Molecular Evolution of Hemagglutinin Genes of H1N1 Swine and Human Influenza A Viruses

Shigeo Sugita,¹ Yasuyuki Yoshioka,¹ Shigeyuki Itamura,¹ Yumi Kanegae,¹ Keiko Oguchi,² Takashi Gojobori,² Kuniaki Nerome,¹ and Akira Oya¹

1 National Institute of Health, 10-35, Kamiosaki, 2-chome, Shinagawa-ku, Tokyo 141

2 National Institute of Genetics, Mishima, 411, Japan

Summary. The hemagglutinin (HA) genes of influenza type A (H1N1) viruses isolated from swine were cloned into plasmid vectors and their nucleotide sequences were determined. A phylogenetic tree for the HA genes of swine and human influenza viruses was constructed by the neighbor-joining method. It showed that the divergence between swine and human HA genes might have occurred around 1905. The estimated rates of synonymous (silent) substitutions for swine and human influenza viruses were almost the same. For both viruses, the rate of synonymous substitution was much higher than that of nonsynonymous (amino acid altering) substitution. It is the case even for only the antigenic sites of the HA. This feature is consistent with the neutral theory of molecular evolution. The rate of nonsynonymous substitution for human influenza viruses was three times the rate for swine influenza viruses. In particular, nonsynonymous substitutions at antigenic sites occurred less frequently in swine than in humans. The difference in the rate of nonsynonymous substitution between swine and human influenza viruses can be explained by the different degrees of functional constraint operating on the amino acid sequence of the HA in both hosts.

Key words: Influenza virus $-$ Swine virus $-$ He mag glutinin gene $-$ Molecular evolution $-$ Neutral theory

Introduction

A number of human H1N1 influenza viruses (type A) isolated in the 1930s are related antigenically to swine influenza viruses, but the exact origin of the pandemic viruses is still unknown (Shope 1931; Murphy and Webster 1990). Analysis by oligonucleotide mapping, nucleotide sequencing, and detailed characterization of the hemagglutinin (HA) antigen with monoclonal antibodies suggests that human H1N1 viruses have evolved faster than swine H1N1 viruses (Nerome et al. 1985a,b; Raymond et al. 1986; Kida et al. 1987, 1988). Analysis of the evolutionary pathway for a number of human epidemic HIN1 viruses isolated between 1950 and 1983 indicated that these viruses contained both mainand side-stream mutations, which have led to many different branches (Raymond et al. 1986). In addition, antigenic characterization of HA has shown that the antigenic structure of the HA of swine influenza viruses had been well conserved since its first isolation in 1930 (Meier-Ewert et al. 1970; Kendal et al. 1977; Nerome et al. 1983; Hinshaw et al'. 1984; Sheerar et al. 1989).

In early 1976, some 500 military recruits in Fort Dix, New Jersey, were infected with swine (H1N1) influenza virus (Palese and Schulman 1976a; Goldfield et al. 1977). H1N1 viruses that were both antigenically and genetically similar to a classical strain of swine influenza virus was isolated from the recruits. In 1988, a pregnant woman 32 years old in the United States died after being infected with a

Offprint requests to: K. Nerome

swine influenza virus (Morbidity and Mortality Weekly Report 1988). These incidents show that swine influenza viruses can infect humans. Therefore, swine $(H1N1)$ influenza viruses are still of much concern. An understanding of the future epidemic and genetic behavior of these viruses is not possible without knowledge of the molecular evolution of field isolates. Swine influenza virus is a useful subject for the understanding and the prediction of the possible appearance of a new pandemic virus in humans.

With the aim of elucidating evolutionary relationships between human and swine influenza viruses, we constructed a phylogenetic tree for these viruses. In particular, we describe a pattern of nucleotide substitutions unique to swine and human influenza viruses.

