JOURNAL OF VIROLOGY, Feb. 1992, p. 1129–1138
0022-538X/92/021129-10$02.00/0

Vol. 66, No. 2

Evolution of the H3 Influenza Virus Hemagglutinin from Human and Nonhuman Hosts

W. J. BEAN,[1]* M. SCHELL,[2] J. KATZ,[1] Y. KAWAOKA,[1] C. NAEVE,[1] O. GORMAN,[1] AND R. G. WEBSTER[1]

Department of Virology and Molecular Biology[1] and Department of Biostatistics,[2] St. Jude Children's Research Hospital, 332 North Lauderdale, P.O. Box 318, Memphis, Tennessee 38101

Received 1 July 1991/Accepted 29 October 1991

The nucleotide and amino acid sequences of 40 influenza virus hemagglutinin genes of the H3 serotype from mammalian and avian species and 9 genes of the H4 serotype were compared, and their evolutionary relationships were evaluated. From these relationships, the differences in the mutational characteristics of the viral hemagglutinin in different hosts were examined and the RNA sequence changes that occurred during the generation of the progenitor of the 1968 human pandemic strain were examined. Three major lineages were defined: one containing only equine virus isolates; one containing only avian virus isolates; and one containing avian, swine, and human virus isolates. The human pandemic strain of 1968 was derived from an avian virus most similar to those isolated from ducks in Asia, and the transfer of this virus to humans probably occurred in 1965. Since then, the human viruses have diverged from this progenitor, with the accumulation of approximately 7.9 nucleotide and 3.4 amino acid substitutions per year. Reconstruction of the sequence of the hypothetical ancestral strain at the avian-human transition indicated that only 6 amino acids in the mature hemagglutinin molecule were changed during the transition between an avian virus strain and a human pandemic strain. All of these changes are located in regions of the molecule known to affect receptor binding and antigenicity. Unlike the human H3 influenza virus strains, the equine virus isolates have no close relatives in other species and appear to have diverged from the avian viruses much earlier than did the human virus strains. Mutations were estimated to have accumulated in the equine virus lineage at approximately 3.1 nucleotides and 0.8 amino acids per year. Four swine virus isolates in the analysis each appeared to have been introduced into pigs independently, with two derived from human viruses and two from avian viruses. A comparison of the coding and noncoding mutations in the mammalian and avian lineages showed a significantly lower ratio of coding to total nucleotide changes in the avian viruses. Additionally, the avian virus lineages of both the H3 and H4 serotypes, but not the mammalian virus lineages, showed significantly greater conservation of amino acid sequence in the internal branches of the phylogenetic tree than in the terminal branches. The small number of amino acid differences between the avian viruses and the progenitor of the 1968 pandemic strain and the great phenotypic stability of the avian viruses suggest that strains similar to the progenitor strain will continue to circulate in birds and will be available for reintroduction into humans.

Influenza viruses of the H3 serotype appeared in humans and caused a major pandemic in 1968. Subsequently, it was found that antigenically related viruses had been present in ducks and horses for a minimum of 5 years earlier (27). It is now known that H3 serotype viruses are widely distributed in waterfowl and have also been associated with occasional outbreaks of swine influenza. Serologic archaeology studies have suggested that antigenically related viruses circulated in humans during the 1890s, leading to the hypothesis that this and other serotypes may periodically recycle through the human population (28).

Previous studies with isolates of the human H3 viruses (5, 6, 14, 43) showed the progressive accumulation of mutations in the prevalent circulating strains and correlated these mutations with the progressive antigenic changes seen in the human strains. Antigenic characterization of H3N2 swine influenza viruses isolates (40, 41) suggested that these viruses had been derived from early human H3N2 viruses and had maintained antigenic characteristics that had been lost in later human virus isolates. Kida et al. (24, 25) analyzed a series of Asian avian and swine viruses and found some of the avian H3 hemagglutinin closely related to both the swine and human H3N2 viruses. The interrelationships of the H3 equine influenza viruses have been studied by Daniels et al. (10) and Kawaoka et al. (22).

We initiated this study to define the detailed evolutionary relationships among the H3 influenza viruses, to determine the mutational characteristics of the virus in different hosts, and to investigate the characteristics of the virus strain proposed to be the progenitor of the human pandemic strain.

MATERIALS AND METHODS

Virus strains and nucleic acid sequencing. The nucleic acid sequences of 40 HA genes from influenza viruses of the H3 serotype and 9 of the H4 serotype were analyzed in this study. The sources of the virus isolate sequences are summarized in Table 1. Those not previously published were obtained from the repository of St. Jude Children's Research Hospital. Virus was grown and purified, and virion RNA was prepared as described previously (3). Sequencing was either directly from virion RNA as described previously (4) or from full-length clones in the pATX vector as described by Kawaoka et al. (22). Of the sequences taken from the literature, 17 reported only the nucleotides coding for the mature peptide. For these strains (Table 1), the sequence coding for the signal peptide was obtained from virion RNA. An aligned compilation of the 40 H3 sequences will be provided on request.

* Corresponding author.

