Sequence Analysis of the Matrix (M2) Protein Gene of Avian Pneumovirus Recovered from Turkey Flocks in the United States

Arshud M. Dar,^{1,2} Shirin Munir,^{1,2} Sagar M. Goyal,³ and Vivek Kapur^{1,2*}

*Biomedical Genomics Center,¹ Department of Veterinary Pathobiology,² and Veterinary Diagnostic Medicine,³
College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota 55108*

Received 17 June 2002/Returned for modification 18 December 2002/Accepted 9 March 2003

We here report the comparative sequence and phylogenetic analysis of the avian pneumovirus subgroup C (APV C) matrix (M2) gene of cell culture-adapted isolates and clinical samples. Limited heterogeneity was observed among the M2 sequences, suggesting that diagnostic tests and vaccines against APV C are likely to exhibit broad cross-reactivity.

Pneumoviruses are associated with respiratory tract infections in humans and animals (5, 9). Avian pneumovirus (APV) established itself in U.S. turkeys as an acute respiratory tract infection in late 1996 (4, 12). Comparative sequence analysis of the N, P, M, and F genes from cell culture-adapted viruses has shown that the U.S. virus is distinct from previously described members of European APV subgroups A and B (APV A and APV B, respectively) (5) and constitutes a third subgroup of APV designated APV subgroup C (APV C) (2, 3, 10, 11; A. M. Dar, S. M. Goyal, and V. Kapur, Abstr. 79th Conf. Res. Workers Anim. Dis., abstr. P101, 1998).

APV was considered the only member of the genus *Metapneumovirus* (15) until the recent discovery of a human metapneumovirus that is more closely related to APV C than to the human pneumoviruses (13). Prior to this report, metapneumoviruses were not associated with infection or disease in mammalian species. Based on a sequence analysis of all open reading frames (ORFs) and on gene organization, the human metapneumovirus shares the highest degree of homology with APV C, which is prevalent only in the United States (13, 14). APV C is therefore thought to be a potential evolutionary source of the human metapneumovirus, although this has not been directly demonstrated. Thus, there is a critical need to develop a better understanding of the molecular epidemiology of this emerging virus as well as sensitive and specific diagnostic tools and broad-spectrum vaccines against all strains of APV C found in the United States.

We here present the complete sequence and phylogenetic analysis of the matrix (M2) genes of three cell culture-adapted isolates (Colorado isolate, APV/CO; Minnesota isolates, APV/MN-1a and APV/MN-1b) and 10 APV-positive clinical samples obtained from U.S. turkey flocks. Previous epidemiological investigations on the F and M protein gene sequences of members of APV C have focused on cell culture-adapted APV isolates (10, 11) that might include sequence variations incorporated as a result of in vitro propagation and adaptation. Nucleotide substitutions were observed in the N and P genes of APV in the non-cell-culture-adapted clinical samples when

they were compared with the N and P genes of cell culture-adapted isolates (3).

The M2 gene has two ORFs (ORF1 and ORF2). The protein products of the M2 gene are involved in the regulation of transcription and the replication of the virus (1). The present investigation describes the comparative analysis of the M2 gene sequences of APV C strains with previously described avian, human, and bovine pneumovirus sequences.

APV/CO was obtained from the National Veterinary Services Laboratories (Ames, Iowa), and APV/MN-1a and APV/MN-1b were initially isolated and maintained in our laboratory (4). Ten clinical samples (tracheal and esophageal swabs) were obtained from APV-affected turkeys from July to November 1998 and were found to be positive for APV as detected by reverse transcription-PCR (2).

Viral RNA was extracted and subjected to reverse transcription-PCR with primers based on previously described APV sequences (GenBank accession number AF085228) (Dar et al., Abstr. 79th Conf. Res. Workers Anim. Dis.). Purified PCR products were sequenced either directly or after they were cloned into a plasmid vector (pGEM-T; Promega, Madison, Wis.). The DNA sequences were assembled and analyzed with EditSeq, MegAlign, and SeqMan computer programs (DNA-STAR, Madison, Wis.). The M2 gene sequences of avian and nonavian pneumoviruses used for analysis with their GenBank accession numbers are as follows: APV/CO, APV/MN-1a, and APV/MN-1b, accession numbers AF176592, AF262572, and AF262573, respectively; APV A strain 3B/85, accession number X63408; human respiratory syncytial virus subgroup A (HRSV-A), accession number U39662; and bovine respiratory syncytial virus (BRSV), accession number M82816. Phylogenetic analysis was performed with PAUP version 4.02b (D. L. Swofford, Illinois Natural History Survey, University of Illinois, Urbana-Champaign, 1999) by the nearest-neighbor interchange option with branch swapping and 1,000 bootstrap replications. The number of synonymous changes per synonymous site (d_s) and the number of coding changes per nonsynonymous site (d_n) in the coding regions were calculated by the unweighted method of Nei and Gojobori (8).

