Polymorphism and Evolution of Influenza A Virus Genes[1]

Naruya Saitou and Masatoshi Nei
Center for Demographic and Population Genetics, University of Texas at Houston

The nucleotide sequences of four genes of the influenza A virus (nonstructural protein, matrix protein, and a few subtypes of hemagglutinin and neuraminidase) are compiled for a large number of strains isolated from various locations and years, and the evolutionary relationship of the sequences is investigated. It is shown that all of these genes or subtypes are highly polymorphic and that the polymorphic sequences (alleles) are subject to rapid turnover in the population, their average age being much less than that of higher organisms. Phylogenetic analysis suggests that most polymorphic sequences within a subtype or a gene appeared during the last 80 years and that the divergence among the subtypes of hemagglutinin genes might have occurred during the last 300 years. The high degree of polymorphism in this RNA virus is caused by an extremely high rate of mutation, estimated to be 0.01/nucleotide site/year. Despite the high rate of mutation, most influenza virus genes are apparently subject to purifying selection, and the rate of nucleotide substitution is substantially lower than the mutation rate. There is considerable variation in the substitution rate among different genes, and the rate seems to be lower in nonhuman viral strains than in human strains. This difference might be responsible for the so-called freezing effect in some viral strains.

Introduction

The influenza virus is a single-stranded RNA virus and is classified into three types, A, B, and C, according to immunological differences (Webster et al. 1982). The type A virus is the major cause of influenza epidemics not only in humans but in other mammals and birds. As do other RNA viruses, this virus has a high mutation rate, the mutation rate per year being approximately one million times higher than that of DNA organisms (Air 1981; Holland et al. 1982). Because of this high mutation rate, the influenza A virus becomes progressively resistant to antibodies made against older viruses. The high mutation rate is also responsible for the high degree of polymorphism observed in the genes in this organism (Nei 1983).

In recent years, the nucleotide sequences of several influenza virus genes have been determined for various strains, including those that have been kept in refrigerators for many years. These data provide a unique opportunity for studying the evolutionary history of polymorphic genes as well as for estimating the rate of nucleotide substitution in evolution.

Previously, Krystal et al. (1983) and Martinez et al. (1983) estimated the rate of nucleotide substitution for some of the genes of this virus, but their estimates are not very reliable because they simply compared genes from two strains isolated in different years. Nei (1983) also estimated the substitution rate by fitting a regression equation

1. Key words: influenza A virus, phylogenetic tree, polymorphism, mutation rate, rate of nucleotide substitution.

Address for correspondence and reprints: Dr. Masatoshi Nei, Center for Demographic and Population Genetics, University of Texas at Houston, P.O. Box 20334, Houston, Texas 77225.

Mol. Biol. Evol. 3(1):57–74. 1986.

0737-4038/86/0301-3101$02.00

to data from various strains isolated in different years. In the presence of polymorphism, however, his method is expected to give an overestimate. A better method of estimating the rate of nucleotide substitution for this case is first to examine the evolutionary relationship of genes obtained from different strains and then use only strains that are closely related by descent.

The main purpose of this paper is to conduct phylogenetic analyses of nucleotide sequences obtained from various strains and to estimate the rate of nucleotide substitution by using the method mentioned above.

Nucleotide Sequences Used

The genome of the influenza A virus consists of eight RNA segments coding for 10 different proteins (e.g., Lamb 1983). In the present study, we used four different genes, i.e., the hemagglutinin gene on segment 4, the neuraminidase gene on segment 6, the matrix protein 1 gene on segment 7, and the nonstructural protein 1 gene on segment 8. The first two genes are known to be responsible for the antigenic variation of this virus. We used only the coding regions of these genes, excluding the initiation codon. In the regions studied, there were no deletions or insertions. We used 46 strains in this study, and they are presented in table 1. Each strain is designated by the abbreviation of the location and the year in which it was isolated (see the legend to table 1).

The hemagglutinin gene is highly variable and can be classified into 13 subtypes (*H1–H13*) according to the immunological differences in hemagglutinin (Webster et al. 1982). The average nucleotide difference among these subtypes is 51% per nucleotide site (Nei 1983). We have therefore treated these subtypes separately, as though they were different genes. Sequence data useful for our analysis were available only for four subtypes, i.e., *H1, H2, H3,* and *H11.* The viral strains used for these four subtypes are given in table 1. The hemagglutinin gene is composed of three functional regions, i.e., signal peptide (*SP*), hemagglutinin 1 (*HA1*), and hemagglutinin 2 (*HA2*), and sequence data for these regions were combined unless otherwise mentioned. The nucleotide sequences available were not always complete, and the number of nucleotides used are presented in table 1.

The neuraminidase gene is also highly variable and can be divided into nine subtypes (*N1–N9*) (Webster et al. 1982). Two of them (*N1* and *N2*) were used here. Since sequence data for the *N1* subtype were available only for the first 168 bases in most strains, our analysis was restricted to this region. There were complete nucleotide sequences (1,404 bases) available for five *N2* subtype strains. The coding region for the matrix protein 1 gene (*MX1*) overlaps with that for the matrix protein 2 gene (*MX2*) on segment 7, and, similarly, the coding region for the nonstructural protein genes 1 (*NS1*) and 2 (*NS2*) overlap with each other. Therefore, we used only the genes for MX1 and NS1. For gene *NS1,* there were two sets of data; partial sequences and complete sequences (see table 1). In the construction of phylogenetic trees, all sequence data were used, excluding unshared nucleotides. In the estimation of the rate of nucleotide substitution, however, only complete sequences were used.

Phylogenetic Trees

Phylogenetic trees were constructed by using the maximum parsimony method (see Fitch 1977). The trees obtained for the four genes are presented in figures 1 and

Table 1
Influenza A Virus Gene Sequences Used in the Present Study

Genes and Subtypes[a]	Strains[b,c]
Hemagglutinin[d]	
H1 subtype: *SP*(48), *HA1*(228)	WIS/30(S), PR/34, NWS/33, BEL/42, FW/50, LO/57, NJ/76 [1]; WSN/33 [2]; USSR/77 [3]
H2 subtype: *SP*(42), *HA1*(207)	RI/57, TOK/67, NED/68, BER/68, GDR/72(D), ALB/77(D), ALB77(P) [4]
H3 subtype: *SP*(45), *HA1*(983), *HA2*(656)[e]	UKR/63(D) [5]; NT/68 [6, 7]; ENG/69, QU/70 [7]; MEM/71 [8]; MEM/72 [9]; HK/71, ENG/72, PC/73, VIC/76, AC/76, TX/77, BA2/79 [10]: VIC/75 [11]; ENG/77 [12]; BA1/79 [13]
H11 subtype: *SP*(45), *HA1*(201)	ENG/56(D), UKR/60(D), AU/75(T), MEM/76(D), NY/78(D) [14]
Neuraminidase:	
N1 subtype (168)	WIS/30(S), PR/34, BEL/42, FW/50, LO/57, NJ/76, ON/77(D), USSR/77, MEM/78 [15]; WSN/33 [16]
N2 subtype (1404)	RI/57 [17]; NT/68 [18]; UD/72 [19]; VIC/75 [20]; BA1/79 [21]
Matrix protein 1 (210)	PR/34 [22]; FPV/34(F) [23]; FW/50, LO/57, RI/57, CG/77 [4]; UD/72 [24]; BA1/79 [25]
Nonstructural protein 1:	
Partial sequences (192)	RI/57, CG/77, AU/78(B) [4]
Complete sequences (687)[f]	PR/34 [26]; FPV/34(F) [27]; FM/47, FW/50, USSR/77 [28]; UD/72 [29]; ALB/76 [30]; AL/77 [31]

[a] Numbers in parentheses are the number of nucleotides.

