Genome sequencing identifies genetic and antigenic divergence of porcine picobirnaviruses

Krisztian Banyai,1 Christiaan Potgieter,2,3 Ákos Gellért,4 Balasubramanian Ganesh,5 Maria Tempesta,6 Eleonora Lorusso,6 Canio Buonavoglia6 and Vito Martella6

1Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
2Virology Division, Onderstepoort Veterinary Institute, Onderstepoort, South Africa
3Department of Biochemistry, Centre for Human Metabonomics, North-West University, Potchefstroom, South Africa
4Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Hungary
5Division of Virology, National Institute of Cholera and Enteric Diseases (NICED), Kolkata, West Bengal, India
6Dipartimento di Medicina Veterinaria, Università di Bari Aldo Moro, Bari, Italy

The full-length genome sequence of a porcine picobirnavirus (PBV) detected in Italy in 2004 was determined. The smaller (S) genome segment was 1730 nt, coding for a putative RNA-dependent RNA polymerase. Two distinct subpopulations of larger (L) genome segment (LA and LB) were identified in the sample, with the sizes ranging from 2351 to 2666 nt. The ORF1, coding for a protein of unknown function, contained a variable number of repetitions of the ExxAxxE motif. The capsid protein-coding ORF2 spanned nt 810–2447 in the LB variants and started at nt 734 in the LA variants. However, a termination codon was present only in one of all the LA segment variants. Three-dimensional modelling of the porcine PBV capsids suggested structural differences in the protruding domain, tentatively involved as antigens in the humoral immune response. Altogether, these findings suggest the simultaneous presence of two different PBV strains sharing the same S segment but displaying genetically diverse L segments. In addition, the sample probably contained a mixture of PBVs with aberrant RNA replication products. Altered structure in the L segments could be tolerated and retained in the presence of functionally integer-cognate genes and represents a mechanism of virus diversification.

The GenBank/EMBL/DDBJ accession numbers for the sequences of the porcine PBV genome determined in this study are KF861767–KF861773.
diarrhoeic infections (Giordano et al., 1998; Grohmann et al., 1993).

Transmission of PBV from one host species to another has been suggested in molecular epidemiological and strain characterization studies, based on very close genetic relatedness between human and porcine or human and equine PBV strains (Bányai et al., 2008; Ganesh et al., 2011; Giordano et al., 2011). Given that PBV is frequently found in wastewater and surface water, a water-borne route of transmission has been also suggested (Ganesh et al., 2011). Indeed, the genetic similarity among heterologous PBV strains has been seen along a short, ~200 bp gene fragment of the S segment. Because the PBV genome is segmented and measures ~ 4 kb, the value of this short fragment to describe molecular epidemiology may be limited. A major obstacle to perform whole-genome-based studies is the lack of genome sequence data for a reasonably high number of strains. So far, the complete genome sequence has been determined only for a human PBV strain and, more recently, a phocine PBV strain (Wakuda et al., 2005; Woo et al., 2012). In addition, complete or partial L segment sequences have become available for a lapine strain and some human strains (Green et al., 1999; Rosen et al., 2000), whilst the sequences of a dozen or so S gene fragments longer than 1000 bp have been determined for human, porcine, bovine and murine PBVs (Ghosh et al., 2009; Phan et al., 2011; Rosen et al., 2000; van Leeuwen et al., 2010; Wang et al., 2012; and unpublished GenBank entries). In this study, sequencing and molecular characterization of the full-length genome of a porcine PBV strain was performed by sequence-independent amplification and high-throughput sequencing of the genomic dsRNA.

During a rotavirus surveillance project in 2004, total RNAs prepared from 37 randomly selected faecal samples collected from diarrhoeic piglets were analysed by PAGE and silver staining to diagnose rotavirus infections. One sample contained bands consistent with PBV dsRNA (Fig. 1a). This porcine PBV strain, designated 221/04-16, was detected from a 45-day-old piglet with diarrhea at a farm in Brescia, Northern Italy, in 2004.

