Letter to the Editor

Phylogenetic Analysis of Nucleoproteins Suggests That Human Influenza A Viruses Emerged from a 19th-Century Avian Ancestor[1]

M. Gammelin, A. Altmüller,* U. Reinhardt,* J. Mandler,* V. R. Harley,† P. J. Hudson,† W. M. Fitch,‡[2] and C. Scholtissek**

*Institut für Virologie, Justus-Liebig-Universität Giessen; †CSIRO, Division of Biotechnology, Parkville, Victoria, Australia; and ‡Department of Biological Sciences, University of Southern California

The nucleoprotein (NP) is one of the main determinants of species specificity of influenza A viruses. Using 25 NP sequences we have constructed evolutionary trees by the strict-parsimony procedure of Fitch (1971). In contrast to the evolutionary-gene tree, the tree based on amino acid sequences unravels remarkable differences between avian and human NPs, differences which are best explained by a strong differential selection pressure on the human NPs. It is speculated that this selection pressure is caused by a change of the host and the (T-cell) immune response. A cautious extrapolation of the tree suggests that the human influenza A virus NPs evolved ~ 50 years ago from an avian ancestor. Furthermore, the ancestral relation between the NPs of influenza A, B, and C viruses was analyzed.

The influenza A virus has a genome consisting of eight RNA segments. Its NP gene seems to play the major role in host range: (1) Rescue of NP ts mutants of fowl plague virus (H7N1, FPV) in chicken cells is possible by avian but not by human H3N2 strains. However, the formation of FPV reassortants with the NP gene of the human Hong Kong virus (H3N2) can easily be achieved in dog kidney cells. These FPV reassortants with the Hong Kong NP do not multiply in chicken cells and are nonpathogenic for chickens (Scholtissek et al. 1978, 1985). Replacement of the other FPV genes by those of the human Hong Kong virus leads to FPV reassortants that replicate well in chicken cells (Scholtissek et al. 1976). Thus, by specific replacement of the NP gene the host range has been changed. (2) This view is strengthened by the observation that the sole replacement of the NP gene of a human influenza A strain by the NP gene of an avian virus is sufficient for an efficient attenuation of the human virus for monkeys (Snyder et al. 1987).

In nature, avian influenza A viruses do not spread in the human population, and human viruses do not spread in birds. In the laboratory, however, ducks can become infected artificially by human H3N2 viruses. These viruses induce relatively high antibody titers in the birds, but the viruses are not excreted. In contrast, ducks infected with duck H3N2 viruses induce only very low antibody titers and excrete these viruses (Scholtissek et al. 1985), which in this way can spread in the duck population.

The NP genes—and not other genes—are responsible for keeping the two large separately evolving reservoirs of influenza A viruses—that in humans and that in water birds (Hinshaw and Webster 1982; Kawaoka et al. 1988)—apart. Therefore it was of interest to search for a common ancestor of these viruses. Furthermore, we wanted to know to what extent the NPs of influenza B and C viruses were related to the NP of influenza A viruses. A thorough comparison of sequences of the NPs of many strains should give an answer.

The sequences of many NP genes of influenza A viruses have been determined (Winter and Fields 1981; Huddleston and Brownlee 1982; van Rumpuy et al. 1982;

1. Key words: influenza virus, birds, humans, phylogeny, nucleoprotein, evolution.

2. Present address: Department of Ecology and Evolutionary Biology, University of California, Irvine.

Address for correspondence and reprints: Dr. C. Scholtissek, Institut für Virologie, Justus-Liebig-Universität Giessen, D-6300 Giessen, West Germany.

Mol. Biol. Evol. 7(2):194–200. 1990.

0737-4038/90/0702-0006$02.00

Tomley and Roditi 1984; Steuler et al. 1985; Buckler-White and Murphy 1986; Reinhardt and Scholtissek 1988; Altmüller et al. 1989; Gammelin et al. 1989; Li et al. 1989; Mandler and Scholtissek 1989). Those NP sequences not listed in these references were obtained by the methods described by Harley et al. (1989) (SHRW; EMBL accession no. X16068) or Gammelin et al. (1989) (SWIT; EMBL accession no. X16067). The influenza A strains belonging to different H/N subtypes and isolated from different species as listed in the legend to figure 1 were investigated.