[b] Numbers in brackets are references: 1 = Air et al. 1981; 2 = Hiti et al. 1981; 3 = Concannon et al. 1984; 4 = Air and Hall 1981; 5 = Fang et al. 1981; 6 = Both and Sleigh 1980; 7 = Sleigh et al. 1981; 8 = Newton et al. 1983; 9 = Sleigh et al. 1980; 10 = Both et al. 1983; 11 = Min Jou et al. 1980; 12 = Hauptman et al. 1983; 13 = Both and Sleigh 1981; 14 = Air 1981; 15 = Blok and Air 1982; 16 = Hiti and Nayak 1982; 17 = Elleman et al. 1982; 18 = Bentley and Brownlee 1982; 19 = Markoff and Lai 1982; 20 = Van Rompuy et al. 1982; 21 = Martinez et al. 1983; 22 = Allen et al. 1980; 23 = McCauley et al. 1982; 24 = Lamb and Lai 1981; 25 = Ortin et al. 1983; 26 = Baez et al. 1980; 27 = Porter et al. 1980; 28 = Krystal et al. 1983; 29 = Lamb and Lai 1980; 30 = Baez et al. 1981; 31 = Buonagurio et al. 1984.

[c] The letter in parentheses denotes the organism from which the virus strain was isolated. B = black duck, D = duck, P = pintail (these three belong to Anatinae); T = tern (Sterninae); F = fowl (including chicken and duck); and S = swine (*Sus scrofa*). Strains without parentheses are from humans.

[d] *SP* = signal peptide; *HA1* = hemagglutinin 1; and *HA2* = hemagglutinin 2.

[e] Ref. 13 does not have *SP* sequences, and Refs. 7 and 10 have only *HA1* sequences, except for *HA2* sequence of the MEM/72 strain. For NT/68, *SP* and *HA2* are from ref. 6, and *HA1* is from ref. 7.

[f] Since the length of open reading frame for three strains (PR/34, FPV/34, and ALB/76) is shorter than that for others, seven unshared codons are not included.

2. When two or more trees with nearly the same total number of substitutions (one or two differences) were obtained for the same set of data, a concensus tree with collapsed node denoted by C was produced (e.g., fig. 1*B*). The root of a tree was located by using an outside strain. For hemagglutinin subtypes (fig. 1*A*–1*D*), Webster et al.'s (1982) dendrogram was used to identify outside strains. In the case of neuraminidase subtypes (fig. 2*A* and 2*B*), there were no such dendrograms available, but for the *N1* subtype the root could be located by using information on the *H1* subtype, because most of the strains studied were the same for the *H1* and *N1* subtypes. For the *N2* subtype, the oldest strain (RI/57) was assumed to be the ancestor (root). For matrix protein 1, FPV/34(F) was used as the outside strain, and for nonstructural protein 1, ALB/76(D) was assumed to be the outside strain. Both strains were isolated from

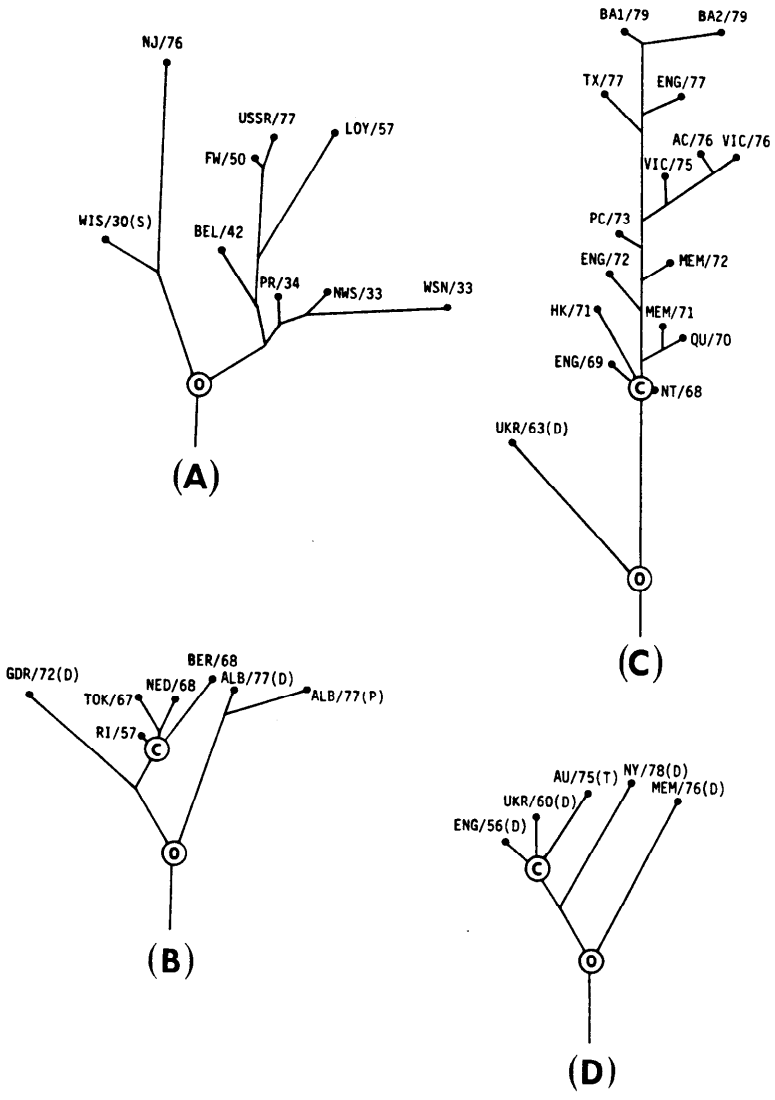

FIG. 1.—Phylogenetic trees for the *H1, H2, H3,* and *H11* subtypes of the hemagglutinin gene reconstructed by the maximum parsimony method. An unresolved node of a consensus tree is denoted by a circle enclosing a C and the root by a circle enclosing an O. For the determination of roots, see text. The letter in parentheses denotes the initial of the organism from which the strain was isolated. Strains without this initial are from humans. For the names of organisms, see table 1. Branch lengths are approximately proportional to the number of nucleotide substitutions. (*A*) = the *H1* subtype (both hemagglutinin 1 [*HA1*] and signal peptide [*SP*] used); (*B*) = the *H2* subtype (*HA1* and *SP* used); (*C*) = the *H3* subtype (only *HA1* used); and (*D*) = the *H11* subtype (*HA1* and *SP* used).

birds and were quite different from other human strains. Since the number of nucleotide substitutions for each branch cannot be estimated uniquely by the maximum parsimony method, only approximate branch lengths are given in figures 1 and 2.

