Virus Genes 30:2, 237–249, 2005

Complete Sequence of the RNA Genome of Pneumonia Virus of Mice (PVM)

CHRISTINE D. KREMPL,[1,2] ELAINE W. LAMIRANDE[2] & PETER L. COLLINS[2,*]

[1]*Department of Virology, Institute for Medical Microbiology and Hygiene, Albert-Ludwigs University, Freiburg im Breisgau, Germany*
[2]*Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892-8007*

Received August 19, 2004; Accepted October 4, 2004

Abstract. Pneumonia virus of mice (PVM) is an enveloped RNA-containing virus of Family *Paramyxoviridae*. Sequences had been determined previously for a number of PVM genes, although these represented cloned cDNAs rather than consensus sequences. Sequences were not available for the 3′ -leader and 5′ -trailer regions that constitute the genome termini or for the large polymerase L gene that accounts for 43% of the genome. Also, the available sequences were from an attenuated variant of strain 15, whereas the present study analyzed the version of strain 15 that is available from the American Type Culture Collection (ATCC) and is highly virulent in mice. Analysis of unclosed RT-PCR products yielded a complete consensus sequence of 14,886 nt (GenBank accession number AY729016). Of the regions for which sequences had been previously reported for the non-pathogenic strain, there were 13 nucleotide differences and 10 amino acid differences compared to the present consensus sequence for the virulent isolate. The various genes of PVM shared 29–62% nucleotide sequence identity and 10–60% amino acid sequence identity with human or bovine respiratory syncytial virus (HRSV and BRSV), its closest relatives.

Key words: mouse virus, paramyxovirus, pneumovirus, RNA virus, viral transcription, viral virulence

Introduction

Pneumonia virus of mice (PVM) is classified in Genus *Pneumovirus,* Subfamily *Pneumovirinae,* Family *Paramyxoviridae,* together with the more extensively characterized human and bovine respiratory syncytial viruses (HRSV and BRSV, respectively) [1]. Subfamily *Pneumovirinae* also contains genus *Metapneumovirus,* represented by avian and human metapneumoviruses (AMPV and HMPV). Members of *Paramyxoviridae* are enveloped viruses that have a non-segmented negative-sense RNA genome of approximately 13–18 kilobases. For subfamily *Pneumovirinae,* complete sequences are available for several strains of HRSV (strain A2, 15,222 nt, M74568; strain S2, 15,190 nt, U39662, strain Bl, 15,225 nt, AF013254), for BRSV (strain ATue51908, 15,140 nt, AF092942), for AMPV subgroup C (13134 nt, AY579780), and for several isolates of HMPV (strain CAN97-83, 13,335 nt, AY297749; strain CAN97-75, 13,280 nt, AY297748; and strain NL/1/99, 13,293 nt, AY525843) [2–12].

The genome of PVM appears to encode 10 mRNAs that are translated into 12 major proteins, 11 of which have counterparts in HRSV and BRSV. Nucleocapsid N protein, phosphoprotein P and large polymerase L protein associate with the viral genome to form the nucleocapsid. Matrix M protein is an internal virion component that probably mediates interaction of the nucleocapsid with the virion envelope. Attachment G glycoprotein, fusion F glycoprotein and small hydrophobic SH glycoprotein are transmembrane virion surface proteins. NS1 and NS2 proteins are putative non-structural species that, by analogy to BRSV and HRSV, might be antagonists of the

*Author for all correspondence:
E-mail: collins@niaid.nih.gov.

host interferon system [13–16]. M2-1 and M2-2 proteins are encoded by overlapping open reading frames (ORFs) of the M2 mRNA and, by analogy to HRSV, might be RNA synthesis factors [17–19]. One PVM protein does not have a counterpart in HRSV or BRSV, namely a 137-amino acid protein of unknown function that is encoded by a second ORF in the P mRNA [20].

PVM was first isolated in 1938 when mice were used in an attempt to isolate new human respiratory viruses from patients experiencing acute non-influenzal respiratory tract disease. Mice were inoculated with clinical nasopharyngeal samples and mouse lung homogenates were passaged from animal to animal. Mice frequently developed morbidity and fatal pneumonia following several passages, and PVM was isolated and identified as a potential murine pathogen rather than a human one [21,22]. However, it remains unclear whether PVM is a virulent pathogen of rodents in nature, since reports of naturally occurring infection and disease are rare and limited to immunocompromised animals, and infection of immunocompetent animals may be inapparent or latent [23,24]. PVM also can infect other rodents, and antibodies specific to PVM or a serologically closely related virus have been detected in many rodent species and also in other mammals including humans [25–29]. Indeed, a study in the United Kingdom indicated that up to 80% of the population had serum antibodies that reacted with PVM [29]. This finding, together with the recent identification of HMPV as a previously-unknown human pathogen [30], raises the possibility that a PVM-like human virus might exist. In addition, PVM infection in mice is used as an experimental model for HRSV infection of humans [31].

PVM strain 15 was originally isolated by Horsfall and colleagues as a virulent virus [21,22] and is the only PVM strain available at the American Type Culture Collection (ATCC). Recently, strain 15 was described as being attenuated in BALB/c mice [32,33]. However, we found that strain 15 obtained from the ATCC and maintained as a low-passage stock was highly virulent in BALB/c mice, and was comparable in virulence to the virulent strain, J3666, which has been maintained by mouse-to-mouse passage [34]. This suggests that the reported non-pathogenic virus is a variant that had sustained one or more mutations during passage in cell culture that resulted in attenuation.

Sequences were determined previously for most of the genes of the attenuated version of PVM strain 15, accounting for more than one-half of the genome [20,32,35–39]. This analysis was based on sequencing cloned cDNAs, a method that has considerable potential for error given the high error rate of viral RNA polymerases (which has been estimated to be 10^{-3}–10^{-5} for RNA viruses in general [40] and 10^{-4} for HRSV based on the frequency of monoclonal antibody-resistant mutants [41], as well as the potential for errors during cDNA synthesis and cloning. Also, sequence information was not available for the 3′ (leader) and 5′ (trailer) ends of the genome, which contain the viral promoters, or for the large polymerase gene L, which accounts for 43% of the genome.

It was of interest to determine a complete consensus sequence for the virulent version of PVM strain 15 that is available from the ATCC. Recently, we reported a consensus sequence for the G gene of virulent strain 15 [34]. This sequence differed from that reported previously for the non-pathogenic version by two mutations: one was a nucleotide substitution that resulted in a change in amino acid coding, while the second involved a single-nt difference in length in the G ORF. Specifically, the G gene of the non-pathogenic strain appeared to have sustained a single-nt insertion that placed the normal translational start site out of frame with the remainder of the ORF. Thus, initiation would have to occur at the next-downstream ATG, and would result in a G protein lacking the N-terminal cytoplasmic tail. To further investigate the differences between virulent strain 15 and its non-pathogenic derivative, we determined a complete consensus sequence for the remainder of the genome of virulent strain 15 virus available from the ATCC.

