Published in final edited form as:

Virus Res. 2009 October ; 145(1): 80–91. doi:10.1016/j.virusres.2009.06.003.

Complete genome sequence of avian paramyxovirus type 7 (strain Tennessee) and comparison with other paramyxoviruses

Sa Xiao¹, Anandan Paldurai¹, Baibaswata Nayak¹, Madhuri Subbiah¹, Peter L. Collins², and Siba K Samal^{1,*}

¹Virginia-Maryland Regional College of Veterinary Medicine, University of Maryland, College Park, Maryland, USA

²Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, USA

Abstract

The complete genome sequence of avian paramyxovirus serotype 7 (APMV-7) prototype strain dove/Tennessee/4/75 was determined. The genome size is 15,480 nucleotides (nt) long and follows the “rule of six”. The genome contains six non-overlapping genes in the order of 3'-N-P/V/W-M-F-HN-L-5'. The 3' leader and 5' trailer sequences of the genome are 55 nt and 127 nt long, respectively. The first 12 nt of the leader and trailer sequences are complementary to each other. The viral genes are flanked by highly conserved gene-start (GS) and gene-end (GE) transcription signals, and in addition the 3' leader sequence contains a sequence (³⁵AAUUAUUUUUU⁴⁵) that is identical to the GE signal present at two of the genes. The genes are separated by intergenic sequences (IGS) ranging between 11 to 70 nt. The phosphoprotein (P) gene contains a conserved RNA editing site (3'-UUUUUCCC-5') presumed to be involved in the production of V and W proteins. The viral fusion (F) protein has a single basic amino acid at the putative cleavage site (¹⁰¹TLPSSR↓F¹⁰⁷); however, the virus did not require exogenous protease for *in vitro* replication. The virus grew in only a few established cell lines, indicating a restricted host range. Sequence alignment and phylogenetic analysis of the predicted amino acid sequence of APMV-7 proteins with the cognate proteins of the viruses of all five genera of the family *Paramyxoviridae* showed that APMV-7 is more closely related to APMV-2, -6, -8 than to APMV-1, -3, -4 and -9. The mean death time in embryonated chicken eggs was found to be more than 144 hrs, indicating APMV-7 to be avirulent for chickens.

Keywords

Avian paramyxovirus; genome sequence; APMV-7; strain Tennessee

1. Introduction

Paramyxoviruses are large pleomorphic enveloped viruses possessing a non-segmented, single-stranded, negative-sense RNA genome. These viruses have been isolated from many

Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.

*Corresponding author: Virginia-Maryland Regional College of Veterinary Medicine, University of Maryland, College Park, MD 20742. Phone: (301) 314-6813. Fax: (301) 314-6855. ssamal@umd.edu .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

species of terrestrial animals, fish, birds, and humans around the world (Lamb and Parks, 2007; Nylund *et al.*, 2008). Paramyxoviruses belong to the family *Paramyxoviridae* of the order *Mononegavirales*. The family is divided into two subfamilies, *Paramyxovirinae* and *Pneumovirinae*, based on structure, genome organization, and sequence relatedness (Lamb *et al.*, 2005). Subfamily *Paramyxovirinae* comprises of five genera; *Respirovirus* (including Sendai virus [SeV] and human parainfluenza virus types 1 and 3 [HPIV-1 and -3]), *Rubulavirus* (including simian virus type 5 [SV5], mumps virus [MuV], and human parainfluenza virus types 2 and 4 [HPIV-2 and -4]), *Morbillivirus* (including measles [MeV] and canine distemper [CDV] viruses), *Henipavirus* (including Hendra [HeV] and Nipah [NiV] viruses), and *Avulavirus* (comprising nine serotypes of avian paramyxoviruses [APMV-1 to -9]). Subfamily *Pneumovirinae* contains two genera, *Pneumovirus* (comprising human respiratory syncytial virus [HRSV] and its animal counterparts) and *Metapneumovirus* (comprising human metapneumovirus [HMPV] and its avian counterpart [AMPV]).

The genomes of all paramyxoviruses range between 15-19 kb in length and contain 6-10 genes that encode up to 12 different proteins (Lamb and Parks, 2007). The total genome length of members of subfamily *Paramyxovirinae* is even multiples of six, known as 'rule of six', which is necessary for efficient RNA replication (Calain and Roux, 1993; Kolakofsky *et al.*, 1998; Samal and Collins, 1996). At the 3' and 5' ends of the genome are short extragenic regions, called leader and trailer, respectively. Each gene starts with a conserved gene start (GS) sequence and ends with a conserved gene-end (GE) sequence. Transcription begins at the 3' leader region and proceeds in a sequential manner by a start-stop mechanism guided by GS and GE signals (Lamb and Parks, 2007). Between the gene boundaries, there are non-coding intergenic sequences (IGS), which are not copied into mRNAs. Viral RNA replication occurs when the GS and GE signals are ignored and a copy (anti-genome) of the genome is synthesized, which serves as the template for synthesis of the progeny genome.

All paramyxoviruses examined to date encode a nucleoprotein (N), a phosphoprotein (P), a matrix protein (M), a fusion glycoprotein (F), a hemagglutinin-neuraminidase (HN) or a major glycoprotein (G) and a large polymerase protein (L) (Lamb and Parks, 2007). The N protein binds to the entire length of viral genomic and antigenomic RNAs, forming highly stable nucleocapsids that function in transcription, replication and virus assembly. The P protein acts as a polymerase co-factor and is required for RNA synthesis (Curran *et al.*, 1995). The M protein is the most abundant protein in the virion and it lines the inner face of the viral envelope (Lamb and Choppin, 1977) and plays a major role in viral morphogenesis (Peeples, 1991). The HN glycoprotein binds to sialic acid-containing cell surface receptors and facilitates virus penetration. The F glycoprotein mediates virus penetration by inducing fusion between the viral envelope and host cell plasma membrane. The large polymerase (L) protein is the viral RNA-dependent RNA polymerase. Most members of the subfamily *Paramyxovirinae* encode two additional proteins, V and W (or I, in case of *Rubulavirus*) from the P gene by a mechanism called RNA editing. RNA editing involves the co-transcriptional insertion of one or more G residues into nascent P mRNA at a conserved RNA editing motif located midway along the P gene. The insertion of one or two G residues results in translational frame shifts that can access alternative open reading frames. Edited mRNAs encode chimeric V or W proteins in which the N-terminal domain is identical to that of the P protein and the C-terminal domain is encoded by an alternative frame. The V protein, which contains a cysteine-rich C-terminal, zinc finger-like domain, has roles in regulating RNA synthesis and in blocking the host type I interferon response (Kolakofsky *et al.*, 1998).

All avian Paramyxoviruses that have been isolated from avian species are placed in the genus *Avulavirus* except for avian metapneumovirus, which is classified in genus

Metapneumovirus (Lamb and Parks, 2007). The viruses in the genus *Avulavirus* are classified into nine serotypes (APMV-1 to -9) based on hemagglutination inhibition (HI) and neuraminidase inhibition (NI) assays (Alexander, 2003). All strains of Newcastle disease virus (NDV) comprise APMV-1. Since NDV is an important poultry pathogen, APMV-1 is the most extensively characterized serotype among APMVs. Recently, the complete genome sequences of several other APMV serotypes 2, 3, 4, 6, 8 and 9 have been reported (Subbiah *et al.*, 2008; Kumar *et al.*, 2008; Nayak *et al.*, 2008; Jeon *et al.*, 2008; Chang *et al.*, 2001; Paldurai *et al.*, 2009; Samuel *et al.*, 2009). However, the complete genome sequence of APMV-5 and -7 are still not available.

