High prevalence and genetic diversity of porcine bocaviruses in pigs in the USA, and identification of multiple novel porcine bocaviruses

Yong-Hou Jiang,1,2 Chao-Ting Xiao,1 Shuang-Hui Yin,1 Priscilla F. Gerber,1,3 Patrick G. Halbur1 and Tanja Opriessnig1,4

1Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA
2College of Life Science, Zhejiang Sci-Tech University, Hangzhou, PR China
3Laboratório de Pesquisa em Virologia Animal, Departamento de Medicina Veterinária Preventiva, Universidade Federal de Minas Gerais, Minas Gerais, Brazil
4The Roslin Institute and The Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

Correspondence
Tanja Opriessnig
Tanja.Opriessnig@roslin.ed.ac.uk

Received 16 July 2013
Accepted 13 November 2013

Viruses in the genus Bocavirus are associated with respiratory and enteric disease in dogs and cattle. In addition, novel porcine bocaviruses (PBoVs) have been identified in domestic and wild pigs in recent years, but are of unknown relevance to date. The objectives of this study were to determine the prevalence rates and genetic diversity of PBoVs in pigs in the USA. Using newly established multiplex real-time PCR assays, 385 lung, lymph node, serum and faecal samples from pigs with various disease conditions were investigated. A high PBoV prevalence rate ranging from 21.3 to 50.8% was identified in the investigated samples and often two or more PBoV species were detected in the same sample. Cloning and sequencing analysis of the partial non-structural protein NS1 and the capsid proteins VP1 and VP2 of DNA samples positive for PBoV groups 1 (n=6), 2 (n=16) and 3 (n=42), including subgroups 3A, 3B or 3C, revealed a high genetic diversity especially for the PBoV G3 VP2 gene, whereas the PBoV group 1 VP1 gene displayed a low nucleotide polymorphism. Using primer walking, 18 partial or nearly complete genomes of PBoVs were obtained and six of the 18 nearly complete genomes represented novel PBoV species. Recombination analysis using partial NS1, VP1 and VP2 genes and the nearly complete genomes indicated possible recombination events within and between PBoVs. Further studies will be required to reveal the possible pathogenic role of these diverse PBoVs.

INTRODUCTION

Several new bocaviruses have been identified recently in domestic and wild pigs and these are known as porcine bocaviruses (PBoVs) (Xiao et al., 2013b). Viruses in the genus Bocavirus belong to the subfamily Parvovirinae of the Parvoviridae family, which is a group of divergent linear ssDNA viruses (Tijssen et al., 2011). Bocaviruses have a genome of ~5 kb and contain three main ORFs, coding for non-structural and capsid proteins (Allander et al., 2005; Chen et al., 1986, 2010; Kapoor et al., 2010a, b; Manteufel & Truyen, 2008; Schwartz et al., 2002). The name bocavirus is derived from the host initials of the two initially identified species in this genus: bovine parvovirus, associated with diarrhoea and mild respiratory disease in calves, and canine minute virus, associated with fetal and neonatal respiratory disease and enteritis in dogs (Binn et al., 1970; Mochizuki et al., 2002; Storz et al., 1978). Several bocaviruses from different hosts have since been discovered and characterized genetically. These include at least four human bocavirus species (HBoV-1 to -4) (Allander et al., 2005; Arthur et al., 2009; Kapoor et al., 2009, 2010b), gorilla bocavirus (GBoV) (Kapoor et al., 2010a), bocavirus-like sequences detected in faecal samples of primates (Sharp et al., 2010), California sea lion bocavirus (CslBoV-1 to -4) (Li et al., 2011) and a novel canine bocavirus (CBoV) (Kapoor et al., 2012). More than five PBoV species have been identified in pigs in various geographical regions, including Sweden, China, the UK and Uganda (Blomström et al., 2009; Cheng et al., 2010; Lau et al., 2011; Li et al., 2012; McKillen et al., 2011; Shan et al., 2011a; Zhai et al., 2010).
The prevalence rates of PBoVs in clinical samples were determined to be 20 genomic copies per microgram of the plasmid for PBoV G1, 600 genomic copies per microgram of the plasmid for PBoV G2, 20 genomic copies per microgram of the plasmid for PBoV G3A, 20 genomic copies for PBoV G3B and 20 genomic copies for PBoV G3C. The detection limits were determined to be 20 genomic copies for PBoV G1, 600 genomic copies per microgram of the plasmid for PBoV G2, 20 genomic copies per microgram of the plasmid for PBoV G3A, 20 genomic copies for PBoV G3B and 20 genomic copies for PBoV G3C. The establishment of multiplex PCR assays was confirmed by BLAST and by testing available samples from pigs in the USA.