Material and Methods

Cells and Viruses

PVM strain 15 was obtained from the ATCC (ATCC VR-25) and was propagated in BHK-21 cells at 32°C. Cells and supernatants were harvested by scraping, and were sonicated and cell debris was removed by low speed centrifugation

($5000 \times g$ for 15 min). The resulting clarified medium supernatant was used as virus stock and was aliquoted and stored at −80°C. BHK-21 cells (ATCC CCL-10) were maintained in Glasgow MEM (Invitrogen GIBCO) supplemented with 10% fetal bovine serum and 4 mM glutamine (Invitrogen GIBCO).

RNA Isolation

Total cellular RNA from PVM infected BHK-21 cells was isolated using Trizol reagent (Invitrogen) according to the manufacturer's instruction except that an additional phenol–chloroform extraction step was included followed by ethanol precipitation. For isolation of virion-associated RNA, virus particles were purified from the supernatants of infected BHK-21 cells by ultracentrifugation through a 30% w/v sucrose cushion (26,000 rpm in a SW28 rotor for 90 min). The virus pellet was resuspended in Trizol reagent and viral RNA was extracted as described above.

RT-PCR, Cloning, and Nucleotide Sequencing

RNA was mixed with 5 pmol of specific primer and incubated for 10 min at 70°C. Reverse transcription (RT) was performed using Superscript II reverse transcriptase (Invitrogen) at 50°C for 1 h followed by digestion of DNA–RNA hybrids with RNaseH at 37°C for 20 min. PCR was performed with LA Taq polymerase (TaKaRa) at 68°C. 5′ RACE (Rapid Amplification of cDNA Ends) was performed with a commercial kit (version 2.0, Invitrogen) as specified by the manufacturer. 5′ RACE involved adding a poly-C tail to the 3′ end of the RT product with terminal transferase followed by PCR amplification using a nested PVM-specific forward primer and the AAP reverse primer that was supplied by the kit and would prime on the poly-C tail.

For confirmation of the 3′ and 5′ ends of genomic RNA, RNA that was isolated from purified virus was circularized by ligation with 10 U of T4 RNA ligase (Epicentre) in a 40 µl reaction for 3 h at 25°C and 3 h at 37°C. RT was performed using a positive-sense primer that annealed near the end of the L gene, and PCR was performed with a second, nested L primer and a negative-sense NS1-specific primer.

All primer sequences are available upon request from the authors. Degenerative primers that were designed based on conserved polymerase motifs in the L protein (see the text) were: Degen1, TCDATWSWYTGRTTRTCMCC, which annealed at what was to become position 10738 in the complete PVM genome sequence; Degen2, AARTYHAWARAVRYATGRTC, which annealed at position 12615; and Degen3, TTTTTTTTTTTTTTTHDWTAAMTA, which annealed at the L gene-end signal.

PCR products were cloned with a TA cloning vector (pCR2.1 Invitrogen) following the instructions of the manufacturer except that blue–white screening was not used.

Nucleotide sequence analysis was performed using an ABI 3100 sequencer with the Big-Dye terminator ready reaction kit v1.1 (Applied Biosystems).

Results and Discussion

Strategy to Determine a Complete Consensus Sequence for PVM Strain 15

Previous studies had sequenced more than one half of the genome of a non-pathogenic isolate of strain 15 that had been attenuated during passage in cell culture [20,32,35–39]. We used this published sequence as a starting point to determine a complete sequence for virulent strain 15, representing the virus that is available from the ATCC. First, we characterized the PVM gene termini and L gene, for which sequences had not been available.

The L gene was cloned and sequenced by several rounds of genome walking and PCR amplification. First, RT primer extension was performed on intracellular viral RNA using a positive-sense primer specific to the upstream F gene. The resulting cDNA was then subjected to 5′ RACE and was cloned, yielding inserts of approximately 500 bp to more than 2000 bp. Sequence analysis of a number of inserts using vector specific primers yielded a sequence that contained the downstream end of the M2 gene (where the PCR forward primer had hybridized) and the adjacent downstream intergenic region, which had not been previously mapped or sequenced, and 1954 nt of the next downstream gene. This gene was identified as L

based on a high degree of sequence relatedness of its partial predicted protein product with the L proteins of HRSV and other paramyxoviruses (see below). Thus, this showed the PVM M2 and L genes are adjacent with no intervening genes, and are separated by a 9-nt intergenic region.

Three additional rounds of genome walking were performed in which RT was directed by a positive-sense primer that hybridized within known sequence and PCR was performed using a nested forward primer and a degenerate reverse primer (successively, Degen1, 2 and 3, see Materials and Methods) that was designed based on conserved polymerase motifs identified by alignment of several paramyxoviruses (Degen1 and 2) [42] or based on the gend-end signals of the other nine PVM genes (Degen3). This yielded the sequence of the complete L gene except that the last 14 nt, which could not be identified because they represent the primer but which were determined in the course of sequencing the genome termini as described below.

To determine the sequence of the 3′ -leader end of the PVM genome, viral RNA was reverse transcribed using a negative-sense primer specific for the NS2 gene and was amplified by 5′ RACE. The PCR products were cloned and 15 cDNA clones were sequenced, eight of which yielded an identical 43-nt sequence for the leader region; the others represented cloned readthrough mRNA or yielded partial leader sequences that likely represented RNA degradation. To determine the sequence of the 5′ -trailer region (and confirm the sequence of the very end of the L gene), viral RNA was reverse transcribed using a positive-sense primer specific for the downstream end of the L gene, and the resulting cDNA was amplified by 5′ RACE and sequenced directly as uncloned material. This yielded a 91-nt sequence for the trailer region. Because the 5′ RACE procedure involves adding a homopolymer tail on the 3′ end of the RT product prior to PCR amplification, it was not possible to unequivocally identify the terminal nt of either the leader or the trailer by that procedure alone. Therefore, we circularized genomic RNA by ligation and copied and amplified the junction region by RT-PCR. The uncloned product was sequenced directly. The resulting consensus sequence confirmed the sequences of the leader and trailer regions and confirmed the identities of the genome ends. This established that the complete gene order of PVM is 3′-NS1-NS2-N-P-M-SH-G-F-M2-L-5′.

The analysis described above, together with published studies, provided a complete sequence for the PVM genome, but one that was considered provisional because it was based mostly on analysis of cloned cDNAs and represented a composite of the non-pathogenic and virulent isolates of strain 15. To determine a complete consensus sequence, primers were designed from the provisional sequence and used to direct RT-PCR of virulent strain 15, and the RT-PCR products were sequenced directly as uncloned material. This yielded a genome sequence that was 14,886 nt in length, shorter than that of HRSV and BRSV (15,140–15,225 nt), the other members of genus *Pneumovirus*. The sequence described here has been deposited in GenBank under accession number AY729016.

Structures of the PVM Genome and Encoded Proteins

The 3′-leader region of PVM is 43 nt in length, compared to lengths of 41–45 nt for the other members of subfamily *Pneumovirinae* (Fig. la). Compared to HRSV and BRSV, its closest counterparts, the PVM leader shares 100% identity for the first 14 nt and 92% identity for the first 25 nt. For non-segmented negative strand viruses in general, the genome contains a single 3′-proximal promoter. Mutational analysis of an HRSV minigenome showed that the first 11 of the leader region represent a major element of the HRSV promoter involved in transcription and RNA replication, whereas other positions within the first 26 nts were less important for these activities [43]. The exact identity between PVM and HRSV for the first 14 nt suggests that PVM contains the same promoter element.

