JOURNAL OF VIROLOGY, Aug. 2004, p. 8771–8779 Vol. 78, No. 16
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.16.8771–8779.2004

Matrix Gene of Influenza A Viruses Isolated from Wild Aquatic Birds: Ecology and Emergence of Influenza A Viruses

Linda Widjaja,[1,2] Scott L. Krauss,[1] Richard J. Webby,[1] Tao Xie,[3] and Robert G. Webster[1,4]*

Division of Virology, Department of Infectious Diseases, St. Jude Children's Research Hospital,[1] and Hartwell Center for Bioinformatics and Biotechnology,[3] Memphis, Tennessee 38105, and Department of Pathology[4] and Integrated Program in Biomedical Sciences,[2] The University of Tennessee Health Science Center, Memphis, Tennessee 38163

Received 11 November 2003/Accepted 8 April 2004

Wild aquatic birds are the primary reservoir of influenza A viruses, but little is known about the viruses' gene pool in wild birds. Therefore, we investigated the ecology and emergence of influenza viruses by conducting phylogenetic analysis of 70 matrix (M) genes of influenza viruses isolated from shorebirds and gulls in the Delaware Bay region and from ducks in Alberta, Canada, during >18 years of surveillance. In our analysis, we included 61 published M genes of isolates from various hosts. We showed that M genes of Canadian duck viruses and those of shorebird and gull viruses in the Delaware Bay shared ancestors with the M genes of North American poultry viruses. We found that North American and Eurasian avian-like lineages are divided into sublineages, indicating that multiple branches of virus evolution may be maintained in wild aquatic birds. The presence of non-H13 gull viruses in the gull-like lineage and of H13 gull viruses in other avian lineages suggested that gulls' M genes do not preferentially associate with the H13 subtype or segregate into a distinct lineage. Some North American avian influenza viruses contained M genes closely related to those of Eurasian avian viruses. Therefore, there may be interregional mixing of the two clades. Reassortment of shorebird M and HA genes was evident, but there was no correlation among the HA or NA subtype, M gene sequence, and isolation time. Overall, these results support the hypothesis that influenza viruses in wild waterfowl contain distinguishable lineages of M genes.

Influenza A viruses can infect a variety of species, including birds and humans. Ecological studies of these viruses have established that wild aquatic birds are the primary source of influenza A viruses (40). These viruses appear to be in evolutionary stasis while residing in asymptomatic aquatic birds, but they rapidly evolve once they cross species barriers and thus cause mild to severe disease in the new hosts.

Avian influenza viruses cause disease in domestic poultry and, more importantly, can also cause human pandemics (23, 25). Among the 15 hemagglutinin (HA) subtypes of influenza virus, two (H5 and H7) have the potential to become highly pathogenic in domestic poultry (41). An early study suggested that the HA gene of highly pathogenic H5 poultry viruses originated in influenza viruses present in wild birds (19). The influenza virus HA gene was shown to have a preferential association with the matrix (M) gene (24, 38). The M gene, an important determinant of species specificity (24, 34), encodes two partly overlapping proteins: a highly conserved 252-amino-acid M1 protein and a 97-amino-acid M2 protein. The evolution of the M gene may reflect host-specific adaptation, since it is probably not subjected to strong selection pressure by the host immune system. Despite the association of the two genes, the involvement of the M gene in the origin of highly pathogenic influenza viruses has been less studied. In addition, the precise type of aquatic-bird origin of viruses causing outbreaks in poultry has remained unclear, even though it was thought that the gene pool of viruses in shorebirds and gulls was the most closely related to the source of virulent avian strains (36, 38). Furthermore, recent cases have revealed a strong connection between avian and human influenza outbreaks. The 1997 influenza outbreak in Hong Kong showed that direct transmission of H5N1 influenza viruses from domestic chickens to humans can occur. The outbreak in humans appeared to be stopped by the slaughter of poultry in the Hong Kong markets during the same year (3, 28, 32). In 2002, H5N1 influenza virus was found, for the first time, to cause severe disease in multiple species of wild waterfowl in two parks in Hong Kong and was subsequently linked to human infections (K. M. Sturm-Ramirez, personal communication). Because measures to contain these viruses are impossible to implement in the wild, understanding the ecology of influenza viruses in wild aquatic birds is crucial if interventions are to be developed to decrease or control the frequent influenza outbreaks in domestic poultry, the annual human epidemics, and, more importantly, the occasional human pandemics, such as the 1957 and 1968 outbreaks in Asia and the 1918 Spanish influenza, which killed >40 million people worldwide (33).

Despite the important role of the gene pools of influenza viruses in wild aquatic birds and the potential threat of pandemic strains that could arise from viruses in aquatic birds, our knowledge of the gene pools is limited (31). Early phylogenetic studies revealed a geographical separation of the avian influenza virus lineage into Eurasian and North American lineages (6, 7, 11, 12). Avian influenza viruses isolated from gulls in North America, previously reported as preferentially linked to the H13 subtype, had been shown to form a distinct lineage, referred as the "gull-specific lineage," in previous publications (7, 11). These studies, however, were based on a relatively

* Corresponding author. Mailing address: Division of Virology, Department of Infectious Diseases, Mail Stop 330, St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105. Phone: (901) 495-3400. Fax: (901) 523-2622. E-mail: robert.webster@stjude.org.

small number of isolates that may not have been representative of viruses in the wider aquatic-bird reservoir. The present theory that there is overlapping of gene pools of viruses in different species of the class Aves (mainly shorebirds, gulls, and ducks) was accepted on the basis of biological and epidemiological studies of a few isolates but without further genetic proof (14, 27). Therefore, present knowledge about these gene pools is incomplete, and there is a lack of substantial genetic evidence, especially in regard to the internal genes of influenza viruses, to support the validity of the theory.

Since the first isolation of influenza A virus in healthy feral ducks in 1974 (30), surveillance has continued in an effort to better understand the ecology of these viruses in birds and their relationship to influenza viruses that appear in other species. We have conducted surveillance of influenza viruses in wild Canadian ducks since 1976 and in shorebirds and gulls of Delaware Bay for the past 18 years and thus have an extensive collection of influenza virus isolates. Therefore, in the present study, we analyzed the M gene partial sequences of 70 viruses of various subtypes from our collection and 61 published M gene sequences of viruses isolated from various host species to gain more detailed knowledge about the genetic features of influenza viruses in their natural hosts, i.e., wild aquatic birds. Our phylogenetic analysis was designed to specifically address six key questions. First, do influenza viruses from wild ducks belong to a sublineage that is separate from the sublineage of shorebird viruses? Second, which type of aquatic bird (shorebirds, gulls, or ducks) is the primary origin of the M genes of poultry viruses, which continue to circulate despite continuous surveillance and frequent eradication of infected poultry populations? Third, if the HA genes of a relatively small number of North American H2 and H6 viruses are more similar to those of Eurasian viruses than to those of H2 or H6 gull, shorebird, and poultry viruses in North America (15, 21, 35), then are similar interregional transfers of M genes evident in these viruses? Fourth, does the evolution of M genes of viruses from wild aquatic birds resemble the evolution of HA genes of North American H5 viruses, which has resulted in multiple sublineages (5)? Fifth, is there evidence of reassortment of M genes occurring between influenza viruses in nature? Sixth, is there any correlation between the M gene, antigenic subtype, and time at which each virus was isolated?