### Table 1. Prevalence of PBoV G1, G2, G3A, G3B and G3C in different sample types obtained from healthy and diseased pigs in the USA

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Age group</th>
<th>Health status</th>
<th>Farms/states</th>
<th>No. samples</th>
<th>PBoV G1 positive (%)</th>
<th>PBoV G1 positive (no.)</th>
<th>PBoV G2 positive (%)</th>
<th>PBoV G2 positive (no.)</th>
<th>PBoV G3A positive (%)</th>
<th>PBoV G3A positive (no.)</th>
<th>PBoV G3B positive (%)</th>
<th>PBoV G3B positive (no.)</th>
<th>PBoV G3C positive (%)</th>
<th>PBoV G3C positive (no.)</th>
<th>Co-infected with two (%)</th>
<th>Co-infected with three (%)</th>
<th>Co-infected with four (%)</th>
<th>Co-infected with all five (%)</th>
<th>Total no. infected pigs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymph node</td>
<td>Diseased</td>
<td>NA</td>
<td>24/6</td>
<td>61</td>
<td>28 (45.9)</td>
<td>4 (6.6)</td>
<td>2 (3.3)</td>
<td>5 (8.2)</td>
<td>8 (13.1)</td>
<td>7 (11.4)</td>
<td>2 (3.3)</td>
<td>2 (3.3)</td>
<td>0 (0)</td>
<td>31 (50.8)</td>
<td>31 (50.8)</td>
<td>31 (50.8)</td>
<td>31 (50.8)</td>
<td>31 (50.8)</td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>Diseased</td>
<td>NA</td>
<td>17/9</td>
<td>111</td>
<td>52 (46.8)</td>
<td>4 (3.6)</td>
<td>1 (0.9)</td>
<td>4 (3.6)</td>
<td>7 (6.3)</td>
<td>2 (1.8)</td>
<td>4 (3.6)</td>
<td>0 (0)</td>
<td>55 (49.5)</td>
<td>55 (49.5)</td>
<td>55 (49.5)</td>
<td>55 (49.5)</td>
<td>55 (49.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum</td>
<td>Diseased</td>
<td>NA</td>
<td>3/1</td>
<td>20</td>
<td>5 (25.0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>9 (45.0)</td>
<td>9 (45.0)</td>
<td>9 (45.0)</td>
<td>9 (45.0)</td>
<td>9 (45.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faeces</td>
<td>Diseased</td>
<td>Suckling</td>
<td>7/3</td>
<td>13</td>
<td>2 (15.4)</td>
<td>1 (7.7)</td>
<td>1 (7.7)</td>
<td>1 (7.7)</td>
<td>1 (7.7)</td>
<td>1 (7.7)</td>
<td>1 (7.7)</td>
<td>3 (23.1)</td>
<td>3 (23.1)</td>
<td>3 (23.1)</td>
<td>3 (23.1)</td>
<td>3 (23.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nursery</td>
<td>26/8</td>
<td>55</td>
<td>31 (56.4)</td>
<td>23 (41.8)</td>
<td>25 (45.5)</td>
<td>35 (63.6)</td>
<td>39 (70.9)</td>
<td>3 (5.5)</td>
<td>16 (29.1)</td>
<td>11 (20)</td>
<td>9 (16.4)</td>
<td>47 (85.5)</td>
<td>47 (85.5)</td>
<td>47 (85.5)</td>
<td>47 (85.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grow-finish</td>
<td>26/6</td>
<td>69</td>
<td>39 (56.5)</td>
<td>27 (39.1)</td>
<td>55 (79.7)</td>
<td>43 (62.3)</td>
<td>58 (84.1)</td>
<td>8 (11.6)</td>
<td>9 (13.0)</td>
<td>16 (23.2)</td>
<td>19 (27.5)</td>
<td>62 (89.9)</td>
<td>62 (89.9)</td>
<td>62 (89.9)</td>
<td>62 (89.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mature</td>
<td>5/3</td>
<td>9</td>
<td>3 (33.3)</td>
<td>2 (22.2)</td>
<td>8 (88.9)</td>
<td>7 (77.8)</td>
<td>9 (100)</td>
<td>2 (22.2)</td>
<td>3 (33.3)</td>
<td>4 (44.4)</td>
<td>0 (0)</td>
<td>9 (100)</td>
<td>9 (100)</td>
<td>9 (100)</td>
<td>9 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>67/12</td>
<td>146</td>
<td>75 (51.4)</td>
<td>54 (37)</td>
<td>89 (61)</td>
<td>86 (58.9)</td>
<td>107 (73.3)</td>
<td>13 (8.9)</td>
<td>28 (19.2)</td>
<td>31 (21.2)</td>
<td>29 (19.9)</td>
<td>121 (82.9)</td>
<td>121 (82.9)</td>
<td>121 (82.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NA, Not applicable.
sample), 1.06 × 10^5 copies g⁻¹ for PBoV G3A (faecal sample), 3.89 × 10^5 copies g⁻¹ for PBoV G3B (faecal sample) and 3.13 × 10^5 copies g⁻¹ for PBoV G3C (faecal sample). In particular, for individual PBoV groups, higher viral loads in PCR-positive pigs were identified in faecal (PBoV G1 and G3C), lung (PBoV G2) and lymph node (PBoV G3A and G3B) samples (Table 2).