The NPs of human influenza B and C viruses were as follows: B/Lee/40 (Briedis and Tobin 1984), B/Ann Arbor/1/66 (B/AA/66; DeBorde et al. 1988), B/Singapore/222/79 (B/Sing/79; Londo et al. 1983), B/Ann Arbor/1/86 (B/AA/86; Rota 1989), and C/California/78 (C/Cal/78; Nakada et al. 1984).

Evolutionary trees were constructed by means of the parsimony methods of Fitch (1971) and Fitch and Farris (1974). More details are given in the legends to figures 1 and 2.

The top panel of figure 1 shows a phylogenetic tree based on the complete sequences of 25 NP genes of influenza A viruses belonging to different glycoprotein (H and N) subtypes and isolated from different species at different times. Two main branches can be differentiated, one including one pig and all human and the other one including all avian strains, another pig virus, and the mink isolate. Two minor branches consist of the two equine strains. This observation is consistent with Bean's (1984) classification of the NP genes according to his hybridization data. Apart from the human portion of the tree, there is not the obvious correlation between year of isolation and genetic relatedness that has been found with strictly human influenza viruses isolated in consecutive years (with the exception of the reappearance of the H1N1 strains in 1977) (Raymond et al. 1983, 1986; Buonagurio et al. 1986; Altmüller et al. 1989). The failure of correlation between year of isolation and position at the upper (avian) part of the tree is difficult to explain. The heterogeneity in species with different rates of replication might play a role. For example, the NPs of NT 60 and SW 6 (fig. 1, top) are closest related, although they were isolated 8 years apart. Furthermore, the geographic distribution of animals might be an important factor. Thus, the mallard duck strain from the United States is relatively different from the other avian strains. A similar observation on low genetic mixing between avian influenza viruses of different geographic regions was made concerning the P protein genes (von Hoyningen-Huene and Scholtissek 1983) and H4 genes (Donis et al. 1989). This could indicate that the flight routes of migratory birds play an important role in the distribution of influenza virus genes.

When a phylogenetic tree of the 25 NPs was constructed at the level of amino acid sequences, a somewhat different picture emerged (fig. 2). There are some minor differences in that the California 78 strain is now closer to the Texas 77 than to the Hong Kong 83 isolate. Furthermore, the classical swine virus from 1931 now clusters with both sw 127 from 1982 and swine Italy from 1981. There are also some minor differences concerning the avian branch. The human side of the protein tree (lower half of fig. 2) extends rightward just as the nucleic acid tree did. However, the horse-bird side of the tree (upper half of fig. 2) simply does not show the same type of extension. The topology of the protein and nucleic acid tree is essentially the same, but the bird sequences are behaving very strangely. All seven other sequences above Mal 78 and FPV 34 descend from a common ancestor identical in sequence to that of duck Bavaria 77, including virus N 49. Thus seven amino acid sequences of strains isolated between 1949 and 1984 (35 years) would all appear to have evolved very little and nearly independently from the same sequence. The nucleic acid tree (fig. 1, top) demonstrates this is not so (i.e., they did not evolve independently and they did mutate as fast as the lower half—but mainly for silent substitutions). The swine group also seems to be slowly evolving compared with the human branch. One explanation for this perseverance of viral protein structure might be that the structure of the NP is