TABLE 1. Influenza virus strains studied in this analysis

Strain designation	GenBank accession no.	Reference or source
A/Memphis/12/85 (H3N2)		Katz et al. (20)[a]
A/Memphis/2/85 (H3N2)		Katz et al. (20)[a]
A/Memphis/6/86 (H3N2)	M21648	Katz and Webster (21)[a]
A/USSR/3/85 (H3N2)		Zhdanov et al. (48)
A/Bangkok/1/79 (H3N2)	J02092	Both and Sleigh (5)
A/England/321/77 (H3N2)		Hauptman et al. (17)
A/Swine/Ukkel/1/84 (H3N2)	M73775	This report
A/Swine/Colorado/1/77 (H3N2)	M73774	This report
A/Victoria/3/75 (H3N2)	J02172	Min Jou et al. (29)
A/Udorn/307/72 (H3N2)		Naeve et al. (30)
A/Memphis/102/72 (H3N2)	V02089	Sleigh et al. (42)[a]
A/Memphis/1/71 (H3N2)	J02132	Newton et al. (33)
A/NT/60/68 (H3N2)	J02135	Both and Sleigh (5)
A/Aichi/2/68 (H3N2)	J02090	Verhoeyen et al. (43)
A/Duck/Ukraine/1/63 (H3N8)	J02109	Fang et al. (12)
A/Duck/Hokkaido/5/77 (H3N2)	M16737	Kida et al. (24)[a]
A/Duck/Hokkaido/9/85 (H3N8)	M16742	Kida et al. (24)[a]
A/Duck/Hokkaido/10/85 (H3N8)	M16743	Kida et al. (24)[a]
A/Duck/Hokkaido/33/80 (H3N8)	M16739	Kida et al. (24)[a]
A/Duck/Hokkaido/8/80 (H3N8)	M16738	Kida et al. (24)[a]
A/Duck/Hokkaido/7/82 (H3N8)	M16740	Kida et al. (24)[a]
A/Swine/Hong Kong/126/82 (H3N2)	M19056	Kida et al. (25)[a]
A/Swine/Hong Kong/81/78 (H3N2)	M19057	Kida et al. (25)[a]
A/Mallard/New York/6874/78 (H3N2)	M73776	This report
A/Duck/Alberta/78/76 (H3N8)	M73771	This report
A/Duck/Memphis/928/74 (H3N8)	M73772	This report
A/Duck/Hokkaido/21/82 (H3N8)	M16741	Kida et al. (24)[a]
A/Equine/Uruguay/1/63 (H3N8)	M24718	Kawaoka et al. (22)
A/Equine/Miami/63 (H3N8)	M24719	Kawaoka et al. (22)
A/Equine/Algiers/72 (H3N8)	M24721	Kawaoka et al. (22)
A/Equine/Tokyo/71 (H3N8)	M24720	Kawaoka et al. (22)
A/Equine/New Market/76 (H3N8)	M24722	Kawaoka et al. (22)
A/Equine/Fontainebleau/76 (H3N8)	M24723	Kawaoka et al. (22)
A/Equine/France/73 (H3N8)	M73773	This report
A/Equine/Romania/80 (H3N8)	M24724	Kawaoka et al. (22)
A/Equine/Santiago/1/85 (H3N8)	M24725	Kawaoka et al. (22)
A/Equine/Kentucky/2/86 (H3N8)	M24727	Kawaoka et al. (22)
A/Equine/Johannesburg/86 (H3N8)		Kawaoka and Webster (23)
A/Equine/Tennessee/5/85 (H3N8)	M24726	Kawaoka et al. (22)
A/Equine/Kentucky/1/87 (H3N8)	M24728	Kawaoka et al. (22)
A/Duck/Alberta/28/76 (H4N6)		Donis et al. (11)
A/Chicken/Alabama/1/75 (H4N8)		Donis et al. (11)
A/Ruddy Turnstone/NJ/47/85 (H4N2)		Donis et al. (11)
A/Turkey/Minnesota/833/80 (H4N2)		Donis et al. (11)
A/Seal/Massachusetts/133/82 (H4N5)		Donis et al. (11)
A/Duck/Czechoslovakia/56 (H4N6)		Donis et al. (11)
A/Budgerigar/Hokkaido/1/77 (H4N6)		Donis et al. (11)
A/Duck/New Zealand/31/76 (H4N6)		Donis et al. (11)
A/Grey Teal/Australia/2/79 (H4N4)		Donis et al. (11)

[a] The published sequence reported only the sequence coding for the mature polypeptide. The sequence of the signal peptide region was determined from viral RNA as described in Materials and Methods.

Phylogenetic analysis. Phylogenetic analysis was primarily performed by using the program PAUP (Phylogenetic Analysis Using Parsimony), version 2.4 (David L. Swofford, Illinois Natural History Survey, Champaign, Ill.). Searches for the most parsimonious topologies for the nucleotide and amino acid sequences of the entire data set (49 taxa) were done using the MULPARS and global swap options. Detailed studies of specified regions of the tree to determine optimal and alternative topologies were done by using the branch and bound algorithm with the BBSAVE options. Phylogenetic analyses of the nucleic acid sequence in which coding substitutions were weighted more heavily than noncoding changes was done by appending each amino acid sequence to its nucleotide sequence and submitting the combined sequence to maximum parsimony analysis. This effectively doubled the weight of each coding substitution. Greater weights for the coding changes were obtained by applying the WEIGHTS option of PAUP to the characters representing the appended amino acids. Hypothetical ancestral sequences were generated as described by Fitch (13).

Statistical analysis. The amino acid change to nucleotide change ratio was estimated for the internal and terminal

branches of the phylogenetic tree for the different host species by regressing amino acid changes on nucleotide changes. Because the variance of amino acid changes for a given branch is proportional to nucleotide changes (binomial theory), the data were weighted by 1/nucleotide change, which is a standard method for assigning weights. Differences in the mutation rates of the HA1 and HA2 domains for different lineages were assessed by using the chi-square test.

Nucleotide sequence accession numbers. The nucleotide sequences for the strains not previously published are available from GenBank under accession numbers M73771 through M73776.

RESULTS

The deduced amino acid sequences for the entire coding regions of the hemagglutinins of the 40 H3 influenza virus strains and that of a hypothetical avian virus ancestor are shown in Fig. 1. The 13 equine virus isolates are 1 amino acid shorter than those from other species, and the alignment used by Daniels et al. (10) is shown. Of the 566 amino acids, 351 are invariant, and 68 others are invariant in all except one of the isolates. The most conserved of the amino acids is tryptophan. Of the 12 tryptophans in the sequence of the ancestral avian virus, 10 are invariant, followed by tyrosine (16 of 20), methionine (7 of 9), histidine (8 of 11), and cysteine (13 of 18). These five conserved amino acids are also among the least abundant. The least conserved amino acid is valine. Of 33 valines in the ancestral sequence, 14 are invariant.

For the estimation of evolutionary relationships among the virus isolates, only the nucleotide sequence of the coding region (nucleotides 30 to 1730) was used in this analysis, because the sequences of the noncoding 3′ and 5′ ends, although highly conserved where they are known, are not available for many of the strains studied. Insertions and deletions of codons were recorded for this analysis to assume that each occurred as a single mutational event rather than as three mutations. The H4 hemagglutinin was previously shown by Air (1) to be more closely related to the H3 hemagglutinins than that of any of the other subtypes. Therefore, nine H4 influenza virus gene sequences were included in these analyses as an outgroup to provide a hypothetical origin for the H3 lineage and to provide further information on the genetic conservation of the avian viruses. Their alignment with the H3 sequences is that used by Donis et al. (11).

The shortest path connecting the 49 nucleotide sequences required 2,817 steps. This pathway split the H3 taxa into three major lineages: one containing only avian viruses; one containing avian, human, and swine virus isolates; and one containing all of the equine virus isolates. Several approaches were used to test the robustness of the topology. Since the number of taxa was too great to allow an exhaustive search of all possible topologies, alternative topologies and shorter paths were searched for by reanalyzing the data after dividing the tree into major sections. The data set was also analyzed by using the amino acid sequences. This resulted in a tree with the same general topology as the tree based on the nucleic acid sequences, but the presence of several short and zero-length internal branches connecting the avian virus isolates resulted in a large number (>50) of equally parsimonious solutions. This problem was eliminated by including the entire RNA sequence and applying greater weight to the coding changes. This allowed the junctions affected by amino acid changes to be studied while preserving those supported only by silent mutations. The coding changes supported a phylogenetic tree differing from that supported by all nucleotide changes at two of the branch points. These are labeled A and B in Fig. 2 and are detailed in Fig. 3. One of these (node B) significantly affects the interpreted origin of the equine lineage and is discussed below. The other has a minor effect on the Asian avian lineage near the origin of the human virus but does not affect the interpretation of the avian-human virus junction. The only other significant ambiguity was at the joining of the two 1968 human virus strains (Aichi and NT/60) with the rest of the tree. In the topology shown, a 3-nucleotide, 0-amino-acid branch joins these two strains with the trunk of the tree. An equally parsimonious solution branches each separately from the same point on the trunk. Analysis of the H3 nucleotide data using the Neighbor Joining Method (36) gave a topology nearly identical to that obtained when all nucleotides were used with the maximum parsimony method. The only difference was that Swine/Ukkel/84 branched from the main trunk immediately before the branch containing Vic/75 and Swine/Col/77, rather than immediately after this branch. The significance of this difference is not known.