Very little is known about the pathogenicity of APMV serotypes 2-9. APMV-2 was shown to affect hatchability and poult yield in turkeys (Bankowski *et al.*, 1981) and caused mild disease in chicken (Alexander, 1980). APMV-3 and -6 are associated with drop in egg production in turkeys (Alexander *et al.*, 1983; Alexander, 1997). APMV-4 has been reported to cause an increase in white-shelled eggs but did not affect the egg production in laying hens (Alexander, 2003). APMV-5 infection in budgerigars caused dyspnoea, diarrhea and death (Nerome *et al.*, 1978). The pathogenicity of APMV-8 and 9 are not known. APMV-7 was first isolated from a hunter killed dove in 1975 in Tennessee, USA and was designed as a new serotype based on HI and NI assays (Alexander *et al.*, 1981). Subsequently, APMV-7 was isolated from a natural outbreak of respiratory tract disease in commercial turkey breeder flocks in Ohio in 1997 (Saif *et al.*, 1997). The experimental infection of APMV-7 in turkeys caused respiratory disease and affected the egg production (Saif *et al.*, 1997). To understand the molecular characteristics of APMV-7, we have determined the complete genome sequence of APMV-7 prototype strain Tennessee and compared its relatedness to other paramyxoviruses.

2. Materials and Methods

2.1. Virus and cells

APMV-7/dove/Tennessee/4/75 was received from the National Veterinary Services Laboratory, Ames, Iowa, USA. The virus was propagated in 9-day-old embryonated specific pathogen free (SPF) chicken eggs. Infected allantoic fluid was harvested 3 days post-inoculation. The titer of the virus was determined by hemagglutination (HA) assay using 1% chicken red blood cells (RBCs). Replication of the virus was evaluated in 11 established cell lines, namely, chicken embryo fibroblast (DF-1), duck embryo fibroblast (DEF), quail fibrosarcoma (QT-35), African green monkey kidney (Vero), baby hamster kidney (BHK-21), Madin Darby Bovine Kidney (MDBK), Madin Darby Canine Kidney (MDCK), Bovine Turbinate (BT), rabbit kidney (RK-13), pig kidney (PK-15) and human cervical carcinoma (HEp-2) cells and one primary cell line, turkey embryo fibroblast (TEF). The cells were grown in Dulbecco's minimum essential medium (DMEM) containing 10% fetal bovine serum (FBS) and incubated at 37°C under 5% CO₂. The cells were infected with 10⁻³ dilution of 2⁸ HA units of egg-grown APMV-7, with or without supplementation of 10% fresh normal allantoic fluid as an exogenous source of protease for the cleavage of viral fusion (F) protein. The cells were observed daily for 7 days for cytopathic effect (CPE) and the HA titer was determined from cell culture supernatant. A total of three serial passages of the virus were made in each cell line. The ability of the virus to produce plaques was examined in the different cell lines using 0.8% methyl cellulose overlay and staining with crystal violet.

2.2. Viral RNA isolation and sequence analysis

APMV-7/dove/Tennessee/4/75 RNA was isolated from the allantoic fluid of virus-infected chicken eggs using RNeasy kit (QIAGEN, USA) according to the manufacturer's

instructions. As a first step towards sequencing the viral genome, a rapid genome sequencing method based on cDNA representational difference analysis (cDNA RDA) was used, as described previously (Sakai *et al.*, 2007 and Mizutani *et al.*, 2007). This strategy involved two steps of amplification. In the first step, a whole transcriptome amplification kit (WTA1, Sigma-Aldrich, USA) was used to synthesize double-stranded cDNA. Briefly, RNA was reverse transcribed as per the manufacturer's instructions using the library synthesis buffer and enzyme provided in the kit. The cDNA was then added to the WTA amplification mix and heated at 94°C for 60 sec, followed by 40 cycles of amplification (denaturation at 94°C for 30 sec, annealing at 65°C for 90 sec and primer extension at 72°C for 90 sec). The first cDNA library thus obtained was digested with HaeIII at 37°C for 1 h. The cDNA was then ligated with EcoRI-NotI-BamHI adaptor using a Rapid Ligation Kit (Roche, USA). This library then served as template for a second round of amplification using primers designed to anneal to the EcoRI-NotI-BamHI adaptor and the partial HaeIII site and which contained four additional nucleotides at the 3' end that varied from primer to primer. The amplification was done in a matrix of reactions with different primer pairs, as described by Mizutani *et al.* (2007), with one partner selected from the published series H1-1 to -13 and the other from the series H9-1, 3, 5, 6, 7, 9, 10, and 13. In those reactions that yielded PCR products, the products were resolved in agarose gels and purified for direct sequencing. The viral sequences thus obtained were located in regions of N, M, HN and L genes as shown in Fig. 1.

To determine the remaining of the viral genome sequence, the following viral specific primers were used. The regions between N and M, M and HN, and those within L gene were amplified using 695 (+), 5'-AGATACTGCTGCGGAGAG (numbered according to position in the viral genome) and 3984 (-), 5'-TCCTCTAATGATCTCGGTTTC; 3865 (+), 5'-CTGGAA GTTGATGTGAAGC and 8004 (-), 5'-TGCTGGAGTACCCTGTGCGAATAA; 9311 (+) 5'-TT TTTGATGTAATTGCTTCTC and 13996 (-), 5'-GTGTTGGTAATGGGCCGATGTTTC and 13827 (+), 5'-GCATGGTATAAAGCAAGTAGTT and 14832 (-), 5'-TTATTTTCCCTCCAAGTGCCAGAT, respectively. Briefly, the first-strand cDNA was synthesized from viral RNA by Superscript-II kit (Invitrogen, USA) with random hexamers according to manufacturer's instructions. PCR was performed using viral specific primers and *Taq* polymerase (Invitrogen). The PCR fragments were cloned in TOPO TA cloning kit (Invitrogen) and the clones were sequenced using vector primers. In addition, selected PCR products were purified by agarose gel electrophoresis and sequenced directly. The DNA sequencing was carried out using BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems, USA) in ABI 3130xl genetic analyzer (Applied Biosystems). Every nucleotide in the genome was sequenced at least three times and once directly from RT-PCR product without cloning, thus ensuring a consensus sequence. All primers were synthesized from Integrated DNA Technologies, Inc.

2.3. Determination of the sequences of genome termini

The end sequences of APMV-7 viral genome were performed by 3' and 5' rapid amplification of cDNA ends (RACE) that has been described previously (Subbiah *et al.*, 2008). Briefly, for the 3' end, viral genomic RNA was ligated with an adaptor 1 (5'-GAAGAGAAGGTGGAAATGGCGTTTTGG-3', 5'-phosphorylated; 3'-blocked), (Li *et al.*, 2005). The cDNAs were synthesized using adaptor-2 which is complementary to adaptor-1 (5'-CCAAAACGCCA TTTCCACCTTCTCTTC-3'). The PCR was performed with adaptor-2 and a viral N-specific reverse primer 7N rev (5'-TGTGAATTCAGGACGCAGACT-3'). The sequence of the trailer region was determined using 5'RACE technique (Invitrogen), a L-specific forward primer 7L6000 fwd (5'-AC TAATAAGACTTGGAGGGAAACC-3') was used for cDNA synthesis and the cDNA was

subsequently poly dATP tailed using T4 terminal deoxynucleotidyl transferase (TdT) according to the manufacturer's protocol (Invitrogen). The PCR was performed using 7L6000fwd primer and oligo (dT) reverse primer (5'-ACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTT-3') using poly adenylated cDNA as template. The PCR products were cloned into the TOPO TA cloning vector and sequenced, and were also directly sequenced.

2.4. Sequence and phylogenetic tree analysis

Sequence compilation and prediction of ORFs were carried out using the SeqMan and EditSeq programs in the Lasergene 8 (DNASTAR) software package (www.dnastar.com). The search for matching protein sequences in GenBank was done using the blastp program of the same package. Phylogenetic analysis was carried out using T-Coffee, (Tree-based consistency objective function for alignment evaluation), a multiple sequence alignment program. The phylogenetic trees were drawn using the same program and applying the "average distance using Blosum62" method (Notredame *et al.*, 2000).

2.5. Database accession numbers

The complete genome sequence of APMV-7 strain Tennessee has been deposited in GenBank under accession no. **FJ231524**. Accession numbers for other paramyxovirus sequences used in this study are given below. Avulaviruses: NDV, AF077761; APMV-2, EU338414; APMV-3, EU403085; APMV-4, EU877976 and FJ177514; APMV-6, NC_003043; APMV-8, FJ215863, APMV-9, EU910942. Rubulaviruses: Human PIV-2 NC_003443; SV5 (also known as parainfluenza 5) NC_006430; MuV NC_002200; SV41 NC006428. Respiroviruses: HPIV-1 NC_003461; HPIV-3 NC_001796; SeV NC_001552. Morbilliviruses: CDV, NC_001921; MeV, AF266288; PDV, NC_006383; RPV, NC_006296. Henipaviruses: NiV, NC_002728; HeV, NC_001906. Pneumoviruses: HRSV, NC_001781; Bovine RSV, NC_001989. Metapneumo-viruses: AMPV, NC_007652; HMPV, NC_004148.

