Virology **295,** 119–132 (2002)
doi:10.1006/viro.2001.1355, available online at http://www.idealibrary.com on IDEAL®

Analysis of the Genomic Sequence of a Human Metapneumovirus

Bernadette G. van den Hoogen, Theo M. Bestebroer, Albert D. M. E. Osterhaus, and Ron A. M. Fouchier[1]

Department of Virology, Erasmus Medical Center Rotterdam, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Received October 8, 2001; returned to author for revision November 2, 2001; accepted January 10, 2002

We recently described the isolation of a novel paramyxovirus from children with respiratory tract disease in The Netherlands. Based on biological properties and limited sequence information the virus was provisionally classified as the first nonavian member of the *Metapneumovirus* genus and named human metapneumovirus (hMPV). This report describes the analysis of the sequences of all hMPV open reading frames (ORFs) and intergenic sequences as well as partial sequences of the genomic termini. The overall percentage of amino acid sequence identity between APV and hMPV N, P, M, F, M2-1, M2-2, and L ORFs was 56 to 88%. Some nucleotide sequence identity was also found between the noncoding regions of the APV and hMPV genomes. Although no discernible amino acid sequence identity was found between two of the ORFs of hMPV and ORFs of other paramyxoviruses, the amino acid content, hydrophilicity profiles, and location of these ORFs in the viral genome suggest that they represent SH and G proteins. The high percentage of sequence identity between APV and hMPV, their similar genomic organization (3′-N-P-M-F-M2-SH-G-L-5′), and phylogenetic analyses provide evidence for the proposed classification of hMPV as the first mammalian metapneumovirus.

Key Words: paramyxovirus; human metapneumovirus; genomic sequence; respiratory tract disease.

INTRODUCTION

Recently we reported the isolation of human metapneumovirus (hMPV) from nasopharyngeal aspirate samples taken from young children in the Netherlands (van den Hoogen *et al.,* 2001). The clinical symptoms of these children were largely similar to the respiratory tract illnesses caused by the human respiratory syncytial virus (hRSV), ranging from mild respiratory problems to severe cough, bronchiolitis, and pneumonia. The newly discovered virus displayed a high percentage of sequence identity with and probably a genomic organization similar to avian pneumovirus (APV), the etiological agent of an upper respiratory tract disease in turkeys (Cook, 2000; Giraud *et al.,* 1986). APV, also known as turkey rhinotracheitis virus (TRTV), belongs to the *Metapneumovirus* genus, which together with the *Pneumovirus* genus constitutes the *Pneumovirinae* subfamily within the *Paramyxoviridae* family (International Committee on Taxonomy of Viruses, 2000). The *Pneumovirus* genus contains the mammalian respiratory syncytial viruses (human, ovine, bovine RSV) and pneumonia virus of mice (PVM). Until the discovery of hMPV, APV was the sole member of the *Metapneumovirus* genus. The classification of the two genera is based primarily on the gene constellation; metapneumoviruses lack nonstructural proteins NS1 and NS2 and the gene order is different from that of pneumoviruses (RSV, 3′-NS1-NS2-N-P-M-SH-G-F-M2-L-5′; APV, 3′-N-P-M-F-M2-SH-G-L-5′) (Collins *et al.,* 1986; Randhawa *et al.,* 1997; Ling *et al.,* 1992; Yu *et al.,* 1992).

Sequence analyses of the nucleoprotein (N), phosphoprotein (P), matrix protein (M), and fusion protein (F) genes of hMPV revealed the highest percentage of sequence identity with APV serotype C, the avian pneumovirus found primarily in birds in the United States. These analyses also revealed the absence of NS1 and NS2 at the 3′ end of the viral genome and positioning of F immediately adjacent to M. Here we present the sequences of the putative 22K protein (M2), the small hydrophobic protein (SH), the attachment glycoprotein (G), and the large polymerase protein (L) genes, the intergenic regions, and the trailer sequence. In combination with the sequences described previously the sequences presented here complete the genomic sequence of hMPV with the exception of the extreme 12–15 nt of the genomic termini and establish the genomic organization of hMPV. Side by side comparisons of the sequences of the hMPV genome with those of APV subtypes A, B, and C, hRSV subgroups A and B, bRSV, PVM, and other paramyxoviruses provide strong evidence for the classification of hMPV in the *Metapneumovirus* genus.

RESULTS

Sequence strategy

hMPV isolate 00-1 (van den Hoogen *et al.,* 2001) was propagated in tertiary monkey kidney (tMK) cells and

[1] To whom correspondence and reprint requests should be addressed at Department of Virology, Erasmus Medical Center Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Fax: +31 10 4089485. E-mail: fouchier@viro.fgg.eur.nl.

0042-6822/02 $35.00

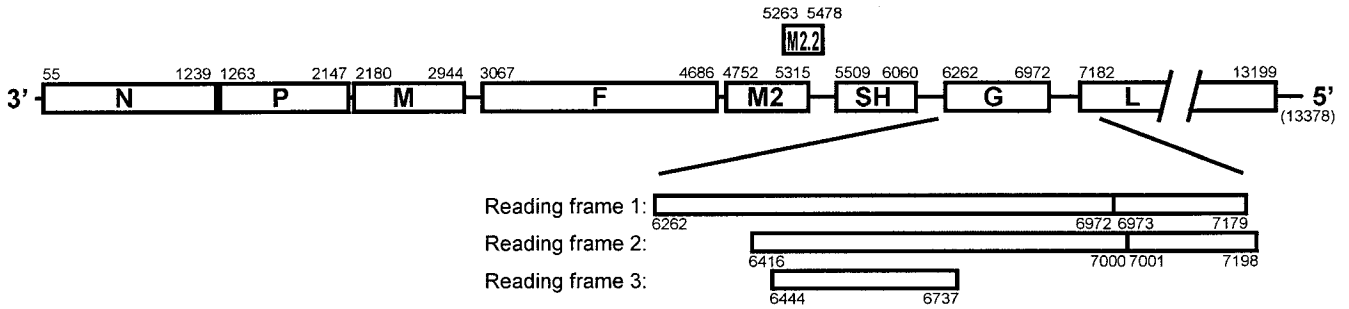

FIG. 1. Genomic map of hMPV isolate 00-1. The putative ORFs and the nt positions of the start and stop codons are indicated. The double lines crossing the L ORF indicate the shortened representation of the L gene. Note that the length of the genome is still uncertain since the sequence of the genomic termini reflects primer sequences. The three reading frames below the map indicate the primary G ORF (nt 6262–6972) and overlapping potential secondary ORFs.

RNA isolated from the supernatant 3 weeks after inoculation was used as template for RT-PCR analyses. Primers were designed on the basis of the partial sequence information available for hMPV 00-1 (van den Hoogen *et al.,* 2001) as well as the leader and trailer sequences of APV and RSV (Randhawa *et al.,* 1997; Mink *et al.,* 1991). Initially, fragments between the previously obtained products, ranging in size from 500 to 4000 bp in length, were generated by RT-PCR amplification and sequenced directly. The genomic sequence was subsequently confirmed by generating a series of overlapping RT-PCR fragments ranging in size from 500 to 800 bp that represented the entire hMPV genome. For all PCR fragments, both strands were sequenced directly to minimize amplification and sequencing errors. The nucleotide and amino acid sequences were used to search for related sequences in the GenBank database using the BLAST software (http://www.ncbi.nlm.nih.gov/BLAST). Putative protein names were assigned to open reading frames (ORFs) based on sequence identity with known viral genes as well as their location in the genome. Based on this information, a genomic map for hMPV was constructed (Fig. 1). The hMPV genome was found to be 13,378 nt in length and its organization was found to be similar to the genomic organization of APV. It should be noted that the sequence of the genomic termini reflects primer sequences and therefore these parts of the genome are still uncertain with respect to size and sequence. Below, we present a comparison between the ORFs and noncoding sequences of hMPV and those of other paramyxoviruses and discuss the important similarities and differences.

The nucleoprotein ORF

As shown previously (van den Hoogen *et al.,* 2001) the first gene in the genomic map of hMPV codes for a 394-aa protein, which resembles the N protein of other pneumoviruses. The length of the N ORF is identical to the length of the N ORF of APV-C (Table 1) and is smaller than those of other paramyxoviruses (Barr *et al.,* 1991). Analysis of the aa sequence revealed the highest percentage of sequence identity with APV-C (88%) and only 7–11% with other paramyxoviruses (Table 2). Barr *et al.*

TABLE 1

Lengths of the Putative ORFs of hMPV and Other Paramyxoviruses

	N[a]	P	M	F	M2-1	M2-2	SH	G	L
hMPV	394	294	254	539	187	71	183	236	2005
APV A	391	278	254	538	186	73	174	391	2004
APV B	391	279	254	538	186	73	—[b]	414	—[b]
APV C	394	294	254	537	184	71	—[b]	—[b]	—[b]
APV D	—[b]	—[b]	—[b]	—[b]	—[b]	—[b]	—[b]	389	—[b]
hRSV A	391	241	256	574	194	90	64	298	2165
hRSV B	391	241	256	574	195	90	65	299	2166
bRSV	391	241	256	574	186	90	81	257	2162
PVM	393	295	257	537	176	98	92	396	—[b]
Others[c]	418–542	225–709	335–393	539–565	—[d]	—[d]	—[d]	—[d]	2183–2262

[a] Lengths in amino acid residues.
[b] Sequence not available.
[c] Human parainfluenza virus types 2 and 3, Sendai virus, measles virus, Nipah virus, phocine distemper virus, and Newcastle disease virus.
[d] ORF absent in viral genome.

