

Avian-to-Human Transmission of the PB1 Gene of Influenza A Viruses in the 1957 and 1968 Pandemics

YOSHIHIRO KAWAOKA,* SCOTT KRAUSS, AND ROBERT G. WEBSTER

Department of Virology/Molecular Biology, St. Jude Children's Research Hospital, 332 North Lauderdale, P.O. Box 318, Memphis, Tennessee 38101-0318

Received 24 April 1989/Accepted 14 July 1989

We determined the origin and evolutionary pathways of the PB1 genes of influenza A viruses responsible for the 1957 and 1968 human pandemics and obtained information on the variable or conserved region of the PB1 protein. The evolutionary tree constructed from nucleotide sequences suggested the following: (i) the PB1 gene of the 1957 human pandemic strain, A/Singapore/1/57 (H2N2), was probably introduced from avian species and was maintained in humans until 1968; (ii) in the 1968 pandemic strain, A/NT/60/68 (H3N2), the PB1 gene was not derived from the previously circulating virus in humans but probably from another avian virus; and (iii) a current human H3N2 virus inherited the PB1 gene from an A/NT/60/68-like virus. Nucleotide sequence analysis also showed that the avian PB1 gene was introduced into pigs. Hence, transmission of the PB1 gene from avian to mammalian species is a relatively frequent event. Comparative analysis of deduced amino acid sequences disclosed highly conserved regions in PB1 proteins, which may be key structures required for PB1 activities.

There are 13 hemagglutinin (HA) and 9 neuraminidase (NA) subtypes of influenza A viruses. Most of these subtypes have been recognized in wild avian species (12), but only three HAs and two NAs (H1, H2, H3, N1, and N2) have been recognized in human influenza viruses. The appearance of new subtypes of influenza A viruses can lead to pandemics in humans. For example, the virus A/Singapore/1/57 (H2N2) (Sing/57), whose HA and NA differed from those of previously circulating viruses, emerged in the 1957 pandemic. Another virus, A/Hong Kong/1/68 (H3N2), with a different HA subtype from the 1957 pandemic virus, was isolated in 1968. The genes responsible for the emergence of these new surface glycoproteins are thought to have been introduced from avian species (24), suggesting genetic reassortment between human and avian influenza viruses. Most other genes of the 1957 and 1968 pandemic strains were derived from previously circulating human viruses. One exception is the PB1 gene of Sing/57, whose origin could not be identified by RNA-RNA hybridization (18).

The PB1 protein of influenza A virus has been shown to form a complex with the other viral polymerase gene products PB2 and PA (10) and to be involved in initiation of transcription and chain elongation (7, 22, 23). Snyder et al. (20) have shown that reassortant viruses with avian PB1 and human PA replicate poorly in MDCK cells and in squirrel monkeys but replicate well in avian tissue culture. Thus, the combination of avian PB1 and human PA may not be conducive to interaction with mammalian host cell factors because of amino acid sequence differences between the two types of proteins.

In contrast to convincing evidence that influenza virus genes for surface glycoproteins can be introduced from avian species into pigs (16) and humans (18), essentially nothing is known about interspecies introduction, species specificities, and evolution of the polymerase genes of this virus. Our aim in the present study was to determine the origin of the PB1 gene of the 1957 pandemic influenza virus, its evolution, and

its species-specific amino acid residues that may be important in determining host specificity.

MATERIALS AND METHODS

Viruses and viral RNA. The viruses used in this study were A/Beijing/11/56 (H1N1) (Beij/56), Sing/57, A/Wisconsin/3523/88 (H1N1) (WI/88), A/Swine/Tennessee/26/77 (H1N1) (Sw/TN/77), A/Swine/Ontario/2/81 (H1N1) (Sw/Ont/81), A/Swine/HongKong/126/82 (H3N2) (Sw/HK/82), A/Equine/London/1416/73 (H7N7) (Eq/Lond/73), A/Equine/Tennessee/5/86 (H3N8) (Eq/TN/86), A/Turkey/Minnesota/833/80 (H4N2) (Ty/MN/80), and A/Gull/Maryland/704/77 (H13N6) (Gull/MD/77). WI/88, obtained from M. Harmon (Centers for Disease Control, Atlanta, Ga.), was isolated from a 32-year-old woman in Wisconsin who died of primary viral pneumonia shortly after giving birth.