The remarkable features of the phylogenetic tree are the close linkage of the human virus lineage to the avian viruses, the rapid divergence of the human viruses from their origin, the great separation of the equine viruses from those of other species, and the strong conservation of the amino acid sequence in the avian lineages.

Analysis of the avian-human junction. The phylogenetic tree indicates that the human virus of the H3N2 serotype originated from a lineage closely related to a series of avian virus isolates from Asia described by Kida et al. (24). To identify the changes that occurred in this virus during its transition from an avian pathogen to a human pandemic strain, the hypothetical ancestral sequences of the last avian and first human strains (Fig. 2, nodes A and C) were determined (13) and their differences were studied in detail. The transition required 13 nucleic acid changes, 7 of which affect the amino acid sequence. The locations of these changes are shown in Fig. 1. One change is not part of the mature protein but is at a highly variable site in the signal peptide; the others are all in the HA1 peptide and in the head of the molecule. These locations are detailed in Fig. 4, and their significance is discussed below.

Rates of accumulation of mutations. The mutation rate of the human sublineage was estimated by plotting the year of isolation against the evolutionary path distances from node A to each isolate (Fig. 5, top). Regression analysis gave mutation rates of 7.9 nucleotides and 3.4 amino acids per year for the coding region (1,701 nucleotides, 566 amino acids). Two swine influenza virus isolates derived from the human virus lineage (Swine/Col/1/77 and Swine/Ukkel/84) were included in these calculations, but the calculated mutation rate is not significantly affected when they are excluded. Extrapolation of the nucleotide regression lines gives dates of 1967 and 1965 for nodes C and A, respectively.

Similar analysis of the equine virus lineage (Fig. 5, bottom) gives substitution rates of 3.1 nucleotides and 0.8 amino acids per year. The date of origin for the lineage (node D, Fig. 2), extrapolated from the nucleotide regression, is 1952. The approximately linear relationship between the number of mutations and the isolation date is valid for the topology shown in Fig. 2 but not for the alternative topology (Fig. 3, top right). With the alternative topology, the 1971 and 1972 isolates are farther from the origin (node B) than are the 1986 strains, and the four earliest isolates appear to form a

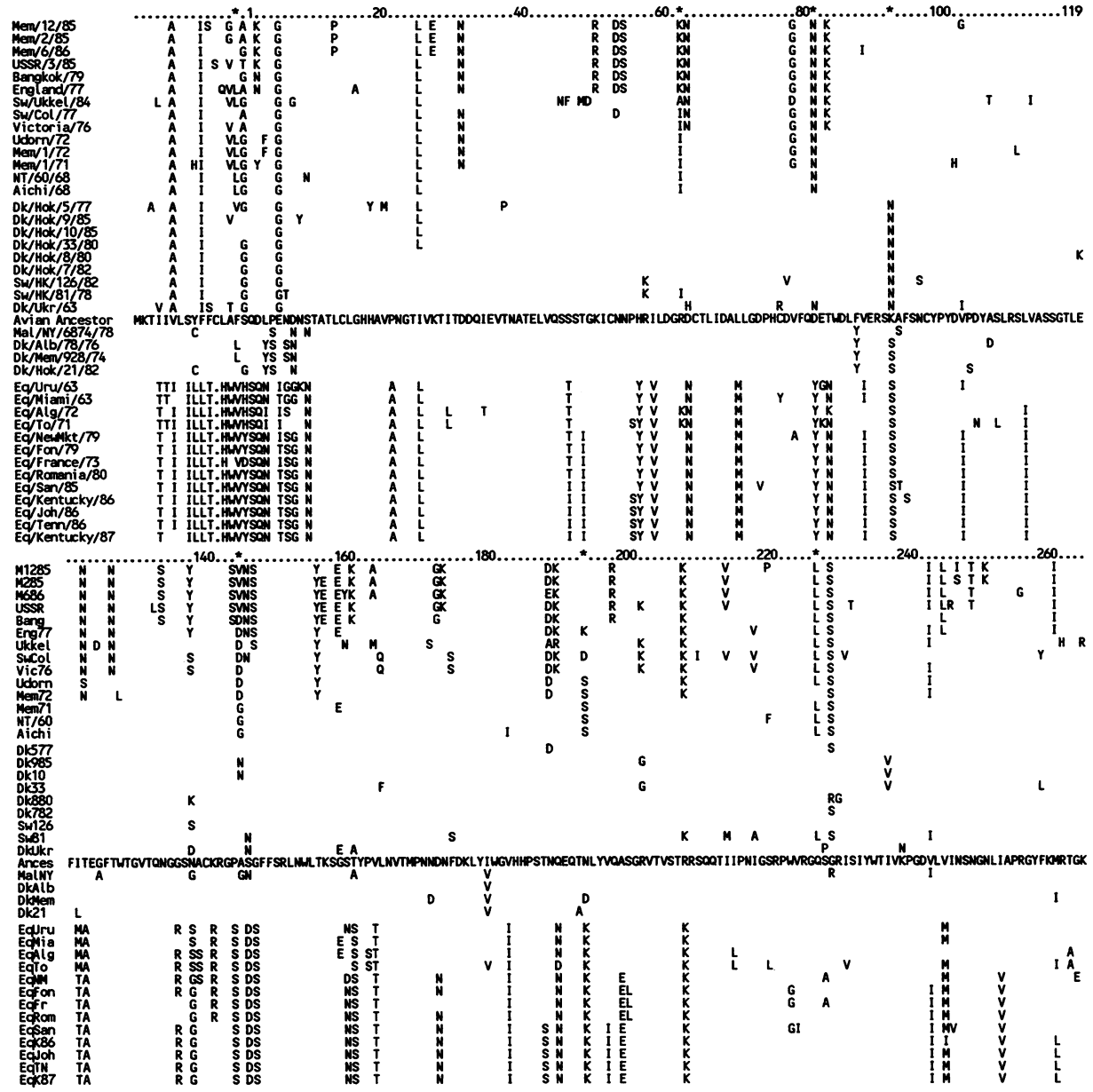

FIG. 1. Predicted amino acid sequences for the H3 influenza virus strains. The sequence of a hypothetical avian ancestral strain (node D, Fig. 2) was reconstructed as described by Fitch (13) and used as a baseline in this figure. For the other sequences, only those that differ from the ancestral strain are shown. The groups defined by Duck/Ukraine, the Asian avian viruses, and the human viruses are displayed above the ancestral strain, while the smaller avian virus group, represented by three North American isolates and Dk/Hok/21/82, is immediately below the ancestral sequence, followed by the equine viruses. Amino acid changes at the avian-human junction (nodes A and C, Fig. 2) are marked with asterisks. Not indicated on the figure is the insertion of an additional asparagine at position 8 in the A/Victoria/3/75 strain.

separate sublineage with an apparent mutation rate about twice as fast as that of the other sublineage (4.8 versus 2.5 nucleotide substitutions per year). Either topology requires several parallel mutations, and neither can be positively excluded on the basis of sequence information from the currently available virus isolates.