(1991) identified three regions of similarity between viruses belonging to the order *Mononegavirales:* A, B, and C (Fig. 2). Here, aa residues are regarded as similar when they are replaced by aa residues from a group sharing some physical or chemical properties: hydrophobic/aliphatic (I, L, V, M, and A); small/neutral (G, P, S, T, and A); acidic/polar (D, E, Q, and N); aromatic (F, Y, W, and H); basic (K and R); thiol (C). Although similarities are highest within a virus family, these regions are highly conserved between virus families. In all three regions hMPV revealed 99.3% aa sequence similarity with APV-C, 98.6% with APV-A, 95.3% with APV-B, and 78–92% with RSV and PVM. The region between aa residues 160 and 340 appears to be highly conserved among metapneumoviruses and to a somewhat lesser extent in the *Pneumovirinae* (Miyahara *et al.,* 1992; Li *et al.,* 1996; Barr *et al.,* 1991). This is in agreement with hMPV being a metapneumovirus, showing 99% similarity with APV-C.

The phosphoprotein ORF

The second ORF in the genome map codes for a 294-aa protein which shares 68% aa sequence identity with the P protein of APV-C and only 22–24% with the P protein of RSV (Table 2). The putative P gene of hMPV contains one substantial ORF and in that respect is similar to P from many other paramyxoviruses (reviewed in Lamb and Kolakofsky, 1996; Sedlmeier and Neubert, 1998).

In contrast to APV-A and -B and PVM and similar to RSV and APV-C the hMPV P ORF lacks cysteine residues. The C-terminus of the hMPV P protein is rich in glutamate residues as has been described for APVs (Ling *et al.,* 1995). Ling *et al.* (1995) suggested that a region of high similarity between all pneumoviruses (aa 185–241) plays a role either in the RNA synthesis process or in maintaining the structural integrity of the nucleocapsid complex. This conserved region may represent the domain that interacts with the polymerase protein. This region of high similarity is also found in hMPV (Fig. 3), showing 100% similarity with APV-C, 93% with APV-A and -B, and approximately 81% with RSV (see the section on the N ORF for the definition of similarity).

The matrix protein ORF

The third ORF of the hMPV genome encodes a 254-aa protein, which resembles the M ORFs of other pneumoviruses. The putative M ORF of hMPV has exactly the same size as the M ORFs of other metapneumoviruses (Table 1) and shows high aa sequence identity with the matrix proteins of APV (76–87%), lower identity with those of RSV and PVM (37–38%), and 10% or less identity with those of other paramyxoviruses (Table 2).

Easton and Chambers (1997) compared the sequences of matrix proteins of all pneumoviruses and found a conserved hexapeptide at residue 14 to 19 that is

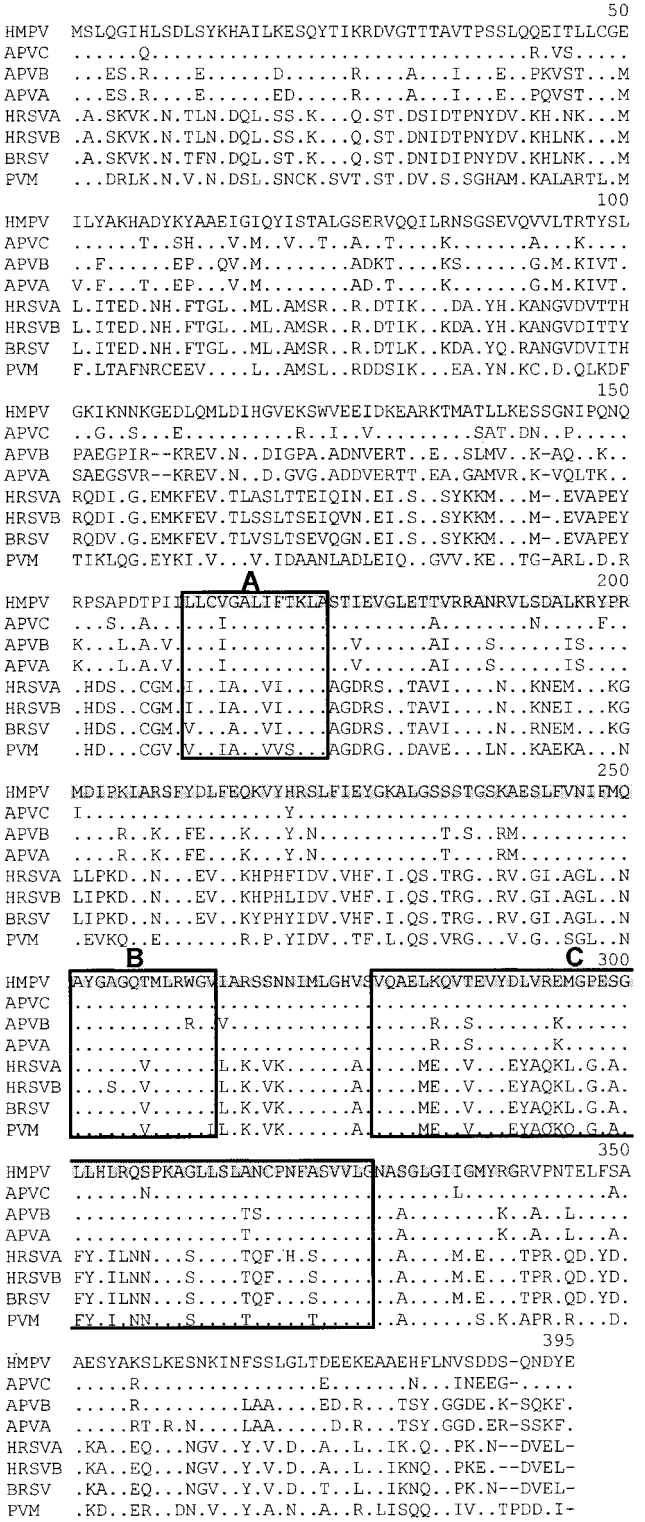

FIG. 2. Amino acid sequence comparison of the putative nucleoprotein ORF of hMPV and other pneumoviruses. The conserved regions (Barr *et al.,* 1991) are represented by boxes and are labeled A, B, and C. The conserved region among pneumoviruses (Li *et al.,* 1996) is shown shaded in gray. Gaps are represented by dashes and periods indicate the positions of identical aa residues compared to hMPV.

TABLE 2

Amino Acid Sequence Identity between the Putative ORFs of hMPV and Those of Other Paramyxoviruses[a]

	N	P	M	F	M2-1	M2-2	L
APV A	69	55	78	67	72	26	64
APV B	69	51	76	67	71	27	—[b]
APV C	88	68	87	81	84	56	—[b]
hRSV A	41	24	38	33	36	18	44
hRSV B	41	23	37	33	35	19	44
bRSV	41	22	38	34	35	13	44
PVM	45	26	37	38	34	12	—[b]
Others[c]	7–11	4–9	7–10	10–18	—[d]	—[d]	13–15

[a] Percentage sequence identity could not be determined for the G and SH proteins because these proteins could not be properly aligned with known G and SH proteins.

[b] Sequence not available.

[c] Human parainfluenza virus types 2 and 3, Sendai virus, measles virus, Nipah virus, phocine distemper virus, and Newcastle disease virus.

[d] ORF absent in viral genome.

also conserved in hMPV (Fig. 4). For RSV, PVM, and APV small secondary ORFs within or overlapping with the major ORF of M have been identified (52 and 51 aa in bRSV, 75 aa in hRSV, 46 aa in PVM, and 51 aa in APV) (Yu *et al.,* 1992; Easton and Chambers, 1997; Samal and Zamora, 1991; Satake and Venkatesan, 1984). We noticed two small secondary ORFs in the M ORF of hMPV. One small ORF of 54 aa residues was found within the major M ORF, starting at nt 2281, and one small ORF of 33 aa residues was found overlapping with the major ORF of M, starting at nt 2893 (data not shown). Similar to the secondary ORFs of RSV and APV there is no significant sequence identity between these secondary ORFs and secondary ORFs of the other pneumoviruses, and apparent start and stop signals are lacking. Evidence for the synthesis of proteins corresponding to these secondary ORFs of APV and RSV has not been reported.

The fusion protein ORF

The putative F ORF of hMPV is located adjacent to the putative M ORF, which is characteristic for members of the *Metapneumovirus* genus. The F gene of hMPV encodes a 539-aa protein, which is 2 aa residues longer than F of APV-C (Table 1). Analysis of the aa sequence revealed 81% sequence identity with APV-C, 67% with APV-A and -B, 33–38% with other pneumovirus F proteins, and only 10–18% with other paramyxoviruses (Table 2). One of the conserved features among F proteins of paramyxoviruses and also seen in hMPV is the distribution of cysteine residues (Morrison, 1988; Yu *et al.,* 1991). The metapneumoviruses share 12 cysteine residues in F1 (7 are conserved among all paramyxoviruses) and 2 in F2 (1 is conserved among all paramyxoviruses). Of the three potential N-linked glycosylation sites present in the F ORF of hMPV, none are shared with RSV and two (positions 66 and 389) are shared with APV. The third, unique, potential N-linked glycosylation site for hMPV is located at position 206 (Fig. 5).