Materials and Methods

Virus Strains andRNAs. Four strains of influenza viruses A/sw/ Iowa/15/30 (H1N1), A/sw/Illinois/63 (H1N1), A/sw/Hong Kong/ 1/74 (HIN1), and A/sw/Ehime/1/80 (HIN2) were propagated in fertile hen's eggs 11 days after being laid by intraallantoic inoculation. Virus particles in the allantoic fluid were harvested after incubation at 34"C for 2 days and were purified by a method described previously (Nerome et al. 1983). Viral RNA was extracted from the purified viruses by the hot-phenol method (Palese and Schulman 1976b).

Serological Test. Hemagglutination inhibition (HI) tests with antisera to reference strains were done as described by Nerome et al. (1983). Antisera were prepared by intravenous inoculation of rabbits with the purified viruses. Injections were repeated four times at weekly intervals.

Cloning and Nucleotide Sequences. Double-strand complementary DNA was synthesized from viral genomie RNA with use of a synthetic oligonucleotide primer. Complementary DNA specific for viral RNA was inserted into the plasmid pBR322 or pUC118. The nucleotide sequences of the HA genes of A/sw/ Iowa/15/30, A/sw/Illinois/63, A/sw/Hong Kong/1/74, and A/sw/ Ehime/l/80 viruses were determined by the dideoxy method (Mizusawa et al. 1986).

Evolutionary Analysis of HA Genes. For the sequences newly determined in this paper and sequences previously published for the HA genes of human and swine influenza viruses, the numbers of synonymous and nonsynonymous substitutions were estimated by the Nei-Gojobori method (Nei and Gojobori 1986). Using the numbers obtained, we constructed phylogenetic trees by the neighbor-joining (N-J) method (Saitou and Nei 1987). The evolutionary rates were computed by the least-squares method with use of the results from the phylogenetic trees.

Results

Sequence Homology at the Nucleotide and Amino Acid Levels

Nucleotide sequences of the HA genes of A/sw/Iowa/ 15/30 (IOWA30), A/sw/Illinois/63 (ILL63), A/sw/ Hong Kong/1/74 (HK74), and A/sw/Ehime/1/80 (EHM80) were determined (data not shown) and deposited in the DNA Data Bank of Japan (DDBJ). The amino acid sequences of the HA from four strains of swine influenza viruses were aligned with other HA sequences available (Fig. 1). Swine influenza viruses are less variable than human influenza viruses at potential antigenic sites, particularly at the Sa and Sb sites (Caton et al. 1982).

Unlike in the human influenza viruses, a signal for potential N-linked glycosylation was absent at the Sa and Sb sites on the surface of the HA in swine influenza viruses. In human influenza viruses, new potential glycosylation sites appeared at positions 127 (PR834, A/Puerto Rico/8/34; USSR77, A/USSR/90/77) and 160 (USSR77). Glycosylation at the antigenic sites of influenza viruses seems to be more important in humans than in swine for escaping from the immune system.

A comparison of the amino acid sequences we determined with those already published data showed that the amino acids at the receptor-binding site were well conserved among different subtypes, H1, H2, H3, H5, H7, and H10 (Feldmann et al. 1988; Weis et al. 1988). In the swine influenza viruses, residues 130-135 and 221-225, which were thought to be on the surface of the receptor-binding pocket, were substituted at only two positions. The proportion of amino acid changes on the surface of the receptor-binding pocket of human influenza viruses was higher; it was calculated to be 54.5% (= 6/11).

The pattern of amino acid substitutions in the HA gene was different between human and swine influenza viruses, which was supported by the serological reactivity of a number of field isolates (Fig. 2). Antisera to the classical swine strain, IOWA30, gave a variety of reaction patterns with swine influenza viruses isolated between 1930 and 1982 (NVD82; A/sw/Nevada/82), but antisera to an earlier human strain, PR834, failed to inhibit hemagglutination of all human influenza viruses isolated from 1947 to 1986.

