Open Systems & Information Dynamics
Vol. 17, No. 3 (2010) 297–310
DOI: 10.1142/S1230161210000199

World Scientific
www.worldscientific.com

How to Classify Influenza A Viruses and Understand Their Severity

Keiko Sato, Tomonori Tanabe, Masanori Ohya

Department of Information Science, Tokyo University of Science
Noda City, Chiba 278–8510, Japan

e-mail: keiko@is.noda.tus.ac.jp

(Received: June 9, 2010)

Abstract. As an application of the chaos degree introduced in the framework of information adaptive dynamics, we study the classification of the Influenza A viruses. What evolutional processes determine the severity and the ability for transmission among human of influenza A viruses? We performed phylogenetic classifications of influenza A viruses that were sampled between 1918 and 2009 by using a measure called *entropic chaos degree*, that was developed through the study of chaos in information dynamics. The phylogenetic analysis of the internal protein (PB2, PB1, PA, NS, M1, M2, NS1, and NS2) indicated that Influenza A viruses adapting to human and transmitting among human were clearly distinguished from swine lineage and avian lineage. Furthermore, the HA, NA, and internal proteins of the influenza strain that caused a pandemic or a severe epidemic with high mortality were phylogenetically different from those from previous pandemic and severe epidemic strains. We have come to the conclusion that the internal protein has a significant impact on the ability for transmission among human. Based of this study, we are convinced that entropic chaos degree is very useful as a measure of understanding the classification and severity of an isolated strain of influenza A virus.

1. Introduction

Influenza pandemics threaten our lives. The 20th century saw three influenza pandemics; the 1918 ("Spanish influenza") H1N1 virus, the 1957 ("Asian influenza") H2N2 virus and the 1968 ("Hong Kong influenza") H3N2 virus. According to the World Health Organization (WHO), the deaths of the 1918, 1957, and 1968 pandemics are estimated at 40–50 million, 2 million, and 1 million people worldwide, respectively. And in the Spring of 2009, a new H1N1 influenza virus emerged and has caused another pandemic. The new 2009 H1N1 virus still remains as a very high transmission among human and has spread worldwide. As of May 23, 2010, WHO reported that at least 18,114 people worldwide have been killed by the 2009 H1N1 infections. Compared with seasonal influenza, mortality was particularly high among infants.

Moreover, though it is not considered a pandemic, an H1N1 strain which

appeared in 1977 (called "Russian influenza") caused severe epidemic in children and young people worldwide. Like the 1977 epidemic, a large mortality was also recorded for the epidemics that occurred in 1928–1929 (H1N1), 1932–1933 (H1N1), 1951–1953 (H1N1), and 1997–1999 (H3N2) [1, 2, 3]. By contrast, the epidemic H1N1 virus which emerged in 1947 was globally distributed like a pandemic virus, but the mortality was relatively low like yearly influenza epidemics [2, 3, 4].

It is very interesting to find the dynamics causing the changes of viruses from one type to others. If one can get a rule to describe the dynamics such as Schrödinger equation of motion and traces the changes of the viruses, it will be a great step to study the genome of the influenza A virus, even more the disease due to that virus. However, it is also very difficult to find such an equation of motion and more difficult to solve the equation because the micro-dynamics of the virus change will be one of the multi-body problems with very complicated interactions. In any case, we have to look for some rule representing the dynamics even in the incomplete microlevel, that is, in a certain macrolevel. Most biologists discuss the changes of viruses by looking at each site of genome or counting the substitution rate of the sites. We use the probability theory and the chaos degree to study such changes, which provide us with slightly more rules than merely counting.

Influenza A viruses have eight pieces of segmented RNA that encode 11 proteins [5]. The antigenic properties in the two viral surface proteins, hemagglutinin (HA) and neuraminidase (NA), are used to classify influenza A viruses into different subtypes. Influenza A viruses representing 16 HA (H1 to H16) and 9 NA (N1 to N9) antigens have been detected in wild birds and poultry all over the world [5, 6]. Wild aquatic birds are considered to be natural reservoirs for all subtypes of Influenza A viruses [7, 8].

Avian influenza A viruses are classified as either highly pathogenic avian influenza (HPAI) or low pathogenic avian influenza (LPAI) viruses, based on a polybasic amino acid cleavage site within the HA and on the severity of disease. To date, only subtypes containing H5 or H7 have been found in the highly pathogenic form. Since 2003, HPAI H5N1 viruses have spread throughout Asia, Europe, Middle East, North and West Africa with outbreaks in poultry and cases of human infection [9, 10]. Despite widespread exposures to poultry infected with the H5N1 viruses, the H5N1 viruses have never circulated widely among humans [11]. According to the WHO, almost all human infections can be linked to contact with infected poultry.

Currently, influenza A viruses circulating among humans are H1N1, H1N2 and H3N2 subtypes. Influenza A viruses continually change by the accumulation of point mutations in the genes that encode the two HA and NA proteins. This continual change is called "antigenic drift" and is associated with seasonal influenza epidemics. Besides antigenic drift, Influenza A viruses suddenly change to form a new influenza A subtype or a virus with HA segment

or HA and NA segments emerging by reassortment that two more influenza virus strains of the same or different subtypes co-infect a single host cell and whereby exchanging RNA segments [8]. This sudden change is called "antigenic shift" and is associated with pandemics. The cause of the 1957 and 1968 pandemics is considered as antigenic shift [4, 12], though the prerequisite condition for pandemic emergence is unknown [13, 14]. The Centers for Disease Control and Prevention reports the cause of the 2009 pandemic to be antigenic shift. Moreover, the cause of the 1947, 1951–1953, and 1997–1999 epidemics is considered as intrasubtypic antigenic shift [3, 14]. Other severe epidemic viruses with high mortality result from antigenic drift, such as seasonal influenza viruses [3].

