Arch Virol (1992) 123: 73–87

Archives of Virology
© Springer-Verlag 1992
Printed in Austria

# Evolutionary pattern of the H3 haemagglutinin of equine influenza viruses: multiple evolutionary lineages and frozen replication

**A. Endo, R. Pecoraro*, S. Sugita, and K. Nerome**

Department of Virology and Rickettsiology, National Institute of Health, Shinagawa-ku, Tokyo, Japan

Accepted August 8, 1991

**Summary.** The nucleotide and deduced amino acid sequences of the haemagglutinin genes coding for the HA1 domain of H3N8 equine influenza viruses isolated over wide regions of the world were analyzed in detail to determine their evolutionary relationships. We have constructed a phylogenetic model tree by the neighbour-joining method using nucleotide sequences of 15 haemagglutinin genes, including those of five viruses determined in the present study. This gene tree revealed the existence of two major evolutionary pathways during a twenty five-year period between 1963 to 1988, and each pathway appeared to consist of two distinct lineages of haemagglutinin genes. Furthermore, our analysis of nucleotide sequences showed that two distinct lineages of equine H3N8 viruses were involved in an equine influenza outbreak during the period of December 1971–January 1972 in Japan. The number of nucleotide changes between strains was proportional to the length of time (in years) between their isolation except for three of the HA genes. However, there are three exceptional strains isolated in 1971, 1987, and 1988, respectively. The haemagglutinin gene in these strains showed a small number of nucleotide substitutions after they branched off around 1963, suggesting an example of frozen replication. Although the estimated rate (0.0094/site/year) of synonymous (silent) substitutions of the haemagglutinin gene of equine H3N8 viruses was nearly the same as that of human H1 and H3 haemagglutinin genes, the rate of nonsynonymous (amino-acid changing) substitutions of the former equine virus gene was estimated to be 0.00041/site/year – that is about 5 times lower than that estimated for the human H3 haemagglutinin gene. The present study is the first demonstration that multiple evolutionary lineages of equine H3N8 influenza virus circulated since 1963.

* Present address: La Plata Veterinary Faculty, University of La Plata, Argentina.

## Introduction

Following the first isolation of H7N7 influenza virus from a horse in Czechoslovakia in 1956 [38], influenza virus H3N8 subtype was isolated in 1963 during a widespread outbreak of equine respiratory disease in the U.S.A. [44]. This isolate was referred to as the prototype strain A/eq/Miami/1/63. Since then, large-scale serological and virological surveillance has been done in many parts of the world [6, 7, 17, 23, 29, 30, 39–42], but the introduction of influenza A subtypes virus other than the above two subtypes has not been reported.

During the course of virological surveillance of a wide variety of bird species, a large number of influenza viruses were isolated, and it was shown that there are 14 haemagglutinin (HA) and 9 neuraminidase (NA) subtypes of influenza A viruses [14, 18]. The sudden appearance of A/Hong Kong (H3N2) virus in man brought research worker's attention to the bird and the horse populations because: the HA antigen of this virus was related immunologically and biochemically to that of avian and equine influenza viruses [24]. The subsequent isolation of H3N2 virus from swine established with certainty a close epidemiological relationship between human and swine influenza [37]. Antigenic analysis, competitive RNA-hybridization, and analysis of nucleotide sequences led to speculation that the HA gene of the 1968 human pandemic strain was most likely to have been introduced from avian species [10, 24, 36]. The extensive sequence information now available for a number of H 3 HA genes of human and animal viruses has made it possible to analyze molecular evolution of influenza viruses from different hosts [10, 20]. On the basis of nucleotide and amino acid sequence analyses, human influenza A viruses appear to evolve rapidly in single lineage [3, 13]. In contrast, the avian influenza viruses evolve at a much slower rate as multiple lineages [19]. A number of nucleotide sequences of H 3 HA genes are now available for a study of the evolution of equine influenza viruses [3, 8, 13, 16]. A phylogenetic tree constructed from the nucleotide sequences of the HA genes indicated that equine H3N8 influenza viruses like human H3N2 viruses evolve as a single lineage [16]. However, our results (unpubl. data) based on the oligonucleotide mapping of the entire RNAs, suggested the cocirculation of the distinct lineage of equine H3N8 virus. We therefore attempted to reconcile the differences between the above two observations by using the neighbour-joining (N-J) method to analyze the equine H 3 HA gene sequences reported here along with those reported previously [27, 33]. In the present communication, authors describe evolutionary pathways and rates of evolution of equine H3N8 influenza virus isolated since 1963.

## Material and methods

### *Viruses*

The following five strains of equine influenza virus were selected and grown in 11 day-old fertile hen's eggs: A/eq/France/67 (FRA/67), A/eq/Tokyo/3/71 (T/3/71), A/eq/Kentucky/1/81 (KEN/81), A/eq/Brazil/87 (BRA/87) and A/eq/La Plata/88 (LPL/88). They were pur-

ified as described previously [28]. In addition to the nucleotide sequences of the above five strains of H3N8 virus, those of the previously sequenced HA genes of the strains of equine H3N8 virus were also used in the present analysis. They were A/eq/Miami/1/63 (MIA/63), A/eq/Uruguay/1/63 (URG/63), A/eq/Tokyo/2/71 (T/2/71), A/eq/Algiers/72 (ALG/72), A/eq/New Market/76 (NM/76), A/eq/Fontainbleau/79 (FON/79), A/eq/Romania/80 (ROM/80), A/eq/Santiago/1/85 (SAN/85), A/eq/Kentucky/2/86 (KEN/86), and A/eq/Kentucky/1/87 (KEN/87) viruses.

### *Cloning and nucleotide sequences of the HA gene*

The cDNA was synthesized from viral RNA with a 12 base universal primer complementary to the 3′ terminus of influenza A virus gene using a cDNA synthesis kit (Amersham International) [21]. Second strand DNA was also prepared similarly in the presence of ribonuclease H and polymerase I. After treatment with T 4 polymerase, the resultant DNAs were cloned into pUC 118 vector using a synthetic Eco RI linker, and the recombinant plasmid was transformed into *Escherichia coli* DH 5α. The plasmid containing H 3 HA DNA was identified by colony hybridization and isolated by the boiling method. Following the transfection of the plasmid possessing HA-DNA into JM 103, single strand DNA was collected for sequencing after infection with helper phage M13K07 as described previously [35]. The nucleotide sequences of the HA were determined by using a series of primers synthesized in our laboratory.