Despite the relatively low percentage of sequence identity with other paramyxoviruses, the F protein of hMPV revealed typical fusion protein characteristics consistent with those described for the F proteins of other *Paramyxoviridae* family members (Morrison, 1988). F proteins of *Paramyxoviridae* members are synthesized as inactive precursors (F0) that are cleaved by host cell proteases which generate N-terminal F2 subunits and large C-terminal F1 subunits. The cleavage site is conserved among all members of the *Paramyxoviridae* family (Collins *et al.,* 1996). The cleavage site of hMPV contains the residues RQSR. Both arginine residues are shared with APV and RSV, but the glutamine and serine residues are shared with other paramyxoviruses such as human parainfluenza virus type 1, Sendai virus, and morbilliviruses (data not shown). In RSV a second cleavage site (RARR) was found which is separated from the first cleavage site immediately upstream of the fusion peptide by a stretch of 27 aa. Cleavage at both sites is required for the acquisition of membrane fusion potential of the F protein. The F proteins of other pneumoviruses such as APV and PVM as well as other paramyxoviruses lack the sequence between the two cleavage sites (Zimmer *et al.,* 2001; González-Reyes *et al.,* 2001). As can be seen in Fig. 5, this second cleavage site and the connecting peptide are absent in the hMPV F protein. The hydrophobic region at the N-terminus of F1 functions as the membrane fusion domain and displays a high degree of sequence similarity among paramyxoviruses and morbilliviruses and to a lesser extent among the pneumoviruses (Morrison, 1988; Horvath and Lamb, 1992). This hydrophobic region (position 137–159, Fig. 5) is conserved between hMPV and APV-C, which is in agreement with this region being highly conserved among the metapneumoviruses (Naylor *et al.,* 1998; Seal *et al.,* 2000). Adjacent to the fusion peptide and transmembrane segment are two regions that contain heptad repeats (HRA and HRB) which are relatively poor in glycines, contain

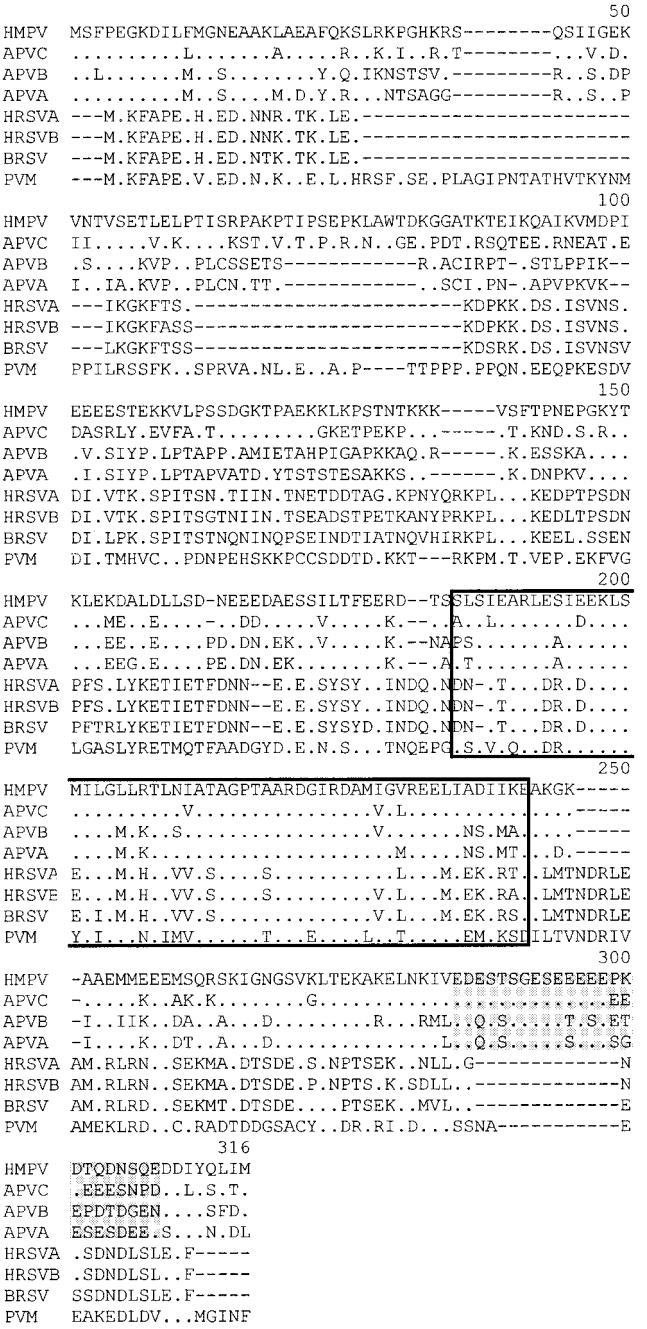

FIG. 3. Amino acid sequence comparison of the putative phosphoprotein ORF of hMPV and other pneumoviruses. The region of high similarity (Ling *et al.,* 1995) is boxed, and the glutamate-rich region is shaded in gray. Gaps are represented by dashes and periods indicate the position of identical aa residues compared to hMPV.

no helix-breaking prolines, and contain charged aa side chains in all heptad positions except a and d. These heptad regions that are necessary for viral fusion (Chambers *et al.,* 1990; Buckland and Wild, 1989; Lamb, 1993; Russell *et al.,* 2001) are also found in the hMPV F gene. Whereas a high percentage of sequence identity is found between HRA of hMPV and all other pneumoviruses (especially in positions a and d of the repeats), a high level of sequence identity in HRB is restricted to the metapneumoviruses (Fig. 5).

Furthermore, for RSV and APV, the signal peptide and anchor domain were found to be conserved within subtypes and displayed high variability between subtypes (Plows and Pringle, 1995; Naylor *et al.,* 1998). At the N-terminus of F2 of hMPV, 11 of 18 aa residues are identical to those of APV-C and lower sequence identity is observed with the signal peptides of other APV or RSV F ORFs. Much more variability is seen in the membrane anchor domain at the C-terminus of F1, although some sequence identity is still seen with APV-C.

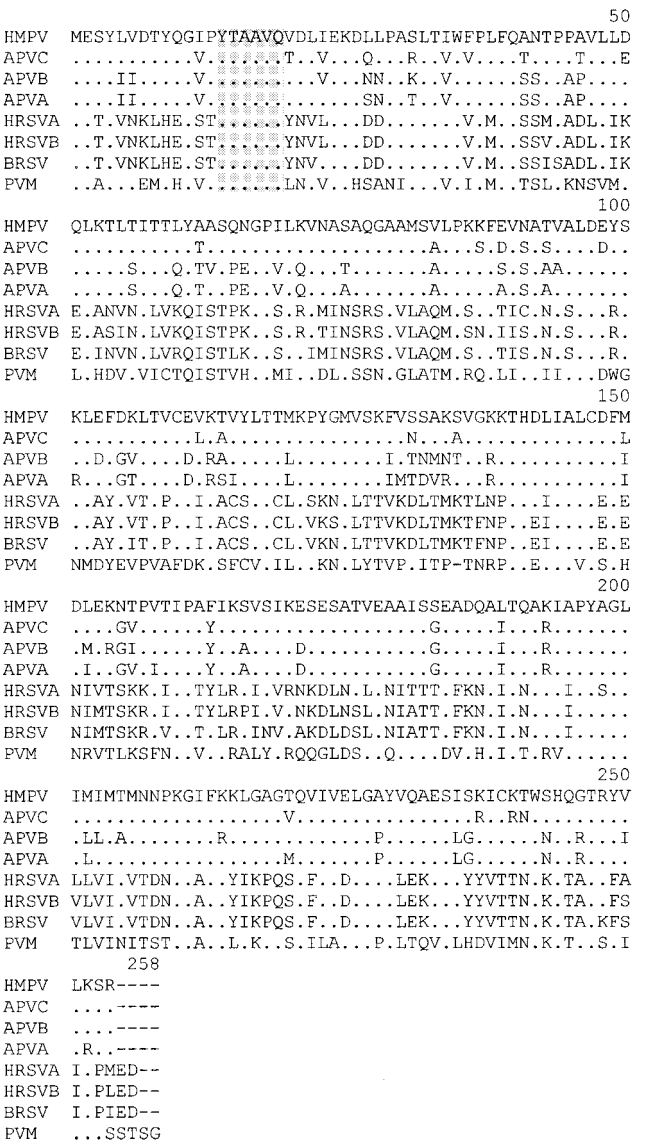

FIG. 4. Amino acid sequence comparison of the putative matrix protein ORF of hMPV and other pneumoviruses. The conserved hexapeptide sequence (Easton and Chambers, 1997) is shaded in gray. Gaps are represented by dashes and periods indicate the position of identical aa residues relative to hMPV.