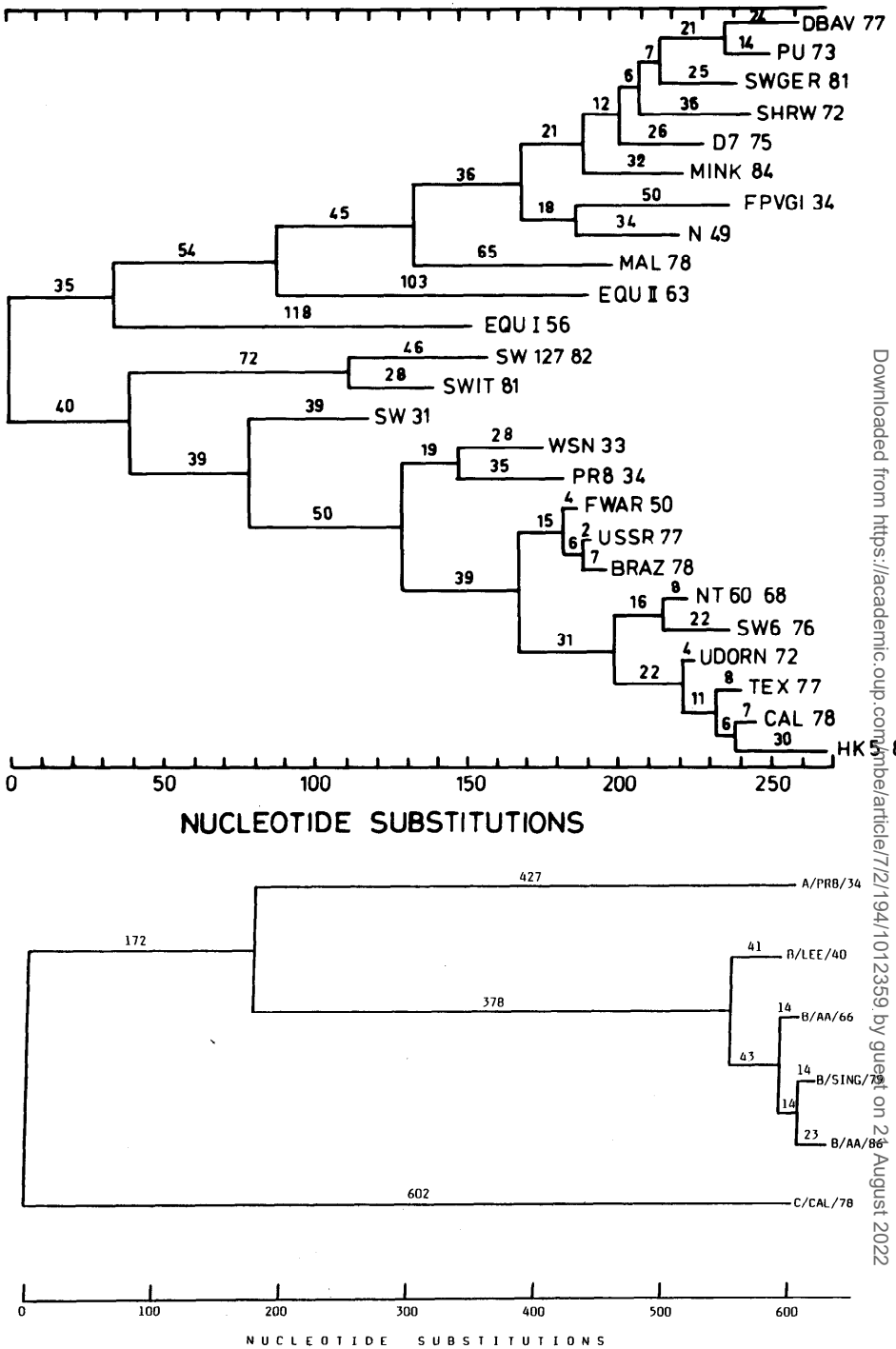

FIG. 1.—*Top,* Phylogenetic tree of 25 NP genes of influenza A viruses derived from the coding region of the nucleotide sequences. The tree is the most parsimonious of ~100 examined. Numbers (uncorrected) on the tree are the number of substitutions required by the strict parsimony procedure of Fitch (1971). The strain designations are as follows: WSN = A/WSN/33, H1N1; PR8 = A/PR/8/34, H1N1; FWAR = A/Fort Warren/1/50, H1N1; USSR = A/USSR/90/77, H1N1; BRAZ = A/Brazil/11/78, H1N1; NT60 = A/NT/60/68, H3N2; UDORN = A/Udorn/307/72, H3N2; TEX = A/Texas/1/77, H3N2; Cal = A/California/10/78, H1N1; HK5 = A/Hong Kong/5/83, H3N2; SW6 = A/Swine/Hong Kong/6/76, H3N2; SWIT = A/

