side, lacks a little more than $\frac{1}{4}$ of the average length. However, this is the only record of a subtype of 1918, which we could not afford to rule it out from our genomic study. All sequences were ordered by year of report, geographical location, and host species (Table 1). Note that we selected 1, 2, or more human shift subtypes per decade.

Table 1. List of influenza A strains associated to different pandemics.

Acc. Numb.	Year	City	Host	Strain
GU186783	1902	Brescia	Chicken	H7N7
CY014998	1927	Dobson	Fowl	H7N7
CY077417	1934	Rostock	Chicken	H7N1
CY014678	1949	Germany	Chicken	H10N7
CY015088	1959	Scotland	Chicken	H5N1
CY045334	1956	Czech_Republic	Duck	H4N6
CY014991	1961	South Africa	Tern	H5N3
CY015071	1963	England	Turkey	H7N3
DQ376870	1972	Taiwan	Duck	H6N1
CY024792	1976	Victoria	Chicken	H7N7
CY015043	1979	Leipzig	Goose	H7N7
CY015080	1983	Pennsylvania	Chicken	H5N2
CY117372	1988	Alberta	Mallard	H2N3
CY025084	1992	Victoria	Chicken	H7N3
CY005836	1994	Hidalgo	Chicken	H5N2
DQ997136	1997	Hubei	Chicken	H5N1
DQ997416	2002	Zhejiang	Duck	H5N1
CY103466	2008	Delaware Bay	Shorebird	H3N2
JX175257	2011	Guangdong	Duck	H3N2

Acc. Numb.	Year	City	Host	Strain
DQ208309	1918	Brevig_Mission	Human	H1N1
CY009611	1933	Wilson-Smith	Human	H1N1
CY020447	1936	Henry	Human	H1N1
CY013275	1940	Hickox	Human	H1N1
CY045779	1946	Melbourne	Human	H1N1
CY009347	1954	Malaysia	Human	H1N1
CY087804	1957	Japan	Human	H2N2
CY032268	1964	Cottbus	Human	H2N2
CY080530	1968	Hong_Kong	Human	H3N2
CY021964	1976	New_Jersey	Human	H1N1
DQ508894	1977	USSR	Human	H1N1
CY021044	1982	Christs_Hospital_UK	Human	H1N1
CY113369	1987	Shanghai	Human	H3N2
CY112844	1997	Hong_Kong	Human	H3N2
AJ278649	1999	Hong_Kong	Human	H9N2
CY006674	2003	New_York	Human	H1N1
CY118918	2006	Malaysia	Human	H3N2
GQ132145	2009	Mexico_InDRE4114	Human	H1N1
CY049984	2009	Mexico_City_WR1087T	Human	H1N1
JX046923	2012	Moscow	Human	H1N1

3. Methods

3.1. An Overview of the Mutual Information Function

Our goal is to quantify the degree of covariation of mutations of the 8 RNA segments of influenza A by using mutual information, a concept from information theory [32,33]. The identification of covarying sites (particular pairings with high mutual information values) is likely to confer a selective advantage in terms of either structure or function that facilitates the propagation of the virus. A formal measure of variability [34] at position i is the Shannon entropy, $H(i)$. $H(i)$ is defined in terms of the probabilities, $P(sym_i)$, of the different symbols, sym , that can appear at sequence position i (e.g., $sym = U, A, G, C, \dots$ for the four nucleotide bases of RNA). $H(i)$ is defined as:

$$H(i) = - \sum_{sym=A,U,G,C,\dots} P(sym_i) \log_2 P(sym_i). \tag{1}$$

Mutual information (MI) is defined in terms of entropies involving the joint probability distribution, $P(sym_i, sym'_j)$, of occurrence of symbol sym at position i , and sym' at position j . The probability, $P(sym_i)$, of a symbol appearing at position i regardless of what symbol appears at position j is defined by $P(sym_i) = \sum_{sym'_j} P(sym_i, sym'_j)$ and similarly, $P(sym'_j) = \sum_{sym_i} P(sym_i, sym'_j)$. Given the

above probability distributions, one can form the associated entropies:

$$H(i) = - \sum_{sym_i} P(sym_i) \log_2 P(sym_i),$$

$$H(j) = - \sum_{sym'_j} P(sym'_j) \log_2 P(sym'_j),$$

and:

$$H(i, j) = - \sum_{sym} \sum_{sym'_j} P(sym_i, sym'_j) \log_2 P(sym_i, sym'_j).$$

The mutual information, $I(i, j)$, is defined as:

$$I(i, j) = H(i) + H(j) - H(i, j). \tag{2}$$

The MI profile is a general measure of correlation between discrete variables, analogous to the Pearson's product-moment correlation coefficient for continuous variables. For RNA (or DNA) symbolic sequences, a MI profile between two symbols separated by a distance k is a function of k , called the mutual information function (MIF) [34]. The MI is particularly useful for analyzing correlation properties of symbolic sequences [34]. Lets denote by $A = \{A, U, G, C\}$ an alphabet and $s = (\dots, a_0, a_1, \dots)$ an infinite string with $a_i \in A$, $i \in \mathbb{Z}$, where \mathbb{Z} represents the set of all integer numbers and the values of a_i can be repeated. The MI of the string s and an identical string shifted k positions upstream is defined as:

$$I(k, s) = \sum_{\alpha \in A} \sum_{\beta \in A} P_{\alpha, \beta}(k, s) \log_2 \left[\frac{P_{\alpha, \beta}(k, s)}{P_{\alpha}(s)P_{\beta}(s)} \right] \tag{3}$$

where $P_{\alpha, \beta}(k, s)$ is the joint probability of having the symbol α followed k sites away by the symbol β on the string s , and $P_{\alpha}(s)$ and $P_{\beta}(s)$ are the marginal probabilities of finding α or β in the string s . By

choosing the logarithm in base 2, $I(k, s)$ is measured in bits. Both the joint probability and the marginal probabilities are estimated throughout the sequence as a global property. The function $I(k, s)$ can be interpreted as the average information over all positions that one can obtain about the actual value of a certain position in the string, given that one knows the actual value of the position k -characters away. The mutual information vanishes if, and only if, the events are statistical independent, *i.e.*, if all 16 joint probabilities $P_{\alpha, \beta}(k, s)$ factorize. Thus, the MI is a function capable of detecting any deviation from statistical independence. It must be noted from Equation (2) $I(i, j) = I(j, s)$, or from Equation (3) that $I(k, s) = I(s, k)$, and that $I(k, s) \geq 0$. The computation of the MI for a given sequence using different shifts of magnitude k provides an autocorrelation profile. Mutual Information values were obtained over the entire sequences of each of the eight segments and the plots illustrate the first 100 positions with the software MATLAB R2010b.

3.2. MI Dendrograms

Using complete sequences of the influenza A virus genome, similarity pairwise tests were performed with the criterion of Euclidian distance. The Euclidean distance between points p and q is the length of the line segment \overline{pq} . In Cartesian coordinates, if $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n -space, then the distance from p to q is given by:

$$d = \|p - q\| = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

The Euclidean distance performs an unbiased pairwise distance between pairs of objects. In MATLAB `pdist(X)` computes by default the Euclidean distance between pairs of objects in the $m \times n$ data matrix X . Columns of X correspond to observations (here the MI distances calculated), and rows correspond to variables (the genomic sequences). Given a $m \times n$ data matrix X , which is treated as m ($1 \times n$) row vectors x_1, x_2, \dots, x_m , the Euclidean distance between the vector x_s and x_t is defined as: $d_{st}^2 = (x_s - x_t)(x_s - x_t)^{transpose}$. The next step was the generation of the dendrogram, a hierarchization method obtained from the linkage of the pairwise distances. The distance matrix was linked with an Un-Weighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm [35], which in evolutionary theory assumes a constant substitution rate for all sites on the given sequence. The UPGMA algorithm is recurrently used in phenetics, but it does not integrate an actual evolutionary model.

3.3. Phylogenetic Reconstructions

Sequences were aligned using MUSCLE v3.7 [36] under default parameters, taking advantage of the `□refine` option in a second iteration. Complete Multiple Sequence Alignments (MSA) were visually inspected with Seaview 4 [37]. The Generalized Time-Reversible (GTR) evolution model [38] was selected relying on jModelTest analyses [39]. The evolutionary history was inferred using the Maximum Likelihood (ML) method [40]. The ML tree inferred from FastTree [41] under default parameters, is taken to represent the evolutionary history of the taxa analyzed. The ML tree is drawn as a cladogram as depicted by FigTree [42] for the respective comparison to MI dendrograms. A cladogram represents only the tree topology and its branch lengths do not represent time or relative amount of character change used to infer the phylogenetic tree. Each analysis involved 20 nucleotide sequences of human Influenza A (Table 1). All positions containing gaps and missing data were

eliminated. The total positions for each final set were: for S1, 2343 positions; for S4, 1837 positions; for S6, 1499 positions; for S8, 892 positions. Original ML trees as well as the alignments can be found in Supplementary Information.

