VIROLOGY 209, 664–670 (1995)

SHORT COMMUNICATION

Do Hemagglutinin Genes of Highly Pathogenic Avian Influenza Viruses Constitute Unique Phylogenetic Lineages?

CAROLIN RÖHM,* TAISUKE HORIMOTO,* YOSHIHIRO KAWAOKA,*,† JOCHEN SÜSS,‡ and ROBERT G. WEBSTER*,†,[1]

*Department of Virology/Molecular Biology, St. Jude Children's Research Hospital, 332 North Lauderdale, P.O. Box 318, Memphis, Tennessee 38101; †University of Tennessee at Memphis, Department of Pathology, 800 North Madison, Memphis, Tennessee 38163; and ‡Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, Diedersdorfer Weg 1, 12277 Berlin, Germany

Received January 10, 1995; accepted March 27, 1995

Avian influenza A viruses of the H5 and H7 subtypes periodically cause severe outbreaks of disease in poultry. The question we wished to address in this study is whether these highly pathogenic strains constitute unique lineages or whether they and related nonpathogenic viruses are derived from common ancestors in the wild bird reservoir. We therefore compared the nucleotide and amino acid sequences of the hemagglutinin (HA) genes of 15 H5 and 26 H7 influenza A viruses isolated over 91 years from a variety of host species in Eurasia, Africa, Australia, and North America. Phylogenetic analysis indicated that the HA genes of H5 and H7 viruses that cause severe disease in domestic birds do not form unique lineages but share common ancestors with nonpathogenic H5 and H7 viruses. These findings predict that highly pathogenic avian H5 and H7 influenza A viruses will continue to emerge from wild bird reservoirs. Another important question is whether H7 influenza viruses found in mammalian species are derived from avian strains. We included eight equine influenza viruses and one seal isolate in the phylogenetic analysis of H7 HA genes. We could show that the HA genes of both, the equine and the seal viruses, shared ancestors with avian H7 HA genes. This indicates that currently circulating H7 viruses with an avian HA gene may have the potential to adapt to mammalian species and to cause an influenza outbreak in the new host.

Wild aquatic birds are considered the primordial reservoir of all 14 HA subtypes of avian influenza A viruses (*1, 2*). In general, the influenza viruses isolated from aquatic birds do not produce symptomatic infections; however, some strains of the H5 and H7 subtypes are highly pathogenic in certain avian species, including domestic chickens, turkeys, and quails.

Although the pathogenicity of influenza viruses is a polygenomic trait (*3, 4*), the HA molecule, which is responsible for virus binding to cell surface receptors and mediating fusion between viral and endosomal membranes, plays a critical role during infection (*5*). In order for a virus to produce infectious progeny, its HA must be cleaved by proteolytic enzymes into HA1 and HA2 subunits (*6, 7*). Highly pathogenic avian viruses possess multiple basic amino acid inserts at the HA cleavage site, making it accessible to intracellular proteases (*8, 9*), in contrast to nonpathogenic viruses which possess only two basic amino acids at the HA cleavage site and are not susceptible to these proteases. Some H5 and H7 avian influenza viruses have been characterized as potentially pathogenic strains (*10, 11*); that is, although they are nonpathogenic, a single point mutation in the HA gene can increase the cleavability of their HA molecules leading to increased pathogenicity.

The source of highly pathogenic avian influenza A viruses remains uncertain. There is strong evidence for transmission from wild birds to domestic poultry, but it is unclear whether highly pathogenic strains constitute unique lineages, which are maintained in a host population where these viruses are nonpathogenic, or whether they are derived from the same ancestors as their nonpathogenic counterparts. To address this question, we determined the phylogenetic relationships among nucleotide and deduced amino acid sequences of the HA1 genes of 15 H5 and 26 H7 viruses, including nonpathogenic and highly pathogenic isolates. A compelling reason to study H7 viruses is that they also affect mammals, including horses and seals, therefore raising the question if viruses with an avian HA have the potential to spread into mammalian hosts.

Our intent was to choose viruses representing a spectrum of geographical locations, dates of isolation, and host species. The viruses and abbreviations which are

[1] To whom reprint requests should be addressed.