3. Results

3.1. Growth characteristics of APMV-7

Avian paramyxovirus type 7 strain Tennessee yielded a titer of 2^{11} HA units in the allantoic cavity of 9-day-old embryonated SPF chicken eggs, three days post-inoculation. However, the virus did not cause death of the chicken embryo until 144 h, indicating that it is avirulent in chickens. Eleven different cell lines (BHK-21, HEp-2, Vero, MDCK, DF-1, DEF, MDBK, PK-15, RK-13, QT-35 and BT) and one primary cell line (TEF) were evaluated to determine the cell type(s) that can support the replication of APMV-7 and whether or not exogenous protease is required for virus replication. APMV-7 replicated in BHK-21, HEp-2, Vero and MDCK cells yielding a titer of 2^4 , 2^4 , 2^3 and 2^1 HA units, respectively. The supplementation of 10 % normal fresh allantoic fluid, as an exogenous source of protease, neither augmented the CPE nor the HA activity of the infected cell culture supernatants, indicating a lack of requirement of exogenous protease for the cleavage of the APMV-7 F protein. The CPE observed were rounding and detachment of cells. Syncytia formation was also observed in BHK-21, Vero and MDCK cells. The virus failed to produce visible plaques under 0.8% methylcellulose overlay in the presence or absence of exogenous protease. There was neither CPE nor HA activity was produced in DF-1, TEF, DEF, MDBK, PK-15, RK-13, QT-35 and BT cells, indicating that these cells did not support the replication of APMV-7.

3.2. Determination of the complete genome sequence of APMV-7

The sequence of the genome of APMV-7 strain Tennessee consists of 15,480 nt (GenBank accession no. **FJ231524**). This was confirmed in its entirety from uncloned RT-PCR products and thus represents a consensus sequence. The genome length is a multiple of six, conforming to the “rule of six” that is characteristic of all other members of the subfamily *Paramyxovirinae* (Kolakofsky *et al.*, 1998). The genome organization of APMV-7 is 3'-leader-N-P/V/W-M-F-HN-L-trailer-5', which is the same as that of other members of genus *Avulavirus* except for APMV-6, which has an additional small hydrophobic (SH) protein gene between F and HN genes (Chang *et al.*, 2001). The molecular features of APMV-7 are shown in Table 1. The genome size (15,480 nt) and its coding percentage (88.4 %) are similar to the average genome size (~15,500 nt) and coding percentage (92%) of other members of subfamily *Paramyxovirinae* (Lamb and Parks, 2007; Miller *et al.*, 2003). The complete genome of APMV-7 has a nucleotide sequence identity of 45.8 %, 52.7 %, 41.2 %, 42.5 %, 47.2 %, 45.9 % and 52.5 % when compared with the complete genome sequences of APMV-1, -2, -3, -4, -6, -8 and -9, respectively. This shows the level of genetic diversity among the members of genus *Avulavirus*. The GC content of the genome of APMV-7 is 39%, which is significantly lower when compared to the GC contents of the genomes of APMV-1 (46%), APMV-2 (47%), APMV-3 (46%), APMV-4 (47%), APMV-6 (46%), APMV-8 (43.3%) and APMV-9 (44.6%).

The 3'-leader sequence of APMV-7 consists of 55 nt, a length that is conserved among almost all the members of the subfamily *Paramyxovirinae*. The genomic terminus of APMV-7 leader has a high degree of nucleotide sequence identity with those of other APMVs (Fig. 2a); for example, the sequence of the first 12 nt of the leader region of APMV-7 is identical to that of APMV-2, -6 and -8. The first 8 nt of the leader sequence (3'-UGGUUUGU-5') are the same for all the members of genus *Avulavirus* sequenced to date except for 1 nt change in the third position of APMV-3, where residue G is replaced with residue A and 2 nt changes at the third and seventh positions of APMV-4, where residue G is replaced by C at the third position and residue G is replaced by residue U at the seventh position. One notable feature in the leader sequence of APMV-7 is the presence of a sequence, ³⁵AAUUAUUUUU⁴⁵ (Fig. 2a), that is identical to the GE signals of the F and HN genes (Table 2). However, the significance of this apparent GE signal sequence in the leader region at present is not known.

The length of the 5'-trailer region of APMV-7 is 127 nt, a length that is variable among the members of the subfamily *Paramyxovirinae*. The sequence of the last 12 nt of the APMV-7 trailer sequence are identical to that of APMV-6, and 11 nt out of the last 12 nt are identical to those of APMV-1 and -9 (Fig. 2b). The first 12 nt of the leader and trailer sequences of APMV-7 are 100% complementary to each other, which is suggestive of conserved promoter elements at the 3' and 5' termini of the genome and antigenome, respectively (Fig. 2c).

3.3. Sequences of the GS and GE motifs and the IGS

The GS sequences of the various genes of APMV-7 strain Tennessee were identical in positions 1-5 and 7-8 except positions 6, 9 and 10 with a consensus sequence of 3'-CUCCNCUNN-5' (Table 2). The GE sequences were identical for positions 1-3 and 6-10 but position 9 is not conserved in the M GE, with a consensus of 3'-AAUNNUUNU₁₋₃-5' (Table 2). Interestingly, the GS and GE signals of avulaviruses sequenced to date showed a high degree of similarity, indicating that this aspect of polymerase recognition of cis-acting sequences is conserved within the genus (Table 3).

The lengths of the IGS of APMV-7 ranged from 11 to 70 nt. The IGS between the N/P, P/M, M/F, F/HN and HN/L genes are 24, 13, 70, 11 and 15 nt, respectively. The IGS of all the avulaviruses sequenced to date vary greatly in length, which shows their evolutionary diversity and implies a flexibility for this genome feature. The range of IGS lengths observed among the known APMVs for the various gene junctions were: N/P, 1-55 nt; P/M, 1-63 nt; M/F, 1-70 nt; F/HN, 8-37 nt, and HN/L, 0-63 nt; in addition, the novel F/SH and SH/HN IGS of APMV-6 are 49 nt and 29 nt, respectively (Table 4).

The hexamer phasing positions of the GS signals of APMV-7 are 2, 2, 4, 1, 2 and 4 (Table 1), which are different from those APMV-1 (2,4,3,3,2 and 5), -2 (2,2,2,3,3 and 3), -3 (2,5,5,2,2 and 1), -4 (2,2,2,6,2 and 2), -6 (2,2,2,2,2,4 and 4), -8 (2,2,2,2,4 and 1) -9 (2,3,4,4,4 and 3). Thus, the hexamer phasing positions of the GS signals in the APMVs sequenced to date are not conserved other than the hexamer position of the first gene being 2.