4. Results

4.1. MI Profiles

For illustrative purposes, here we show a set of MI profiles that were obtained by using complete sequences of segments S1, S4, and S6. We selected the recently reconstructed 1918 Spanish flu virus [43] and compared it with the MI profile of isolates of virus in Mexico from the 2009 influenza pandemics [44,45] (Figure 1). In each plot only the first 100 positions are displayed for the sake of clarity. Note that both profiles of S1 A (H1N1) from 1918 and 2009 subtypes, which contains the gene for PB2, show an irregular oscillatory behavior along their whole sequences (Figure 1A). It is noteworthy to mention that there are regions (from positions 1 to 70) in which the oscillations are practically similar; but there are other regions (positions 75–90) in which there are clearly departures from each other. In contrast, the MI profiles for S4 (Figure 1B), which contains the gene for HA, are in general different from each other, being the 1918 strain the one that displays higher peaks than the 2009 strain. In both segments the behavior of the MI shows an irregular oscillatory behavior although they are roughly parallel. The discrepancies between the peaks of each segment correspond to relevant changes in information that presumably correspond to mutations that have occurred close to a century of evolution of A (H1N1). When comparing the MI's of S6 (Figure 1C), which contains the gene for NA, it is notorious that the MI of the subtype 2009 is ~tenfold higher in extent of information than the MI of flu virus of 1918, and it shows more peaks at intermediate and long distances. Despite the difference in the level of information, they are essentially parallel. This finding was unexpected but it is coherent with both the dendrogram and the phylogenetic analyses (See Section 4.2), since they are not closely similar in terms of MI values, and they are not closely related along sampling times. A MI profile of zero would indicate that mutations would be independent events and this would correspond to the case of neutral mutations which can be used as a control for neutral or nearly neutral evolution.

4.2. MI Dendrograms and Phylogenetic Reconstructions

Using all the values of MI profiles, dendrograms for each of the eight segments of the 39 influenza A subtypes, were calculated. Here we illustrate the dendrograms for S1, S4, S6, and S8 together with their respective Maximum Likelihood Phylogeny visualized as cladograms (Figure 2). Original ML trees as well as their corresponding alignments are provided in Supplementary Information. Dendrograms are rotated 90 degrees counterclockwise so that the abscissa is at the bottom and the units are in bits. Hence the number of bits associated for each split can be observed (Figure 2). The MI dendrogram of S1, that code for PB2 (a basic RNA polymerase) is presented in Figure 2A1. PB2 is a protein of the replication-transcription complex of influenza virus. PB2 is considered to be an evolutionary conserved molecule, meaning that small amount of changes should be expected.

Figure 1. Mutual Information of A(H1N1) from 1918 (red curves) and from 2009 (black curves): **(A)** S1; **(B)** S4; **(C)** S6.

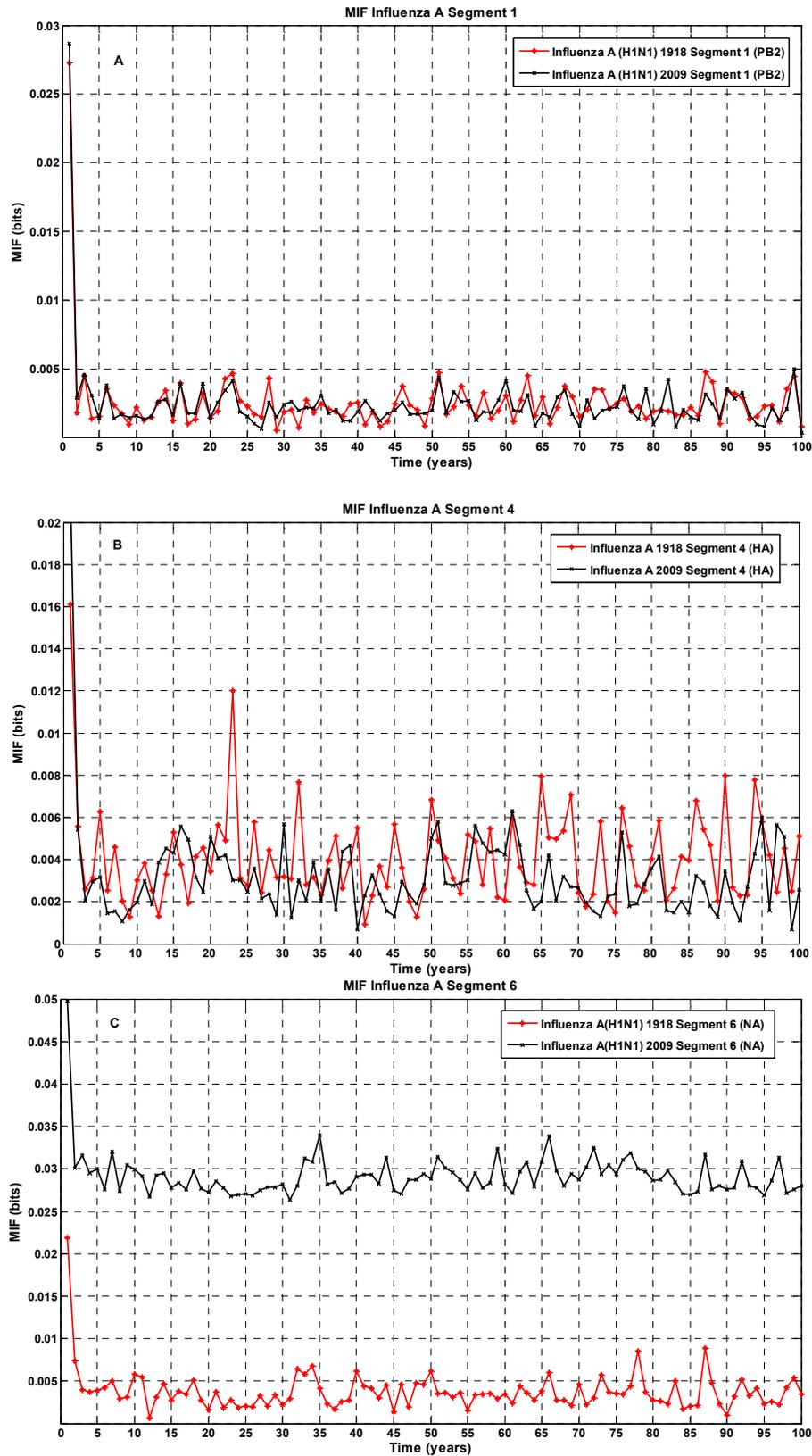


Figure 2. Dendrograms of the Mutual Information for: S1 (A1); S4 (B1); S6 (C1); and S8 (D1) segments of the RNA genome of influenza A. Maximum Likelihood Phylogenetic Reconstructions visualized as cladograms for: S1 (A2) LogLk (Log-likelihood) = -9150.268; S4 (B2) LogLk = -12865.908; S6 (C2) LogLk = -8565.225; and S8 (D2) LogLk = -3181.513. The numbers at each split in ML trees correspond to the Shimodaira-Hasegawa reliability estimate test which is part of the default parameter values of FastTree.