The phylogenetic trees in figures 1 and 2 are quite different from ordinary trees

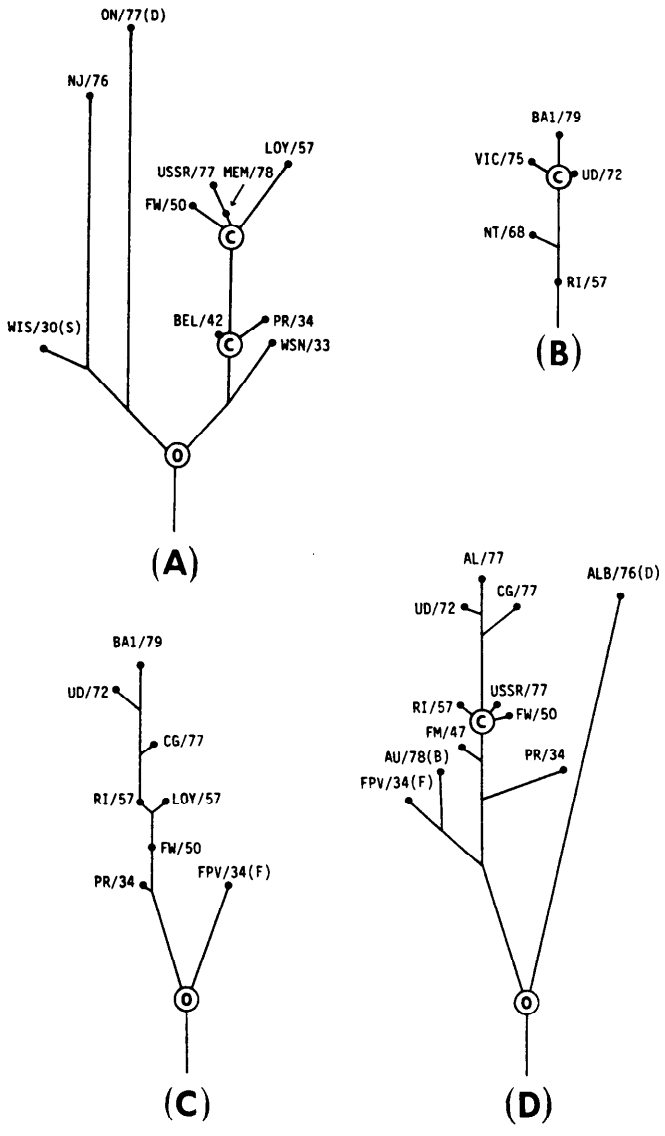

FIG. 2.—Phylogenetic trees for the *N1* and *N2* subtypes of the neuraminidase gene, for the matrix protein 1 gene and for the nonstructural protein 1 gene. (*A*) = the *N1* subtype; (*B*) = the *N2* subtype; (*C*) = matrix protein 1; and (*D*) = nonstructural protein 1. See fig. 1 for explanation of symbols.

because the strains used here were isolated in different years and mutations accumulated rapidly. The differences in evolutionary pattern among the phylogenetic trees in figures 1 and 2 largely depend on whether there was polymorphism in the past. If a gene were always monomorphic, one would expect all strains isolated in different years to be located on a line without branching, as in the case of FW/50 and RI/57 in figure 2(*C*) (matrix protein 1). On the other hand, if a gene is highly polymorphic and polymorphic alleles stay in the population for a long time, we would expect the type of trees represented by figures 1*A* and 1*D* (hemagglutinin).

Figures 1(*A*), 1(*B*), 1(*C*), and (*D*) show the phylogenetic trees for the *H1, H2, H3,* and *H11* subtypes of the hemagglutinin gene, respectively. A substantial amount of polymorphism exists in all subtypes, and, in general, mutations accumulate more or less linearly with time. However, figure 1(*A*) shows one anomaly. That is, strain USSR/77 did not accumulate mutations very much after it separated from FW/50. This anomaly has been previously noted by Nakajima et al. (1978) and will be discussed later. Subtypes *H4–H13* have been found only in nonhuman species, but the other three subtypes infect humans as well as other organisms. Except for NJ/76 of the *H1* subtype, strains of subtypes *H1, H2,* and *H3* isolated from humans are monophyletic and form one cluster (fig. 1*A*). NJ/76 is known to be close to swine strains (Blok and Air 1982), though it was isolated from humans, and our result, showing this strain clustered with a swine strain (WIS/30), confirms Blok and Air's earlier conclusion. The recent *H3* subtype strains in humans (Hong Kong flu) are all derived from NT/68, which in turn shares a common ancestor with UKR/63 isolated from duck. A main trunk is noticed in the phylogenetic tree, suggesting that in each year there was a dominant strain. The topology of the tree of *H3* subtype is similar to that of Both et al. (1983) if we compare the same set of 12 strains. The *H11* subtype infects only nonhuman species and seems to be highly polymorphic.

Figures 2(*A*) and 2(*B*) present the phylogenetic trees for the *N1* and *N2* subtypes of the neuraminidase gene, respectively. The tree for the *N1* subtype shares seven strains with that for the hemagglutinin *H1* subtype (fig. 1*A*). This is because the subtypes that were dominant in humans until 1957 are *H1* for the hemagglutinin gene and *N1* for the neuraminidase gene. Consequently, the topologies of these two trees are similar. In the tree of the *N2* subtype (fig. 2*B*), a trunk is identified, as in the case of the *H3* subtype of the hemagglutinin gene.

Figures 2(*C*) and 2(*D*) represent phylogenetic trees for the matrix protein 1 (*MX1*) and nonstructural protein 1 (*NS1*) genes, respectively. The tree for *NS1* is similar to that of Buonagurio et al. (1984) if we consider the same set of strains. The trees for *MX1* and *NS1* share six strains, and the topological relationships of these six strains are identical, if the branch linking strains FW/50 and RI/57 is eliminated. This similarity of topology suggests that these two sequences have evolved together, although they are located on two different RNA segments. Note that the strains used for constructing a tree for the matrix protein 1 and nonstructural protein 1 do not have the same subtypes of the hemagglutinin and neuraminidase genes. For example, in figure 2(*C*) the subtypes of the hemagglutinin and neuraminidase genes for strains PR/34, RI/57, and BA1/79 are *H1N1, H2N2,* and *H3N2,* respectively. This is because the genes of influenza A virus are segmented and occasionally reassorted (Webster et al. 1982).

Pattern of Accumulation of Nucleotide Substitutions

We studied the pattern of nucleotide substitution using the main evolutionary lines identified by the above phylogenetic analysis. The oldest strains, which were at the root (RI/57 for the *N2* subtype) or near the root (PR/34 for *H1, N1, MX1,* and *NS1;* RI/57 for *H2;* and NT/68 for *H3*), were treated as the ancestral strains, and strains that diverged earlier than the appearance of the ancestral strains were excluded from the analysis to minimize the effect of polymorphism. Most of the strains excluded were those isolated from nonhuman vertebrates, and the strains used for the analysis

were all from humans. The *H11* subtype was not included in the analysis, since no main evolutionary line was identified. The number of nucleotide substitutions per site between two strains was estimated by means of Jukes and Cantor's (1969) formula.

Figure 3 presents the pattern of accumulation of nucleotide substitutions for the genes or subtypes (*H1, N1, MX1,* and *NS1*) in which the ancestral strain was identified as PR/34. The results for the other subtypes (*H2, H3,* and *N2*) are given in figures 4 and 5. In the *H1* subtype, data for *SP* and *HA1* were combined, since they were similar. There are two sets (partial and complete sequences) of data for *NS1,* and the results from the comparison of complete sequences (687 bases) are shown, the data for strains RI/57, CG/77, and AU/78 being excluded.

The accumulation of nucleotide substitutions is approximately linear for all genes and subtypes examined. Particularly, in the *H3* subtype of the *HA1* gene, where the largest number of strains is used, the linearity is quite satisfactory (fig. 5). However, there are exceptional strains that do not follow the pattern of linear accumulation. They are USSR/77 and its close relative MEM/78 (marked by open circles in fig. 3). The *H1, N1,* and *NS1* genes in these strains show a small number of nucleotide substitutions after they branched off about 1950. A slowdown of nucleotide substitution