0042-6822/95 $6.00

used in place of complete strain designations are listed in Table 1. The RNA was extracted by a phenol:chloroform procedure (*12*) and amplified by PCR (*13*). The nucleotide sequences of the 17 HA genes were determined by the dideoxynucleotide-chain termination method (*14*), using the fmol DNA sequencing system (Promega, Madison, WI). Primer sequences are available upon request. The nucleotide sequences have been submitted to GenBank. HA sequences of 24 viruses were obtained from GenBank or literature. Accession numbers or references are listed in Table 1.

The HA1 genes from 10 nonpathogenic avian H7 influenza A viruses, each comprising 1038 nucleotides, were found to have a single open reading frame that extended from position 22 to the end of the HA1 and encoded a polypeptide of 339 amino acids. Both of the highly pathogenic viruses (TyEn63 and CkLe79) were distinguished from the nonpathogenic strains by a 6-nucleotide insertion at the HA cleavage site. Analysis of the deduced amino acid sequences for H7 HA1s identified the first 18 amino acids as the N-terminal signal peptide. Each HA1 subunit from nonpathogenic viruses comprised 321 amino acids, as compared with 323 in the HA1s of TyEn63 and CkLe79. As predicted by nucleotide analysis, HA cleavage sites in nonpathogenic H7 strains included only two basic amino acids (K-X-R), whereas TyEn63 and CkLe79 possessed multiple basic amino acids (K-R-R-R-R and K-K-K-K-R), which is in accord with the HA sequences of the other avian H7 viruses used in this study.

The HA1 gene of the H5N2 virus DkHK78 comprised 1026 nucleotides, as did the same gene from other nonpathogenic H5 viruses examined in this study. The pathogenic TeSA61 strain, by contrast, had a 12-nucleotide insertion at the cleavage site. The two HA1 genes encoded proteins comprising 342 and 346 amino acids, respectively. DkHK78 featured a cleavage site sequence typical for nonpathogenic avian H5 viruses (R-E-T-R), whereas the pathogenic TeSA61 was characterized by multiple basic amino acids at this site (R-E-T-R-R-Q-K-R) like the other pathogenic H5 viruses used in this study. Only partial sequences were obtained for the HA genes of four nonpathogenic H5 isolates: i.e., nucleotide positions 545–775 (TyCO72, DkMI80, and DkPA84) and 77-355 (ShAU75).

The phylogenetic analysis was performed by the maximum parsimony method to determine the minimum number of mutations needed to account for the sequence differences (*16*). The shortest tree connecting the 26 H7 HA nucleotide sequences required 1592 steps (Fig. 1A). We obtained three slightly different trees of the same length; the only topological features affected were the terminal branches in the North American avian lineage. The primary analysis shows (i) the divergence of H7 HAs into two geographically distinct avian lineages, (ii) a close phylogenetic relationship between pathogenic and nonpathogenic avian viruses that includes common progenitors, and (iii) a distinct equine lineage that shares a common ancestor with all Eurasian avian viruses.

Due to their geographical separation, the HA genes of avian H7 influenza A viruses diverge into two major lineages, one restricted to North American viruses and the other to viruses from Eurasia, Africa, and Australia. The latter (or Eurasian) lineage splits into two major branches, one formed by early European viruses that circulated at the beginning of this century and the other by more recent strains from Eurasia and Africa. Interestingly, the Australian avian viruses are diverging from an ancestor shared with the early European viruses toward a separate Australian sublineage.

Evaluation of the position of highly pathogenic avian viruses in this phylogenetic tree was complicated by the absence of nonpathogenic counterparts for the early pathogenic isolates. Hence, we found it necessary to focus on more recent isolates, none of which form a distinct lineage relative to nonpathogenic strains, but rather share common ancestors with closely related nonpathogenic strains. The pathogenic Australian virus CkVi85, for example, arose from the same progenitor as the nonpathogenic DkVi76 strain also isolated in Australia. A similar relationship was found between a pathogenic German virus, CkLe79, and two of its benign relatives from the same geographical area (TePo79 and SnPo81). Although located on a separate branch, TyEn63 also diverged from a common ancestor shared with all the more recent Eurasian and African viruses, most of which are nonpathogenic.

