Full-length genome sequence and genetic relationship of two paramyxoviruses isolated from bat and pigs in the Americas

L.-F. Wang[1]**, E. Hansson**[1]**, M. Yu**[1]**, K. B. Chua**[2]**, N. Mathe**[1]**, G. Crameri**[1]**,
B. K. Rima**[3]**, J. Moreno-López**[4]**,** and **B. T. Eaton**[1]

[1] CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Vic., Australia
[2] National Public Health Laboratory, Selangor, Malaysia
[3] School of Biomedical Sciences, The Queen's University of Belfast, Belfast, UK
[4] Department of Biomedical Sciences and Veterinary Public Health,
  Swedish University of Agricultural Sciences, Uppsala, Sweden

Summary

Mapuera virus (MPRV) was isolated from a fruit bat in Brazil in 1979, but its host range and disease-causing potential are unknown. Porcine rubulavirus (PoRV) was identified as the aetiological agent of disease outbreaks in pigs in Mexico during early 1980s, but the origin of PoRV remains elusive. In this study, the completed genome sequence of MPRV was determined, and the complete genome sequence of PoRV was assembled from previously published protein-coding genes and the noncoding genome regions determined from this study. Comparison of sequence and genome organization indicated that PoRV is more closely related to MPRV than to any other members of the genus *Rubulavirus*. In the P gene coding region of both viruses, there is an ORF located at the 5′ end of the P gene overlapping with the P protein coding region, similar to the C protein ORF present in most viruses of the subfamily *Paramyxovirinae*, but absent in other known rubulaviruses. Based on these findings, we hypothesise that PoRV may also originate from bats, and spillover events from bats to pigs, either directly or via an intermediate host, were responsible for the sporadic disease outbreaks observed in Mexico.

Introduction

Paramyxoviruses are well-known pathogens of humans and animals. In the last few decades, many novel paramyxoviruses have been isolated from a variety of animal species and their emergence has, on occasion, been accompanied by spectacular outbreaks of disease. Members of the family *Paramyxoviridae* are enveloped viruses possessing a nonsegmented negative-strand genome and are divided into two subfamilies, *Paramyxovirinae* and *Pneumovirinae*. Currently, there are five genera within the subfamily *Paramyxovirinae*: *Rubulavirus*, *Avulavirus*, *Respirovirus*, *Morbillivirus*, and *Henipavirus* [22].

Since 1994, four novel bat-associated paramyxoviruses have emerged, three of which have caused

  Author's address: Dr. Lin-Fa Wang, CSIRO Australian Animal Health Laboratory, PO Bag 24, Geelong, Victoria 3220, Australia. e-mail: linfa.wang@csiro.au

outbreaks of disease in both humans and animals [47]. In September 1994 in Brisbane, Australia, Hendra virus (HeV) caused an outbreak of severe respiratory disease resulting in the death of 13 horses and their trainer [32]. Since then, sporadic HeV outbreaks involving both horse and human infections have been reported [13]. A related virus, Nipah virus (NiV) was isolated and subsequently identified as the cause of an outbreak of encephalitis in humans in Malaysia that began in September 1998 and resulted in 105 fatalities [9]. The outbreak was associated with respiratory illness in pigs, and over one million pigs were culled in its control. Following the discovery of NiV in Malaysia, outbreaks of NiV have been detected in Bangladesh and India between 2001 and 2005 [7, 12, 17]. Serological and virological studies suggest that similar viruses exist in the fruit bat populations in Cambodia, Thailand and Indonesia [36, 38, 45]. In 1997, Menangle virus (MenPV) was isolated from stillborn piglets affected by foetal deformities, at a commercial piggery in New South Wales, Australia [34], and serological evidence indicated that it may have caused influenza-like illness in two humans who had been in close contact with infected piglets [8]. The presence of neutralizing antibodies in several species of fruit bats suggests that they may be the natural host of this virus [34]. In 2000, Tioman virus (TioPV) was isolated from urine collected beneath a fruit bat colony on Tioman Island, Malaysia [10]. Molecular characterization revealed that MenPV and TiPV are closely related novel members of the genus *Rubulavirus* [4, 5, 10, 11].

Mapuera virus (MPRV) was isolated from the salivary glands of an apparently healthy fruit bat (*Sturnira lilium*), suborder *Microchiroptera*, family *Phyllostomidae,* captured in the tropical rainforest of Brazil in 1979 [19]. Its initial placement within the family *Paramyxoviridae* was made on the basis of its morphology and ability to agglutinate guinea pig erythrocytes at 4 °C [50]. The classification of MPRV as a member of the *Paramyxoviridae* was corroborated by a later study in which all of the seven virus-encoded proteins were identified and the gene coding for the nucleocapsid protein (N) was cloned and sequenced [14]. Analysis based on the N gene sequence suggested that MPRV

should be classified as a new member of the genus *Rubulavirus*. To date, it is not known whether MPRV is capable of infecting and causing disease in other animal species.

Originally called La Piedad Michoacan Mexico virus (LPMV) or blue-eye disease virus, porcine rubulavirus (PoRV) was responsible for outbreaks of disease in pigs in Mexico during the early 1980s [31, 39]. Piglets 2–15 days old were the most susceptible, manifesting nervous signs and clinical symptoms such as lethargy and conjunctivitis, and in a small proportion (up to 10%), either unilateral or bilateral corneal opacity (blue eye). In litters farrowed during an outbreak, morbidity and mortality were high (up to 50 and 90%, respectively). Older pigs were also susceptible, but the proportion affected was much smaller, 1–4%, and the mortality was low. First discovered in 1980 on a commercial farm in La Piedad, Michoacan State, the disease spread in subsequent years to other parts of central Mexico [28], but to date, it seems to be confined to this country, with no report of similar disease in other countries. The virus was initially classified as a paramyxovirus on the basis of its morphology and capacity to agglutinate a wide range of red blood cells and spontaneously elute from them at 37 °C. Subsequent sequence analysis revealed significant genetic relatedness with several rubulaviruses, resulting in the creation of the new name porcine rubulavirus. Since the first recognised outbreak in 1980, sporadic outbreaks continued to occur, but the origin of the virus remains unknown [28].

One unique feature of the PoRV genome is the presence of a small open reading frame (ORF) overlapping with the 5′ coding region of the P gene in the second transcription unit [2]. The presence of an overlapping C ORF (coding for a C protein) in this region is a distinguishing feature of most *Paramyxovirinae* members with the exception of rubulaviruses.

In this study, we complete the genome sequences of MPRV and PoRV and examine the coding strategies of the P gene in MPRV to determine whether, like PoRV, a C-like ORF exists and whether a C-like protein is expressed in virus-infected cells. Considering the recent discovery of novel bat paramyxovirus pairs (HeV and NiV, and MenPV and TioPV)

in Australia and Southeast Asia, we were interested to examine whether a similar close relationship could be established for MPRV and PoRV, both of which were isolated on the American continent. Such information may facilitate the search for the natural reservoir host of PoRV and better define risk factors for the prevention of future disease outbreak among pig populations.

Materials and methods

Virus culture and RNA isolation

MPRV (isolate BeAnn 370284) and PoRV (isolate LPMV) were grown in Vero cells as previously described [10]. At 3 days post infection (dpi) the infected cell culture medium was decanted and centrifuged at $1400 \times g$ for 10 min to remove cellular debris. Virus was pelleted in a Beckman type 52.2 rotor at 50,000 rpm for 25 min and total RNA extracted using the RNeasy Mini Kit (Qiagen, Germany) and quantified by the GeneQuant II RNA/DNA Calculator (Pharmacia, USA). When MPRV genomic RNA was prepared for use in the cDNA subtraction protocol, PoRV genomic RNA was prepared concurrently under similar conditions.