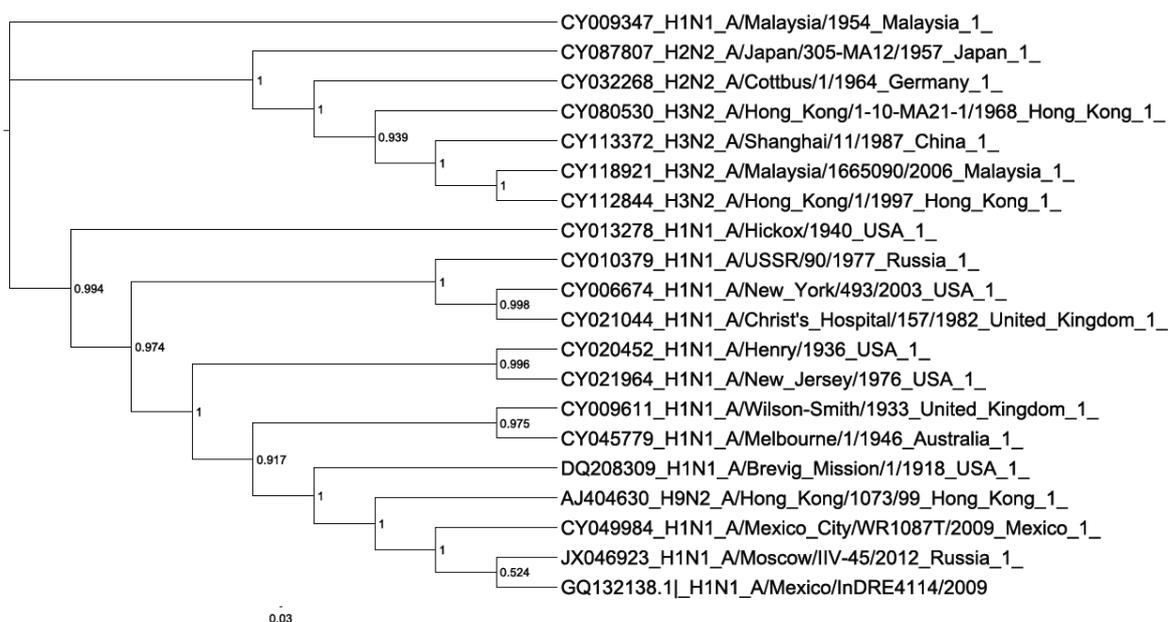
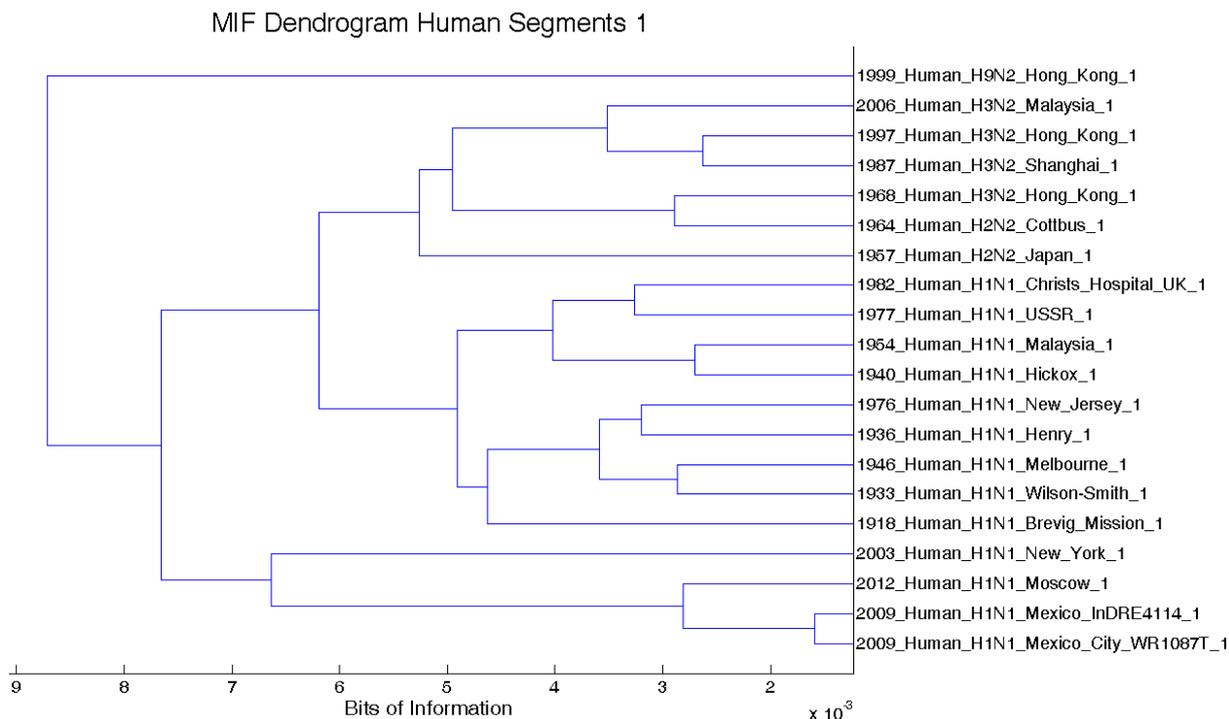
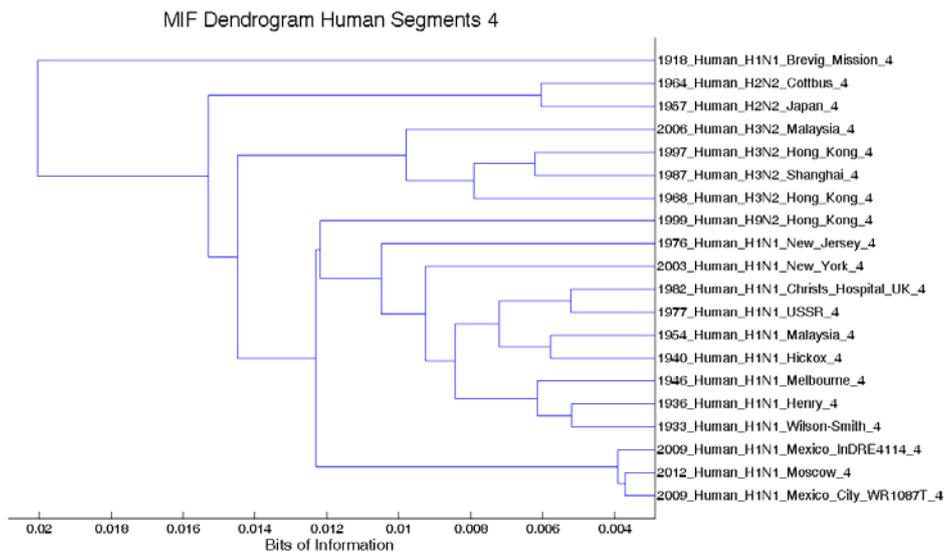
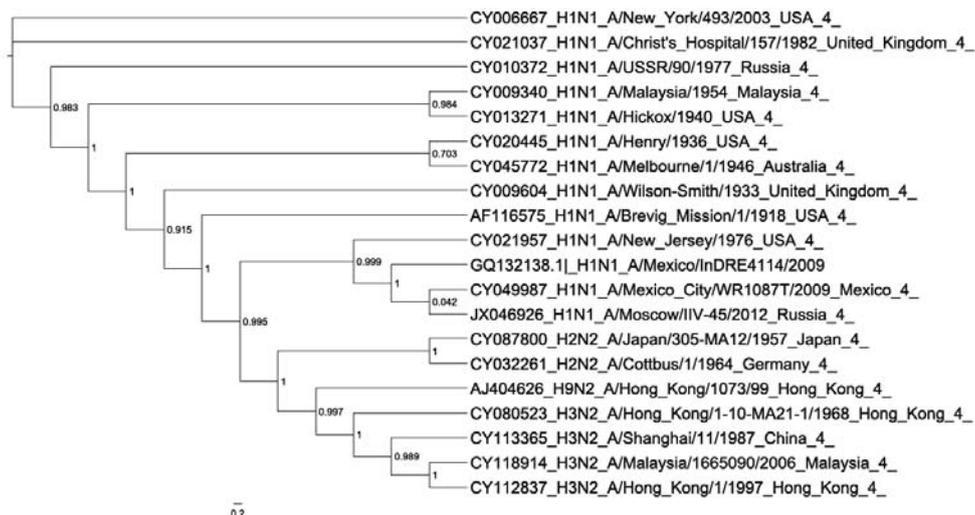


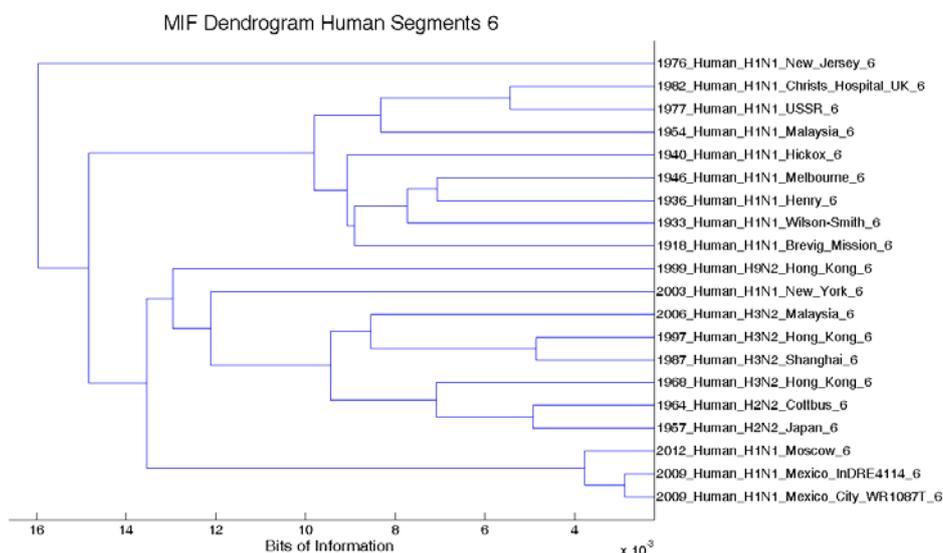
Figure 2. Cont.