Phylogenetic Analysis

To examine the evolutionary pathway between human and swine influenza viruses, a phylogenetic tree was *constructed* by the N-J method using a total of 26 nucleotide sequences of the HA gene (Fig. 3). The value given to each branch of the tree in the figure represents the number of nucleotide substitutions per synonymous site. The branching features of the HA of swine influenza virus isolates were quite different from those of human HA genes, although classical swine virus, IOWA30, has been taken to be a representative of swine and human

Fig. 1. Comparison of the amino acid sequences of the HAs between swine and human influenza viruses. For comparative analysis of amino acid substitutions, the HA gene of IOWA30 was used as the reference sequence. S.P. represents N-terminal signal sequences. Potential glycosylation sites are boxed. Antigenic regions characterized in the PR834 (Mt. Sinai) strain are marked as follows: a, Sa; b, Sb; c, Cb; 1, Ca₁; and 2, Ca₂ (Caton et al. 1982). Residues corresponding to receptor binding site are indicated by R (Weis et al. 1988). It is probable that a change at residues 153 and 155 would alter host specificity of the virus. These residues are indicated by S (Both et al. 1983; Kilbourne et al. 1988a,b).

Reduction in HI **titer**

Fig. 2. Comparison of the antibody response among field isolates of swine and human influenza viruses. Antisera to IOWA30 and PR834 were treated with receptor destroyed enzymes (RDE). Nonspecific HA in the sera was removed by addition of washed chicken red blood cells. The following strains were used in the HI tests: IOWA30, 197631 (A/sw/1976/31), ILL63, NJ76, EHM80, NVD82 (A/sw/Nevada/82), NWS33 (A/NWS/33), PR834, FM47 (A/FM/1/47), KJY52 (A/Kojiya/4/52), OMC53, (A/Ohmachi/l/53), SJK56 (A/Shinjuku/l/56), and YMG86 (AJ Yamagata/120/86).

influenza viruses from 1918 to 1930 (Nakajima et al. 1984).

The tree for the HA genes of human influenza viruses isolated between 1933 and 1983 showed much variety in terms of genealogy. Two human influenza viruses from the 1933 epidemic seemed to be on the independent branches. The tree also showed that there were multiple pathways of evolution even during the same period. Of the branches containing epidemic strains prevalent during the A prime era (1946-1956), the branch giving rise to A/Fort Warren/I/50 (FW50) was most closely related to USSR substrains.

As already reported, USSR77 seems to have reemerged in humans in 1977 after a 25-year absence (Nakajima et al. 1978; Hayashida et al. 1985; Buonagurio et al. 1986; Saitou and Nei 1986). As shown in the evolutionary tree, USSR77 and FW50 viruses seemed to be genetically very similar with each other. A/Brazil/11/78 (BRZ78), A/Texas/29/82 (TX2982), A/Dunedin/6/83 (DUN683), and A/Hong Kong/32/ 83 (HK83) appeared to be representatives of four evolutionary pathways arising in or after 1977. Vi-

* The rates of nucleotide substitutions per site per year were calculated from the phylogenetie tree presented in Fig. 3

^{\dagger} Estimates obtained in the present study. Accumulations of synonymous substitutions in swine and human are shown in Fig. 4

 \ddagger After the report of Hayashida et al. (1985)

ruses isolated from many parts of the world in 1983 showed multiple pathways of evolution.

To estimate the year of divergence and the rate of synonymous substitution, the numbers of nucleotide substitutions were calculated for all of the HA sequences of human and swine influenza viruses. The numbers of synonymous substitutions for human and swine influenza viruses from each putative origin were respectively plotted against the year of isolation. As shown in Fig. 4, linear regression analysis indicated that, with the exception of USSR substrains, the number of synonymous substitutions for swine and human influenza viruses was proportional to the year of isolation. The result obtained was consistent with the previous report that nucleotide substitutions in human influenza virus genes can be used as a good molecular clock (Hayashida et al. 1985). However, the USSR substrains were on a regression line different from other human strains.