The 5′-trailer of PVM is 91 nt in length (Fig. 1b), compared to lengths of 40–161 for the other members of *Pneumovirinae*. The PVM trailer is identical to that of HRSV and BRSV for 11 of the terminal 12 nt, after which there was little apparent identity (Fig. 1c). The trailer encodes the 3′ end of the antigenome and thus encodes the antigenome promoter. Alignment of the 3′ ends of the genome and antigenome (Fig. 1d) showed that the two are identical for

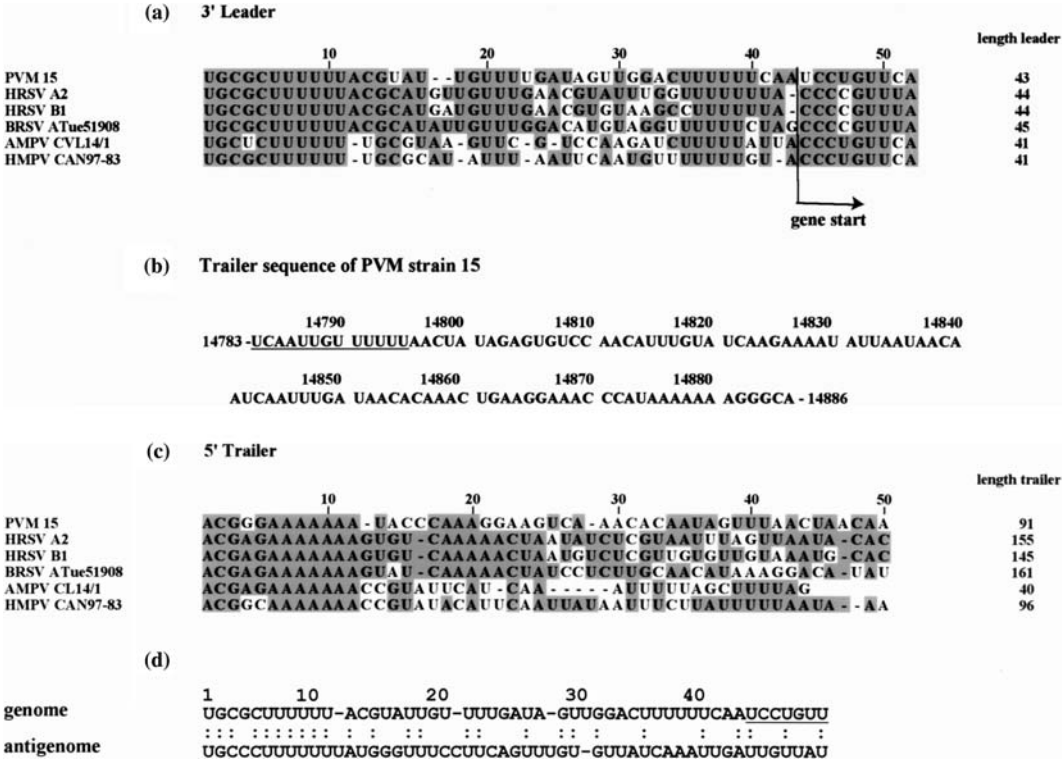

Fig. 1. Sequences of the 3′ and 5′ termini of PVM genomic RNA. (a) Alignment of the 3′-leader region of PVM (negative-sense, 3′-5′) with its counterparts in HRSV strain A2 (representing HRSV antigenic subgroup A), HRSV strain B1 (antigenic subgroup B), BRSV A Tue51908, avian metapneumovirus (AMPV) strain CVL14/1 (representing subgroup A), and human metapneumovirus (HMPV) strain CAN97-83. Gaps were introduced to optimize the alignment, conserved nt are shaded, and the transcription gene-start signal of the leader-proximal gene is indicated. (b) Sequence of the 5′-trailer region of PVM (negative-sense, 3′-5′). The gene-end signal of the L gene is underlined. (c) Alignment of the 5′ -trailer region of PVM (negative-sense, 5′-3′) with its counterparts in the indicated viruses. Gaps were introduced to maximize the alignment. Only the terminal ~50 nt are shown for each trailer because there was no evident sequence relatedness among the remainder of the sequences. (d) Alignment of the 3′ termini of PVM genomic and antigenomic RNA. Nt identities are indicated with dots. The gene start signal of NS1 is underlined.

10 of the first 11 positions, comparable to the degree of sequence identity observed between the 3′ ends of the genome and antigenome for HRSV [4]. This likely identified a promoter element that is conserved between the genomic promoter, which directs both transcription and RNA replication, and the antigenomic promoter, which directs only RNA replication.

A map of the complete PVM genome is shown in Fig. 2a, together with coordinates for the genes and the nucleotide lengths of the genes and extragenic regions and the amino acid lengths of the encoded proteins. Figure 2b shows the sequences of the PVM gene-start motifs that mark the start of each gene, the gene-end motifs that mark the end of each gene, and the intergenic regions that lie between the genes. The generation of the gene map and identification of gene junction sequences shown in Fig. 2 also drew heavily from published work [36], which is summarized here together with the present new data to provide a complete gene map and complete set of signals and intergenic regions. By analogy to HRSV and other paramyxoviruses, the genes are transcribed sequentially in their 3′ to 5′ map order. The gene-start and gene-end motifs are transcription signals that direct, respectively, initiation and termination/polyadenylation of each individual mRNA. By analogy to HRSV, the intergenic regions are thought to be spacers that do not contain any polymerase signals but which are traversed as the polymerase moves from one gene to the next.