3.4. The Nucleoprotein (N) gene

The N gene of APMV-7 is 1,644 nt long and encodes a N protein of 463 amino acids (aa) with a predicted molecular weight (MW) of 50,711 and isoelectric point (pI) of 5.29. The N protein contain a highly conserved motif, 324-FAPANYTLLYSYAMG-338 (F-X4-Y-X3-Φ-S-Φ-A-M-G, where, X is any amino acid and Φ is any aromatic amino acid) that has been identified in other members of the subfamily *Paramyxovirinae* and is thought to be responsible for N-N self assembly during genomic RNA binding (Yu *et al.*, 1998; Morgan, 1991). The amino acids at position 325, 326 and 333 in the motif are alanine (A), proline (P) and tyrosine (Y), respectively, which are conserved in all known avulaviruses (Table 5). The N protein of APMV-7 has an amino acid (aa) sequence identity of 40.6%, 55.3%, 35.2%, 37.0%, 53.5%, 54.4% and 41.3% with the N proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). The APMV-7 N protein has an aa sequence identity of 35.6%, 17.9%, 25.9%, 27.2%, and 8.6%, with the N proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

3.5. The Phosphoprotein (P) gene

The P gene of APMV-7 is 1,453 nt long and encodes a P protein of 394 aa with a MW of 42,636 and pI of 5.139. The predicted P protein contains 14 potential sites for phosphorylation, which include 9 predicted sites of serine (68S, 73S, 128S, 131S, 170S, 291S, 292S, 310S, and 342S), 4 predicted sites of threonine (44T, 88T, 160T and 380T) and 1 predicted site of tyrosine (311Y), as predicted by the NetPhos 2.0 program of ExPASy proteomics server. The P protein of APMV-7 has an aa sequence identity of 20.8%, 25.6%, 23.4%, 21.1%, 23.3%, 27.2% and 21.6% with the P proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). The P protein of APMV-7 has an aa sequence identity of 19.8%, 10.2%, 11.2%, 11.4%, and 5.8%, with the P proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

The P gene contains a putative editing site 3'-UUUUUCCC-5' (genome sense) at nucleotide positions 430 to 437 in the P gene (that corresponds to nucleotide position 2153 to 2160 in the genome). The sequence of the APMV-7 editing is identical to those of APMV-1, -2, -6, -8 and -9. However, it differed from those of APMV-3 and -4, which contain the nucleotide sequence 3'-AA UUUCCC-5' (genome sense) (Fig. 3a). The addition of a single G residue at the APMV-7 editing site would produce a V mRNA encoding a V protein of 250 aa, with a predicted MW of 26,998. The N-terminal 124 aa of the V protein are identical to the corresponding domain in the P protein and are followed by a V-specific C-terminal domain of 126 aa. This V-specific domain contains the seven conserved invariantly spaced cysteine

residues, as reported for other members of subfamily *Paramyxovirinae* (Fig. 3b). The addition of two G residues at the editing site would produce a W mRNA encoding a W protein of 125 aa, with a predicted MW of 13,421. The putative W protein contains an N-terminal domain of 124 aa that is identical to that of P followed by a single additional residue that is unique to W.

3.6. The Matrix (M) gene

The M gene of APMV-7 is 1,211 nt long and encodes a M protein of 363 aa with a predicted MW of 39,809. The pI value of the M protein is 9.323, which underlines the basic property of this protein that is necessary for ionic interactions with the acidic N protein (Lamb and Parks, 2007). The M protein contains a unique motif **KKTNASGKPRSLED MRKKVR**, corresponding to the aa positions 242 to 261, which consists of bipartite clustering of basic residues. This motif is thought to serve as a nuclear localization signal (NLS) for the protein (Coleman and Peeples, 1993; Peeples *et al.*, 1992). A protein-protein interaction domain called the “late domain”, which contains the motif FPIV and is involved in assembly and budding of viruses, was first identified in the paramyxovirus SV5 (Schmitt *et al.*, 2005). The M protein of APMV-7 contains a potential late domain motif FPII at aa positions 22-25. Similar motifs are present in the M proteins of other APMV, namely: 22-YPLI-25, APMV-2; 23-FPLI-26, APMV-3; 22-FPLI-25, APMV-4; 22-FPIV-25, APMV-6; 22-FPLV-25, APMV-8; and 23-FPIV-26, APMV-9 (Table 5). In the late domain motif, the amino acid proline (P) in the second position, is conserved in all the members of genera *Avulavirus* and *Rubulavirus*, with the consensus pattern α -P- Φ - Φ (α represents an aliphatic amino acid and Φ represents an aromatic amino acid). However, this motif is not observed in the M proteins of members of genera *Respirovirus*, *Morbillivirus* and *Henipavirus* of the subfamily *Paramyxovirinae*. The M protein of APMV-7 has an aa sequence identity of 33.3%, 42%, 29%, 28.9%, 44.5%, 43.5% and 33.3% with the cognate M proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). The APMV-7 M protein has an aa sequence identity of 30.6%, 19.3%, 20.9%, 21.2%, and 6.1%, with the M proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

3.7. The Fusion (F) gene

The F gene of APMV-7 is 1,620 nt long and encodes a F protein of 539 aa with a predicted MW of 58,107 and pI of 4.981. The F protein is predicted to be a type I transmembrane protein similar to the F proteins of other members of the family *Paramyxoviridae*. It contains a signal sequence of 25 aa (positions 1-25) at the N-terminal end and a predicted transmembrane domain between aa sequence positions 488 to 512. The putative cleavage site of the APMV-7 unprocessed F protein (F0) is T-L-P-S-S-R↓F (aa positions 101-107) giving rise to F1 and F2 peptides, wherein the residue F is present at the N-terminus of the F1 fusion peptide. The cleavage site sequence of APMV-7 does not conform to the favored sequence motif for cleavage by the intracellular protease furin, R-X-K/R-R↓F (Hosaka *et al.*, 1991), and instead has a single basic residue (underlined). A comparison of F protein cleavage sites of other members of genera *Avulavirus* is given in Table 7. The F1 polypeptide of APMV-7 contains heptad repeats at aa positions 120-162 and 451-483, as predicted by the LearnCoil-VMF program, which correspond to the characteristic HRA and HRB heptad repeats that are found in paramyxovirus F proteins and are thought to be important for virus-cell membrane fusion (Morrison *et al.*, 2003). The APMV-7 F protein contains six potential N-linked glycosylation sites located at positions 67 and 79 of the F2 subunit and at aa positions 186, 461, 479 and 524 of the F1 subunit, as predicted by the NetNGlyc 1.0 program of the ExPASy proteomics server. The F protein of APMV-7 has an aa sequence identity of 39.3%, 38.4%, 27.4%, 29.5%, 36.6%, 39.0% and 35.8% with the cognate F proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). The APMV-7

F protein has an aa sequence identity of 30.1%, 23.4%, 26.7%, 24.3%, and 12.4%, with the F proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

3.8. The Hemagglutinin-Neuraminidase (HN) gene

The HN gene of APMV-7 is 2,021 nt long and encodes a HN protein of 569 aa with a predicted MW of 62,446 and pI of 6.059. The HN protein is predicted to be a type II integral membrane protein and has a predicted hydrophobic transmembrane domain located between aa residues 14-36 at the N-terminus. There are nine potential N-linked glycosylation sites predicted at aa positions 48, 109, 135, 268, 333, 367, 482, 511 and 562. The HN protein contains the unique motif N-R-K-S-C-S at aa positions 224-229, which is thought to be involved in sialic acid binding at the cell surface (Varghese *et al.*, 1983; Mirza *et al.*, 1994). This motif is conserved in the HN proteins among avulaviruses sequenced to date with the exception of APMV-6, whose HN protein has a N residue at aa position 245 (Table 5). Comparing the six conserved residues of the proposed neuraminidase active site found in the HN protein of NDV (Takimoto, 2000; Langedijk *et al.*, 1997; Panda *et al.*, 2004), the corresponding residues in the APMV-7 HN are: R(164), E(391), R(406), R(496), Y(524) and E(545). The HN protein of APMV-7 has an aa sequence identity of 35.3%, 41.4%, 32.4%, 34.8%, 40.6%, 41.5% and 34.6% with the HN proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). APMV-7 HN protein has an aa sequence identity of 32.5%, 21.2%, 11.1%, 18.1%, and 7.9%, with the HN proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

3.9. The Large Polymerase (L) gene

The L gene of APMV-7 is 6,920 nt long and encodes a L protein of 2,227 aa with a predicted MW of 251,716 and pI of 7.384. The sequence alignment of the APMV-7 L protein showed six linear conserved domains, as found in other members of the family *Paramyxoviridae* (Poch *et al.*, 1989) (data not shown). The previously conserved GDNQ sequence motif in domain III, which is widely conserved and is thought to be involved in L protein transcriptional activity (Schnell and Conzelmann, 1995; Malur *et al.*, 2002), also was present in the L protein of APMV-7 (aa sequence positions 767-771, Table 5). The L protein of APMV-7 has an aa sequence identity of 38.2%, 44.2%, 36.2%, 34.9%, 43.8%, 45.2% and 38.4% with the L proteins of APMV-1, -2, -3, -4, -6, -8 and -9, respectively (Table 6). The APMV-7 L protein has aa sequence identity of 35.9%, 26.8%, 27.1%, 26.9%, and 13.2%, with the L proteins of *Rubulavirus* (MuV), *Respirovirus* (SeV), *Morbillivirus* (MeV), *Henipavirus* (NiV) and *Metapneumovirus* (AMPV), respectively.