Total RNA was extracted from the faecal specimen using the commercial TRI-REAGENT–LS (Molecular Research Centre), according to the manufacturer’s protocol. ssRNA was removed by precipitation with 2 M LiCl (Sigma) at 4 °C for 16 h followed by centrifugation at 16 000 g for 30 min. dsRNA was purified from the resulting supernatant using a MinElute gel extraction kit (Qiagen). The complete nucleotide sequence was determined for the S and L segments of this strain by sequence-independent amplification and deep sequencing as described previously (Potgieter et al., 2002, 2009; Lambden et al., 1992). Briefly, an anchor primer was ligated to dsRNA and used for reverse transcription and subsequent amplification of virus genome segments. Amplified cDNA products were then subjected to sequencing with GS FLX technology at Inqaba Biotech.

Lasergene7 software from DNASTAR was used for de novo assembly of the contigs. Files containing the sequence information, quality values and flowgrams (sff files) were loaded into the Seqman 7 program of the Lasergene software. Contigs resulting from the assembly were checked manually and their consensus sequences were exported as FASTA files. Consensus sequences were aligned to known sequences using MEGALIGN (Lasergene7). Finally, sequences were subjected to BLASTN analysis using the National Center for Biotechnology Information (NCBI) website. The NCBI’s ORF finder was used to identify protein-coding regions (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Nucleotide and amino acid sequence alignments were prepared by Multalin (Corpet, 1988) and manually adjusted to codons. Phylogenetic analysis was performed by using MEGAS (Tamura et al., 2011).

The S segment of strain 221/04-16 was 1730 bp. In comparison with other full-length PBV S segments, this segment was shorter than those of the human strain Hy005102 (1745 bp), the bovine strain RUVB-P/IND/20 (1758 bp) and longer than those of the human strains 1-CHN-97 (1696 bp) and GPBV6C1 (1711 bp), and the otarine PBV (1687 bp). The 5’ and 3’ UTRs were 74 and 49 bp, respectively, and they were structurally heterogeneous when compared with the reference strains, except for the common GTAAA and ACTGC pentanucleotides, located at the termini of the RNA segments and conserved among the various PBV strains. The S segment of strain 221/04-16 contained a single 1607 nt ORF (nt 75–1682). Alternative ORFs were also predicted (data not shown). However, the predicted proteins were too short and their presence was not consistently found in the S segments of other strains; therefore, these putative ORFs were ignored in this analysis. The protein encoded by the S segment ORF was 535 aa and was predicted to be the viral RdRp. It had conservative motifs at aa 266–271, 328–333 and 364–366 defining the GDD sequence motif. When comparing the 221/04-16 S gene sequence with the Chinese strain SD, a single-nucleotide deletion was observed at position 768 that was compensated for at position 848 by a single-nucleotide insertion. This insertion caused a codon shift, altering the translation of a 28 aa stretch between the two porcine PBVs (Fig. 1c).

Phylogenetic analysis was first performed using an ~200 bp fragment of the RdRp used to amplify GGI PBV strains in diagnostic PCR (Rosen et al., 2000). This analysis identified a Chinese porcine PBV strain as presenting the greatest nucleotide sequence identity (95 %) with that of the Italian strain. Other relatively closely related sequences (up to 89 %) included viral sequences identified from US water samples. The Italian porcine PBV strain shared a lower nucleotide similarity to European and American porcine PBVs (up to 75 %) and shared comparable similarities with European, Asian and American human PBVs (up to 88 %). Another partially overlapping region of the S segment of ~280 bp used specifically to amplify porcine PBVs (Carruyo et al.,
was analysed separately. This analysis demonstrated that the Italian strain shared somewhat greater similarities with Argentinean porcine and human PBV strains (up to 97% nucleotide identity; data not shown). Analysis of the whole ORF of several strains of porcine, human, murine, phocine and bovine origin identified comparable sequence identity data when these strains were compared along the commonly characterized ~200 bp fragment. The same observation was made when comparing the Italian strain with two GGII strains, 4-GA-91 and GPBV6G2. However, it should be considered that extrapolating nucleotide similarities obtained from a short fragment may not be relevant in each case as exemplified by the codon shift described above. Phylogenetic analysis of an ~1200 bp fragment of selected strains with longer S genome segment sequences confirmed the similarity-based prediction of the genetic relatedness between the Italian strain and the Chinese SD strain (Fig. 1b).