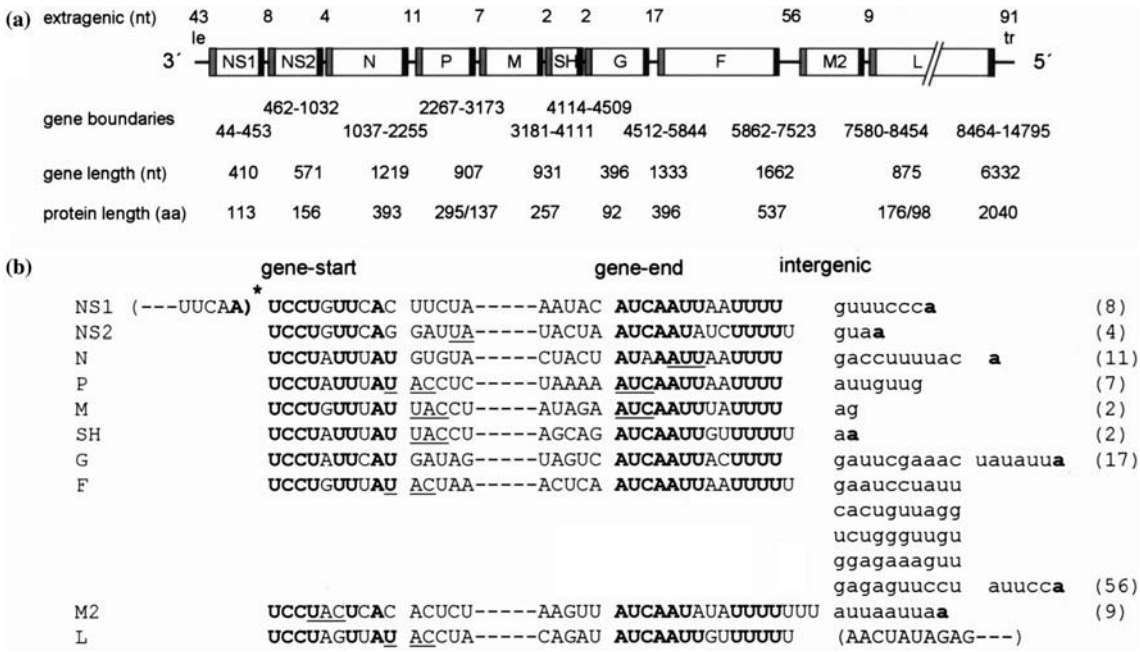

Fig. 2. PVM gene map and transcription signals. (a) A map of the PVM genome is shown (not to scale), drawn 3′ to 5′ as negative-sense RNA such that the direction of transcription is from left to right. Genes are shown as rectangles, with the shaded bars at each end representing the gene-start (lighter shading) and gene-end (darker shading) transcription signals. The horizontal line represents the extragenic regions, specifically the 3′-leader (le), 5′-trailer (tr) and intergenic regions. Nt lengths for the extragenic region are shown above the map; below the map are the gene co-ordinates as well as nt lengths for the encoded mRNAs and amino acid lengths of the predicted proteins. The P mRNA contains overlapping ORFs that have been shown to encode separate proteins, and the M2 mRNA contains two overlapping ORFs that, by analogy to HRSV, also would encode separate proteins. (b) Alignment of the PVM gene-start and gene-end transcription signals and intergenic regions, shown 3′-5′ as negative-sense RNA. Gene-start signals: the 10-nt gene-start signal that marks the beginning of each gene is shown followed by a space and the next 5 nt of each gene; the initiating nt for mRNA synthesis is indicated with an asterisk and residues that are conserved among 7 or more of the 10 genes are in bold. The complements of predicted translational start sites that fall within these sequences are underlined. The sequence in parentheses that precedes the NS1 gene-start site represents the last 5 nt of the leader region; the terminal A residue is in bold because 8 of the 10 gene-start signals are preceded by A. Gene-end signals: the 13-16-nt gene-end signal that marks the end of each gene is shown preceded by a space and the preceding 5 nt of each gene; residues that are conserved among 7 or more of the 10 genes are in bold. The complements of translational termination signals that fall within these sequences are underlined; in the case of M, this termination signal corresponds to a second ORF whose significance is not known. Intergenic regions: the complete intergenic region following each gene is shown in lower case letters; the terminal A residue is in bold because 8 of the 10 gene-start signals are preceded by A. The length of each intergenic region is shown in parentheses. The sequence in parentheses that follows the L gene-end signal represents the adjacent 10 nt of the trailer region.

The PVM gene start signal consists of a 10-nucleotide sequence of which 8 positions are highly conserved among the 10 different PVM genes and the first 4 positions are invariant (3′-UCCU). In addition, 8 of the 10 PVM gene start signals are preceded by an A residue present in the respective upstream leader or intergenic region, as also frequently occurs with HRSV. The initiating nucleotide within the gene start signals for the first 9 genes had been mapped previously by primer extension [36], and the initiating nucleotide of the tenth gene, L, was identified presumptively based on sequence relatedness. The PVM gene-start signal has some similarity to that of HRSV, which consists of the 10-nt sequence 3′-CCCCGUUUAU that is almost exactly conserved among the 9 smaller HRSV genes whereas the HRSV L gene has the variant 3′-CCCUGUUUUA, differences underlined [1]. The gene-start signal of the PVM L gene is not similarly distinct.

The PVM gene-end signal is 13–16 nt in length and consists of the highly conserved heptamer 3′ -AUCAAUU, followed by two non-conserved nucleotides and a poly-U tract of 4–7 residues that

encodes the poly A tail of the mRNA by reiterative copying. The PVM gene-end signal is very similar to that of HRSV, which consists of the pentamer 3′-UCAAU followed in order by 3 or 4 divergent nt and a tract of 4–5 U residues that, when the adjacent divergent nt are U residues, can be in the context of up to 7 uninterrupted U residues [1]. The PVM intergenic regions range from 2 to 56 nt in length and lack obvious conserved elements except that the last downstream nt usually is A, as noted above. Similarly, the intergenic regions of HRSV range in length from 1 to 56 nt, and intergenic regions in other members of subfamily *Pneumovirinae* are up to 190 nt in length [11]. Whereas the M2 and L mRNAs of HRSV and BRSV are transcribed from overlapping genes [2], the M2 and L genes of PVM do not overlap.

The PVM L gene encodes an mRNA of 6332 nt, exclusive of poly A. The predicted L protein is 2040 amino acids in length, compared with amino acid lengths for other pneumoviruses of 2165 (HRSV A2 and Bl), 2159 (BRSV), 2005 (HMPV) and 2004 (AMPV) (3,6,7,11,44). The percentage amino acid sequence identity was 52–53% for PVM L versus HRSV and BRSV, representing the same genus, and 49–50% versus HMPV and AMPV, representing the other genus of subfamily *Pneumovirinae,* and 19.5% for Sendai virus, representing the heterologous *Paramyxovirinae* subfamily (Table 1). It is somewhat curious that PVM L is closer in size to the L proteins of AMPV and HMPV of the heterologous *Metapneumovirus* genus than to the L proteins of the homologous *Pneumovirus* genus, and there is not much difference in the percent amino acid identity. Thus, the L protein sequence alone does not provide unambiguous identification of PVM as being a member of genus *Pneumovirus*.

An alignment of the sequences of the L proteins of PVM and HRSV (Fig. 3) showed that there was extensive sequence relatedness along most of the length of L. We previously prepared an alignment of the L proteins of HRSV and six other nonsegmented negative strand RNA viruses and noted that the HRSV L protein appeared to contain an N-terminal extension of approximately 70 amino acids [3]. Since the HRSV L gene is unique among this group in that it initiates within its upstream neighbor, and since the L coding sequence extends into this overlap, we previously suggested that the apparent N-terminal extension might reflect an ancestral extension of the L gene and its ORF due to the acquisition of a new gene-start signal within its upstream neighbor. However, the present data show that the first 32 amino acids of PVM L are highly conserved (78% identity) with the first 31 amino acids of HRSV, including one stretch in which 17 of 18 residues are identical (Fig. 3). This region is also correspondingly highly conserved in BRSV, HMPV and AMPV (not shown), and thus the L protein of HRSV does not have an apparent N-terminal extension relative to the other pneumoviruses. Since the L gene does not overlap with its upstream neighbor in PVM (Fig. 2) or in HMPV or AMPV, it seems very unlikely that the HRSV L N-terminus had been extended by a shifting gene-start signal. Instead, we suggest that the gene-end signal of the HRSV M2 gene (the upstream neighbor) originally had been upstream of the L gene (as in PVM, HMPV and AMPV), but became silenced due to point mutations, and a new gene-end signal appeared downstream within the L gene. This event presumably occurred after the evolutionary split between HRSV/BRSV (which have the gene overlap) and PVM (which does not), and before the split between HRSV and BRSV.