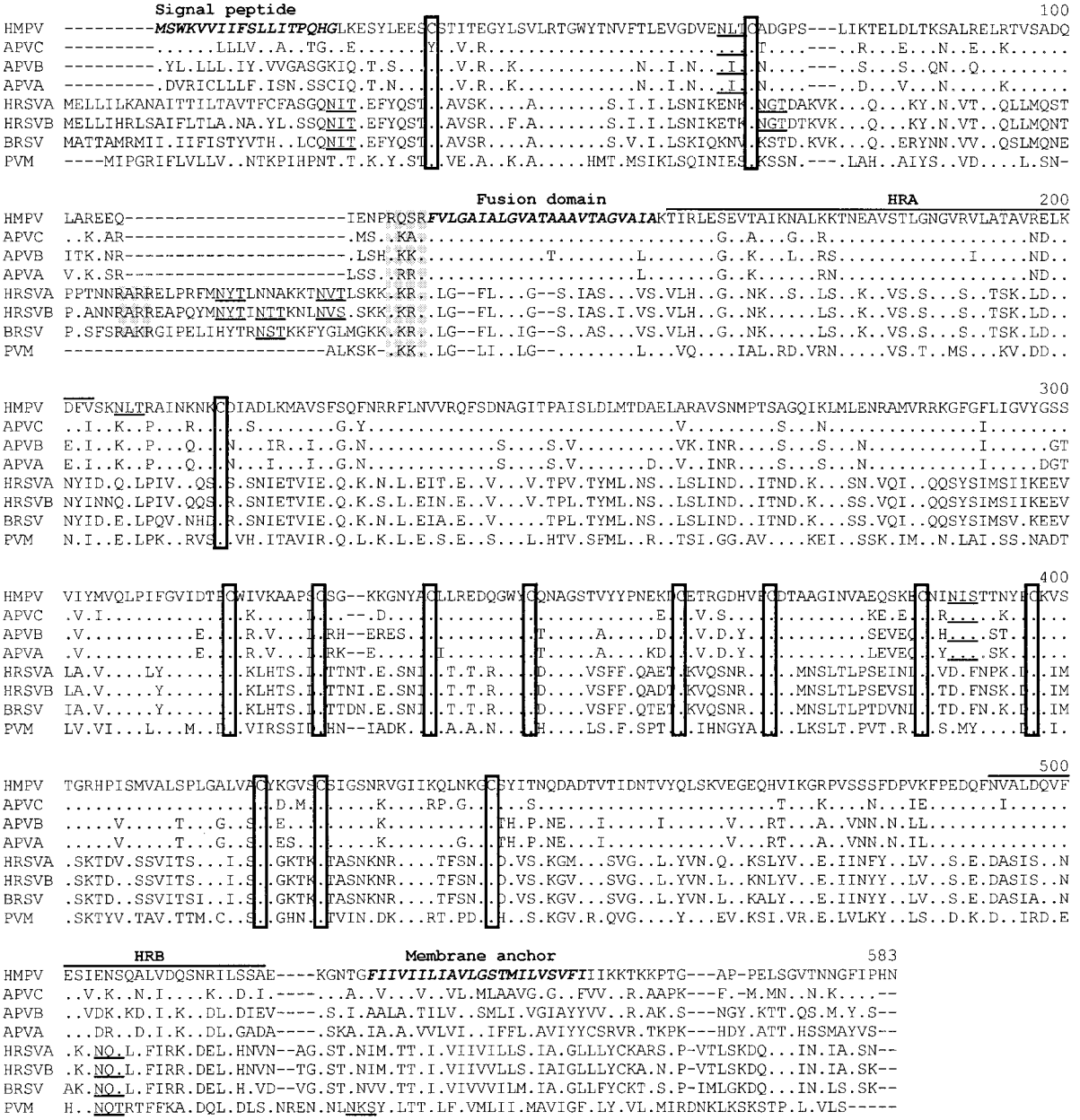

FIG. 5. Amino acid sequence comparison of the putative fusion protein ORF of hMPV and other pneumoviruses. The cleavage sites are shown shaded in gray and three hydrophobic domains (the putative signal peptide, fusion domain, and membrane anchor domain) are shown italicalized in boldface type. The conserved cysteine residues are boxed, putative N-linked glycosylation sites are underlined, and domains HRA and HRB are indicated with a line above the sequence. Gaps are represented by dashes and periods indicate the position of identical aa residues relative to hMPV.

The 22K protein ORF

The M2 gene is unique to the members of the *Pneumovirinae* subfamily and two overlapping ORFs have been observed in all pneumoviruses. The first major ORF represents the M2-1 protein which enhances the processivity of the viral polymerase (Collins *et al.,* 1995, 1996) and its read-through of intergenic regions (Hardy and Wertz, 1998; Fearns and Collins, 1999). The putative M2-1 gene for hMPV, located adjacent to the F gene, encodes a 187-aa protein (Table 1) and reveals the highest percentage of sequence identity with M2-1 of APV-C (84%, Table 2). Comparison of all pneumovirus M2-1 proteins revealed the highest level of conservation in the N-terminal half of the protein (Collins *et al.,* 1990; Zamora and Samal, 1992; Ahmadian *et al.,* 1999), which is in agreement with the observation that hMPV displays 100% similarity with APV-C in the first 80 aa residues of the protein (Fig. 6A). The hMPV M2-1 protein contains 3 cysteine residues located within the first 30 aa residues that are conserved among all pneumoviruses. Such a concentration of cysteines is frequently found in zinc-binding proteins (Ahmadian *et al.,* 1999; Cuesta *et al.,* 2000).

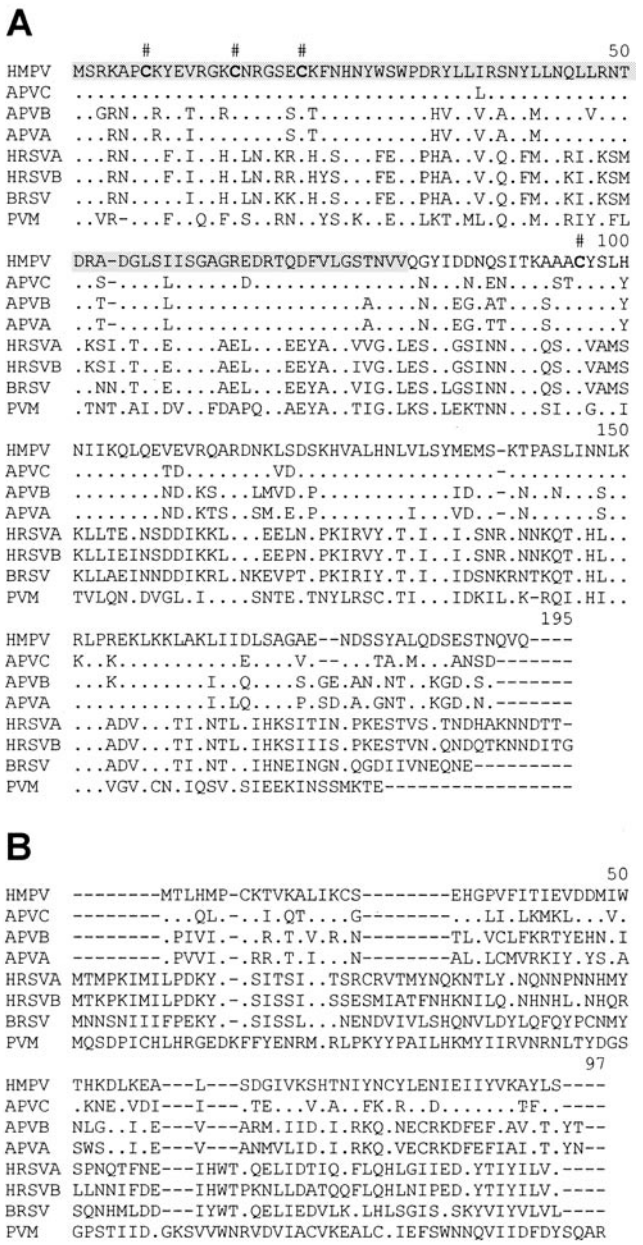

FIG. 6. Amino acid sequence comparison of the putative M2 ORFs of hMPV and other pneumoviruses. (A) Alignment of M2-1 ORFs, with the conserved N-terminus (Collins *et al.,* 1990; Zamora and Samal, 1992) shown shaded in gray. The conserved cysteine residues are printed in boldface type and are indicated by pound signs. (B) Alignment of M2-2 ORFs. Gaps are represented by dashes and periods indicate the position of identical aa residues relative to hMPV.

The secondary ORFs (M2-2) that overlap with the M2-1 ORFs of pneumoviruses are conserved in location but not in sequence and are thought to be involved in the control of the switch between virus RNA replication and transcription (Collins and Wertz, 1985; Collins *et al.,* 1990; Elango *et al.,* 1985; Baybutt and Pringle, 1987; Ling *et al.,* 1992; Zamora and Samal, 1992; Alansari and Potgieter, 1994; Ahmadian *et al.,* 1999; Bermingham and Collins, 1999). For hMPV, the putative M2-2 ORF starts at nt 512 in the M2-1 ORF (Fig. 1), which is exactly the same start position as for APV-C. The lengths of the M2-2 ORFs are the same for APV-C and hMPV, 71 aa residues (Table 1). Sequence comparison of the M2-2 ORF (Fig. 6B) revealed 56% aa sequence identity between hMPV and APV-C and only 26–27% aa sequence identity between hMPV and APV-A and -B (Table 2).

The small hydrophobic protein ORF

The gene located adjacent to M2 of hMPV probably encodes a 183-aa SH protein (Figs. 1 and 7). There is no discernible sequence identity between this ORF and other RNA virus genes or gene products. This is not surprising since sequence similarity between pneumovirus SH proteins is generally low. The putative SH ORF of hMPV is the longest SH ORF known to date (Table 1). The aa composition of the SH ORF is relatively similar to that of APV, RSV, and PVM, with a high percentage of threonine and serine residues (22, 18, 19, 20, 21, and 28% for hMPV, APV, RSV A, RSV B, bRSV, and PVM,