optimal and that there is no selection pressure. However, it is the more remarkable in that the human sequences behave differently. Therefore a plausible and important explanation to resolve these differences would be that, in contrast to the avian NPs, the human NPs are under a strong selection pressure. There must be a memory concerning the selection pressure, since the development is unidirectional. It might be speculated that this selection pressure is caused by changing the host and/or by the immune defense of the host, possibly by T-cell receptor/class I MHC recognition (Townsend et al. 1986). Nonvirulent avian influenza viruses are enterotropic. Therefore there is no significant immune response and thus no selection pressure. (Chicken infected with pantropic fowl plague viruses do not survive an infection, and pigs kept in farms seldom survive longer than a year.) This differential selection pressure is consistent with the divergence of the NPs into two subtypes (Gammelin et al. 1989), which results in the incompatibility of the NPs during reassortment and ensuing species specificity.

A cautious calculation from the phylogenetic tree of figure 2 suggests the common ancestor of the human and animal influenza A virus NPs might have existed only ~150 years ago: if we assume that the avian NP has reached its optimum and is no longer under a selection pressure, we can take the duck Bavaria 77 isolate as prototype strain. The WSN strain isolated in 1933 from a human exhibits 25 amino acid replacements when compared with DBAV 77, while the human Hong Kong isolate from 1983 (HK5) has 40 replacements. The replacements of the NPs of all the other human strains, including the A/Ann Arbor/6/60 (H2N2) strain from 1960 (Cox et al. 1988) and the A/Kiev/59/79 (H1N1) strain from 1979 (Beklemishe et al. 1986), fall on a straight line. A linear-regression analysis (least-square method) of the year of isolation versus the amino acid differences of the human isolates in respect to DBAV 77 extrapolates the common ancestor to the year 1837. Around this time the common ancestor should have existed, if it is assumed that the mutation rate in the first 100 years was the same as during the final 50 years.

According to the bottom panel of figure 1, influenza B and C viruses, after a corresponding alignment (Nakada et al. 1984), have a common root with influenza A viruses. Influenza B and—apart from a rare isolation from pigs (Guo et al. 1983)—influenza C viruses were found only in humans. These viruses might have emerged also from an avian influenza A ancestor a correspondingly long time ago and, by the selection pressure specific for humans, might have developed finally their own type.

Although influenza is known from medieval times, the rate of evolution appears sufficiently great and the human and avian sequences appear sufficiently close that it is likely that a xenologous exchange (interspecies NP gene transfer) has been occurring slowly for some time. In this context an observation of some kind of "biological evolution" of human H1N1 strains might be significant. All human H1N1 strains up to the USSR/77 strain are able to rescue the FPV-NP ts mutants in chicken embryo cells, although at a much lower frequency than do avian strains, while all later H1N1 isolates, including the Brazil/78 strain, cannot (Altmüller et al. 1989). This would imply that, from time to time, whole branches of human influenza A virus NPs reconstitute their own type. The other viral genes—the gene products of which have to

Swine/Italy/147/81, H1N1; SW127 = A/Swine/Hong Kong/127/82, H3N2; SW = A/Swine/1976/31, H1N1; SWGER = A/Swine/Germany/2/81, H1N1; DBAV = A/Duck/Bavaria/2/77, H1N1; D7 = A/Duck/Hong Kong/7/75, H3N2; PU = A/Parrot/Ulster/73, H7N1; SHRW = A/Shearwater/Australia/72, H6N5; N = A/chicken/Germany "N"/49, H10N7; FPVGi = A/FPV/Rostock/34 Giessen-isolate, H7N1; MINK = A/Mink/Sweden/84, H10N4; MAL = A/Mallard/NY/6750/78, H2N2; EQUII = A/Equi/Miami/1/63, H3N8; EQUI = A/Equi/Prague/56, H7N7. The year of isolation is included with the strains. *Bottom,* Phylogenetic tree of NP genes of influenza A, B, and C viruses, based on the alignment by Nakada et al. (1984). For further details of the procedure, see top panel. The mutation rate for influenza A NP genes is 3.3 nucleotide substitutions/year, while the corresponding value for B viruses is 1.15 substitutions/year. For strain designation, see text.