Among the avian viruses, no consistent relationship was seen with respect to the date of isolation and the path lengths from common nodes and thus no meaningful divergence dates for other interior nodes could be calculated.

Phenotypic change and conservation. It is clear from the examination of Fig. 2 that the number of amino acid changes relative to the number of nucleotide changes is lower among the H3 and H4 avian viruses than among the mammalian viruses. Table 2 shows ratios of amino acid changes versus nucleotide changes for the human, equine, swine, and avian regions of the tree. The data are also separated into terminal branches, connecting actual viral sequences with the tree, and internal branches, connecting nodes within the tree. The ratios shown were calculated from the sums of nucleotide

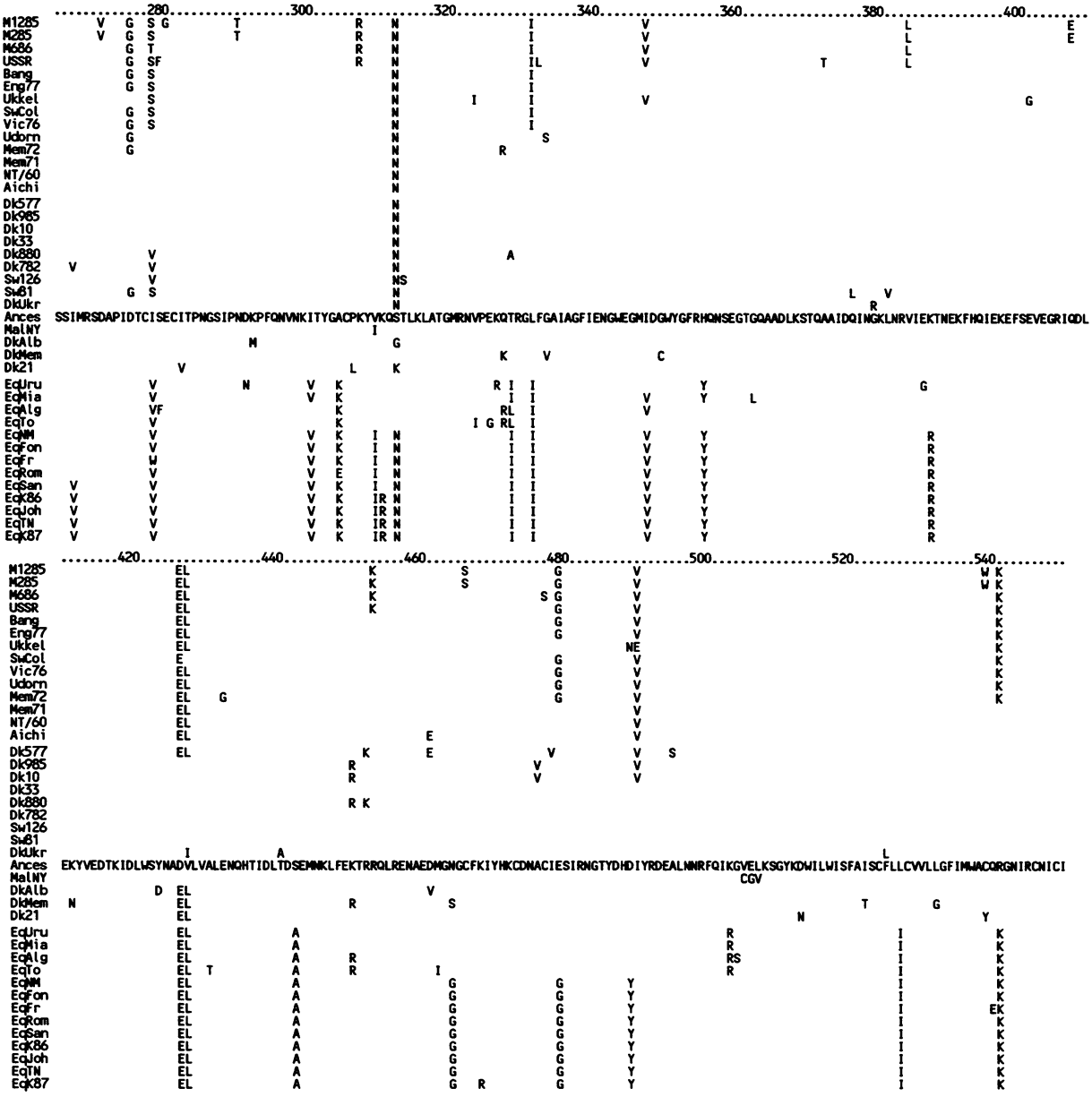

FIG. 1—*Continued.*

and amino acid branch lengths for each branch type and range from 0.46 for the internal human virus isolates to 0.074 for the internal branches of the H4 virus strains. The ratios of coding to total changes of the human internal branches are greater than those for the human terminal branches. The opposite is the case for the equine viruses; none of the differences among the mammalian virus branches is statistically significant. However, the ratios for the internal branches of the avian virus lineages are significantly lower than those for the terminal branches, and the differences between the avian and mammalian branches are also significant.

To allow examination of the distribution of coding and noncoding mutations on the H3 hemagglutinin polypeptide, the hypothetical ancestral sequences of the H3 strains were reconstructed as described by Fitch (13) and each variable position in the RNA sequence was examined. Figure 6 shows each mutation plotted according to its position in the protein sequence, its host lineage, and whether it is coding or noncoding.

The paucity of coding changes in the avian virus lineage and their concentration between amino acids 50 and 300 in the human strains are readily apparent. The partition of the coding and noncoding changes between HA1 and HA2 in the various species and branch types of the H3 strains was analyzed in detail (Table 3). The coding changes for each of the mammalian lineages show a significant deviation from a random distribution, whereas the mutations in the avian virus lineages do not. No coding regions were found notably lacking in silent mutations.