Viruses were grown in 11-day-old embryonated chicken eggs and purified by differential sedimentation through 25 to 70% sucrose gradients in a Beckman SW28 rotor. Virion RNA was isolated by treatment of purified virus with proteinase K and sodium dodecyl sulfate followed by extraction with phenol-chloroform (1:1), as described previously (5).

Cloning of the PB1 gene. Full-length cDNA was prepared by reverse transcription of virion RNA by previously described methods (14). A 12-base synthetic primer complementary to the 3' terminus of the negative-strand RNA was phosphorylated with T4 polynucleotide kinase and then used to prime reverse transcription of the total virion RNA in the presence of [α -³²P]dATP. Second-strand DNA synthesis was performed with a phosphorylated 13-base synthetic primer complementary to the 3' end of the cDNA and the Klenow fragment of *Escherichia coli* DNA polymerase I. Full-length double-stranded copies of the PB1 gene were blunt-end ligated into the *Pvu*II site of pATX, obtained from C. Naeve (St. Jude Children's Research Hospital, Memphis, Tenn.).

Nucleic acid sequencing. Nucleotides of the cloned PB1 genes were sequenced by the method of Chen and Seeburg (8) by using alkali-denatured DNA templates. Oligonucleotide primers complementary to the PB1 gene segment were

* Corresponding author.

synthesized on an Applied Biosystems model 280A DNA synthesizer by the solid-phase phosphoramidite method. The reaction products were resolved on 6% polyacrylamide-7 M urea thin gels containing a 1× to 5× TBE (90 mM Tris-borate [pH 8.0], 1 mM EDTA) gradient. The sequences of the oligonucleotides used as primers will be provided on request. Nucleotide sequences were analyzed by the maximum-parsimony method to determine the minimum number of mutations needed to account for sequence differences (11).

RESULTS

Introduction of avian PB1 gene into humans and pigs. To determine the relatedness of the influenza virus PB1 genes from different species, we cloned the genes into a plasmid and sequenced them by the chain termination method. Each PB1 gene consisted of 2,341 nucleotides; there were no deletions or insertions. These sequences have been entered in the GenBank database as accession numbers m25924 to m25936. Table 1 shows a relatively high degree of nucleotide sequence homology (82 to 99%) among PB1 genes from different influenza A viruses compared with HA genes of different subtypes (26 to 80%) (1). When these PB1 genes were compared with those of influenza B virus (15), approximately 60% homology was detected.

The nucleotide sequences were analyzed by the maximum-parsimony method, and an evolutionary tree was constructed (Fig. 1). The diagram indicates four distinct lineages for the PB1 gene, the first comprising avian (Gull/MD/77, A/Mallard/New York [H2N2] [Mal/NY/78], and Ty/MN/80), human H2N2 (Sing/57, A/Ann Arbor/6/60 [H2N2] [AA/60], and A/Korea/426/68 [Korea/68]) and H3N2 (A/Northern Territory/60/68 [H3N2] [NT/68] and A/Memphis/8/88 [Mem/88]), and swine H3N2 (Sw/HK/82) viruses; the second comprising equine viruses (Eq/Lond/73 and Eq/TN/86); the third comprising H1N1 human viruses (A/Puerto Rico/8/34 [H1N1] [PR8/34] and Beij/56); and the fourth comprising H1N1 swine (Sw/TN/77 and Sw/Ont/81) and H1N1 human (WI/88) viruses. The PB1 genes of H2N2 and H3N2 human influenza viruses are more closely related to those of avian viruses than to those of the H1N1 viruses that were previously circulating in humans (Table 1 and Fig. 1). This indicates that the avian PB1 as well as the HA and NA genes (18) were introduced into humans prior to the 1957 pandemic. The PB1 gene of the late H2N2 viruses (Korea/68) arose directly from the Sing/57-like PB1. The PB1 genes of human H3N2 viruses (NT/68 and Mem/88) are also more closely related to those of avian viruses than to those of H1N1 human viruses. However, they belong to a different lineage than do those of human H2N2 viruses (Sing/57, AA/60, and Korea/68). The evolutionary tree clearly illustrates that the PB1 gene of the early human H3N2 virus (NT/68) was not derived from that of the previously circulating H2N2 virus, e.g., Korea/68. The current human H3N2 virus (Mem/88) inherited the PB1 gene from the early H3N2 virus, e.g., NT/68.