**VP1, VP2 and NS1 gene analysis**

The VP1, VP2 and NS1 genes of 42 PBoV-positive samples were amplified and cloned for sequencing. To account for the possibility of sequencing artefacts and to simplify the analysis, the sequences obtained from the same sample with >99.5% identity were considered to be the same and only one of them was selected for subsequent analysis. Seventy-three sequences based on the 1340 nt partial VP1 gene were obtained and used to construct a phylogenetic tree (Fig. S1). Among all recovered VP1 sequences, 97.3% (71/73) were unique, with a pairwise nucleotide distance ranging from 0.001 to 0.18, whereas 5.5% (4/73) of the sequences (IA34-1, IA18-2, IN109-1 and IA160-1) were 100% identical. Phylogenetic analysis revealed that these VP1 sequences formed two large clusters and each cluster could be divided into several clades. Some sequences from the same sample were classified as different clades or clusters (Fig. S1). Furthermore, even if the sequences belonged to the same branches, they showed a genetic distance ranging from 0.001 to 0.101 (smallest branches), which increased up to 0.175 (larger clusters) (Fig. S1). To further evaluate the genetic diversity of PBoVs within the NS1 and the VP2 genes, 62 partial NS1 (Fig. 1) and 61 partial VP2 sequences identified in this study were used to construct phylogenetic trees (data not shown). Pairwise distances of 1–24% (partial NS1 gene) and 3.4–32.4% (VP2 gene) were obtained. These results suggested that the presently circulating PBoVs exhibited considerable genetic diversity within the same sample and between different pigs (Table S2).

Partial VP1 and VP2 gene sequences of six PBoV G1- and 16 PBoV G2-positive samples were also amplified, cloned and sequenced. Alignment results showed that the amplified nearly full-length VP1 region (1678 nt) of PBoV G1 was conserved highly and displayed 97–99.8% sequence identity within this group. In contrast, alignment results showed that the amplified partial VP2 region (1310 nt) of PBoV G2 was less conserved (89.8–99.2% sequence identity within PBoV G2 in the USA and 90.3–98.4% to other PBoV G2).