Isolation and characterization of viral cDNA using cDNA subtraction

The ClonTech PCR-Select cDNA Subtraction Kit (ClonTech, USA) was used to select virus-specific fragments from viral genomic RNA harvested from tissue culture medium as previously described [5, 10, 18, 29]. Double-stranded cDNA was made using random hexamer oligonucleotide primers and total RNA from pelleted MPRV (tester cDNA) and TioPV (driver cDNA), respectively. Digestion, adaptor ligation, hybridization and PCR reactions were then carried out as described in the instructions provided with the kit. Nested PCR products from both subtractions were size-purified on a 1% agarose gel in three fractions (0.2–0.6, 0.6–1.0 and 1.0–2.0 kb) using the QIA-Quick PCR Gel Extraction Kit (Qiagen). The three purified fractions were cloned, and plasmids with insert sizes 100 bp or greater were randomly selected and sequenced.

Amplification of viral genomic RNA using virus-specific primers

Virus propagation and RNA extraction were carried out as above. Random cDNA was synthesized using the Omniscript RT Kit (Qiagen) with random hexamer primers. Virus-specific primers were designed either from published sequences (for PoRV) or cDNA subtraction-derived sequences (for MPRV), and produced by a commercial provider (GeneWorks, Australia). The Platinum PCR SuperMix

Kit (Invitrogen, USA) was used to perform PCR on the cDNA template synthesized as described above. PCR products were visualized with ethidium bromide on 1–2% agarose gels and purified, using either the QIAquick PCR Purification Kit (Qiagen) or the QIAquick Gel Purification Kit (Qiagen), prior to DNA sequencing.

Characterization of genome termini

The protocol for determining the exact sequences of the 5′ genome and 5′ anti-genome termini was adapted from a published 5′ rapid amplification of cDNA ends (5′ RACE) method [43]. Virus was propagated in and harvested from Vero cell monolayers as previously described [10]. Total RNA was extracted from pelleted virus (containing both genome and anti-genome RNA) using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was carried out using the Thermoscript RT-PCR System kit (Life Technologies, USA) and virus-specific primers located within approximately 100–200 nt of each end of the genome. Reverse transcriptase reactions were incubated at 37 °C for 60 min followed by RT inactivation at 85 °C for 5 min and treatment with Rnase H. The first-strand cDNA was purified using the QIAquick PCR Purification Kit (Qiagen) prior to ligation with a single-strand anchor oligonucleotide (5′-GAAGAGAAGGTGGA AATGGCGTTTTGG, 5′-phosphorylated and 3′-blocked) using T4 RNA ligase (New England BioLabs, USA). The ligated product was amplified by PCR using a virus-specific primer, nested with respect to the first primer used for cDNA synthesis, and a 27-nt primer complementary in sequence to the adaptor. When required, a hemi-nested PCR, using the same adaptor primer and an additional (third) nested virus-specific primer, was also performed. PCR products obtained were gel purified as described earlier and either sequenced directly or cloned before sequencing, in which case at least six individual clones were sequenced to ensure a reliable consensus sequence.

DNA sequencing

Purified PCR products and plasmid DNA were sequenced using the BigDye® Terminator v1.0 Kit (Applied Biosystems, USA) and an ABI PRISM 377 DNA Sequencer (Applied Biosystems). Every nucleotide in the genome was sequenced with a minimum of three-fold redundancy, at least once in each sense and at least once directly from PCR template without cloning.

Sequence analysis

Sequence data were routinely managed and aligned using the Clone Manager and Align Plus programs in the Sci Ed Central software package (Scientific and Educational Software, USA). Sequence similarity searches were conducted using the BLAST service at the National Center for Bio-

technology Information (NCBI). Phylogenetic trees were constructed using the neighbour-joining algorithm with bootstrap values determined by 1000 replicates in the MEGA3 software package [21].

Database accession numbers

The full-length genome sequences of MPRV and PoRV have been deposited into GenBank under the accession EF095490 and BK005918, respectively. Accession numbers for other sequences used in this study are listed below. For viruses where full-length genome sequence was not available, individual protein sequences were used and are indicated by the abbreviated gene letter in parentheses following the accession number. The new naming convention for paramyxovirus abbreviation, as proposed in the 8th ICTV report [22], was used in this paper. Hendra virus (HeV) AF017149; human parainfluenza virus 2 (hPIV2) X57559; human parainfluenza virus 4a (hPIV4a) M32982(N), M55975(P/V), D10241(M), D49821(F), M34033(HN); human parainfluenza virus 4b (hPIV4b) M32983(N), M55976 (P/V) D10242(M), D49822(F), AB006958(HN); measles virus (MeV) AB016162; Menangle virus (MenPV) AF326114; mumps virus (MuV) AB040874; Newcastle disease virus (NDV) AF077761; Sendai virus (SeV) AB005795; simian virus 5 (SV5) AF052755; simian virus 41 (SV41) X64275; Tioman virus (TioPV) AF298895.

Production of rabbit antisera against recombinant C proteins

The coding regions for the C-like proteins in MPRV and PoRV were isolated by PCR using primers with *Asc*I and *Not*I sites incoporated at the 5′ and 3′ ends of the gene fragments, respectively. These fragments were cloned into His$_6$-, GST- and biotin-fusion vectors, which were derived from pRSET (Invitrogen), pGD [49], and pDW363 [44], respectively. Induction of expression and purification of protein by preparative SDS-PAGE were conducted as previously described [49]. Four female New Zealand white rabbits (Institute of Medical and Veterinary Science, Adelaide, Australia) were each inoculated with 100 µg of purified recombinant C protein mixed with an equal volume of montanide ISA 50 adjuvant (Tall-Bennett Group, Australia) in a final volume of 1 ml. Two boosts were performed at two-week intervals. Serum samples were collected via marginal ear vein bleeds before each inoculation. Terminal bleeds were conducted via cardiac puncture under anesthetisation.

Detection of C protein expression in virus-infected cells

Three different assays were used to detect expression of C protein. For the immunofluorescent antibody test, Vero and PK15 cells were seeded on 8-well chamber slide (Nalge Nunc, Germany) and infected at 80% confluency

with MPRV and PoRV, respectively. At 48 hr post-infection, the slides were fixed and processed as described [10]. For Western blot analysis, cells were harvested 48 hr post-infection by PBS washing and trypsin treatment. Cell pellets were lysed in 1% SDS with heating at 75 °C for 15 min, followed by sonication on ice for 1 min using a Microson™ Ultrasonic XL 2000 cell disruptor equipped with a P-3 probe at strength setting 3 (Misonix, USA). Separation of total protein lysate and subsequent blotting and antibody incubation were carried out using standard Western blot procedures as described [49]. For radioimmunoprecipitation, protocols described by Bossart et al. [6] were followed. Briefly, virus infected and mock-infected cells were radiolabelled with 200 µCi/ml Trans^{35}S-label (ICN Biomedicals, USA) for 5 hr at 37 °C. After washing and lysis, immunoprecipitation was carried out using Dynabeads® Protein A magnetic beads (Dynal A.S, Norway) together with different dilutions of various antibodies. At the end of the incubation, the bead-IgG-protein C complexes were washed three times, and elution was achieved by incubating the beads in 30 µl 0.1 M citric acid (pH 2.8). A 20-µl aliquot of 1 M Tris–HCl (pH 8.0) was added to each eluant to neutralise the solution. Subsequent separation of proteins by SDS-PAGE and autoradiography were carried out as described [6].