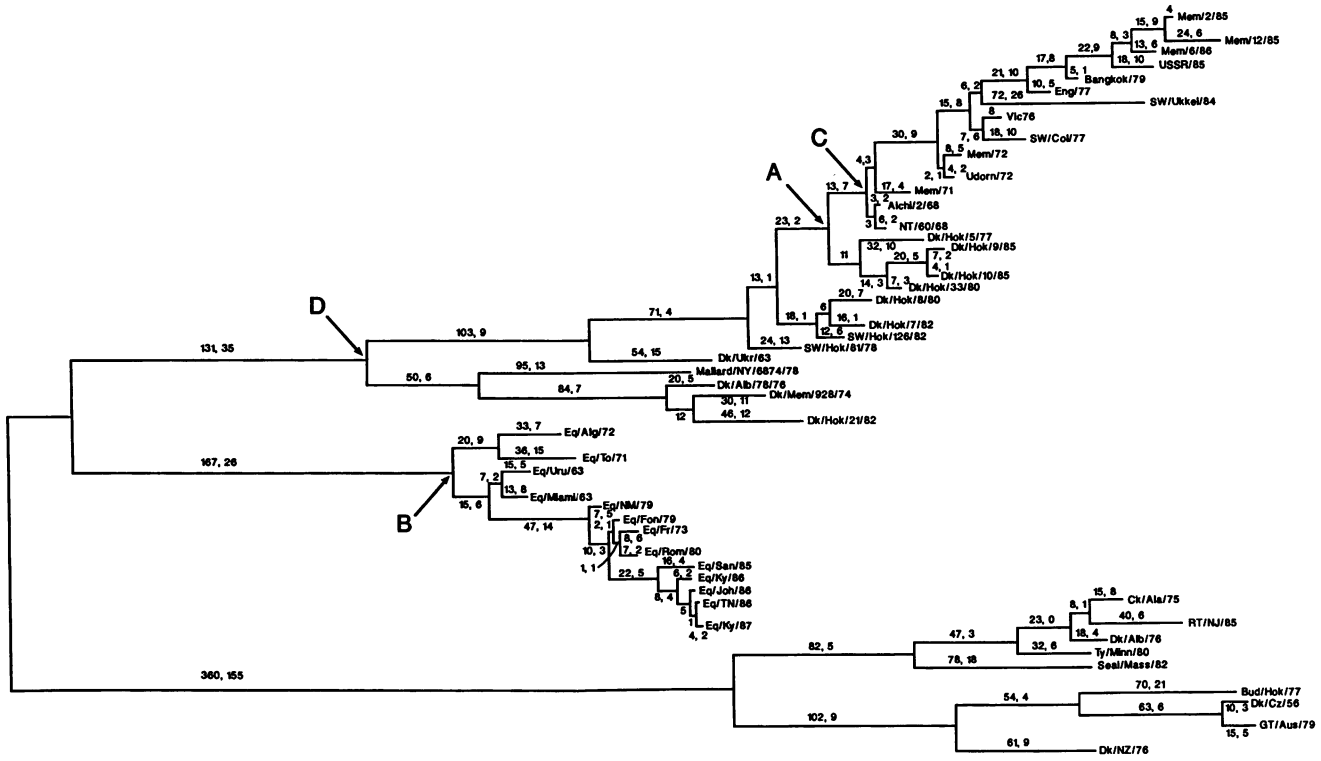

FIG. 2. Hypothetical phylogenetic tree joining the H3 and H4 influenza A virus hemagglutinin genes. This proposed phylogeny requires 2,122 nucleotide changes and 571 amino acid changes. The number of nucleotide changes and amino acid changes are indicated on each branch. Branches with one number have no amino acid changes. Alternative joinings for two ambiguous junctions (A and B) are detailed in Fig. 3. The proposed first human virus is indicated as C. The hypothetical avian ancestral strain used as a baseline in Fig. 1 is indicated as D.

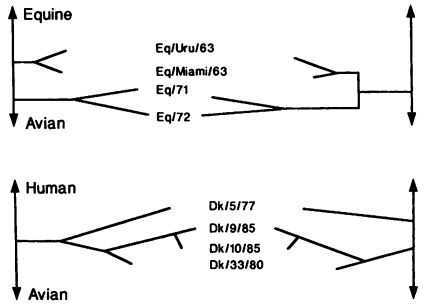

FIG. 3. Alternative topologies for two ambiguous junctions of the phylogenetic tree. Differences between the most parsimonious trees calculated from the nucleotide sequences and slightly longer trees requiring fewer amino acid changes are shown. (Top) Equine-avian junction (Fig. 2, node B). The shortest tree is shown on the right. With this topology, the four earliest equine viruses share a common divergence from the main lineage and require 2,817 nucleotide changes for the 49 taxa. With the topology on the left, Eq/Tokyo/71 and Eq/Algeria/72 branch from the main lineage before the two 1963 isolates. This topology requires two fewer amino acid changes but five additional nucleotide changes. The topology on the left is used in Fig. 1 and was chosen because it provides a consistent relationship between the date of virus isolation and distance from the origin of the equine virus lineage. (Bottom) Four avian taxa near the avian-human junction (Fig. 2, node A). The joining on the right requires three additional nucleotide substitutions but one less amino acid substitution. The topology on the left was used in Fig. 1.

DISCUSSION

This study was initiated to define the source and characteristics of the progenitor of the hemagglutinin of the current H3 human viruses and to provide a better understanding of the genetic interrelationships and mutational constraints of the virus strains in different hosts. The H3 hemagglutinin phylogeny calculated from the sequence data clearly shows that there is a remarkably close relationship among the first human H3 viruses and some of the virus strains still circulating in avian species. The results indicate that the human hemagglutinin gene was very recently introduced into the virus infecting humans and that it underwent only a few mutations in its transition from an avian virus to a human pandemic strain. One of these mutations (position 226) has been previously implicated in host specificity (30). Of the others, three (positions 62, 144, and 193) are among the most variable in the molecule, each having three or four different amino acids at the site within the human lineage. These locations have been shown to be affected by antigenic drift or host range selection (5, 6, 20, 21, 44). One of the sites (position 193), which changes from Asn to Ser during the transition, later reverts to the ancestral form. The changes at these positions may not have been part of the adaptation of the avian virus to humans but may have been the early stages of antigenic drift that occurred before the virus had been detected.

The calculated date of 1965 for the introduction of the H3 gene into a human virus is made on the basis of assumptions that the mutation rate of the virus has been constant, that the

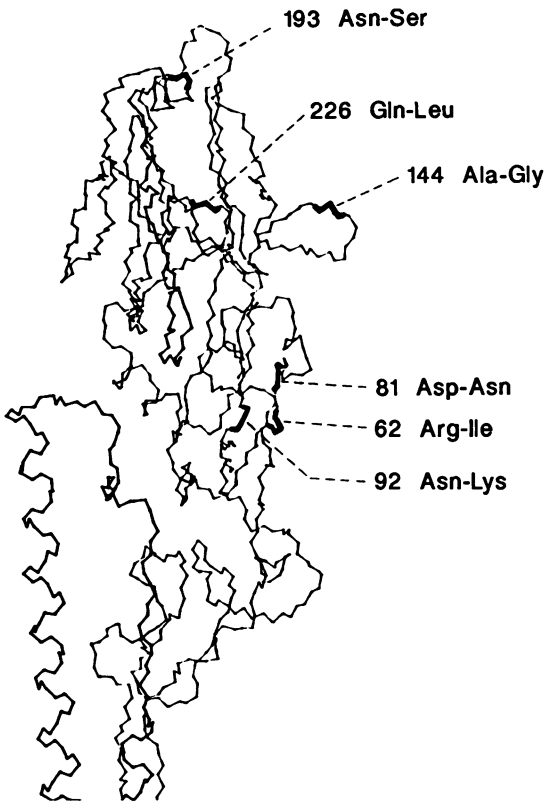

FIG. 4. Amino acid changes in the hemagglutinin proposed to have been associated with the early adaptation of the avian hemagglutinin to humans. The figure shows the human H3 hemagglutinin structure, as determined by Wilson et al. (46). The locations of the six amino acid changes required between nodes A and C of Fig. 2 are indicated. The position number is followed by the amino acids present in the avian and human forms, respectively.