These results strongly suggest that the PB1 gene was introduced from avian species into humans prior to the 1957 pandemic and that it was maintained in humans until 1968. During generation of the 1968 pandemic strain, however, a new avian PB1 gene was introduced into humans and has been maintained since that time.

That the PB1 gene of H3N2 swine virus (Sw/HK/82) is more closely related to the avian PB1 gene than to the H1N1 swine PB1 genes (Sw/TN/77 and Sw/Ont/81) (Fig. 1) pro-

TABLE 1. Nucleotide and amino acid sequence homologies among the PB1 genes of human, swine, avian, and equine influenza A viruses

Strain	% Homology of PB1 ^a																	
	Sing/57	AA/60	Korea/68	NT/68	Mem/88	Gull/MD/77	Mal/NY/78	Ty/MN/80	Sw/HK/82	Eq/Lond/73	Eq/TN/86	PR8/34	Beij/56	Sw/TN/77	Sw/Ont/81	WI/88	B/Lee/40	
Sing/57	99.5																	
AA/60	99.1	99.4																
Korea/68	99.1	98.8	97.9															
NT/68	98.5	98.3	97.9	98.3														
Mem/88	98.2	97.9	97.5	98.3	98.3													
Gull/MD/77	99.1	98.9	98.4	98.9	98.4	99.7												
Mal/NY/78	99.2	98.9	98.5	99.1	98.4	99.7	99.7											
Ty/MN/80	99.1	98.8	98.4	98.9	98.3	99.7	99.3	99.3										
Sw/HK/82	98.7	98.4	98.0	98.5	97.9	99.3	99.3	99.3	90.7									
Eq/Lond/73	97.1	96.8	96.6	96.7	96.3	97.6	97.4	97.2	96.8	83.1								
Eq/TN/86	97.4	97.1	96.8	97.0	96.6	97.8	97.6	97.2	96.8	97.2	97.2							
PR8/34	96.7	96.4	96.0	96.6	96.0	97.1	97.2	97.4	96.7	95.5	95.5	83.7						
Beij/56	96.2	95.9	95.6	96.0	96.0	96.8	96.8	97.1	96.4	94.6	94.8	96.4	94.4					
Sw/TN/77	95.8	95.5	95.2	95.5	95.5	96.2	96.3	96.3	95.8	94.2	94.5	94.5	95.0	87.2				
Sw/Ont/81	95.6	95.4	95.1	95.1	95.0	95.8	95.9	95.9	95.4	94.1	94.2	94.3	94.7	98.8	94.8			
WI/88	95.2	95.0	94.7	95.0	94.8	95.8	95.6	95.6	95.1	94.3	94.5	94.3	94.5	98.3	99.2			
B/Lee/40	59.3	59.4	59.6	58.8	58.9	59.7	59.4	59.6	59.4	58.3	58.5	58.8	59.0	59.2	59.3	59.3		

^a Nucleic acid sequence data are presented in the top right triangle, and amino acid sequence data are in the bottom left triangle.