## Results

### *Nucleotide and deduced amino acid sequences of the HA gene of equine H3N8 viruses*

In order to analyze the evolutionary patterns of equine H 3 HA genes, the complete nucleotide sequences of five HA genes were determined. The sequence of their HA 1 domains are presented in Fig. 1, along with those of the ten previously sequenced HA genes of H3N8 viruses. The equine influenza virus HA gene of the H 3 subtype was 1762 nucleotides long with a single open reading frame of 1650 nucleotides (Fig. 2), identical in size to the previous reports [8, 16]. No insertions or deletions were observed in any of the fifteen HA genes. The HA gene of equine H 3 virus encodes a polypeptide of 565 amino acids, 15 of which comprise the signal peptide, 329 comprise the HA 1 and 221 comprise the HA 2. Although a large number of the nucleotide changes were scattered over the HA 1 domain, amino acid changes were mainly clustered between residues 137 and 257. It was evident that three antigenic sites [45] exist. By analogy with receptor specificity of the H 3 HA, the presence of glutamine at residue 226 of all equine virus HAs of the H 3 subtype suggests that these viruses bind Neu α-2–3 Gal [31]. There are six to eight potential N-linked glycosylation sites on the HA 1 molecule. In addition, the HA 2 molecule contained only one potential glycosylation site (Fig. 2), coinciding with previous reports [8, 16].

```
                 10        20        30        40        50        60        70        80        90       100
MIA/63  AGCAAAAGCAGGGGATATTTCTGTCAATCATGAAGACAACCACTATTTTGATACTACTGACCCATTGGGTCCACAGTCAAAACCCAACCGGTGGCAACAA
URG/63  ...........................................................................................T.........A..
FRA/67  ..........................................T............................T.................A..........
T/3/71  ..............................................................................T.....................
T/2/71  .................................................................................T.....T.AA..A......
ALG/72  ..........................................T...............A.T...............C....T.....TTA...A......
NM/76   ..........................................T............................T...............T.A..........
FON/79  ..........................................T............................T.................A..........
ROM/80  ..........................................T............................T.................A..........
KEN/81  ..........................................T...........................GT.................A..........
SAN/85  ..........................................T............................T.................A.......T..
KEN/86  ..........................................T............................T.................A..........
KEN/87  ..........................................T.G..........................T.................A..........
BRA/87  ..............................................................................T.....................
LPL/88  ..............................................................................T.....................
                110       120       130       140       150       160       170       180       190       200
MIA/63  CACAGCCACACTGTGCCTGGGACACCATGCAGTAGCAAATGGAACACTGGTAAAAACAATAACTGATGACCAGATTGAGGTGACAAATGCTACTGAATTA
URG/63  ....................................................................................................
FRA/67  ............A..T..............................T.........................A...........................
T/3/71  ....................................................................................................
T/2/71  ............A..............................................T........................................
ALG/72  ............A..............................................T..............C.........................
NM/76   ...............T.............................G..........................A...........................
FON/79  ............A..T........................................................A........................C..
ROM/80  ............A..T..............................T.........................A...........................
KEN/81  ............A..T..............................T.........................A...........................
SAN/85  ..........T.A..T..A...........................T......................T..A...........................
KEN/86  ..........T.A..T..............................T.........................A...........................
KEN/87  ..........T.A..T....................................................................................
BRA/87  ....................................................................................................
LPL/88  ....................................................................................................
                210       220       230       240       250       260       270       280       290       300
MIA/63  GTTCAGAGCACTTCAACAGGGAAAATATGCAACAACCCATATAGGGTCCTAGATGGAAGAAACTGCACATTAATAGATGCAATGCTGGGAGATCCCCATT
URG/63  ....................................................................T...............................
FRA/67  ................T..............................T......................................A...........C.
T/3/71  ....................................................................................................
T/2/71  ........T...........A...........T...T..........T..........A............G............................
ALG/72  .....A..T...........A...........T.......................G.A.........................................
NM/76   ................T..............................T......................................A...........C.
FON/79  ................T..............................T......................................A...........C.
ROM/80  ................T..............................T......................................A...........C.
KEN/81  ................T..............................T......................................A...........C.
SAN/85  ..........T.....T..............................T......................................A...T.C.....C.
KEN/86  ..........T.....T...................T..........T..............T.......................A.....C.....C.
KEN/87  ..........T.....T...................T..........T..............T.......................A.....C.....C.
BRA/87  ....................................................................................................
LPL/88  ....................................................................................................
                310       320       330       340       350       360       370       380       390       400
MIA/63  ATGATGTTTTTCSGTATGAGAATTGGGACCTCTTCATTGAAAGAAGCAGTGCTTTCAGCAATTGCTACCCATATGACGTCCCTGACTATGCATCGCTCCG
URG/63  G.................G..........................................................A......................
FRA/67  G...........A........................A...........C...........................A......................
T/3/71  G.A.........A.......................................................................................
T/2/71  G..................................G...............................................A.........T......
ALG/72  G..................................G.....................................C...........................
NM/76   G.....C...............G.....T........A...........C...........................A......................
FON/79  G....................................A...........C...........................A......................
ROM/80  G....................................A...........C...........................A......................
KEN/81  G...........A........................A...........C...........................A......................
SAN/85  G....................................A...........CA..........................A......................
KEN/86  G....................................A...........C...C.......................A......................
KEN/87  G....................................A...........C...........................A......................
BRA/87  G.A.........A.......................................................................................
LPL/88  G.A.........A.......................................................................................
                410       420       430       440       450       460       470       480       490       500
MIA/63  GTCTCTTGTGGCATCTTCAGGAACGTTAGAATTCATGGCAGAGGGATTCACATGGACAGGTGTCACTCAAAACGGAGGAAGTAGCGCCTGCAGAAGGGGA
URG/63  ...........................................................................GA.......................
FRA/67  ....A...................A..........CA.............................................G.................
T/3/71  ....................................................................................................
T/2/71  ....A................................................................................T..............
ALG/72  ....A....A.....C.............................................................A........T..............
NM/76   ....A...................A..........CA........................................A.....G..T..............
FON/79  ....A...................A..........CA........................................A.....G.................
ROM/80  ....A...................A..........CA.............................................G.................
KEN/81  ....A...................A..........CA.............................................G.A..............G
SAN/85  ....A...................A..........CA........................................A.....G.A.......A.......
KEN/86  ....A..........C........A..........CA........................................A.....G.A.......A.......
KEN/87  ....A..........C........A..........CA........................................A.....G.A.......A.......
BRA/87  ....................................................................................................
LPL/88  ....................................................................................................
                510       520       530       540       550       560       570       580       590       600
MIA/63  TCAGCCGATAGTTTCTTTAGCCGACTGAATTGGCTAACACAATCTGAAAGTTCCTACCCCACATTGAATGTGACAATGCCTAACAATGACAATTTCGATA
URG/63  ....................T..................A......G..A..................................................
FRA/67  .......................................A......G..A...T..............................................
T/3/71  .......................................A......G.....................................................
T/2/71  ...................................G...A......G..........T.....C...................T................
ALG/72  ........C..............................A.................T..........................................
NM/76   .......................................A......G.GA...T.................................A............
FON/79  .......................................A......G..A...T.................................A............
ROM/80  .......................................A......G..A...T.................................A............
KEN/81  .......................................A......G..A...T.................A...............A............
SAN/85  .......................................A......G..A...T.................................A............
KEN/86  .......................................A......G..A...T.................................A............
KEN/87  .......................................A......G..A...T.................................A............
BRA/87  .......................................A......G.....................................................
LPL/88  .......................................A......G.....................................................
```