Amplification and cloning of the L segment yielded a heterogeneous population of cDNA products, comprising two major types and several minor sequence variants within each type. The major sequence types were designated LA and LB. The LB variants (i.e. LB-1, LB-3 and LB-4) were

**Fig. 1.** (a) RNA profile of PBV strain 221/04-16 identified in an Italian pig with diarrhoea (right lane), with the L and S RNA segments indicated by arrowheads. The genomic pattern of a rotavirus A strain (left lane) electrophoresed in parallel with the PBV strain is also shown, with RNA segments indicated by arrowheads. (b) Nucleotide (Nt) and amino acid (Aa) sequence-based phylogeny of the partial S genome segment. The gene fragment used in the phylogeny encodes a ~400 aa polypeptide (nt 290–1414, reference strain Hy005102). Bootstrap values ≥ 50% are indicated. GGI and GGII refer to the genogroup assignments of PBV strains. Strain 221/04-16 is indicated by an arrowhead. Bar, substitutions per site. (c) Partial S genome segment of two porcine PBV strains, the Italian strain 221/04-16 and the Chinese strain SD. The alignment shows the relative position of the single-nucleotide insertion and deletion resulting in remarkable amino acid changes within the affected region. Numbers above the alignment refer to the nucleotide positions of strain 221/04-16. (d) Number and position of the sequence motif ExxRxNxxxE (x = any amino acid), indicated by shading, in the ORF1-encoded protein (L segment) of various PBV strains. Hu, human; La, lapine; Ot, otarine; Po, porcine. Hu-1 is VS10, Hu-2 is Hy005102, Po-1 is LA-5 and Po-2 is LB-3. Numbers on the right indicate the length of the protein encoded by ORF1, ranging from 106 aa (lapine PBV) to 224 aa (Hy005102). The protein sequences are not aligned.
2524 bp and had two predicted ORFs at nt 195–794 (ORF1) and 810–2447 (ORF2), encoding proteins of 199 and 545 aa, respectively. The LA variants, LA-2, LA-5 and LA-7, were 2361, 2666 and 2351 bp, respectively. The ORF1 started at nt 185 and ended at nt 721, coding for a polypeptide of 175 aa. The ORF2 started at nt 734 in all LA sequences. However, whilst LA-5 was terminated with a stop codon, LA-2 and LA-7 seemed to be pseudogenes without translation termination signals. The protein encoded by the LA-5 sequence was longer (619 aa) than those of other PBV CPs (i.e. 590 aa in the lapine PBV strain, 552 aa in the human strain Hy005102 and 575 aa in the otarine strain). The 3’ UTRs ranged from 184 nt (LA segments) to 194 nt (LB segments). The 5’ UTRs ranged from 76 nt (LB segments) to 84 nt (LA-5).

The ORF1-encoded proteins of the LA and LB clones were extremely diverse, yet retained common genetic signatures. One shared feature was the recently recognized 10 aa sequence motif (Da Costa et al., 2011), ExxRxNxxxE, where x can be any amino acid. Four and six repeats of this motif were identified in LB clones (e.g. LB-3) and LA clones (e.g. LA-5), respectively. The amino acid distance between two adjacent motifs was variable. When comparing the ORF1-encoded proteins with other reference strains, we found similar diversities in the number and position of the ExxRxNxxxE motif (Fig. 1d).