MATERIALS AND METHODS

Viruses. The 70 influenza viruses from wild ducks, shorebirds, and gulls used in this study were collected as part of surveillance programs in the Delaware Bay and Alberta, Canada, since 1985 and 1976, respectively. These samples were selected from the repository at St. Jude Children's Research Hospital, Memphis, Tenn., on the bases of time of isolation and antigenic subtype. Viruses isolated between 1985 and 1999 from shorebirds and gulls were selected, as were viruses isolated from ducks between 1976 and 1999. All HA subtypes except H8, H14, and H15 and all nine neuraminidase (NA) subtypes were represented (Table 1).

(i) Viruses isolated from shorebirds and gulls. Viruses were collected from shorebirds and gulls in New Jersey on the Delaware Bay from mid-May to early June every year since 1985. This period coincides with the stopover of shorebirds to feed on the abundant horseshoe crab eggs in Delaware Bay during their migration from the southern tip of South America to their breeding grounds in the Canadian arctic. Some of the viruses collected from 1985 through 1989 were obtained from the cloacae of shorebirds caught in the mist nets, but most viruses were obtained from fecal samples. Since 1990, viruses have been collected from fecal samples; each sample consisted of three fecal deposits placed in a single vial. Since the birds tend to group together in the sample collection area, the strain host designation is made with some degree of confidence. Most viruses isolated from gulls were obtained from fresh feces deposited on beaches and agricultural land where only one gull species was present. The shorebirds and gulls from which virus samples were collected belong to the order *Charadriiformes* and represent two families: (i) *Scolopacidae*—ruddy turnstone (*Arenaria interpres*), red knot (*Calidris canutus*), sanderling (*Calidris alba*), and semipalmated sandpiper (*Calidris pusilla*); and (ii) *Laridae*—laughing gull (*Larus atricilla*) and herring gull (*Larus argentatus*).

(ii) Viruses isolated from wild ducks. Virus samples were collected annually from wild ducks in Alberta, Canada, from late July through early September since 1976 (27). This period coincides with the completion of the breeding season, when many adult ducks have molted and the birds are staging before they migrate south to their winter habitat. In this study, we analyzed viruses from cloacal samples of wild ducks that are species of mallard (*Anas platyrhynchos*), pintail (*Anas acuta*), and blue-winged teal (*Anas discors*). Each species belongs to the family *Anatidae* and the order *Anseriformes*.

(iii) Sample collection. At the collection site, each virus sample was placed in transport medium containing 50% glycerol in phosphate-buffered saline, pH 7.2, containing antibiotics (penicillin G, 1,000 U/ml; streptomycin sulfate, 0.2 mg/ml; gentamicin sulfate, 0.24 mg/ml; polymyxin B, 100 U/ml; and mycostatin, 50 U/ml). The vials containing samples isolated from wild ducks were placed immediately into liquid nitrogen, while those containing fecal and cloacal samples collected from shorebirds and gulls were stored on ice. The samples were shipped by air, approved for importation by license from the U.S. Department of Agriculture, transported to St. Jude Children's Research Hospital, and stored at −70°C.

(iv) Virus isolation. The samples were thawed, and 0.3 ml of fluid was mixed with an equal volume of phosphate-buffered saline containing antibiotics (penicillin G, 4,000 U/ml; streptomycin sulfate, 800 U/ml; polymyxin B, 400 U/ml; gentamicin sulfate, 0.1 mg/ml). The inoculum for each sample was propagated in 11-day-old embryonated chicken eggs as previously described (9). The presence of agglutinating virus in the allantoic fluid of each egg was tested by hemagglutination assay. Bacterial contamination was checked by streaking blood agar plates, followed by incubation for 48 h at 37°C. To remove bacterial contamination, the allantoic fluid was passed through a 0.2-μM-pore-size syringe filter.

(v) Subtype determination. Samples that were positive for HA were tested in HA inhibition and NA inhibition assays to determine the HA type (18) and the NA type (1), respectively. The antigenic characteristics of all hemagglutinating agents were identified using monospecific antisera prepared against the surface antigens of reference influenza A viruses (39) or chicken or rabbit antisera against the whole virus antigen.

If the HA or NA inhibition assay could not determine the subtype, the sample was identified as influenza A virus by negative staining and scanning by electron microscope, by reverse transcription-PCR of the M gene (29), or by both methods. Once the isolate was identified as influenza A, the HA and NA types were determined by reverse transcription-PCR (10) and sequence analysis. Nucleotide sequences were edited and compiled using Lasergene sequence analysis software (DNASTAR, Madison, Wis.), and their homologies were determined against archived sequence data in GenBank.

(vi) M gene sequences obtained from GenBank. M gene sequences of 61 additional influenza A viruses were obtained from GenBank. These sequences were from 23 poultry viruses, 12 nonpoultry avian viruses, 5 equine viruses, and 13 human influenza A viruses.

RNA extraction and PCR. Viral RNA was extracted from 200 μl of infective allantoic fluid by using the RNeasy kit (Qiagen, Valencia, Calif.) according to the manufacturer's instructions. The RNA was then converted to full-length cDNA by using reverse transcriptase. PCR with M gene-specific primers (sequences are available upon request) was performed to amplify the cDNA. The PCR products were purified by using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocols.

Sequencing and sequence analysis of M genes. Purified PCR products containing M genes were sequenced by the Hartwell Center for Bioinformatics and Biotechnology at St. Jude by using synthetic oligonucleotides and rhodamine or dRhodamine dye terminator Cycle-Sequencing Ready-Reaction kits with AmpliTaq DNA polymerase FS (Perkin-Elmer Applied Biosystems, Inc., Foster City, Calif.). Samples were separated by electrophoresis and analyzed on Perkin-Elmer Applied Biosystems Inc. model 373 or model 377 DNA sequencers. Nucleic acids in positions 129 through 903 were sequenced. The DNA sequences were compiled and edited by using the Lasergene sequence analysis software package. Initial multiple sequence alignments were made with the Megalign program (DNASTAR), which used the CLUSTAL X algorithm. Phylogenetic trees were constructed by the neighbor-joining method using nucleotide sequences within the Neighbor program of PHYLIP (Phylogeny Inference Package) version 3.57c, and TreeView programs were used to view the distance tree.