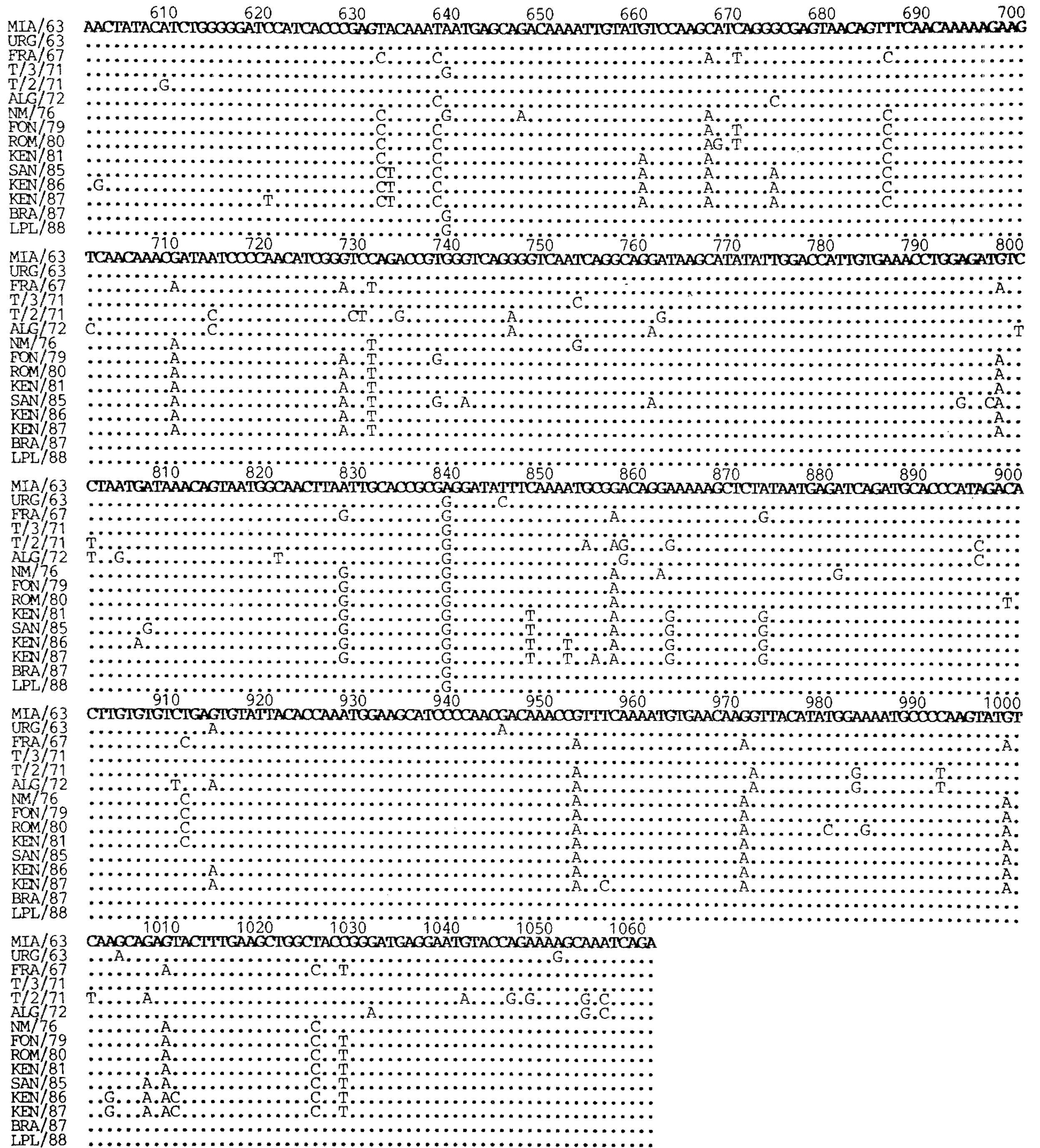

**Fig. 1.** Nucleotide sequences of the complete HA 1 domain of the H 3 HA gene of equine influenza viruses. The nucleotide sequences of the each HA gene were aligned together with those of the previously sequenced HA genes. The oldest equine (H3N8) virus HA gene was used as the prototype strain but only nucleotides that differ from those of the former prototype strain are shown