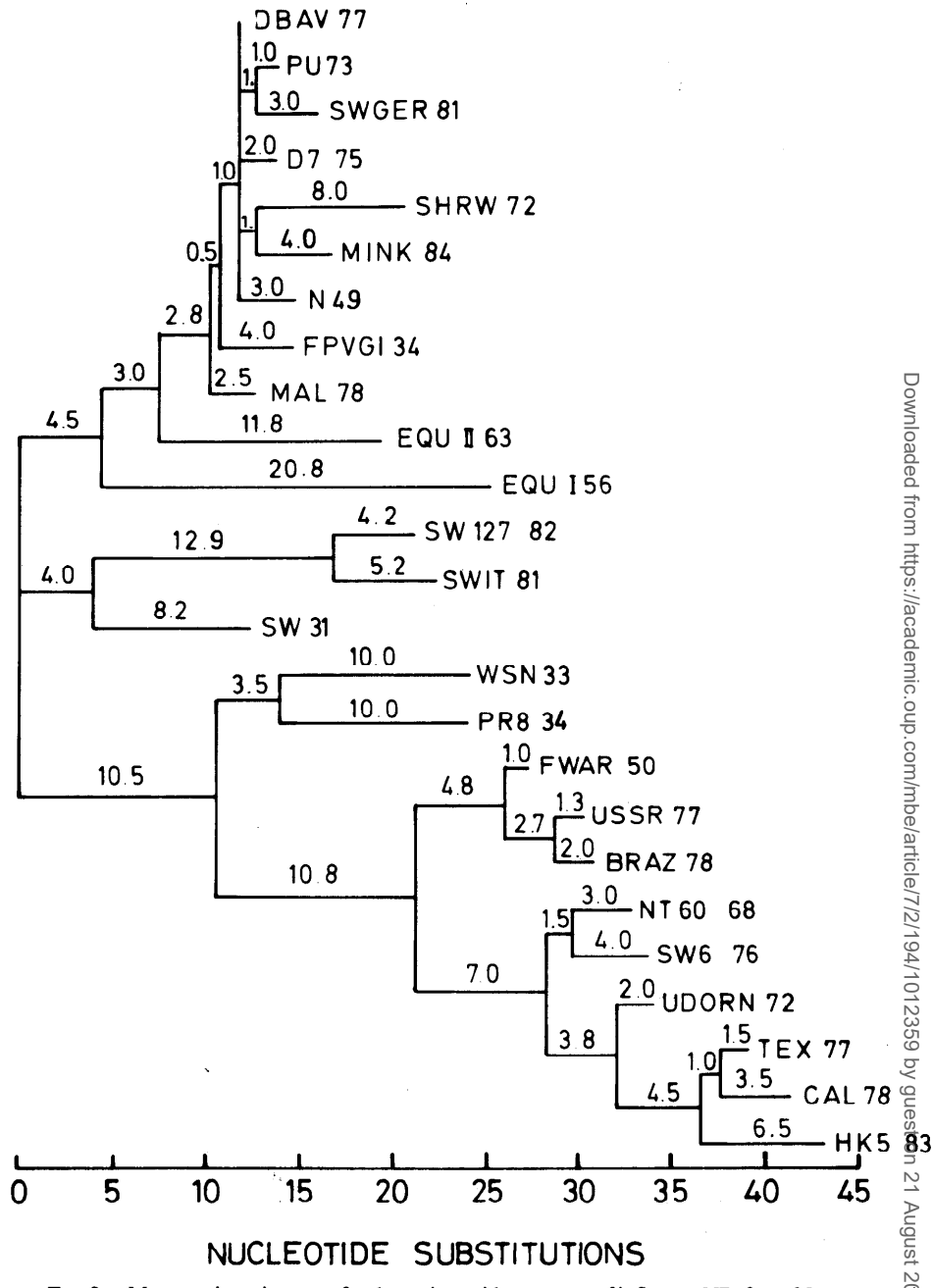

FIG. 2.—Most parsimonious tree for the amino acid sequences of influenza NPs from 25 type A strains. The sequences were analyzed by the method of Fitch (1971) and Fitch and Farris (1974), which obtains the most parsimonious tree in terms of nucleotide substitutions by back-translating the changed amino acids into ambiguous codons. The tree is shown unrooted because the three major lineages (humans, swine, and others) are of nearly equal depth (the greater depth of the human lineage is the result of a process discussed in the text). Numbers on the branches are the number of nucleotide substitutions in the interval, averaged over all possible ways of distributing the 203 substitutions. Horizontal lengths are proportional to the number of substitutions on the branches. Strain designations are identical to those in the top panel of fig. 1. The five avian (or groups of avian) descendants from the ancestor that are 11.8 substitutions from the "root" can be resolved into 120 different, equally parsimonious, strictly bifurcating trees.