(B1)

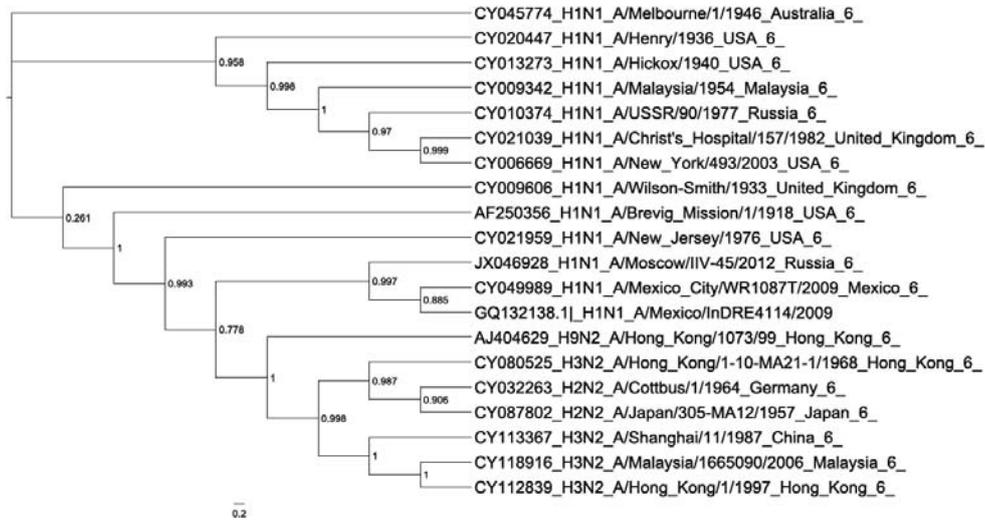


(B2)



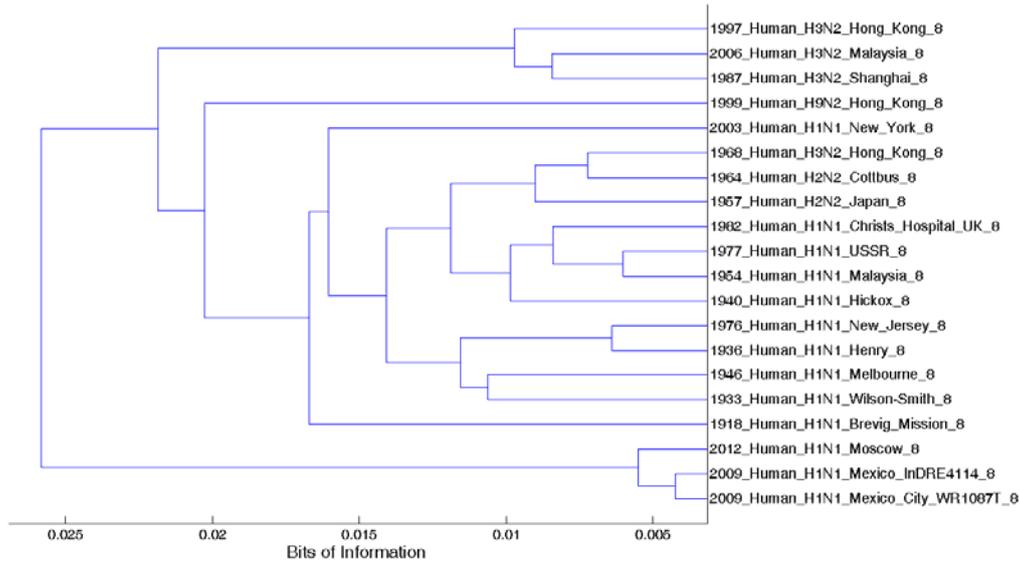
(C1)

Figure 2. Cont.

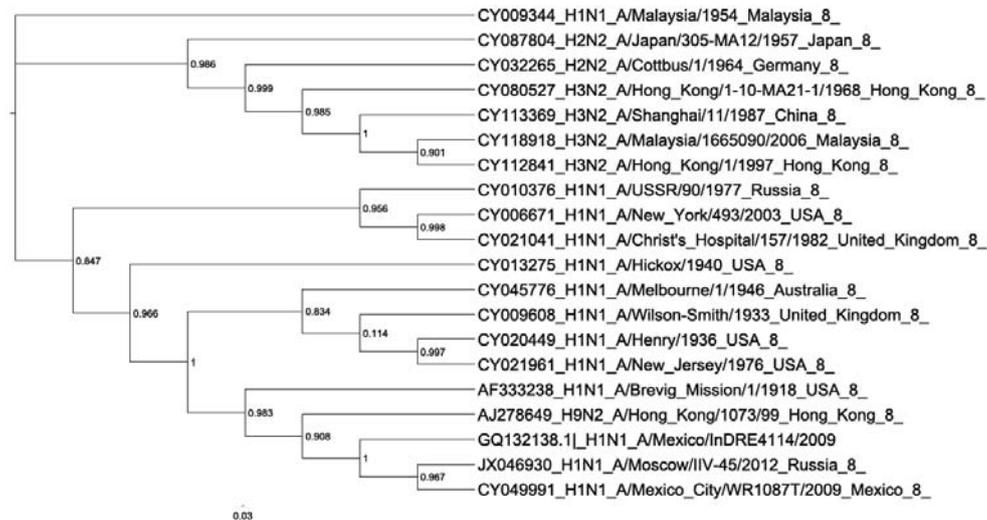


(C2)

MIF Dendrogram Human Segments 8



(D1)



(D2)

4.2.1. Dendrogram of S1

There is a clear separation between two groups: one group comprises H1 subtypes recently isolated. The 2nd group can still be subdivided into two subgroups: the subgroup of H1 subtypes isolated during the period 1918–1982, and the subgroup where all the non-H1 subtypes are encountered. The subtype 1999_H9N2 is located as the most dissimilar in this MI dendrogram, suggesting that PB2 of H9 is entirely different from the other sampled strains.

4.2.2. Phylogenetic Analysis of S1

The corresponding phylogeny also discerns between two groups: one group comprises N2 subtypes, meaning H3N2 and H2N2 subtypes with the exception of the H1N1_1954, which will repeat this grouping at the S8 phylogeny. The 2nd group comprises the H1N1 subtypes with the exception of the H9N2_1999 subtype.

It is remarkable how the MI dendrogram resembles the accurate discrimination between subtypes, in this case based at N1 and N2 viral classification. Also interesting is the similar clusterization of the most recent viral isolates, meaning that MI dendrogram may also capture vestiges of ancestry even when no evolutionary model is taken into account.

4.2.3. Dendrogram of S4

The similarities among the genomic segments of human influenza A that codes for HA are illustrated in a MI dendrogram (Figure 2B1). We remark that in this work we use the whole segments and not only the coding sequence. HA is an important viral protein, localized at the surface of the virion and it participates on the recognition of host immune cells and in cell invagination. It constitutes the 25% of the viral mass. Note that in the MI dendrogram (Figure 2B1) there is a clear distinction between the subtypes classified as H1 and the remaining ones. The MI dendrogram can discern between the types of HA by grouping exactly all H2 subtypes in one group, all H3 subtypes in another group, and all H1 subtypes still in another group. It is of interest that H9 is linked with the groupings of H1. On the other hand, the subtype H1_1918 is the most dissimilar sequence and it is located in the most outer part of the dendrogram. This could be attributable to the large number of carbohydrates that are expressed in the subtype H1_1918 in comparison with the remaining ones [46]. It is of notice the grouping of the subtypes 2009-2012 which is consistent with the dendrogram of S1.

4.2.4. Phylogenetic Analysis of S4

Overall, the topology of the phylogenetic trees for the S1, S2, S6, and S8 is ladder-type except for the vicinity of the most recent subtypes with the isolates during 1933–1946 (Figure 2A2, B2, C2, D2). Note also that the topology of the phylogenetic trees for S1, S4, S6, and S8, are not identical among them. In the MI dendrogram of S4 coding for HA, (Figure 2B1), we can observe that the 1918 H1N1 influenza A strain is arranged as the most dissimilar sequence, in this case the most ancestral one.