Examination of the aligned nt sequences (not shown) of PVM, HRSV, BRSV, and HMPV provides some support for the idea that, in the HRSV/BRSV progenitor, an upstream M2 gene-end signal was silenced and a new one appeared downstream. For example, the HRSV and BRSV sequences that align with the PVM M2 gene-end signal have features similar to that of a gene-end signal (for example, the sequence in HRSV is 3′-U<u>A</u>AUUAAGUUU<u>G</u>UU, and, when compared to all of the HRSV gene-end signal, differences are found at only two positions, underlined). Thus, there might have been a functional gene-end signal at this location in the HRSV/BRSV progenitor that was silenced by nt substitutions. In addition, the PVM sequence that corresponds to the present HRSV M2 gene-end signal differs from this motif at only three positions (3′-UC<u>G</u>AU<u>G</u>AAUUU<u>C</u>, differences underlined), and a similar gene-end-like sequence is found in the corresponding position of HMPV. Thus, a (non-functional) gene-end-like sequence might have existed at this position (within the L gene)

Table 1. Amino acid or nucleotide sequence identity (%) between the indicated pneumoviruses or Sendai virus for the proteins and ORFs[a]

Viruses[b] compared	Percent amino acid sequence identity for indicated protein (percent nucleotide sequence identity for indicated ORF)										
	NS1	NS2	N	P	M	SH	G	F	M2-1	M2-2	L
PVM/HRSV	15.8 (47.7)	20.4 (35.2)	60.2 (62.1)	33.1 (40.9)	42.2 (54.4)	22.6 (30.1)	11.6 (32.3)	42.9 (53.6)	42.9 (52.2)	10.1 (38.0)	53.2 (58.5)
PVM/BRSV	14.9 (43.3)	17.8 (32.9)	59.9 (60.8)	34.8 (40.4)	42.2 (53.9)	10.8 (38.0)	11.8 (29.0)	45.0 (55.2)	42.9 (54.2)	12.1 (41.1)	52.4 (60.0)
PVM/HMPV	-	-	45.4 (54.1)	29.7 (45.5)	38.8 (50.3)	19.4 (46.2)	9.3 (23.9)	39.8 (50.6)	37.9 (50.3)	12.1 (35.7)	49.4 (56.9)
PVM/AMPV	-	-	44.2 (51.4)	30.1 (45.6)	38.0 (50.9)	23.7 (43.4)	10.8 (38.5)	40.1 (51.6)	39.5 (49.7)	13.1 (32.7)	50.5 (56.3)
PVM/SeV	-	-	20.1 (41.8)	24.7 (42.5)	17.1 (42.6)	-	-	12.1 (43.9)	-	-	19.5 (43.6)
HRSV/BRSV	67.1 (72.1)	84.0 (75.7)	93.9 (80.5)	80.6 (78.1)	89.1 (79.9)	46.2 (60.0)	26.8 (46.3)	80.0 (75.0)	76.9 (72.6)	41.8 (54.6)	84.1 (77.1)
HRSV/HMPV	-	-	42.3 (53.2)	35.1 (50.4)	37.7 (52.5)	23.1 (50.3)	15.4 (37.5)	33.0 (48.3)	34.4 (46.7)	16.5 (37.4)	44.6 (53.6)
HRSV/AMPV	-	-	68.9 (56.5)	56.6 (60.3)	76.9 (70.8)	16.7 (42.6)	20.0 (47.4)	68.1 (65.9)	73.9 (68.3)	25.0 (53.2)	64.2 (63.5)

[a]Amino acid sequences of the complete predicted proteins or nucleotide sequences of the corresponding ORFs were aligned and sequence identities were calculated using the ClustalW algorithm of the MegAlign tool of the Lasegene/DNASTAR 5.5 siftware (Wisconsin).

[b]sequences of virus strains that were used for the alignments: PVM strain 15, HRSV strain A2 (M74568), BRSV Atue51908 (AF092942), HMPV CAN97-83 (AY297749), AMPV type A (U39295 for N, X58639 fir M, U22110 for P, D00850 fir F, X63408 for M2, U65312 for L, S40185 for SH), Sendai virus strain Z (M30202).

in the HRSV/BRSV progenitor, and might have been rendered functional by nt substitutions at the underlined positions.

As noted above, the HRSV (and BRSV) L proteins were somewhat longer than that of PVM: most of the difference in length occurred in the segments following amino acids 122 and 1598 in the PVM sequence, as illustrated by the gaps introduced in the alignment (Fig. 3). A multiple alignment of all of the available *Pneumovirinae* L protein sequences resulted in similarly located gaps introduced into the AMPV and HMPV sequences (not shown). This is another characteristic by which the L protein of PVM was more closely similar to those of HMPV/AMPV rather than HRSV/BRSV.

Previous sequence alignments of the L proteins of five non-segmented negative strand RNA viruses representing subfamily *Paramyxovirinae* or another family of non-segmented negative strand viruses, *Rhabdoviridae,* identified six regions (called I–VI) that contained segments of conserved residues and thus might represent functional domains [42]. Some (but not all) of these segments also were conserved in PVM and HRSV and are underlined in Fig. 3. In particular, the sequence segments III-A, III-B, III-C and III-D (boxed in Fig. 3) correspond to conserved motifs identified by comparison of a large panel of RNA-dependent RNA polymerases [42,45]. These segments are thought to represent conserved functional elements involved in interaction with the template or in RNA chain elongation.

The percent amino acid and nucleotide sequence identity between individual genes and proteins are shown in Table 1 for PVM versus HRSV and BRSV of *Pneumovirus* and versus AMPV and HMPV of *Metapneumovirus*. PVM was the most closely related to HRSV and BRSV. The N and L proteins were the most highly related (52–60% identity); the P, M, F and M2-1 proteins were intermediate (33–45% identity), and the NSl, NS2, SH, G and M2-2 proteins were the most divergent (<23% identity). Based on amino acid and nucleotide sequence identity, there was more divergence between PVM versus HRSV/BRSV than between either HRSV versus BRSV or HMPV versus AMPV. In addition, Table 1 shows that the percent amino acid identity between PVM and HRSV/BRSV for the various proteins was not