Given this, we have the following questions:

(1) What are the conditions required for human to human transmission of influenza?
(2) What evolutional processes determine the severity of influenza A viruses?
(3) What are the conditions leading to differences between severe epidemic influenza viruses, pandemic influenza viruses and seasonal influenza viruses?
(4) If the reason for the differences is antigenic change in HA segment or HA and NA segments, we would like to know how much change HA and NA proteins or other proteins are needed to cause severe epidemics or pandemics.
(5) In particular, we are very much interested in knowing that once a new influenza A virus appears, is it mild or severe for human.

We try to find the solutions of these questions. For this purpose, we applied *entropic chaos degree* to the processes of sequence changes in 10 proteins (PB1, PB2, PA, HA, NP, NA, M1, M2, NS1, and NS2) of influenza A viruses that were sampled between 1918 and 2009.

2. Methods

2.1. Entropic chaos degree

Entropic chaos degree (ECD for short) has been used to characterize the chaotic aspects of the dynamics leading to sequence changes [15, 16]. We will briefly explain the ECD. The ECD for the amino acid sequences is given as follows [17]. Let us take the amino acid sequences of two influenza A viruses X and Y. We would like to find a rule by which X changes to Y if X is supposed to be ahead of Y. As we explained above, it is almost impossible to write down the equation of motion in the very micro level, say the level of quantum mechanics. However, it is true that there exists some

micro-dynamics causing the change from X to Y even if we do not know the exact form of the dynamics. We denote this dynamics by Λ^*_{micro} and its extension to a macro scale properly considered by Λ^*. Even when the micro-dynamics Λ^*_{micro} is hidden, we can somehow find the macrodynamics Λ^*. The macrodynamics we consider here is the one in usual discrete probability theory, in which we can apply the usual Shannon's information theory. So the macro-dynamics Λ^* is nothing but a channel from the probability space of X to the probability space of Y.

After two amino acid sequences X, Y are aligned, they are denoted by the same symbols X and Y. The complete event system of X is determined by the occurrence probability p_i of each amino acid a_i, $i = 1, \ldots, 20$, and p_0 of the gap $*$,

$$(X, p) \;=\; \begin{pmatrix} * & a_1 & \ldots & a_{20} \\ p_0 & p_1 & \ldots & p_{20} \end{pmatrix},$$

where $\sum_{i=1}^{20} p_i = 1$. In the same way, the complete event system Y is denoted by

$$(Y, \bar{p}) \;=\; \begin{pmatrix} * & a_1 & \ldots & a_{20} \\ \bar{p}_0 & \bar{p}_1 & \ldots & \bar{p}_{20} \end{pmatrix}.$$

The compound event system of X and Y is denoted by

$$(X \times Y, r) \;=\; \begin{pmatrix} ** & *a_1 & \ldots & a_{20}a_{20} \\ r_{00} & r_{01} & \ldots & r_{2020} \end{pmatrix},$$

where r_{ij} represents the joint probability for the event a_i of X and the event a_j of Y satisfying

$$\sum_{j=0}^{20} r_{ij} \;=\; p_i\,, \qquad \sum_{i=0}^{20} r_{ij} \;=\; \bar{p}_j\,.$$

So the dynamics Λ^* describing the change of a sequence from X to Y is given by a certain mapping called a channel sending the probability distribution p to $\bar{p} = \Lambda^* p$. As stated above, it is difficult to know the details of this dynamics in the course of sequence changes. The ECD can be used to measure the complexity without knowing the exact dynamics [18], which is one of the aspects due to the sequence change.

The ECD for the amino acid sequences is given by the following formula:

$$\mathrm{ECD}(X, Y) \;\equiv\; \sum_i p_i S(\Lambda^* \delta_i)\,,$$

where S is the Shannon entropy and $p = \sum_i p_i \delta_i$, $\delta_i(j)$ is the Kronecker delta. Note that $\mathrm{ECD}(X, Y)$ can be written as $\mathrm{ECD}(p, \Lambda^*)$ to indicate p and

Λ^* explicitly. When p, $\bar{p}$ and r are obtained by some proper means as above, the ECD is represented as

$$\mathrm{ECD} \;=\; \sum_{ij} r_{ij} \log \frac{p_i}{r_{ij}} \,.$$

In this case, this ECD is nothing but the conditional entropy associated with the dynamics Λ^*.

This chaos degree is originally designed to measure how much chaos is produced by the dynamics [18, 19], that is,

(1) it produces chaos iff ECD > 0,
(2) it does not produce chaos iff ECD $= 0$.

Moreover, the chaos degree $\mathrm{ECD}(X, Y)$ provides a certain difference between X and Y through a change from X to Y, so that the chaos degree characterizes the dynamics changing X to Y.