4.2.5. Dendrogram of S6

In Figure 2C1 the dendrogram corresponding to S6 containing the gene for the NA molecule is shown. This protein has also antigenic binding sites. The MI dendrogram discerns between the 2 main groups associated to N1 and N2. One group contains solely the N1 subtypes that appeared during 1918 to 1982. The 2nd group can be subdivided into two subgroups: to wit, those of subtype N2 that are linked with H1N1_2003. There is an interesting progression in this group which goes from 1957_H2N2 until 2006_H3N2 and then it culminates with the incorporation of the subtype 1999_H9N2. Interestingly enough, the recent H1 subtypes (2009–2012) are clustered in the 2nd subgroup and this is in harmony with the MI dendrograms of S2 and S4. The subtype 1976_H1N1 appears as the most dissimilar one.

4.2.6. Phylogenetic Analysis of S6

The comparison between the MI dendrogram of S6 (Figure 2C1) with its cladogram (Figure 2C2), as derived from bioinformatics methodologies, shows that the former resembles crisply and reasonable, similar subtype grouping. This is also observed in the dendrograms of S1, S4, S6, and S8. Bioinformatics methodologies organize subtypes adequately by homology (Figure 2A2, B2, C2, D2).

4.2.7. Dendrogram of S8

The segment S8 is interesting per se since it is bi-cistronic, *i.e.*, it encodes for 2 different proteins: to wit, NS1 which is the only Non-Structural protein of the virion which inhibits the mRNA splicing and the nuclear export of cellular and viral mRNA; it also plays a fundamental role in evading the immune response since it is an antagonist of interferon γ ; and NEP which is a protein with a signal of nuclear export and it is expressed from edited mRNA by splicing mechanisms. The topology of this dendrogram (Figure 2D1) is more complex than the previous ones: notwithstanding the MI dendrogram separates H1 and N2 subtypes, there are four groups and three sequences that are left without a pairwise similar subtype. Once more, a group is formed that comprises the most recent subtype 2009–2012 in agreement with the MI dendrograms of S1, S4, and S6. This group of recent ones is the most dissimilar of the whole dendrogram. The subtype that corresponds to influenza A from 1918 is set alone, sharing similarity with the 20th century H1N1 subtypes. The central group is the most complex as it generates two subgroups of H1 (1982, 1977, 1954, 1940, and 1933, 1946, 1936, 1976) flanking the subgroup formed by the subtypes N2 (1968, 1964, 1957). There is another subtype N2 (1999_H3N2) situated as an out-group-like of the central group. On the other hand, there is a group of the subtype N2 that turns out to be the 2nd less similar of the whole dendrogram and that involves the subtypes H3N2 of 1997, 2006, and 1987.

Overall, for the four dendrograms there are ~4.5-, ~6.6-, ~16-, and ~10 – fold dissimilarities with S1, S4, S6, and S8, respectively, with respect to the amount of information. In the MI dendrogram of S1 and S8 (Figure 2A1, D1), we are neither dealing with HA nor NA, but with PB2 and NS1-NEP. Yet, it can be observed in the latter figures that there are three informational clades with the nomenclature for HA and NA. One formed by the N2 subtypes from 1957 to 2006, the other formed by H1 and H9 subtypes from 1918 to 2009, and the third one comprises H9 as well as N2. Albeit there is

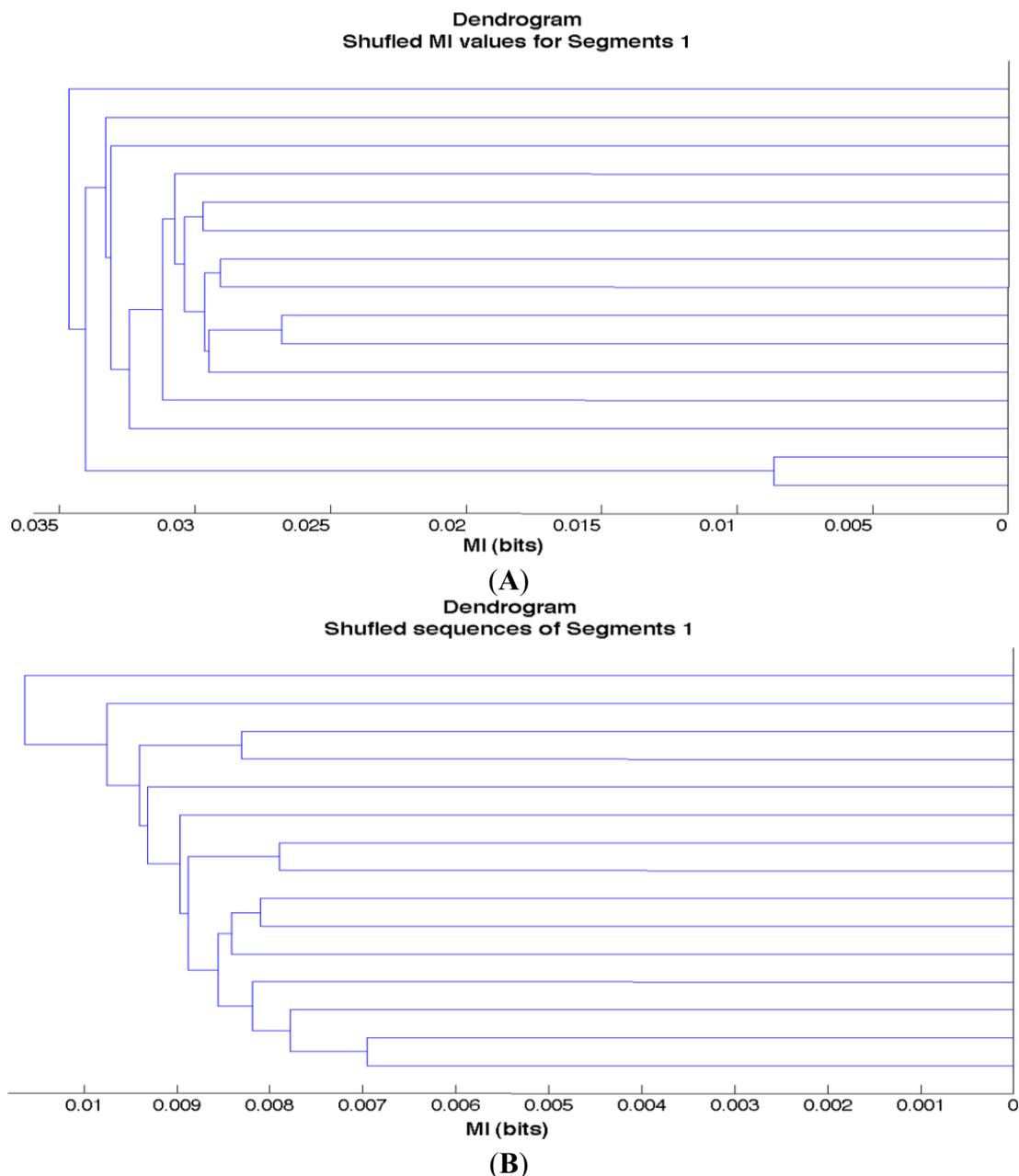
no nomenclature for polymerase subtypes of influenza A nor for the bi-cistronic gene, there seems to be a correlation with the canonical classification according to the types of HA's and NA's. By observing the corresponding phylogenetic reconstructions, there is a striking similarity to the informational cluster of the MI dendrogram. *In the human case the informational groups of the MI dendrogram correspond to the phylogenetic clades* (Figure 2A2, B2, C2, D2). Therefore, MI dendrograms capture evolutionary relevant information just as phylogenetic reconstructions do. Further statistical analysis show how the presence of the canonical subtypes HA and NA are related to subtypes of the remaining proteins (not shown).

In summary, all dendrograms of the MIs and the phylogenetic trees distinguish groups between subtypes and isolation times. Both clusterizations are surprisingly similar showing the capabilities of the MI to extract evolutionary important aspects of the actual RNA viral sequences. MI dendrograms for each segment reflected biologically coherent informational groups and were consistent with their corresponding clusters in the phylogenetic reconstructions. The MI dendrograms for the avian segments S1, S4, S6, and S8, as well as their corresponding phylogenetic reconstructions reinforce the same conclusions (not shown).