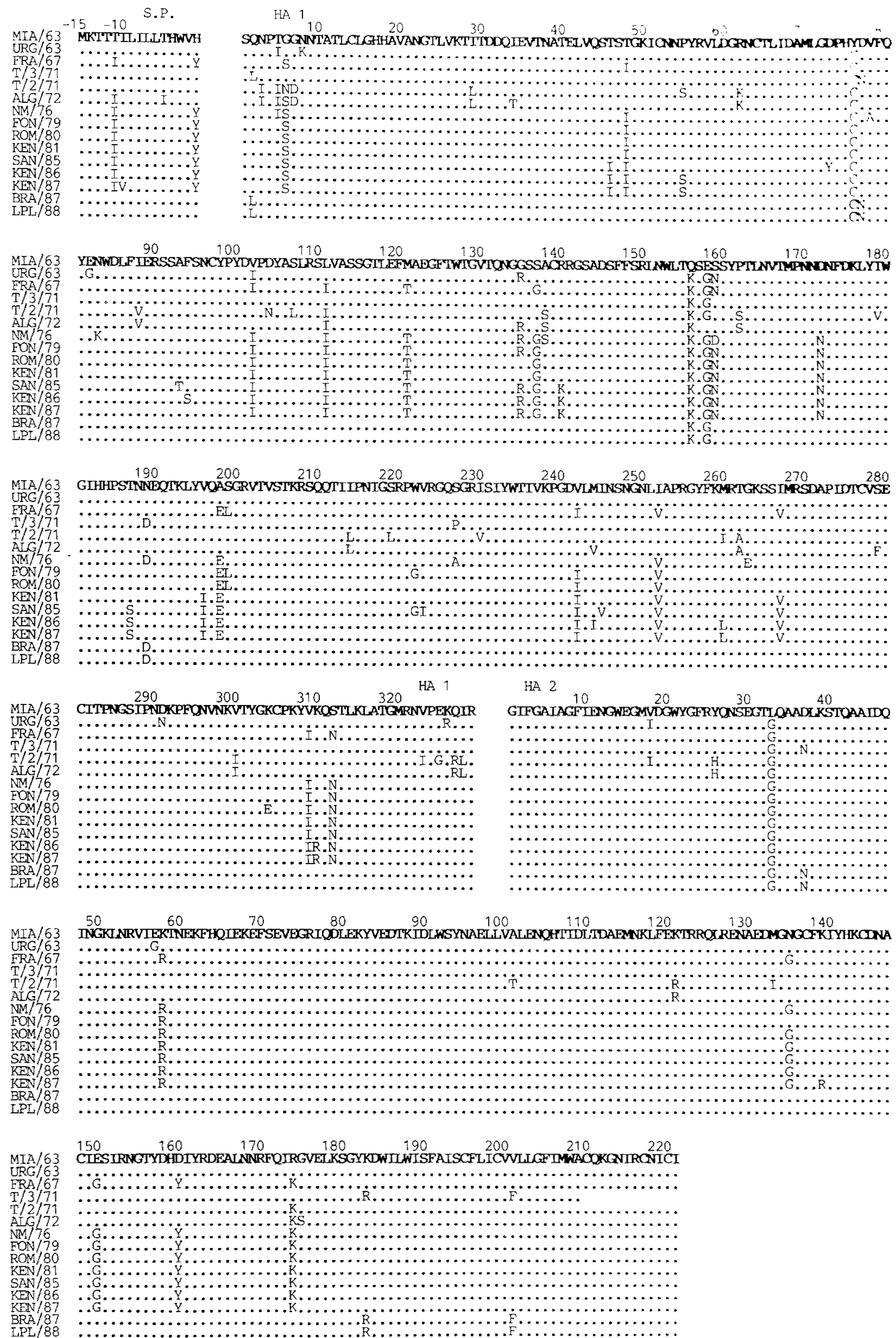

**Fig. 2.** Comparison of deduced amino acid sequences of the HA 1 domain presented in Fig. 1. All of the amino acid sequences were aligned with that of the oldest strain MIA/63, but only nucleotides that differ from those of MIA/63 virus are shown

**Table 1.** Comparison of nucleotide and amino acid sequence homologies amongst haemagglutinin gene coding for HA 1 domain of H 3 equine influenza viruses isolated during twenty five-year period between 1963 and 1988

| Strain | Nucleotide and amino acid sequence homology (%) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MIA/63 | URG/63 | FRA/67 | T/2/71 | T/3/71 | ALG/72 | NM/76 | FON/79 | ROM/80 | KEN/81 | SAN/85 | KEN/86 | KEN/87 | BRA/87 | LPL/88 |
| MIA/63 | – | *98.2* | 95.7 | 95.1 | *99.1* | 95.1 | 95.1 | 95.5 | 95.3 | 95.1 | 93.6 | 93.7 | 93.8 | *99.2* | *99.2* |
| URG/63 | 96.6 | – | 95.1 | 94.3 | 98.2 | 94.5 | 95.0 | 95.1 | 94.8 | 94.6 | 93.4 | 93.5 | 93.6 | 98.3 | 98.3 |
| FRA/67 | 94.8 | 94.5 | – | 92.8 | 95.7 | 92.6 | 97.8 | *99.3* | *99.3* | *98.9* | 97.4 | 97.5 | 97.2 | 95.8 | 95.8 |
| T/2/71 | 91.7 | 90.8 | 89.3 | – | 95.1 | 95.5 | 92.4 | 92.6 | 92.4 | 92.2 | 91.1 | 91.4 | 91.5 | 95.1 | 95.1 |
| T/3/71 | 97.8 | 96.3 | 94.5 | 91.4 | – | 94.9 | 95.4 | 95.5 | 95.3 | 95.1 | 93.6 | 93.7 | 93.8 | *99.8* | *99.8* |
| ALG/72 | 93.6 | 92.6 | 90.8 | 95.4 | 92.6 | – | 92.2 | 92.6 | 92.2 | 92.0 | 91.0 | 91.3 | 91.4 | 95.0 | 95.0 |
| NM/76 | 92.9 | 93.6 | 96.4 | 88.7 | 93.6 | 90.8 | – | 98.0 | 97.6 | 97.2 | 95.9 | 96.0 | 95.9 | 95.4 | 95.4 |
| FON/79 | 94.2 | 94.5 | 98.7 | 88.7 | 93.9 | 90.8 | 96.6 | – | *99.1* | *98.5* | 97.4 | 97.3 | 97.2 | 95.6 | 95.6 |
| ROM/80 | 94.5 | 94.2 | 99.0 | 89.0 | 94.2 | 90.5 | 96.3 | 99.0 | – | *98.5* | 97.0 | 97.1 | 96.8 | 95.4 | 95.4 |
| KEN/81 | 94.5 | 94.2 | 99.0 | 89.0 | 94.2 | 90.5 | 96.3 | 98.4 | 98.7 | – | 97.6 | 97.7 | 97.4 | 95.2 | 95.2 |
| SAN/85 | 91.7 | 92.0 | 96.3 | 86.2 | 91.4 | 88.4 | 94.2 | 96.6 | 96.0 | 97.2 | – | 97.9 | 97.6 | 93.7 | 93.7 |
| KEN/86 | 91.7 | 92.0 | 96.3 | 87.2 | 91.4 | 88.7 | 94.2 | 96.3 | 96.0 | 97.2 | 96.9 | – | *99.0* | 93.8 | 93.8 |
| KEN/87 | 92.3 | 92.6 | 96.9 | 87.8 | 92.0 | 89.0 | 94.8 | 96.9 | 96.6 | 97.8 | 97.5 | 99.3 | – | 93.9 | 93.9 |
| BRA/87 | 98.1 | 96.6 | 94.8 | 91.7 | 99.7 | 92.9 | 93.6 | 94.2 | 94.5 | 94.5 | 91.7 | 91.7 | 92.3 | – | *100* |
| LPL/88 | 98.1 | 96.6 | 94.8 | 91.7 | 99.7 | 92.9 | 93.6 | 94.2 | 94.5 | 94.5 | 91.7 | 91.7 | 92.3 | 100 | – |