**Full- or partial-genome sequence analysis**

Overall, 13 nearly full genomes of 4911–5243 nt and five partial genomes of 3803–4785 nt were obtained. Further sequencing of the terminal regions was hampered by formation of hairpin structures as described for other parvoviruses (Cheng et al., 2010; Kapoor et al., 2012). After aligning with other members of the genus *Bocavirus* with full- or near-full-length genomes available in GenBank, three distinct genetic lineages (G1–G3) were evident within the PBoV sequences obtained and those from GenBank. Three sublineages within G3 were revealed, instead of five subclades as proposed previously (Xiao et al., 2013b; Yang et al., 2012). This is likely in part because more sequences were available in the present study (Fig. 2). Clusters G1 and G2 shared a more recent common ancestry with HBoVs. However, cluster G3 formed a distinct, large clade, related distantly to clusters G1 and G2.

The genome organization of PBoVs identified in this study is similar to that of bocaviruses in other hosts, with three major ORFs encoding two non-structural proteins (NS1 and NP1) and two capsid proteins (VP1 and VP2). To further examine the newly identified PBoV genomes, the NS1, NP1 and VP1/2 genes of PBoV were compared with other available PBoV sequences (Table S2). Similar to a previous study (Allander et al., 2005), NS1 appeared to be the most conserved gene, whereas VP1/2 had the most nucleotide polymorphisms.

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Age group</th>
<th>PBoV G1</th>
<th>PBoV G2</th>
<th>PBoV G3A</th>
<th>PBoV G3B</th>
<th>PBoV G3C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymph node</td>
<td>NA</td>
<td>2.22 ± 0.28</td>
<td>4.31 ± 0.49</td>
<td>2.86 ± 0.82</td>
<td>3.17 ± 0.33</td>
<td>1.98 ± 0.45</td>
</tr>
<tr>
<td>Lung (sick pigs)</td>
<td>NA</td>
<td>1.89 ± 0.18</td>
<td>5.59 ± 0.55</td>
<td>1.22</td>
<td>1.67 ± 0.25</td>
<td>1.66 ± 0.83</td>
</tr>
<tr>
<td>Lung (healthy)</td>
<td>NA</td>
<td>0.3 ± 0.08</td>
<td>0</td>
<td>0.57</td>
<td>2.18 ± 0.15</td>
<td>1.13 ± 0.49</td>
</tr>
<tr>
<td>Serum</td>
<td>NA</td>
<td>0.32 ± 0.09</td>
<td>2.88 ± 0.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Faeces</td>
<td>Suckling</td>
<td>2.43 ± 0.41</td>
<td>2.03 ± 0.67</td>
<td>1.62</td>
<td>0.49</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Nursery</td>
<td>2.23 ± 0.19</td>
<td>3.14 ± 0.26</td>
<td>2.75 ± 0.45</td>
<td>2.44 ± 0.25</td>
<td>3.05 ± 0.20</td>
</tr>
<tr>
<td></td>
<td>Grow-finish</td>
<td>2.67 ± 0.21</td>
<td>3.12 ± 0.19</td>
<td>2.55 ± 0.15</td>
<td>2.11 ± 0.25</td>
<td>2.12 ± 0.15</td>
</tr>
<tr>
<td></td>
<td>Mature</td>
<td>2.18 ± 0.05</td>
<td>3.06 ± 0.15</td>
<td>3.38 ± 1.03</td>
<td>2.35 ± 0.50</td>
<td>2.13 ± 0.53</td>
</tr>
</tbody>
</table>

NA, Not applicable.

Table 2. Mean ± SE viral loads of tissue homogenate, serum or faecal suspensions of PBoVs in different sample types as determined by quantitative real-time PCR

Only PCR-positive pigs were included in the means.
Recently, Kapoor et al. (2010b) and Zeng et al. (2011) showed the presence of highly conserved potential RNA splicing signals in NS1 of all HBoV species and PBoV G1 SX. The presence of a possible splicing signal in other PBoVs was investigated by multiple alignments and comparison; however, a similar stretch of encoded amino acids as present in HBoVs was only identified in PBoV G1 MN307.

