Cloning of the nucleocapsid protein gene of peste-des-petits-ruminants virus: relationship to other morbilliviruses

A. Diallo,[1]* T. Barrett,[2] M. Barbron,[1] G. Meyer[1] and P. C. Lefèvre[1]

[1] *CIRAD/E.M.V.T., 10 rue Pierre Curie, 94704 Maisons-Alfort Cédex, France and* [2] *AFRC Institute for Animal Health, Pirbright Laboratory, Pirbright, Surrey GU24 ONF, U.K.*

Two independent cDNA clones, identified as representing the mRNA of the nucleocapsid protein gene of peste-des-petits-ruminants virus, were sequenced. The longest insert was 1662 nucleotides, not counting the poly(A) tail, and it was estimated that about 21 nucleotides were missing from the complete gene sequence. The sequence contained one long open reading frame encoding a protein of 525 amino acids with a predicted relative molecular mass of 58008. Comparisons of the nucleic acid and protein sequences of all the morbillivirus nucleoproteins so far determined indicated two major subgroups in the morbillivirus genus of the *Paramyxoviridae*: one group included canine and phocine distemper viruses, and the other rinderpest, measles and peste-des-petits-ruminants viruses. Peste-des-petits-ruminants virus was found to be slightly more related to canine and phocine distemper viruses than were measles and rinderpest viruses.

Peste-des-petits-ruminants (PPR) is an acute and highly contagious virus disease of small ruminants. It is endemic in parts of Africa, the Middle East and also in Southern India (Lefèvre & Diallo, 1990; Shaila *et al.*, 1989). Clinically it resembles rinderpest in cattle and is characterized by pyrexia, ocular and nasal discharges, erosive stomatitis and diarrhoea. Like rinderpest, the causative agent of PPR is a virus belonging to the *Morbillivirus* genus in the family *Paramyxoviridae*. The other members of the genus include measles virus (MV), canine (CDV) and phocine (PDV) distemper viruses and new morbilliviruses isolated from dolphins and porpoises (Barrett *et al.*, 1993; Visser *et al.*, 1993).

The morbilliviruses are enveloped pleomorphic particles which contain six structural proteins: the nucleocapsid protein (N), the phosphoprotein, the polymerase or large protein, the matrix protein, the fusion protein and the haemagglutinin protein (Diallo, 1990). The genome consists of single-stranded RNA of negative polarity. Only the MV and CDV genomes have been entirely sequenced; they are 15892 and 15616 nucleotides long, respectively (Crowley *et al.*, 1988; Sidhu *et al.*, 1993). They are divided into six transcription units separated from each other by a conserved trinucleotide sequence (Sidhu *et al.*, 1993). This genome organization appears to be the same for other morbilliviruses (Curran *et al.*, 1991) and is likely to be the same in the case of PPR virus (PPRV) since, in previous reports, we have shown similar electrophoretic migration profiles of proteins and mRNAs synthesized in cells infected with different viruses from this group (Barrett & Underwood, 1985; Diallo *et al.*, 1987). Here we report for the first time the nucleotide sequence of a cDNA derived from the mRNA for a PPRV protein: the N protein.

Poly(A)$^+$ RNA was isolated from Vero cells infected with the PPRV vaccine strain (Diallo *et al.*, 1989*a*). A cDNA library was made from this RNA in a pUC13 plasmid vector as previously described (Diallo *et al.*, 1989*b*). Virus-specific clones were identified using a probe made from ^{32}P-labelled infected cell mRNA mixed with a 10-fold excess of unlabelled uninfected cell mRNA. Plasmid DNAs were purified from positive clones and were used to select ^{32}P-labelled virus-induced mRNA according to the technique described by Barrett & Mahy (1984). Clone B2-6 selected a labelled message corresponding in size to the nucleoprotein mRNA of PPRV (Barrett & Underwood, 1985). Partial sequence analysis of the clone confirmed its homology with the N mRNA of other morbilliviruses but the AUG initiator codon sequence was missing. In order to obtain a full-length N gene clone, another cDNA library was made in plasmid pT7T3D (Directional Cloning kit, Pharmacia) and screened using the cDNA insert of plasmid B2-6 as a probe. Four positive clones were selected but only one clone (NB71) was analysed further since partial sequence analysis had shown that it was the only one containing