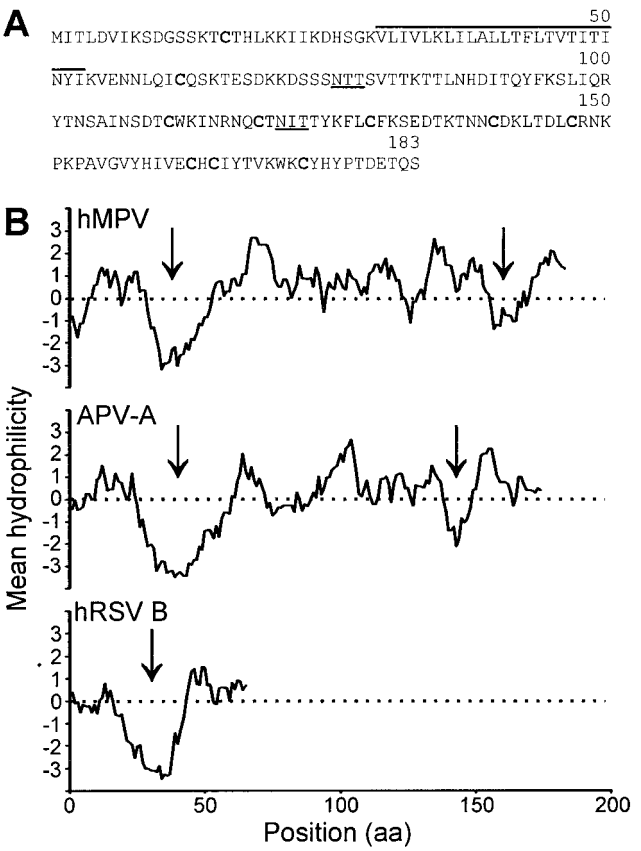

FIG. 7. Amino acid sequence analysis of the putative SH ORF of hMPV. (A) Amino acid sequence of the putative SH ORF of hMPV, with cysteine residues in boldface type, potential N-linked glycosylation sites underlined, and the most hydrophobic region represented by a line above the sequence. (B) Alignment of the hydrophilicity plots of the SH proteins of hMPV, APV-A, and hRSV-B. The procedure of Kyte and Doolittle (1982) was used with a window size of 17 aa residues. Arrows indicate strong hydrophobic domains.

respectively). The SH ORF of hMPV contains 10 cysteine residues, whereas APV SH contains 16 cysteine residues. The SH ORF of hMPV contains two potential N-linked glycosylation sites (aa 76 and 121), whereas APV has one, RSV has two or three, and PVM has four.

The hydrophilicity profiles for the putative hMPV SH protein and SH of APV and RSV revealed similar characteristics (Fig. 7B). The SH ORFs of APV and hMPV have a hydrophilic N-terminus, a central hydrophobic domain which can serve as a potential membrane-spanning domain (aa 30–53 for hMPV), a second hydrophobic domain (aa 155–170), and a hydrophilic C-terminus. In contrast, RSV SH appears to lack the C-terminal part of the APV and hMPV ORFs. In all pneumovirus SH proteins the hydrophobic domain is flanked by basic aa residues, which are also found in the SH ORF for hMPV (aa 29 and 54).

The attachment glycoprotein ORF

The putative G ORF of hMPV is located adjacent to the putative SH gene and encodes a 236-aa protein (nt 6262–6972; Fig. 1). A secondary small ORF is found immediately following this ORF, potentially coding for 68 aa residues (nt 6973–7179) but lacking a start codon. A third potential ORF in the second reading frame of 194 aa residues overlaps both of these ORFs but also lacks a start codon (nt 6416–7000). This ORF is followed by a potential fourth ORF of 65 aa residues in the same reading frame (nt 7001–7198), again lacking a start codon. Finally, a potential ORF of 97 aa residues (but lacking a start codon) is found in the third reading frame (nt 6444–6737, Fig. 1). Unlike the first ORF, the other ORFs do not have apparent gene start or gene end sequences (see below). Although the 236 aa G ORF probably represents at least a part of the hMPV attachment protein, the possibility cannot be excluded that the additional coding sequences are expressed as separate proteins or as part of the attachment protein through some RNA editing event. It should be noted that for APV and RSV no secondary ORFs after the primary G ORF have been identified but that both APV and RSV have secondary ORFs within the major ORF of G. However, evidence for expression of these ORFs is lacking and there is no sequence identity between the predicted aa sequences for different viruses (Ling *et al.,* 1992). The secondary ORFs in hMPV G do not reveal characteristics of other G proteins and whether the additional ORFs are expressed requires further investigation.

BLAST analyses with all ORFs revealed no discernible sequence identity at the nucleotide or aa sequence level with other known virus genes or gene products. This is in agreement with the low percentage of sequence identity found for other G proteins such as those of hRSV A and B (53%) (Johnson *et al.,* 1987) and APV-A and -B (38%) (Juhasz and Easton, 1994).

Whereas most of the hMPV ORFs resemble those of APV in both length and sequence, the putative G ORF of 236 aa residues of hMPV is considerably smaller than the G ORF of APV (Table 1). The aa sequence revealed a serine and threonine content of 34%, which is even higher than the 32% for RSV and 24% for APV. The putative G ORF also contains 8.5% proline residues, which is higher than the 8% for RSV and 7% for APV. The unusual abundance of proline residues in the G proteins of APV, RSV, and hMPV has also been observed in glycoproteins of mucinous origin where it is a major determinant of the proteins' three-dimensional structures (Collins and Wertz, 1983; Wertz *et al.,* 1985; Jentoft, 1990). The G ORF of hMPV contains five potential N-linked glycosylation sites, whereas hRSV has seven, bRSV has five, and APV has three to five.

The predicted hydrophilicity profile of hMPV G revealed characteristics similar to those of other pneumoviruses. The N-terminus contains a hydrophilic region followed by a short hydrophobic area (aa 33–53 for hMPV) and a mainly hydrophilic C-terminus (Fig. 8B). This overall organization is consistent with that of an anchored type II transmembrane protein and corresponds well with these regions in the G protein of APV and RSV. The putative G ORF of hMPV contains only 1 cysteine residue in contrast to RSV and APV (5 and 20, respectively). Of note, only two of the four secondary ORFs in the G gene contained 1 additional cysteine residue and these four potential ORFs revealed 12–20% serine and threonine residues and 6–11% proline residues.

The polymerase protein ORF

In analogy to other negative-strand viruses, the last ORF of the hMPV genome is the RNA-dependent RNA polymerase component of the replication and transcription complexes. The L gene of hMPV encodes a 2005-aa protein, which is one residue longer than the APV-A L protein (Table 1). The L protein of hMPV shares 64% aa sequence identity with APV-A, 44% with RSV, and 13–15% with other paramyxoviruses (Table 2). Poch *et al.* (1989, 1990) identified six conserved domains within the L proteins of nonsegmented negative-strand RNA viruses, from which domain III contained the four core polymerase motifs that are thought to be essential for polymerase function. These motifs (A, B, C, and D) are well conserved in the hMPV L protein: hMPV shares nearly 100% aa sequence similarity with other pneumoviruses. For the entire domain III (aa 625–847 in the hMPV L ORF), hMPV shares 83% aa sequence identity with APV, 67–68% with RSV, and 26–30% with other paramyxoviruses (Fig. 9). In addition to these polymerase motifs the pneumovirus L proteins contain a sequence which conforms to a consensus ATP-binding motif, $K(X)_{21}GEGAGN(X)_{20}K$ (Stec *et al.,* 1991). The hMPV L ORF contains a similar

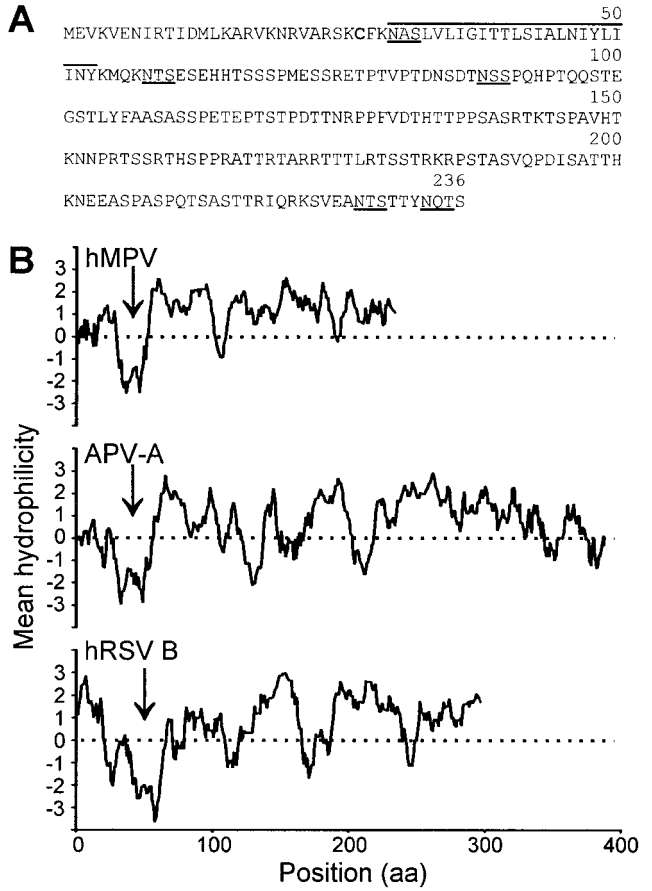

FIG. 8. Amino acid sequence analysis of the putative G ORF of hMPV. (A) Amino acid sequence of the putative G ORF of hMPV, with the cysteine residue in boldface type, potential N-linked glycosylation sites underlined, and the hydrophobic region represented by a line above the sequence. (B) Alignment of the hydrophilicity plots of the G proteins of hMPV, APV-A, and hRSV-B. The procedure of Kyte and Doolittle (1982) was used with a window size of 17 aa residues. Arrows indicate hydrophobic domains.

motif as APV, in which the spacing of the intermediate residues is off by 1: $K(X)_{22}GEGAGN(X)_{19}K$.