Table 1 shows the rates of nucleotide substitutions for swine and human virus HA genes. The rates of synonymous substitution in swine and human influenza viruses were about the same. The results are consistent with the value obtained for human H3 HA gene reported by Hayashida et al. (1985). The rates of substitution at the third position ofa codon were 0.00565 and 0.00717 for swine and humans, respectively. Because most of the nucleotide substitutions at the third position of a codon do not cause amino acid changes, the substitution rate at the third position of a condon for the HA gene of the swine influenza viruses was about the same as that of the human influenza viruses. The estimation of the synonymous substitution rate allowed us to calculate the time of divergence of the root leading to the two different genealogies for swine and human influenza viruses; it was around 1905.

The rates of nucleotide substitution at the first and second positions of a codon for the genes of the swine influenza viruses were estimated to he 0.0006

the phylogenetic tree: IOWA30, ILL63, HK74, NJ76, EHM80, WSN33, PR834, FW50,
ENG51 (A/England/1/51), FLW52 (A/Fort Leonard Wood/1/52), QSL54 (A/Queensland/ 34/54), DEN57 (A/Denver/1/57), USSR77, LACK78 (A/Lackland/3/78), BRZ78,

root between the swine and human virus groups was inferred from the relationship between the rate and the year of isolation.

20

Fig. 4. Linearity of the number of synonymous substitutions with time for the HA gene of swine and human viruses. Nucleotides 4-1062 in HA were used for linear regression analysis. The branch length was estimated from the trees and was plotted against the year of isolation.

and 0.00017 per site per year. These values were about one-third and one-fifth, respectively, of those for human influenza viruses. These results are consistent with our observation that the rate of nonsynonymous substitution of swine influenza viruses was about three times lower than that of human influenza viruses. For both human and swine influenza viruses, the rate of synonymous substitution was much higher than that of nonsynonymous substitution (Table 1). This feature is consistent with the neutral theory of molecular evolution (Kimura 1968, 1983), contrary to the general belief that positive selection dominates evolutionary changes in human influenza A viruses (Air et al. 1990).

Discussion

A series of H₁ H_A genes of swine and human influenza viruses was analyzed in detail from both the

Fig. 5. Relationship between the number of nonsynonymous substitutions and that of synonymous substitutions at antigenic sites (also see Fig. 1). Closed squares represent swine viruses and open circles represent human viruses. The mean values of N/S for human and swine influenza viruses were 0.927 and 0.296, respectively. The N/S value between WSN33 and IOWA30 is given with a closed triangle.

evolutionary and the phylogenetic points of view. Raymond et al. (1986) classified the evolutionary pathways of the human influenza viruses into mainand side-stream mutations by phylogenetic tree. However, we did not follow their classifications because our tree showed that there was no main stream in evolutionary pathways for human influenza vi-Fuses.

Many of nucleotide sequences have been published by several authors (Raymond et al. 1986; Kida et al. 1987, 1988). They showed that the number of nucleotide substitutions for the HA genes of animal influenza viruses was lower than for the HA genes of human influenza viruses. However, our phylogenetic tree indicated that the rates of synonymous substitutions for swine and human influenza viruses were about the same.

Influenza A virus has the highest rate of all the animal viruses examined so far. We estimated the rate of synonymous substitution to be $0.0080 - 0.011$ per site per year, which was 106 times the rate for mammalian genes (Hayashida et al. 1985). This may be due to a higher error rate of the RNA polymerase, a lack of a repair mechanism for the RNA genome, and a shorter generation time for the influenza A viruses.

For the swine influenza viruses, the rate of nonsynonymous substitution of the HA gene does not seem to be proportionally related to that of synonymous substitution at antigenic sites (Fig. 5). This is different from the evolutionary pattern of the human viral genes, which had a proportional relationship between rates of nonsynonymous and synonymous substitution at antigenic sites. This is explained by the fact that for swine influenza viruses, the rate of synonymous substitution is extremely small. However, it was not clear whether the slower rate of nonsynonymous substitution was due to absence of antigenic selection for swine influenza viruses.

To examine the mechanism by which the antigenic sites in HA of swine influenza viruses are conserved, the ratio of the number of nonsynonymous substitutions to that of synonymous substitutions (N/S) was calculated. We assumed that when the N/S ratio was greater than 1, the amino acid change at the antigenic site of the HA would occur through the mechanism of positive selection. However, the N/S ratio at antigenic sites in the swine virus HA was always I or less. Therefore, the stronger conservation of amino acids at the antigenic sites might be due to the more severe functional constraint against the HA of swine influenza viruses.