Results

Determination of full-length genome sequences of MPRV and PoRV

To obtain initial MPRV-specific sequence, cDNA subtraction strategies were used. As shown in Fig. 1, more than ten virus-specific sequence segments were obtained from one subtraction experiment without any prior sequence information required. MPRV-specific oligonucleotide primers were then designed using the sequence data obtained from the subtraction experiments. PCR products were generated to fill in the sequence 'gaps' between cDNA clones, as well as to cover the entire length of the genome, using randomly primed cDNA generated from viral genomic RNA as template. The genome terminal sequences were obtained using a modified 5′ RACE technique as described in Materials and Methods. For PoRV, most of the protein-coding gene sequences have been published previously [1–3, 41, 42]. In this study, we conducted the 5′ RACE to determine the sequences of the genome termini. We also determined the sequences of all the untranslated regions of the genome to compile the final full-length genome sequence.
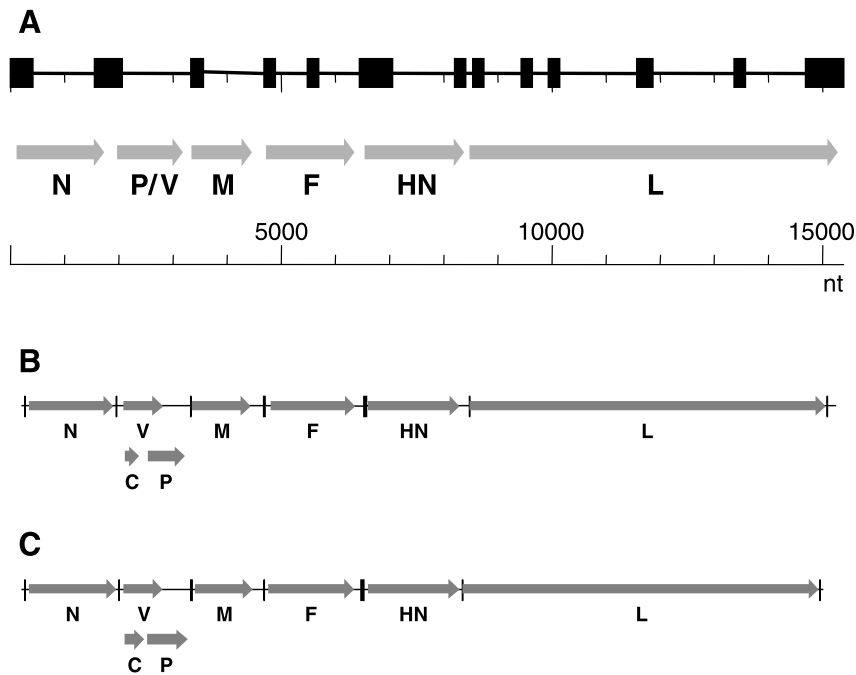
Fig. 1. Schematic diagram of MPRV and PoRV genomes. **A** Diagram illustrating the MPRV genome fragments obtained by cDNA subtraction. The dark solid boxes (top row) represent the cDNA fragments obtained by subtraction and their relative positions to the six major genes (shown below). The genome size scale is given at the bottom. The genome organization and coding strategy for MPRV and PoRV are shown in **B** and **C**, respectively. The vertical lines represent the location and size of IGRs or gene boundaries

Table 1. Molecular features of MPRV and PoRV genes and their deduced gene products

Gene[a]		mRNA features (nt)			IGR (nt)	ORF and deduced protein			
		Total length	5′ UTR	3′ UTR		Coding frame[b]	Size (aa)	MW (kDa)	pI
N	MPRV	1750	78	61	7	+2	537	60.0	5.12
	PoRV	1785	90	60	1	+2	545	60.1	5.12
P/V (V)	MPRV	1400	118	111	3	+2	251	27.5	4.90
	PoRV	1372	85	77	27	+1	249	26.1	5.06
P/V (C)	MPRV	1400-2[c]	–	–	–	+2	92	10.5	10.94
	PoRV	1372-4[c]	–	–	–	+2	126	14.7	10.39
P/V (W)	MPRV	1401	–	–	–	+2/+1	201	22.1	6.25
	PoRV	1373	–	–	–	+1/+3	174	18.1	4.95
P/V (P)	MPRV	1403	–	–	–	+2/+3	391	42.2	5.24
	PoRV	1374	–	–	–	+1/+2	404	42.4	5.33
M	MPRV	1398	38	24	6	+2	371	41.8	9.38
	PoRV	1340	38	195	22	+3	369	41.7	9.72
F	MPRV	1881	108	162	58	+3	537	58.7	6.07
	PoRV	1838	74	141	46	+3	541	58.3	6.15
HN	MPRV	1954	37	171	8	+2	582	64.8	7.84
	PoRV	1862	72	61	3	+2	576	63.3	7.99
L	MPRV	6808	8	41	–	+2	2253	255.6	6.01
	PoRV	6786	8	25	–	+2	2251	255.3	6.31

[a] For the P/V gene of PoRV, there were three different types of mRNA made via RNA editing, each encoding a different protein product, i.e., V (non-edited mRNA), W (+1G) and P (+2G) [8].

[b] Out of the three possible coding frames in the positive-sense sequence (i.e., the antigenome sequence), the one in-frame with the first residue of the antigenome sequence is defined as +1, those in-frame with the second and third residues as +2 and +3 coding frames, respectively.

[c] The putative C protein can be made from any of the 3 possible mRNA species.

Genome features of MPRV and PoRV

The structure of the MPRV genome, comprising six genes in the order: 3′-N-P/V-M-F-HN-L-5′ is identical to that of PoRV (Fig. 1B) and resembles that of most other viruses belonging to the genus *Rubulavirus*. Neither MPRV nor PoRV contains the SH gene present in two known rubulaviruses, MuV and SV5. At 15180 nt in length, the PoRV genome is the smallest in the subfamily *Paramyxovirinae*, and 6 nt smaller than the previously smallest genome of Newcastle disease virus (NDV) at 15186 nt [22]. The size of the MPRV genome, at 15486 nt, is within the range of other rubulavirus genomes. Both viruses have a genome length which is divisible by six, suggesting that these viruses most likely conform to the rule-of-six [20].

The sizes of the encoded MPRV proteins are similar to those of PoRV (Table 1). Approximately 90.5% of the MPRV genome is used for protein coding in comparison to 92.6% for PoRV. Both viruses have a genome coding capacity similar to the average coding percentage (92%) of the subfamily [29].

As with all other members of *Paramyxovirinae*, MPRV and PoRV have a 3′ leader of 55 nt. The 5′ trailer of MPRV is 159 nt, significantly longer than the 43-nt trailer of PoRV. The first and last 11–13 nt

in the genomes of paramyxoviruses are highly conserved and complementary in nature and are critical elements of the genome and anti-genome viral RNA polymerase promoter [23]. This is also true for MPRV (Fig. 2A) and PoRV (Fig. 2B). When the 3′ leader sequences of MPRV and PoRV were aligned with those of other rubulaviruses (Fig. 2C), it was evident that MPRV and PoRV have almost identical genome terminal sequences as that of other known rubulaviruses, and they share significant sequence homology with each other and with MuV.