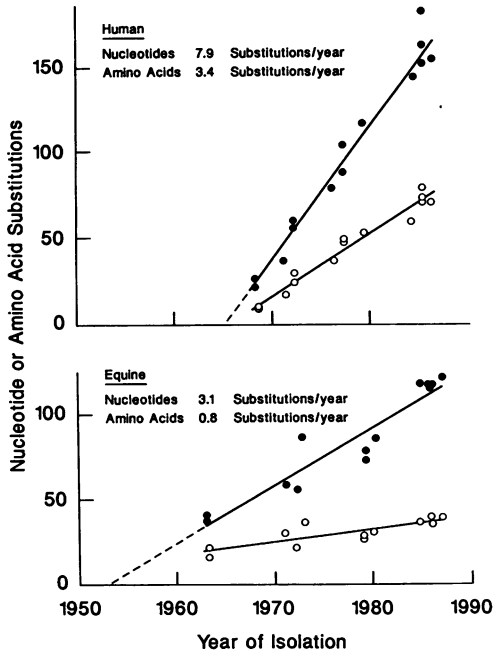

FIG. 5. Rates of accumulation of nucleotide and amino acid changes in the human and equine viruses. Rates were calculated by the regression of the date of isolation and the total branch distance from nodes A and B for the human and equine virus isolates, respectively (Fig. 2). The amino acid and nucleotide distances are plotted as open and closed circles, respectively.

introduction occurred at node A on the phylogenetic tree, and that there were no intermediate hosts. The assumption that the introduction occurred at node A rather than at node C (Fig. 2) is based on the high ratio of coding to noncoding changes in the branch linking nodes A and C which is characteristic of the virus in humans rather than in birds.

The evolutionary record provides no evidence for an intermediate host between the avian and human sublineages, but it cannot be ruled out that one may have existed, and the possibility of swine intermediaries in the generation of human influenza virus strains has been previously considered (15, 19, 26, 31, 32, 37–39, 45). Of the four swine virus isolates included in this analysis, two were introduced into swine from the human virus lineage and two are recent introductions from avian viruses. This finding and previous evidence for the transmission of virus between birds and swine (19, 34), as well as documented transmissions of influenza virus from swine to humans (18, 35), leaves open the possibility that a swine intermediate could have been involved. The ratios of the coding changes to mutations in the four swine virus isolates included in this study were similar to those of the human strains. Thus, the data are consistent with a transfer of the hemagglutinin from the avian reservoir into swine at node C and then into humans at node A. Regardless of the initial mammalian host, if the amino acid substitutions during this period were selected by antibody pressure, the virus is likely to have been the predominant strain in a population, sufficiently large to have maintained the virus but limited enough to have escaped detection.

In contrast to the progressive changes of both the nucleotide and amino acid sequences of the mammalian virus lineages, the avian viruses show far less variation and no clear relationship between the position on the phylogenetic tree and the date of isolation. Additionally, most of the coding changes in the avian lineages have occurred in the terminal branches, whereas in the mammalian lineages the terminal and internal branches have similar ratios of coding and noncoding changes. This fundamental difference in the

TABLE 2. Ratios of coding changes to total nucleotide changes in the terminal and internal branches of the avian and mammalian evolutionary lineages

Branch type[a]	No. of branches	Amino acid changes (AAC)	Nucleotide changes (NC)	Ratio (AAC/NC)
Avian (H3), int.	12	39	423	0.092[b,c]
Avian (H3), ter.	11	82	334	0.246[c]
Avian (H4), int.	7	28	380	0.074[b,c]
Avian (H4), ter.	9	80	338	0.234[c]
Human, int.	13	77	168	0.458
Human, ter.	12	43	120	0.358
Equine, int.	11	46	132	0.348
Equine, ter.	13	59	143	0.413
Swine, ter.	4	55	126	0.436

[a] Phylogenetic tree branches (Fig. 2). ter., terminal branches connecting virus isolates to the tree; int., internal branches connecting nodes on the tree.
[b] Significantly less than corresponding terminal branches ($P = 0.01$).
[c] Significantly less than corresponding mammalian branches ($P = 0.01$).

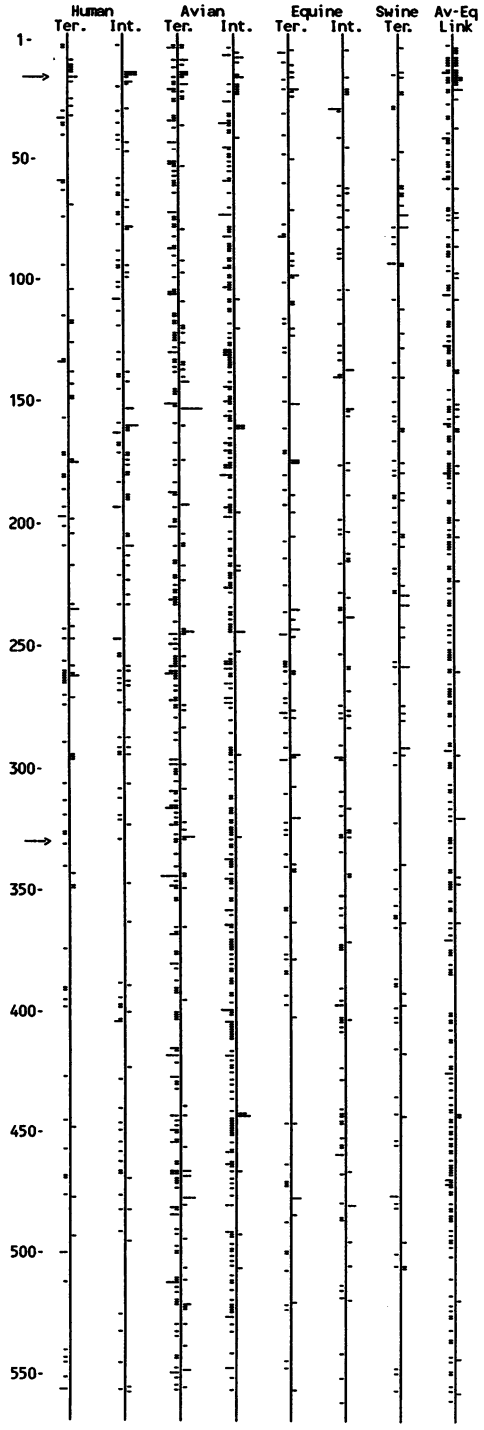

FIG. 6. Distribution of coding and noncoding mutations of the phylogenetic tree for the H3 hemagglutinin. Separate plots were constructed for the terminal and internal branches for human, avian, and equine lineages, for terminal branches for swine, and for the link between the avian and equine lineages. The numbers on the vertical axis represent amino acid positions 1 to 556 in the hemagglutinin molecule, with the initiating methionine as 1. Dashes on the left and right of the vertical axes represent noncoding and coding mutations, respectively.