The mammalian strains in this study represent a distinct equine lineage and a single seal isolate in the North American branch of the H7 HA tree (Fig. 1). Because of the topological variation in the North American branch, we could not trace the HA gene of SeMA80 to a single hypothetical ancestor; however, SeMA80 is linked to the North American avian strains with an external node that emphasizes the recent divergence of this gene from an ancestor common to avian viruses.

The equine H7 viruses form a distinct lineage that is linked to the Eurasian avian lineage; an internal node represents a hypothetical ancestor shared with the HA genes of all avian viruses, except the North American isolates, as well as equine viruses worldwide. The equine HAs differ by only a few nucleotide changes; however, two sublineages are apparent, one containing the prototype-like viruses EqPr56 and EqCa63, and the other more recent equine isolates.

By comparing the nucleotide and amino acid tree (Figs. 1A and 1B), one finds that the numbers of amino acid changes relative to the numbers of nucleotide changes differ between avian and mammalian strains. For the equine H7 lineage, an average of 1.67 nucleotide

TABLE 1

Influenza A Virus Strains Used for the Phylogenetic Analysis

Strain	Abbreviation	Pathogenicity	GenBank No. or reference
Avian H5 viruses			
A/chicken/Scotland/59 (H5N1)	CkSc59	+[a]	X07826
A/tern/South Africa/61 (H5N3)[b]	TeSA61	+	U20460 (this study)
A/turkey/Ontario/7732/66 (H5N9)	TyON66	+	M30122
A/turkey/Colorado/72 (H5N2)[b,c]	TyCO72	−	U20472 (this study)
A/shearwater/Australia/75 (H5N3)[c]	ShAU75	−	J02160
A/duck/Hong Kong/342/78 (H5N2)[b]	DkHK78	−	U20475 (this study)
A/duck/Michigan/80 (H5N2)[b,c]	DkMI80	−	U20474 (this study)
A/chicken/Pennsylvania/1/83 (H5N2)	CkPA1/83	−	M18001
A/chicken/Pennsylvania/1370/83 (H5N2)	CkPA13/83	+	M10243
A/duck/Ireland/113/83 (H5N8)	DkIr83	+	M18450
A/turkey/Ireland/1378/83 (H5N8)	TyIr83	+	M18451
A/duck/Pennsylvania/84 (H5N2)[b,c]	DkPA84	−	U20473 (this study)
A/ruddy turnstone/Delaware/244/91 (H5N2)	RTDE91	−	U05330
A/chicken/Florida/25717/93 (H5N2)	CkFL93	−	U05332
A/chicken/Pennsylvania/13609/93 (H5N2)	CkPA93	−	U05331
Avian H7 viruses			
A/chicken/Brescia/1902 (H7N7)	CkBr02	+	U20471 (this study)/ (15)[d]
A/FPV/Weybridge/27 (H7N7)	FPVW27	+	(15)
A/FPV/Rostock/34 (H7N1)	FPVR34	+	M24457
A/turkey/England/63 (H7N3)[b]	TyEn63	+	U20462 (this study)
A/turkey/Oregon/71 (H7N3)	TyOR71	−	M31689
A/duck/Victoria/76 (H7N?)[e]	DkVi76	−	U20463 (this study)
A/duck/Hong Kong/293/78 (H7N2)[b]	DkHK78	−	U20461 (this study)
A/chicken/Leipzig/79 (H7N7)[f,g]	CkLe79	+	U20459 (this study)
A/tern/Potsdam/342/6/79 (H7N7)[f]	TePo79	−	U20470 (this study)
A/turkey/Minnesota/1237/80 (H7N3)[b]	TyMN80	−	U20466 (this study)
A/swan/Potsdam/63/6/81 (H7N7)[f]	SnPo81	−	U20467 (this study)
A/chicken/Victoria/1/85 (H7N7)	CkVi85	+	M17735
A/duck/Heinersdorf/S495/6/86 (H7N7)[f]	DkHe86	−	U20465 (this study)
A/chicken/Jena/1816/87 (H7N7)[f]	CkJe87	−	U20469 (this study)
A/ostrich/South Africa/5352/92 (H7N1)[b]	OSA92	−	U20458 (this study)
A/softbill/South Africa/142/92 (H7N4)[b]	SbSA92	−	U20464 (this study)
A/rhea/North Carolina/39482/93 (H7N1)[b]	RhNC93	−	U20468 (this study)
Mammalian H7 viruses			
A/seal/Massachusetts/1/80 (H7N7)	SeMA80		K00429
A/equine/Prague/1/56 (H7N7)	EqPr56		X62552
A/equine/Cambridge/1/63 (H7N7)	EqCa63		X62553
A/equine/Detroit/1/64 (H7N7)	EqDe64		X61627
A/equine/Lexington/1/66 (H7N7)	EqLe66		X62556
A/equine/Switzerland/137/72 (H7N7)	EqCH72		X62557
A/equine/London/1416/73 (H7N7)	EqLo73		M58657
A/equine/Sao Paulo/1/76 (H7N7)	EqSP76		X62559
A/equine/Newmarket/1/77 (H7N7)	EqNM77		X62554