For classification of influenza A virus strains that were sampled between 1918 and 2009, we used the *rate of entropic chaos degree* (RECD) defined by

$$\mathrm{RECD}(X, Y) \;\equiv\; \frac{\mathrm{ECD}(X, Y)}{S(Y)} \,.$$

2.2. Evolution analysis of Influenza A viruses by ECD

To study the evolutional dynamics of influenza A viruses, we calculated RECD between influenza A virus and another one. Amino acid sequences of 10 proteins (i.e., PB2, PB1, PA, HA, NP, NA, M1, M2, NS1, and NS2, except PB1–F2) from all 8 genomic segments that were sampled between 1918 and 2009 were downloaded from Influenza Virus Resources at the National Center for Biotechnology Information (NCBI) and Influenza Research Database at the Biodefense and Public Health Database (BioHealthBase) Bioinformatics Resource Center (BRC). In this study, we used 10 encoded protein sequences from available influenza A virus strains, where all 8 genomic segments were completely sequenced. All strain names are listed in Table 1.

For each of the 10 proteins, we aligned the amino acid sequences [20] and evaluated the RECD between the sequences. The $\mathrm{RECD}_{\mathrm{HA}}(X, Y)$, $\mathrm{RECD}_{\mathrm{NA}}(X, Y)$, and $\mathrm{RECD}_{\mathrm{inter}}(X, Y)$ represent the RECD of HA protein, NA protein, and internal viral protein (PB2, PB1, PA, NS, M1, M2, NS1, and NS2) for a strain X with another strain Y, respectively. The $\mathrm{RECD}_{\mathrm{inter}}(X, Y)$ was defined as

$$\begin{aligned}\mathrm{RECD}_{\mathrm{inter}}(X, Y) \;=\; \frac{1}{8}\Big[&\mathrm{RECD}_{\mathrm{PB2}}(X, Y) + \mathrm{RECD}_{\mathrm{PB1}}(X, Y) \\ &+ \mathrm{RECD}_{\mathrm{PA}}(X, Y) + \mathrm{RECD}_{\mathrm{NP}}(X, Y) + \mathrm{RECD}_{\mathrm{M1}}(X, Y) \\ &+ \mathrm{RECD}_{\mathrm{M2}}(X, Y) + \mathrm{RECD}_{\mathrm{NS1}}(X, Y) + \mathrm{RECD}_{\mathrm{NS2}}(X, Y)\Big] \,.\end{aligned}$$

Table 1: All strain names of Influenza A viruses used in this study

A/Brevig Mission/1/1918(H1N1)	A/chicken/Pennsylvania/1/1983(H5N2)
A/crow/Kyoto/53/2004(H5N1)	A/swine/1931(H1N1)
A/laughing gull/New Jersey/75/1985(H2N9)	A/swine/Henan/wy/2004(H5N1)
A/Wilson-Smith/33(H1N1)	A/Memphis/12/1986(H1N1)
A/Thailand/16/2004(H5N1)	A/Melbourne/35(H1N1)
A/mallard/Ohio/48/1986(H3N2)	A/Viet Nam/1203/2004(H5N1)
A/Phila/1935(H1N1)	A/swine/Virginia/671/1987(H1N1)
A/swine/Spain/53207/2004(H1N1)	A/Henry/1936(H1N1)
A/ruddy turnstone/DE/2378/1988(H7N7)	A/wood duck/Ohio/423/2004(H5N1)
A/Hickox/1940(H1N1)	A/chicken/New York/28263/1989(H6N3)
A/turkey/Italy/4479/2004(H7N3)	A/Bellamy/1942(H1N1)
A/mallard duck/ALB/155/1990(H6N3)	A/chukar/New York/11653-1/2005(H7N2)
A/Cameron/1946(H1N1)	A/blue-winged teal/Alberta/141/1992(H1N1)
A/duck/Viet Nam/12/2005(H5N1)	A/Fort Monmouth/1/1947(H1N1)
A/chicken/Hidalgo/28159-232/1994(H5N2)	A/duck/Viet Nam/1/2005(H5N1)
A/Fort Worth/50(H1N1)	A/chicken/New York/13828-3/1995(H2N2)
A/Indonesia/5/2005(H5N1)	A/Albany/13/1951(H1N1)
A/Goose/Guangdong/1/96(H5N1)	A/chicken/Nigeria/641/2006(H5N1)
A/Albany/1618/1951(H1N1)	A/chicken/Hubei/wj/1997(H5N1)
A/Indonesia/CDC594/2006(H5N1)	A/duck/England/1/1956(H11N6)
A/Hong Kong/483/97(H5N1)	A/China/GD01/2006(H5N1)
A/Japan/305/1957(H2N2)	A/chicken/Chis/15224/1997(H5N2)
A/environment/Maryland/1189/2006(H5N1)	A/Denver/57(H1N1)
A/blue wing teal/Ohio/31/1999(H3N2)	A/Cygnus cygnus/Iran/754/2006(H5N1)
A/chicken/Scotland/1959(H5N1)	A Virus A/New South Wales/8/1999(H3N2)
A/duck/Italy/194659/2006(H3N2)	A/tern/South Africa/1961(H5N3)
A/Hong Kong/1774/99(H3N2)	A/Shanghai/1/2006(H5N1)
A/equine/Sao Paulo/6/1963(H3N8)	A/New York/146/2000(H1N1)
A/grey heron/Hong Kong/3088/2007(H5N1)	A Virus A/Albany/10/1968(H3N2)
A/swine/Bakum/1832/2000(H1N2)	A/Jiangsu/1/2007(H5N1)
A Virus A/Beijing/1/68(H3N2)	A/duck/Guangxi/xa/2001(H5N1)
A/shorebird/Delaware/472/2007(H5N1)	A/Hong Kong/1-1-MA-12/1968(H3N2)
A/swine/Spain/33601/2001(H3N2)	A/turkey/Saudi Arabia/6732-6/2007(H5N1)
A/equine/Sachiyama/1/1971(H3N8)	A/swine/Spain/39139/2002(H3N2)
A/swine/Minnesota/SG-00239/2007(H1N2)	A/Memphis/1/1971(H3N2)
A/New York/78/2002(H1N2)	A/Taiwan/70120/2008(H3N2)
A/Memphis/102/1972(H3N2)	A/chicken/Hebei/108/02(H5N1)
A/Pennsylvania/PIT43/2008(H3N2)	A/equine/Kentucky/1a/1975(H7N7)
A/chicken/Jilin/hd/2002(H5N1)	A/whooper swan/Akita/1/2008(H5N1)
A/duck/Hong Kong/7/1975(H3N2)	A/duck/Zhejiang/bj/2002(H5N1)
A/District of Columbia/WRAMC-1154047/2008(H1N1)	A/pintail duck/ALB/86/1976(H3N2)
A/blue-winged teal/Ohio/908/2002(H3N2)	A/Taiwan/70167/2008(H1N1)
A/swine/Colorado/1/77(H3N2)	A/chicken/Shantou/3744/2003(H5N1)
A/peregrine falcon/Hong Kong/2142/2008(H5N1)	A/mallard duck/ALB/127/1977(H1N1)
A/Beijing/01/2003(H5N1)	A/whooper swan/Hokkaido/1/2008(H5N1)
A/swine/Arizona/148/1977(H1N1)	A/duck/Guangxi/12/2003(H5N1)
A/swine/Hong Kong/294/2009(H1N2)	A/USSR/92/77(H1N1)
A/duck/Guangxi/27/2003(H5N1)	A/California/04/2009(H1N1)
A/swine/Tennessee/25/1977(H1N1)	A/goose/Jilin/hb/2003(H5N1)
A/New York/3002/2009(H1N1)	A/Hong Kong/117/77(H1N1)
A/quail/Italy/4610/2003(H7N2)	A/Mexico/47N/2009(H1N1)
A/Tientsin/78/1977(H1N1)	A/swine/Spain/51915/2003(H1N1)
A/New York/3227/2009(H1N1)	A/mallard/Alberta/77/1977(H2N3)
A/Swine/Spain/50047/2003(H1N1)	A/Italy/05/2009(H1N1)
A/California/10/1978(H1N1)	A/New York/32/2003(H3N2)
A/Shanghai/37T/2009(H1N1)	A/red-necked stint/Australia/4189/1980(H4N8)
A/New York/61A/2003(H3N2)	A/Guangdong/02/2009(H1N1)
A/pheasant/MN/917/1980(H7N3)	A/swine/Ontario/52156/03(H1N2)
A/Nanjing/2/82(H3N2)	A/chicken/Kyoto/3/2004(H5N1)