Nucleotide and amino acid sequence homologies among HA 1 domains of haemagglutinin genes were shown in lower left half and upper right half of the table, respectively

### *Nucleotide and deduced amino acid sequence homologies*

The differences in nucleotide and deduced amino acid sequences of the HA gene based on a pairwise comparison, are summarized in Table 1. With the exception of strains T/3/71, BRA/87, and LPL/88, equine H 3 HA genes tend to show nucleotide changes proportional to the amount of time between isolations. However, the number of nucleotide and amino acid sequence differences between T/3/71 and MIA/63 were smaller than those between FRA/67 and MIA/63. Similarly, despite the sixteen to seventeen-year interval between their isolations, nucleotide sequences of two viruses isolated in Brazil and Argentina (BRA/87, LPL/88) differed by only 1 nucleotide from T/3/71 derived from the outbreak in Japan. Our suggestion is that the above two strains, had been "frozen" in nature or elsewhere.

On the basis of considerable nucleotide sequence similarity (98.2–99.2% homology), the HA genes of MIA/63, URG/63, T/3/71, BRA/87, and LPL/88 are thought to belong to the same lineage. Similarly, another equine lineage showing a high nucleotide sequence homology (97.2–98.0%) was found to include strains NM/76, FON/79, ROM/80, and KEN/81. These results suggest that at least two mainstreams of equine H3N8 viruses, distinguished by nucleotide differences, were involved in influenza outbreaks among horses in different countries during 1963–1981.

### *Evolutionary relationships among the HA genes of equine influenza virus*

The genetic diversity of the equine H 3 HA genes and possible existence of multiple pathways led us to analyze their evolutionary patterns. As seen in the phylogenetic tree constructed from nucleotide sequences (Fig. 3), the branching pattern for fifteen isolates was different from that reported by Kawaoka et al. [16]. The values given for each branch in the tree represent the expected branch length (the expected number of nucleotide substitutions per site). Each main branch can be distinguished by an evolutionary tree, one including seven strains isolated between 1963 and 1988, and the other one including the remaining eight strains. It was also shown that each major branch cluster contains two minor branches consisting of several strains of H3N8 virus. These results indicated the possibility that equine H 3 HA genes have evolved into four lineages.

The phylogenetic tree clearly indicated that the branch giving rise to the HA gene of the oldest strain MIA/63 belonging to the first lineage had previously been divided off from the second lineage containing the T/2/71 and ALG/72 strains. Probably, they have a common putative root with that of the oldest strain of H3N8 virus, suggesting their derivation from a common ancestral virus which had been prevalent in the early 1960s. These observations are essentially consistent with the evidence obtained from the phylogenetic tree constructed by the maximum parsimony method [16]. Although the second branch cluster consists of eight strains, they can be further divided into two or

