Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 110 (2005) 119–124
Virus Research
www.elsevier.com/locate/virusres

Full genome sequence of peste des petits ruminants virus, a member of the *Morbillivirus* genus☆

Dalan Bailey[a], Ashley Banyard[a], Pradyot Dash[a], Aykut Ozkul[b], Tom Barrett[a,*]

[a] *Pirbright Laboratory, Institute for Animal Health, Ash Road, Woking, Surrey GU24 ONF, UK*
[b] *Faculty of Veterinary Medicine, University of Ankara, Diskapi 06110, Ankara, Turkey*

Received 25 November 2004; received in revised form 25 January 2005; accepted 25 January 2005
Available online 19 March 2005

Abstract

Peste des petits ruminants virus (PPRV) causes an acute febrile illness in small ruminant species, mostly sheep and goats. PPRV is a member of the *Morbillivirus* genus which includes measles, rinderpest (cattle plague), canine distemper, phocine distemper and the morbilliviruses found in whales, porpoises and dolphins. Full length genome sequences for these morbilliviruses are available and reverse genetic rescue systems have been developed for the viruses of terrestrial mammals, with the exception of PPRV. This paper presents the first published full length genome sequence for PPRV. The genome was found to be consistent with the rule-of-six and open reading frames (ORFs) were identified that encoded the eight proteins characteristic of morbilliviruses. At the nucleotide (nt) level, the full length genome of PPRV was most similar to that of rinderpest, the other ruminant morbillivirus. However, at the protein level five of the six structural proteins and the V protein showed a greater similarity to the dolphin morbillivirus (DMV) while only the C and L proteins showed a high relationship to rinderpest.

Keywords: *Morbillivirus*; Genome sequence; Peste des petits ruminants; PPR

Peste des petits ruminants (PPR) is an economically significant disease of small ruminants such as sheep and goats (Dhar et al., 2002). The virus is classified in the *Morbillivirus* genus of the *Paramyxoviridae* family. This genus includes the type virus measles (MV) as well as rinderpest virus (RPV), the agent of cattle plague, canine distemper virus (CDV) and the phocine, porpoise and dolphin distemper viruses that infect marine mammals (Barrett, 2001). Morbilliviruses are non-segmented, single stranded, negative sense RNA viruses with genomes approximately 15–16 kb in size. PPRV was first isolated in West Africa in the 1940s (Gargadennec and Lalanne, 1942) but is now also known to be present in a broad belt of sub-Saharan Africa, Arabia, the Middle East and Southern Asia. Major outbreaks in Turkey and India in recent years have indicated a marked rise in the global incidence of PPRV (Nanda et al., 1996; Ozkul et al., 2002; Shaila et al., 1996). Epidemics in sheep and goats, the mainstay of subsistence farming in the developing world can cause mortality rates of 50–80% in naïve populations (Kitching, 1988). Antelope and other small wild ruminant species can also be severely affected (Abu Elzein et al., 2004).

Following infection there is a 3–4 day incubation period during which the virus replicates in the draining lymph nodes of the oro-pharynx before spreading via the blood and lymph to other tissues and organs in the animal, including the lungs causing a primary viral pneumonia. This pneumonia is often exacerbated by secondary bacterial infections which are a major feature of the pathology of this disease. Other overt clinical signs include ocular and nasal discharges, which usually become mucopurulent, as well as conjunctivitis, necrotic stomatitis and severe diarrhoea (Gibbs et al., 1979).

Full length genome sequences are available for MV, RPV, CDV and the dolphin morbillivirus (DMV). These data have been used to establish reverse genetics for MV, RPV and

☆ Genome accession number AJ849636 for PPRV Tu/00.
* Corresponding author. Tel.: +44 1483 231009; fax: +44 1483 232448.
E-mail address: tom.barrett@bbsrc.ac.uk (T. Barrett).

0168-1702/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2005.01.013

CDV, a technology critical for negative sense RNA virus research (Nagai, 1999; Neumann et al., 2002). In this paper, we present the first full genome sequence of a wild-type PPRV, isolated from a sick sheep in Turkey in 2000 (Tu/00), and compare it to other morbillivirus genomes. This virus, like most PPRV strains, was isolated by passage in Vero cells to obtain sufficient RNA for amplification and sequencing. We cannot be sure whether a low number of passages (<5) will affect pathogenicity; however, during development of the live attenuated PPRV vaccine the strain was not attenuated by passage 20 in Vero cells (Diallo et al., 1989). Vero cells were infected at an MOI of less than 0.01 with a low passage virus and incubated until significant cytopathic effect was observed. RNA was extracted using Trizol reagent (Invitrogen) and Superscript II Reverse Transcriptase (Invitrogen) used, with gene specific primers, to generate cDNAs which provided a template for PCR amplification. The 16 kb genome of PPRV was divided into six overlapping regions and amplified by gradient PCR using a blend of *Taq* (Invitrogen) and *Pfu* (Stratagene) polymerases at a ratio of 8:1. Several PCR products were cloned into either pGEM-T (Promega) or pT7 Blue (Novagen). Around 99% of the genetic information for PPRV was generated from these clones. The remainder, the 3′ and 5′ genome termini, was determined by rapid amplification of cDNA ends (RACE) using Moloney's murine leukaemia virus reverse transcriptase, Mu-MLV RT (Promega) to generate cDNA which was tailed with cytosine residues using terminal deoxy-nucleotidyl transferase (TdT) (Promega) (Baron and Barrett, 1995). This polyC_n tailed cDNA was used as a template for *Taq* PCR in combination with a polyG_n primer. The RACE generated product (subsequently cloned into pGEM-T) represented true sequences for the termini of the PPRV genome.