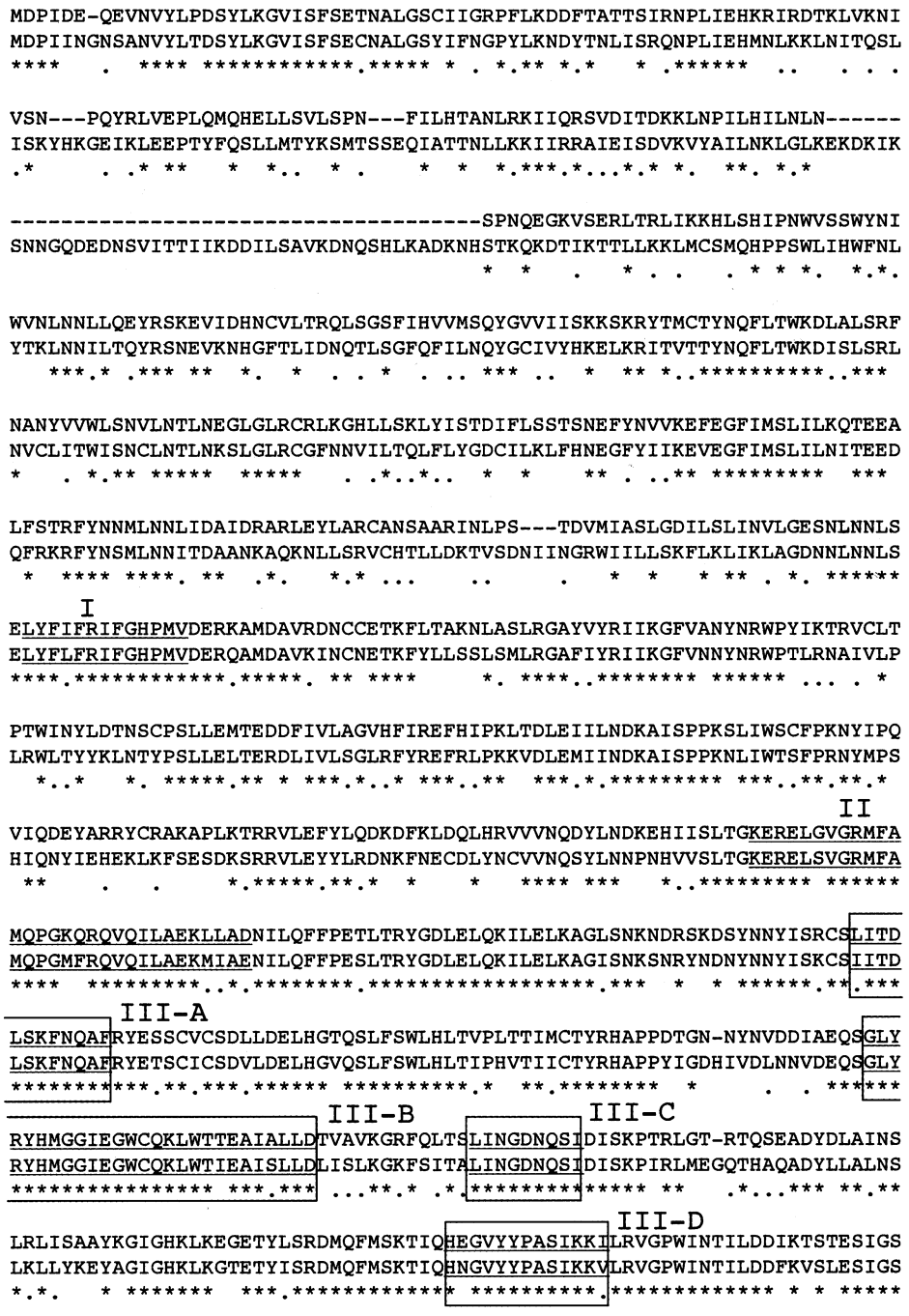

Fig. 3. Alignment of the amino acid sequence of the L protein of PVM (top) and HRSV (bottom) deduced from the nt sequences of the genes. Gaps were introduced to optimize the alignment. Sequence identities are indicated by asterisks and similarity by dots. Underlined sequences labeled I, II, III-A, III-B, III-C, III-D, IV, V and VI refer to sequence segments that also were conserved among five members of subfamily *Paramyxovirinae* and family *Rhabdoviridae* [42]. In particular, segments III-A, III-B, III-C and III-D (boxed) are thought to represent polymerase motifs that are conserved in diverse RNA-dependent RNA polymerases [45].

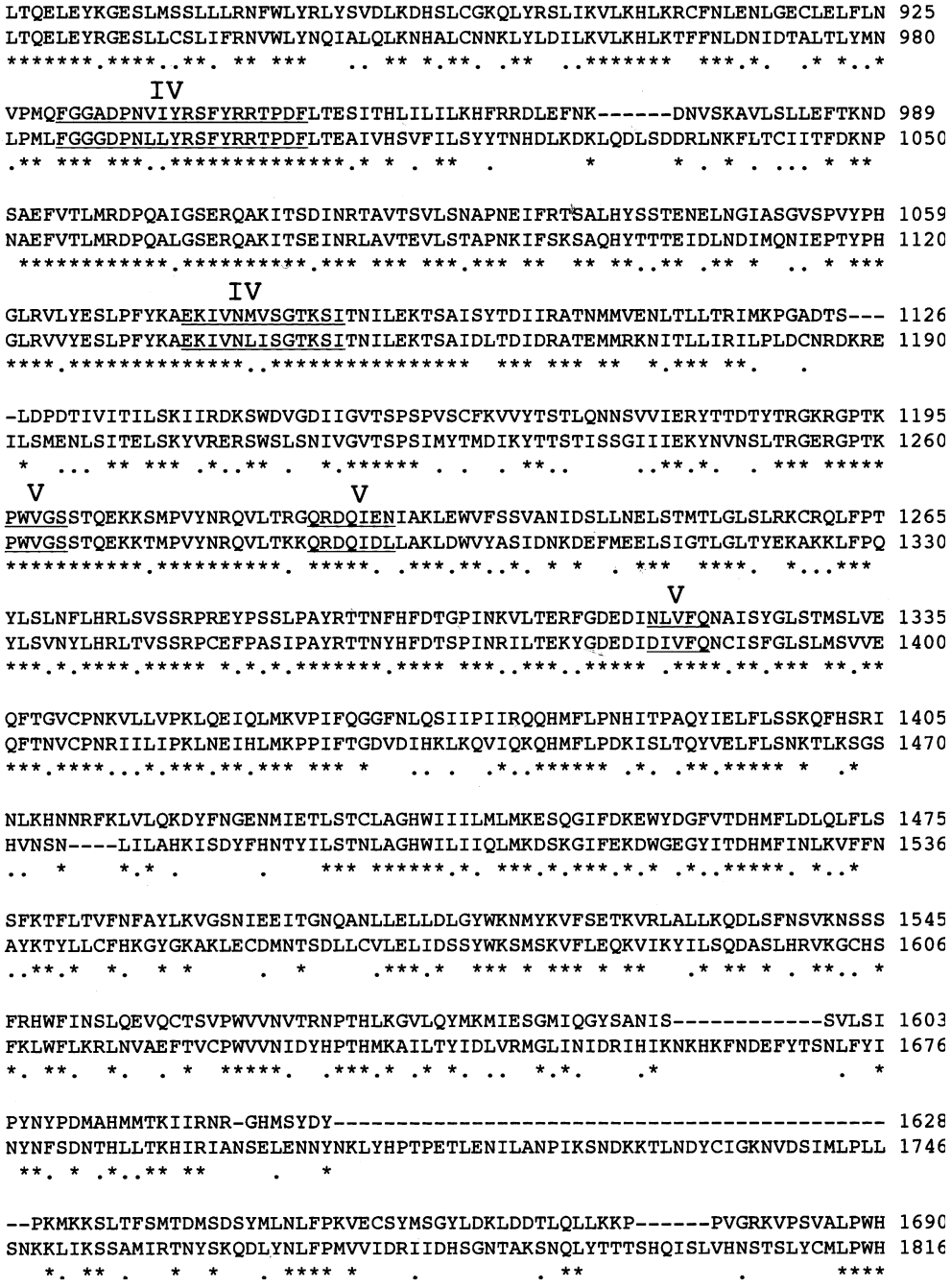

Fig. 3. Continued.