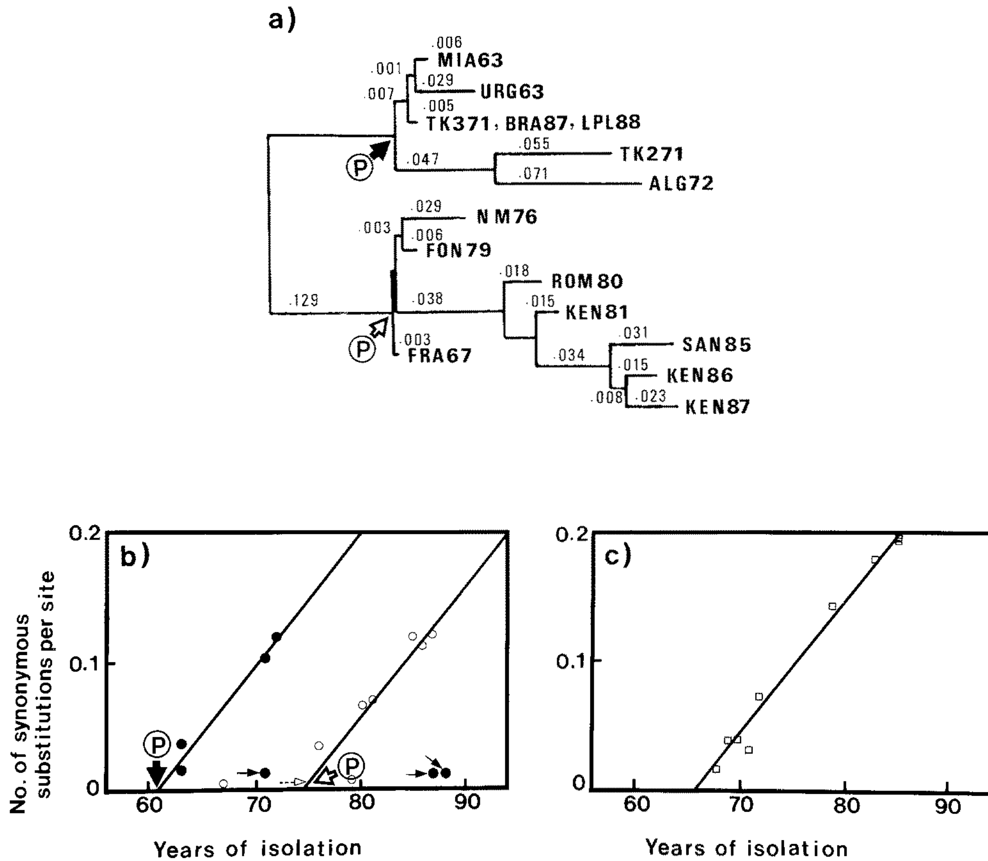


**Fig. 3.** Phylogenetic tree for equine H3N8 influenza virus HA genes (**a**). The tree was constructed by the neighbour-joining method [25, 32]. Genomic distance was estimated on the basis of number of synonymous substitutions as described by Nei and Gojubori [27]. The length of all the branches were calculated by the principle of minimum evolution. Linear regression analysis of the synonymous substitution of equine (**b**) and human (**c**) H 3 HA genes. Closed circles represent the strains MIA/63, URG/63, T/2/71, and ALG/72, respectively. Closed circles indicated by arrows indicate the position of strains T/3/71, BRA/87, and LPL/88 (strains whose replication was frozen during a certain period). Strains FRA/67, NM/76, FON/79, ROM/80, KEN/81, SAN/85, KEN/86, and KEN/87 are shown from left by open circles. The regression line (**c**) was prepared by the following human H3N2 viruses: A/Northern Territory/60/68, A/Aichi/2/68, A/England/878/69, A/Queensland/7/70, A/Memphis/1/71, A/Memphis/102/72, A/Bangkok/1/79, A/Philippines/2/82, A/Oita/3/83, A/New Jersey/4/85, A/Michigan/1/85, A/Stockholm/4/85, A/Yamagata/96/85, A/Bangkok/2467/85, A/Bangkok/2746/85, A/Tonga/23/85, A/Yokohama/C-5/85, A/Gunma/346/85, A/Yamanashi/497/85, A/Fukuoka/C-29/85. *P* indicated by the closed arrow represents the position of a putative root corresponding to the first main evolutionary stream. *P* indicated by an open arrow represents the putative root belonging to the second main stream. Closed circles indicate the strains belonging to the first stream and open circles are the strains involved in the second stream. Each square shows the position of the HA genes of human H3N2 virus. The abscissa and the ordinate represent the year of isolation and the number of synonymous substitutions of H3N8 influenza virus HA genes used in the test respectively

three minor branches. Two strains of H3N8 virus isolated in England and France, NM/76 and FON/79, were distinguished from the remaining six strains, suggesting that they belong to the third lineage. In contrast, the branch giving rise to SAN/85, KEN/86, and KEN/87 were found to have previously divided off from the lineage containing FRA/67 around mid-1960s. Although the recent two strains ROM/80 and KEN/81 are relatively distantly related to the above three strains and to the NM/76 and FON/79 viruses, evolutionary pathways suggested that the former two strains, together with SAN/85, KEN/86, and KEN/87, are on the fourth evolutionary lineage.

### *Comparison of evolutionary rates*

The phylogenetic tree constructed by the N-J method allows us to estimate the evolutionary rate of the HA genes of equine viruses. The evolutionary rate, based on the synonymous and nonsynonymous substitutions/site/year, was calculated by plotting the year of virus isolation against the branch length to each ancestor node. The HA genes of the strains BRA/87 and LPL/88 (like the T/3/71 strain) were omitted from the calculation because they differ by only eight nucleotides compared with the HA gene of the oldest strain MIA/63 isolated 24–25 years earlier (indicated by closed arrows) (Fig. 3 b). The rate of nucleotide substitutions of the human HA gene is clearly proportional to the year of virus isolations (Fig. 3 c), indicating that these genes evolve steadily in a single lineage. In contrast to this, two regression lines were seen for the HA genes of equine influenza viruses, reflecting the existence of two major evolutionary streams (Fig. 3 b). With the exception of three strains belonging to the first stream (T/3/71, BRA/87, and LPL/88), and one strain (FRA/67) belonging to the second stream, two regression lines showed that the number of nucleotide changes in the equine HA gene is roughly proportional to the year of virus isolation. Because there were two main evolutionary lines, the strains of H3N8

**Table 2.** Comparative analysis of the rate of evolution at the synonymous and nonsynonymous substitutions between HA gene derived from equine and human viruses

| Positions | Rate of evolution | |
|---|---|---|
| | Equine H 3 | Human H 3 |
| Synonymous (silent) | 0.00994–0.00989 | 0.01075 |
| 1st | 0.00177 | 0.00236 |
| 2nd | 0.00024 | 0.00234 |
| 3rd | 0.00530 | 0.00549 |
| Nonsynonymous (amino acid changing) | 0.00041 | 0.00193 |

Number of synonymous and nonsynonymous substitutions per site per year was calculated based on the phylogenetic tree shown in Fig. 3. Accumulation of synonymous substitution in equine and human HA are shown in Fig. 3 b and c

virus including the upper and lower branch clusters were used separately for estimating the evolutionary rate. Accordingly, we estimated the rate of synonymous substitution for the two groups of the equine HA gene to be 0.0094–0.00989 (Table 2). This value is nearly the same as that of the human H 3 HA gene. In contrast, the rates of nonsynonymous substitutions/site/year were estimated to be 0.00041 for the HA gene of equine virus, a value approximately one-fifth of the rate (0.00193) for the human H 3 HA gene. Our phylogenetic analysis of the HA gene of equine H3N8 viruses suggested that unique evolutionary pattern different from that of human and avian influenza A viruses exists for equine influenza viruses.

## Discussion

In order to examine the pattern of evolution for the equine H3N8 influenza virus and to compare it with patterns for other influenza viruses, nucleotide sequences of the HA gene were analyzed based both on pairwise sequence comparisons and on an evolutionary tree. Since the N-J method is reported to find pairs of operational taxonomic units (OTUs) that minimize the sum of branch lengths at each stage of clustering of OTUs beginning with a starlike tree [33], we used this method for construction of the above phylogenetic tree.