**Characterization of individual recombination events**

Six potentially significant recombination events were identified by analysing all obtained partial VP2 gene sequences, with a high degree of confidence ($P < 9.83 \times 10^{-3}$) judged by the eight recombination detection methods. These were further confirmed by phylogenetic trees using MEGA5 constructed on the non-recombinant and recombinant regions. Fig. 3 outlines the recombination event that occurred between OH110-1 (GenBank accession no. KF025395) and IA5F159-6 (GenBank accession no. KF025436), which led to the recombinant IA34-1 (GenBank accession no. KF025378). The bootstrap plot of this recombination event is shown in Fig. 3(a), which used OH110-1 and IA5F159-6 as the parental strains, leading to the recombinant IA34-1. To confirm this recombination event, the relevant strains were analysed by the neighbour-joining
method with MEGA5. Fig. 3(b, c) presents two trees constructed based on the non-recombinant (positions 259–1243) and recombinant regions (positions –258), respectively. In Fig. 3(b), IA34-1 clustered closely with OH110-1, whilst Fig. 3(c) shows discordant phylogenetic relationships compared with Fig. 3(b), where IA34-1 clustered with IA5F159-6. The phylogenetic analysis results confirmed the recombination event that occurred between OH110-1 and IA5F159-6, leading to the recombinant IA34-1. Moreover, a recombination event within a pig was also detected between IN5F109-5 and
IN5F109-4, which led to the recombinant cluster IN109-1. This recombination event, as well as four others, was further confirmed by phylogenetic trees (Fig. 4). Several significant recombination events were also identified in the partial VP1 gene \( (n=3) \), partial NS1 gene \( (n=4) \) and the whole genome \( (n=3) \), which were confirmed using MEGA5 by evolutionary trees constructed on the non-recombinant region and the recombinant region (data not shown).

**DISCUSSION**

Emerging parvoviruses have been discovered recently through the application of random PCR and next-generation sequencing techniques (viral metagenomic analysis), followed by bioinformatics analysis of large numbers of the sequences of the resulting clones (Allander et al., 2001, 2005; Shan et al., 2011a). In the present study, six new species and 12 new strains were identified and characterized by partial or nearly complete genome comparison.

To date, >11 PBoV species, including the species identified in the present study, have been characterized with partial or nearly full-length genomes. Although the phylogenetic trees constructed on different ORFs or the whole genome are incongruent, PBoVs clustered into three groups, named provisionally G1, G2 and G3. From the phylogenetic tree of the NS1 gene and genomes, PBoV G3 can be further divided into three subgroups: G3A, G3B and G3C.

Members of the genus Bocavirus are pathogens that have been associated with various disease manifestations, including respiratory and enteric disease (Allander et al., 2005; Chen et al., 2010; Kapoor et al., 2010a, b; Manteufel & Truyen, 2008). All four genotypes of HBoV have been identified in healthy children and adults or persons with diarrhoea, and HBoV1 and HBoV2 have been detected in respiratory tract samples and stool samples (Xu et al., 2012; Kapoor et al., 2010b). The pathogenicity of PBoVs has not been investigated extensively, although some reports suggested that some PBoVs may be associated with respiratory signs (Blomström et al., 2009; Zhai et al., 2010). Detection and differentiation of PBoVs is important to better understand the potential role of PBoVs in disease. In the present study, we identified three main PBoV groups G1, G2 and G3, with further differentiation of G3 into G3A, G3B and G3C.

PBoV displayed a lower prevalence rate in suckling pigs (Table 1), similar to reports in China (Zhai et al., 2010). PBoV G1 was detected with higher prevalence rates in weaning pigs (Cadar et al., 2011); likewise, HBoV tended to infect primarily infants <2 years of age with only a few adult infections (Allander et al., 2005; Fry et al., 2007; Xu et al., 2012).

All five targeted groups and subgroups of PBoV were detected in the clinical samples tested in this study, with prevalence rates ranging from 17.2 to 43.1 %. Any prevalence estimation is highly dependent on the sensitivity of the detection assay utilized and reasons for the low prevalence of PBoV G2 might include the less-optimal PCR sensitivity compared with the other groups. Nevertheless, our findings were comparable with the 63.2 % PBoV1-H18 (PBoV G1) and 64.4 % PBoV2-A6 (PBoV G2) positive rates in stool samples (Shan et al., 2011a) and 38.7 % PBoV1 (PBoV G1) in tissue samples (Zhai et al., 2010), although only one or two groups of PBoV were investigated in those studies.