[a] +, Highly pathogenic; −, nonpathogenic.
[b] From the repository of St. Jude Children's Research Hospital (Memphis, TN).
[c] Partial sequence.
[d] Nucleotides 766–831 were done in this study to complete the sequence published earlier (*15*).
[e] From the repository of Commonwealth Scientific and Industrial Research Organization, Australian Animal Health Laboratory.
[f] From the repository of the Bundesinstitut für Veterinärmedizin (Berlin, Germany).
[g] Virus was isolated by the former Bezirksinstitut für Veterinärmedizin (Leipzig, Germany).

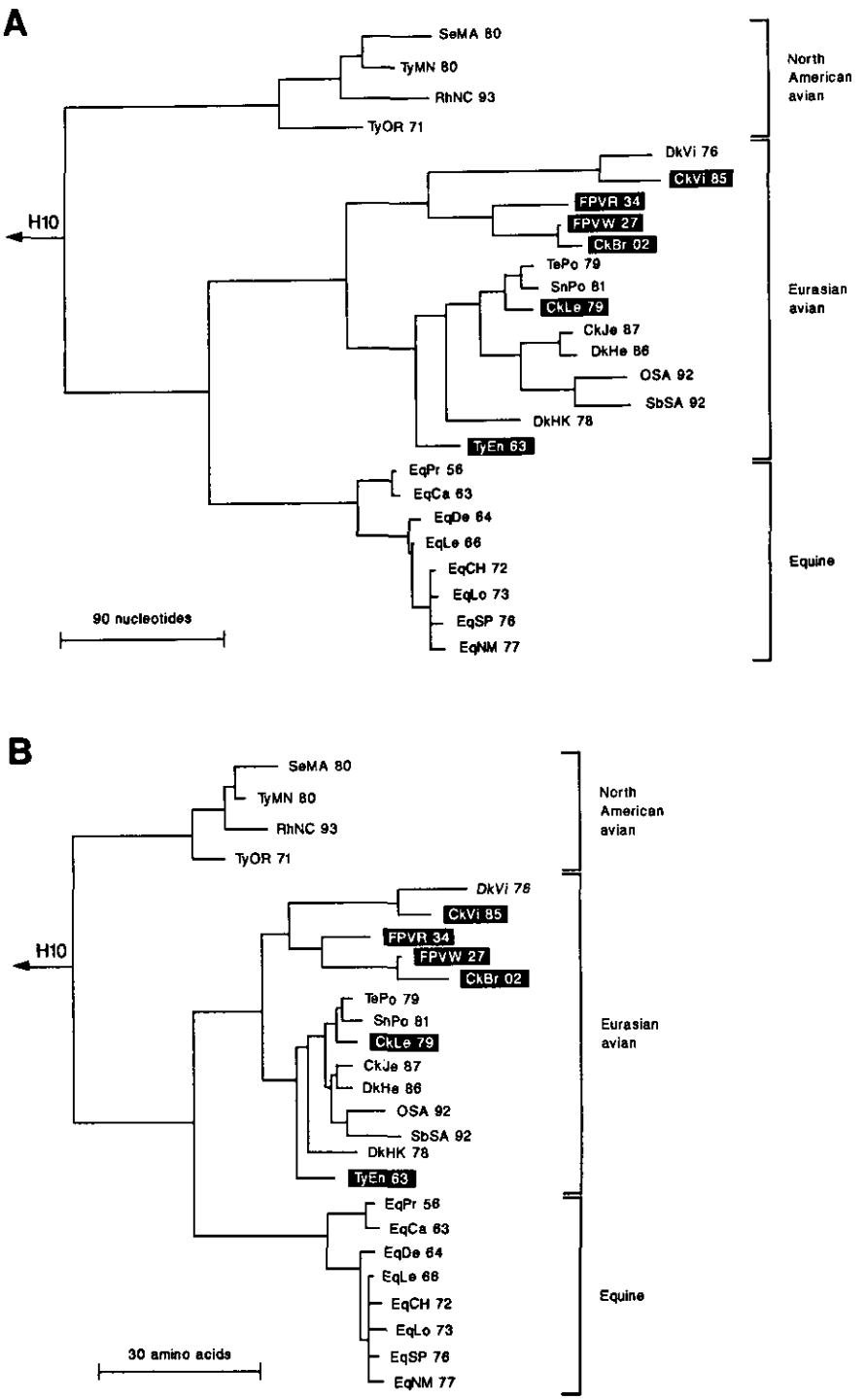

FIG. 1. Phylogenetic trees of the H7 influenza virus HA genes. The phylogenetic tree constructed for the 26 H7 viruses included complete sequences of the HA1 portion of the HA gene and 51 nucleotides of the HA2 portion; the tree was rooted to the H10 HA1 sequence from A/chicken/Germany/N/49 (H10N7) (GenBank No. M21646). The signal peptides (54 nucleotides) and the insertions at the cleavage sites were excluded, as were two insertions (three nucleotides each) in the H10 HA sequence. The phylogenetic analysis was performed by the maximum parsimony method (*16*). Searches for the most parsimonious topologies were done with PAUP software, version 2.4 (David L. Swofford, Illinois Natural History Survey, Champaign, IL). The nucleotide tree (A) requires 1592 nucleotide changes, while the amino acid tree (B) requires 426 amino acid changes. Horizontal distances are proportional to the number of nucleotide and amino acid changes required to join nodes and H7 HA sequences. Vertical lines are for spacing branches and labels. The arrow at the left indicates the direction of the A/chicken/Germany/N/49 from the root node. Abbreviations of viruses are listed in Table 1. Pathogenic viruses are marked with black boxes.