Transcription of individual genes of *Paramyxoviridae* members is carried out by a stop-and-reinitiation mechanism controlled by conserved sequences at the gene borders [23]. Members of the genera *Respirovirus*, *Morbilllivirus* and *Henipavirus* have a conserved trinucleotide intergenic region (IGR), whereas viruses in the genera *Rubulavirus* and *Avulavirus* have an IGR of variable length and sequence. The gene start, stop and IGR sequences of MPRV and PoRV are listed in Table 2 (shown as cDNA sequence in the antigenome 5′-3′ sense). As expected for all rubulaviruses, the IGRs of MPRV and PoRV vary in size. For MPRV, all but one IGRs start with the C residue, an observation that has been previously made only for TioPV IGRs [11].
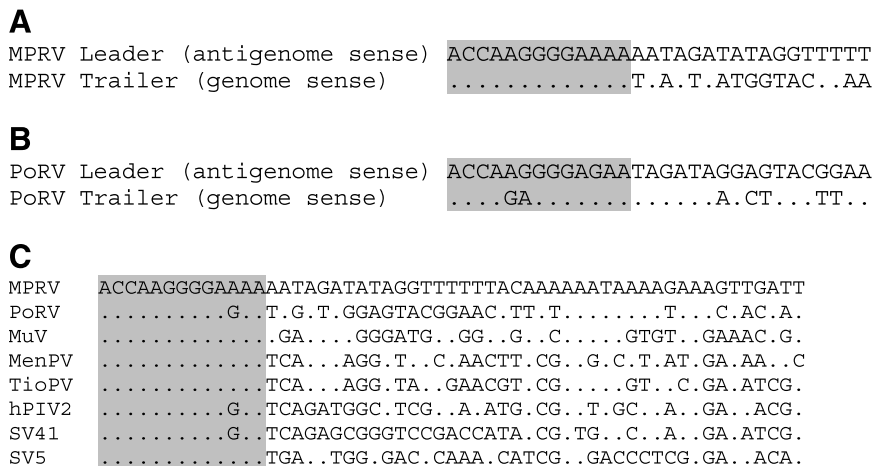
A

```
MPRV Leader (antigenome sense)  ACCAAGGGGAAAAAATAGATATAGGTTTTT
MPRV Trailer (genome sense)     .............T.A.T.ATGGTAC..AA
```

B

```
PoRV Leader (antigenome sense)  ACCAAGGGGAGAATAGATAGGAGTACGGAA
PoRV Trailer (genome sense)     ....GA..............A.CT...TT..
```

C

```
MPRV    ACCAAGGGGAAAAAATAGATATAGGTTTTTTACAAAAAATAAAAGAAAGTTGATT
PoRV    ..........G..T.G.T.GGAGTACGGAAC.TT.T........T...C.AC.A.
MuV     ...............GA....GGGATG..GG..G..C.....GTGT..GAAAC.G.
MenPV   .............TCA...AGG.T..C.AACTT.CG..G.C.T.AT.GA.AA..C
TioPV   .............TCA...AGG.TA..GAACGT.CG.....GT..C.GA.ATCG.
hPIV2   ..........G..TCAGATGGC.TCG..A.ATG.CG..T.GC..A..GA..ACG.
SV41    ..........G..TCAGAGCGGGTCCGACCATA.CG.TG..C..A..GA.ATCG.
SV5     .............TGA..TGG.GAC.CAAA.CATCG..GACCCTCG.GA..ACA.
```

Fig. 2. Alignment of genome terminal sequences. Dots indicate identical residues. See Materials and Methods for full virus names and database accession numbers. **A** Alignment of the 3′ leader and 5′ trailer sequences of MPRV. **B** Alignment of the 3′ leader and 5′ trailer sequences of MPRV. **C** Alignment of 3′ leader sequences of members of the genus *Rubulavirus*. The 3′ leader is represented as anti-genome written in the 5′→3′ sense, and the 5′ trailer is represented as genome written in the 5′→3′ sense. The first 13 nt sequence of each terminus is shaded

Table 2. Sequence and size of intergenic regions (IGR) and transcriptional start and stop signals of MPRV and PoRV

Virus	Genes	Gene stop	IGR	Gene start
MPRV	/N			AGGCCCGAAG
	N/P	TTAAGAAAAAA	CAAAGAT	AGGCCCGAAT
	P/M	TTAAGAAAAAA	CAT	AGGCCCGACT
	M/F	TTTAGAAAAAA	CCATAT	AGGCCCGACT
	F/HN	TTAATAAAAAA	CCGG..50 nt..AAGT	AGGCCCGACT
	HN/L	TTAATAAAAAA	ATCATAGT	AGGCCAGAAT
	L/	TTAAGAAAAAA	CCC..	
PoRV	/N			AGGCCCGGAA
	N/P	TTAAGAAAAAA	T	AGGCCCGAAA
	P/M	TTAAGAAAAAA	TTCA..19 nt..CCGC	AGGCCCGAAC
	M/F	TTTAGAAAAAA	TTTC..14 nt..AGAC	AGGCCCGAAC
	F/HN	TTAAGAAAAAA	CCAT..38 nt..AACT	AGGCCCGAAG
	HN/L	TTAAGAAAAAA	CGT	AGGCCGGAAT
	L/	TTAAGAAAAAA	CTT..	

P gene coding strategy and expression studies

Similar to other rubulaviruses, the non-edited mRNA of the P gene codes for the V protein for both MPRV and PoRV. The P gene RNA editing site for the two viruses is identical and has a stretch of six G residues (TTAAGAGGGGGG). To express the P protein, insertion of two G residues is required to produce the P mRNA. If one G residue is inserted, the mRNA has the capacity to code for a 201-aa W protein for MPRV. For PoRV, the W mRNA contains a stop codon downstream to the RNA editing site, resulting in a much shorter putative W protein of 174 aa (see Table 1 for comparison).

One major difference between rubulaviruses (and avulaviruses) and viruses in the other three genera of the subfamily is the lack of coding capacity for a C protein. The PoRV P gene was unique in that it had an ORF in the 5′ end of the P gene coding for a putative C-like protein of 126 aa [2]. Interestingly, a similar ORF C was also present in the MPRV P gene, albeit with a smaller size of 96 aa. To examine whether the putative C proteins were produced in cells infected with these viruses, rabbit antisera were produced against recombinant C proteins expressed in *E. coli*. Three versions of N-terminally tagged fusion proteins were made for both MPRV and PoRV C proteins, which were His_6-C, GST-C, and biotin-C, respectively. The His_6-C was expressed at the highest level for both viruses (Fig. 3A) and was used as an immunogen for raising antibo-

dies in rabbits. After three injections, the rabbits produced antisera which reacted not only to His_6-C, but also to GST-C or biotin-C (Fig. 3B, C), indicating that C-specific antibodies were successfully produced. However, when MPRV- or PoRV-infected cells were analyzed using these antibodies, no C protein was detected using the rabbit antisera in any of the three assays conducted, immunofluorescent antibody test, Western blot and radio immunoprecipitation (data not shown).

Molecular features of structural proteins

The N proteins of MPRV and PoRV share a 56% sequence identity. The most variable region is located in the C-terminus. The sequence identity increases to 67% when the C-terminal regions are removed from the comparison. The highly conserved *Paramyxovirinae* sequence F-X4-Y-X3-Φ-S-Φ-A-M-G, where X represents any amino acid residue and Φ represents an aromatic amino acid residue, is present within the central domain of the MPRV and PoRV N proteins. The predicted MW and pI of the two proteins are almost identical, as shown in Table 1.