TABLE 1. Influenza virus strains used in phylogenetic analysis, excluding those obtained from GenBank, and their subtypes and abbreviations

Virus strain	Subtype
A/Ruddy Turnstone/Delaware/134/99[a]	H7N7
A/Ruddy Turnstone/Delaware/130/99	H7N2
A/Ruddy Turnstone/Delaware/142/99	H3N2
A/Shorebird/Delaware/276/99	H9N2
A/Shorebird/Delaware/86/99	H11N2
A/Sanderling/Delaware/65/99	H3N8
A/Ruddy Turnstone/Delaware/259/98	H9N9
A/Ruddy Turnstone/Delaware/116/98	H9N8
A/Ruddy Turnstone/Delaware/142/98	H2N8
A/Ruddy Turnstone/Delaware/105/98	H6N8
A/Ruddy Turnstone/Delaware/265/98	H12N8
A/Ruddy Turnstone/Delaware/67/98	H12N4
A/Ruddy Turnstone/Delaware/106/98	H6N2
A/Shorebird/Delaware/194/98	H6N4
A/Shorebird/Delaware/24/98	H2N1
A/Shorebird/Delaware/27/98	H5N8
A/Shorebird/Delaware/207/98	H5N8
A/Herring gull/Delaware/281/98	H5N8
A/Shorebird/Delaware/222/97	H3N6
A/Shorebird/Delaware/224/97	H13N6
A/Shorebird/Delaware/127/97	H6N2
A/Shorebird/Delaware/122/97	H2N1
A/Shorebird/Delaware/138/97	H2N1
A/Shorebird/Delaware/111/97	H2N1
A/Ruddy Turnstone/Delaware/34/93	H2N1
A/Ruddy Turnstone/Delaware/81/93	H2N1
A/Ruddy Turnstone/Delaware/244/91[a]	H5N2
A/Red knot/Delaware/2552/87	H9N5
A/Red knot/Delaware/2561/87	H10N5
A/Ruddy Turnstone/Delaware/2589/87[a]	H11N4
A/Ruddy Turnstone/Delaware/2764/87	H10N7
A/Ruddy Turnstone/Delaware/2770/87	H7N5
A/Ruddy Turnstone/Delaware/2762/87	H11N2
A/Laughing gull/Delaware/2838/87	H7N2
A/Sanderling/New Jersey/766/86[a]	H2N7
A/Sanderling/Delaware/1258/86[a]	H6N6
A/Herring gull/New Jersey/780/86	H1N3
A/Herring gull/New Jersey/782/86	H13N2
A/Herring gull/Delaware/471/86	H13N7
A/Laughing gull/New Jersey/798/86	H2N7
A/Ruddy Turnstone/New Jersey/51/85	H11N1
A/Ruddy Turnstone/New Jersey/65/85	H7N3
A/Laughing gull/New Jersey/75/85	H2N9
A/Mallard/Alberta/213/99	H4N6
A/Mallard/Alberta/199/99	H3N6
A/Mallard/Alberta/110/99	H10N7
A/Mallard/Alberta/111/99	H4N6
A/Mallard/Alberta/125/99	H11N6
A/Mallard/Alberta/122/99	H11N9
A/Pintail/Alberta/37/99	H3N8
A/Mallard/Alberta/30/98	H4N6
A/Mallard/Alberta/205/98	H2N9
A/Mallard/Alberta/226/98	H2N3
A/Mallard/Alberta/71/98	H10N7
A/Mallard/Alberta/223/98	H10N8
A/Mallard/Alberta/119/98	H1N1
A/Mallard/Alberta/47/98	H4N1
A/Mallard/Alberta/167/98	H3N8
A/Blue-winged teal/Alberta/16/97	H2N9
A/Mallard/Alberta/90/97	H10N7
A/Mallard/Alberta/114/97	H3N8
A/Mallard/Alberta/52/97	H12N5
A/Mallard/Alberta/223/77	H12N5
A/Mallard/Alberta/42/77	H1N6
A/Mallard/Alberta/300/77[a]	H4N3
A/Mallard/Alberta/297/77	H10N7
A/Mallard/Alberta/77/77	H2N3
A/Pintail/Alberta/293/77	H2N9
A/Blue-winged teal/Alberta/96/76	H4N8
A/Mallard/Alberta/22/76	H3N6

[a] Strain whose M2 protein sequence was not included in the phylogenetic analysis.

Errors associated with the generation of phylogenetic trees were evaluated by bootstrap analyses with 100 bootstrapping replicas.

We used SNAP (Synonymous/Nonsynonymous Analysis Program [http://www.hiv.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html]), which implemented the method of Nei and Gojobori (16), to calculate synonymous and nonsynonymous substitution (ds/dn) ratios for M1 and M2. Statistical analysis of the ds/dn ratio value was performed using the *t* test (two tailed; alpha < 0.01).

Nucleotide sequence accession numbers. The nucleotide sequences presented in this article have been submitted to GenBank under accession numbers AY664420 through AY664489.

RESULTS

M gene phylogenetic tree. Each of the 131 M gene sequences comprised 775 nucleotides (nucleic acid positions 129 through 903). No insertions or deletions were observed. The nucleotide sequence homology ranged from 82.7 to 100%. The results of our phylogenetic analysis of these sequences are presented as a phylogenetic tree rooted to A/Equine/Prague/1/56 (Fig. 1).

The phylogenetic tree confirmed that the evolution of M genes has resulted in the development of major influenza A virus lineages, each of which could be associated with a particular host (Fig. 1): human- and classical swine-like viruses, gull-like viruses, nongull avian-like viruses, and equine-like viruses.

North American and Eurasian avian M lineages: multiple sublineages and intermixing. Excluding the presence of a number of shorebird viruses isolated between 1997 and 1999 that formed a distinct group (bootstrap value, 98) (Fig. 1C), there was no clear separation of isolates into host lineages.

Within the lineages of North American and Eurasian avian-like viruses, divisions into sublineages seem to be apparent. The Eurasian avian-like lineage comprised a major sublineage I (bootstrap value, 99) and a minor sublineage II, less well supported by a bootstrap value of 19 (45) (the numbers in parentheses represent bootstrap values based on phylogenetic trees generated without outlier sequences) (Fig. 1B). The North American avian-like viruses seem to segregate in two sublineages, sublineage III (bootstrap value, 66 [100]) and sublineage IV (bootstrap value, 72) (Fig. 1C). Major sublineage I of the Eurasian avian-like lineage included viruses isolated from quail, pheasants, humans, and chickens in Hong Kong from 1997 through 2001. Minor sublineage II of the Eurasian avian-like lineage comprised other avian isolates from Europe