very much higher than between PVM and HMPV/AMPV. Thus, amino acid sequence relatedness alone does not provide a compelling case for classifying PVM with HRSV/BRSV rather than HMPV/AMPV. Rather, the most compelling factor is that the PVM gene map, NS1-NS2-N-P-M-SH-F-G-M2-L, is an exact match with that of HRSV/BRSV (apart from the absence of a gene overlap in PVM) and contains differences from that of HMPV/AMPV, which is N-P-M-F-M2-

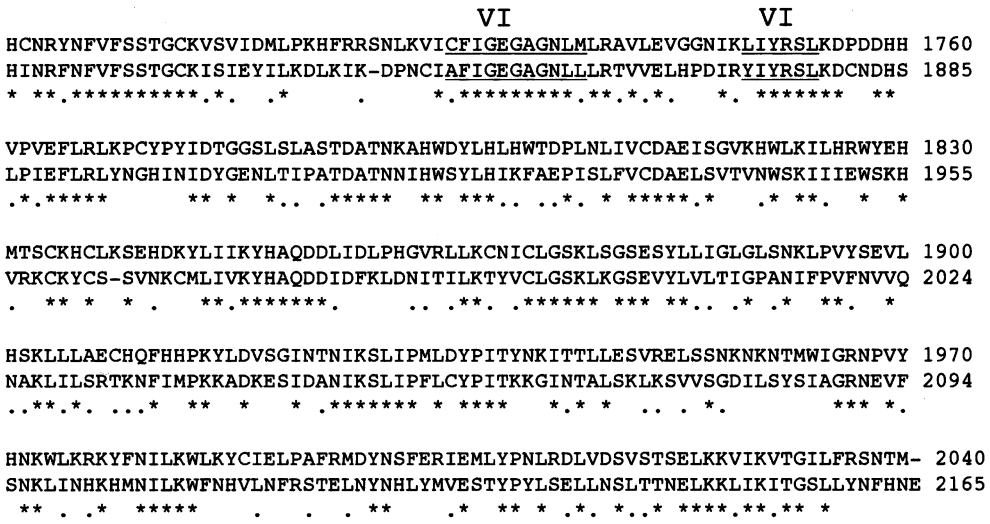

Fig. 3. Continued.

SH-G-L and thus lacks NS1 and NS2 genes and has the F, M2, SH and G genes in the order F-M2-SH-G rather than SH-G-F-M2. Thus, the split between PVM and HRSV/BRSV may be almost as old as that between PVM/HRSV/BRSV and HMPV/AMPV.

Compared to the previously published non-consensus sequences for an attenuated derivative of strain 15, the present sequence of the virulent ATCC virus contained 14 nt differences involving the N, M, SH, G, F and M2-1 genes (Table 2). No differences were observed in the NS1, NS2, P or M2-2 genes. Of the 14 nt differences, 13 were substitutions and resulted in a total of 10 amino acid substitutions involving the N, M, SH, G, F and M2-1 proteins. The remaining mutation involved a difference in nt length in the G ORF, where the previously reported sequence contained a tract of 5 U residues beginning at position 169 in the G gene whereas the virulent strain contained 4 U residues, as recently reported [34]. As noted in the Introduction and described elsewhere [34], the additional U residue in the non-pathogenic strain introduced a frame shift that resulted in the production of an N-terminally truncated G protein. This mutation and any of the others reported in Table 2 are candidates to be involved in the attenuation phenotype of the non-pathogenic version of PVM. The mutations in SH probably are less likely to have an important impact, since HRSV, BRSV and HMPV derivatives that lack the SH gene replicate efficiently *in vitro* and, in the case of HRSV and HMPV, *in vivo* [46,47]. However, whether SH is dispensable for PVM remains to be determined. The attenuating mutation(s) among this panel can be identified once a reverse genetics system is developed for PVM. It might be possible to use such mutations to develop corresponding attenuating mutations in HRSV for vaccine purposes, since attenuating mutations identified in one virus frequently are attenuating when introduced into the corresponding site of a heterologous but related virus [48].

In summary, we have determined a complete consensus sequence for PVM. In particular, this sequence represents virulent strain 15, which is the only strain available from the ATCC, whereas previously-described PVM sequences mostly represent a non-pathogenic variant that was generated during passage in cell culture. The present sequence is a consensus derived from uncloned cDNA, and hence is the majority sequence and provides the basis for developing a reverse genetics system for PVM. PVM is the murine counterpart to human respiratory syncytial virus (HRSV), which is an important respiratory pathogen. HRSV is only semi-permissive in most experimental animals such as mice, and thus investigation of HRSV disease and the evaluation of HRSV vaccines cannot be reliably studied in these experimental systems. Since PVM is virulent in mice and might be a natural mouse pathogen, it

Table 2. Nucleotide and amino acid differences of the virulent strain of PVM with previously available sequence data (genes NS1 to M2) form a non-pathogenic derivative of PVM stain 15

Gene	Protien	Assignment in the present versus published sequence	
		Nucleotide[a] (Position[b])	Amino acid (Position[b])
NS1	nonstructural protein 2	none[c]	none[c]
NS2	nonstructural protein 2	none[c]	none[c]
N	nucleoprotein	U, C[c] (732)	K, R[c] (234)[d]
		C, U[c] (733)	K, R[c] (234)[d]
P	phosphoprotien	none[e]	none[e]
M	matrix protein	G, U[f] (510)	S, Y[f] (167)
		U, C[f] (525)	D, G[f] (172)
		C, U[f] (616)	silent
SH	small hydrophobic protein	A, G[f] (80)	Y, H[f] (24)
		A, G[f] (105)	L, P[f] (32)
G	attachment protein	UUUU[g], UUUU[h] (169–173)	Restores reading frame [g]
		G[g], C[h] (953)	Q[g], E[h] (291)
F	fusion protein	A, U[i] (92)	F, Y[i] (28)
		C, G[i] (669)	silent
		C, G[i] (1416)	L, F[i] (469)
M2	matrix protein 2 ORF1	C, G[j] (171)	M, I[j] (56)
		U, A[j] (416)	H, L[j] (138)
	matrix protein 2 ORF2	none[i]	none[i]

[a]Nucleotide sequences are in negative (genome) sense.
[b]Position relative to the consensus sequence described in the present paper.
[c]Genbank accession no./literature reference: [c]D10331
[d]The two nt chages in N involved the same codon.
[e–j]Genebank accession no./literature reference:, [e]U09649, [f]U66893, [g]ref. 34, [h]D11129, [i]D11129, [i]D11128, [j]ref.32.

can serve as a surrogate for HRSV that can be studied in the convenient and well-defined mouse model [31].