3.10. Phylogenetic tree analysis

Phylogenetic trees were constructed from a sequence alignment of N, P, M, F, HN and L proteins of APMV-7 strain Tennessee with the cognate proteins of representative viruses of all the five genera of the subfamily *Paramyxovirinae* (Fig 4). The phylogenetic trees clearly indicated that the APMV-7 was most closely related to the known APMVs, consistent with its inclusion in the genus *Avulavirus*. Furthermore, the phylogenetic analysis of these 6 proteins revealed a closer genetic relatedness of APMV-7 to APMV-2, APMV-6 and APMV-8 than to the other APMV serotypes within the genus *Avulavirus*.

4. Discussion

The APMVs are frequently isolated from a wide variety of avian species and are represented by nine different serological types. The molecular biology and pathogenicity of these viruses are not well known, except for APMV-1. Recently, the complete genome sequences of

APMV-2,-3,-4,-6,-8 and -9 have been determined (Subbiah *et al.*, 2008; Kumar *et al.*, 2008; Jeon *et al.*, 2008; Nayak *et al.*, 2008; Chang *et al.*, 2001; Paldurai *et al.*, 2009 and Samuel *et al.*, 2009). Here, we have characterized APMV-7 prototype strain dove/Tennessee/4/75, and determined its complete genome sequence. APMV-7 did not cause death of chicken embryos until 144 h of inoculation with 2^8 HA units of virus, indicating that it is avirulent for chickens. Experimental infection of turkeys by APMV-7 has also shown the avirulent nature of this virus (Saif *et al.*, 1997). However, the pathogenicity of APMV-7 needs to be evaluated directly in chickens and other avian species.

The genome of APMV-7 is 15,480 nt in length, which is longer than those of APMV-1 (15,186 nt), APMV-2 (14,904 nt), APMV-4 (15,054 nt), APMV-8 (15,342 nt) and APMV-9 (15,438 nt), but shorter than those of APMV-3 (16,272 nt) and APMV-6 (16,236 nt). The length of the APMV-7 genome also is typical of most members of family *Paramyxoviridae* (approximately 15,500 nt). The genome length of APMV-7 also follows the “rule of six” that is characteristic of subfamily *Paramyxovirinae* (Calain and Roux, 1993; Kolakofsky *et al.*, 1998). The genome contains six genes (N, P, M, F, HN and L) that are also found in all members of genus *Avulavirus* sequenced to date, with the exception that APMV-6 has an additional gene potentially encoding the small hydrophobic SH protein (Chang *et al.*, 2001).

The leader region of APMV-7 is 55 nt long, a length that is generally conserved among members of the subfamily *Paramyxovirinae* (Lamb and Parks, 2007). In general, the first 12-13 nt of the leader region are most closely related among members of a genus of the subfamily *Paramyxovirinae*. Comparison of the sequences of leader regions of APMV serotypes sequenced to date showed that the first 12 nt are identical among APMV-7, -2, -6 and -8, indicating a close evolutionary relationship. But, comparison of the leader sequence of APMV-7 with those of APMV-1, -3, -4 and -9 showed 2, 4, 6 and 1 nt changes in the first 12 nt, respectively, indicating also a diversity among members of genus *Avulavirus*. The APMV-7 leader sequence is unique in having a stretch of sequence (35 AAUUAU₆) that is identical to the GE signal of its F and HN genes. It will be very interesting to study whether this sequence has any functional significance in viral replication and transcription. The trailer region of APMV-7 is 127 nt in length, which is longer than the typical length (40-60 nt) for most members of the family *Paramyxoviridae* (Shioda *et al.*, 1986), but is shorter than that of APMV-3 (707 nt) (Kumar *et al.*, 2008). The terminal sequences of the 3' leader and 5' trailer regions showed 100% complementarity for the first 12 nt, suggestive of conserved elements in the 3' promoter region of the genome and antigenome.

Comparison of the GS and GE sequences of APMV-7 with those of the other members of genus *Avulavirus* showed that the GE sequences are mostly conserved among the different serotypes, with the first 3 positions being AAU and the sequence ending with a run of 5-7 U residues except for APMV-7 M GE (negative-sense, Table 3). There was somewhat more variability among the GS sequences of these serotypes with the C residues in positions 3 and 5 were completely conserved and the U in position 4 was also conserved except APMV-3 (Table 3). This suggests that the viral RNA polymerase of each APMV serotype has a specificity to recognize its GS signal for transcription. The UTRs of paramyxoviruses show a broad diversity in their length and sequence, both within a virus and among viruses within the family. The lengths of the 3'UTRs (relative to the mRNA) are longer than those of the 5'UTRs in all genes of APMV-7; which is similar to other APMVs, except for UTRs of APMV-2 F (3'UTR, 42 nt and 5'UTR, 52 nt) and APMV-4 HN (3'UTR, 42 nt and 5'UTR, 69 nt) genes. Although the role of UTR length in APMV-7 is not known, UTRs have been shown to modulate transcription and translation in NDV (Yan *et al.*, 2009) and canine distemper virus (Anderson *et al.*, 2008; Takeda *et al.*, 2005). The IGS of APMV-7 vary from 11 to 70 nt (Table 1). The variable, nonconserved nature of the IGS of APMV-7 is similar to

that of the members of *Rubulavirus* and *Avulavirus* and differ from the typical conserved trinucleotide GAA of *Respirovirus*, *Morbillivirus*, and *Henipavirus* (Nylund *et al.*, 2008).

The APMV-7 P gene contained a putative editing site (UUUUUCCC, genome sense), which is similar to those observed in APMV-1, -2, -6, -8 and -9 but differed with those of APMV-3 and -4 wherein the first two positions were A instead of U. The APMV-7 V protein contained 250 aa with a C-terminus 126 aa which contained the seven conserved invariantly spaced cysteine residues, as reported for other members of the subfamily *Paramyxovirinae*. The existence of W protein in APMV-7 remains to be confirmed.

The paramyxovirus F protein is synthesized as an inactive precursor (F0) that is cleaved by host protease into the biologically-active form that consists of disulfide-linked F1-F2 subunits (Lamb and Parks, 2007). The F protein cleavage site is a major determinant of NDV pathogenesis in chickens. Virulent NDV strains have a polybasic cleavage site (R-X-K/R-R↓F), which is recognized by furin-like intracellular proteases, and the cleavage site is followed by an F residue at the beginning of the F1 subunit. The avirulent NDV strains have one or a few basic residues at the cleavage site and a leucine (L) residue at the first position of F1 subunit, and are cleaved by secretory proteases found only in the respiratory or enteric tracts (or, in cell culture, by exogenous protease supplementation). Interestingly, the ability of NDV F to be cleaved intracellularly was blocked by the presence of leucine as the first amino acid of the F1 subunit, indicating that the identity of this residue contributes to the cleavage phenotype (Morrison *et al.*, 1993). In contrast to the NDV paradigm, analysis of recently-available sequences of putative cleavage sites of other APMV serotypes showed that the cleavage site sequences of some serotypes are not necessarily predictive of the protease activation phenotype. For example, the F protein cleavage site of APMV-8 (PQTR↓L) contains a single basic residue (underlined) and leucine as the N-terminal amino acid of F1, and requires supplementation of acetylated trypsin (1 µg/ml) for growth in cell lines but not in the primary chicken embryo kidney (CEK) cells (Paldurai *et al.*, 2009). The F protein cleavage site of APMV-9 (REGR↓I) contains two basic residues (like the avirulent strains of NDV) and a phenylalanine at the N-terminus of F1 (like the virulent strains of NDV), but requires 10% allantoic fluid supplementation for growth in cell culture (Samuel *et al.*, 2009). In contrast, the F protein cleavages site of APMV-2 (PASR↓F) and APMV-4 (IQPR↓F) also contain a single basic residue (R), but do not require exogenous protease for growth in cell culture (Subbiah *et al.*, 2008; Nayak *et al.*, 2008). Similar to the examples of APMV-2 and -4, the putative APMV-7 F protein cleavage site (PSSR↓F) contains a single basic residue (R) and phenylalanine (F) residue at the N-terminus of F1 subunit and also grew in cell culture without the supplementation of exogenous protease (Table 7). One possibility is that the cleavage might occur in endosomal compartment, as is the case in Nipah virus (Moll *et al.*, 2004; Diederich *et al.*, 2005).