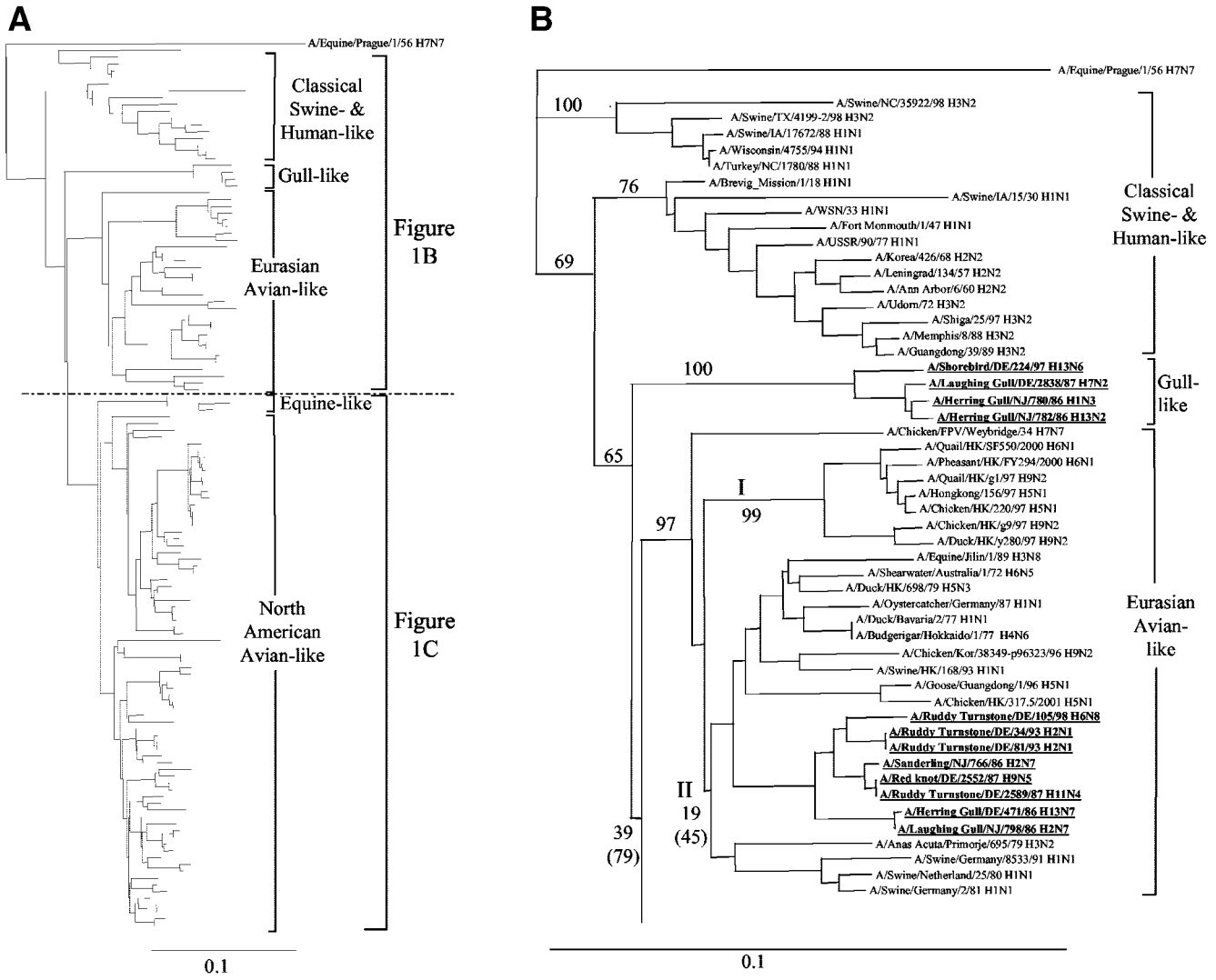

FIG. 1. Phylogenetic tree for influenza A virus M genes rooted to that of A/Equine/Prague/1/56. (A) Complete simplified nucleotide phylogram showing influenza A virus lineages associated with particular hosts and/or geographical regions: classical swine- and human-like, gull-like, Eurasian avian-like, equine-like, and North American avian-like lineages. (B) Part of the phylogram illustrating the classical swine- and human-like, gull-like, and Eurasian avian-like lineages. (C) Part of the phylogram representing the equine-like and North American avian-like lineages. The numbers at each major fork represent bootstrap values, and the roman numerals (I to IV) represent different sublineages. The numbers in parentheses represent bootstrap values based on phylogenetic trees generated without outlier sequences, i.e., without A/Chicken/FPV/Weybridge/34 H7N7 for sublineage II, without equine-like viruses for the North American avian-like lineage and sublineage III, and without gull-like viruses for the common branch of the Eurasian and North American avian-like lineages. The strains sequenced in this study are indicated in boldface; strains isolated from ducks are distinguished from those of shorebirds and gulls (shorebird and gull lineages are underlined). The scale shows the number of nucleotide substitutions per site. Vertical lines are used for spacing the branches and labels; their lengths are not significant. DE, Delaware; NJ, New Jersey; HK, Hong Kong; Kor, Korea; NC, North Carolina; TX, Texas; IA, Iowa.

and Asia and a relatively small number of swine isolates; interestingly, a group of strains consistently isolated from gulls and shorebirds in North America between 1986 and 1998 belong to this group (Fig. 1B). Sublineage III included the prototypical turkey strain isolated in 1966 in California, more recent isolates (H7N2) from chickens and turkeys in Pennsylvania, and viruses that represented subtypes H2 through H7 and H9 through H12 that were isolated from shorebirds, ducks, and poultry between 1966 and 1999. Sublineage IV included the prototypical strain A/Chicken/Pennsylvania/1370/83 (H5N2), a strain isolated during the 1994 outbreaks in chickens in Mexico, H6N2 isolates obtained in 2000 and 2001 from chickens in California, and viruses that represented subtypes H1 through H7 and H9 through H12 that were isolated from shorebirds, ducks, and poultry between 1977 and 2001.

Six viruses isolated from North American shorebirds carried HA genes that are of Eurasian lineage (15, 21). Among these viruses, we identified two strains bearing HA and M genes of the Eurasian avian-like lineage and four strains with HA genes of the Eurasian avian-like lineage and M genes of the North American avian-like lineage (Table 2).