Although a number of the influenza A virus nucleotide changes are reported to be directly proportional to the year of virus isolations [3], examination of fifteen equine H 3 HA genes in a pairwise comparison indicated that the substitution of nucleotide and deduced amino acid sequences are not always related to the amount of time between their isolation. For example, the MIA/63 and two strains (BRA/87, LPL/88) isolated 24–25 years apart differ by only eight nucleotides (0.75%). Similarly, the number of nucleotide and amino acid sequences between the HA of the above oldest strain MIA/63 and T/3/71 isolated in Japan was very small. As seen in Fig. 3 b, the T/3/71, BRA/87, and LPL/88 viruses are outlier from the linear regression line. There are two possible explanations: first, the oldest strain (MIA/63) may contaminate the above three strains; second, MIA/63 virus, which showed a striking similarity with T/3/71, BRA/87, and LPL/88, had been frozen during a certain period and recently had been introduced into horse populations. However, it appears unlikely that the H3N8 virus isolated in 1971 in Japan, in 1987 in Brazil and in 1988 in Argentina are laboratory contaminants, because the MIA/63 strain was not handled in the laboratories where the above-mentioned viruses were isolated. With the exception of the T/3/71, BRA/87, and LPL/88 viruses, the nucleotide substitutions of the HA gene of the remaining five strains increased linearly with the time of virus isolation, suggesting the presence of a molecular clock in their HA genes. Although in a recent study the HA of H 3 equine influenza virus is reported to evolve in a single lineage on the basis of phylogenetic tree constructed by the maximum parsimony method [16], our tree showed the existence of two main evolutionary lineages. Each of the two lineages was also found to contain more branch points than previously reported [16]. For ex-

ample, our phylogenetic tree constructed by the N-J method revealed that the seven strains (MIA/63, URG/63, T/3/71, BRA/87, LPL/88, T/2/71, ALG/72) are included in the first main evolutionary lineage, that appeared to contain two branches. The data, based on a pairwise comparison, matched the evolutionary pathways shown by our phylogenetic analysis. The HA of two strains from Japan and Algiers (T/2/71, ALG/72) were shown to be on a branch cluster different from that containing MIA/63, URG/63, T/3/71, BRA/87, LPL/88 because they belong to the second group of the first evolutionary lineage, coinciding to the evidence reported by Kawaoka et al. [16]. It was of interest to note that the H3N8 equine viruses belonging to two distinct branches could cocirculate in the same epizootic (T/3/71 and T/2/71).

In recent years, the nucleotide sequences of HA and nonstructural protein (NS) genes of human influenza C virus were determined and analyzed from an evolutionary point of view [4]. On the basis of pairwise comparison, the extent of nucleotide substitutions in the HA gene of C virus do not appear to accumulate with time [4]. For example, no amino acid substitutions were observed between the HAs of C/Great Lakes/1167/54 and C/Yamagata/10/81 viruses despite the thirty one-years interval between their isolation. Similar evidence was also obtained for the NS gene of influenza C viruses. In this case, the frozen period was sixteen years [5]. These results led us to conclude that human influenza C and equine influenza A viruses may be frozen for a considerable period, suggesting that their evolutionary mechanisms may be similar to each other.

The high degree of genetic similarity (97.2 to 99.3% nucleotide sequence homology) in the HA genes of the FRA/67, NM/76, FON/79, ROM/80, KEN/81, SAN/85, KEN/86, and KEN/87 viruses reflects on their close evolutionary relationships. It is, therefore, interesting to know the branching patterns of these eight strains. As understood from Fig. 3a, they appear to form a second evolutionary lineage. The branch containing NM/76 and FON/79 viruses is considered to divide off from that containing of the remaining six strains. Unlike human influenza A viruses [3, 46], equine influenza viruses are characterized by multiple evolutionary lineages similar to that of influenza B and C viruses [4, 15, 48]. Thus, evolutionary patterns based on the phylogenetic model tree shown in the present study matched the pairwise comparison data.

The construction of a phylogenetic tree by the N-J method allows us to compare the evolutionary rate of the HA genes of equine and human viruses. In recent papers it has been reported for a variety of genes that substitutions at the third codon position occur at a higher rate than for substitutions at the other two codon positions [12, 22, 34]. In agreement with this evidence, the rate of synonymous substitutions of equine HA gene at the third codon position is apparently higher than for the other two codon positions.

Furthermore, nucleotide substitutions at the first and second codon positions frequently led to amino acid substitutions. Coupled with this evidence, it was of interest to find that the HA of equine influenza virus (based on the non-

synonymous substitutions, and nucleotide substitutions at the first and the second codon positions) evolve more slowly than those of human influenza virus. The present study demonstrated that equine influenza viruses have evolved in multiple lineages in a fashion similar to influenza B and C viruses. This evolutionary feature might lead to the recent problem that vaccination is not controlling influenza in horses.

## Acknowledgements

This work was supported by grants from the Ministry of Health and Welfare, Japan. The authors are grateful to Dr. N. J. Cox, Centers for Disease Control, for kindly reading the manuscript. We would particularly like to thank Dr. F. Koide, Division of Veterinary Medicine, Nihon University, for the kindness with which he gave to one (A.E.) of the authors an opportunity to study on influenza virus as a doctoral thesis.

## References

1. Air GM, Giggs AJ, Laver WG, Webster RG (1990) Evolutionary lineages in influenza B are not primarily governed by antibody selection. Proc Nat Acad Sci USA 87: 3884–3888
2. Both GW, Sleigh MJ (1981) Conservation and variation in the haemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J Virol 39: 663–672
3. Both GW, Sleigh MJ, Cox N, Kendal AP (1983) Antigenic drift in influenza virus H 3 haemagglutinin from 1968 to 1980: multiple evolutional pathway and sequential amino acid changes at key antigenic sites. J Virol 48: 52–60
4. Buonaguro DA, Nakada S, Desselberger U, Krystal M, Palese P (1985) Noncumulative sequence changes in the haemagglutinin genes of influenza C virus isolates. Virology 146: 221–232
5. Bunagurio DA, Nakada S, Fitch WM, Palese P (1986) Epidemiology of influenza C virus in man: multiple evolutionary lineages and low rate of change. Virology 153: 12–21
6. Burk F, Lamatsch O (1981) Surveillance for immunity against equine influenza virus infections. Comp Immun Microbiol Infect Dis 4: 267–278
7. Burrows R, Goodrige D, Denyer M, Hutchings G, Frank CJ (1982) Equine influenza infections in Great Britain, 1979. Vet Rec 110: 494–497
8. Daniels RS, Skehel JJ, Wiley DC (1985) Amino acid sequences of haemagglutinins of influenza viruses of the H 3 subtype isolated from horses. J Gen Virol 66: 457–464
9. Donis RO, Bean WJ, Webster RG (1989) Distinct lineages of influenza virus H 4 haemagglutinin genes in different regions of the world. Virology 169: 406–417
10. Fang R, Min Jou W, Huylebroeck D, Devos R, Fiers W (1981) Complete structure of A/duck/Ukraina/1/63 influenza haemagglutinin gene: animal viruses progenitor of human Hong Kong 1968 influenza haemagglutinin. Cell 25: 315–323
11. Gorman OT, Bean EJ, Kawaoka Y, Webster RG (1990) Evolution of the nucleoprotein gene of influenza A virus. J Virol 64: 1487–1497
12. Grunstein M, Schedle P, Kedes L (1976) Sequence analysis and evolution of sea urchin (*Lytechinus pictus* and *Strongylocentrorus purpuratus*) histone H 4 messenger RNAs. J Mol Biol 104: 351–369
13. Hayashida H, Itoh H, Kikuno R, Miyata T (1985) Evolution of influenza virus genes. Mol Biol Evol 2: 389–303
14. Hinshaw VS, Webster RG, Rodriguez RJ (1981) Influenza viruses: combinations of haemagglutinin and neuraminidase subtypes isolated from animals and other sources. Arch Virol 67: 191–201