intimately cooperate, directly or indirectly, with the NP, [as, e.g., does the nonstructural protein NS1 (Scholtissek and Spring 1982)]—have to coevolve correspondingly (Buonagurio et al. 1986). A new NP gene then becomes reimplanted again from an animal reservoir, giving rise to a strain highly attenuated for humans (Snyder et al. 1987) with a lag phase in virulence reminiscent of the antigenic shift and slowly adapting to the new host. Such a mechanism cannot apply to influenza B viruses, since for the B type no animal reservoir is known. The evolution of other influenza A virus genes, especially the development of the various subtypes of the viral glycoproteins, presumably follows somewhat different rules, although the principal mechanism might be the same.

Acknowledgments

We thank Beate Schröder and Martina Kunerl for excellent technical assistance. This work was supported by Deutsche Forschungsgemeinschaft grants SFB 47 and SFB 272, the Fonds der Chemischen Industrie, and National Science Foundation grant BSR-8796183 to W.M.F.

LITERATURE CITED

ALTMÜLLER, A., W. M. FITCH, and C. SCHOLTISSEK. 1989. Biological and genetic evolution of the nucleoprotein (NP) gene of human influenza A viruses. J. Gen. Virol. **70**:2111–2119.

BEAN, W. J. 1984. Correlation of influenza A virus nucleoprotein genes with host specificity. Virology **133**:438–442.

BEKLEMISHE, A. B., V. M. BLOYNOV, S. K. VASSILENKO, S. Y. GOLOVIN, V. A. KARGINOV, E. V. MAMAYEW, S. V. NETESOV, N. A. PETROV, and P. F. SAFRONOV. 1986. Nucleotide sequence of a full-length DNA copy of the influenza A virus A/Kiev/59/79 (H1N1) nucleoprotein gene. Bioorg. Khim. **12**:369–374.

BRIEDIS, D. J., and M. TOBIN. 1984. Influenza B virus genome: complete sequence of the influenza B/Lee/40 virus genome RNA segment 5 encoding the nucleoprotein and comparison with the B/Singapore/222/79 nucleoprotein. Virology **133**:448–455.

BUCKLER-WHITE, A. J., and B. R. MURPHY. 1986. Nucleotide sequence analysis of the nucleoprotein of an avian and a human influenza virus strain identifies two classes of nucleoproteins. Virology **155**:345–355.

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.

COX, N. J., F. KITAME, A. P. KENDAL, H. F. MAASSAB, and C. W. 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.

DEBORDE, D. C., A. M. DONABEDIAN, M. L. HERLOCHER, C. W. NAEVE, and H. F. MAASSAB. 1988. Sequence comparison of wild-type and cold-adapted B/Ann Arbor/1/66 influenza virus genes. Virology **163**:429–443.

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.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.

FITCH, W. M., and J. S. FARRIS. 1974. Evolutionary trees with minimum nucleotide replacements from amino acid sequences. J. Mol. Evol. **3**:263–278.

GAMMELIN, M., J. MANDLER, and C. SCHOLTISSEK. 1989. Two subtypes of nucleoproteins (NP) of influenza A viruses. Virology **170**:71–80.

GUO, Y.-J., F.-G. JIN, P. WANG, M. WANG, and J.-M. ZHU. 1983. Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. J. Gen. Virol. **64**: 177–182.

HARLEY, V. R., C. W. WARD, and P. J. HUDSON. 1989. Molecular cloning and analysis of the N5 neuraminidase subtype from an avian influenza virus. Virology **169**:239–243.

HINSHAW, V. S., and R. G. WEBSTER. 1982. The natural history of influenza viruses. Pp. 79–104 *in* A. S. BEARE, ed. Basic and applied influenza research. CRC Press, Boca Raton, Fla.