The DNA sequence of the clones was obtained on both strands more than once; the average number of readings per consensus character being 9.24. An automated sequencer (Beckman CEQ8000) was used to generate 278 readings, of which an average of 543.78 base pairs was used to calculate consensus information. The total number of used characters in the entire database was 151,171 and genome assembly was performed using the Staden package (Staden et al., 1998). Data was processed by pregap4 (Phred) to remove poor quality readings and vector sequence. Contiguous assembly of the sequence data was accomplished in the gap4 program using the Phred assembly algorithm (set to quality and difference clip the data). Initial sequence data were used to design more primers, the process eventually resulting in a full length genome sequence for PPRV.

This revealed that the genome of PPRV was analogous in structure to all other morbilliviruses for which complete sequences are available (Fig. 1). The genome sequence was 15,948 nucleotides in length and encoded the same eight proteins as the type virus (MV), also its length was divisible by six, a feature shared with other deoxyvirinae (Calain and Roux, 1993). The genome was most similar at the nucleotide (nt) level to that of RPV, another virus of ruminant species (Table 1). More sequence data would obviously be required to investigate evolutionary relationships further. The exact lengths of morbillivirus genomes differ due to a variably sized junction between the M and F genes, not because of varied protein lengths. This junction has a particularly high GC content (65% plus) but no obvious role in replication has been shown (Liermann et al., 1998; Radecke et al., 1995).

A graphical representation of the similarity between PPRV and other members of the morbillivirus genus is provided (Fig. 1). This plot shows the regions of high and low nucleotide conservation in morbillivirus genomes. Regions of high conservation include the L and M genes. Non-coding regions, such as the GC-rich M/F gene junction, show very little overall similarity. The nucleocapsid (N) protein open reading frame (ORF) began at UAC (nt 108) and terminated at AUU (nt 1685) (becoming AUG and UAA in the positive sense transcript). The N protein has a similar RNA encapsidation function across the whole of the paramyxovirus family. It binds to the genome once every six nucleotides (a hexamer unit) forming the ribo-nucleoprotein (RNP) complex.

A polyU tract (nt 1739–1744), responsible for polyadenylation of the positive sense transcripts produced by the viral RNA-dependent RNA polymerase, was located 52 bases downstream of the N ORF stop. This sequence, U_6GAAUCC, is highly conserved in the morbilliviruses and acts as part of a gene stop and polyadenylation signal. Reduction in size of the polyU tract of the paramyxovirus simian virus 5 (SV5) from six residues to four was shown to diminish downstream initiation to 20–30% of wild-type levels indicating a possible role as a critical spacer region between gene stops and starts (Rassa et al., 2000). This sequence was maintained throughout the genome of PPRV except at the M/F, and F/H junctions where the U tract is interrupted by a G residue becoming UUUGUUUUGAAUCC. It is unknown if this G insertion has any modulating effect on the polymerase stuttering mechanism employed to polyadenylate nascent mRNAs, or on the transcription of the downstream mRNA(s). Immediately following the polyU tract was a conserved triplet (GAA) that has been shown to be an intergenic region which is not transcribed during mRNA synthesis but which is an essential signal for the activity of the viral polymerase since mutations or deletions in this region can reduce or abrogate viral replication (Kolakofsky et al., 1998).

The intergenic triplet was followed by the phosphoprotein (P) gene start (nt 1748). This second transcription unit encodes the P, C and V proteins. The P ORF began at nucleotide 1807 coding through to 3336, producing a protein of 509 amino acids (aa). The P protein of morbilliviruses interacts with both the N and L proteins to form the viral polymerase. Position 1807 is also the start of the V protein (297 aa) ORF. The N terminus of V is identical to P but polymerase slipping at the editing site (nt 2487–2498) can result in a frameshift whereby an inserted G residue in the mRNA directs the production of an alternative 67 aa C terminus for the protein (Mahapatra et al., 2003). The hexamer phasing of the P gene editing site is also thought to play a critical