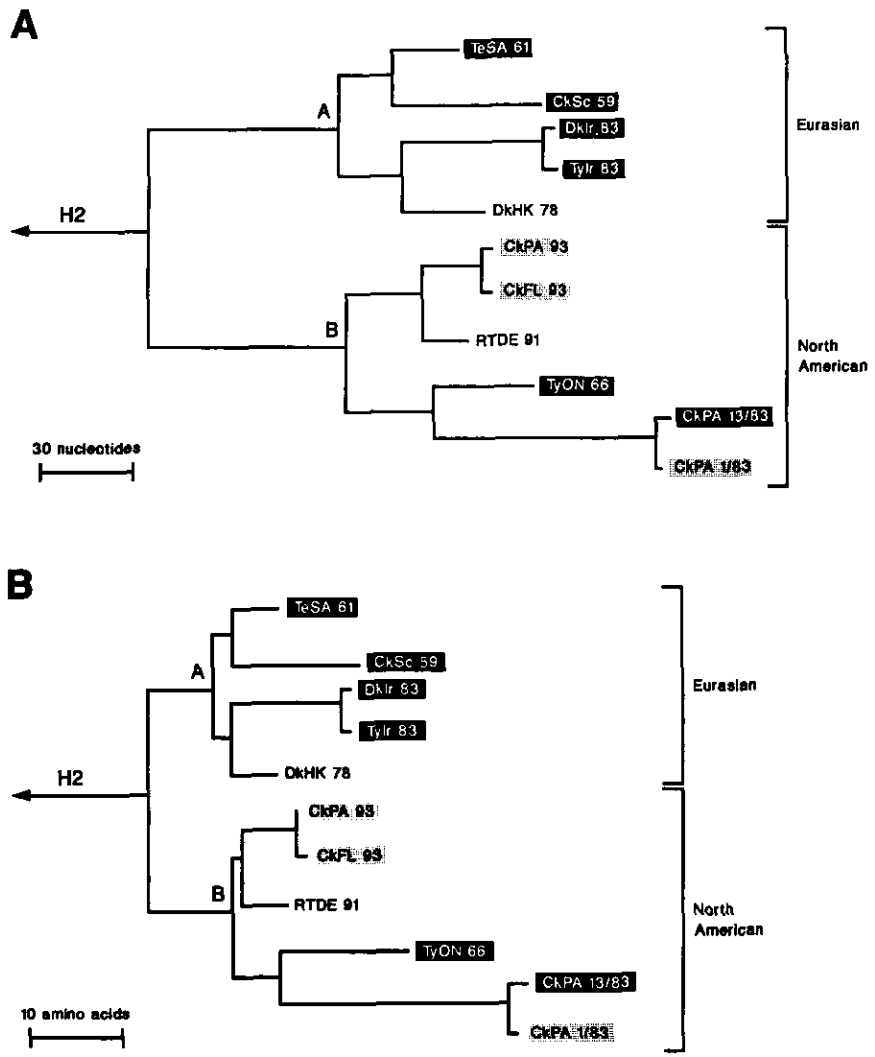

FIG. 2. Phylogenetic trees of the H5 influenza virus HA genes. The phylogenetic tree of the H5 viruses included 11 sequences of the HA1 subunit and was rooted to the H2 HA sequence of A/pintail/Praimoric/625/76 (H2N2) (GenBank No. L11141). The signal peptides (54 nucleotides) and 15 nucleotides at the cleavage site were excluded, as were three nucleotides inserted into the H5 HA1 sequences. An additional three nucleotides were excluded because of a deletion in the HA gene of CkPA 1/83 and CkPA 13/83. The nucleotide tree (A) requires 750 nucleotide changes, while the amino acid tree (B) requires 200 amino acid changes. Horizontal distances are proportional to the number of nucleotide and amino acid changes required to join nodes and H5 HA sequences. Vertical lines are for spacing branches and labels. The arrow at the left indicates the direction of the A/pintail/Praimoric/625/76 (H2N2) HA gene from the root node. Abbreviations of viruses are listed in Table 1. Pathogenic viruses are marked with black boxes, viruses with a potential to become pathogenic with grey boxes.

changes in the terminal branches leads to one amino acid change. This result contrasts with findings for equine H3 viruses, which accumulate 2.42 nucleotide changes per each coding change (*17*). Avian H7 viruses exhibit greater disparity between nucleotide and amino acid changes. That is, an average of 3.71 mutations is acquired for each coding change in the terminal branches.

In the phylogenetic analysis of the H5 viruses the shortest path connecting the 11 H5 HA sequences was 750 steps for the nucleotide sequences and 200 steps for amino acid sequences (Figs. 2A and 2B). In this analysis, which includes only avian influenza viruses, we again found a strict separation into North American and Eurasian lineages. Both branches of the phylogenetic tree include highly pathogenic as well as nonpathogenic viruses with common ancestors. To demonstrate the mixed character of both avian lineages, we performed additional analyses with partial HA sequences, including several from nonpathogenic H5 viruses not examined earlier (Fig. 3). The two pathogenic viruses from Ireland (DkIr83, TyIr83) share a common ancestor with a nonpathogenic isolate from Hong Kong (DkHK78) (Fig. 2). The other pathogenic isolates in the Eurasian branch (TeSA61, CkSc59) diverged from the same ancestor as the nonpathogenic ShAU75 (Fig. 3A). The two pathogenic viruses