The M gene of gull viruses. Influenza viruses of the H13 subtypes isolated from North American gulls had previously been shown to form a distinct lineage (11). In this study, one

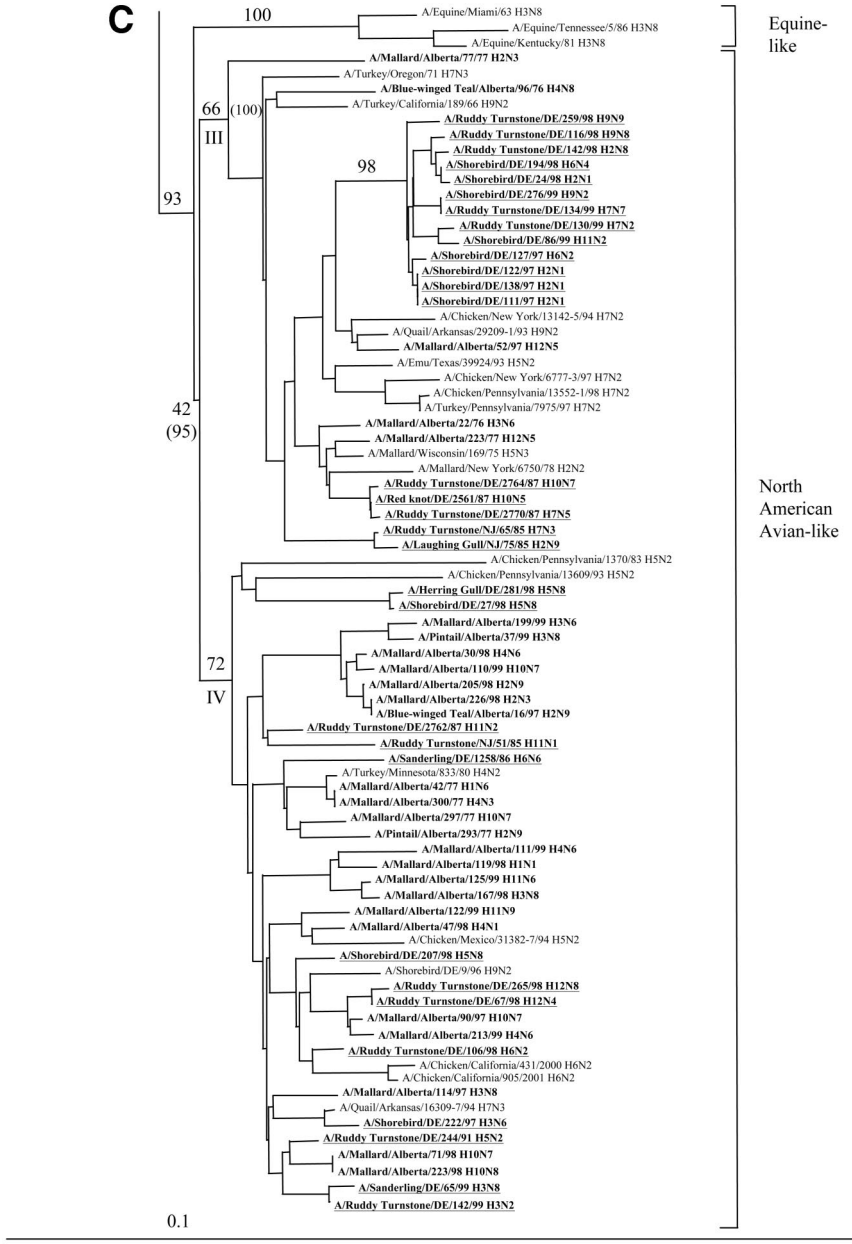

FIG. 1—*Continued.*

strain, A/Herring Gull/Delaware/471/86 (H13N7), belonged to the Eurasian avian-like M gene lineage and shared a branch with one gull virus, A/Laughing Gull/New Jersey/798/86 (H2N7), isolated in the same year (Fig. 1B, sublineage II). In addition, A/Laughing Gull/Delaware/2838/87 (H7N2) and A/Herring Gull/New Jersey/780/86 (H1N3) were grouped into the gull-like lineage (Fig. 1B). Two viruses, A/Laughing Gull/New Jersey/75/85 (H2N9) and A/Herring Gull/Delaware/98 (H5N8), which were isolated 13 years apart, were not part of the gull-like lineage but belonged to the North American avian-like lineage (Fig. 1C). The only H13 virus isolated from a North American shorebird, A/Shorebird/Delaware/224/97 (H13N6), has an M gene that is similar to that of gull viruses but has diverged from those of other shorebird isolates. These data showed that gulls' M genes do not link solely to the H13 HA subtype or cluster exclusively in a single discrete lineage.

M gene of poultry viruses. To identify the exact origin of viruses that infect North American poultry, we included the M gene sequences of viruses isolated from North American commercial poultry in our analysis. Most viruses isolated from ducks, shorebirds, gulls, and poultry shared a common M gene lineage (Fig. 1C).

On the basis of the phylogenetic relationships and the percentage of M gene sequence homology among isolates, we attempted to deduce the virus (or viruses) whose M gene was the closest probable precursor of the M genes of some of the analyzed poultry viruses. Our analysis revealed that the M genes of the two isolates from chickens during the 2000-2001

TABLE 2. North American avian strains that have Eurasian-like HA genes, M genes, or both

North American avian strain	Subtype	HA[a]	M[b]
A/Ruddy Turnstone/Delaware/142/98	H2N8	Eurasian	North American
A/Shorebird/Delaware/24/98	H2N1	Eurasian	North American
A/Shorebird/Delaware/111/97	H2N1	Eurasian	North American
A/Shorebird/Delaware/138/97	H2N1	Eurasian	North American
A/Ruddy Turnstone/Delaware/34/93	H2N1	Eurasian	Eurasian
A/Ruddy Turnstone/Delaware/81/93	H2N1	Eurasian	Eurasian

[a] Makarova et al. (15).
[b] This report.

influenza outbreaks in California (A/Chicken/California/431/2000 [H6N2] and A/Chicken/California/905/2001 [H6N2]) were derived from shorebird-like viruses of the same subtype (Fig. 1C). Accordingly, the M genes of these two chicken viruses were highly similar to that of A/Ruddy Turnstone/Delaware/106/98 (H6N2) (range, 98.1 to 98.3%), whereas the M gene of the most similar duck strain, A/Mallard/Alberta/71/98 (H10N7), was only 96.8 to 97% similar to the genes of the two chicken viruses. However, because of the high degree of similarity, ranging from 82.7 to 100%, the degree of homology did not always exactly correlate with the phylogenetic relationships. Other poultry isolates, including A/Chicken/Pennsylvania/1370/83 (H5N2), A/Chicken/Pennsylvania/13609/93 (H5N2), and A/Turkey/Oregon/71 (H7N3), did not have high sequence similarities with the shorebird, gull, and duck strains sequenced (Fig. 1C).

Analysis of the M1 and M2 protein sequences. To further investigate the evolutionary stasis of the M genes of viruses from wild aquatic birds, we translated the M1 and M2 open reading frames and analyzed the protein sequences (Table 3). Because the entire amino acid sequences of the M proteins were not available for all viruses, we analyzed the region between amino acid positions 43 and 252 of M1 and 10 and 88 of M2. The phylogenetic trees and the predicted amino acid sequences of M1 and M2 proteins are available online at http://www.stjuderesearch.org/data/flu2/index.html.

The topology and distribution of viruses in the phylogram based on the nucleotide sequences of the M genes differed greatly from those in the phylograms based on the amino acid sequences of the M proteins. In fact, the trees based on the M1 and M2 amino acid sequences had fewer distinct lineages than did the nucleotide-based tree, and clear lineage demarcation was not evident (http://www.stjuderesearch.org/data/flu2/index.html).