Phylogenetic analyses

As an indicator for the relationship between hMPV and members of the *Pneumovirinae,* phylogenetic trees based on the N, P, M, and F ORFs were constructed previously (van den Hoogen *et al.,* 2001) and revealed a close relationship between hMPV and APV-C. Because of the low level of sequence identity of the hMPV SH and G genes with those of other paramyxoviruses, reliable phylogenetic trees for these genes could not be constructed. In addition, the distinct genomic organization between members of the *Pneumovirus* and *Metapneumovirus* genera makes it difficult to generate phylogenetic trees based on the entire genomic sequence. We therefore constructed phylogenetic trees for only the M2-1 and L genes in addition to those previously published. Both these trees confirmed the close relationship between APV and hMPV within the *Pneumovirinae* subfamily (Fig. 10).

hMPV noncoding sequences

The noncoding regions between the ORFs of pneumoviruses contain gene end signals, intergenic regions,

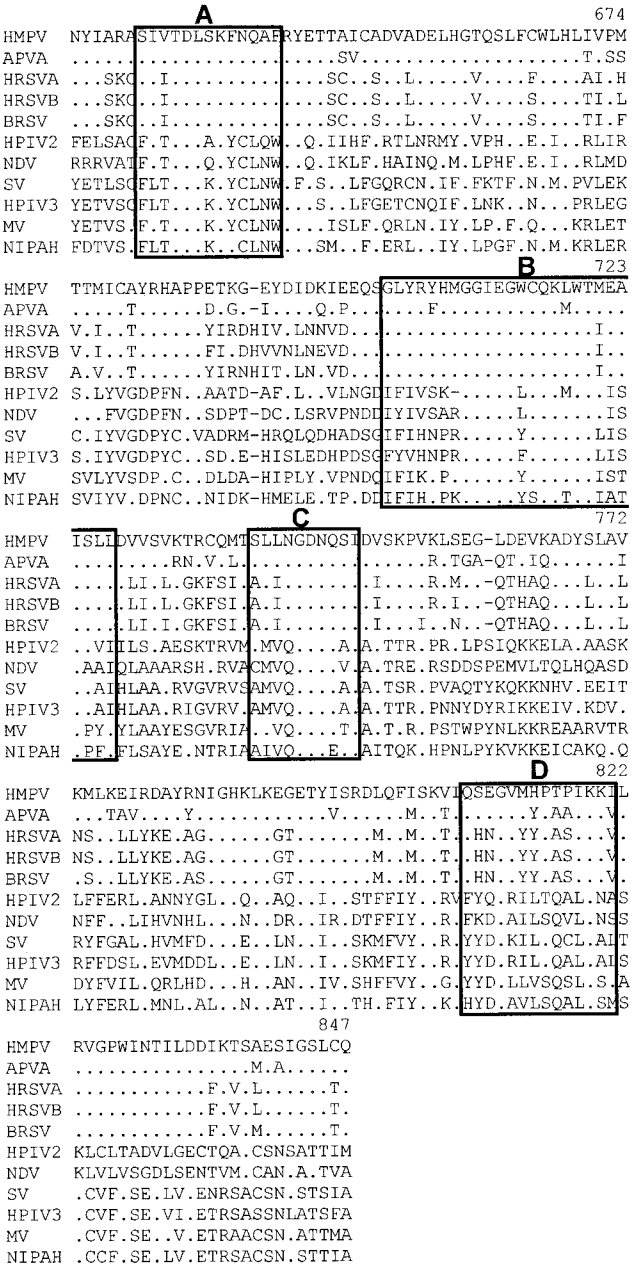

FIG. 9. Amino acid sequence comparison of a conserved domain of the putative polymerase ORF of hMPV and other paramyxoviruses. Domain III is shown with the four conserved polymerase motifs (A, B, C, and D) boxed (Poch *et al.,* 1989, 1990). Gaps are represented by dashes and periods indicate the position of identical aa residues relative to hMPV. HPIV2 and HPIV3, human parainfluenza virus types 2 and 3; SV, Sendai virus; NDV, Newcastle disease virus; MV, measles virus; NIPAH, Nipah virus.

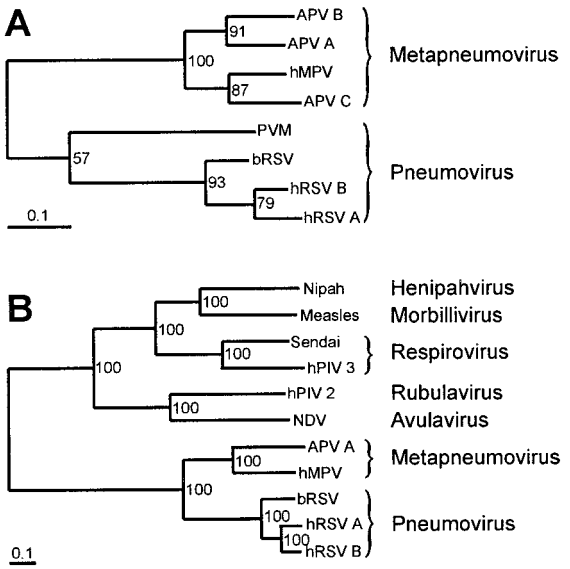

FIG. 10. Phylogenetic analyses of the M2-1 and L ORFs of hMPV and selected paramyxoviruses. The putative M2-1 (A) and L (B) ORFs of hMPV were aligned with those from selected paramyxoviruses (see legend to Fig. 9) and phylogenetic trees were generated by maximum-likelihood analyses using 100 bootstraps and three jumbles. The most probable trees are shown with numbers representing the bootstrap values calculated from the consensus trees. The scale bars roughly represent 10% of nucleotide changes between close relatives.

and gene start signals, and they vary in size and sequence (Curran and Kolakofsky, 1999; Blumberg *et al.,* 1991; Collins and Wertz, 1983). The noncoding sequences between the hMPV ORFs range in size from 23 to 209 nt (Fig. 11A) and do not reveal significant sequence identity with those of APV and RSV. The region between the M and F ORFs of hMPV contains part of a potential secondary ORF, which starts in the primary M ORF (see above). The region between SH and G ORFs of 201 nt does not appear to have coding potential based on the presence of numerous stop codons in all three reading frames. The region between G and L of 209 nt may encode additional ORFs in two reading frames (see above). Interestingly, the start of the L ORF is located in these secondary ORFs. Whereas the L gene of APV does not start in the preceding G ORF, the L ORF of RSV starts in the preceding M2 gene. Comparison of the noncoding sequences between the ORFs of hMPV revealed a consensus sequence for the gene start signal of the N, P, M, F, M2, and G genes, GGGACAA A/G U (Fig. 11A), which is similar to the consensus gene start signal of the APV (Ling *et al.,* 1992; Yu *et al.,* 1992; Li *et al.,* 1996; Bäyon-Auboyer *et al.,* 2000). The gene start signals for the SH and L genes of hMPV were found to be slightly different from this consensus (SH, GGGAUAAAU; L, GAGACAAAU). For APV the gene start signal of L was also found to be different from the consensus: AGGACCAAT (APV-A) (Randhawa *et al.,* 1996) and GGGACCAGT (APV-D) (Bäyon-Auboyer *et al.,* 2000).

In contrast to the similar gene start sequences of hMPV and APV, the consensus gene end sequence of APV, UAGUUAAUU (Chambers *et al.,* 1990; Randhawa *et al.,* 1996), could be found only in the F and L genes of hMPV. Variants of this motif, with 1 to 3 nt substitutions, were found in the N, P, M, M2, and SH genes, but no homologue was found in G. Another repeated sequence (UAAAAA U/A/C) was found downstream of each of the hMPV ORFs except G. The gene end signal described for G of APV Fr/85/1 (Bäyon-Auboyer *et al.,* 2000) was found also at the end of the G ORF of hMPV (Fig. 11A). Although these sequences may represent (part of) the gene end signal, sequence analyses of viral mRNA are required to formally delineate the hMPV gene end signals.

Short extragenic regions at the 3′ and 5′ extremities of the genome of paramyxoviruses are referred to as the leader and trailer sequences, and approximately the first 12 nt of the leader and last 12 nt of the trailer are complementary, probably because each of them contains basic elements of the viral promoter (Curran and Kolakofsky, 1999; Blumberg *et al.,* 1991; Mink *et al.,* 1991). The 3′ leaders of hMPV and APV are both 41 nt in length, and some sequence identity is seen in the region between nt 16 and 41 of both viruses (18 of 26 nt) (Fig. 11B). As mentioned before, the first 15 nt of the hMPV genomic map is based on a primer sequence derived from the APV genome. At the 5′ end of the viral genome of hMPV after the stop codon of the L ORF 179 nt of noncoding sequence is found, including the trailer sequence. Alignment of the APV trailer sequence with the extreme 45 nt of the hMPV genome revealed 24 out of 33 nt to be identical, excluding the extreme 12 nt which represents primer sequence based on the genomic sequence of APV (Fig. 11B).

DISCUSSION

Previously we proposed the classification of hMPV in the *Metapneumovirus* genus based on limited sequence information for this virus. The sequence information described here is in agreement with this classification. Our sequence analyses revealed the absence of NS1 and NS2 genes at the 3′ end of the genome and a genomic organization resembling the organization of metapneumoviruses (3′-N-P-M-F-M2-SH-G-L-5′). The high percentage of sequence identity found between hMPV and APV ORFs further emphasizes the close relationship between these two viruses. For the N, P, M, F, M2-1, and M2-2 ORFs of hMPV an overall aa identity of 80% is found with APV-C. In fact, for these genes APV-C and hMPV revealed percentages of sequence identity which are in the same range as those found between subgroups of other genera, such as RSV-A and -B or APV-A and -B. This close relationship between APV-C and hMPV is also seen in the phylogenetic analyses which revealed hMPV and APV-C to be always in the same branch, separate

FIG. 11. Noncoding sequences of hMPV isolate 00-1. (A) The noncoding sequences between the putative ORFs and at the genomic termini are shown in the positive sense. From left to right, stop codons of indicated ORFs are shown, followed by the noncoding sequences, the gene start signals, and start codons of the indicated subsequent ORFs. Numbers indicate the first position of start and stop codons in the hMPV map. Sequences that display similarity to published gene end signals are underlined and sequences that display similarity to UAAAAAU/A/C are indicated with a line above the sequence. (B) Nucleotide sequences of the genomic termini of hMPV. The genomic termini of hMPV are aligned with each other and with those of APV. Underlined regions represent the primer sequences used in RT-PCR assays which are based on the 3′ and 5′ end sequences of APV and RSV (Randhawa *et al.,* 1997; Mink *et al.,* 1991). Boldface italicalized nucleotides are part of the gene start signal of the N gene. Le, leader; Tr, trailer.

from the branch containing APV-A and -B. The identical genomic organization, the high percentage of sequence identity, and phylogenetic analyses are all in favor of the classification of hMPV as the first mammalian member in the *Metapneumovirus* genus.