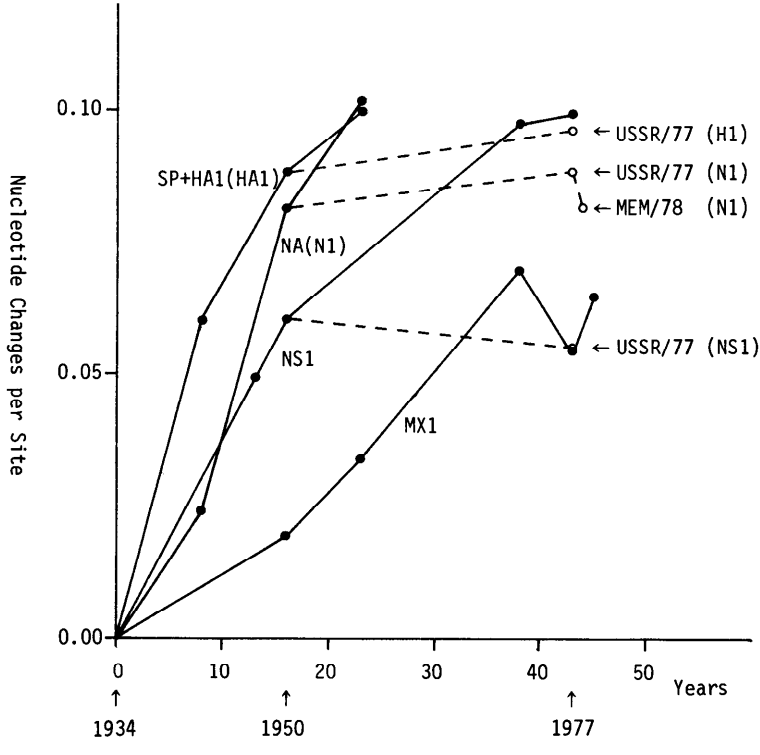

FIG. 3.—Accumulation of nucleotide substitutions in influenza A virus genes. Signal peptide (*SP*) and hemagglutinin 1 (*HA1*) of the *H1* subtype, neuraminidase (*NA*) of the *N1* subtype, nonstructural protein 1 (*NS1*), and matrix protein 1 (*MX1*) are presented. The year of isolation of the ancestral strains is 1934 for all four genes. Two strains (USSR/77 and MEM/78) that were excluded from the regression analysis are denoted by open circles, and they are connected by dashed lines with strains isolated in 1950. All nucleotides of codons were used.

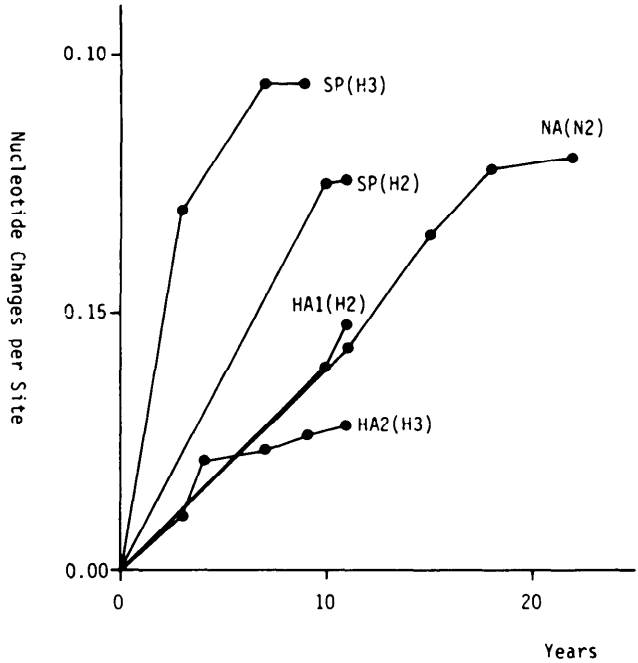

FIG. 4.—Accumulation of nucleotide substitutions in influenza A virus genes. *SP* and *HA1* of the *H2* subtype, *SP* and *HA2* of the *H3* subtype, and *NA* of the *N2* subtype are presented. The years of isolation for the ancestral strains are 1957 for the *H2* and *N2* subtypes and 1968 for the *H3* subtype.

is also observed for the *MX1* gene of strain CG/77, though it could arise from stochastic errors. All genes of the USSR/77 strain have been shown to be quite similar to those strains circulating about 1950 (Nakajima et al. 1978), and the phylogenetic trees for *H1, N1,* and *NS1* confirm this finding (see Discussion). Since USSR/77 and MEM/78 are clearly abnormal, they were excluded from the following analysis.

Rate of Nucleotide Substitution

We applied two types of regression analyses to estimate the rate of nucleotide substitution. One was the regression through the origin as used by Nei (1983). This method seems to be suitable for the *H2* and *H3* subtypes of the hemagglutinin gene, the *N2* subtype of the neuraminidase gene, and the *MX1* gene, where the ancestral strains of these genes are at the root or quite close to the root. On the other hand, when the ancestral strain branched off the root and if the branch length is not negligibly small as in the case of the *NS1* gene from strain PR/34, this method may give an overestimate of the substitution rate. Therefore, the usual regression method, which would alleviate this problem, was also used for the *H1, N1,* and *NS1* genes. The regression coefficients for the former and the latter methods are given by

$$b_I = \sum x_i y_i / \sum x_i^2$$

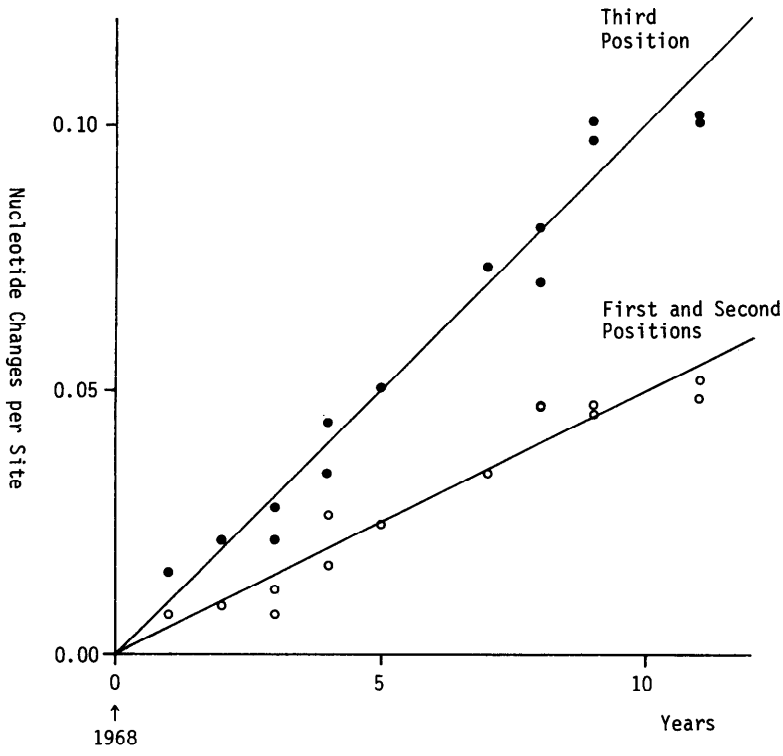

FIG. 5.—Accumulation of nucleotide substitutions in the *H3* subtype of the hemagglutinin 1 gene of the influenza A virus genes. The patterns of the accumulation of nucleotide substitutions for the first and second positions (open circles) and for the third position (black dots) are shown. The year of isolation of the ancestral strain is 1968. For each case, the regression lines are superimposed. The regression coefficient is 0.0050/site/year for the first and second position and 0.0104/site/year for the third position.

and

$$b_2 = [\sum x_i y_i - (\sum x_i \sum y_i)/n]/[\sum x_i^2 - (\sum x_i)^2/n],$$

respectively, where x_i is the time of isolation of the ith strain measured from the origin, y_i is the number of nucleotide substitutions between the ith strain and the ancestral strain, and n is the number of strains used excluding the ancestral one. (Note that b_1 usually has a smaller sampling error than does b_2 because the number of degrees of freedom for b_1 is larger than that for b_2.) The regression coefficients thus obtained were used as estimates of the rate of substitution. These estimates were obtained for each of the three nucleotide positions of codons separately. The results obtained are shown in table 2. In this table, the average rates for the first and second positions are combined, since these two positions show similar rates.