To perform a direct comparison with the previous data on the evolution of the influenza A virus M gene, the M1 and M2 amino acid sequences of all isolates were compared with that of A/Turkey/Minnesota/80 as a baseline sequence, as in the report by Ito et al. (11). Comparison of the predicted amino acid sequences with that of A/Turkey/Minnesota/833/80 (H4N2) (http://www.stjuderesearch.org/data/flu2/index) showed that M1 proteins of 4 of 29 duck viruses and of 8 of 38 shorebird isolates possessed substitutions of 1 to 3 residues over a span of 210 residues; in contrast, 11 of 14 strains isolated from poultry had M1 sequences with substitutions of 1 to 4 residues over a similar-size region. Comparison of the predicted amino acid sequences showed changes in 77 (36.7%) of 210 amino acids of the M1 proteins and in 51 (65.4%) of 78 amino acids of the M2 proteins. Although the changes in the M1 proteins were uniformly distributed, most of the changes in the M2 proteins were within residues 10 to 28, 54 to 57, and 77 to 88; these changes are consistent with those described in a previous report (11).

The predicted amino acid sequences indicated that viruses isolated from poultry underwent more amino acid mutations than did viruses isolated from shorebirds, gulls, and ducks (http://www.stjuderesearch.org/data/flu2/index). Upon adaptation in domestic poultry species, the changes in the predicted M1 and M2 amino acid sequences of poultry viruses did not occur in discrete locations; rather, they were uniformly distributed compared to the sequences of viruses isolated from shorebirds, gulls, and ducks (http://www.stjuderesearch.org/data/flu2/index.html).

To illustrate the relative influence of selection pressure in the evolution of M1 and M2 proteins, the ratio of synonymous to nonsynonymous substitutions was calculated (Table 3). Nucleotide substitutions that result in an amino acid change are nonsynonymous; those that do not result in an amino acid change are synonymous. The relative proportion of synonymous to nonsynonymous substitutions (the ds/dn ratio) was higher for M1 than for M2 in viruses from all host species. The ds/dn value for M1 of poultry viruses was significantly lower than for those of shorebird, gull, and duck viruses and significantly higher than for those of human viruses, but not significantly different than for those of swine and equine viruses. The ds/dn value for M2 of poultry viruses was significantly lower than for those of shorebird, duck, and human viruses but not significantly different than for those of gull, equine, and swine viruses.

DISCUSSION

Despite the importance of the ecology of influenza virus to animal and human health, only a few genes from the natural reservoir of these viruses have been sequenced; hence, there has not been a sufficient database for comparisons and further interpretations. In the present study, we sought to provide a better representation of existing influenza A viruses in their wild-aquatic-bird reservoir by analyzing a large number of viruses that were of various subtypes and isolated from wild birds at various times over several years.

Several important points can be deduced from this study. First, influenza viruses from wild ducks share common sublineages with shorebird viruses. Second, the primary origin of the

TABLE 3. Ratio of synonymous to nonsynonymous substitutions for M1 and M2

Host	ds/dn ratio	
	M1	M2
All	64.8	4.1
Poultry	38.4	3.5
Shorebird	266.9	4.5
Gull	275.9	3.9
Duck	143.2	6
Swine	33.2	2.6
Equine	31.4	4.2
Human	17.7	6

M genes of poultry viruses can come from either shorebirds, ducks, or gulls or may not have been identified yet. Third, interregional transmission of M genes is evident among the North American and Eurasian shorebird, gull, and wild-duck viruses. Fourth, the evolution of M genes of viruses from wild aquatic birds has resulted in multiple sublineages. Fifth, our results support the occurrence of reassortment of M genes between influenza A viruses in nature. Sixth, there was no evidence that showed any correlation among the M gene, antigenic subtype, and time at which each virus was isolated.

The overall segregation of the influenza A viruses in clusters reflected a divergent cumulative evolution of M genes with respect to the host and, for the avian strains, the geographical distribution (North America and Eurasia), although this conclusion may not hold up when a more extensive data set is obtained. Within the lineage of avian virus M genes, the existence of sublineages of viruses seems to be apparent. A similar feature was found in the analysis of the HA and nonstructural (NS) genes of H5 avian influenza A viruses in North America (5). We did not find any clear correlation between each sublineage and other analyzed variables, such as the HA subtype or date of isolation.

Sublineage II of the Eurasian avian-like M gene lineage contained both Eurasian and North American avian viruses, providing evidence of the transmission of the influenza virus M gene from Eurasian to North American avian-like viruses. Conversely, none of the Eurasian viruses belonged in the North American avian-like M gene lineage. This finding suggests that North American wild birds, mainly shorebirds and gulls, can support replication of viruses bearing at least the M gene of the Eurasian avian-like lineage. Thus, our study confirms the results of previous phylogenetic analyses of the HA genes of H2 and H6 influenza viruses (15, 21, 35). Although the details of these gene transfers between strains are not yet clear, the existence of such transfers has important implications. The Eurasian swine and avian influenza viruses are most closely related to human viruses and are potential sources of genes for future pandemic influenza strains (24). These facts are compelling reasons for the further study and surveillance of avian strains circulating in Eurasia and North America.

We described a previously unknown element of diversity in the M gene of viruses obtained from gulls. Previous phylogenetic studies of NP (nucleoprotein) and M genes had shown that the M genes of the viruses isolated from gulls preferentially associated with the H13 subtype and grouped together as a distinct gull lineage (7, 11). In contrast, we report that two samples collected from gulls belong to subtypes other than H13, that two gull isolates each belong to the Eurasian and North American avian-like lineages, and that a virus of the H13 subtype (A/Shorebird/DE/224/97 [H13N6]) was obtained from a shorebird rather than a gull. Our results showed that gulls' M genes do not preferentially associate with the H13 subtype and are compatible with HA genes that are not of the H13 subtype. The present analyses also demonstrated that gulls' M genes do not form a single lineage. A similar finding was revealed by the analysis of NS genes of H13 viruses that carried the NS gene of the North American avian-like lineage (12). Previous reports (7) demonstrated that for the NP gene, an important determinant of species specificity (22), such mixing between lineages does not occur, although this conclusion may not hold up as more extensive data are obtained. Like NS genes, the M genes may not play as much of a role as the NP gene in host range restriction. Nonetheless, because of the nature of the sampling method and the still relatively small number of samples, further investigation is recommended.

The M gene lineages of viruses isolated from poultry and those isolated from shorebirds, gulls, and ducks suggest that the M genes of these virus groups share a common origin and do not form host-specific lineages until the terminal branch of the tree. This finding is consistent with the proposals that wild aquatic birds, both ducks and shorebirds, serve as the host of the precursor(s) of past and future influenza A viruses that are transmitted to domestic poultry.