The P and V proteins of MPRV are very similar in size to PoRV and those of other rubulaviruses. At 391 aa and 251 aa, respectively, they are much smaller than the P and V proteins of respiroviruses, morbilliviruses and henipaviruses. The sequence identities between P and V proteins of the two
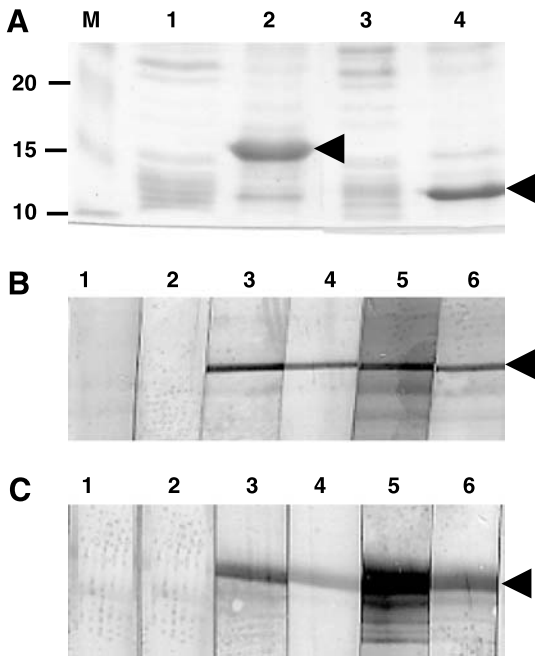
Fig. 3. Production of rabbit antisera against recombinant C proteins of MPRV and PoRV. **A** Expression of recombinant His$_6$-C proteins in *E. coli*. Shown here is a Coomassie-stained gel for PoRV His$_6$-C (*1* and *2*) and MPRV His$_6$-C (*3* and *4*). Lanes *1* and *3* were soluble fractions (*1* and *3*), whereas *2* and *4* contained insoluble fractions. Pre-stained molecular weight markers (Bio-Rad, USA) was loaded in lane M, and the size of each marker band is indicated on the left. **B** and **C** are Western blots of recombinant PoRV GST-C and MPRV Biotin-C proteins reacted with rabbit sera raised against their His$_6$-tagged C protein counterparts. In each panel, strips *1* and *2* were incubated with pre-immune sera, *3* and *4* with second-bleed sera, and *4* and *5* with final-bleed sera. Sera were used at 1:1000 dilution for strips *1*, *3* and *5* and 1:10,000 for *2*, *4* and *6*. Arrowheads in each panel indicate the location of the predicted recombinant fusion protein band

viruses are 38 and 36%, respectively, with the C-terminal regions (the P- or V-specific regions) showing a much higher sequence conservation than the N-terminal part. Both the P and V proteins display a slightly acidic nature. The predicted MW of MPRV and PoRV P proteins is approximately 42 kDa, but their observed relative MW is around 50 kDa [14,

40]. This is true for all P proteins of paramyxoviruses, which have a slower mobility on SDS-PAGE than predicted from their molecular size [48].

The M protein of MPRV is 56% identical to that of PoRV, and the sequence homology was evenly distributed along the whole molecule. Both proteins have a pI greater than 9 and have an almost identical predicted MW at 41–42 kDa.

The F protein of MPRV, like the F proteins of other paramyxoviruses, is predicted to be a type I integral membrane protein. The predicted cleavage site of the MPRV F protein, RRKKR, is highly basic and conforms to the sequence specificity for cleavage by furin, R-X-K/R-R [16]. This sequence is almost identical to the cleavage site of the PoRV F protein, HRKKR. The predicted cleavage site of the MPRV F protein is immediately followed by a highly conserved 25-aa putative hydrophobic fusion peptide as for all known paramyxovirus F proteins. The MPRV F protein contains 7 potential sites for N-linked glycosylation, 2 of which are located within the F$_2$ subunit (N44, N72) before the cleavage site and 5 in the F$_1$ subunit (N181, N351, N426, N456, N518). Five of these sites are conserved in the PoRV F protein, and two sites (N44 and N518) are absent.

Like other members of the *Paramyxovirinae*, the MPRV HN protein is predicted to be a type II integral membrane protein with a hydrophobic domain located at the N-terminal region functioning both as the signal sequence and the transmembrane anchor. The six common neuraminidase active site residues identified for respirovirus and rubulavirus HN proteins [24] are conserved in the MPRV HN protein. Also present in the MPRV HN protein is the hexapeptide NRKSCS, thought to form part of the sialic acid binding site [30]. Five potential N-linked glycosylation sites were found, at N281, N350, N488, N504 and N528, in comparison to four sites in PoRV HN. In contrast to the highly conserved location of the potential N-linked glycosylation sites among the F proteins of the two viruses, there is only

➤

Fig. 4. Phylogenetic trees based on complete protein sequences of six major proteins of all viruses in the genus *Rubulavirus* together with the prototype virus of the other four genera within the subfamily *Paramyxovirinae*. Numbers at nodes indicate levels of bootstrap support calculated from 1000 trees. Branch lengths represent relative genetic distances, and all trees were drawn to the same scale. Full virus names and database accession numbers are provided in Materials and methods