The M2 gene in the three isolates and 10 clinical samples was present downstream (towards the 5' end) of the F gene, a gene order that is similar to that of APV A. The M2 gene in the

* Corresponding author. Mailing address: Department of Microbiology and Biomedical Genomics Center, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108. Phone: (612) 625-7712. Fax: (612) 625-5203. E-mail: vkapur@umn.edu.

FIG. 1. (A) Alignment of the deduced amino acid sequences of the M2 ORF1 proteins of APV/CO (GenBank accession number AF176592), APV/MN-1a (AF262572), APV/MN-1b (AF262573), APV A strain UK/3B 85 (X63408), HRSV A (U39662), and BRSV (M82816). The sequences were aligned using the CLUSTAL V method (DNASTAR). Amino acid sequence identities with the APV C sequence are indicated by periods, and polymorphic sites are indicated by the substituting amino acid. The deletions are shown by minus signs. (B) Alignment of the deduced amino acid sequences of the M2 ORF2 proteins of APV/CO (AF176592), APV/MN-1a (AF262572), APV/MN-1b (AF262573), APV A strain UK/3B 85 (X63408), HRSV A (U39662), and BRSV (M82816). The sequences were aligned as described above. Amino acid sequence identities with the APV/CO sequence are indicated by periods, and polymorphic sites are indicated by the substituting amino acid. The deletions are shown by minus signs.

three laboratory isolates was 1 bp longer (751 bp) than that in the nonpropagated in clinical samples at 750 bp, revealing a deletion of 1 bp (APV A) at position 749. ORF1 and ORF2 in all 13 APVs (APV C) partially overlapped in the M2 gene from positions 15 to 569 and 526 to 741, respectively, and encoded two proteins designated M2-1 and M2-2, respectively. In all 13 viruses, the lengths of the M2 genes (750 to 751 nucleotides) were 40 to 41 nucleotides shorter than that of the M2 gene of APV A (7).

The predicted APV C M2-1 and M2-2 proteins contained 184 and 71 amino acid residues (Fig. 1) with a predicted molecular mass of ~21 and ~8 kDa, respectively. The length of the predicted M2-1 protein (184 residues) was similar to that of the corresponding product of APV A, whereas the predicted APV C M2-2 protein (71 residues) was 2 residues shorter than that of APV A. The comparison of nucleotide sequences of APV/CO ORF1 and ORF2 with those of APV A showed 193 and 96 polymorphic sites and a total nucleotide diversity (π) of 0.3496 ± 0.0203 and 0.5333 ± 0.0372 , respectively. The d_n/d_s ratio was determined to evaluate the nature of the selection pressure driving the evolution of the APV M2 gene. This ratio was 0.85 and 3.6 for ORF1 and ORF2, respectively. Different d_n/d_s values for ORF1 and ORF2 of the M2 gene probably suggest that purifying selection acts to maintain the primary structure of ORF1 and that positive selection occurs in ORF2 (6). The biological significance of these observations is not known.

The alignment of the nucleotide sequences of the M2 genes of all members of APV C showed six substitutions in APV/MN-1a and APV/MN-1b and seven substitutions in clinical samples compared with the M2 gene of APV/CO (Table 1). Six of the seven substitutions were common to all clinical samples, but the seventh change at position 467 was noted only in

clinical sample number 8. Four changes in the ORF1 nucleotide sequence of all APVs C were nonsynonymous in nature. The fifth polymorphic site in ORF1 of the M2 gene of clinical sample number 8 was synonymous. The nucleotide change at position 599 in ORF2 in APV/MN-1a and APV/MN-1b was nonsynonymous, whereas another change at position 609 was synonymous. Similarly, clinical samples showed one change at position 594 from C to T and one change at position 731 from

TABLE 1. Positions of polymorphic sites in the M2 genes of U.S. APVs

Virus ^b	Nucleotide at position ^a :									
	100	294	341	452	467	594	599	609	731	749
APV/CO	G	A	C	T	C	C	A	A	T	A
APV/MN-1a	A	G	T	C	.	.	<u>G</u>	G	.	.
APV/MN-1b	A	G	T	C	.	.	<u>G</u>	G	.	.
APV/CS5	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS6	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS7	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS8	A	G	T	C	A	T	.	.	<u>C</u>	—
APV/CS9	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS10	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS11	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS17	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS25	A	G	T	C	.	T	.	.	<u>C</u>	—
APV/CS53	.	A	G	T	C	.	.	.	<u>C</u>	—
APV/CS56	A	G	T	C	.	.	.	.	<u>C</u>	—

^a Nucleotide sequence identities with nucleotides of APV/CO are indicated by periods, deletions are shown by minus signs, and polymorphic sites are indicated by the substituting nucleotide. Underlined residues represent nonsynonymous substitutions.