Acknowledgments

We thank Kim C. Tran for excellent technical assistance and Christopher T. Hanson for maintaining the automated sequencing facility.

References

1. Collins P.L., Chanock R.M., and Murphy B.R. in Knipe P.M., Howly P.M. (eds), Fields Virology. Lippincott Williams and Wilkins, Philadelphia, 2001, pp. 1443–1485.
2. Collins P.L., Olmsted R.A., Spriggs M.K., Johnson P.R., and Buckler-White A.J., Proc Natl Acad Sci USA *84*, 5134–5138, 1987.
3. Stec D.S., Hill M.G., III, and Collins P.L., Virology *183*, 273–287, 1991.
4. Mink M.A., Stec D.S., and Collins P.L., Virology *185*, 615–624, 1991.
5. Tolley K.P., Marriott A.C., Simpson A., Plows D.J., Matthews D.A., Longhurst S.J., Evans J.E., Johnson J.L., Cane P.A., Randolph V.B., Easton A.J., and Pringle C.R., Vaccine *14*, 1637–1646, 1996.
6. Karron R.A., Buonagurio D.A., Georgiu A.F., Whitehead S.S., Adamus J.E., Clements-Mann M.L., Harris D.O., Randolph V.B., Udem S.A., Murphy B.R., and Sidhu M.S., Proc Natl Acad Sci USA *94*, 13961–13966, 1997.
7. Buchholz U.J., Finke S., and Conzelmann K.K., J Virol *73*, 251–259, 1999.
8. Seal B.S., Virus Res *58*, 45–52, 1998.
9. Seal B.S., Sellers H.S., and Meinersmann R.J., Virus Res *66*, 139–147, 2000.
10. Alvarez R., Lwamba, H.M., Kapczynski D.R., Njenga, M.K., and Seal B.S., J Clin Microbiol *41*, 1730–1735, 2003.
11. Biacchesi S., Skiadopoulos M.H., Boivin G., Hanson C.T., Murphy B.R., Collins P.L., and Buchholz U.J., Virology *315*, 1–9, 2003.
12. Herfst S., de Graaf M., Schickli J.H., Tang R.S., Kaur J., Yang C.F., Spaete R.R., Haller A.A., van den Hoogen B.G., Osterhaus A.D., and Fouchier R.A., J Virol *78*, 8264–8270, 2004.

13. Schlender J., Bossert B., Buchholz U., and Conzelmann K.K., J Virol *74*, 8234–8242, 2000.
14. Bossert B., and Conzelmann K.K., J Virol *76*, 4287–4293, 2002.
15. Bossert B., Marozin S., and Conzelmann K.K., J Virol *77*, 8661–8668, 2003.
16. Spann K.M., Tran K.C., Chi B., Rabin R.L., and Collins P.L., J Virol *78*, 4363–4369, 2004.
17. Collins P.L., Hill M.G., Cristina J., and Grosfeld H., Proc Natl Acad Sci USA *93*, 81–85, 1996.
18. Bermingham A., and Collins P.L., Proc Natl Acad Sci USA *96*, 11259–11264, 1999.
19. Jin H., Cheng X., Zhou H.Z., Li S., and Seddiqui A., J Virol *74*, 74–82, 2000.
20. Barr J., Chambers P., Harriott P., Pringle C.R., and Easton A.J., J Virol *68*, 5330–5334, 1994.
21. Horsfall F.L., and Hahn R.G., J Exp Med *71*, 391–408, 1940.
22. Horsfall F.L., and Hahn R.G., Proc Soc Exp Biol Med *40*, 684–686, 1939.
23. Richter C.B., Thigpen J.E., Richter C.S., and Mackenzie J.M., Jr., Lab Anim Sci *38*, 255–261, 1988.
24. Weir E.C., Brownstein D.G., Smith A.L., and Johnson E.A., Lab Anim Sci *38*, 133–137, 1988.
25. Eaton M.D., and van Herick W., Proc Soc Exp Biol Med *54*, 89–92, 1944.
26. Horsfall F.L., and Curnen E.C., J Exp Med *83*, 43–64, 1946.
27. Miyata H., Kishikawa M., Kondo H., Kai C., Watanabe Y., Ohsawa K., and Sato H., Exp Anim *44*, 95–104, 1995.
28. Pearson H. E., and Eaton M.D., Proc Soc Exp Biol Med *45*, 677–679, 1940.
29. Pringle C.R., and Eglin R.P., J Gen Virol *67*, 975–982, 1986.
30. van den Hoogen B.G., de Jong J.C., Groen J., Kuiken T., de Groot R., Fouchier R.A., and Osterhaus A.D., Nat Med *7*, 719–724, 2001.
31. Easton A.J., Domachowske J.B., and Rosenberg H.F., Clin Microbiol Rev *17*, 390–412, 2004.
32. Randhawa J.S., Chambers P., Pringle C.R., and Easton A.J., Virology *207*, 240–245, 1995.
33. Domachowske J.B., Bonville C.A., Easton A.J., and Rosenberg H.F., J Infect Dis *186*, 8–14, 2002.
34. Krempl C.D., and Collins P.L., J Virol, *78*, 13362–13365, (2004).
35. Ahmadian G., Chambers P., and Easton A.J., J Gen Viro1 *80*, 2011–2016, 1999.
36. Chambers P., Matthews D.A., Pringle C.R., and Easton A.J., Virus Res *18*, 263–270, 1991.
37. Chambers P., Pringle C.R., and Easton A.J., J Gen Virol *72*, 2545–2549, 1991.
38. Chambers P., Pringle C.R., and Easton A.J., J Gen Virol *73*, 1717–1724, 1992.
39. Easton A.J., and Chambers P., Virus Res *48*, 27–33, 1997.
40. Domingo E., and Holland J.J., Annu Rev Microbiol *51*, 151–178, 1997.
41. Beeler J.A., and van Wyke Coelingh K., J Virol *63*, 2941–2950, 1989.
42. Poch O., Blumberg B.M., Bougueleret L., and Tordo N., J Gen Virol *71*, 1153–1162, 1990.
43. Fearns R., Peeples M.E., and Collins P.L., J Virol *76*, 1663–1672, 2002.
44. Randhawa J.S., Wilson S.D., Tolley K.P., Cavanagh D., Pringle C.R., and Easton A.J., J Gen Virol *77*, 3047–3051, 1996.
45. Poch O., Sauvaget I., Delarue M., and Tordo N., Embo J *8*, 3867–3874, 1989.
46. Bukreyev A., Whitehead S.S., Murphy B.R., and Collins P.L., J Virol *71*, 8973–8982, 1997.
47. Biacchesi S., Skiadopoulos M.H., Yang L., Lamirande E., Tran K.C., Murphy B.R., Collins P.L., and Buchholz U.J., J Virol, *78*, 12877–12887, 2004.
48. Newman J.T., Riggs J.M., Surman S.R., McAuliffe J.M., Mulaikal T.A., Collins P.L., Murphy B.R., and Skiadopoulos M.H., J Virol *78*, 2017–2028, 2004.