4.3. Random Controls

To test if the dendrograms of the MI could be the result of randomness we used two different controls. In the first one, we shuffled the actual biological sequences and then we calculated the MI dendrogram (Figure 3A). In the second one, the MI values obtained from the original viral sequences were randomized (Figure 3B). In both cases, we have disrupted the correlations that are detected in the actual biological sequences by the MI and therefore the autocorrelations selected during evolutionary pressures in these random controls disappear. Note that in both cases the hierarchies of the splits are soon generated in the most dissimilar possible zones, which lie in the truly random region. Henceforth all clusters are not alike to each other and these results validate the fact that our previous dendrograms cannot be the result of either randomness or of artifacts. We included these random controls to test whether the different subtypes of influenza A show evolutionary selection either for or against sequences that covary among them and favor particular informational groups. As the randomized controls do not possess biological meaningful information, we presumed that these controls are evolutionarily neutral with respect to viral evolution or genotype, such that preferred groupings would occur by chance. These results strongly argue that the viral genomes have evolved to favor particular informational groups according to the viral subtype.

Figure 3. Control experiments: (A) MI dendrogram from randomization of the actual viral sequences; (B) MI dendrogram from randomization of the MI values from the original viral sequences.



4.4. Flow of Information During Viral Evolution

Since evolutionary processes entail loss and gain of relevant information for survival, we estimated the flow of information during a century of evolution of influenza by determining the average information content for each segment in each influenza A subtype as a function of time irrespective of the subtype (Figure 4). Note that the content of information of PB2, for both human and avian strains, has oscillated throughout the sample years of registry (Figure 4A1). For human strains the range of oscillation is of the order of ~ 1.26 whereas for the avian strains this value is of the order of ~ 1.16 . Notice that the average value of information at the beginning of the XX century lies in the range of the

most recent values of information. Each trough and peak, in principle, is associated to antigenic shifts originating a pandemic at its time of appearance. In regard to the average extent of information of S4, its value in 1918 and 1988 was the highest among all human and avian subtypes, respectively, all along a century (Figure 4B1). There is an oscillatory behavior in both avian and human strains with most frequent oscillations in the latter than in the former (Figure 4B1 and inset). Note that over almost a century, the average information content of S6, containing the NA gene, appears to have cycles of small amplitude for the human strains whereas for the avian subtypes the pattern is also oscillating but with a large peak in 1963 (Figure 4C1). This abrupt increase is of the order of ~ 1.57 in regard to the remaining oscillating values. In regard to S8 of human subtypes its information content was the highest in 1918 and then oscillated in a lower range whereas in the avian subtypes the oscillations were in a still lower range except at peaks in 1988 and 2008. The prototypic pandemic H1N1 influenza virus emerged in 1918 and then gave rise to periodic seasonal strains that began to diminish in frequency during the late 1950. In all segments it can be observed a peak in 1977 indicating a resurgence of H1N1 viruses, thus reestablishing the H1N1 seasonal strains that are currently in circulation. Pandemic viruses typically evolve into seasonal forms (not shown in Figure 4) that develop resistance to antibody neutralization. These seasonal strains would form a slow oscillating trunk from which the pandemic strains depart as branches. This episodic pattern is consistent with a “cactus-like” structure of evolution [15] in which there are periods of neutral mutation during inter-pandemic periods interrupted by sudden bursts of change during antigenic shifts.

There seems not to be a singular characteristic which could herald the large change that was about to happen in the Mexican pandemic of 2009 in spite of its corresponding antigenic shift. The same holds true for any other pandemic.

Figure 4. Average amount of information for: S1 (A1); S4 (B1 and inset at other scale); S6 (C1); and S8 (D1) as a function of time from 1918 to 2009 in human (red curves) and from 1902 to 2011 in avian (blue curves) subtypes of the RNA genome of influenza A.

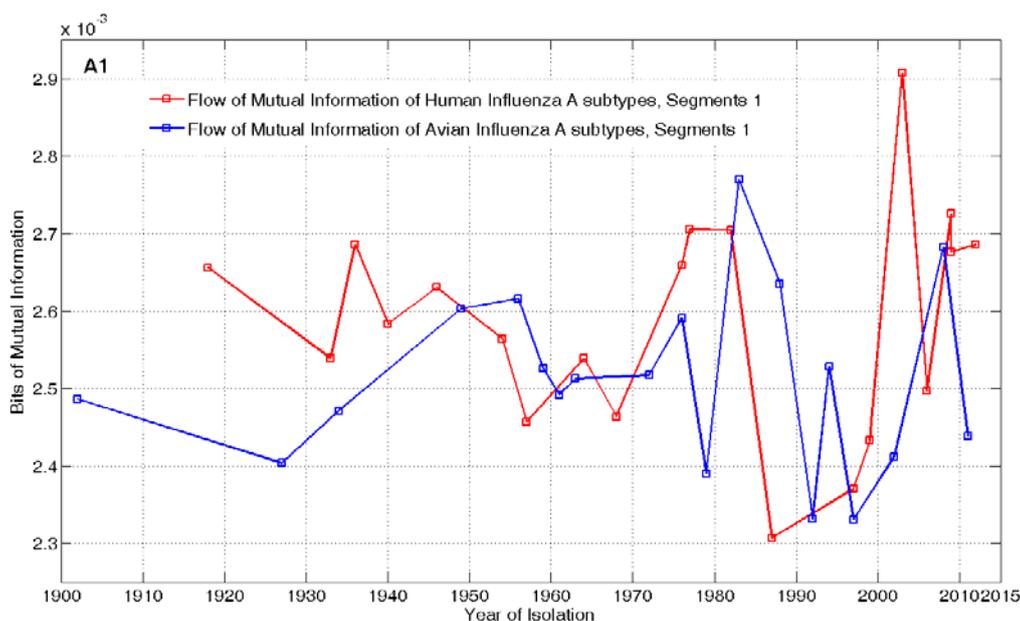
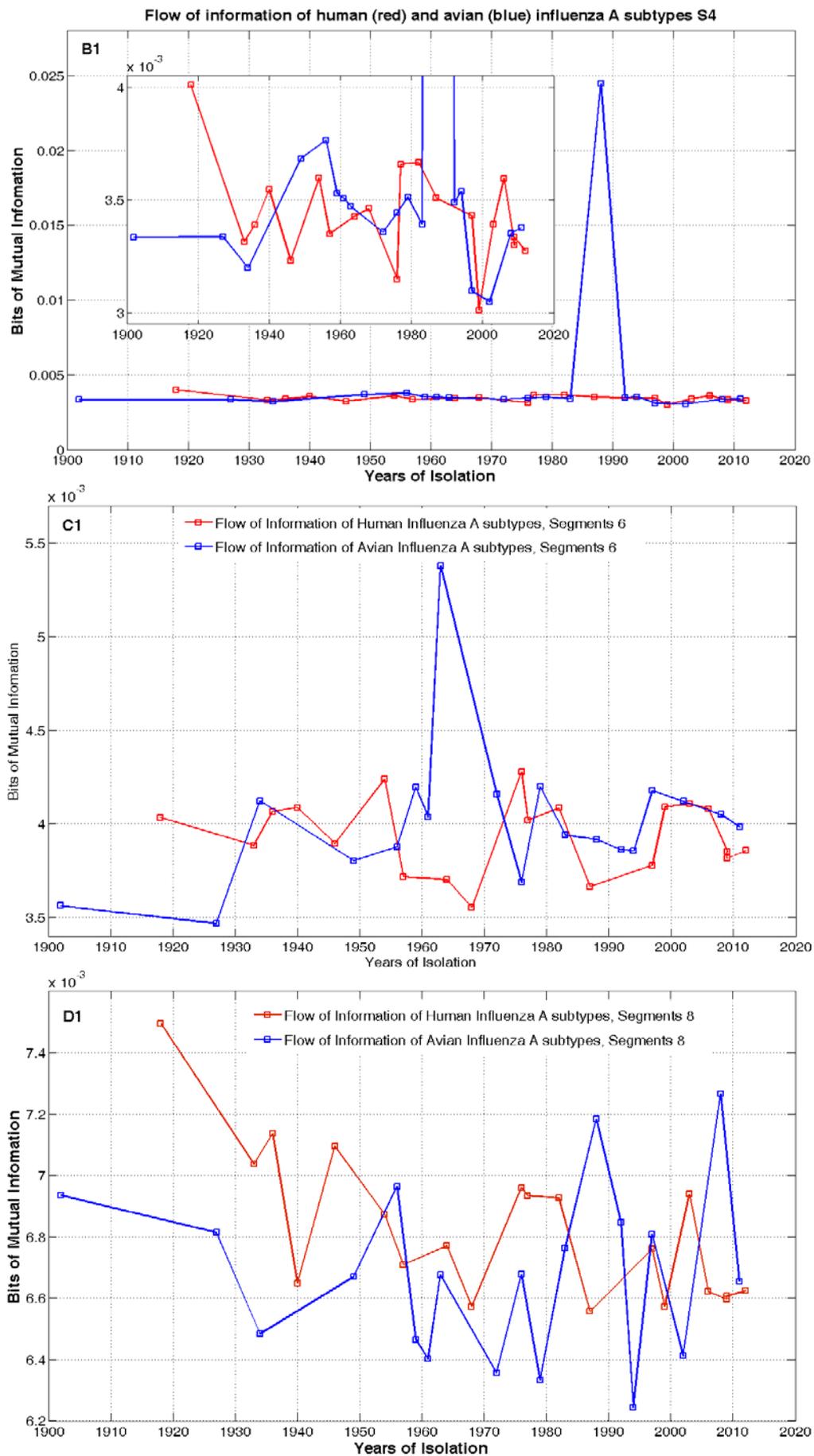


Figure 4. Cont.