^b APV/CO, Colorado isolate of APV; APV/MN-1a and APV/MN-1b, Minnesota isolates of APV; APV/CS5 through APV/CS56, non-cell-culture-adapted APV clinical samples.

TABLE 2. Amino acid sequence identities of ORF1 and ORF2 in the M2 genes of isolates of APV^a

Virus ^b	% Amino acid sequence identity in M2 ORF protein of:			
	APV/CO	APV/MN-1a	APV/MN-1b	APV A
APV/CO		98.4	98.4	70.8
APV/MN-1a	97.2		99.5	70.8
APV/MN-1b	97.2	98.6		70.8
APV-A	20.8	22.2	22.2	

^a See the text for accession numbers of M2 ORF1 and ORF2 proteins.

^b APV/CO, Colorado isolate of APV; APV/MN-1a and APV/MN-1b, Minnesota isolates of APV; APV A, European APV A.

^c Numbers above the diagonal indicate percentages of amino acid sequence identity of ORF1, and numbers below the diagonal indicate percentages of amino acid sequence identity of ORF2.

T to C as being synonymous and nonsynonymous substitutions, respectively, in ORF2 of the M2 gene (Table 1).

The Minnesota isolates (APV/MN-1a, APV/MN-1b) showed percentages of amino acid sequence divergence of the M2-1 and M2-2 proteins from the corresponding proteins in APV/CO of 1.6 and 2.8%, respectively. The analysis of the M2-1 and M2-2 proteins of the clinical samples showed 0 and 2.9% amino acid sequence divergence from those of APV/MN-1a and APV/MN-1b, respectively. Nine of 10 substitutions in the nucleotide sequences of the M2 genes of the clinical samples were transitions (A→G or T→C), and 1 was a transversion (C→A) at position 467 in ORF1 of the M2 gene (Table 1). These data suggest the existence of limited genetic heterogeneity between cell culture-adapted APV isolates and clinical samples, the latter representing native field viruses recovered from U.S. turkey flocks.

ORF1 and ORF2 of the M2 gene of APV/CO shared 62 and 34% nucleotide sequence identity, respectively, and 71 and 21% amino acid sequence identity, respectively, to ORF1 and ORF2 of the M2 gene of APV A (Table 2). In contrast, the sequence variation observed in ORF1 and ORF2 in the M2

genes of the APV C isolates and clinical samples was considerably less. ORF1 and ORF2 in the M2 genes of the three isolates (APV/CO, APV/MN-1a, and APV/MN-1b) shared a nucleotide and amino acid sequence identity of ~99% (Table 2). An amino acid sequence alignment of both ORFs of the M2 genes of the three APV isolates revealed the presence of one conservative amino acid replacement (S↔N) in ORF1 at position 29 and one conservative change (K↔R) in ORF2 at position 38 (Fig. 1).

The nucleotide sequence alignment of the M2 genes of APV/CO and APV A showed a large number of insertions and deletions (indels) in the APV/CO ORF2 region. For instance, 36 nucleotide indels were in the nonoverlapping region of ORF2, 6 indels were in the region common to ORF1 and ORF2, and only 1 indel was in ORF1 of APV/CO. An amino acid sequence alignment revealed two deletions at positions 177 to 178 in ORF1 and four deletions at positions 54 to 57 in ORF2 of APV/CO compared with those in APV A (Fig. 1).

Phylogenetic analyses using parsimony of ORF1 and ORF2 in the M2 genes of the three APV isolates and the clinical samples examined here and those of previously described APV A, HRSV, and BRSV suggested the grouping of all APV C (APV/CO, APV/MN-1a, APV/MN-1b, and 10 clinical samples) in one branch (Fig. 2). The phylogenetic analysis of the deduced amino acid sequences of M2 ORF1 and ORF2 in avian and mammalian pneumoviruses also shows that the three APV isolates (APV/CO, APV/MN-1a, and APV/MN-1b) and the clinical samples are related to, but distinct from, APV A. The closer relationship between the three U.S. isolates and the clinical samples with APV A (and not with mammalian pneumoviruses) is also indicated by high bootstrap confidence levels (Fig. 2).