Table 2 shows that the rates of nucleotide substitution obtained from b_2 are generally smaller than those obtained from b_1, as expected. The rate of nucleotide substitution varies considerably with the gene examined. The highest rate (0.0124) for "all positions" is observed for the signal peptide of the *H3* subtype, and the lowest

Table 2
Rates of Nucleotide Substitution in Influenza A Virus Genes

Gene	Subtype	N	Mean ± SE Substitution Rate/Site/Year[b]			Ratio
			1st + 2d	3d	All	(1st + 2d/3d)
Signal peptide:	*H1*	(4)	0.29 ± 0.11	1.14 ± 0.16	0.55 ± 0.08	0.26
			(−0.01 ± 0.15)	(1.16 ± 0.23)	(0.47 ± 0.11)	...
	H2	(4)	0.35 ± 0.20	1.48 ± 0.05	0.71 ± 0.14	0.24
	H3	(4)	1.39 ± 0.21	0.95 ± 0.27	1.24 ± 0.23	1.46
Mean[b]			0.68	1.18	0.83	0.58
Hemagglutinin 1: ...	*H1*	(4)	0.29 ± 0.06	0.91 ± 0.13	0.48 ± 0.08	0.32
			(0.10 ± 0.08)	(0.54 ± 0.18)	(0.23 ± 0.11)	(0.19)
	H2	(4)	0.12 ± 0.02	1.08 ± 0.23	0.42 ± 0.07	0.11
	H3	(15)	0.50 ± 0.02	1.04 ± 0.03	0.68 ± 0.01	0.48
Mean[b]			0.41	1.02	0.61	0.40
Hemagglutinin 2	*H3*	(6)	0.10 ± 0.02	0.70 ± 0.07	0.30 ± 0.03	0.14
Neuraminidase:	*N1*	(4)	0.38 ± 0.01	0.61 ± 0.09	0.45 ± 0.04	0.62
			(0.39 ± 0.02)	(0.80 ± 0.12)	(0.52 ± 0.05)	(0.49)
	N2	(5)	0.29 ± 0.02	0.63 ± 0.03	0.40 ± 0.02	0.46
Mean[b]			0.30	0.63	0.41	0.48
Nonstructural ..... protein 1		(5)	0.17 ± 0.02	0.44 ± 0.04	0.26 ± 0.03	0.39
			(0.10 ± 0.02)	(0.31 ± 0.05)	(0.17 ± 0.03)	(0.32)
Matrix protein 1		(7)	0.05 ± 0.01	0.37 ± 0.04	0.15 ± 0.01	0.14
Grand mean[b]			0.31	0.71	0.41	0.44

[a] Values of λ and its SE are multiplied by 100. Figures in parentheses are b_2 and its SE. 1st, 2d, and 3d denote the first, second, and third nucleotide positions of codons, respectively.

[b] Mean of b_1 weighted by the number of nucleotides.

rate (0.0015) is observed in the gene for matrix protein 1, the former being about eight times higher than the latter when all nucleotide positions were compared. The rate for the signal peptide of the *H3* subtype is exceptionally high, and the other subtypes for the signal peptide show lower rates. This difference results mainly from the high rate in the first and second positions for the *H3* subtype. Similar difference among subtypes is observed in the hemagglutinin 1 gene. However, even the average rate (0.0083/site/year) for three subtypes of the signal peptide is more than five times higher than that of the matrix protein 1 gene. The average rate for all genes examined is 0.0041/site/year.

Martinez et al. (1983) estimated the rate of nucleotide substitution for the *N2* subtype of the neuraminidase gene to be 0.0033–0.0070. Our estimate for this subtype (0.0040) is closer to their minimum estimate. Krystal et al.'s (1983) estimates for nonstructural proteins and for the *H3* subtype of hemagglutinins 1 and 2 are 0.0022–0.0034 and 0.0045–0.0065, respectively. Our estimates for the *NS1* gene (0.0026 by b_1) and for the hemagglutinin gene (the average of estimates for *HA1* and *HA2* of the *H3* subtype weighted by the number of nucleotides = 0.0053) are again closer to the lower ends of the range of their estimates. This indicates that the estimate obtained

from a simple comparison of the two strains tends to give an overestimate. Nei's (1983) estimate for *MX1* (0.0013) and *NS1* (0.0027) are both close to the estimates obtained in the present study. In the case of *NS1*, Nei (1983) used Air and Hall's (1981) shorter (192 nucleotides) sequence data for four strains, whereas we used longer (687 nucleotides) sequence data (see table 1) for five strains. The agreement between the two estimates suggests that the rate of nucleotide substitution for *NS1* is rather uniform for all gene regions.

It is seen from table 2 that the average rate of nucleotide substitution for the first and second nucleotide positions of codons is considerably lower than that for the third position except for the *H3* signal peptide. Indeed, the ratio of the former to the latter is 0.62 or less. This is because certain amino acids in proteins are highly conserved in the evolutionary process. This is known to be the case even with this rapidly evolving virus (Nei 1983).

The above estimates of the rate of nucleotide substitution were obtained using human strains only. It is difficult to obtain reliable estimates of the substitution rate for strains derived from nonhuman vertebrates, but there are some indications that the rate for nonhuman strains is lower than that for human strains. For example, strain NJ/76 and a swine strain WIS/30 share a common ancestor in both the *H1* subtype of *HA1* (fig. 1*A*) and the *N1* subtype of neuraminidase (fig. 2*A*). (NJ/76 is known to be essentially a swine strain, though it was isolated from humans; Blok and Air 1982.) Comparison of these two strains is expected to give an overestimate of the rate, but the estimates obtained (0.0024 and 0.0031 for *H1* and *N1*, respectively) are lower than those obtained from human strains. All five strains of the *H11* subtype of hemagglutinin were derived from birds. Because there was no main evolutionary line for this gene (fig. 1*D*), these strains were not used for estimating substitution rate. However, if we use MEM/76 as a reference strain, the rate can be estimated by the regression analysis (b_2) mentioned earlier. This gives an estimate of 0.0019 (the signal peptide and *HA1* genes combined), which is approximately one-third of the average rate for three *HA1* genes. Comparison of two bird strains, FPV/34(F) and AU/78(B), in the *NS1* gene (fig. 2*D*) also gives a rate (0.0008) that is approximately one-third to one-half of the rate (0.0026 by b_1 or 0.0017 by b_2) for human strains (table 2). It is not clear why the rate of nucleotide substitution is lower in nonhuman viral strains than in human strains, if this difference is real.

Discussion

Rate of Nucleotide Substitution in RNA Genomes

We have seen that the rate of nucleotide substitution is of the order of 10^{-3}/site/year for most influenza A virus genes studied. Using the oligonucleotide mapping technique, Takeda et al. (1984) estimated the substitution rate for the enterovirus type 70 (single-stranded RNA virus) to be 4×10^{-3}/site/year. Gojobori and Yokoyama (1985) estimated that the substitution rates for the *gag* gene (gene for internal proteins) and for the *v-mos* gene (oncogene) in a retrovirus (single-stranded RNA virus) are 6.3×10^{-4} and 1.31×10^{-3}, respectively. These high rates of nucleotide substitution in RNA viruses are believed to arise from the absence of proofreading exonucleases for correcting replication error (Holland et al. 1982). In DNA viruses such as the papovavirus, the rate of nucleotide substitution is much lower and of the order of 10^{-9} (Soeda and Maruyama 1982). The same order of substitution rate [$(2–4) \times 10^{-9}$]

has been observed for many mammalian genes (Li et al. 1985). Since DNA viruses use the replication system of the host, the mutation rate is expected to be similar to that of the host.