TABLE 3. Comparison of the occurrence of mutations in the HA1 and HA2 domains of different host lineages

Branch type[a]	No. of mutations		*P* value[b]
	HA1	HA2	
Human, TN	44	22	0.09
Human, TC	32	6	0.0004
Human, IN	58	23	0.004
Human, IC	51	10	<0.0001
Avian, TN	146	93	0.08
Avian, TC	53	28	0.09
Avian, IN	174	116	0.12
Avian, IC	20	8	0.11
Equine, TN	47	24	0.09
Equine, TC	41	12	0.001
Equine, IN	44	39	0.44
Equine, IC	35	10	0.003
Swine, TN	34	27	0.84
Swine, TC	42	9	<0.0001
Avian-equine, IN	137	95	0.39
Avian-equine, IC	32	8	0.002

[a] Branch type (from Fig. 6). TN, terminal branches, noncoding mutations; TC, terminal branches, coding mutations; IN, internal branches, noncoding mutations; IC, internal branches, coding mutations.

[b] The *P* values were determined by the chi-square test. Significant results indicate differences in the occurrence of mutations in the two domains.

evolution of the avian viruses suggests that their long-term survival favors those that have maintained the original phenotype. If survival favors those that have not changed, then virus populations in environments that undergo relatively few replication cycles would be more likely to yield progeny that do not have deleterious mutations. Those replicating in other environments or mutants in the original population might have a temporary selective advantage in a particular host or environment, but the accumulation of mutations in these subpopulations would be deleterious in other circumstances. Thus, the original population (perhaps often in a very small minority) would have a selective advantage as hosts or environmental conditions change.

The striking difference in the topologies of the H3 and H4 avian lineages in comparison with the human virus lineage is apparently due to a heavy positive selective pressure on viruses replicating in humans that is not seen when replicating in birds. Fitch et al. (14) have studied the unusual "cactuslike" topology of the human influenza A viruses and have shown that the proportion of amino acid changes affecting antigenic regions of the hemagglutinin (HA1) is greater on the main trunk than on the side branches of the tree. This provided convincing evidence for positive Darwinian selection on the HA1 molecule, mediated by immune pressure. They calculated the age of the nonsurviving side branches (the distance from the main lineage trunk to the branch tips) to be 1.6 years. The present study, using a very different data set and the entire coding sequence, gives a very similar value of 1.5 years. The increased mutation rate seen in the hemagglutinin when transferred from an avian to a mammalian host is paralleled by smaller increases in the mutation rates of other genes (9, 14, 16).

The topology of the avian virus lineages resembles that described for the human influenza C virus (7, 8, 47), with multiple cocirculating lineages and little correlation between tree position and date of isolation. In both cases this may reflect a long-established equilibrium between the virus and the host. The influenza B viruses show a regular relationship between tree position and isolation date but a slower rate of

change than the human influenza A strains and cocirculating branches that survive much longer than the short side branches of the influenza A viruses (14, 47). Air et al. (2) have analyzed the proportion of silent and nonsilent mutations in the human A and B viruses and have proposed that the evolution of the B viruses is not primarily driven by immune selection.

In the present study, the only internal avian branches with a large proportion of coding changes on the phylogenetic tree are the long links connecting the H3 equine viruses and the H4's with the rest of the taxa. However, these lengths and their connection points to each other must be interpreted with caution. The large number of silent mutations, particularly in the link to the H4 viruses, make it likely that the total number of mutations is underestimated, since any of them could have mutated multiple times. Lacking evolutionary intermediates and with no information on what selective pressures or time periods were involved in the separation of the HA subtypes, we cannot estimate the actual lengths of these branches with any certainty.

The stability of the avian strains suggests that the virus has reached an adaptive optimum in birds that has not been achieved in humans. There are several possible explanations for this difference. One obvious possibility is that because this gene was only recently introduced into the human virus, it has not had time to reach an optimal configuration. Another possibility is that when the virus is in humans there is sufficient flexibility in the interaction of the hemagglutinin with the host that no single configuration is optimal and that mutations selected by antibody may not significantly affect other properties or functions of the protein. This implies that in an immunologically naive population, the viruses derived by continued immunological selection would not be at a selective disadvantage if placed in competition with the original strain. A third possibility, and perhaps the most likely, is that the H3 hemagglutinin was already in its optimal configuration after its initial adaptation as a human pathogen. Subsequent mutations in response to antibody pressures in the human population, while essential for the continued survival of the virus lineage, would put the virus at a disadvantage if forced to compete with the original strain in an immunologically naive population. This question remains to be tested directly but is consistent with the observed periodic replacement of influenza virus strains in the human population with another serotype and the hypothesized recycling of influenza virus strains (28). If the recycling of influenza virus serotypes is to occur, it requires a mechanism for maintaining the virus while it is not circulating in humans, and the avian virus reservoir clearly provides one. The highly conserved phenotype of the H3 hemagglutinin in birds suggests that strains very similar to the progenitor of the 1968 pandemic will continue to circulate in birds and will be available for reintroduction into mammalian hosts in the future.

ACKNOWLEDGMENTS

We thank Raphael Onwuzuruigbo and Evelyn Stigger for technical assistance.

This work was supported by Public Health Service grants AI-20591, AI-08831, AI-29680, AI-29599, and AI-27497 from the National Institute of Allergy and Infectious Diseases, by Cancer Center Support Grant (CORE) CA-21765, and by the American Lebanese Syrian Associated Charities.