For the HAs of human influenza viruses, some of the *N/S* ratios at the antigenic sites were more than 1. In particular, 10 of the 14 values between WSN33 and PR834 (called the A0 group) and other human influenza viruses (the A1 group) were more than i, Thus, antigenic drift between the A0 and A 1 groups might have occurred by positive selection of the antibodies. However, this may be exceptional because the N/S values for most of the substitutions at antigenic sites were less than 1 and they are explained by neutral evolution (Kimura 1968, 1983). Therefore, we propose that positive selection may not be operating very strongly on antigenic sites for the antigenic drift of human and animal viruses.

One of the aims of this study was to conduct a phylogenetic analysis of swine and human influenza viruses in relation to the origin of these viruses. We found that the H1N1 viruses from these two hosts had diverged from a common ancestor in around 1905, long before the 1918-1919 outbreak of influenza viruses in humans. Many variants of swine influenza viruses continue to circulate in many parts of the world. Our study showed that their stable conservation in the swine population was closely related to an unusual fashion of evolution explained above. The study of molecular evolution in relation to functional constraint of the HA gives clues to help in the understanding of the mechanisms by which influenza A viruses circulate in a range of animal species including humans.

Acknowledgment. We thank Dr. Nancy J. Cox, Centers for Disease Control, for her helpful comments on the manuscript.

References

- Air GM, Gibbs, AJ, Laver WG, Webster RG (1990) Evolutionary changes in influenza B are not primarily governed by antibody selection. Proc Natl Acad Sci USA 87:3884-3888
- Both GW, Shi CH, Kilbourne ED (1983) Hemagglutinin of swine influenza virus: a single amino acid change pleiotropically affects viral antigenicity and replication. Proc Natl Aead Sci USA 80:6996-7000
- Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM (1986) Evolution of human influenza A virus over 50