To classify influenza A viruses according to the value of the RECD, phylogenetic tree was constructed with the neighbour-joining method of PHYLIP version 3.69 (Felsenstein J. PHYLIP: Phylogeny Inference Package) [21, 22] using the difference measured by the RECD. We carried out phylogenetic analysis of the HA, NA and internal virus protein sequences and examined the evolutionary process of influenza A viruses to find a factor causing various degree of influenza activity.

3. Results and Discussion

Phylogenetic analysis of the internal viral protein gave the following results shown in Fig. 1:

(1) H5N1 viruses isolated from human in Eurasia between 2003 and 2007 were closely related to H5N1 viruses isolated from birds in Eurasia after 2001.
(2) Influenza A viruses adapting to human and transmitting among human were distinguished from swine lineage and avian lineage, which is one of the most important results.
(3) Human influenza A viruses of human lineage fell into separate clusters (1918, 1933–1936, 1940–2008, 1968–1982, 1977, 1999–2008, and 2009).

We insist that influenza A viruses were clearly classified into three lineages (i.e., Avian lineage, Human lineage, and Swine lineage) for the phylogenetic analysis by the RECD, which gives better classification compared with a tree obtained by another measure called the substation rate.

We will explain (3) above in more details. The 1933 H1N1 severe epidemic strain (A/Wilson-Smith/1933(H1N1)) belonged to the 1933 – 1936 cluster that contains H1N1 strains isolated between 1933 and 1936. The 1957 H2N2 pandemic strain (A/Japan/305/1957(H2N2)) belonged to the 1940 – 2008 cluster, together with H1N1 strains from 1940 – 2008 including the A/Fort Monmouth/1/1947(H1N1), A/Fort Worth/50(H1N1) and A/Albany/13/1951 (H1N1). Human seasonal influenza H1N1 strains currently circulating fell into the 1940 – 2008 cluster. Note that the 1977 H1N1 severe epidemic strain (A/USSR/92/1977) did not belong to the 1940 – 2008 cluster. The A/USSR/92/1977(H1N1) was quite distinct from that cluster. The 1968 H3N2 pandemic strain (A/Hong Kong/1-1-MA-12/1968(H3N2)) belonged to the 1968 – 1982 cluster that comprises H3N2 strains isolated between 1968 and 1982. The 1999 H3N2 severe epidemic strain (A/New South Wales/8/1999(H3N2)) belonged to the 1999 – 2008 cluster, together with H1N2 strain isolated in 2002 and H3N2 strains isolated between 1999 and 2008. Human seasonal influenza H3N2 strains currently circulating fell into the 1999-2008 cluster. The 2009 H1N1 pandemic strain (A/California/04/2009(H1N1)) belonged to the 2009 cluster that contains 2009 H1N1 strains.