To identify all the potential reservoirs of influenza viruses infecting poultry, it is important to study the relationships among all genes of viruses from all available hosts, including wild aquatic birds. Since evolution occurs rapidly in poultry species compared to wild aquatic birds and due to the high degree of similarity among the M genes of viruses isolated from shorebirds, gulls, and ducks, the phylogenetic analysis of viruses from wild aquatic birds was unable to determine which species was the donor reservoir for the poultry viruses. Therefore, we concluded that the M genes of viruses from either shorebirds, gulls, or ducks could potentially be the precursor of the M gene of virus causing outbreaks in poultry. Alternatively, the possible M gene precursors of the pathogenic poultry viruses may not yet have been identified. Although the structural features in the HA glycoprotein are a primary determinant of pathogenicity (26, 37), the highly pathogenic phenotypes of particular avian influenza viruses have been linked to multiple genes (13, 20). Therefore, further studies of the relationships among all eight genes of poultry viruses would best pinpoint the primary origin of these viruses.

Genetic reassortment of influenza A viruses has been demonstrated in vivo, in vitro, and in nature. Indeed, reassortment of gene segments between influenza A viruses of different human subtypes in nature (8, 42) has been reported, as has gene reassortment between swine subtypes of influenza A viruses (17). Reassortant viruses characterized by different mixtures of human and avian genes have been isolated from humans and swine (2). So far, several reports of reassortment in nature between subtypes in avian species have been described (4, 34). Likewise, reassortment of gene segments between duck viruses has been suggested to occur in nature (5). In the present study, we found that shorebirds were a site of reassortment of gene segments between viruses having Eurasian or North American avian-like HA genes, M genes, or both. The fact that these reassortant viruses are of only the H2 subtype may reflect the limited number of isolates analyzed rather than a true association between a certain subtype and the occurrence of reassortment, such as the reassortment of H6 viruses in chickens. The rationale for an exclusive, geographically based classification of different virus strains appears to be less strong than in the past. Why and how this intermixing between lineages occurs in nature is still uncertain.

The ratio of synonymous to nonsynonymous substitutions in a protein-coding gene reflects the relative influences of selection pressure and evolution. Our observation suggests that the M2 protein evolves under much stronger positive selection than the M1 protein. This is true for all host species tested. Our

data suggest that viruses in poultry species are under selection pressure more similar to those seen in mammalian species than to those seen in avian species. Since only partial sequences were used in this analysis, further studies are recommended to eliminate any biased illustration of the overall evolutionary tendency of these proteins with respect to different hosts and other viral proteins.

The collapse of lineage divergence among avian virus proteins confirms that the cumulative genetic differences are noncoding changes and provides further evidence of a remarkable degree of genotypic stability among gene pools of geographically separated viruses. The fact that poultry viruses undergo more amino acid changes than do viruses in their natural reservoir (wild aquatic birds) supports the theory of the evolutionary stasis of influenza viruses in their natural reservoir and the rapid accumulation of mutations once they infect other species. The greater divergence of the M2 protein sequences confirms the independent evolution of the M1 and M2 proteins (11).

Although the functional interaction between the HA and M genes of influenza viruses has not been resolved, previous reports have shown a preferential association between the two types of genes (25, 39). Our findings do not confirm this observation. Although some groups of specific HA subtypes are located further away from the root of the M nucleotide phylogenetic tree than are others, in most cases, the M genes of strains representative of most HA subtypes are clustered together. For example, multiple viruses of subtypes H2 through H7 and H9 through H12 isolated from shorebirds, gulls, ducks, and poultry are clustered together within sublineage III (Fig. 1C). Since the proteins encoded by the M gene were relatively conserved, the M1-M2 phylogenetic tree was also used to address this query. Likewise, we found that there is no specific group of virus with certain HA subtypes that were grouped together. These data suggest that there is no specific correlation between the HA and M genes of the viruses analyzed.

Overall, this study enhances the present understanding of the M gene pool and further supports the hypothesis that influenza A viruses in aquatic wild birds contain multiple distinguishable M genes yet maintain remarkable homology. The phylogeny of each virus gene represents only a partial history of virus evolution. Therefore, consideration of the phylogenies of all eight genes may reveal a more complete picture of the evolution of viruses in their natural reservoir, i.e., wild aquatic birds. An important implication of phylogenetic studies is that the ancestral viruses that caused outbreaks in humans, including those that provided gene segments for viruses that caused the 1918, 1957, and 1968 pandemics, continue to circulate in wild birds with few, if any, mutations. Therefore, intensive surveillance of influenza viruses in aquatic birds can provide information about future outbreaks in domestic species and humans.

ACKNOWLEDGMENTS

These studies were supported by Public Health Research Grants AI-95357 and CA-21765 and by the American Lebanese Syrian Associated Charities (ALSAC).

We thank Paul Pryor and the Canadian Wildlife Service for collection of field samples from wild ducks and The New Jersey Department of Environmental Protection, Division of Fish and Wildlife Endangered and Nongame Species Program, for facilitating the collection of shorebird samples. We thank Todd Hatchette for his help with the phylogenetic analysis and Julia Cay Jones for scientific editing.