The sequence alignment and phylogenetic analysis showed that APMV-7 clustered with the other APMVs within genus *Avulavirus*, justifying its classification and consistent with the International Committee on Taxonomy of Viruses (ICTV) statement on classification that the amino acid sequence relationships are the main criteria for grouping viruses into genera within the family *Paramyxoviridae* (Lamb *et al.*, 2000). Phylogenetically, APMV-7 is more closely related to APMV-2, APMV-6 and APMV-8 than to the other known APMVs. This is not consistent with the results of cross HI and NI tests, which suggested the grouping of nine APMV serotypes into two broad subgroups, the first subgroup consisting of APMV-2 and -6 while the second subgroup consisting of APMV-1, -3, -4, -7, -8 and -9 (Lipkind and Shihmanter, 1986). Since the phylogenetic analysis is based on the sequences of all of the viral genes, whereas the HI/NI analysis is based on antibody interaction with the functional domain of a single protein (HN), the phylogenetic analysis should provide a more accurate picture of evolutionary relationships. It will be important to group all of the APMVs

according to phylogenetic analysis when the complete genome sequence of all the nine serotypes becomes available. The availability of the complete genome sequence of APMV-7 will be helpful to study the molecular biology, genetics and pathogenicity of this little known APMV.

Acknowledgments

We thank Sachin Kumar for helpful suggestions in the present complete genome sequencing experiment. We thank Dianel Rockemann, Flavia Dias and all our laboratory members for their excellent technical assistance and help. We also thank Ireen Dryburgh-Barry for proofreading the manuscript. “This research was supported by NIAID contract no.N01A060009 (85% support) and NIAID, NIH Intramural Research Program (15% support). The views expressed herein do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.”

References

- Alexander DJ. Avian Paramyxoviruses. *The Veterinary Bulletin*. 1980; 50:737–751.
- Alexander DJ, Hinshaw VS, Collins MS. Characterization of viruses from doves representing a new serotype of avian paramyxoviruses. *Arch Virol*. 1981; 68:265–269. [PubMed: 6168246]
- Alexander DJ, Pattison M, Macpherson. Avian paramyxoviruses of PMV-3 serotype in British turkeys. *Avian Pathology*. 1983; 12:469–482. [PubMed: 18766806]
- Alexander, DJ. Newcastle disease and other avian *Paramyxoviridae* infections. In: Calnek, BW., editor. *Diseases of Poultry*. Iowa State University Press; Ames: 1997. p. 541-569.
- Alexander, DJ. Avian Paramyxoviruses 2-9. In: Saif, YM., editor. *Diseases of poultry*. 11th edn. Iowa State University Press; Ames: 2003. p. 88-92.
- Anderson DE, von Messling V. The Region between the Canine Distemper Virus M and F Genes Modulates Virulence by Controlling Fusion Protein Expression. *J Virol*. 2008; 82:10510–10518. [PubMed: 18753197]
- Bankowski RA, Almquist J, Dombrucki J. Effect of paramyxovirus Yucaipa on fertility, hatchability, and poult yield of turkeys. *Avian Diseases*. 1981; 25:517–520. [PubMed: 7259687]
- Calain P, Roux L. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J. Virol*. 1993; 67:4822–4830. [PubMed: 8392616]
- Chang PC, Hsieh ML, Shien JH, Graham DA, Lee MS, Shieh HK. Complete nucleotide sequence of avian paramyxovirus type 6 isolated from ducks. *J Gen Virol*. 2001; 82:2157–2168. [PubMed: 11514725]
- Coleman NA, Peeples ME. The matrix protein of Newcastle disease virus localizes to the nucleus via a bipartite nuclear localization signal. *Virology*. 1993; 195:596–607. [PubMed: 8337834]
- Curran J, Marq JB, B. J, Kolakofsky D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol*. 1995; 69:849–855. [PubMed: 7815552]
- Diederich S, Moll M, Klenk HD, Maisner K. The nipah virus fusion protein is cleaved within the endosomal compartment. *J Biol Chem*. 2005; 280:29899–29903. [PubMed: 15961384]
- Hosaka M, Nagahama M, Kim WS, Watanabe T, Hatsuzawa K, Ikemizu J, Murakami K, Nakayama K. Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. *J Biol Chem*. 1991; 266:12127–12130. [PubMed: 1905715]
- Jeon WJ, Lee EK, Kwon JH, Choi KS. Full-length genome sequence of avian paramyxovirus type 4 isolated from a mallard duck. *Virus Genes*. 2008; 37:342–50. [PubMed: 18770019]
- Kolakofsky D, Pelet T, Garcin D, Hausmann S, Curran J, Roux L. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J Virol*. 1998; 72:891–9. [PubMed: 9444980]
- Kumar S, Nayak B, Collins PL, K S. Samal. Complete genome sequence of avian paramyxovirus type 3 reveals an unusually long trailer region. *Virus Res*. 2008; 137:189–97. [PubMed: 18691616]
- Lamb RA, Choppin PW. The synthesis of Sendai virus polypeptides in infected cells. III. Phosphorylation of polypeptides. *Virology*. 1977; 81:382–397. [PubMed: 197698]