The nucleotide sequence data reported in this paper have been assigned the accession no. X74443 in the EMBL database.

```
   1                                   M  A  T  L  L  K  S  L  A  L    20
   1  GGGGAGAGGAGGAGGAGCAAGATCTTTGACCATGGCTACTCTCCTTAAAAGCTTGGCATT    60

  21  F  K  R  N  K  D  K  A  P  T  A  S  G  S  G  G  A  I  R  G    40
  61  GTTCAAGAGGAACAAAGACAAAGCGCCTACTGCGTCGGGTTCAGGAGGGGCCATCCGGGG   120

  41  I  K  N  V  I  I  V  P  I  P  G  D  S  S  I  I  T  R  S  R    60
 121  GATTAAGAATGTTATCATAGTCCCCATTCCCGGGGACTCATCCATCATTACCCGTTCAAG   180

  61  L  L  D  R  L  V  R  L  A  G  D  P  D  I  N  G  S  K  L  T    80
 181  ACTGCTCGACAGGCTTGTCAGATTGGCCGGAGATCCTGACATCAACGGGTCAAAGCTGAC   240

  81  G  V  M  I  S  M  L  S  L  F  V  E  S  P  G  Q  L  I  Q  R   100
 241  CGGCGTGATGATCAGCATGTTATCTTTGTTCGTGGAGTCACCCGGGCAATTGATACAGCG   300

 101  I  T  D  D  P  D  V  S  I  R  L  V  E  V  V  Q  S  T  R  S   120
 301  GATCACAGATGATCCAGATGTTAGCATCCGCCTTGTTGAGGTAGTTCAAAGTACTAGGTC   360

 121  Q  S  G  L  T  F  A  S  R  G  A  D  L  D  N  E  A  D  M  Y   140
 361  CCAGTCCGGGTTGACCTTTGCATCACGTGGTGCTGATTTGGACAATGAGGCAGATATGTA   420

 141  F  S  T  E  G  P  S  S  G  S  K  K  R  I  N  W  F  E  N  R   160
 421  TTTTTCAACTGAAGGACCCTCGAGTGGAAGTAAGAAAAGGATCAACTGGTTTGAGAACAG   480

 161  E  I  I  D  I  E  V  Q  D  A  E  E  F  N  M  L  L  A  S  I   180
 481  AGAAATAATAGACATAGAGGTGCAAGATGCAGAAGAGTTCAATATGTTGTTAGCCTCCAT   540

 181  L  A  Q  V  W  I  L  L  A  K  A  V  T  A  P  D  T  A  A  D   200
 541  CTTAGCACAAGTTTGGATCCTCCTGGCCAAGGCGGTTACGGCACCGGATACGGCAGCTGA   600

 201  S  E  L  R  R  W  V  K  Y  T  Q  Q  R  R  V  I  G  E  F  R   220
 601  CTCAGAACTGAGAAGGTGGGTTAAATACACACAACAAAGGAGAGTGATTGGGGAATTTCG   660

 221  L  D  K  G  W  L  D  A  V  R  N  R  I  A  E  D  L  S  L  R   240
 661  CCTTGACAAAGGGTGGCTGGACGCAGTCCGCAACAGGATTGCAGAAGATCTATCACTTCG   720

 241  R  F  M  V  S  L  I  L  D  I  K  R  T  P  G  N  K  P  R  I   260
 721  GCGGTTCATGGTATCTCTCATACTTGACATCAAAAGGACCCCCGGCAACAAGCCAAGGAT   780

 261  A- E  M  I  C  D  I  D  N  Y  I  V  E  A  G  L  A  S  F  I   280