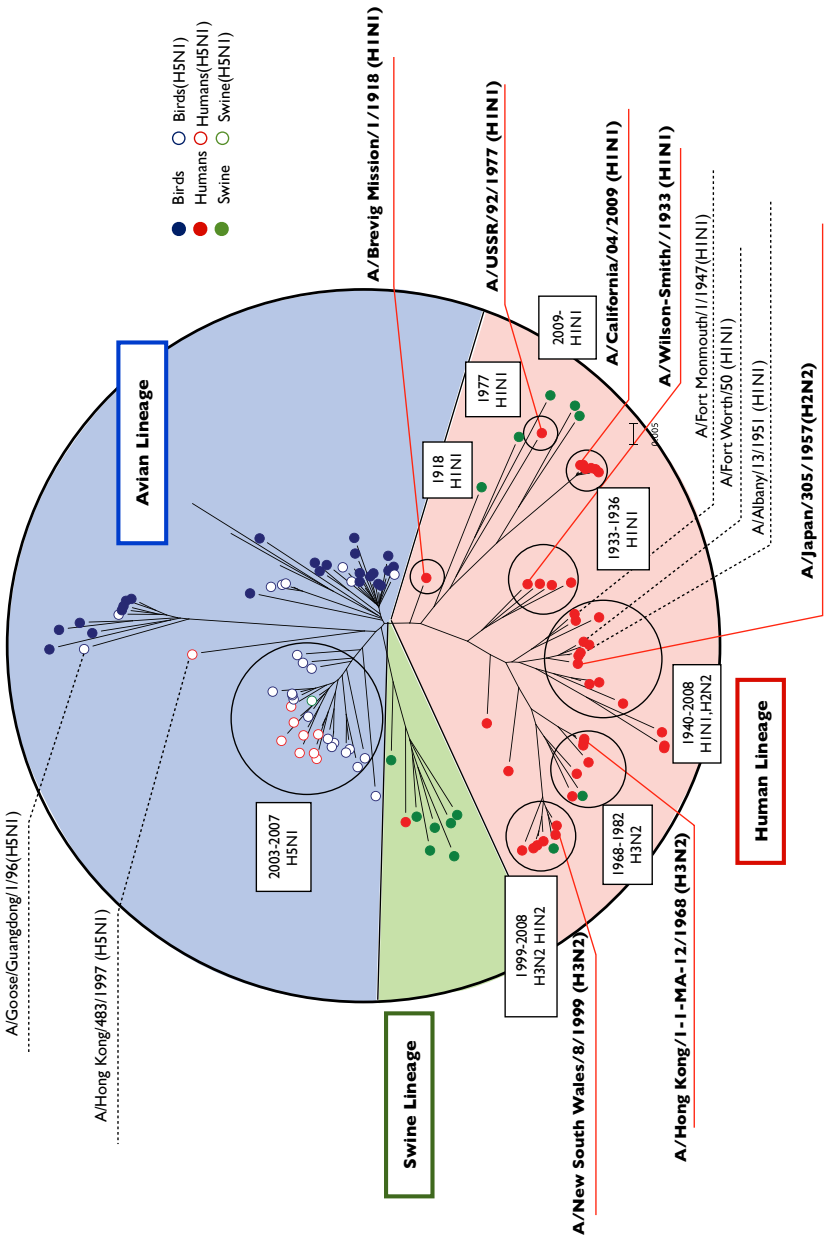

Fig. 1: Phylogenetic classification of the internal protein of influenza A viruses. Avian, human, and swine influenza A virus strains are represented by blue, red, and green circles. Influenza A virus strains that caused pandemics or severe epidemics with high mortality are indicated in boldface with red lines.

In contrast, phylogenetic trees for HA and NA proteins (Figs. 2 and 3) showed influenza A viruses were distinguished not by the species but the subtype of their strains. All H5N1 strains isolated from human in Eurasia between 1997 and 2007 were grouped into the same cluster, together with all H5N1 strains isolated from birds in the same region since 1996. The human H1 and H3 subtype viruses in Fig. 2 formed several clusters (1918, 1933 – 1942, 1940 – 2000, 1977, 2002 – 2008, 2009 for the H1, and 1968 – 1982 and 1999 – 2008 for the H3). The human N1 and N2 subtype viruses in Fig. 3 also formed several clusters (1918, 1933 – 1947, 1950 – 2008, 1977, 1977 – 2007, 2009 for the N1, and 1957, 1968 – 1999 and 1999 – 2008 for the N2).

These results indicate that the influenza strains that caused severe epidemics with high mortality or pandemics emerged under the following conditions:

(1) The HA and NA proteins of such epidemic or pandemic strain are necessary to be different from those of human strains that had already appeared prior to that strain, namely new HA protein and new NA protein.

(2) As for the internal protein (PB2, PB1, PA, NS, M1, M2, NS1, and NS2), the influenza strain that caused such epidemic or pandemic is necessary to be different from previous severe epidemic and pandemic strains.