5. Discussion and Conclusions

In this work, we have ushered in the use of the MI in symbolic series of genomic RNA sequences for capturing evolutionary relevant changes in information. The use of information theory in biology is not new. For example, the use of entropy and mutual information has been used for detecting gene-gene associations [47], multivariate entropy distance method has been developed for prokaryotic gene identification [48], the construction of phylogenetic trees using relative information between DNA sequences based on Lempel-Ziv complexity (particularly useful when multiple alignment based strategies fail) [49]; for detecting heart arrhythmias [50] and for determining nucleosome positioning motifs [51]. We use as an example 39 viral genomes comprising 312 sequences of influenza A that span over almost a century. We have applied standard information theory techniques to actual data on the history of influenza pandemics and their respective RNA symbolic sequences, and we have compared the results with Maximum Likelihood Phylogenetic methodology for determining evolutionary reconstruction histories of influenza A. We here report the use of the MI for detecting relevant informational changes which were known to be directly associated to the emergence of influenza pandemics associated to antigenic shifts in the eight segments of influenza A. In contrast, the analysis of more than 4,000 influenza A/H3N2 HA sequences from 1968 to 2010 using mutual information (MI)-based machine-learning model to design a site transition network for each amino acid site of HA to predict antigenic drifts has been reported [18]. The rate of prediction accuracy of this method is of the order of 70% [18]. In our present work we note that during antigenic shifts episodic changes in MI occur in all segments of influenza A. It has been found that positive selections are ongoing most of the time (*i.e.* not sporadic), and multiple mutations at antigenic sites cumulatively enhance antigenic drift [18,52]. It is worthwhile to mention that pandemic flu varieties evolve into seasonal flu varieties that are the ones that undergo antigenic drift.

The MIF is sensible enough to detect punctuated changes in a given viral sequence. The use of MI dendrograms certainly opens up new research avenues, such as elaborating a possible taxonomic classification of DNA and RNA viruses from an informational approach since it can clearly distinguish among them (ongoing work). MI dendrograms of S1, S2, S3, and S4, turned out to be practically identical or very similar in the clusterization, but not to the topology, of their corresponding Maximum Likelihood Phylogeny. A dendrogram is equivalent to a similarity analyses in evolutionary biology and is useful when the evolutionary dynamics of all the sequences are not fully understood. The combined use of the MI profiles with a simple hierarchical algorithm casted light upon the importance of relevant mutations associated to antigenic shifts. This was clearly illustrated for the case of influenza A (H1N1) of 1918 and 2009 (Figures 1 and 2). There are several evolutionary processes that might give rise to influenza strains with pandemic potential, and evaluating the significance of each of these has attracted much research effort. Molecular analyses suggest that the two previous pandemics involved reassortments between human-adapted and avian-adapted influenza viruses [43,44,53]. The 1957 pandemic probably involved reassortment between an avian H2N2 influenza virus and a human H1N1 influenza virus [54–56] and the 1968 pandemic probably involved reassortment between an avian H3Nx influenza virus and a human H2N2 influenza virus [54,57]. Even in these relatively well-documented cases, however, the species in which reassortment took place is not known. In this work we show that the MI profiles shed light upon the evolutionary process of influenza A. In particular we used segments

S1, S4, S6, and S8 which are clinically, pharmacologically, and immunologically of great interest. We highlight that the dendrograms of the MI profiles, constructed upon different strains of influenza A over time, indicate how the information flows among different viral subtypes during the evolution of these segments. It is clear that the evolutionary pattern that is observed in one segment is not equivalent to the rest of the viral genome. This means that focusing in changes in only one segment or a gene does not represent the whole evolutionary processes of the whole organism. For example, S1 and S8 show significant changes in information during a century of evolution (Figures 2 and 4). In other words, it is necessary to include all segments of influenza A genome to discern an accurate and more comprehensive picture of all evolutionary aspects that can act in part or in concert. The phylogenies of the HA genes of a given specific subtype like the human influenza A (H3N2) and human measles virus appear immediately distinct, with influenza exhibiting a ladder-like tree with a long trunk and very short side branches and measles presenting a bushier tree with deeper branchings [58]. For a specific subtype of influenza, the trunk branch corresponds to the progenitor lineage; mutations that occur along the trunk are eventually fixed, persisting until “over-written” by subsequent mutations. In contrast, mutations that appear on side branches are eventually lost from the population. Lack of powerful technologies and cutting-edge software programs are no longer an excuse for avoiding evolutionary analysis for whole genomes.

The flows of information over a century for the four segments were in agreement with their dendrograms in particular the dendrogram of S6. There is a significant abrupt increase in the information level of S6 of A (H1N1) subtype 2009 when compared with all the others (Figure 1C). This may have far-reaching implications in regard to the immune response, its contribution to the appearance of the pandemic, and for antiviral drug efficacy. Neuraminidase (NA; 1,407 nucleotides, S6), is involved in budding new virions from infected cells. Antiviral drugs are available for prevention and treatment of influenza. The influenza virus neuraminidase inhibitors zanamivir and oseltamivir were introduced into clinical practice in various parts of the world from 1999 through 2002 [58]. Oseltamivir limits replication of both influenza A and B viruses [59]. In most European countries, neuraminidase inhibitors are not widely used to treat seasonal influenza, but they are being stockpiled in many countries as part of their pandemic influenza preparedness. The drugs include amantadine, which inhibits the uncoating of virions by interfering with M2, and oseltamivir, which inhibits the release of virions from infected cells by interfering with NA [60]. According to our results, we anticipate that the efficacy of oseltamivir for A (H1N1) 2009 treatment might have changed.

The flows of information over a century for the 4 segments were in agreement with their dendrograms. *Mutual information may reflect but not predict relevant mutations of antigenic shifts driving influenza virus pandemics.* Every phylogeny depicts these processes and our analysis shows that MI dendrograms captures remarkably well the same evolutionary dynamics. Therefore, we are successfully connecting an evolutionary process with gains and losses of information. For each segment of the virus, we can say that during its evolution, new correlations arise or old correlations vanish, and they are all relevant for the evolutionary strategies of the virus. Correlations appear and disappear over time, and this is information conveyed through evolutionary processes. To our knowledge no other evolutionary tests are capable of determining the gains and losses in terms of bits of mutual information. The approach of the MI dendrograms was validated by phylogenetic reconstructions and both can certainly enhance our understanding of them. Whole-genome phylogenies

show the coexistence of multiple viral lineages, particularly on a limited spatial and temporal scale. This indicates that the transitions among antigenic types do not always proceed in a simple linear manner, that reassortments among coexisting lineages is relatively frequent, and that, for these reasons, predicting the path of influenza virus evolution from sequence data alone is inherently difficult. Interestingly, the resulting clusters of MI dendrograms turned out to be consistent with methodologies such as neighbor joining using MEGA5 (not shown) [61–63] as well as Maximum Likelihood methodologies [40,41]. Therefore we propose the use of the MI as a complement of subsequent bioinformatics methodologies.

Acknowledgments

MVJ was financially supported by PAPIIT-IN107112, UNAM, México. VSS was granted a doctoral fellowship from CONACYT and he also was partially supported by Coordinación de Estudios de Posgrado, UNAM.

Conflict of Interest

The author has no conflict of interest to declare.