We have seen that the rate of nucleotide substitution at the first and second positions of codons is much lower than that at the third position except in the gene for the *H3* signal peptide of hemagglutinin. In this exceptional gene, all three nucleotide positions have essentially the same substitution rate, and the rate is ~0.01/site/year. These values suggest that for some reason there is little purifying selection operating for this gene and that nucleotide substitution occurs at the same rate as does the mutation rate. They also suggest that the relatively high substitution rates for *HA1* and neuraminidase are not caused by positive selection pressure owing to the *host immune system* against this virus but *are simply* the result of a high mutation rate. The lower rate for *HA2* than for *HA1* apparently occurs because hemagglutinin 2 constitutes the supporting leg for hemagglutinin 1 (Wilson et al. 1981) and thus is subject to stronger purifying selection than hemagglutinin 1. The lowest rate observed for the matrix protein 1 gene is also understandable from the function of this protein, since this protein underlies the lipid bilayer of the viral core and is important for viral construction and budding (Lamb 1983). Generally speaking, there seems to be a good correlation between the rate of nucleotide substitution and the level of functional constraint in the influenza A virus genes, as is true in the case of DNA genomes (Kimura 1983).

Deceleration of Substitution Rate

Since Nakajima et al. (1978) reported the striking similarity between USSR/77 and the strains that were circulating during the 1950s, the problem of deceleration of nucleotide substitution has attracted attention from many authors. Krystal et al. (1983) sequenced *NS1* genes from two strains isolated about 1950 and compared these sequences with that of USSR/77. They showed that only five nucleotide substitutions occurred during more than 20 years. Our lower estimate (via b_2) of the substitution rate for the *NS1* gene is 0.0017 (table 2). Therefore, if nucleotide substitution occurred at this rate, there should have been approximately 30 substitutions from 1950 to 1977 in the sequence of 687 nucleotides. This indicates how slow the nucleotide substitution was in the USSR/77 lineage. The same slow rate of substitution is also observed for the MEM/78 strain of the *N1* subtype of neuraminidase (see fig. 3). The reason for this unusual deceleration is not known at present. The deceleration of the rate does not seem to be an inherent property of these strains, since after their reappearance they started to evolve quickly (Young et al. 1979). Palese and Young (1983) speculated that USSR/77 might have originated from a laboratory stock that had been kept frozen for a long time. Another possibility is that about 1950 the ancestor of this strain infected some nonhuman organism, one in which the replication rate of this virus was reduced, and that it later reinfected humans.

Age of Sequence Polymorphism

The phylogenetic trees in figures 1 and 2 indicate that human and nonhuman strains can coexist for a long time showing sequence polymorphism. For example, the polymorphism of strains (or lineages) NJ/76 (a "swine" strain) and LOY/57 in *H1* (fig. 1*A*) apparently existed for more than 27 years (i.e., from 1930 to 1957). It is

possible to estimate the time of divergence between these two strains under the assumption of a constant rate of evolution. Let t_1 and t_2 be the years of isolation of strains 1 and 2 respectively, with $t_1 > t_2$. Then, the expected number of nucleotide substitutions between the two strains (d) may be expressed as $d = (2T + t_1 - t_2)\lambda$, where T is the time of duration of polymorphism, i.e., the time between t_2 and the year of divergence (the root denoted by O), and λ is the rate of nucleotide substitution per site per year. The d value may be computed by Jukes and Cantor's (1969) formula. Therefore, T can be estimated by the equation

$$T = d/(2\lambda) - (t_1 - t_2)/2.$$

In the case of the *H1* subtype (fig. 1*A*), we do not know the λ value for nonhuman strains, but if we use $\lambda = 0.005$ (the average for *SP* and *HA1*) from table 2, we obtain $T = 16$ years. This value is certainly an underestimate and apparently caused by the fact that λ is smaller in nonhuman vertebrates than in humans. In the case of nonstructural protein 1 in figure 2(*D*), a more reasonable result is obtained even if we use the λ value from human viral strains. In this case the polymorphism of AL/77 (human strain) and ALB/76 (duck strains) existed for at least 42 years (i.e., from 1934 to 1976), and the d value between the two strains is 0.383. If we use $\lambda = 0.0026$, we obtain $T = 72$. This suggests that the two strains have coexisted since 1904 or earlier.

Polymorphism of Hemagglutinin Genes

As mentioned earlier, 13 different subtypes have been identified in the hemagglutinin gene, and the extent of nucleotide divergence is very large. Nevertheless, they are polymorphic "alleles" in classical genetics and have the same biological function. It is known that they are usually host specific, and three of them (*H1, H2,* and *H3*) are carried by strains infecting humans. However, host specificity is not absolute, and switching of hosts occurs occasionally (e.g., Webster et al. 1982). The high degree of genetic diversity among the subtypes seems to result largely from the high rate of mutation in this virus in association with host specificity (Nei 1983). It is therefore interesting to know the evolutionary history of these subtypes. For this purpose we constructed a dendrogram for 13 subtypes of this gene by using the unweighted pair-grouping method (UPGMA). Previously, Hinshaw et al. (1982) and Webster et al. (1982) constructed similar dendrograms, but they used various strains that were isolated in different years. In our study we used only one strain for each subtype and chose the one that was isolated in the same or nearly the same year. In this study we used amino acid sequences rather than nucleotide sequences because the former give more reliable results when genetic divergence is large. We estimated the number of amino acid replacements per site for each pair of subtypes by the equation $d = -\log_e(1 - p)$, where p is the proportion of different amino acids. In this computation we used a region of 78 amino acids (the shared segment of hemagglutinin 1 excluding deletions/insertions; see Air 1981 and Hinshaw et al. 1982) but excluded 10 amino acids that are apparently invariable. Thus, a total of 68 amino acids were used. The proportion of different amino acids (p) and the number of amino acid replacements per site (d) for all pairs of subtypes are given in table 3, and the dendrogram obtained by UPGMA from this d matrix is presented in figure 6. This dendrogram need not represent the

Table 3
Amino Acid Differences among 13 Hemagglutinin Subtypes

	H1	H2	H5	H11	H6	H13	H8	H9	H12	H7	H10	H4	H3
H1		0.28	0.34	0.47	0.40	0.57	0.62	0.54	0.60	0.76	0.74	0.82	0.81
H2	0.33		0.22	0.44	0.44	0.54	0.60	0.53	0.59	0.72	0.75	0.82	0.78
H5	0.41	0.25		0.44	0.46	0.57	0.62	0.56	0.57	0.74	0.74	0.81	0.78
H11	0.64	0.58	0.58		0.44	0.43	0.56	0.53	0.56	0.78	0.82	0.82	0.81
H6	0.51	0.58	0.61	0.58		0.65	0.57	0.56	0.54	0.75	0.76	0.76	0.76
H13	0.85	0.79	0.85	0.56	1.04		0.54	0.54	0.60	0.78	0.82	0.82	0.90
H8	0.96	0.92	0.96	0.82	0.85	0.79		0.35	0.40	0.78	0.76	0.81	0.88
H9	0.79	0.75	0.82	0.75	0.82	0.79	0.44		0.35	0.78	0.79	0.75	0.84
H12	0.92	0.89	0.85	0.82	0.79	0.92	0.51	0.44		0.78	0.78	0.81	0.81
H7	1.45	1.28	1.33	1.51	1.39	1.51	1.51	1.51	1.51		0.50	0.68	0.62
H10	1.33	1.39	1.33	1.74	1.45	1.74	1.45	1.58	1.51	0.69		0.62	0.59
H4	1.74	1.74	1.66	1.74	1.45	1.74	1.66	1.39	1.66	1.13	0.96		0.59
H3	1.66	1.51	1.51	1.66	1.45	2.27	2.14	1.82	1.66	0.96	0.89	0.89	

NOTE.—Figures above the diagonal are the observed proportions of amino acid differences (p) in a sequence of 68 amino acids compared. Figures below the diagonal are the estimated number (d) of amino acid replacements per site. Subtypes are arranged in the order of appearance in the dendrogram of fig. 6.

true evolutionary tree, since the rate of amino acid replacements might vary with host and host switching occasionally occurs. Nevertheless, it gives a rough idea of the evolutionary divergence of the 13 subtypes of the hemagglutinin gene. Essentially the same tree topology was obtained by using Farris's (1972) distance Wagner method (tree not shown).