years: rapid, uniform rate of change in NS gene. Science 232: 980-982

- Caton A, Brownlee GG, Yewdell JW, Gerhard W (1982) The antigenic structure of influenza virus A/PR/8/34 hemagglutinin (HI subtype). Cell 31:417--427
- Concannon P, Cummings IW, Salser WA (1984) Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene. J Virol 49:276-278
- Fang R, Jou WM, Huylebroeck D, Devos R, Fiers W (1981) Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1986 influenza hemagglutinin. Cell 25:315-323
- Feldmann H, Kretzschmar E, Klingeborn B, Rott R, Klenk H-D, Garten W (1988) The structure of serotype H10 hemagglutinin of influenza A virus: comparison of an apathogenic avian and a mammalian strain pathogenic for mink. Virology 165: 428-437
- Goldfield M, Bartley JD, Pizzuti W, Black HC, Altman R, Halperin WE (1977) Influenza in New Jersey in 1976: isolation of influenza A/New Jersey/76 virus at Fort Dix. J Infect Dis 136(suppl):S347-S355
- Hayashida H, Toh H, Kikuno R, Miyata T (1985) Evolution of influenza virus genes. Mol Biol Evol 2:289-303
- Jabbar MA, Sivasubramanian N, Nayak DP (1985) Influenza viral (A/WSN/33) hemagglutinin is expressed and glycosylated in the yeast *Saccharomyces cerevisiae.* Proc Natl Acad Sci USA 82:2019-2023
- Hinshaw VS, Alexander DJ, Aymard M, Bachmann PA, Easterday BC, Hannoun C, Kida H, Lipkind M, MacKenzie JS, Nerome K, Schild GC, Scholtissek C, Senne DA, Shortridge KF, Skehel JJ, Webster RG (1984) Antigenic comparisons of swine-influenza-like H1N1 isolates from pigs, birds and humans: an international collaborative study. Bull WHO 62: 871-878
- Kendal AP, Noble GR, Dowdle WR (1977) Swine influenza viruses isolated in 1976 from man and pig contain two coexisting subpopulations with antigenically distinguishable hemagglutinins. Virology 82:111-121
- Kida H, Kawaoka Y, Naeve CW, Webster RG (1987) Antigenic and genetic conservation of H3 influenza virus in wild ducks. Virology 159:109-119
- Kida H, Shortridge KF, Webster RG (1988) Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China. Virology 162:160-166
- Kilbourne ED, Easterday BC, McGregor S (1988a) Evolution to predominance of swine influenza virus hemagglutinin mutants of predictable phenotype during single infections of the natural host. Proc Natl Acad Sei USA 85:8098-8101
- Kilbourne ED, Taylor AH, Whitaker CW, Sahai RR, Caton AJ (1988b) Hemagglutinin polymorphism as the basis for lowand high-yield phenotypes of swine influenza virus. Proc Natl Acad Sci USA 85:7782-7785
- Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624-626
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- Meier-Ewert H, Gibbs AJ, Dimmock NJ (1970) Studies on antigenic variations of haemagglutinin and neuraminidase of swine influenza virus isolates. J Gen Virol 6:409-419
- Mizusawa S, Nishimura S, Seela F (1986) Improvement of the chain termination method of DNA sequencing by use of
- deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res 14:1319-1324
- Morbidity and Mortality Weekly Report (1988) Human infection with swine influenza virus--Wisconsin. In: Epidemiologic notes and reports, vol 37, p 661
- MurphyBR, Webster RG (1985) Orthomyxoviruses. In: Fields BN (ed) Virology. Raven Press, New York, pp 1091-1152
- Nakjima K, Desselberger D, Palese P (1978) Recent human influenza A (HIN1) viruses are closely related genetically to strains isolated in 1950. Nature 274:334-339
- Nakajima K, Nobusawa E, Nakajima S (1984) Genetic relatedness between A/swine/Iowa/15/30(H1N1) and human influenza viruses. Virology 139:194-198
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426
- Nerome K, Sakamoto S, Yano N, Yamamoto T, Kobayashi S, Webster RG, Oya A (1983) Antigenic characteristics and genome composition of a naturally occurring recombinant influenza virus isolated from a pig in Japan. J Gen Virol 64: 2611-2620
- Nerome K. Yoshioka Y, Sakamoto S, Yasuhara H, Oya A (1985a) Characteristics of a swine recombinant influenza virus isolated in 1980: recombination between swine and the earliest Hong Kong (H3N2) viruses. Vaccine 3(suppl):267-273
- Nerome K, Yoshioka Y, Sakamoto S, Yasuhara H, Oya A (1985b) Characterization ofa 1980-swine recombinant influenza virus possessing H1 hemagglutinin and N2 neuraminidase similar to that of the earliest Hong Kong (H3N2) virus. Arch Virol 86:197-211
- Palese P, Schulman JL (1976a) RNA pattern of'swine' influenza virus isolated from man is similar to those of other swine influenza viruses. Nature 263:528-530
- Palese P, Schulman JL (1976b) Differences in RNA patterns of influenza A viruses. J Virol 17:876-884
- Raymond FL, Caton AJ, Cox NJ, Kendal AP, Brownlee GG (1986) The antigenicity and evolution of influenza H1 haemagglutinin, from 1950-1957 and 1977-1983: two pathways from one gene. Virology 148:275-287
- Saitou N, Nei M (1986) Polymorphism and evolution of influenza A virus genes. Mol Biol Evol 3:57-74
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425
- Sheerar MG, Easterday BC, Hinshaw VS (1989) Antigenic conservation of H1NI swine influenza viruses. J Gen Virol 70: 3297-3303
- Shope RE (1931) Swine influenza: III. Filtration experiments and etiology. J Exp Med 54:373-380
- Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature 333:426-431
- Winter G, Fields S, Brownlee GG (1981) Nucleotide sequence of the haemagglutinin gene of a human influenza virus H 1 subtype. Nature *292:72-75*

Received June 28, 1990/Revised July 30, 1990