In addition, a prerequisite for human-human transmission is that the internal protein is located in the human lineage.

The 1933 H1N1 strain that caused a severe epidemic with high mortality was distinct from the 1918 H1N1 strain that appeared prior to the 1933 strain in the internal protein, the HA protein, and the NA protein as shown in Figs. 1, 2, and 3, respectively. The HA and NA proteins of the 1957 H2N2 pandemic strain were different from those of human strains that appeared prior to the 1957 strain. The internal protein of the 1957 strain was clearly different from that of the 1918 and 1933 strains. The origin of strains belonging to the 1940 – 2008 cluster was the HicKox/1940 H1N1 strain, but the internal protein of the 1957 H2N2 strain that evolved from the HicKox/1940 H1N1 strain was the first adaptation to the H2N2 subtype with new HA and NA proteins. The HA protein of the 1957 H2N2 pandemic strain clustered with that of avian H2 subtype viruses in Fig. 2 and the NA protein of that strain clustered with that of avian N2 subtype viruses in Fig. 3. However, the internal protein of the 1957 H2N2 strain fell into human lineage. On the other hand, the H5N1 strains isolated from human were closely related to the H5N1 strains isolated from birds in the HA, NA and internal proteins. This difference may explain why H5N1 viruses are less susceptible to human and less likely to spread among human.

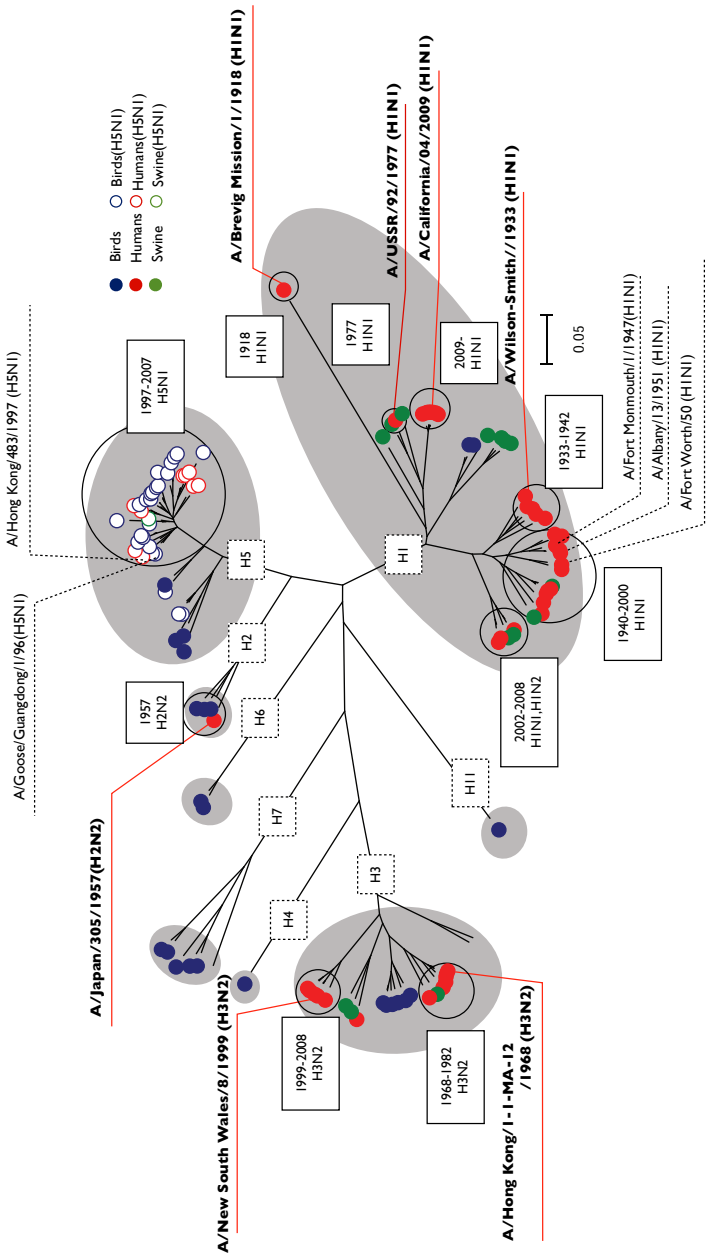

Fig. 2: Phylogenetic classification of the HA protein of influenza A viruses. Avian, human, and swine influenza A virus strains are represented by blue, red, and green circles. Influenza A virus strains that caused pandemics or severe epidemics with high mortality are indicated in boldface with red lines.