Due to the low amount of sequence data derived from cognate genes, a meaningful phylogenetic analysis of the CP-coding genes could not be performed. However, as the three-dimensional structure of a rabbit PBV has been determined using recombinant CP (Duquerroy et al., 2009), attempts were made to gain insight into the structure of the CP of porcine PBVs by homology modelling. The deduced amino acid sequences of the LB-3 and LA-5 clones were subjected to protein modelling employing the Prime module of Schrödinger Suite (http://www.schrodinger.com) molecular modelling software package using the rabbit PBV CP X-ray structure (Protein Data Bank ID: 2VF1) as the homo-dimer template. Sequence similarity values of LB-3 and LA-5 with the rabbit CP were 38.3 and 39.6 %, respectively. The models built for the rabbit CP model (Fig. 2b–d, g). These additional surface loops were located on the most protruded edge of the coat protein, raising the possibility that these peptide insertions may act as potential immune epitopes for the B-cell-mediated response.

The porcine PBV genome described represents the third complete sequence of this highly diverse virus family. Until now, only one human and one phocine PBV genome sequence had been determined (Wakuda et al., 2005; Woo et al., 2012). In this study, we used sequence-independent amplification of genomic dsRNA and a next-generation sequencing strategy. This seemed important to maximize the depth of sequence information and minimize any bias inherent to sequence variation seen in PBV genomes, given that porcine PBVs frequently occur as a mixed-strain population in the same host and there is evidence indicative of the quasi-species nature of PBVs (Bányai et al., 2008).

The porcine stool specimen from which strain 221/04-16 was isolated contained two major PBV subpopulations, each sharing the same RdRp gene and expressing different CP genes. The conservation of the RdRp gene sequence suggested that one strain may be a reassortant. The likelihood that another RdRp gene was also present but remained hidden in the sample is low, given that sequence-independent amplification coupled with deep sequencing were employed, which on the one hand showed no bias in recognition of the template genomic dsRNA segment and on the other hand should detect minor sequence variants. This was further illustrated with the truncated L segments of type LA clones. In this study, we did not determine whether pseudogenes among the LA sequences were generated as an artefact during RNA processing or whether the shorter L segments encoding pseudogenes were incorporated into the capsid. However, PBV genes with altered structure (pseudogenes) may be tolerated/maintained if functionally integer-cognate genes are retained (Bányai et al., 2008), and we were able to determine at least one LA segment variant (LA-5) with a correct ORF architecture in the porcine PBV genome.

Sequencing of the whole genome of strain 221/04-16 uncovered another key mechanism probably contributing to the great sequence diversity seen even among PBV strains collected from the same host species (Bányai et al., 2003, 2008; Carruyo et al., 2008; Fregolente et al., 2009; Phan et al., 2011; Rosen et al., 2000; van Leeuwen et al., 2010; Wang et al., 2012). We found further proof that insertions and deletions (in/dels) occur in the PBV genome, as evidenced by the sequence length heterogeneity seen in the UTRs and the coding regions of both the S and L segments. The effect of such mutations on phylogenetic analysis and protein evolution was demonstrated by the observation of a deleterious mutation in the RdRp gene, which was compensated for by a single-nucleotide insertion several nucleotides downstream (Fig. 1c). This mutation resulted in a highly diverse deduced amino acid sequence in the flanking region. It is possible that, in addition to the accumulation of single point mutations, these compensatory in/del mutations are also common in the PBV genome, generating variants of viral proteins with modified antigenic structure, more efficient replication and/or increased viral fitness (Ganesh et al., 2012).
In conclusion, this study provides a more detailed insight into the evolutionary mechanisms of PBVs. How different mechanisms contribute to better adaptation to a host or allow interspecies transmission should be the subject of future studies. Until adequate cell-culture and model animals are established, whole-genome sequencing of PBVs from
variou s host species remains the only alternative to investigate this intriguing question.

Acknowledgements

K.B. was supported by the Momentum Programme (Hungarian Academy of Sciences). Á.G. was the recipient of a János Bolyai fellowship from the Hungarian Academy of Sciences. The licensing of the Schrödinger Suite software package was funded from an OTKA grant under agreement no. 108793.

References