 781  TGCAGAAATGATCTGCGACATTGACAACTATATTGTAGAAGCCGGACTCGCCAGTTTCAT   840

 281  L  T  I  K  F  G  I  E  T  M  Y  P  A  L  G  L  H  E  F  A   300
 841  TCTTACTATCAAATTTGGTATTGAAACCATGTATCCTGCATTAGGCCTTCACGAGTTCGC   900

 301  G  E  L  S  T  I  E  S  L  M  N  L  Y  Q  Q  L  G  E  V  A   320
 901  CGGGGAATTGTCCACTATTGAATCCTTGATGAACTTGTATCAACAGCTAGGAGAGGTTGC   960

 321  P  Y  M  V  I  L  E  N  S  I  Q  N  K  F  S  A  G  A  Y  P   340
 961  ACCCTACATGGTGATTCTAGAGAACTCAATTCAGAACAAGTTTAGTGCAGGAGCCTATCC  1020

 341  L  L  W  S  Y  A  M  G  V  G  V  E  L  E  N  S  M  G  G  L   360
1021  TCTCCTCTGGAGCTATGCGATGGGTGTCGGAGTCGAGTTGGAGAACTCAATGGGGGGCCT  1080

 361  N  F  G  R  S  Y  F  D  P  A  Y  F  R  L  G  Q  E  M  V  R   380
1081  GAACTTTGGCAGGTCATATTTTGACCCGGCCTATTTCCGTCTCGGACAGGAGATGGTCAG  1140

 381  R  S  A  G  K  V  S  S  V  I  A  A  E  L  G  I  T  A  E  E   400
1141  AAGATCTGCAGGAAAGGTCAGCTCTGTAATCGCGGCTGAGCTTGGTATCACAGCAGAGGA  1200

 401  A  K  L  V  S  E  I  A  S  Q  T  G  D  E  R  T  V  R  G  T   420
1201  AGCCAAACTAGTCTCGGAAATCGCCTCACAGACTGGGGATGAACGAACCGTTAGAGGGAC  1260

 421  G  P  R  Q  A  Q  V  S  F  L  Q  H  K  T  D  E  G  E  S  P   440
1261  TGGGCCTCGACAGGCGCAGGTCTCCTTCCTCCAGCATAAAACAGATGAGGGAGAGTCGCC  1320

 441  T  P  A  T  R  E  E  V  K  A  A  I  P  N  G  S  E  G  R  D   460
1321  TACACCAGCGACCAGAGAAGAAGTCAAAGCTGCGATCCCAAATGGGTCCGAAGGAAGGGA  1380

 461  T  K  R  T  R  S  G  K  P  R  G  E  T  P  G  Q  L  L  P  E   480
1381  CACAAAGCGAACACGCTCAGGAAAGCCCAGAGGAGAAACTCCCGGCCAACTGCTTCCGGA  1440

 481  I  M  Q  E  D  E  L  S  R  E  S  S  Q  N  P  R  E  A  Q  R   500
1441  GATCATGCAAGAGGATGAACTCTCGCGAGAGTCTAGTCAAAACCCTCGTGAGGCTCAAAG  1500

 501  S  A  E  A  L  F  R  L  Q  A  M  A  K  I  L  E  D  Q  E  E   520
1501  ATCGGCTGAGGCACTCTTCAGGCTGCAGGCCATGGCCAAGATTCTGGAGGACCAGGAGGA  1560

 521  G  E  D  N  S  Q  I  Y  N  D  K  D  L  L  S                   540
1561  GGGAGAAGACAACAGTCAGATCTACAACGACAAGGATCTCCTCAGCTGAGCAGACGCACC  1620

1621  CTCCGTCCAAATCAGTGACAAGACATCGCCCGCCAGTATTAT  1662
```