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**:1029–1038.

KAWAOKA, Y., T. M. CHAMBERS, W. L. SLADENY, and R. G. WEBSTER. 1988. Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? Virology **163**: 247–250.

LI, R., P. PALESE, and M. KRYSTAL. 1989. Complementation and analysis of an NP mutant of influenza virus. Virus Res. **12**:97–112.

LONDO, D. R., A. R. DAVIS, and D. P. NAYAK. 1983. Complete nucleotide sequence of the nucleoprotein gene of influenza B virus. J. Virol. **47**:642–648.

MANDLER, J., and C. SCHOLTISSEK. 1989. Localization of the temperature-sensitive defect in the nucleoprotein of an influenza A/FPV/Rostock/34 virus. Virus Res. **12**:113–122.

NAKADA, S., R. S. CREAGER, M. KRYSTAL, and P. PALESE. 1984. Complete nucleotide sequence of the influenza C/California/78 virus nucleoprotein gene. Virus Res. **1**:433–441.

RAYMOND, F. L., A. J. CATON, N. J. COX, A. P. KENDAL, and G. G. BROWNLEE. 1983. Antigenicity and evolution amongst recent influenza viruses of H1N1 subtype. Nucleic Acids Res. **11**:7191–7203.

———. 1986. The antigenicity and evolution of influenza H1 hemagglutinins, from 1950–1957 and 1977–1983: two pathways from one gene. Virology **148**:275–287.

REINHARDT, U., and C. SCHOLTISSEK. 1988. Comparison of the nucleoprotein genes of a chicken and a mink influenza A H10 virus. Arch. Virol. **103**:139–145.

ROTA, P. A. 1989. Sequence of a cDNA clone of the nucleoprotein gene of influenza B/Ann Arbor/1/86. Nucleic Acids Res. **17**:3595.

SCHOLTISSEK, C., E. HARMS, W. ROHDE, M. ORLICH, and R. ROTT. 1976. Correlation between RNA fragments of fowl plague virus and their corresponding gene functions. Virology **74**: 332–344.

SCHOLTISSEK, C., I. KOENNECKE, and R. ROTT. 1978. Host range recombinants of fowl plague (influenza A) virus. Virology **91**:79–85.

SCHOLTISSEK, C., H. BÜRGER, O. KISTNER, and K. F. SHORTRIDGE. 1985. The nucleoprotein as a possible major factor in host specificity of influenza H3N2 viruses. Virology **147**:287–294.

SCHOLTISSEK, C., and S. B. SPRING. 1982. Extragenic suppression of temperature-sensitive mutations in RNA segment 8 by replacement of different RNA segments with those of other influenza A virus prototype strains. Virology **118**:28–34.

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.

STEULER, H., B. SCHRÖDER, H. BÜRGER, and C. SCHOLTISSEK. 1985. Sequence of the nucleoprotein gene of influenza A/parrot/Ulster/73. Virus Res. **3**:35–40.

TOMLEY, F. M., and I. J. RODITI. 1984. Nucleotide sequence of RNA segment 5, encoding the nucleoprotein of influenza A/FPV/Rostock/34. Virus Res. **1**:625–630.

TOWNSEND, A. R. M., J. ROTHBARD, F. M. GOTCH, G. BAHACHER, D. WRAITH, and A. J. MCMICHAEL. 1986. The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. Cell **44**:958–968.

VAN RUMPUY, L., W. MIN JOU, D. HUYLEBROECK, R. DEVOS, and W. FIERS. 1982. Complete nucleotide sequence of the nucleoprotein gene from the human influenza strain A/PR/8/34 (H0N1). Eur. J. Biochem. **116**:645.

VON HOYNINGEN-HUENE, V., and C. SCHOLTISSEK. 1983. Low genetic mixing between avian influenza viruses of different geographic regions. Arch. Virol. **76**:63–67.

WINTER, G., and S. FIELDS. 1981. The structure of the gene encoding the nucleoprotein of human influenza A/PR/8/34. Virology **114**:423–428.

JAN KLEIN, reviewing editor

Received August 4, 1989; revision received September 13, 1989

Accepted September 26, 1989