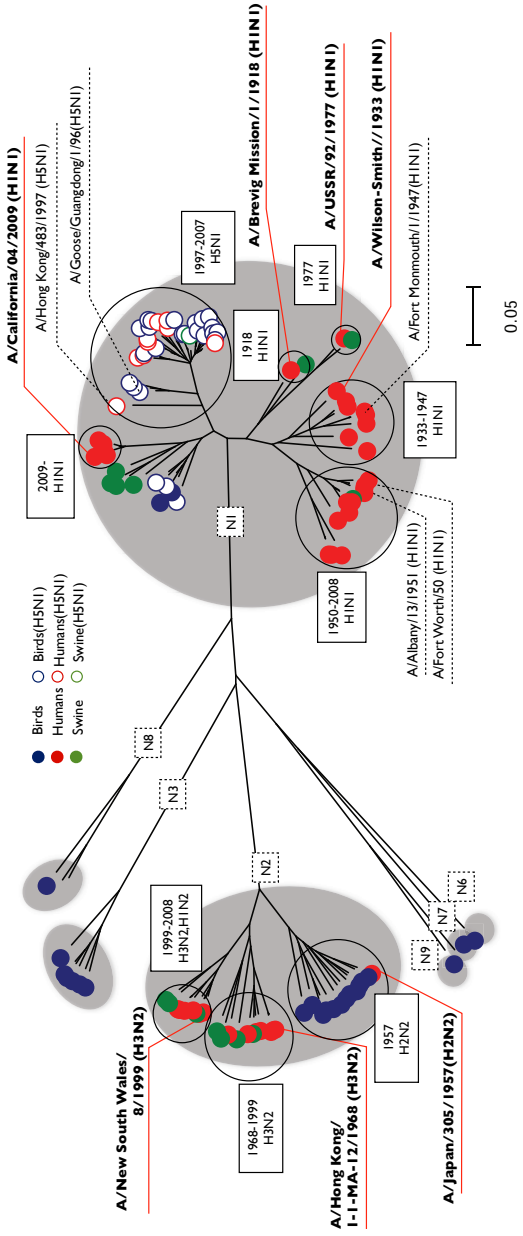

Fig. 3: Phylogenetic classification of the NA protein of influenza A viruses. Avian, human, and swine influenza A virus strains are represented by blue, red, and green circles. Influenza A virus strains that caused pandemics or severe epidemics with high mortality are indicated in boldface with red lines.

The HA protein of the 1968 H3N2 pandemic strain was different from that of human strains that had appeared prior to that strain, which was the origin of strains that belonged to the 1968 – 1982 cluster in the HA phylogenetic analysis. In addition to the HA protein, the NA protein of the 1968 H3N2 strain was the origin of strains belonging to the 1968 – 1999 cluster in the NA phylogenetic analysis. The internal protein of the 1968 H3N2 strain with new HA and NA proteins formed the 1968 – 1982 cluster in the phylogenetic analysis of internal protein, which was also the origin of strains belonging to the cluster. Other strains such as the 1977 H1N1 severe epidemic, the 1999 H3N2 severe epidemic, and the 2009 H1N1 pandemic appeared as the 1968 H3N2 pandemic strain.

In contrast to the severe epidemic or pandemic strains, each of the seasonal influenza strains was not the origin of strains belonging to a cluster for the HA protein and/or the NA protein. In addition, almost all the seasonal influenza strains were derived from a preexisting human strain for the internal protein.

The Fort Monmouth/1947 H1N1 strain that caused a global epidemic with low mortality was a descendant of the HicKox/1940 H1N1 strain for the HA protein, the 1933 H1N1 severe epidemic strain for the NA protein, and the 1940 H1N1 strain for the internal protein. Since the 1947 H1N1 strain was similar to the evolutional pattern of most seasonal influenza A viruses, we consider that the mortality of the 1947 H1N1 epidemic was relatively low like seasonal influenza epidemics. Similarly, the Fort Worth/1950 H1N1 and Albany/1951 H1N1 strains did not satisfy the conditions for the emergence of pandemics and severe epidemics with high mortality. It is not known whether the Fort Worth/1950 H1N1 and/or Albany/1951 H1N1 strains were derived from the severe 1951 epidemic. The 1951 epidemic caused an unusually large mortality in England, Wales and Canada. By contrast, that epidemic was not particularly severe in the United States [23, 24]. Although the 1950 and 1951 H1N1 strains in the NA protein were different from every cluster of human strains that appeared prior to both strains, those strains in both the HA and internal proteins were descendants of the HicKox/1940 H1N1 strain. Our result suggests that the Fort Worth/1950 H1N1 and Albany/1951 H1N1 strains were not trigger of the severe 1951 epidemic with high mortality.

4. Conclusions

We performed the classification of influenza A viruses using the difference between sequences measured by means of the rate of entropic chaos degree. Phylogenetic analysis of the internal protein (PB2, PB1, PA, NS, M1, M2, NS1, and NS2) revealed that influenza A viruses can be divided into three lineages (i.e., Avian lineage, Human lineage, and Swine lineage) and evolve independently in each lineage. In this study, we have come to the conclusion

that the internal protein has a significant impact on the ability for transmission among human. Although the HA protein of influenza A viruses is known to be responsible for the restriction of interspecies transmission, we found that the internal protein plays a key role in determining host species of influenza A viruses. Furthermore, our present results indicate that a pandemic strain or a severe epidemic strain emerges in a combination of new HA, new NA, and new internal proteins that are phylogenetically distinct from those of previous pandemic and severe epidemic strains.

Based on this study, we are convinced that entropic chaos degree describes the dynamics hidden in the evolution of the influenza A virus and can be a useful measure of understanding the classification and severity of an isolated strain of influenza A virus.

Bibliography

[1] W. P. Glezen, *Emerging Infections: Pandemic Influenza*, Epidemiol. Rev. **18**, 64 (1996).

[2] M. I. Nelson, C. Viboud, L. Simonsen, R. T. Bennett, S. B. Griesemer, K. S. George, J. Taylor, D. J. Spiro, N. A. Sengamalay, E. Ghedin, J. K. Taubenberger, and E. C. Holmes, *Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus Since 1918*, PLoS Pathog **4** (2), 1 (2008).