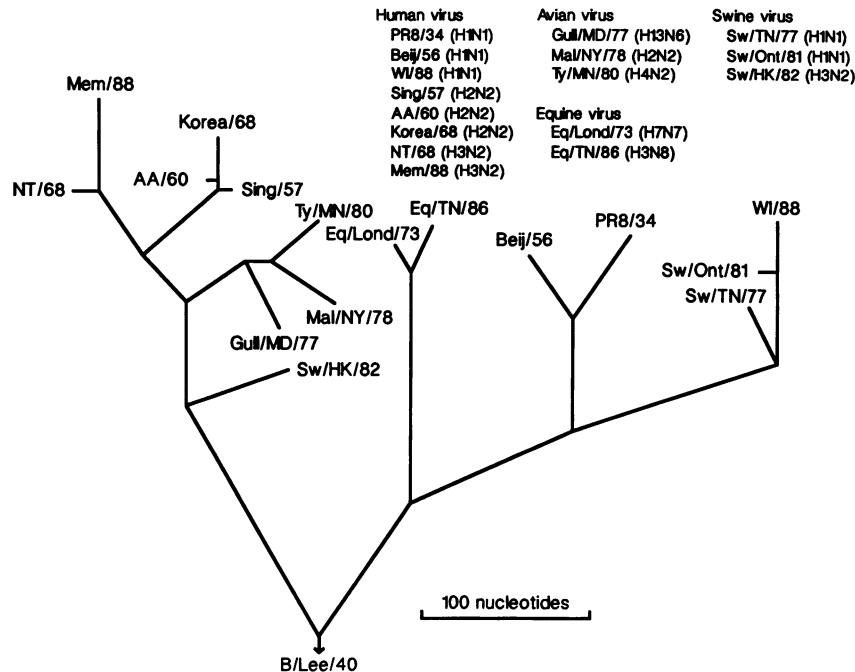


FIG. 1. Evolutionary tree for PB1 genes. The nucleotide sequences were analyzed, and the minimal tree length was determined by using the branch-and-bound algorithm of PAUP (David Swafford, Illinois Natural History Survey). The lengths of the trunk and side branches of the diagram are proportional to the number of substitutions required to account for the differences in sequences. The nucleotide sequences of the B/Lee/40 PB1 gene were aligned according to Kenedi et al. (15) and used to define the root of the tree. Nucleotide sequences for PR8/34 (H1N1) (25), NT/68 (H3N2) (6), AA/60 (H2N2) (9), and Mal/NY/78 (J. Treanor, Y. Kawaoka, R. Miller, R. G. Webster, and B. Murphy, *Virus Res.*, in press) are from published sources.

vides further evidence of introduction of the avian gene into mammalian species and supports data showing that the HA gene of this swine virus was recently introduced from avian influenza viruses (16). The WI/88 PB1 gene is closely related to the H1N1 swine PB1 gene (Sw/TN/77 and Sw/Ont/81), confirming direct introduction of a swine influenza virus gene into a human (P. Rota et al., unpublished antigenic and oligonucleotide mapping analyses).

Determination of functionally important regions in the PB1 protein. To determine the functionally important regions in the PB1 protein, we aligned the deduced amino acid sequences of the influenza A virus PB1 genes from different hosts (Fig. 2). All of the PB1 genes have the potential to encode a polypeptide of 757 amino acids. Comparison of the amino acid sequences showed a high degree of homology among all of the influenza A virus PB1 genes; although nucleotide sequence homology was as low as 82%, more than 94% homology was observed in the amino acid sequences. For the 757 amino acid residues of the PB1, amino acid substitutions were observed at 90 positions. Some of these were clustered; for instance, in the regions of residues 152 to 182 and 358 to 401, more than 30% of the residues were variable. By contrast, the rest of the regions were relatively conserved. Some of these regions are also highly conserved by comparison with the corresponding regions of the PB1 gene of influenza B virus (15), and three of these conserved regions (residues 224 to 243, 266 to 282, and 402 to 421) are relatively wide.

Results of the nucleotide sequence analysis suggest that the PB1 genes of Sing/57, WI/88, and Sw/HK/82 were recently introduced from other species, providing a unique opportunity to identify amino acid substitutions required for adaptation to new hosts. Most of the amino acid substitu-

tions in these viruses do not appear to be important for adaptation because there was no correlation between the host species and the substituted amino acid residues. However, an amino acid substitution that occurred in the WI/88 PB1 at position 375 might be critical for adaptation in the new species. The PB1 of WI/88, which seems to result from the direct introduction of a swine influenza virus into a human, has serine at amino acid residue 375 as in all other human PB1 proteins except those of Beij/56, whereas other H1N1 swine PB1 proteins have glycine at position 375 (Fig. 2). Furthermore, this is one of the positions at which more than three different amino acids are found in the different PB1 proteins. There are other amino acid substitutions found in the WI/88 PB1 compared with swine PB1 proteins (e.g., at positions 175, 214, and 435). However, they do not seem to be important for adaptation of the virus to replicate in humans; the amino acid residues found in these positions differ between the PB1 proteins of WI/88 and other human viruses.

The observations described above show that there are highly conserved regions in the PB1 protein that may be functionally important and that the amino acid residues at 375 might be critical for host specificity of the protein.