PBoV G1 was found to be almost twice as prevalent in pigs affected by porcine circovirus-associated disease (PCVAD) than in non-PCVAD pigs in Sweden from 2003 to 2007 (Blomström et al., 2010). Similarly, PBoV in China was more prevalent in diseased pigs (Zhai et al., 2010). However, in the present study and another study (Lau et al., 2011), no significant difference was noted in the detection rate for PBoV G3A and PBoV G3B in faecal or lung samples from healthy and diseased pigs.

We could not rule out the possibility that diseases in these pigs were caused by other viruses, consistent with reports that the frequency of HBoV infection was higher when another virus was present (Lau et al., 2007; Manning et al., 2006). Further studies are required to determine whether PBoV plays a causative role, acts as an exacerbating factor that increases the severity of infections caused by other pathogens or is entirely non-pathogenic.

Parvoviral capsid proteins contain determinants of immunogenicity and tissue tropism. Minor genetic changes in these proteins are known to alter the host range and pathogenic potential of parvoviruses (Hoelzer et al., 2008a, b; Kapoor et al., 2012; Parrish & Kawaoka, 2005). Kapoor et al. (2012) observed that CBoV with a unique deletion in the VP2 gene was significantly more prevalent in animals with respiratory diseases compared with healthy animals. Similarly, residues 93 or 300 in the VP2 protein of canine parvovirus (CPV) bind to the cellular transferrin receptor, which determines the infectivity of CPV in dogs (Parker & Parrish, 1997). In the present study, the prevalence rates of PBoVs differed depending on sample type investigated (Table 1) and similar copy numbers for PBoVs were present in different sample types, except for PBoV G2 (Table 2), suggesting that PBoV has a wide tissue tropism. In serum samples, PBoVs was detected frequently, which...
(a) Tract of sequence with a recombinant origin IA5F76-1–IA5F282-2

(b) Tract of sequence with a recombinant origin IN5F109-7–IA5F34-10

(c) Tract of sequence with a recombinant origin IN5F109-5–IN5F109-1

(d) Tract of sequence with a recombinant origin IA5F34-2–IA5F34-8

(e) Tract of sequence with a recombinant origin IA5F34-2–IA5F109-4
sugests that the virus may be associated with systemic infection. Moreover, a high variation in the VP2 gene of PBoVs was detected in this study, suggestive of continuous or persistent infection. Sampling individual pigs over time should be done to understand if persistence occurs. In vitro propagation of these viruses and experimental inoculation of pigs needs to be done to better understand the pathogenic potential of PBoVs. Some progress has been made to culture successfully PBoV G3A and PBoV G3B in primary pig kidney cell lines (McKillen et al., 2011).

Evidence suggests that rapid virus evolution due to genome sequence variation is not restricted to RNA viruses and retroviruses (Grigoras et al., 2010). Recent research indicates that paroviruses have nucleotide substitution rates that are as high as those of some RNA viruses (Duffy et al., 2008; Shackelton et al., 2005). In this study, combined analyses of genetic diversity and PCR prevalence data indicate that PBoVs are genetically diverse and 18 novel species or genotypes with nearly complete genomes were identified. Similar results were also described in a recent report, although only two species were identified based on four samples from three sick pigs (Lau et al., 2011). The presence of highly similar sequence variants in the same pig (based on one sample collection) but seldom between pigs in that study implies that the observed diversity is due to newly arising mutations in a single strain or co-infection with different strains. These findings suggest that these viruses are in the process of adaptation and can undergo rapid evolution to generate new genotypes or species.

Reports have shown that inter-genotype and intra-genotype recombinations are present amongst bocaviruses (Fu et al., 2011). Currently, only recombination breakpoints bounding large tracts of exchanged sequence are identifiable, whilst complex recombinants may never be detected in nature (van der Walt et al., 2009). In this study, inter-species and intra-species recombination events were identified in the VP1, VP2 and NS1 genes and the whole genome. These findings are in line with a recent report that showed the presence of recombination among different gene regions by analysis of all the complete genomes of bocaviruses available in GenBank (Fu et al., 2011). High rates of co-infection of PBoVs observed in this study might increase the chance of recombination between PBoVs as detected in the present study. Although high mutation and recombination frequencies do not necessarily translate into high evolutionary rates, our results from genetic diversity and recombination analysis imply the huge evolutionary potential of PBoVs. This is further supported by the extensive host range and emerging new species of bocaviruses.