Taken together, the results of the present investigation suggest that there is only limited genetic heterogeneity among APVs recovered from U.S. turkey flocks and that the clinical samples representing native field viruses show limited poly-

FIG. 2. (A) Dendrogram showing phylogenetic relationships among ORF1 proteins of the M2 genes of three APV C isolates and 10 clinical samples with non-U.S. APV A, HRSV, and BRSV. The figure was generated with PAUP version 4.02b (Swofford, 1999). The dendrogram was constructed using the nearest-neighbor interchange option with branch swapping and 1,000 bootstrap replications. The values adjacent to each node represent the percentages of 1,000 bootstrap trees that supported the clustering to the right. (B) Dendrogram showing phylogenetic relationships among ORF2 proteins of the M2 genes of APV C isolates and clinical samples (a total of 13 APVs) with non-U.S. APV A, HRSV, and BRSV. The figure was generated with PAUP version 4.02b (Swofford, 1999), and the dendrogram was constructed as described above.

morphism compared with that of the cell culture-adapted isolates. This information has important implications for the development of improved molecular and immunologic diagnostic tests for APV infections. Furthermore, the present data also provide evidence to support previous findings (3, 10, 11; Dar et al., Abstr. 79th Conf. Res. Workers Anim. Dis.) that the APV C circulating in the United States is closely related to but distinct from previously described European APV A.

This research was supported in part by the Minnesota Turkey Growers Association, the Minnesota Agricultural Experiment Station, and the Rapid Agricultural Response Fund.

REFERENCES

1. Collins, P. L., M. G. Hill, J. Cristina, and H. Grosfeld. 1996. Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc. Natl. Acad. Sci. USA* **93**:81–85.
2. Dar, A. M., K. Tune, S. Munir, B. Panigrahy, S. M. Goyal, and V. Kapur. 2001. PCR-based detection of an emerging avian pneumovirus in US turkey flocks. *J. Vet. Diagn. Investig.* **13**:201–205.
3. Dar, A. M., S. Munir, S. M. Goyal, M. S. Abrahamsen, and V. Kapur. 2001. Sequence analysis of the nucleocapsid and phosphoprotein genes of avian pneumovirus circulating in the US. *Virus Res.* **79**:15–25.
4. Goyal, S. M., S. J. Chiang, A. M. Dar, K. V. Nagaraja, D. P. Shaw, D. A. Halvorson, and V. Kapur. 2000. Isolation of an avian pneumovirus from an outbreak of respiratory illness in Minnesota turkeys. *J. Vet. Diagn. Investig.* **12**:166–168.
5. Juhasz, K., and A. J. Easton. 1994. Extensive sequence variation in the attachment (G) protein of avian pneumovirus: evidence for two distinct subgroups. *J. Gen. Virol.* **71**:2873–2880.
6. Kimura, M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**:275–276.
7. Ling, R., A. J. Easton, and C. R. Pringle. 1992. Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene order different from that of other pneumoviruses. *J. Gen. Virol.* **73**:1709–1715.
8. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the number of synonymous and nonsynonymous substitutions. *Mol. Biol. Evol.* **3**:418–426.
9. Pringle, C. R. 1987. Paramyxoviruses and diseases. *Soc. Gen. Microbiol. Symp.* **40**:91–138.
10. Seal, B. 1998. Matrix protein gene nucleotide and predicted amino acid sequence demonstrate that the first US avian pneumovirus is distinct from European strains. *Virus Res.* **58**:45–52.
11. Seal, B., H. S. Sellers, and R. J. Meinersmann. 2000. Fusion protein predicted amino acid sequence of the first US avian pneumovirus isolate and lack of heterogeneity among other US isolates. *Virus Res.* **66**:139–147.
12. Senne, D. A., R. K. Edson, J. C. Pederson, and B. Panigrahy. 1997. Avian pneumovirus update, p. 190. *Proceedings of the 134th Annual Convention of the American Veterinary Medical Association*, Reno, Nevada. American Veterinary Medical Association, Washington, D.C.
13. van den Hoogen, B. G., J. C. de Jong, J. Groen, T. Kuiken, R. de Groot, R. A. M. Fouchier, and A. D. M. E. Osterhaus. 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* **7**:719–724.
14. van den Hoogen, B. G., T. M. Bestebroer, A. D. M. E. Osterhaus, and R. A. M. Fouchier. 2002. Analysis of the genome sequence of a human metapneumovirus. *Virology* **295**:119–132.
15. van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, et al. (ed.). 2000. *Virus taxonomy: classification and nomenclature of viruses*. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.