The topology of the dendrogram obtained is quite similar to those of Hinshaw et al. (1982) and Webster et al. (1982), though these authors used several strains for

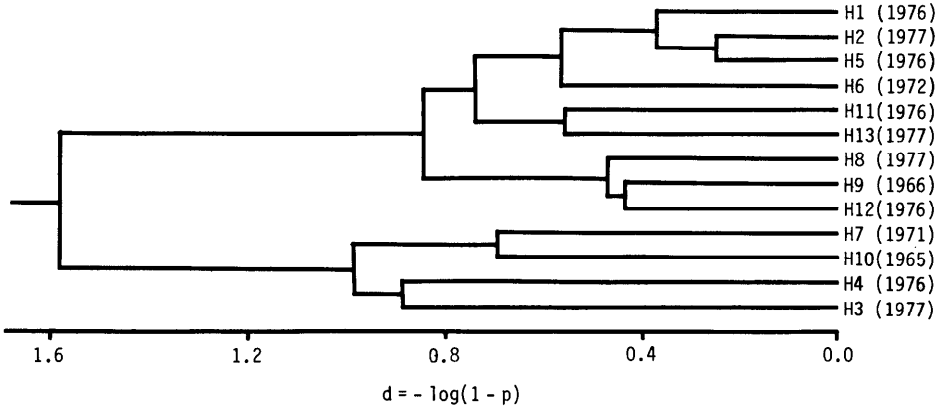

FIG. 6.—Phylogenetic tree of 13 hemagglutinin subtypes of the influenza A virus made by UPGMA. Sixty-eight amino acid sequences are used. *H1* = New Jersey/11/76 (Air et al. 1981); *H2* = duck/Alberta/77/77 (Air and Hall 1981); *H3* = Texas/1/77 (Webster et al. 1982); *H4* = duck/Alberta/28/76 (Air 1981); *H5* = shearwater/Australia/28/76 (Air 1981); *H6* = shearwater/Australia/72 (Air 1981); *H7* = turkey/Oregon/71 (Air 1981); *H8* = duck/Alberta/283/77 (G. M. Air, unpublished data); *H9* = turkey/Wisconsin/1/66 (Air 1981); *H10* = quail/Italy/1117/65 (G. M. Air, unpublished data); *H11* = duck/Memphis/546/76 (Air 1981); *H12* = duck/Alberta/60/76 (Air 1981); and *H13* = gull/Md/704/77 (Hinshaw et al. 1982).

most subtypes. This is apparently because the intrasubtype divergence is generally much smaller than the intersubtype divergence (Air 1981). However, the branch lengths of our dendrogram are considerably longer than those of the dendrograms of these authors. Air (1981) obtained a similar dendrogram for 12 subtypes of the hemagglutinin gene.

How long did it take for these subtypes to diverge from each other? This is a difficult question to answer, since the rate of amino acid replacement seems to vary with host and the extent of amino acid difference is so high. It is possible that some subtypes have existed for a long time and that the remaining similarity between them results from functional constraint rather than from lack of time to diverge. Nevertheless, it is interesting to know how long it would take for the observed level of divergence to occur under the assumption of a constant rate of amino acid replacement. For this purpose we estimated the rate of amino acid replacement for the *H1, H2, H3,* and *H11* subtypes using the same method (b_2) as that used to estimate nucleotide substitution. The estimates obtained were 0.0010, 0.0005, 0.0102, and 0.0009/amino acid site/year for subtypes *H1, H2, H3,* and *H11*, respectively, the average being 0.0035. If we accept this replacement rate, the divergence between the two largest clusters of hemagglutinins in figure 6 corresponds to 200–300 years. One may compare this result with the age of polymorphic alleles at the alcohol dehydrogenase locus in *Drosophila melanogaster.* In this case the average nucleotide difference between two randomly chosen alleles (DNA sequences) is only 0.007, but the coalescence time of polymorphic alleles (time of the earliest divergence) has been estimated to be ~1 Myr (Stephens and Nei 1986). This indicates how rapidly gene divergence has occurred in the influenza A virus.

Note added in proof.—After the submission of this paper, Hayashida et al. (Mol. Biol. Evol. **2**:289–303, 1985) reported an analysis of influenza A virus genes. Their general conclusions (extremely high mutation rate and constant rate of evolution) are virtually the same as ours as well as Nei's (1983). There is, however, one notable difference. Hayashida et al. located the Sage (A/duck/Ontario/77) strain as a direct descendant of a 1950 strain (FW) in their figure 4. In our analysis, however, FW [=FW/50] is not the ancestor of the Sage [=ON/77(D)] strain but both are derived from a common ancestor (fig. 2A). Therefore, Hayashida et al.'s conclusion (9 years of frozen period) for the Sage strain is not supported by our analysis.

Acknowledgments

We thank Dr. G. M. Air for sending us two unpublished sequences of *HA1* subtypes. We also thank Drs. A. K. Roychoudhury, J. C. Stephens, D. Graur, T. Gojobori, P. Pamilo, and W. M. Fitch for their comments. This study was supported by research grants from the National Institutes of Health and the National Science Foundation to M. Nei.

LITERATURE CITED

AIR, G. M. 1981. Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. Proc. Natl. Acad. Sci. USA **78**:7639–7643.

AIR, G. M., J. BLOK, and R. M. HALL. 1981. Sequence relationships in influenza viruses. Pp. 225–239 *in* D. H. L. BISHOP and R. W. COMPANS, eds. The replication of negative strand viruses. Elsevier North-Holland, New York.

AIR, G. M., and R. M. HALL. 1981. Conservation and variation in influenza gene sequences.

Pp. 29–44 *in* D. NAYAK and C. F. FOX, eds. Genetic variation among influenza viruses: ICN-UCLA symposia on molecular and cellular biology. Vol. **22.** Academic Press, New York.

ALLEN, H., J. MCCAULEY, M. WATERFIELD, and M. J. GETHING. 1980. Influenza virus RNA segment 7 has the coding capacity for two polypeptides. Virology **107**:548–551.

BAEZ, M., R. TAUSSIG, J. J. ZAZRA, J. F. YOUNG, P. PALESE, A. REISFELD, and A. M. SKALKA. 1980. Complete nucleotide sequence of the influenza A/PR/8/34 virus NS gene and comparison with the NS genes of the A/Udorn/72 and A/FPV/Rostock/34 strains. Nucleic Acids Res. **8**:5845–5858.

BAEZ, M., J. J. ZAZRA, R. M. ELLIOTT, J. F. YOUNG, and P. PALESE. 1981. Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus NS RNA: conservation of the NS1/NS2 overlapping gene structure in a divergent influenza virus RNA segment. Virology **113**:397–402.

BENTLEY, D. R., and G. G. BROWNLEE. 1982. Sequence of the N2 neuraminidase from influenza virus A/NT/60/68. Nucleic Acids Res. **10**:5033–5042.

BLOK, J., and G. M. AIR. 1982. Sequence variation at the 3′ end of the neuraminidase gene from 39 influenza type A viruses. Virology **121**:211–229.

BOTH, G. W., and M. J. SLEIGH. 1980. Complete nucleotide sequence of the haemagglutinin gene from a human influenza virus of the Hong Kong subtype. Nucleic Acids Res. **8**:2561–2575.

———. 1981. Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J. Virol. **39**:663–672.