It should be noted that previously found sequence variation between different virus isolates of hMPV in the N, M, F, and L genes revealed the possible existence of different genotypes (van den Hoogen *et al.,* 2001). Here we describe the genomic sequence of virus isolate 00-1, the prototype of one of the subgroups. Sequence analysis of the genome of a member of the other subgroup is in progress to further investigate the relationship between distinct hMPV isolates and APV.

The close relationship between hMPV and APV-C is not reflected in the host range, since APV infects birds in contrast to hMPV (van den Hoogen *et al.,* 2001). This difference in host range may be determined by the differences between the SH and the G proteins of both viruses. The SH and G proteins of hMPV did not reveal significant aa sequence identity with SH and G proteins of any other virus. Although the aa content and hydrophilicity plots are in favor of defining these ORFs as SH and G, experimental data are required to formally prove this. Such analyses will also shed light on the role of the additional overlapping ORFs in the putative G gene. Sequence analysis of a second hMPV isolate that belongs to a distinct genetic lineage (van den Hoogen *et al.,* 2001) revealed the presence of a similar primary G ORF and the presence of secondary ORFs (data not shown). However, the positions and sizes of the secondary ORFs were not identical to those of isolate 00-1. Sequence analyses of the SH and G genes of APV-C might provide more insight into the function of the putative SH and G

proteins of hMPV and their relationship with those of APV-C. Some of the noncoding regions of hMPV were found to be similar to those of APV. The 3′ leader and 5′ trailer sequences of APV and hMPV displayed a high percentage of sequence identity. Although the lengths of the intergenic regions were not always the same for APV and hMPV, the consensus gene start signals of most of the genes were found to be identical and some similarity may exist between potential gene end signals of APV and hMPV. Sequence analysis of viral mRNAs is required to formally delineate the gene end sequences of hMPV. It should also be noted that sequence information for 15 nt at the extreme 3′ end and 12 nt at the extreme 5′ end is still missing. This information will be of crucial importance for the development of reverse genetics systems for hMPV, as was shown recently for RSV and APV (Marriott *et al.,* 2001). Nevertheless, the sequence information provided here will be of importance for the generation of diagnostic tests, vaccines, and antivirals for hMPV.

MATERIALS AND METHODS

Sequence analysis

Virus isolate 00-1 was propagated to high titers (approximately 10,000 $TCID_{50}$/ml) on tertiary monkey kidney cells as described previously (van den Hoogen *et al.,* 2001). Viral RNA was isolated from supernatants from infected cells using a High Pure RNA Isolating Kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands). Primers were designed based on sequences published previously (van den Hoogen *et al.,* 2001) in addition to sequences published for the leader and trailer of APV/RSV (Randhawa *et al.,* 1997; Mink *et al.,* 1991) and are available upon request from the authors. RT-PCR assays were conducted with viral RNA, using a one-tube assay in a total volume of 50 μl with 50 mM Tris, pH 8.5, 50 mM NaCl, 4.5 mM $MgCl_2$, 2 mM DTT, 1 μM forward primer, 1 μM reverse primer, 0.6 mM dNTPs, 20 units RNAsin (Promega, Leiden, The Netherlands), 10 U AMV reverse transcriptase (Promega), and 5 units *Taq* polymerase (PE Applied Biosystems, Nieuwerkerk aan den IJssel, The Netherlands). Reverse transcription was conducted at 42°C for 30 min, followed by 8 min of inactivation at 95°C. The cDNA was amplified during 40 cycles of 95°C for 1 min, 42°C for 2 min, 72°C for 3 min, with a final extension at 72°C for 10 min. After examination on a 1% agarose gel, the RT-PCR products were purified from the gel using a Qiaquick Gel Extraction Kit (Qiagen, Leusden, The Netherlands) and sequenced directly using a Dyenamic ET Terminator Sequencing Kit (Amersham Pharmacia Biotech, Roosendaal, The Netherlands) and an ABI 373 automatic DNA sequencer (PE Applied Biosystems), according to the instructions of the manufacturer.

Sequence alignments were made using the Clustal software package available in the software package of BioEdit version 5.0.6 (Hall, 1999) (http://jwbrown.mbio.ncsu.edu/Bioedit/bioedit.html).

Phylogenetic analysis

To construct phylogenetic trees, DNA sequences were aligned using the ClustalW software package and maximum-likelihood trees were generated using the DNA-ML software package of the Phylip 3.5 program using 100 bootstraps and three jumbles. Bootstrapping is a method for deriving confidence values for the groupings in a tree, which involves making random samples of sites from the alignment, drawing trees from each sample, and counting how many times each grouping from the original tree occurs in the sample trees. Bootstrap values were computed for consensus trees created with the Consensus package (Felsenstein, 1989).

The hMPV genomic sequence is available from GenBank under accession No. AF371337. All other sequences used here are available from GenBank under Accession Nos. AB046218 (measles virus, all ORFs), NC001796 (human parainfluenza virus type 3, all ORFs), NC001552 (Sendai virus, all ORFs), X57559 (human parainfluenza virus type 2, all ORFs), NC002617 (Newcastle disease virus, all ORFs), NC002728 (Nipah virus, all ORFs), NC001989 (bRSV, all ORFs), M11486 (hRSV A, all ORFs except L), NC001803 (hRSV, L ORF), NC001781 (hRSV B, all ORFs), D10331 (PVM, N ORF), U09649 (PVM, P ORF), U66893 (PVM, M ORF), U66893 (PVM, SH ORF), D11130 (PVM, G ORF), D11128 (F ORF), AF176590 (APV-C, N ORF), U39295 (APV-A, N ORF), U39296 (APV-B, N ORF), AF262571 (APV-C, M ORF), U37586 (APV-B, M ORF), X58639 (APV-A, M ORF), AF176591 (APV-C, P ORF), AF325443 (APV-B, P ORF), U22110 (APV-A, P ORF), AF187152 (APV-C, F ORF), Y14292 (APV-B, F ORF), D00850 (APV-A, F ORF), AF176592 (APV-C, M2 ORF), AF356650 (APV-B, M2 ORF), X63408 (APV-A, M2 ORF), U65312 (APV-A, L ORF), S40185 (APV-A, SH ORF). The PVM M2 ORF was taken from Ahmadian *et al.* (1999).

ACKNOWLEDGMENTS

We thank M. Spronken, E. Fries, and O. Schaap for excellent technical assistance. R.F. is a fellow of the Royal Dutch Academy of Arts and Sciences.

REFERENCES

Ahmadian, G., Chambers, P., and Easton, A. J. (1999). Detection and characterization of proteins encoded by the second ORF of the M2 gene of pneumoviruses. *J. Gen. Virol.* **80,** 2011–2016.

Alansari, H., and Potgieter, L. N. (1994). Molecular cloning and sequence analysis of the phosphoprotein, nucleocapsid protein, matrix protein and 22K (M2) protein of the ovine respiratory syncytial virus. *J. Gen. Virol.* **75,** 3597–3601.

Barr, J., Chambers, P., Pringle, C. R., and Easton, A. J. (1991). Sequence of the major nucleocapsid protein gene of pneumonia virus of mice:

Sequence of comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. *J. Gen. Virol.* **72,** 677–685.

Baybutt, H. N., and Pringle, C. R. (1987). Molecular cloning and sequencing of the F and 22K membrane protein genes of the RSS-2 strain of respiratory syncytial virus. *J. Gen. Virol.* **68,** 2789–2796.

Bäyon-Auboyer, M. H., Arnauld, C., Toquin, D., and Eterradossi, N. (2000). Nucleotide sequences of the F, L and G protein genes of two non-A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup. *J. Gen. Virol.* **81,** 2723–2733.

Bermingham, A., and Collins, P. L. (1999). The M2–2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proc. Natl. Acad. Sci. USA* **96,** 11259–11264.

Blumberg, B. M., Chan, J., and Udem, S. A. (1991). Function of paramyxovirus 3′ and 5′ end sequences: In theory and practice. *In* "The Paramyxoviruses" (D. Kingsbury, Ed.), pp. 235–247. Plenum, New York.

Buckland, R., and Wild, F. (1989). Leucine zipper motif extends. *Nature* **338,** 547.

Chambers, P., Pringle, C. R., and Easton, A. J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J. Gen. Virol.* **71,** 3075–3080.

Collins, P. L., and Wertz, G. W. (1983). cDNA cloning and transcriptional mapping of nine polyadenylylated RNAs encoded by the genome of human respiratory syncytial virus. *Proc. Natl. Acad. Sci. USA* **80,** 3208–3212.

Collins, P. L., and Wertz, G. W. (1985). The envelope-associated 22K protein of human respiratory syncytial virus: Nucleotide sequence of the mRNA and a related polytranscript. *J. Virol.* **54,** 65–71.