DISCUSSION

The influenza virus genes encoding surface glycoproteins can be introduced from avian into mammalian species (16, 18), but this route of transmission has not been demonstrated for genes encoding internal proteins. Results of the present study suggest an avian origin for the PB1 genes of human influenza A viruses in the 1957 and 1968 pandemics and confirm a previous study in which the PB1 gene of WI/88 was

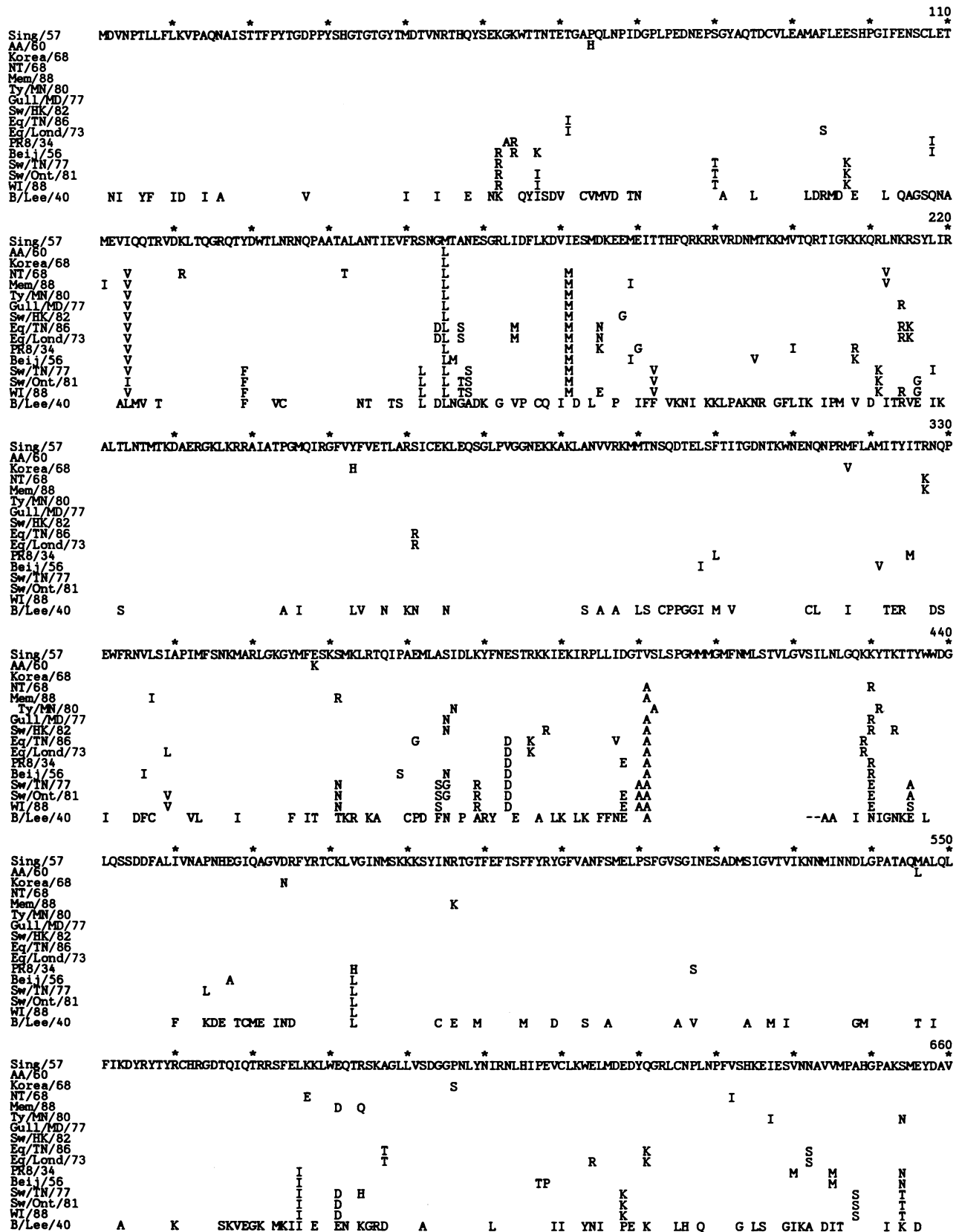


FIG. 2. Amino acid sequences of the PB1 proteins of influenza viruses from different hosts. The amino acid sequences were deduced from the nucleotide sequences and are written in full for Sing/57 by using the single-letter amino acid code. Only amino acid substitutions are given for the other strains. Dashes are included to adjust the alignment for B/Lee/40.