N

P

M

F

HN

L
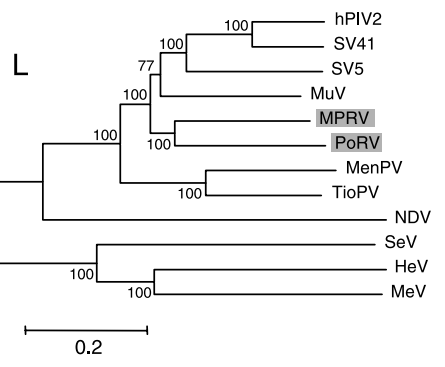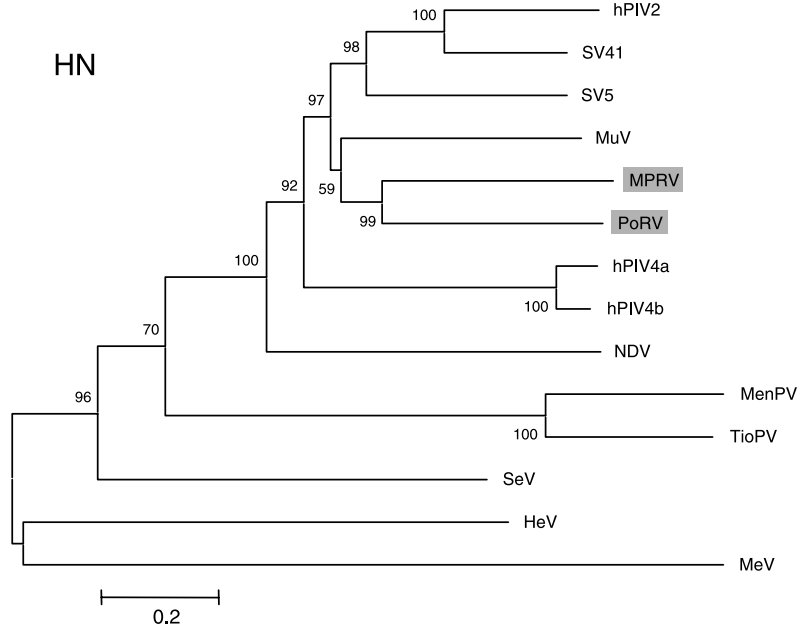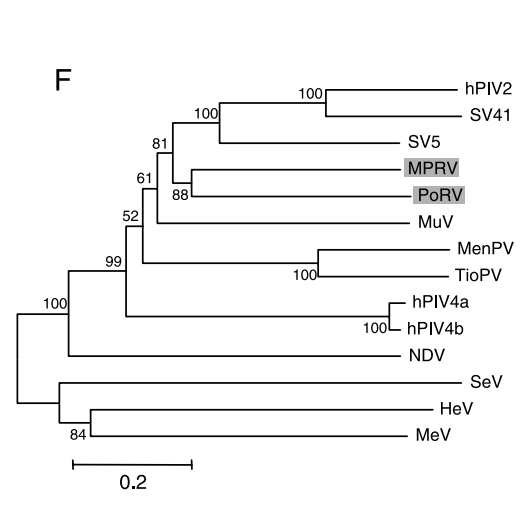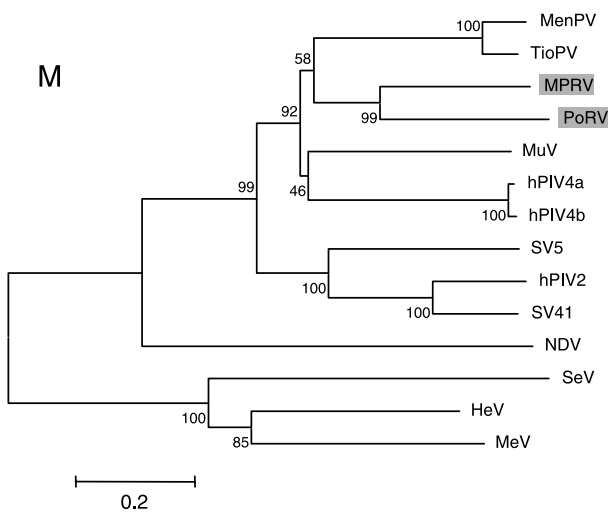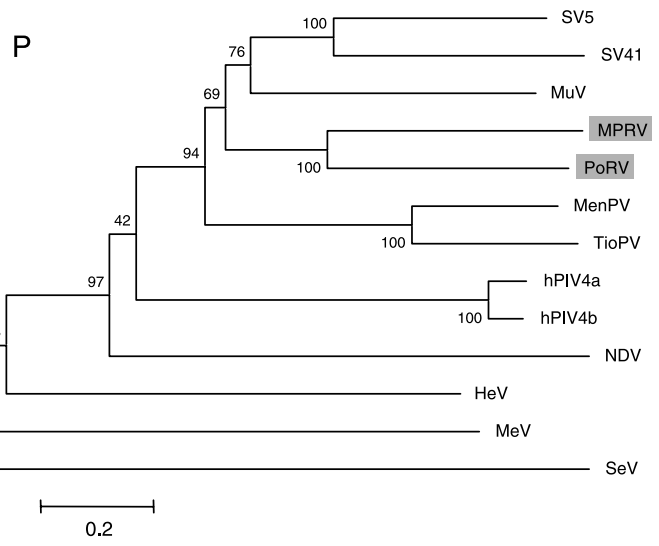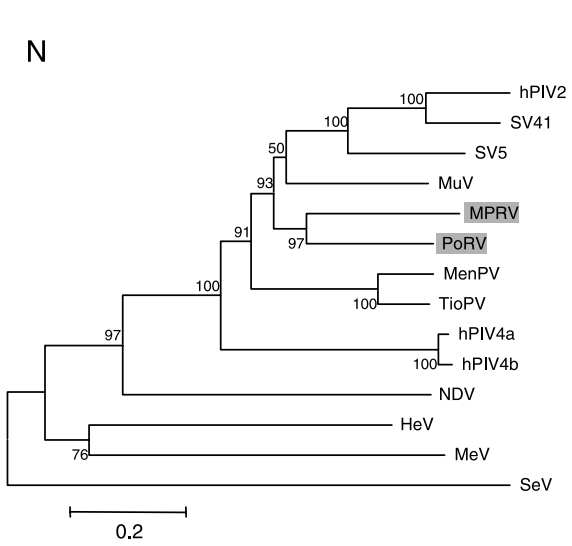
one site conserved among the HN proteins of MPRV and PoRV, N281 for MPRV and N277 for PoRV.

The L proteins of MPRV and PoRV share more than 60% sequence identity. The six strongly conserved linear domains identified within the L proteins of nonsegmented negative-strand (NNS) RNA viruses by Poch et al. [35] can also be identified within these two L proteins. In the most conserved domain III, both L proteins contained the highly conserved GDNQ sequence motif, which is conserved in all known NNS RNA viruses with the exception of HeV, NiV, Mossman virus and Tupaia paramyxovirus [29].

Phylogenetic analysis

Phylogenetic trees were generated by neighbour-joining methods from sequence alignments of full-length N, P, M, F, HN and L proteins of MPRV and PoRV with cognate proteins of other members of the genus *Rubulavirus* and the prototype virus from each of the other four genera in the subfamily *Paramyxovirinae*. All six trees displayed similar topologies (Fig. 4). These trees strongly support the classification of MPRV and PoRV as members of the genus *Rubulavirus*, and clearly indicate their close genetic relationship.

Discussion

Bats, probably the most abundant, diverse and geographically dispersed vertebrates on earth, have recently been shown to be the reservoir hosts of a variety of emerging viruses responsible for severe animal and human disease outbreaks. In addition to the four new bat paramyxoviruses (HeV, NiV, MenPV and TioPV) and MPRV, at least one other paramyxovirus was isolated from fruit bats in the 1970s in India [15, 33]. Recently, we and another group independently identified horseshoe bats (genus *Rhinolophus*) as the reservoir host for a group of genetically diverse SARS-like coronaviruses [25, 27], suggesting that the etiological agent responsible for SARS had a bat origin. The exact origin of Ebola virus remains unknown, but recent serological and molecular studies suggested a potential link between Ebola virus and bats [26].

Although PoRV was initially isolated from pigs during outbreaks in the early 1980s, the origin of the virus remains elusive. The lack of wide-spread and sustained PoRV infection among pig populations mimics the epidemiological pattern of the MenPV outbreak in Australia. Although MenPV was never isolated from bats, there was serological evidence suggesting that the same virus or a closely related virus was circulating in the fruit bat population near the pig farm where the outbreaks occurred [34]. This notion was further supported by the isolation of TioPV from fruit bats in Malaysia, which has the closest genetic relatedness to MenPV among all known rubulaviruses [4, 5, 10, 11].

We were interested to know whether a similar scenario exists between PoRV, a virus of unknown origin capable of causing diseases in pigs in Mexico, and MPRV, isolated from fruit bats in Brazil with unknown disease-causing status in other animal species. As a first step, we embarked on the characterization of the full genomes of both viruses to establish their evolutionary relationship among rubulaviruses.