REFERENCES

1. **Air, G. M.** 1981. Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A viruses. Proc. Natl. Acad. Sci. USA **78**:7639–7643.
2. **Air, G. M., A. J. Gibbs, W. G. Laver, and R. G. Webster.** 1990. Evolutionary changes in influenza B are not primarily governed by antibody selection. Proc. Natl. Acad. Sci. USA **87**:3884–3888.
3. **Bean, W. J., G. Sriram, and R. G. Webster.** 1980. Electrophoretic analysis of iodine-labeled influenza virus RNA segments. Anal. Biochem. **102**:228–232.
4. **Bean, W. J., S. C. Threlkeld, and R. G. Webster.** 1989. Biologic potential of amantadine-resistant influenza A virus in an avian model. J. Infect. Dis. **159**:1050–1056.
5. **Both, G. W., and M. J. Sleigh.** 1981. Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J. Virol. **39**:663–672.
6. **Both, G. W., M. J. Sleigh, N. 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.
7. **Buonagurio, D. A., S. Nakada, W. M. Fitch, and P. Palese.** 1986. Epidemiology of influenza C virus in man: multiple evolutionary lineages and low rate of change. Virology **153**:12–21.
8. **Buonagurio, D. A., S. Nakada, U. Desselberger, M. Krystal, and P. Palese.** 1985. Noncumulative sequence changes in the hemagglutinin gene of influenza C virus isolates. Virology **146**:221–232.
9. **Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch.** 1986. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science **232**:980–982.
10. **Daniels, R. S., J. J. Skehel, and D. C. Wiley.** 1985. Amino acid sequences of haemagglutinins of influenza viruses of the H3 subtype isolated from horses. J. Gen. Virol. **66**:457–464.
11. **Donis, R. O., W. J. Bean, Y. Kawaoka, and R. G. Webster.** 1989. Distinct lineages of influenza virus H4 hemagglutinin genes in different regions of the world. Virology **169**:408–417.
12. **Fang, R., W. Min Jou, D. Huylebroeck, R. Devos, and W. Fiers.** 1981. Complete structure of A/Duck/Urkraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin. Cell **25**:315–323.
13. **Fitch, W. M.** 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.
14. **Fitch, W. M., J. M. E. Leiter, L. Xingqiang, and P. Palese.** 1991. Positive Darwinian evolution in human influenza viruses. Proc. Natl. Acad. Sci. USA **88**:4270–4274.
15. **Gorman, O. T., W. J. Bean, Y. Kawaoka, I. Donatelli, Y. Gou, and R. G. Webster.** 1990. Evolution of the influenza A virus nucleoprotein genes: implications for the origins of H1N1 human and classical swine viruses. J. Virol. **65**:3704–3714.
16. **Gorman, O. T., W. J. Bean, Y. Kawaoka, and R. G. Webster.** 1990. Evolution of the nucleoprotein gene of influenza A virus. J. Virol. **64**:1487–1497.
17. **Hauptman, 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.
18. **Hinshaw, V. S., W. J. Bean, R. G. Webster, and B. C. Easterday.** 1978. The prevalence of influenza viruses in swine and the antigenic and genetic relatedness of influenza viruses from man and swine. Virology **84**:51–62.
19. **Hinshaw, V. S., R. G. Webster, W. J. Bean, J. C. Downie, and D. A. Senne.** 1983. Swine influenza-like viruses in turkeys: a potential source of virus for humans? Science **220**:206–208.
20. **Katz, J. M., C. W. Naeve, and R. G. Webster.** 1987. Host cell-mediated variation in H3N2 influenza viruses. Virology **156**:386–395.
21. **Katz, J. M., and R. G. Webster.** 1988. Antigenic and structural characterization of multiple subpopulations of H3N2 influenza virus from an individual. Virology **165**:446–456.
22. **Kawaoka, Y., W. J. Bean, and R. G. Webster.** 1989. Evolution

of the hemagglutinin of equine H3 influenza viruses. Virology **169**:283–292.

23. **Kawaoka, Y., and R. G. Webster.** 1989. Origin of the hemagglutinin on A/Equine/Johannesburg/86 (H3N8): the first known equine influenza outbreak in South Africa. Arch. Virol. **106**: 159–164.
24. **Kida, H., Y. Kawaoka, C. W. Naeve, and R. G. Webster.** 1987. Antigenic and genetic conservation of H3 influenza in wild ducks. Virology **159**:109–119.
25. **Kida, H., K. F. Shortridge, and R. G. Webster.** 1988. Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China. Virology **162**:160–166.
26. **Kundin, W. D.** 1970. Hong Kong A2 influenza virus infection among swine during a human epidemic in Taiwan. Nature (London) **228**:857.
27. **Laver, W. G., and R. G. Webster.** 1973. Studies on the origin of pandemic influenza. III. Evidence implicating duck and equine influenza viruses as possible progenitors of the Hong Kong strain of human influenza. Virology **51**:383–391.
28. **Masurel, N., and W. M. Marine.** 1973. Recycling of Asian and Hong Kong influenza A virus hemagglutinins in man. Am. J. Epidemiol. **97**:44–49.
29. **Min Jou, W., M. Verhoeyen, R. Devos, et al.** 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.
30. **Naeve, C. W., V. S. Hinshaw, and R. G. Webster.** 1984. Mutations in the hemagglutinin receptor-binding site can change the biological properties of an influenza virus. J. Virol. **51**:567–569.
31. **Nakajima, K., E. Nabusawa, and S. Nakajima.** 1984. Genetic relates between A/Swine/Iowa/15/30 and human influenza viruses. Virology **139**:194–198.
32. **Nerome, K., M. Ishida, A. Oya, and K. Oda.** 1982. The possible origin of H1N1 (HswlN1) virus in the pig population of Japan and antigenic analysis of isolates. J. Gen. Virol. **62**:171–175.
33. **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.
34. **Pensaert, M., K. Ottis, J. Vandeputte, M. M. Kaplan, and P. A. Bachman.** 1981. Evidence for the natural transmission from wild ducks to swine and its potential importance for man. Bull W.H.O. **59**:75–78.
35. **Rota, P. A., E. P. Rocha, and M. W. Harmon.** 1989. Laboratory characterization of a swine influenza virus isolated from a fatal case of human influenza. J. Clin. Microbiol. **27**:1413–1416.
36. **Saitu, N., and M. Nei.** 1987. The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
37. **Scholtissek, C., H. Burger, O. Kistner, and K. F. Shortridge.** 1985. The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. Virology **147**: 287–294.
38. **Scholtissek, C., W. Rohde, V. Von Hoynigen, and R. Rott.** 1978. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology **87**:13–28.
39. **Schultz, U., W. M. Fitch, S. Ludwig, J. Mandler, and C. Scholtissek.** 1991. Evolution of pig influenza viruses. Virology **183**:61–73.
40. **Shortridge, K. F., A. Cherry, and A. P. Kendal.** 1979. Further studies on the antigenic properties of H3N2 strains of influenza A isolated from pigs in Southeast Asia. J. Gen. Virol. **44**:251–254.
41. **Shortridge, K. F., R. G. Webster, W. K. Butterfield, and C. H. Campbell.** 1977. Persistence of Hong Kong influenza virus variants in pigs. Science **196**:1454–1455.
42. **Sleigh, M. J., G. W. Both, P. A. Underwood, and V. J. Bender.** 1981. Antigenic drift in the hemagglutinin of Hong Kong influenza subtype: correlation of amino acid changes with changes in viral antigenicity. J. Virol. **37**:845–853.
43. **Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers.** 1980. Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature (London) **286**:7771–7776.
44. **Wang, M., J. M. Katz, and R. G. Webster.** 1989. Extensive heterogeneity in the hemagglutinin of egg-grown influenza viruses from different patients. Virology **171**:275–279.
45. **Webster, R. G., and W. G. Laver.** 1975. Antigenic variation of influenza viruses, p. 296–314. *In* E. D. Kilbourne (ed.), The influenza viruses and influenza. Academic Press, Inc., New York.
46. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. The hemagglutinin membrane glycoprotein of influenza virus: structure at 3 Å resolution. Nature (London) **289**:366–373.
47. **Yamashita, M., M. Krystal, W. Fitch, and P. Palese.** 1988. Influenza B virus evolution: cocirculating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. Virology **163**:112–122.
48. **Zhdanov, V. M., N. A. Petrov, A. A. Grinev, M. A. Yakhno, V. A. Ishachenko, Y. A. Gorbunov, I. A. Vtorushina, S. V. Netesov, S. K. Vaselenko, and L. S. Sandakhchiev.** 1989. Primary structure of influenza A viruses (H3N2) isolated in USSR, 1985. Vopr. Virusol. **2**:155–160.