- Lamb, RA.; Collins, PL.; Kolakofsky, D.; Melero, JA.; Nagai, Y.; Oldstone, MBA.; Pringle, CR.; Rima, BK. Family *Paramyxoviridae*. In: van Regenmortel, MHV.; Fauquet, CM.; Bishop, DHL.; Carstens, EB.; Estes, MK.; Lemon, SM.; Maniloff, J.; Mayo, MA.; Mc-Geoch, DJ.; Pringle, CR.; Wickner, RB., editors. *Virus Taxonomy: Classification and Nomenclature of Viruses*. The Seventh report of the International Committee for Taxonomy of Viruses. Academic Press; San Diego: 2000. p. 549-561.
- Lamb RA, Collins PL, Kolakofsky D, Melero JA, Nagai Y, Oldstone MBA, Pringle CR, Rima BK. Fauquet CM. Family *Paramyxoviridae*. *Virus Taxonomy: The Classification and Nomenclature of Viruses*. The Eighth Report of the International Committee in Taxonomy of Viruses. 2005
- Lamb, RA.; Parks, GD. *Paramyxoviridae*: the viruses and their replication. In: Knipe, DM.; Howley, PM., editors. *Fields Virology*. 5th ed. Lippincott Williams and Wilkins; Philadelphia. USA: 2007. p. 1449-1496.
- Langedijk JP, Daus FJ, van Oirschot JT. Sequence and structure alignment of Paramyxoviridae attachment proteins and discovery of enzymatic activity for a morbillivirus hemagglutinin. *J Virol*. 1997; 71:6155–6167. [PubMed: 9223510]
- Li Z, Yu M, Zhang H, Wang HY, Wang LF. Improved rapid amplification of cDNA ends (RACE) for mapping both the 5' and 3' terminal sequences of paramyxovirus genomes. *J Virol Meth*. 2005; 130:154–156.
- Lipkind M, Shihmanter E. Antigenic relationships between avian paramyxoviruses. I. Quantitative characteristics based on hemagglutination and neuraminidase inhibition tests. *Arch Virol*. 1986; 89:89–111. [PubMed: 2424406]
- Malur AG, Gupta NK, De Bishnu P, Banerjee AK. Analysis of the mutations in the active site of the RNA-dependent RNA polymerase of human parainfluenza virus type 3 (HPIV3). *Gene Expr*. 2002; 10:93–100. [PubMed: 12064576]
- Miller PJ, Boyle DB, Eaton BT, Wang LF. Full-length genome sequence of Mossman virus, a novel paramyxovirus isolated from rodents in Australia. *Virology*. 2003; 317:330–344. [PubMed: 14698671]
- Mirza AM, Deng R, Iorio RM. Site-directed mutagenesis of a conserved hexapeptide in the paramyxovirus hemagglutinin-neuraminidase glycoprotein: effects on antigenic structure and function. *J Virol*. 1994; 68:5093–5099. [PubMed: 8035509]
- Mizutani T, Endoh D, Okamoto M, Shirato K, Shimizu H, Arita M, Fukushi S, Saijo M, Sakai K, Lim CK, Ito M, Nerome R, Takasaki T, Ishii K, Suzuki T, Kurane I, Morikawa S, Nishimura H. Rapid genome sequencing of RNA viruses. *Emerg Infect Dis*. 2007; 13:322–324. [PubMed: 17479903]
- Moll M, Diederich S, Klenk HD, Czub M, Maisner A. Ubiquitous activation of the Nipah virus fusion protein does not require a basic amino acid at the cleavage site. *J. Virol*. 2004; 78:9705–9712. [PubMed: 15331703]
- Morgan, EM. Evolutionary relationships of paramyxovirus nucleocapsid-associated proteins. In: Kingsbury, DW., editor. *The paramyxoviruses*. Plenum Press; New York: 1991. p. 163-179.
- Morrison TG. Structure and function of a paramyxovirus fusion protein. *Biochim Biophys Acta*. 2003; 11:73–84. [PubMed: 12873767]
- Morrison T, McQuain C, Sergel T, McGtnnes L, Reitter J. The role of the amino terminus of F 1 of the Newcastle disease virus fusion protein in cleavage and fusion. *Virology*. 1993; 193:997–1000. [PubMed: 8460504]
- Nayak B, Kumar S, Collins PL, Samal SK. Molecular characterization and complete genome sequence of avian paramyxovirus type 4 prototype strain duck/Hong Kong/D3/75. *Virol J*. 2008; 5:124. [PubMed: 18937854]
- Nerome K, Nakayama M, Ishida M, Fukumi H. Isolation of a new avian paramyxovirus from budgerigar (*Melopsittacus undulatus*). *J Gen Virol*. 1978; 38:293–301. [PubMed: 627875]
- Notredame C, Higgins D, Heringa J. T-Coffee: A novel method for multiple sequence alignments. *J Mol Biol*. 2000; 302:205–217. [PubMed: 10964570]
- Nylund S, Karlsen M, Nylund A. The complete genome sequence of the Atlantic salmon paramyxovirus (ASPV). *Virology*. 2008; 373:137–148. [PubMed: 18155122]

- Paldurai A, Subbiah M, Kumar S, Collins PL, Samal SK. Complete genome sequences of avian paramyxovirus type 8 strains goose/Delaware/1053/76 and pintail/Wakuya/20/78. *Virus Res.* 2009 doi:10.1016/j.virusres.2009.02.003.
- Panda A, Elankumaran S, Krishnamurthy S, Huang Z, Samal SK. Loss of N-linked glycosylation from the hemagglutinin-neuraminidase protein alters virulence of Newcastle disease virus. *J Virol.* 2004; 78:4965–4975. [PubMed: 15113876]
- Peeples ME. Paramyxovirus M Proteins: Pulling it all Together and Taking it on the Road. Plenum Press; New York: 1991.
- Peeples M, Wang C, Gupta KC, Coleman N. Nuclear entry and nucleolar localization of the Newcastle disease virus (NDV) matrix protein occur early in infection and do not require other NDV proteins. *J. Virol.* 1992; 66:3263–3269. [PubMed: 1560547]
- Poch O, Blumberg BM, Bougueleret L, Tordo N. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol.* 1990; 71:1153–1162. [PubMed: 2161049]
- Saif YM, Mohan R, Ward L, Senne DA, Panigrahy B, Dearth RN. Natural and experimental infection of turkeys with avian paramyxovirus-7. *Avian Dis.* 1997; 41:326–329. [PubMed: 9201395]
- Sakai K, Mizutani T, Fukushi S, Saijo M, Endoh D, Kurane I, Takehara K, Morikawa S. An improved procedure for rapid determination of viral RNA sequences of avian RNA viruses. *Arch Virol.* 2007; 152:1763–1765. [PubMed: 17541697]
- Samal SK, Collins PL. RNA replication by a respiratory syncytial virus RNA analog does not obey the rule of six and retains a nonviral trinucleotide extension at the leader end. *J Virol.* 1996; 70:5075–5082. [PubMed: 8764015]
- Samuel AS, Kumar S, Madhuri S, Collins PL, Samal SK. Complete sequence of the genome of avian paramyxovirus type 9 and comparison with other paramyxoviruses. *Virus Res.* 2009 doi:10.1016/j.virusres.2008.12.016.
- Schmitt AP, Leser GP, Morita E, Sundquist WI, Lamb RA. Evidence for a new viral late-domain core sequence, FPIV, necessary for budding of a paramyxovirus. *J Virol.* 2005; 79:2988–2997. [PubMed: 15709019]
- Schnell MJ, Conzelmann KK. Polymerase activity of *in vitro* mutated rabies virus L protein. *Virology.* 1995; 214:522–530. [PubMed: 8553554]
- Shioda T, Iwasaki K, Shibuta H. Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of the F, HN and L proteins. *Nucl. Acids Res.* 1986; 14:1545–1563. [PubMed: 3005975]
- Subbiah M, Xiao S, Collins PL, Samal SK. Complete sequence of the genome of avian paramyxovirus type 2 (strain Yucaipa) and comparison with other paramyxoviruses. *Virus Res.* 2008; 137:40–48. [PubMed: 18603323]
- Takeda M, Ohno S, Seki F, Nakatsu Y, Tahara M, Yanagi Y. Long untranslated regions of the measles virus M and F genes control virus replication and cytopathogenicity. *J Virol.* 2005; 79:14346–14354. [PubMed: 16254369]
- Takimoto T, Taylor GL, Crennell SJ, Scroggs RA, Portner A. Crystallization of Newcastle disease virus hemagglutinin-neuraminidase glycoprotein. *Virology.* 2000; 270:208–214. [PubMed: 10772993]
- Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature.* 1983; 303:35–40. [PubMed: 6843658]
- Yan Y, Rout SN, Kim SH, Samal SK. Role of Untranslated Regions of Hemagglutinin-Neuraminidase Gene in Replication and Pathogenicity of Newcastle Disease Virus. *J Virol.* 2009 doi:10.1128/JVI.00188-09.
- Yu M, Hansson E, Shiell B, Michalski W, Eaton BT, Wang LF. Sequence analysis of the Hendra virus nucleoprotein gene: comparison with other members of the subfamily *Paramyxovirinae*. *J. Gen. Virol.* 1998; 79:1775–1780. [PubMed: 9680142]

Fig. 1. Schematic diagram of APMV-7 genome and location of PCR products of viral cDNA (denoted as discrete blocks) amplified using the RDA (cDNA representational difference analysis) methods. Two products from the HN gene were contiguous.

a.

3' UGGUUUGUUCUGACGUUAUCGUCACAUUUUGAAAAUUUUUUUAUUGAAAGCA (55)

APMV-7CU..U.G.C.C..AAUG.UA...UCCGC..CC.CG...ACUUC...G (55)

APMV-2U.UCC.....UG.....A...U.CU.AU...GG...CUU.G... (55)

APMV-3 ..A.....CUU.U.AU...GAACA.UCA.CA.CGCG.A...AGGCGA.U.UA. (55)

APMV-4 ..C...U.CUUCUUAU.U.C....UUCGGAAA.UUU.CC..GGGACCC..C.... (55)

APMV-6UUG...A..AAC.CC..AA.UGC.CU.GCGAAUA.UUU.... (55)

APMV-8U..U..UC.G..UG.CC..GA..UU.AU...G..GCUU..UA. (55)

APMV-9U..UUAAC....UAUG..A...CUGGC..CA.CG..GACUG.... (55)

b.