Fig. 1. Nucleotide sequence of the PPRV N gene. The nucleotide sequence of clone NB71 is presented in the message sense along with the deduced amino acid sequence in the single-letter code.

an AUG codon in a position corresponding to those of the other morbillivirus N genes. Both B2-6 and NB71 were subcloned into plasmid pBluescript KS$^+$ and were submitted to unidirectional deletion using exonuclease III and mung bean nuclease (pBluescript Exo/Mung DNA sequencing system, Stratagene) to obtain overlapping deletions for sequence analysis of the double-stranded DNA. Sequencing was carried out using the dideoxynucleotide chain termination method described for double-stranded DNA using T7 DNA polymerase (T7 sequencing kit, Pharmacia) with the M13 forward and reverse primers (Murphy & Kavanagh, 1988). Some specific primers, based on the B2-6 sequence, were also

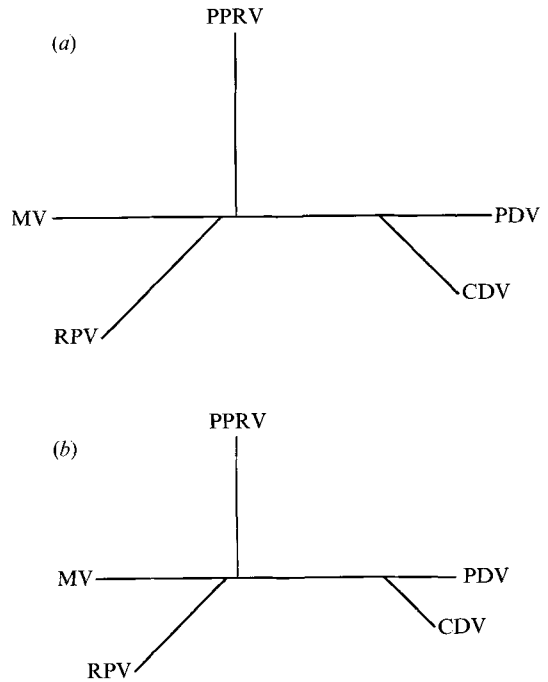

Fig. 2. Phylogenetic analysis of the morbillivirus N protein nucleotide (*a*) and amino acid (*b*) sequences. Sequences were aligned using the PILEUP program (Genetics Computer Group, 1991) and analysed using the PHYLIP programs DNADIST and PROTDIST. The distance matrices were used to generate the phylogenetic trees using the KITSCH program (Felsenstein, 1989). The branch lengths are proportional to the estimated mutational distances between the sequences and the hypothetical common ancestors that existed at the nodes in the trees.

synthesized (Genset) and used to sequence part of clone NB71.

The sequence of the PPRV N gene, excluding the poly(A) tail, is shown in the message sense together with the deduced amino acid sequence in Fig. 1. The sequences of four other morbillivirus N protein genes (MV, CDV, RPV and PDV) have been published (Cattaneo *et al.*, 1988; Kamata *et al.*, 1991; Blixenkrone-Möller *et al.*, 1992; Sidhu *et al.*, 1993). The N genes of RPV and MV are both 1683 nucleotides long and those of CDV and PDV are 1677 nucleotides long. Alignment of the N gene sequences of the five morbilliviruses (data not shown) revealed that the insert of clone NB71, the longest cDNA clone of PPRV N we isolated, started at position 22. Therefore, it can be assumed that about 21 nucleotides of the complete gene are missing. The overall nucleotide identity of the PPRV N gene with those of RPV, MV, CDV and PDV was 68·8%, 66·3%, 64·5% and 64·2%, respectively. Phylogenetic analysis using the PHYLIP DNADIST and PROTDIST programs (Felsenstein, 1989) showed that the morbilliviruses formed two distinct subgroups: CDV and PDV in one and RPV and MV in another, as was noted in previous investigations