BOTH, G. W., M. J. SLEIGH, N. J. COX, and A. P. KENDAL. 1983. Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J. Virol. **48**:52–60.

BUONAGURIO, D. A., M. KRYSTAL, P. PALESE, D. C. DEBORDE, and H. F. MAASSAB. 1984. Analysis of an influenza A virus mutant with a deletion in the NS segment. J. Virol. **49**: 418–425.

CONCANNON, P., I. W. CUMMINGS, and W. A. SALSER. 1984. Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene. J. Virol. **49**:276–278.

ELLEMAN, T. C., A. A. AZAD, and C. W. WARD. 1982. Neuraminidase gene from the early Asian strain of human influenza virus, A/RI/5^{-}/57 (H2N2). Nucleic Acids Res. **10**:7005–7015.

FANG, R., W. MIN JOU, D. HUYLEBROECK, R. DEVOS, and W. FIERS. 1981. Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin. Cell **25**: 315–323.

FARRIS, S. J. 1972. Estimating phylogenetic trees from distance matrices. Am. Nat. **106**: 645–668.

FITCH, W. M. 1977. On the problem of discovering the most parsimonious tree. Am. Nat. **111**: 223–257.

GOJOBORI, T., and S. YOKOYAMA. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. Proc. Natl. Acad. Sci. USA **82**:4198–4201.

HAUPTMANN, R., L. D. CLARKE, R. C. MOUNTFORD, H. BACHMAYER, and J. W. ALMOND. 1983. Nucleotide sequence of the haemagglutinin gene of influenza virus A/England/321/77. J. Gen. Virol. **64**:215–220.

HINSHAW, V. S., G. M. AIR, A. J. GIBBS, L. GRAVES, B. PRESCOTT, and D. KARUNAKARAN. 1982. Antigenic and genetic characterization of a novel hemagglutinin subtype of influenza A viruses from gulls. J. Virol. **42**:865–872.

HITI, A. L., A. R. DAVIS, and D. P. NAYAK. 1981. Complete sequence analysis shows that the hemagglutinins of the H0 and H2 subtypes of human influenza virus are closely related. Virology **111**:113–124.

HITI, A. L., and D. P. NAYAK. 1982. Complete nucleotide sequence of the neuraminidase gene of human influenza virus A/WSN/33. J. Virol. **41**:730–734.

HOLLAND, J., K. SPINDLER, F. HORODYSKI, E. GRABAU, S. NICHOL, and S. VANDEPOL. 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism III. Academic Press, New York.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KRYSTAL, M., D. BUONAGURIO, J. F. YOUNG, and P. PALESE. 1983. Sequential mutations in the NS genes of influenza virus field strains. J. Virol. **45**:547–554.

LAMB, R. A. 1983. The influenza virus RNA segments and their encoded proteins. Pp. 21–69 *in* P. PALESE and D. W. KINGSBURY, eds. Genetics of influenza viruses. Springer, Vienna and New York.

LAMB, R. A., and C.-J. LAI. 1980. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell **21**:475–485.

———. 1981. Conservation of the influenza virus membrane protein (M_1) amino acid sequence and an open reading frame of RNA segment 7 encoding a second protein (M_2) in H1N1 and H3N2 strains. Virology **112**:746–751.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

MCCAULEY, J. W., B. W. J. MAHY, and S. C. INGLIS. 1982. Nucleotide sequence of fowl plague virus RNA segment 7. J. Gen. Virol. **58**:211–215.

MARKOFF, L., and C.-J. LAI. 1982. Sequence of the influenza A/Udorn/72 (H3N2) virus neuraminidase gene as determined from cloned full-length DNA. Virology **119**:288–297.

MARTINEZ, C., L. DEL RIO, A. PORTELA, E. DOMINGO, and J. ORTIN. 1983. Evolution of the influenza virus neuraminidase gene during drift of the N2 subtype. Virology **130**:539–545.

MIN JOU, W., M. VERHOEYEN, R. DEVOS, E. SAMAN, R. FANG, D. HUYLEBROECK, W. FIERS, G. THRELFALL, C. BARBER, N. CAREY, and S. EMTAGE. 1980. Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell **19**:683–696.

NAKAJIMA, K., U. DESSELBERGER, and P. PALESE. 1978. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. Nature **274**:334–339.

NEI, M. 1983. Genetic polymorphism and the role of mutation in evolution. Pp. 165–190 *in* M. NEI and R. K. KOEHN, eds. Evolution of genes and proteins. Sinauer, Sunderland, Mass.

NEWTON, S. E., G. M. AIR, R. G. WEBSTER, and W. G. LAVER. 1983. Sequence of the hemagglutinin gene of influenza virus A/Memphis/1/71 and previously uncharacterized monoclonal antibody-derived variants. Virology **128**:495–501.

ORTIN, J., C. MARTINEZ, L. DEL RIO, M. DAVILA, C. LOPEZ-GALINDEZ, N. VILLANUEVA, and E. DOMINGO. 1983. Evolution of the nucleotide sequence of influenza virus RNA segment 7 during drift of the H3N2 subtype. Gene **23**:233–239.

PALESE, P., and J. F. YOUNG. 1983. Molecular epidemiology of influenza virus. Pp. 321–336 *in* P. PALESE and D. W. KINGSBURY, eds. Genetics of influenza viruses. Springer, Vienna and New York.

PORTER, A. G., J. C. SMITH, and J. S. EMTAGE. 1980. Nucleotide sequence of influenza virus RNA segment 8 indicates that coding regions for NS_1 and NS_2 proteins overlap. Proc. Natl. Acad. Sci. USA **77**:5074–5078.

SLEIGH, M. J., G. W. BOTH, G. G. BROWNLEE, V. J. BENDER, and B. A. MOSS. 1980. The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of cloned DNA

copies. Pp. 69–79 *in* G. LAVER and G. M. AIR, eds. Developments in cell biology. Vol. 5. Structure and variation in influenza virus. Elsevier North-Holland, New York.

SLEIGH, M. J., G. W. BOTH, P. A. UNDERWOOD, and V. J. BENDER. 1981. Antigenic drift in the hemagglutinin of the Hong Kong influenza subtype: correlation of amino acid changes with alterations in viral antigenicity. J. Virol. **37**:845–853.

SOEDA, E., and T. MARUYAMA. 1982. Molecular evolution in papova viruses and in bacteriophages. Adv. Biophys. **15**:1–17.

STEPHENS, J. C., and M. NEI. 1986. Phylogenetic analysis of polymorphic DNA sequences at the *ADH* locus in *Drosophila melanogaster* and its sibling species. J. Mol. Evol. (accepted)

TAKEDA, N., K. MIYAMURA, T. OGINO, K. NATORI, S. YAMAZAKI, N. SAKURAI, N. NAKAZONO, K. ISHII, and R. KONO. 1984. Evolution of enterovirus type 70: oligonucleotide mapping analysis of RNA genome. Virology **134**:375–388.

VAN ROMPUY, L., W. MIN JOU, D. HUYLEBROECK, and W. FIERS. 1982. Complete nucleotide sequence of a human influenza neuraminidase gene of subtype N2 (A/Victoria/3/75). J. Mol. Biol. **161**:1–11.

WEBSTER, R. G., W. G. LAVER, G. M. AIR, and G. C. SCHILD. 1982. Molecular mechanisms of variation in influenza viruses. Nature **296**:115–121.

WILSON, I. A., J. J. SKEHEL, and D. C. WILEY. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3A resolution. Nature **289**:366–373.

YOUNG, J. F., U. DESSELBERGER, and P. PALESE. 1979. Evolution of human influenza A viruses in nature: sequential mutations in the genomes of new H1N1 isolates. Cell **18**:73–83.

WALTER M. FITCH, reviewing editor

Received April 18, 1985; revision received August 1, 1985.