15. Kanegae Y, Sugita S, Endo A, Ishida M, Senya S, Osaka K, Nerome K, Oya A (1990) Evolutionary pattern of the haemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. J Virol 64: 2860–2865
16. Kawaoka Y, Bean W, Webster RG (1989) Evolution of the haemagglutinin of equine H 3 influenza viruses. Virology 169: 283–292
17. Kawaoka Y, Webster RG (1989) Origin of the haemagglutinin on A/eq/Johannesburg/86 (H3N8): the first known equine influenza outbreak in South Africa. Arch Virol 106: 159–164
18. Kawaoka Y, Yamnikova S, Chambers TM, Lvov DK, Webster RG (1990) Molecular characterization of a new haemagglutinin, subtype H 14, of influenza A virus. Virology 179: 759–767
19. Kida H, Kawaoka Y, Naeve C, Webster RG (1987) Antigenic and genetic conservation of H 3 influenza virus in wild ducks. Virology 159: 109–119
20. Kida H, Shortridge KF, Webster RG (1988) Origin of the haemagglutinin gene of H3N2 influenza from pigs in China. Virology 162: 160–166
21. Kimmel AR, Berger SL (1987) Preparation of cDNA and generation of cDNA libraries. In: Abelson JN, Simon MI (eds) Methods in enzymology, vol 152. Academic Press, London, pp 307–316
22. Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267: 275–276
23. Kumanomido T, Okuda Y, Akiyama T, Hirasawa K, Tabuchi E (1972) Studies on outbreak of equine influenza in Japan in 1971-virus isolation. Exp Res Equine Health Lab Tokyo 9: 35–43
24. Laver WG, Webster RG (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
25. Min Jou W, Verhoeyen M, Devos R, Saman E, Fang R, Huylebroeck D, Fiers W (1980) Complete structure of the haemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell 19: 683–696
26. Nei M (1975) Molecular population genetics and evolution. Elsevier/North-Holland, New York
27. Nei M, Gojubori T (1986) Simple methods for estimating the number of synonymous and nonsynonymous substitutions. Mol Biol Evol 3: 418–426
28. Nerome K, Sakamoto S, Yano N, Yamamoto T, Kobayashi S, Webster RG, Oya A (1983) Antigenic characteristics and genomic composition of naturally occurring recombinant influenza virus isolated from a pig in Japan. J Gen Virol 64: 2611–2620
29. Powell DG, Spooner P, Thompson GR, Plowright W, Burrow R, Schild GC (1974) The outbreak of equine influenza in England 1973. Vet Rec 94: 282–287
30. Powell DG (1975) Equine infections respiratory disease. Vet Rec 96: 300–304
31. Rogers GN, Paulson JC, Daniels RS, Skehel JJ, Wilson IA, Wiley DC (1983) Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. Nature: 76–78
32. Rota PA, Walls TR, Harmon MW, Rota JS, Kendal AP, Nerome K (1990) Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. Virology 175: 59–68
33. Saitou M, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenic trees. Mol Biol Evol 4: 406–425
34. Salser W, Isaacson JS (1976) Mutation rates in globin genes: the genetic load and haldane's dilemma. Prog Nucleic Acids Res Mol Biol 19: 205–220
35. Sambrook J, Fritsch EF, Maniatis T (1989) Commonly used plasmid vector. In: Sambrook J, Fritsch EF, Maniatis T (eds) Molecular cloning, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, p 1.14

36. Scholtissek C, Rohde W, Von Hoyningen V, Rott R (1978) On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology 87: 13–20
37. Shortridge KF, Webster RG, Bufferfield WK, Campbell CH (1976) Persistence of Hong Kong influenza virus variants. Science 196: 1454–1455
38. Sovinova O, Tumova B, Pouska F, Nemec J (1958) Isolation of a virus causing respiratory disease in horses. Acta Virol 2: 52–60
39. Stuart-Harris CH, Schild GC (1976) Influenza viruses of lower animals and birds. In: Stuart-Harris CH, Schild GC (eds) Influenza, the viruses and disease. Edward Arnold, London, pp 83–85
40. Thompson GR, Mumford JA, Spooner PR, Burrow R, Powell DG (1977) The outbreak of equine influenza in England, January 1976. Vet Rec 100: 465–468
41. Uppal PK, Yadav MP (1987) Outbreak of equine influenza in India. Vet Rec 121: 569–570
42. Van Oirshot JT, Masuel N, Huffels ADNHJ, Auker WJJ (1981) Equine influenza in the Netherlands during the winter of 1978–1979: antigenic drift of the A-equi 2 virus. Vet Q 3: 80–84
43. Verhoyen M, Fang R, Min Jou W, Devos R, Huylebroeck D, Saman E, Friers W (1980) Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature 286: 771–776
44. Waddel GH, Teigland MB, Sigel MM (1963) A new influenza virus associated with equine respiratory disease. J Am Vet Med Assoc 143: 587–594
45. Ward CW, Dopheide TA (1981) Evolution of the Hong Kong influenza A subtype. Structural relationship between the haemagglutinin from A/duck/Ukraina/63 (Hav 7) and the Hong Kong (H 3) haemagglutinins. Biochem J 195: 337–340
46. Webster RG, Laver WG, Air GM, Schild GC (1982) Molecular mechanisms of variation in influenza viruses. Nature 296: 115–121
47. Wiley DC, Wilson IA, Skehel JJ (1981) Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289: 373–378
48. Yamashita M, Krystal M, Fitch WM, Palese P (1988) Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. Virology 163: 112–122

Authors' address: Dr. K. Nerome, Department of Virology and Rickettsiology, National Institute of Health, Kamiosaki, Shinagawa-ku, Tokyo 141, Japan.

Received December 5, 1990