REFERENCES

1. **Aymard-Henry, M., W. R. Coleman, W. R. Dowdle, G. C. Laver, G. C. Schild, and R. G. Webster.** 1973. Influenza neuraminidase and neuraminidase-inhibition test procedures. Bull. W. H. O. **48:**199–202.
2. **Castrucci, M. R., I. Donatelli, L. Sidoli, G. Barigazzi, Y. Kawaoka, and R. G. Webster.** 1993. Genetic reassortment between avian and human influenza A viruses in Italian pigs. Virology **193:**503–506.
3. **Claas, E. C., A. D. Osterhaus, R. van Beek, J. C. De Jong, G. F. Rimmelzwaan, D. A. Senne, S. Krauss, K. F. Shortridge, and R. G. Webster.** 1998. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Lancet **351:**472–477.
4. **Desselberger, U., K. Nakajima, P. Alfino, F. S. Pederson, W. A. Haseltine, C. Hannoun, and P. Palase.** 1978. Biochemical evidence that "new" influenza virus strains in nature may arise by recombination (reassortment). Proc. Natl. Acad. Sci. USA **75:**3341–3345.
5. **Garcia, M., D. L. Suarez, J. M. Crawford, J. W. Latimer, R. D. Slemons, D. E. Swayne, and M. L. Perdue.** 1997. Evolution of H5 subtype avian influenza A viruses in North America. Virus Res. **51:**115–124.
6. **Gorman, O. T., R. O. Donis, Y. Kawaoka, and R. G. Webster.** 1990. Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. J. Virol. **64:**4893–4902.
7. **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.
8. **Guo, Y. J., X. Y. Xu, and N. J. Cox.** 1992. Human influenza A (H1N2) viruses isolated from China. J. Gen. Virol. **73:**383–387.
9. **Hinshaw, V. S., R. G. Webster, and B. Turner.** 1978. Novel influenza A viruses isolated from Canadian feral ducks, including strains antigenically related to swine influenza (Hsw1N1) viruses. J. Gen. Virol. **41:**115–127.
10. **Hoffmann, E., J. Stech, Y. Guan, R. G. Webster, and D. R. Perez.** 2001. Universal primer set for the full-length amplification of all influenza A viruses. Arch. Virol. **146:**2275–2289.
11. **Ito, T., O. T. Gorman, Y. Kawaoka, W. J. Bean, and R. G. Webster.** 1991. Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins. J. Virol. **65:**5491–5498.
12. **Kawaoka, Y., O. T. Gorman, T. Ito, K. Wells, R. O. Donis, M. R. Castrucci, I. Donatelli, and R. G. Webster.** 1998. Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. Virus Res. **55:**143–156.
13. **Kawaoka, Y., and R. G. Webster.** 1988. Molecular mechanism of acquisition of virulence in influenza virus in nature. Microb. Pathog. **5:**311–318.
14. **Kawaoka, Y., T. M. Chambers, W. L. Sladen, 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.
15. **Makarova, N. V., N. V. Kaverin, S. Krauss, D. Senne, and R. G. Webster.** 1999. Transmission of Eurasian avian H2 influenza virus to shorebirds in North America. J. Gen. Virol. **80:**3167–3171.
16. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
17. **Nerome, K., S. Sakamoto, N. Yano, T. Yamamoto, S. Kobayashi, R. G. Webster, and A. Oya.** 1983. Antigenic characteristics and genome composition of a naturally occurring recombinant influenza virus isolated from a pig in Japan. J. Gen. Virol. **64:**2611–2620.
18. **Palmer, D. F., M. T. Coleman, W. R. Dowdle, and G. C. Schild.** 1975. Advanced laboratory techniques for influenza diagnosis. U. S. Department of Health, Education, and Welfare, Washington, D.C.
19. **Rohm, C., T. Horimoto, Y. Kawaoka, J. Suss, and R. G. Webster.** 1995. Do hemagglutinin genes of highly pathogenic avian influenza viruses constitute unique phylogenetic lineages? Virology **209:**664–670.
20. **Rott, R., M. Orlich, and C. Scholtissek.** 1979. Correlation of pathogenicity and gene constellation of influenza A viruses. III. Non-pathogenic recombinants derived from highly pathogenic parent strains. J. Gen. Virol. **44:**471–477.
21. **Schafer, J. R., Y. Kawaoka, W. J. Bean, J. Suss, D. Senne, and R. G. Webster.** 1993. Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible progenitors in the avian reservoir. Virology **194:**781–788.
22. **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.
23. **Scholtissek, C., V. von Hoyningen, and R. Rott.** 1978. Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). Virology **89:**613–617.
24. **Scholtissek, C., J. Stech, S. Krauss, and R. G. Webster.** 2002. Cooperation between the hemagglutinin of avian viruses and the matrix protein of human influenza A viruses. J. Virol. **76:**1781–1786.
25. **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.

26. **Senne, D. A., B. Panigrahy, Y. Kawaoka, J. E. Pearson, J. Suss, M. Lipkind, H. Kida, and R. G. Webster.** 1996. Survey of the hemagglutinin (HA) cleavage site sequence of H5 and H7 avian influenza viruses: amino acid sequence at the HA cleavage site as a marker of pathogenicity potential. Avian Dis. **40:**425–437.
27. **Sharp, G. B., Y. Kawaoka, S. M. Wright, B. Turner, V. Hinshaw, and R. G. Webster.** 1993. Wild ducks are the reservoir for only a limited number of influenza A subtypes. Epidemiol. Infect. **110:**161–176.
28. **Shortridge, K. F., N. N. Zhou, Y. Guan, P. Gao, T. Ito, Y. Kawaoka, S. Kodihalli, S. Krauss, D. Markwell, K. G. Murti, M. Norwood, D. Senne, L. Sims, A. Takada, and R. G. Webster.** 1998. Characterization of avian H5N1 influenza viruses from poultry in Hong Kong. Virology **252:**331–342.
29. **Shu, L. L., W. J. Bean, and R. G. Webster.** 1993. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. J. Virol. **67:**2723–2729.
30. **Slemons, R. D., D. C. Johnson, J. S. Osborn, and F. Hayes.** 1974. Type-A influenza viruses isolated from wild free-flying ducks in California. Avian Dis. **18:**119–124.
31. **Suarez, D. L.** 2000. Evolution of avian influenza viruses. Vet. Microbiol. **74:**15–27.
32. **Subbarao, K., A. Klimov, J. Katz, H. Regnery, W. Lim, H. Hall, M. Perdue, D. Swayne, C. Bender, J. Huang, M. Hemphill, T. Rowe, M. Shaw, X. Xu, K. Fukuda, and N. Cox.** 1998. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science **279:**393–396.
33. **Taubenberger, J. K., A. H. Reid, A. E. Krafft, L. E. Bijwaard, and T. G. Fanning.** 1997. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science **275:**1793–1796.
34. **Tian, S. F., A. J. Buckler-White, W. T. London, L. J. Reck, R. M. Chanock, and B. R. Murphy.** 1985. Nucleoprotein and membrane protein genes are associated with restriction of replication of influenza A/Mallard/N.Y./78 virus and its reassortants in squirrel monkey respiratory tract. J. Virol. **53:**771–775.
35. **Webby, R. J., P. R. Woolcock, S. L. Krauss, and R. G. Webster.** 2002. Ressortment and interspecies transmission of North American H6N2 influenza viruses. Virology **295:**44–53.
36. **Webster, R. G., G. B. Sharp, and E. C. Claas.** 1995. Interspecies transmission of influenza viruses. Am. J. Respir. Crit. Care Med. **152:**S25–S30.
37. **Webster, R. G., and R. Rott.** 1987. Influenza virus A pathogenicity: the pivotal role of hemagglutinin. Cell **50:**665–666.
38. **Webster, R. G., Y. Kawaoka, and W. J. Bean.** 1989. What is the potential of avirulent influenza viruses to complement a cleavable hemagglutinin and generate virulent strains? Virology **171:**484–492.
39. **Webster, R. G., V. A. Isachenko, and M. Carter.** 1974. A new avian influenza virus from feral birds in the USSR: recombination in nature? Bull. W. H. O. **51:**324–332.
40. **Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka.** 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. **56:**152–179.
41. **Wood, G. W., J. W. McCauley, J. B. Bashiruddin, and D. J. Alexander.** 1993. Deduced amino acid sequences at the haemagglutinin cleavage site of avian influenza A viruses of H5 and H7 subtypes. Arch. Virol. **130:**209–217.
42. **Young, J. F., and P. Palese.** 1979. Evolution of human influenza A viruses in nature: recombination contributes to genetic variation of H1N1 strains. Proc. Natl. Acad. Sci. USA **76:**6547–6551.