4. **Bean, W. J.** 1984. Correlation of influenza A virus nucleoprotein genes with host species. *Virology* **133**:438–442.
5. **Bean, W. J., G. Sriram, and R. G. Webster.** 1980. Electrophoretic analysis of iodine-labeled influenza virus RNA segments. *Anal. Biochem.* **102**:228–232.
6. **Bishop, D. H. L., J. A. Huddleston, and G. G. Brownlee.** 1982. The complete sequences of RNA segment 2 of influenza A/NT/60/68 and its encoded P1 protein. *Nucleic Acids Res.* **10**:1335–1343.
7. **Braam, J., I. Ulmanen, and R. M. Krug.** 1983. Molecular model of a eucaryotic transcription complex: function and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34**:609–618.
8. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoiled sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**:165–170.
9. **Cox, N. J., F. Kitame, A. P. Kendal, H. F. Maassab, and C. Naeve.** 1988. Identification of sequence changes in the cold-adapted, live attenuated influenza vaccine strain, A/Ann Arbor/6/60 (H2N2). *Virology* **167**:554–567.
10. **Detjen, B. M., C. St. Angelo, M. G. Katze, and R. M. Krug.** 1987. The three influenza virus polymerase (P) proteins not associated with viral nucleocapsids in the infected cell are in the form of a complex. *J. Virol.* **61**:16–22.
11. **Fitch, W. M.** 1971. Toward defining the course of evolution minimum change from a specific tree topology. *Syst. Zool.* **20**:406–415.
12. **Hinshaw, V. S., R. G. Webster, and R. J. Rodriguez.** 1981. Influenza A viruses: combinations of hemagglutinin and neuraminidase subtypes isolated from animals and other sources. *Arch. Virol.* **67**:191–201.
13. **Honda, A., K. Ueda, K. Nagata, and A. Ishihama.** 1988. RNA polymerase of influenza virus: role of NP in RNA chain elongation. *J. Biochem.* **104**:1021–1026.
14. **Huddleston, J. A., and G. G. Brownlee.** 1982. The sequence of the nucleoprotein gene of human influenza A virus strain, A/NT/60/68. *Nucleic Acids Res.* **10**:1029–1037.
- 14a. **Kawaoka, Y., W. J. Bean, and R. G. Webster.** 1989. Evolution of the hemagglutinin of equine H3 influenza viruses. *Virology* **169**:283–292.
15. **Kemdirim, S., J. Palefsky, and D. J. Briedis.** 1986. Influenza B virus PB1 protein: nucleotide sequence of the genome RNA segment predicts a high degree of structural homology with the corresponding influenza A virus polymerase protein. *Virology* **152**:126–135.
16. **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.
17. **Nakajima, K., E. Nobusawa, and S. Nakajima.** 1984. Genetic relatedness between A/Swine/Iowa/15/30 (H1N1) and human influenza viruses. *Virology* **139**:194–198.
18. **Scholtissek, C., W. Rohde, V. Von Hoyningen, and R. Rott.** 1978. On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* **87**:13–20.
19. **Shapiro, G. I., and R. M. Krug.** 1988. Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. *J. Virol.* **62**:2285–2290.
20. **Snyder, M. H., A. J. Buckler-White, W. T. London, E. L. Tierney, and B. R. Murphy.** 1987. The avian influenza virus nucleoprotein gene and a specific constellation of avian and human virus polymerase genes each specify attenuation of avian-human influenza A/Pintail/79 reassortant viruses for monkeys. *J. Virol.* **61**:2857–2863.
21. **Tian, S. F., A. J. Buckler-White, W. T. London, L. J. Reck, R. M. Chanock, and B. R. Murphy.** 1985. Nucleoproteins and membrane protein genes are associated with restriction of replication of influenza A/Mallard/NY/78 virus and its reassortants in squirrel monkey respiratory tract. *J. Virol.* **53**:771–775.
22. **Ulmanen, I., B. A. Broni, and R. M. Krug.** 1981. The role of two of the influenza virus core P proteins in recognizing cap 1 structures (m⁷GpppNm) on RNAs and in initiation of viral RNA transcription. *Proc. Natl. Acad. Sci. USA* **78**:7355–7359.
23. **Ulmanen, I., B. A. Broni, and R. M. Krug.** 1983. Influenza virus temperature-sensitive cap (m⁷GpppNm)-dependent endonuclease. *J. Virol.* **45**:27–35.
24. **Webster, R. G., and W. G. Laver.** 1972. The origin of pandemic influenza. *Bull. W.H.O.* **47**:449–452.
25. **Winter, G., and S. Field.** 1982. Nucleotide sequence of human influenza A/PR/8/34 segment 2. *Nucleic Acids Res.* **10**:2135–2143.