[3] D. M. Morens, J. K. Taubenberger, and A. S. Fauci, *The persistent legacy of the 1918 influenza virus*, N. Engl. J. Med. **361** (3), 225 (2009).

[4] S. M. Zimmer and D. S. Burke, *Historical perspective — Emergence of influenza A (H1N1) viruses*, N. Engl. J. Med. **361** (3), 279 (2009).

[5] B. Olsen, V. J. Munster, A. Wallensten, J. Waldenström, A. D. M. E. Osterhaus, and R. A. M. Fouchier, *Global patterns of influenza a virus in wild birds*, Science **312**, 384 (2006).

[6] S. Krauss, C. A. Obert, J. Franks, D. Walker, K. Jones, P. Seiler, L. Niles, S. P. Pryor, J. C. Obenauer, C. W. Naeve, L. Widjaja, R. J. Webby, and R. G. Webster, *Influenza in Migratory Birds and Evidence of Limited Intercontinental Virus Exchange*, PLoS Pathog **3** (11), 1684 (2007).

[7] R. A. M. Fouchier, V. Munster, A. Wallensten, T. M. Bestebroer, S. Herfst, D. Smith, G. F. Rimmelzwaan, B. Olsen, and A. D. M. E. Osterhaus, *Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls*, J. Virol. **79** (5), 2814 (2005).

[8] M. I. Nelson and E. C. Holmes, *The evolution of epidemic influenza*, Nature Reviews Genetics **8**, 196 (2007).

[9] V. G. Dugan, R. Chen, D. J. Spiro, N. Sengamalay, J. Zaborsky, E. Ghedin, J. Nolting, D. E. Swayne, J. A. Runstadler, G. M. Happ, D. A. Senne, R. Wang, R. D. Slemons, E. C. Holmes, and J. K. Taubenberger, *The Evolutionary Genetics and Emergence of Avian Influenza Viruses in Wild Birds*, PLoS Pathog **4** (5), 1 (2008).

[10] B. Pattnaik, A. K. Pateriya, R. Khandia, C. Tosh, S. Nagarajan, S. Gounalan, H. V. Murugkar, B. P. Shankar, N. Shrivastava, P. Behera, S. Bhagat, J. S. M. Peiris, and H. K. Pradhan, *Phylogenetic analysis revealed genetic similarity of the H5N1 avian influenza viruses isolated from HPAI outbreaks in chickens in Maharashtra, India with those isolated from swan in Italy and Iran in 2006*, Science **96** (1), 77 (2006).

[11] A. N. Abdel-Ghafar, T. Chotpitayasunondh, Z. Gao, F. G. Hayden, D. H. Nguyen, M. D. de Jong, A. Naghdaliyev, J. S. Peiris, N. Shindo, S. Soeroso, and T. M. Uyeki, *Update on avian influenza A (H5N1) virus infection in human*, N. Engl. J. Med. **358** (3), 261 (2008).

[12] G. J. D. Smith, J. Bahl, D. Vijaykrishna, J. Zhang, L. L. M. Poon, H. Chen, R. G. Webster, J. S. M. Peiris, and Y. Guan, *Dating the emergence of pandemic influenza viruses*, Proc. Natl. Acad. Sci. USA **106** (28), 11709 (2009).

[13] T. M. Tumpey and J. A. Belser, *Resurrected Pandemic Influenza Viruses*, Annual Review of Microbiology **63**, 79 (2009).

[14] J. K. Taubenberger and D. M. Morens, *1918 Influenza: the Mother of All Pandemics*, CDC EID **12** (1), 15 (2006).

[15] M. Ohya, *Information Dynamics and Its Applications to Optical Communication Processes*, Lect. Note Phys. **378**, 81 (1991).

[16] M. Ohya, *Complexities and their application to characterization of chaos*, Int. J. Theor. Phys. **37**, 495 (1998).

[17] K. Sato and M. Ohya, *Analysis of the disease course of HIV-1 by entropic chaos degree*, Amino Acids **20**, 155 (2001).

[18] M. Ohya, *Adaptive Dynamics and its Applications to Chaos and NPC Problem*, QP–PQ: Quantum Probability and White Noise Analysis, Quantum Bio-Informatics, 2007, pp. 181–216.

[19] K. Inoue, M. Ohya, and K. Sato, *Application of chaos degree to some dynamical systems*, Chaos, Solitons and Fractals **11**, 1377 (2000).

[20] T. Hara, K. Sato, and M. Ohya, *MTRAP: Pairwise sequence alignment algorithm by a new measure based on transition probability between two consecutive pairs of residues*, BMC Bioinformatics 2010, pp. 235–245.

[21] J. Felsenstein, PHYLIP Phylogeny Inference Package, http://evolution.genetics.washington.edu/phylip/doc/main.html.

[22] J. Felsenstein, Neighbor-Neighbor-Joining and UPGMA methods, http://evolution.genetics.washington.edu/phylip/doc/neighbor.html.

[23] C. Viboud, T. Tam, D. Fleming, A. Handel, M. A. Miller, and L. Simonsen, *Transmissibility and mortality impact of epidemic and pandemic influenza, with emphasis on the unusually deadly 1951 epidemic*, Vaccine **24** (44–46), 6701 (2006).

[24] C. Viboud, T. Tam, D. Fleming, M. A. Miller, and L. Simonsen, *1951 Influenza Epidemic, England and Wales, Canada, and the United States*, Emerg. Infect. Dis. **12** (4), 661 (2006).