5' ACCAAACAAGGAAGACAAUAGUACAUAUUAAUUAUUUUUUUAAAA (127)

APMV-7A..UUUGGUG.AUGA.G.GACUAC.C.CAAG.....U.G.GCGC.. (114)

APMV-2G...U...UA.C.A.GC.A.GUAACUCAG.GA..AUGC.GUGC..AU.UU (154)

APMV-3 ..U.....AA.GUUAU.AUGAU.UUUUA.UCAG.C..CA.CCUUU.G..CUUUG (707)

APMV-4 ..G...U..AAG.CAU.----- (17)

APMV-6U..UA.GCU.UUUGACUAUGA.CC..GUCG.C...CC.GG- (54)

APMV-8G.A.U...UA..GG.AAUA.GG.A.U.UU.AC.AAU.C.UU....UU. (171)

APMV-9A..UUUGGCA.CAUA.GUGAC.CC...AAA.CGU..CUG----- (47)

c.

Leader 3' UGGUUUGUUCUGACGUUAUCGUCACAUUUUGAAAAUUUUUUUAUUGAAAGCA

+++++ + + + + + + +

Trailer 5' ACCAAACAAGGAAGACAAUAGUACAUAUUAAUUAUUUUUUUAAAA

Fig. 2. Sequence alignments (negative-sense) of the (a) 3'-leader (b) 5'-trailer regions of APMV-7 with other avulaviruses, and (c) terminal complementarity between the leader and trailer of APMV-7. Dots in (a) and (b) indicate identical nucleotides to APMV-7; crosses in (c) indicate complementarity. Gaps are indicated by dashes. A GE-like sequence in the APMV-7 is underlined.

Fig. 3.

Sequence alignments of (a) the RNA editing site in the P gene (shown in negative-sense) and (b) the C-terminal of the V protein of APMV-7 with members of genus *Avulavirus*. The editing site is underlined. Dots indicate nucleotides identical to that of APMV-7. The conserved cysteine residues in V protein are marked by stars.

N protein

P protein

M protein

F protein

Fig. 4. Phylogenetic analysis of the N, P, M, F, HN and L proteins of members of family *Paramyxoviridae*. The numbers represent the amino acid sequence distance between APMV-7 to other viruses. The phylogenetic trees were analyzed by average distance using Blosum62 in the T-Coffee program.

Table 1

Molecular features of genes, mRNA and deduced proteins of APMV-7

Gene	Hexamer phasing position at gene-start	mRNA features (nt)			Protein size (aa)	Intergenic region (nt)
		Total length	5'UTR	ORF	3'UTR	
N	2	1644	75	1392	155	24
P	2	1453	55	1185	191	13
M	4	1211	41	1092	58	70
F	1	1916	82	1620	193	11
HN	2	2021	54	1710	236	15
L	4	6920	5	6684	210	-

Table 2The gene-start and gene-end sequences of APMV-7[†]

Gene-start	Gene	Gene-end
CUCCACUUA	N	AAUUAUUUUUU
CUCCCCUUG	P	AAUUCUUUUUU
CUCCUCUGG	M	AAUUAUUUAU
CUCCGCUUA	F	AAUUAUUUUUU
CUCCGCUUU	HN	AAUUAUUUUUU
CUCCCCUGG	L	AAUUCUUUUUU
CUCCNCUNN [*]		AAUNNUUNU ₁₋₃ [*]

[†] Gene-start and gene-end sequences are given in negative or genome sense^{*} Consensus sequence

Table 3

The consensus sequences of gene-start and gene-end signals of APMVs[‡]

APMV	Gene-start	Gene-end
APMV-7	CUCCNCUNN	AAUNNUUNU ₁₋₃
APMV-1	UGCCCAUCUU	AAUCU ₆₋₇
APMV-2	CCCCCGCUGU	AAUUCU ₆
APMV-3	UCCUCGCCUU	AAUUAU ₆
APMV-4	CACCCCUUCC	AAUUAU ₅
APMV-6	CUCCCCCUUC	AAUUAU ₅₋₇
APMV-8	CCCCCGCUGG	AAUUCU ₆
APMV-9	UGCCCAUCUU	AAUNU ₆

[‡]Gene-start and gene-end sequences are given in negative or genome sense

Table 4

Comparison of the IGS lengths of APMVs

APMV	N/P	P/M	M/F	F/HN	F/SH	SH/HN	HN/L
APMV-7	24	13	70	11	-	-	15
APMV-1	1	1	1	31	-	-	47
APMV-2	7	7	23	9	-	-	3
APMV-3	55	63	31	34	-	-	59
APMV-4	9	34	14	37	-	-	42
APMV-6	7	2	59	-	49	29	63
APMV-8	2	30	25	8	-	-	1
APMV-9	19	9	30	23	-	-	0
Range	1-55	1-63	1-70	8-37	49	29	0-63

Table 5

Comparison of the domains/motifs observed in N, M, HN and L proteins of APMVs

APMV	N self-assembly	M late domain	HN sialic acid binding	L domain III
	FXXXXYXXXΦSΦAMG	αPΦΦ	NRKSCS	QGDNQ
APMV-7	³²⁴FAPANYTLLYSYAMG³³⁸	²²FPII²⁵	²²⁴NRKSCS²²⁹	⁷⁶⁷QGDNQ⁷⁷¹
APMV-1	³²² ...E.AQ...F... ³³⁶	²³ ...V ²⁶	²³⁴ ²³⁹	⁷⁴⁹ ⁷⁵³
APMV-2	³²⁴FST..... ³³⁸	²² Y.L. ²⁵	²³⁵ ²⁴⁰	⁷⁷³ ⁷⁷⁷
APMV-3	³²² ...G..S..... ³³⁶	²³ ..L. ²⁶	²³⁷ ²⁴²	⁷⁴⁴ ⁷⁴⁸
APMV-4	³²² ...G.FPHM..... ³³⁶	²² ..L. ²⁵	²²⁹ ²³⁴	⁷⁵⁶ ⁷⁶⁰
APMV-6	³²⁴ ...G..P.M..... ³³⁸	²² ...V ²⁵	²⁴⁰N ²⁴⁵	⁷⁷⁶ ...E. ⁷⁸⁰
APMV-8	³²⁴STM..... ³³⁸	²² ..LV ²⁵	²³⁶ ²⁴¹	⁷⁷³ ⁷⁷⁷
APMV-9	³²² ...E.AQ..... ³³⁶	²³ ...V ²⁶	²³⁴ ²³⁹	⁷⁴⁹ ⁷⁵³

Table 6

The percent amino acid sequence identities observed in the proteins of APMV-7 with those of other APMVs

APMV	Amino acid sequence identity (%)			
	N	P	M	L
APMV-1	40.8	21.1	33.3	38.4
APMV-2	54.9	26.8	42.3	39.0
APMV-3	35.7	23.3	29.0	28.0
APMV-4	36.5	19.6	28.9	29.3
APMV-6	53.5	24.4	44.5	36.2
APMV-8	54.4	27.2	43.5	39.0
APMV-9	41.3	21.6	33.3	35.8

Table 7

Comparison of the putative cleavage sites of F protein and exogenous protease requirements of APMVs

APMV	F protein cleavage site [†]	Requirement of exogenous protease [‡]
APMV-7	TL<u>PSS</u>R ↓ FAGLVVGA	—
APMV-1(Avirulent)	GGR <u>R</u> QG. ↓ LI.AII.G	+
APMV-1(Virulent)	GRR <u>R</u> QK. ↓ .I.AII.S	—
APMV-2	D <u>K</u> .A.. ↓ .V.AII.S	—
APMV-3	A <u>R</u> . <u>R</u> G. ↓ LF.P...S	+
APMV-4	ADIQP. ↓ .I.AII.AT	—
APMV-6	PA.EP. ↓ LI.AII.T	—*
APMV-8	.Y.QT. ↓ LI.A.I.S	+
APMV-9	<u>R</u> I <u>R</u> E <u>G</u> . ↓ IF.AII..G	+

[†] Basic amino acids (R/K) are underlined and downward arrow indicates the site of cleavage

[‡] Virus replication in cell culture requires the supplementation of 10% allantoic fluid or acetylated trypsin (1µg/ml)

* APMV-6 does not require exogenous protease for growth in primary chicken embryo kidney cells (Xiao and Samal, unpublished data).