References

1. Breitbart, M.; Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **2005**, *13*, 278–284.
2. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **2006**, *117*, 5–16.
3. Cornelis, G.; Heidmann, O.; Bernard-Stoecklin, S.; Reynaud, K.; Vêrond, G.; Mulote, B.; Dupressoir, A.; Heidmann, T. Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E432–E441.
4. Norrby, E. Nobel Prizes and the emerging virus concept. *Arch. Virol.* **2008**, *153*, 1109–1123.
5. Miller, E.S.; Kutter, E.; Mosig, G.; Arisaka, F.; Kunisawa, T.; Rûger, W. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 86–156.
6. Forterre, P.; Gribaldo, S.; Gadelle, D.; Serre, M.C. Origin and evolution of DNA topoisomerases. *Biochimie* **2007**, *9*, 427–446.
7. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **2008**, *9*, 267–276.
8. Eigen, M. The origin of genetic information: Viruses as models. *Gene* **1993**, *135*, 37–47.
9. Smith, R.A.; Loeb, L.A.; Preston, B.D. Lethal mutagenesis of HIV. *Virus Res.* **2005**, *107*, 215–228.
10. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **1999**, *401*, 877–884.
11. Sharp, P.M. Origins of human virus diversity. *Cell* **2002**, *108*, 305–312.
12. Rambaut, A.; Posada, P.; Crandall, K.A.; Holmes, E.C. The causes and consequences of HIV evolution. *Nat. Rev. Gen.* **2004**, *5*, 52–61.

13. Pybus, O.G.; Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Gen.* **2009**, *10*, 540–550.
14. Fitch, W.M.; Leiter, J.M.E.; Li, X.; Palese, P. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4270–4274.
15. Wolf, Y.I.; Viboud, C.; Holmes, E.C.; Koonin, E.V.; Lipman, D.J. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct* **2006**, *1*, 34.
16. Nelson, M.I.; Simonsen, L.; Viboud, C.; Miller, M.A.; Taylor, J.; St George, K.; Griesemer, S.B.; Ghedin, E.; Sengamalay, N.A.; Spiro, D.J.; *et al.* Stochastic processes are key determinants of short-term evolution in influenza A virus. *PLoS Pathog.* **2006**, *2*, e125.
17. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523.
18. Eigen, M.; Schuster, P. *The hypercycle A principle of natural self-organization*; Springer-Verlag: Heidelberg, Germany, 1979.
19. Stich, M.; Briones, C.; Manrubia, S.C. Collective properties of evolving molecular quasispecies. *BMC Evol. Biol.* **2007**, *7*, 110–123.
20. Wilke, C.O. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* **2005**, *5*, 44–52.
21. Kimura, M.; Maruyama, T. The mutational load with epistatic gene interactions in fitness. *Genetics* **1966**, *54*, 1337–1351.
22. Nelson, M.I.; Holmes, E.C. The evolution of epidemic influenza. *Nat. Rev. Gen.* **2007**, *8*, 196–205.
23. Hilleman, M.R. Realities and enigmas of human viral influenza: Pathogenesis, epidemiology and control. *Vaccine* **2002**, *20*, 3068–3087.
24. Murphy, B.R.; Webster, R.G. Orthomyxoviruses. In *Fields Virology*, 3rd ed.; Fields, B.N., Knipe, D.M., Howley, P.M., Eds.; Lippincott-Raven Publishers: Philadelphia, PA, USA, 1996; pp. 1397–1445.
25. De Jong, J.C.; Rimmelzwaan, G.F.; Fouchier, R.A.; Osterhaus, A.D. Influenza virus: A master of metamorphosis. *J. Infect.* **2000**, *40*, 218–228.
26. Ferguson, N.M.; Galvani, A.P.; Bush, R.M. Ecological and immunological determinants of influenza evolution. *Nature* **2003**, *422*, 428–433.
27. Shannon, C.E.A. Mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
28. Xia, Z.; Jin, G.; Zhu, J.; Zhou, R. Using a mutual information-based site transition network to map the genetic evolution of influenza A/H3N2 virus. *Bioinformatics* **2009**, *25*, 2309–2317.
29. Influenza Virus Resource. Available online: <http://www.ncbi.nlm.nih.gov/genomes/FLU/> (accessed on 15 November 2012).
30. Avian influenza A (H5N1)-update 31: Situation (poultry) in Asia: Need for a long-term response, comparison with previous outbreaks. Available online: http://www.who.int/csr/don/2004_03_02/en/ (accessed on 15 November 2012).
31. Influenza A virus (A/Brevig Mission/1/1918(H1N1)). Available online: <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=88776/> (accessed on 15 November 2012).
32. Kullback, S. *Information Theory and Statistics*; John Wiley and Sons: New York, NY, USA, 1959.
33. Blahut, R.E. *Information Theory and Statistics*; Addison-Wesley: Reading, MA, USA, 1987.

34. Li, W. Mutual information function vs. correlation functions. *J. Stat. Phys.* **1990**, *60*, 823–837.
35. Hartl, D.L. *A Primer of Population Genetics*, 3rd ed.; Sinauer Associates: Sunderland, MA, USA, 2000.
36. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
37. Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221–224.
38. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.
39. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **2008**, *25*, 1253–1256.
40. Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **1981**, *17*, 368–376.
41. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **2010**, *5*, e9490.
42. FigTree. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 10 June 2013).
43. Reid, A.H.; Fanning, T.G.; Janczewski, T.A.; Taubenberger, J.K. Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6785–6790.
44. Reid, A.H.; Taubenberger, J.K. The origin of the 1918 pandemic influenza virus: A continuing enigma. *J. Gen. Virol.* **2003**, *84*, 2285–2292.
45. Garten, R.J.; Davis, C.T.; Russell, C.A.; Shu, B.; Lindstrom, S.; Balish, A.; Sessions, W.M.; Xu, X.; Skepner, E.; Deyde, V.; *et al.* Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* **2009**, *325*, 197–201.
46. Wei, C.-J.; Boyington, J.C.; Dai, K.; Houser, K.V.; Pearce, M.B.; Kong, W.-P.; Yang, Z.-Y.; Tumpey, T.M.; Nabel, J.G. Cross-neutralization of 1918 and 2009 influenza viruses: Role of glycans in viral evolution and vaccine design. *Sci. Transl. Med.* **2010**, *2*, 24ra21.
47. Butte, A.J.; Kohane, I.S. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomp.* **2000**, *5*, 415–426.
48. Ouyang, Z.Q.; Zhu H.Q.; Wang J.; She Z.S. Multivariate entropy distance method for prokaryotic gene identification. *J. Bioinf. Comp. Biol.* **2004**, *2*, 353–373.
49. Out, H.H.; Sayood, K. A new sequence distance measure for phylogenetic tree reconstruction. *Bioinformatics* **2003**, *16*, 2122–2130.
50. Núñez-Acosta, E.; Lerma, C.; Márquez, M.F.; José, M.V. Mutual information analysis reveals bigeminy patterns in Andersen-Tawil syndrome and in subjects with history of sudden cardiac death. *Physica A: Statist. Mech. Appl.* **2012**, *391*, 693–707.
51. Sosa, D.; Miramontes, P.; Li, W.; Mireles, V.; Bobadilla, J.R.; José, M.V. Periodic distribution of a putative nucleosome positioning motif in human, non-human primates, and Archaea: Mutual information analysis. *Int. J. Genomics* **2013**, 963956.
52. Shih, A.C.; Hsiao, T.C.; Ho, M.S.; Li, W.H. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 6283–6288.
53. Webster, R.G.; Bean, W.J.; Gorman, O.T.; Chambers, T.M.; Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **1992**, *56*, 152–179.
54. Scholtissek, C.; Burger, H.; Bachmann, P.A.; Hannoun, C. Genetic relatedness of hemagglutinins of the H1 subtype of influenza A viruses isolated from swine and birds. *Virology* **1983**, *129*, 521–523.

55. Kawaoka, Y.; Krauss, S.; Webster, R.G. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J. Virol.* **1989**, *63*, 4603–4608.
56. Schafer, J.R.; Kawaoka, Y.; Bean, W.J.; Suss, J.; Senne, D.; Webster, R.G.; et al. Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible progenitors in the avian reservoir. *Virology* **1993**, *194*, 781–788.
57. Bean, W.J.; Schell, M.; Katz, J.; Kawaoka, Y.; Naeve, C.; Gorman, O.; Webster, R.G.; et al. Evolution of the H3 influenza virus hemagglutinin from human and non-human hosts. *J. Virol.* **1992**, *66*, 1129–1138.
58. Bedford, T.; Cobey, S.; Pascual, M. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evol. Biol.* **2011**, *11*, 220.
59. Hayden, F.G. Pandemic influenza: Is an antiviral response realistic? *Pediatr. Infect. Dis. J.* **2004**, *23*, S262–S269.
60. Moscona, A. Neuraminidase Inhibitors for Influenza. *N. Engl. J. Med.* **2005**, *353*, 1363–1373.
61. Tamura, K.; Nei, M.; Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 11030–11035.
62. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
63. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).