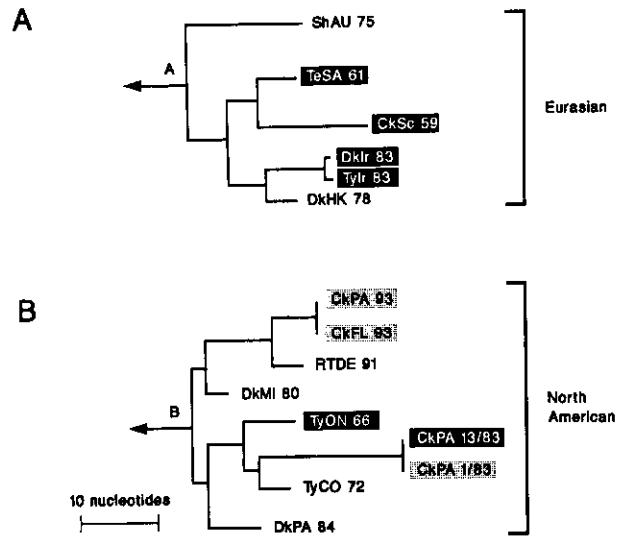

FIG. 3. Hypothetical Eurasian (A) and North American (B) branches of phylogenetic trees constructed with partial H5 HA nucleotide sequences. Nucleotides 77–355 of the HA1 were used for the Eurasian branch, and nucleotides 545–775 for the North American branch. In each case the viruses of the other branch were used as an outgroup. The arrows at the left indicate the direction of the outgroup from the root node. Nodes A and B correspond to the same nodes in Fig. 2.

from North America (CkPA13/83 and TyON66) both evolved from the same progenitors as did the nonpathogenic TyCO72 and DkPA84 strains (Fig. 3B). Moreover, three of the H5 viruses (CkPA93, CkFL93, and CkPA1/83) occupy an intermediate position with respect to pathogenicity, in that a single point mutation in their HA genes could lead to greatly increased cleavability (*10, 11*). These intermediate strains are found in different sublineages. CkFL93 and CkPA93 are closely related to a nonpathogenic strain, RTDE91, as previously described (*11*), whereas CkPA1/83 is located next to CkPA13/83, which caused a devastating influenza outbreak in poultry (Fig. 2).

Comparison of the H5 nucleotide and amino acid trees (Figs. 2A and 2B) yielded similar results to those for avian H7 viruses. In the terminal branches of the phylogenetic trees, avian H5 viruses are characterized by a ratio of nucleotide to amino acid changes of 3.18.

The central aim of this study was to define the genetic relationship between the HAs of highly pathogenic and nonpathogenic avian influenza A viruses of the H5 and H7 subtypes. This information was then used to address a key question in contemporary influenza virus research: Are the HA genes of pathogenic viruses derived from progenitors shared with their nonpathogenic counterparts in the wild bird population or do they represent separate and distinct lineages which may be maintained in a different host species? Our phylogenetic results (Figs. 1–3) clearly support the first alternative. Although we could not trace the origin of the early European H7 viruses (CkBr02, FPVW27, FPVR34) due to a lack of nonpathogenic counterparts, study of the HA genes of more recent avian H5 and H7 viruses revealed a common linkage of pathogenic and nonpathogenic viruses to one hypothetical ancestor. For some highly pathogenic viruses, such as CkLe79, CkVi85, TyIr83, DkIr83, and CkPA13/83, we found a direct relationship to nonpathogenic viruses from the same geographical region; others were linked to earlier ancestors from which an entire branch of pathogenic and nonpathogenic viruses emerged (TyEn63, TeSA61, CkSc59, and TyON66). These findings predict that the HAs of nonpathogenic avian viruses which hypothetically circulated in Europe at the beginning of this century would constitute a sublineage with the pathogenic early European strains that are included in our survey.

Thus far, we have not been able to identify a progenitor virus for the mixed lineages containing pathogenic and nonpathogenic viruses. But, we do have H5 viruses whose HA genes occupy an intermediate position in terms of their cleavability (hence pathogenicity): two isolates from live bird markets, CkPA93 and CkFL93 (*11*), share a common ancestor for their HAs with the nonpathogenic shorebird isolate RTDE91 (Fig. 2); CkPA1/83 and its pathogenic counterpart CkPA13/83 carry HAs derived from a common ancestor with the nonpathogenic TyCO72 (Fig. 3B). The phylogenetic relationships in these two cases strongly suggest that HA genes introduced into chickens from nonpathogenic shorebird or turkey isolates can mutate unfavorably, resulting in highly pathogenic new strains of influenza virus. Thus, an avian influenza A virus with a highly cleavable HA could arise from the wild bird reservoir at any time, and at any geographical site, and produce another devastating influenza outbreak in poultry. Therefore, all H5 and H7 influenza viruses are potentially suspect for becoming pathogenic.