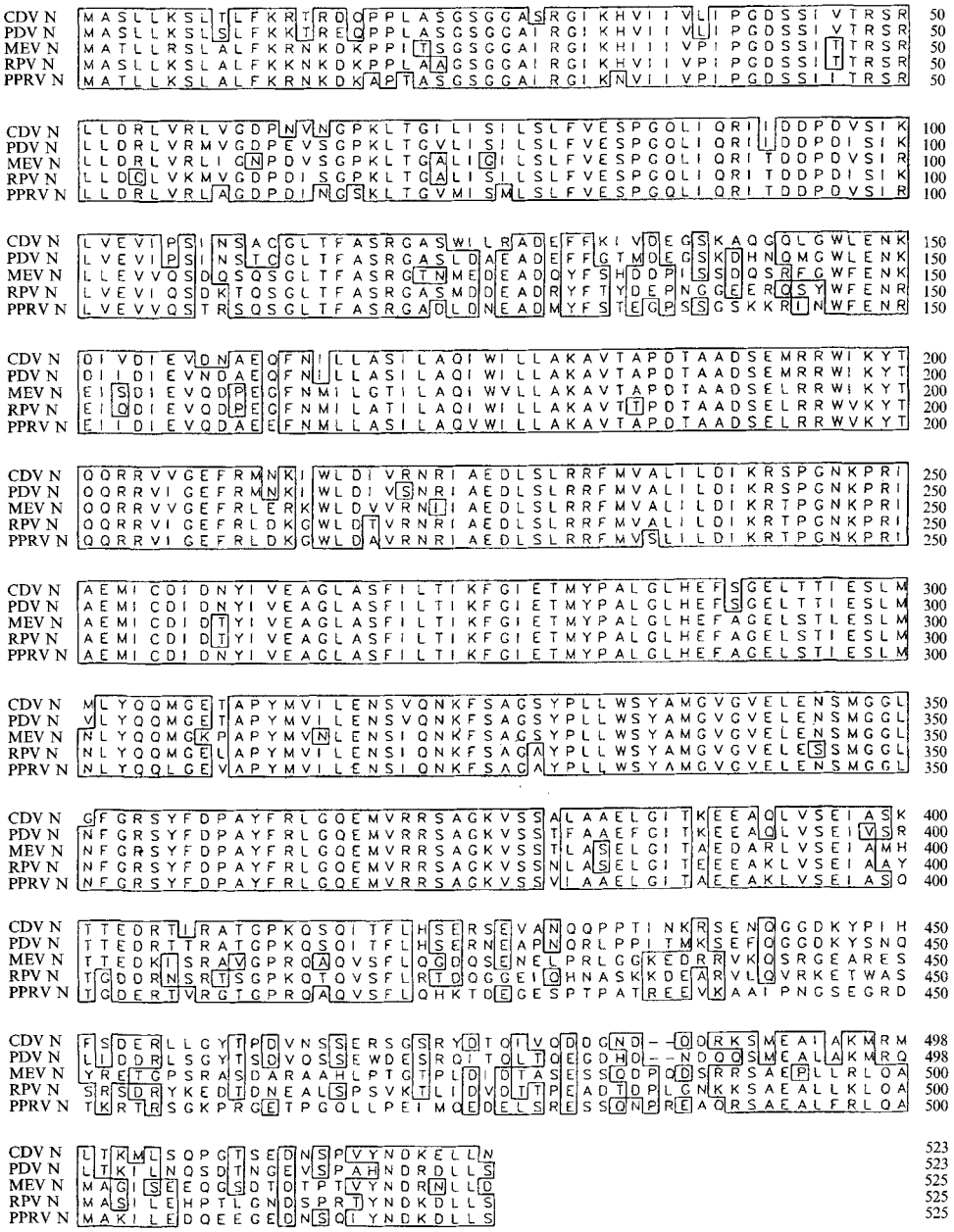

Fig. 3. Alignment of the amino acid sequences of the morbillivirus N proteins. Conserved amino acids are boxed.

(Blienkrone-Möller *et al.*, 1992; Barrett *et al.*, 1993). However, PPRV was slightly more related to MV and RPV than to CDV and PDV (see Fig. 2).

Clone NB71 had a single long open reading frame (ORF) starting at nucleotides 32 to 34 and ending with TGA at positions 1607 to 1609, leaving 53 untranslated nucleotides at the 3′ end of the mRNA, a feature shared with all other morbilliviruses. The protein encoded by this ORF consisted of 525 amino acid residues, as is the case for the N proteins of RPV and MV, whereas CDV and PDV have N proteins composed of only 523 residues. The predicted M_r of the PPRV N protein (58008) was close to the value (60K) estimated by polyacrylamide gel electrophoresis (Diallo *et al.*, 1987). It contained 22 proline residues, of which 13 are in identical positions in all morbillivirus N proteins, and five conserved tryptophan and one conserved cysteine residue at position 255. However, PDV and CDV have another cysteine at position 112 and the RPV lapinized vaccine (L strain) one at position 54. Another RPV strain (RBOK vaccine; EMBL no. X68311) has only the one conserved cysteine (M. D. Baron, personal communication). Therefore, since the presence of a second cysteine is dependent on the strain of the virus, this residue probably does not

contribute to the final conformation of the protein by making a disulphide bond with the conserved one.

Comparison of the PPRV N protein amino acid sequence with those of other morbilliviruses showed that the lowest similarity was with CDV (67·2%) followed by PDV (68·0%) whereas that with RPV and MV was slightly higher at 72·9% and 72·2%, respectively. The same comparison carried out for RPV and MV N proteins gave 74·3% similarity, confirming the closer relationship between RPV and MV than between RPV and PPRV. Phylogenetic analysis based on the protein sequences is shown in Fig. 2(*b*). Based on a comparative study of the different morbilliviruses using a panel of monoclonal antibodies (MAbs) to CDV and MV, McCullough *et al.* (1986) suggested that PPRV evolved relatively recently from RPV, which they proposed as the archevirus of the morbillivirus group (Norrby *et al.*, 1985). The present analysis, together with other sequence data for the PPRV P and F genes (Barrett *et al.*, 1993; Meyer, 1993), suggests that PPRV separated from the supposed archevirus long before MV.