Data obtained from this study unequivocally indicate that PoRV is more closely related to MPRV than to any other member of the genus *Rubulavirus*. Both viruses have identical genome organizations, highly conserved genome terminal sequences and gene start and stop sequences. The 55-nt 3′ leader sequences of the two viruses were also the most closely related among all rubulaviruses. As shown in Table 1, the major structural proteins of the two viruses have almost identical predicted molecular weights and pI values. The sequence identity between cognate proteins of the two viruses varies from approximately 40% (for P proteins) to 60% (for L proteins). Although the percentage identities were not as high as those between MenPV and TioPV (54–84%), phylogenetic trees generated from all of the six major structural proteins clearly confirmed that PoRV is more closely related to MPRV than to any other member in the genus.

The P gene coding strategy has been used as one of the major molecular features for classification of viruses in the subfamily *Paramyxovirinae* [23, 46]. For example, the lack of the C-protein-coding capacity has been used as a genetic marker

for classification of paramyxoviruses into the genera *Rubulavirus* and *Avulavirus*. In this regard, it was surprising to find that both PoRV and MPRC contain a putative C-protein-coding region in their P genes. Although paramyxovirus C proteins are highly variable in size and sequence, they do share two common molecular characteristics. They are encoded by the P gene in an alternative reading frame overlapping with the N-terminal coding region of the P protein, and most of them are basic proteins with high pI [29]. It is worth noting that the putative C proteins of MPRV and PoRV have a very similar pI and both are basic (Table 1).

In order to determine whether the putative C proteins were expressed in cells infected with PoRV or MPRV, rabbit antisera were generated against His$_6$-C fusion proteins for both viruses. Although we were able to demonstrate that C-specific antibodies were generated in both cases, we were unable to detect C proteins in either PoRV- or MPRV-infected cells using three different assays, including the most sensitive radioimmunoprecipitation method. There are several possibilities to account for these results. First, it is possible that the C proteins in these viruses are expressed only under certain physiological conditions in their hosts, and they were not expressed under the *in vitro* tissue culture conditions used in this study. Second, they might be expressed only in very minute amounts and our reagents and assays were not sensitive enough to detect the expression. Compared with the optimal Kozak consensus sequence (GCCRCCATGG, R = A or G), the predicted Kozak sequences for the C ORF of MPRV (AGATCGATGA) and PoRV (TCTCTGATGG) appeared to be weaker than those of the V/P ORF of MPRV (CGGGCCATGG) and PoRV (GCTGAGATGG). This is different from other paramyxoviruses known to express C proteins, which usually have a stronger Kozak sequence for the C ORF than the P ORF, allowing translational re-initiation from the downstream C ORF start codon. Third, our antibodies were generated against an *E. coli*-derived recombinant fusion protein which was insoluble and purified in the presence of SDS. These antibodies may not be able to recognize the native viral C proteins. Further investigation is required to obtain more conclusive results.

Recently, a parallel study [37], conducted while our genome characterization work was being carried out, found that at least one bat species in Mexico was seropositive to PoRV. In the study, 108 serum samples from 13 different bat species were analysed for antibodies to PoRV and one was positive. The positive sample came from one of three individuals of the species *Rhogeessa parvula major*. Although there was only one positive sample detected in this study, considering that only three bats were sampled for this particular species, the preliminary finding is encouraging. It will be important to conduct follow-up field studies into this and related bat species in the region to better assess the potential of bats as reservoir hosts of PoRV.

In summary, we initiated this study to examine the genetic relationship of MPRV and PoRV and found that these two viruses shared both genomic and protein features which placed them as the two most closely related viruses in the genus *Rubulavirus*. Although this study alone is unable to conclusively indicate the origin of PoRV, it does suggest a possible bat origin of the virus. The recent discovery of a single seropostive bat in Mexico, although inconclusive, is encouraging and will undoubtedly provoke more field investigation in Mexico and neighbouring countries. An alternative approach worth considering would be to conduct experimental infection of *Rhogeessa parvula* and related bat species with PoRV to ascertain the susceptibility of this and related bat species to PoRV.

Acknowledgements

References

1. Berg M, Bergvall AC, Svenda M, Sundqvist A, Moreno-López J, Linné T (1997) Analysis of the fusion protein gene of the porcine rubulavirus LPMV: comparative analysis of paramyxovirus F proteins. Virus Genes 14: 55–61
2. Berg M, Hjertner B, Moreno-López J, Linné T (1992) The P gene of the porcine paramyxovirus LPMV encodes three possible polypeptides P, V and C: the P protein mRNA is edited. J Gen Virol 73: 1195–1200

3. Berg M, Sundqvist A, Moreno-López J, Linné T (1991) Identification of the porcine paramyxovirus LPMV matrix protein gene: comparative sequence analysis with other paramyxoviruses. J Gen Virol 72: 1045–1050

4. Bowden TR, Boyle DB (2005) Completion of the full-length genome sequence of Menangle virus: characterisation of the polymerase gene and genomic 5′ trailer region. Arch Virol 150: 2125–2137

5. Bowden TR, Westenberg M, Wang L-F, Eaton BT, Boyle DB (2001) Molecular characterization of Menangle virus, a novel paramyxovirus which infects pigs, fruit bats, and humans. Virology 283: 358–373

6. Bossart KN, Wang LF, Eaton BT, Broder CC (2001) Functional expression and membrane fusion tropism of the envelope glycoproteins of Hendra virus. Virology 10: 121–135

7. Chadha MS, Comer JA, Lowe L, Rota PA, Rollin PE, Bellini WJ, Ksiazek TG, Mishra A (2006) Nipah virus-associated encephalitis outbreak, Siliguri, India. Emerg Infect Dis 12: 235–240

8. Chant K, Chan R, Smith M, Dwyer DE, Kirkland P (1998) Probable human infection with a newly described virus in the family Paramyxoviridae. Emerg Infect Dis 4: 273–275

9. Chua KB, Bellini WJ, Rota PA, Harcourt BA, Tamin A, Lam SK, Ksiazek TR, Rollin PE, Zaki SR, Shieh WJ, Goldsmith CS, Gubler DJ, Roehrig JT, Eaton B, Gould AR, Olson J, Field H, Daniels P, Ling AE, Peters CJ, Anderson LJ, Mahy WJ (2000) Nipah virus: a recently emergent deadly paramyxovirus. Science 288: 1432–1435

10. Chua KB, Wang L-F, Lam SK, Crameri G, Yu M, Wise T, Boyle D, Hyatt AD, Eaton BT (2001) Tioman virus, a novel paramyxovirus isolated from fruit bats in Malaysia. Virology 283: 215–229

11. Chua KB, Wang L-F, Lam SK, Eaton BT (2002) Full length genome sequence of Tioman virus, a novel paramyxovirus in the genus *Rubulavirus* isolated from fruit bats in Malaysia. Arch Virol 147: 1323–1348

12. Epstein JH, Field HE, Luby S, Pulliam JR, Daszak P (2006) Nipah virus: impact, origins, and causes of emergence. Curr Infect Dis Rep 8: 59–65

13. Field HE, Breed AC, Shield J, Hedlefs RM, Pittard K, Pott B, Summers PM (2007) Epidemiological perspectives on Hendra virus infection in horses and flying foxes. Aust Vet J (in press)

14. Henderson GW, Laird C, Dermott E, Rima BK (1995) Characterization of Mapuera virus: structure, proteins and nucleotide sequence of the gene encoding the nucleocapsid protein. J Gen Virol 76: 2509–2518

15. Hollinger FB, Pavri KM (1971) Bat parainfluenza virus. Immunological, chemical, and physical properties. Am J Trop Med Hyg 20: 131–138