Another important question is whether there are instances in the evolutionary history of avian H7 influenza A viruses to indicate a potential for spread into mammalian hosts? Our phylogenetic data suggest at least two such events: The first appears to have occurred at about the time the equine H7 lineage diverged from the Eurasian avian lineage by a single introduction of an HA gene from the avian Eurasian gene pool into the horse population; the second was the relatively recent introduction of an avian HA into the North American seal population. We would propose from our findings that viruses carrying the ancestral HA gene of the currently circulating avian H7 strains had the potential to establish a continuing lineage in equine hosts. Furthermore, H7 viruses with an avian HA, which is probably still circulating in the wild bird reservoir, have the potential to adapt to seals. Thus, we should be aware of the possibility of a spread of hazardous H7 viruses to mammalian populations.

In this article, we suggest that the HA genes of highly pathogenic and nonpathogenic avian H5 and H7 viruses evolve from a common ancestor in wild aquatic birds. As we lack the means to eradicate this virus pool, we can expect to see new H5 and H7 viruses in domestic poultry which are potentially suspect for becoming pathogenic. Therefore, consideration should be given to eradication of any H5 or H7 viruses which emerge in domestic poultry.

ACKNOWLEDGMENTS

We thank Drs. Volker Pohle and A. J. Della-Porta for providing viruses, Krisna Wells and Scott Krauss for excellent technical assistance, Pat Eddy for computer support, John Gilbert for scientific editing, and Dayna Anderson for typing the manuscript. This work was supported by Public Health Service Grants AI-08831, AI-29599, and AI-29680 from the National Institute of Allergy and Infectious Diseases, by Cancer Center Support (CORE) Grant CA-21765, and by the American Lebanese Syrian Associated Charities (ALSAC).

REFERENCES

1. Hinshaw, V. S., Webster, R. G., and Turner, B., *Can. J. Microbiol.* **26,** 622–629 (1980).
2. Süss, J., Schäfer, J., Sinnecker, H., and Webster, R. G., *Arch. Virol.* **135,** 101–114 (1994).
3. Rott, R., Orlich, M., and Scholtissek, C., *J. Virol.* **19,** 54–60 (1976).
4. Scholtissek, C., Rott, R., Orlich, M., Harms, E., and Rohde, W., *Virology* **81,** 74–80 (1977).
5. Webster, R. G., and Rott, R., *Cell* **50,** 665–666 (1987).
6. Klenk, H.-D., Rott, R., Orlich, M., and Blödorn, J., *Virology* **68,** 426–439 (1975).
7. Rott, R., Reinacher, M., Orlich, M., and Klenk, H.-D., *Arch. Virol.* **65,** 123–133 (1980).
8. Bosch, F. X., Garten, W., Klenk, H.-D., and Rott, R., *Virology* **113,** 725–735 (1981).
9. Kawaoka, Y., Nestorowicz, A., Alexander, D. J., and Webster, R. G., *Virology* **158,** 218–227 (1987).
10. Kawaoka, Y., Naeve, C. W., and Webster, R. G., *Virology* **139,** 303–316 (1984).
11. Saito, T., Horimoto, T., Kawaoka, Y., Senne, D., and Webster, R. G., *Virology* **201,** 277–284 (1994).
12. Bean, W. J., Sriram, G., and Webster, R. G., *Anal. Biochem.* **102,** 228–232 (1980).
13. Katz, J. M., Wang, M., and Webster, R. G., *J. Virol.* **64,** 1808–1811 (1990).
14. Sanger, F., Nicklen, S., and Coulson, A. R., *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467 (1977).
15. Klimov, A., Prösch, S., Schäfer, J., and Bucher, D., *Arch. Virol.* **122,** 143–161 (1992).
16. Fitch, W. M., *Syst. Zool.* **20,** 406–416 (1971).
17. Bean, W. J., Schell, M., Katz, J., Kawaoka, Y., Naeve, C., Gorman, O., and Webster, R. G., *J. Virol.* **66,** 1129–1138 (1992).