Alignment of morbillivirus N proteins (Fig. 3) defined four regions with varying degrees of homology. Region I, amino acids 1 to 122, was quite well conserved with PPRV sharing 82·8%, 81·8%, 78·5% and 75·2% amino acid sequence identity with MV, RPV, CDV and PDV, respectively. Region II, amino acids 123 to 144, showed only low homology with amino acid identities of 40% or less. Region III, amino acids 145 to 420, was the most highly conserved area with 90·5% identity to RPV, and 85 to 86% to the other morbilliviruses. Finally, region IV, amino acid 421 to the C terminus, was the least conserved with 30·5%, 26·6%, 18·4% and 17·4% identity to MV, RPV, CDV and PDV, respectively.

Epitope mapping studies carried out by Buckland *et al.* (1989) showed that all epitopes recognized by anti-N MAbs specific to MV were located mainly in regions IV (C-terminal) and II, the least conserved areas of the protein. It is probable, therefore, that anti-N MAbs specific for either RPV (Libeau & Lefèvre, 1990; McCullough *et al.*, 1990) or PPRV (G. Libeau, personal communication) will map to the same regions. When the N protein sequences of the morbillivirus and the paramyxovirus genera were compared six areas with identical or highly conserved amino acid residues were identified, all in hydrophobic stretches of the proteins. These were at the following amino acid residues: 73 to 79 (area 1), 173 to 178 (area 2), 271 to 280 (area 3), 329 to 344 (area 4), 352 to 358 (area 5) and 383 to 387 (area 6). Only the first conserved block was in the morbillivirus region I, the others being in the highly conserved region III, the most conserved part of the morbillivirus N proteins. Previous comparisons had identified three conserved clusters corresponding to areas 2, 3 and 4 in our alignments (Morgan, 1991; Lyn *et al.*, 1991). These highly conserved sequences may be in parts of the protein which are so essential for the function of the protein that strict constraints are imposed and only minor amino acid changes are tolerated.

This work was supported by a grant from EEC-DGXII (contract no. TS2A-0178-F).

References

BARRETT, T. & MAHY, B. W. J. (1984). Molecular cloning of the nucleoprotein gene of canine distemper virus. *Journal of General Virology* **65**, 549–557.

BARRETT, T. & UNDERWOOD, B. (1985). Comparison of messenger RNAs induced in cells infected with each member of the Morbillivirus group. *Virology* **145**, 195–199.

BARRETT, T., VISSER, I. K. G., MAMAEV, L., GOATLEY, L., VAN BRESSEM, M.-F. & OSTERHAUS, A. D. M. E. (1993). Dolphin and porpoise morbilliviruses are genetically distinct from phocine distemper. *Virology* **193**, 1010–1012.

BLIXENKRONE-MÖLLER, M., SHARMA, B., VARSANYI, T. M., HU, A., NORRBY, E. & KÖVAMEES, J. (1992). Sequence analysis of the genes encoding the nucleocapsid protein and phosphoprotein (P) of phocine distemper virus, and editing of the P gene transcript. *Journal of General Virology* **73**, 885–893.

BUCKLAND, R., GIRAUDON, P. & WILD, F. (1989). Expression of measles virus nucleoprotein in *Escherichia coli*: use of deletion mutants to locate the antigenic sites. *Journal of General Virology* **70**, 435–441.

CATTANEO, R., SCHMIDT, A., BILLETER, M. A., SHEPPARD, R. D. & UDEM, S. A. (1988). Multiple mutations rather than host factors cause defective measles virus gene expression in a subacute sclerosing panencephalitis cell line. *Journal of Virology* **62**, 1388–1397.

CROWLEY, J. C., DOWLING, P. C., MENONNA, J., SILVERMAN, J. I., SCHUBACK, D., COOK, S. D. & BLUMBERG, B. M. (1988). Sequence variability and function of measles virus 3′ and 5′ ends and intercistronic regions. *Virology* **164**, 498–506.

CURRAN, M. D., CLARKE, D. K. & RIMA, B. K. (1991). The nucleotide sequence of the gene encoding the attachment protein H of the canine distemper virus. *Journal of General Virology* **72**, 443–447.

DIALLO, A. (1990). Morbillivirus group: genome organisation and proteins. *Veterinary Microbiology* **23**, 155–163.