16. Hosaka M, Nagahama M, Kim WS, Watanabe T, Hatsuzawa K, Ikemizu J, Murakami K, Nakayama K (1991) Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. J Biol Chem 266: 12127–12130

17. Hsu VP, Hossain MJ, Parashar UD, Ali MM, Ksiazek TG, Kuzmin I, Niezgoda M, Rupprecht C, Bresee J, Breiman RF (2004) Nipah virus encephalitis reemergence, Bangladesh. Emerg Infect Dis 10: 2082–2087

18. Jack PJM, Boyle DB, Eaton BT, Wang L-F (2005) The complete genome sequence of J virus reveals a unique genome structure in the family Paramyxoviridae. J Virol 79: 10690–10700

19. Karabatsos N (1985) International Catalogue of Arboviruses, 3rd edn. American Society of Tropical Medicine and Hygiene, San Antonio, Texas

20. Kolakofsky D, Pelet T, Garcin D, Hausmann S, Curran J, Roux L (1998) Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. J Virol 72: 891–899

21. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5: 150–163

22. Lamb RA, Collins PL, Kolakofsky D, Melero JA, Nagai Y, Oldstone MBA, Pringle CR, Rima BK (2005) Family *Paramyxoviridae*. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: 8th Report of the International Comittee on Taxonomy of Viruses. Elsevier Acedemic Press, San Diego, pp 655–668

23. Lamb RA, Kolakofsky D (2001) *Paramyxoviridae*: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Lippincott Williams & Wilkins, Philadelphia, pp 305–1340

24. Langedijk JP, Daus FJ, van Oirschot JT (1997) Sequence and structure alignment of Paramyxoviridae attachment proteins and discovery of enzymatic activity for a morbillivirus hemagglutinin. J Virol 71: 6155–6167

25. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci USA 102: 14040–14045

26. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Delicat A, Paweska JT, Gonzalez JP, Swanepoel R (2005) Fruit bats as reservoirs of Ebola virus. Nature 438: 575–576

27. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang L-F (2005) Bats are natural reservoirs of SARS-like coronaviruses. Science 310: 676–679

28. Linné T, Berg M, Bergvall AC, Hjertner B, Moreno-López J (1992) The molecular biology of the porcine paramyxovirus LPMV. Vet Microbiol 33: 263–273

29. Miller PJ, Boyle DB, Eaton BT, Wang L-F (2003) Full-length genome sequence of Mossman virus, a novel paramyxovirus isolated from rodents in Australia. Virology 317: 330–344

30. Mirza AM, Deng R, Iorio RM (1994) Site-directed mutagenesis of a conserved hexapeptide in the paramyxovirus hemagglutinin-neuraminidase glycoprotein: effects on antigenic structure and function. J Virol 68: 5093–5099

31. Moreno-López J, Correa-Giron P, Martinez A, Ericsson A (1986) Characterization of a paramyxovirus isolated from the brain of a piglet in Mexico. Arch Virol 91: 221–231

32. Murray K, Selleck P, Hooper P, Hyatt A, Gould A, Gleeson L, Westbury H, Hiley L, Selvey L, Rodwell B, Ketterer P (1995) A morbillivirus that caused fatal disease in horses and humans. Science 268: 94–97

33. Pavri KM, Singh KR, Hollinger FB (1971) Isolation of a new parainfluenza virus from a frugivorous bat, *Rousettus leschenaulti*, collected at Poona, India. Am J Trop Med Hyg 20: 125–130

34. Philbey AW, Kirkland PD, Ross AD, Davis RJ, Gleeson AB, Love RJ, Daniels PW, Gould AR, Hyatt AD (1998) An apparently new virus (family Paramyxoviridae) infectious for pigs, humans, and fruit bats. Emerg Infect Dis 4: 269–271

35. Poch O, Blumberg BM, Bougueleret L, 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

36. Reynes JM, Counor D, Ong S, Faure C, Seng V, Molia S, Walston J, Georges-Courbot MC, Deubel V, Sarthou JL (2005) Nipah virus in Lyle's flying foxes, Cambodia. Emerg Infect Dis 11: 1042–1047. Trans R Soc Trop Med Hyg. 2004 Oct; 98(10): 577–584

37. Salas-Rojas M, Sánchez-Hernández C, Romero-Almaraz MDL, Schnell GD, Schmid RK, Aguilar-Setién A (2004) Prevalence of rabies and LPM paramyxovirus antibody in non-hematophagous bats captured in the Central Pacific coast of Mexico. Trans R Soc Trop Med Hyg 98: 577–584

38. Sendow I, Field HE, Curran J, Darminto, Morrissy C, Meehan G, Buick T, Daniels P (2006) Henipavirus in Pteropus vampyrus bats, Indonesia. Emerg Infect Dis 12: 711–712

39. Stephano HA, Gay GM, Ramírez TC (1998) Encephalomyelitis, reproductive failure and corneal opacity (blue eye) in pigs, associated with a paramyxovirus infection Vet Rec 122: 6–10. [Published erratum appears in Vet Rec 1988 122: 420]

40. Sundqvist A, Berg M, Hernandez-Jauregui P, Linné T, Moreno-López J (1990) The structural proteins of a porcine paramyxovirus (LPMV). J Gen Virol 71: 609–613

41. Sundqvist A, Berg M, Moreno-López J, Linné T (1992) The haemagglutinin-neuraminidase glycoprotein of the porcine paramyxovirus LPMV: comparison with other paramyxoviruses revealed the closest relationship to simian virus 5 and mumps virus. Arch Virol 122: 331–340

42. Svenda M, Berg M, Moreno-López J, Linné T (1997) Analysis of the large (L) protein gene of the porcine rubulavirus LPMV: identification of possible functional domains. Virus Res 48: 57–70

43. Tillett D, Burns BP, Neilan BA (2000) Optimized rapid amplification of cDNA (RACE) for mapping bacterial mRNA transcripts. Biotechniques 28: 448–456

44. Tsao K-L, DeBarbieri B, Michel H, Waugh DS (1996) A versatile plasmid expression vector for the production of biotinylated proteins by site-specific, enzymatic modification in *Escherichia coli*. Gene 169: 59–64

45. Wacharapluesadee S, Lumlertdacha B, Boongird K, Wanghongsa S, Chanhome L, Rollin P, Stockton P, Rupprecht CE, Ksiazek TG, Hemachudha T (2005) Bat Nipah virus, Thailand. Emerg Infect Dis 11: 1949–1951

46. Wang L-F, Chua KB, Yu M, Eaton BT (2003) Genome diversity of emerging paramyxoviruses. Curr Genom 4: 263–273

47. Wang L-F, Eaton BT (2001) Emerging paramyxoviruses. Infect Dis Rev 3: 52–69

48. Wang L-F, Michalski WP, Yu M, Pritchard LI, Crameri G, Shiell B, Eaton BT (1998) A novel P/V/C gene in a new member of the Paramyxoviridae family, which causes lethal infection in humans, horses, and other animals. J Virol 72: 1482–1490

49. Wang L-F, Yu M, White JR, Eaton BT (1996) BTag: a novel six-residue epitope tag for surveillance and purification of recombinant proteins. Gene 169: 53–58

50. Zeller HG, Karabatsos N, Calisher CH, Digoutte JP, Murphy FA, Shope RE (1989) Electron microscopy and antigenic studies of uncharacterized viruses. I. Evidence suggesting the placement of viruses in families Arenaviridae, Paramyxoviridae, or Poxviridae. Arch Virol 108: 191–209