Collins, P. L., Dickens, L. E., Buckler-White, A., Olmsted, R. A., Spriggs, M. K., Camargo, E., and Coelingh, K. V. W. (1986). Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proc. Natl. Acad. Sci. USA* **83,** 4594–4598.

Collins, P. L., Hill, M. G., and Johnson, P. R. (1990). The two open reading frames of the 22K mRNA of human respiratory syncytial virus: Sequence comparison of antigenic subgroups A and B and expression *in vitro*. *J. Gen. Virol.* **71,** 3015–3020.

Collins, P. L., Hill, M. G., Camargo, E., Grosfeld, H., Chanock, R. M., and Murphy, B. R. (1995). Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5′ proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development. *Proc. Natl. Acad. Sci. USA* **92,** 11563–11567.

Collins, P. L., McIntosh, K., and Chanock, R. M. (1996). Respiratory syncytial virus. *In* "Fields Virology" (B. N. Knipe and P. M. Howley, Eds.), pp. 1313–1351, Lippincott–Raven, Philadelphia.

Cook, J. K. (2000). Avian rhinotracheitis. *Rev. Sci. Tech.* **19,** 602–613.

Cuesta, I., Geng, X., Asenjo, A., and Villanueva, N. (2000). Structural phosphoprotein M2–1 of the human respiratory syncytial virus is an RNA binding protein. *J. Virol.* **74,** 9858–9867.

Curran, J., and Kolakofsky, D. (1999). Replication of paramyxoviruses. *Adv. Virus Res.* **50,** 403–422.

Easton, A. J., and Chambers, P. (1997). Nucleotide sequence of the genes encoding the matrix and small hydrophobic proteins of pneumonia virus of mice. *Virus Res.* **48,** 27–33.

Elango, N., Satake, M., and Venkatesan, S. (1985). mRNA sequence of three respiratory syncytial virus genes encoding two nonstructural proteins and a 22K structural protein. *J. Virol.* **55,** 101–110.

Fearns, R., and Collins, P. L. (1999). Role of the M2–1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *J. Virol.* **73,** 5852–5864.

Felsenstein, J. (1989). PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5,** 164–166.

Giraud, P., Bennejean, G., Guittet, M., and Toquin, D. (1986). Turkey rhinotracheitis in France: Preliminary investigations on a ciliostatic virus. *Vet. Rec.* **119,** 606–607.

González-Reyes, L., Ruiz-Arguello, M. B., Garcia-Barreno, B., Calder, L., Lopez, J. A., Albar, J. P., Skehel, J. J., Wiley, D. C., and Melero, J. A. (2001). Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. *Proc. Natl. Acad. Sci. USA* **98,** 9859–9864.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98.

Hardy, R. W., and Wertz, G. W. (1998). The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J. Virol.* **72,** 520–526.

Horvath, C. M., and Lamb, R. A. (1992). Studies on the fusion peptide of a paramyxovirus fusion glycoprotein: Roles of conserved residues in cell fusion. *J. Virol.* **66,** 2443–2455.

International Committee on Taxonomy of Viruses (2000). Family paramyxoviridae. *In* "Virus Taxonomy, Seventh Report of the International Committee on Taxonomy of Viruses" (M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner, Eds.), pp. 549–561, Academic Press, San Diego.

Jentoft, N. (1990). Why are proteins O-glycosylated? *Trends Biochem. Sci.* **15,** 291–294.

Johnson, P. R., Jr., Olmsted, R. A., Prince, G. A., Murphy, B. R., Alling, D. W., Walsh, E. E., and Collins, P. L. (1987). Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: Evaluation of the contributions of F and G glycoproteins to immunity. *J. Virol.* **61,** 3163–3166.

Juhasz, K., and Easton, A. J. (1994). Extensive sequence variation in the attachment (G) protein gene of avian pneumovirus: Evidence for two distinct subgroups. *J. Gen. Virol.* **75,** 2873–2880.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157,** 105–142.

Lamb, R. A. (1993). Paramyxovirus fusion: A hypothesis for changes. *Virology* **197,** 1–11.

Lamb, R. A., and Kolakofsky, D. (1996). *Paramyxoviridae:* The viruses and their replication. *In* "Fields Virology" (B. N. Knipe and P. M. Howley, Eds.), pp. 1177–1204, Lippincott–Raven, Philadelphia.

Li, J., Ling, R., Randhawa, J. S., Shaw, K., Davis, P. J., Juhasz, K., Pringle, C. R., Easton, A. J., and Cavanagh, D. (1996). Sequence of the nucleocapsid protein gene of subgroup A and B avian pneumoviruses. *Virus Res.* **41,** 185–191.

Ling, R., Easton, A. J., and Pringle, C. R. (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.

Ling, R., Davis, P. J., Qingzhong, Y., Wood, C. M., Pringle, C. R., Cavanagh, D., and Easton, A. J. (1995). Sequence and *in vitro* expression of the phosphoprotein gene of avian pneumovirus. *Virus Res.* **36,** 247–257.

Marriot, A. C., Smith, J. M., and Easton, A. (2001). Fidelity of leader and trailer sequence usage by the respiratory syncytial virus and avian pneumovirus replication complexes. *J. Virol.* **75,** 6265–6272.

Mink, M. A., Stec, D. S., and Collins, P. L. (1991). Nucleotide sequences of the 3′ leader and 5′ trailer regions of human respiratory syncytial virus genomic RNA. *Virology* **185,** 615–624.

Miyahara, K., Kitada, S., Yoshimoto, M., Matsumura, H., Kawano, M., Komada, H., Tsurudome, M., Kusagawa, S., Nishio, M., and Ito, Y. (1992). Molecular evolution of human paramyxoviruses. Nucleotide sequence analysis of the human parainfluenza type 1 virus NP and M protein genes and construction of phylogenetic trees for all the human paramyxoviruses. *Arch. Virol.* **124,** 255–268.

Morrison, T. G. (1988). Structure, function, and intracellular processing of paramyxovirus membrane proteins. *Virus Res.* **10,** 113–135.

Naylor, C. J., Britton, P., and Cavanagh, D. (1998). The ectodomains but not the transmembrane domains of the fusion proteins of subtypes A

and B avian pneumovirus are conserved to a similar extent as those of human respiratory syncytial virus. *J. Gen. Virol.* **79,** 1393–1398.

Plows, D. J., and Pringle, C. R. (1995). Variation in the fusion glycoprotein gene of human respiratory syncytial virus subgroup A. *Virus Genes* **11,** 37–45.

Poch, O., Blumberg, B. M., Bougueleret, L., and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: Theoretical assignment of functional domains. *J. Gen. Virol.* **71,** 1153–1162.

Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8,** 3867–3874.

Randhawa, J. S., Marriott, A. C., Pringle, C. R., and Easton, A. J. (1997). Rescue of synthetic minireplicons establishes the absence of the NS1 and NS2 genes from avian pneumovirus. *J. Virol.* **71,** 9849–9854.

Randhawa, J. S., Wilson, S. D., Tolley, K. P., Cavanagh, D., Pringle, C. R., and Easton, A. J. (1996). Nucleotide sequence of the gene encoding the viral polymerase of avian pneumovirus. *J. Gen. Virol.* **77,** 3047–3051.

Russell, C. J., Jardetzky, T. S., and Lamb, R. A. (2001). Membrane fusion machines of paramyxoviruses: Capture of intermediates of fusion. *EMBO J.* **20,** 4024–4034.

Samal, S. K., and Zamora, M. (1991). Nucleotide sequence analysis of a matrix and small hydrophobic protein dicistronic mRNA of bovine respiratory syncytial virus demonstrates extensive sequence divergence of the small hydrophobic protein from that of human respiratory syncytial virus. *J. Gen. Virol.* **72,** 1715–1720.

Satake, M., and Venkatesan, S. (1984). Nucleotide sequence of the gene encoding respiratory syncytial virus matrix protein. *J. Virol.* **50,** 92–99.

Seal, B. S., Sellers, H. S., and Meinersmann, R. J. (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.

Sedlmeier, R., and Neubert, W. J. (1998). The replicative complex of paramyxoviruses: Structure and function. *Adv. Virus Res.* **50,** 101–139.

Stec, D. S., Hill, M. G., and Collins, P. L. (1991). Sequence analysis of the polymerase L gene of human respiratory syncytial virus and predicted phylogeny of nonsegmented negative-strand viruses. *Virology* **183,** 273–287.

van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A., and Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* **7**(6), 719–724.

Wertz, G. W., Collins, P. L., Huang, Y., Gruber, C., Levine, S., and Ball, L. A. (1985). Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc. Natl. Acad. Sci. USA* **82,** 4075–4079.

Yu, Q., Davis, P. J., Barrett, T., Binns, M. M., Boursnell, M. E., and Cavanagh, D. (1991). Deduced amino acid sequence of the fusion glycoprotein of turkey rhinotracheitis virus has greater identity with that of human respiratory syncytial virus, a pneumovirus, than that of paramyxoviruses and morbilliviruses. *J. Gen. Virol.* **72,** 75–81.

Yu, Q., Davis, P. J., Li, J., and Cavanagh, D. (1992). Cloning and sequencing of the matrix protein (M) gene of turkey rhinotracheitis virus reveal a gene order different from that of respiratory syncytial virus. *Virology* **186,** 426–434.

Zamora, M., and Samal, S. K. (1992). Sequence analysis of M2 mRNA of bovine respiratory syncytial virus obtained from an F-M2 dicistronic mRNA suggests structural homology with that of human respiratory syncytial virus. *J. Gen. Virol.* **73,** 737–741.

Zimmer, G., Budz, L., and Herrler, G. (2001). Proteolytic activation of respiratory syncytial virus fusion protein. Cleavage at two furin consensus sequences. *J. Biol. Chem.* **276,** 31642–31650.