DIALLO, A., BARRETT, T., LEFÈVRE, P. C. & TAYLOR, W. P. (1987). Comparison of proteins induced in cells infected with rinderpest and peste-des-petits-ruminants viruses. *Journal of General Virology* **68**, 2033–2038.

DIALLO, A., TAYLOR, W. P., LEFÈVRE, P. C. & PROVOST, A. (1989*a*). Atténuation d'une souche de virus de la peste-des-petits-ruminants: candidat pour un vaccin homologue vivant. *Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux* **42**, 311–319.

DIALLO, A., BARRETT, T., BARBRON, M., SUBBARAO, M. S. & TAYLOR, W. P. (1989*b*). Differentiation of rinderpest and peste-des-petits-ruminants viruses using specific cDNA clones. *Journal of Virological Methods* **23**, 127–136.

FELSENSTEIN, J. (1989). *Phylip 3.2 Manual.* Berkeley: University of California Herbarium.

KAMATA, H., TSUKIYAMA, K., SUGIYAMA, M., KAMATA, Y., YOSHIKAWA, Y. & YAMANOUCHI, K. (1991). Nucleotide sequence of cDNA to the rinderpest virus mRNA encoding the nucleocapsid protein. *Virus Genes* **5**, 5–15.

LEFÈVRE, P. C. & DIALLO, A. (1990). Peste-des-petits-ruminants. *Revue Scientifique et Technique de l'Office International des Epizooties* **9**, 951–965.

LIBEAU, G. & LEFÈVRE, P. C. (1990). Comparison of rinderpest and peste-des-petits-ruminants viruses using anti-nucleoprotein monoclonal antibodies. *Veterinary Microbiology* **25**, 1–16.

LYN, D., GILL, D. S., SCROGGS, R. A. & PORTNER, A. (1991). The nucleoproteins of human parainfluenza virus type 1 and Sendai virus

share amino acid sequences and antigenic and structural determinants. *Journal of General Virology* **72**, 983–987.

McCULLOUGH, K. C., SHESHBERADARAN, H., NORRBY, E., OBI, T. U. & CROWTHER, J. R. (1986). Monoclonal antibodies against morbilliviruses. *Revue Scientifique et Technique de l'Office International des Epizooties* **5**, 411–427.

McCULLOUGH, K. C., OBI, T. U. & SHESHBERADARAN, H. (1991). Identification of epitope(s) on the internal proteins of rinderpest virus which are absent from peste-des-petits-ruminants virus. *Veterinary Microbiology* **26**, 313–321.

MEYER, G. (1993). *Clonage et sequençage du gène codant pour la protéine de fusion du virus de la peste-des-petits-ruminants*. Thèse Doctorat Vétérinaire, Université Paul-Sabatier, Toulouse.

MORGAN, E. M. (1991). Evolutionary relationships of paramyxovirus nucleocapsid-associated proteins. In *The Paramyxoviruses*, pp. 163–179. Edited by D. W. Kingsbury. New York: Plenum Press.

MURPHY, G. & KAVANAGH, T. (1988). Speeding-up the sequencing of double-stranded DNA. *Nucleic Acids Research* **16**, 5198.

NORRBY, E., SHESHBERADARAN, H., McCULLOUGH, K. C., CARPENTER, W. C. & ORVELL, C. (1985). Is rinderpest virus the archevirus of Morbillivirus genus? *Intervirology* **23**, 228–232.

SHAILA, M. S., PURSHOTHAMAN, V., BHAVASAR, D., VENUGOPAL, K. & VENKATESAN, R. A. (1989). Peste-des-petits-ruminants of sheep in India. *Veterinary Record* **125**, 602.

SIDHU, M. S., HUSAR, W., COOK, S. D., DOWLING, P. C. & UDEM, S. A. (1993). Canine distemper terminal and intergenic non-protein coding nucleotide sequences: completion of the entire CDV genome sequence. *Virology* **193**, 66–72.

VISSER, I. K. G., VAN BRESSEM, M.-F., DE SWART, R. L., VAN DE BILDT, M. W. G., VOS, H. W., VAN DER HEIJDEN, R. W. J., SALIKI, J. T., ÖRVELL, C., KITCHING, P., BARRETT, T. & OSTERHAUS, A. D. M. E. (1993). Characterization of morbilliviruses isolated from dolphins and harbour porpoises in Europe. *Journal of General Virology* **74**, 631–641.

(*Received 7 June 1993; Accepted 3 September 1993*)