To our knowledge, this is the first study to investigate systematically the prevalence of PBoVs and report nearly complete genomes, genomic characteristics and genetic diversity of several novel PBoV species or strains in pigs in the USA. Our findings indicate that genetically diverse populations of PBoVs are circulating and many new PBoV species exist in pigs in the USA. The results of this study will provide a foundation for future work on PBoVs.

**METHODS**

**Specimen collection and DNA extraction.** A total of 365 pig samples, including serum, lymph node, faecal and lung samples, were selected randomly from routine diagnostic cases submitted to the Iowa State University Veterinary Diagnostic Laboratory during 2011–2012, with most of the pigs having a history of enteric or respiratory disease. These samples originated from 97 farms located in 14 states in the USA (Colorado, Iowa, Illinois, Indiana, Michigan, Minnesota, Missouri, North Carolina, North Dakota, Nebraska, Ohio, Oklahoma, Virginia and Wisconsin) with the ages of the pigs ranging from suckling to adult pigs. Twenty lung samples from healthy pigs were also included in this study (Table 1).

**Sample processing and viral nucleic acid extraction.** The lymph node, lung and faecal samples were processed as described previously (Xiao et al., 2013a). Briefly, tissue samples were minced and diluted 1:10 (w/v) in Dulbecco’s modified Eagle’s medium, homogenized using a Stomacher® 80 (Seward Laboratory Systems) and centrifuged at 1500 g for 10 min to obtain the supernatant. Faecal samples were resuspended 1:10 (w/v) in PBS, vortexed for 30 s and centrifuged at 1500 g for 10 min. Viral DNA extraction was done on 30 µl tissue, faecal supernatant or serum using the 5 x MagMAX* 96 Viral Isolation Kit (Ambion) according to the manufacturer’s instructions on an automated extraction platform (KingFisher Flex; Thermo Fisher Scientific). The extracted DNA was stored at −80 °C until use.

**Development of a multiplex real-time PCR assay to investigate the prevalence of PBoV G1, G2 and G3.** Initially, available nucleotide sequences of PBoVs were aligned by the Clustal W method (DNASTAR). Based on the alignment and phylogenetic tree of all the available nucleotide sequences of PBoVs, three distinct groups, G1 (represented by GenBank accession no. HQ223038), G2 (represented by GenBank accession no. HM053694) and G3 (represented by GenBank accession no. NC_016031), were identified, similar to the group classification described previously (Xiao et al., 2013b; Yang et al., 2012). To screen the samples for the presence of these three PBoV groups, a multiplex real-time PCR assay was developed (Table S3). To detect, differentiate and quantify the DNA of these three main PBoV groups in a single step, three pairs of degenerate primers and three different TaqMan probes were designed based on conserved VP1 regions (Table 3). The PCRs were carried out in 96-well plates. Each reaction consisted of a total volume of 25 µl, containing 12.5 µl QuantiTec Multiplex PCR Master Mix (Qiagen), 1.5 µl sample or standard DNA, 0.3 (G1), 0.4 (G2) and 0.2 (G3) μM primers, 0.2 μM each of the three probes and 5 µl distilled water. Standards were run in triplicate. Amplification
Table 3. Primers and probe information for PBoV detection and genome sequencing

<table>
<thead>
<tr>
<th>PBoV group and subgroup</th>
<th>Primer/probe sequence (5’−3’)*</th>
<th>Position</th>
<th>Reference sequence (GenBank accession no.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoV detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBoV G1</td>
<td>F: TTTCCTGTGGTACGCTCC</td>
<td>3085</td>
<td>HQ291308</td>
</tr>
<tr>
<td></td>
<td>R: CTTTAAGTCCACTTGCTGTG</td>
<td>3384</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: CAL Fluor Orange 560-CCTCTTTTGCCGCTGTTT-BHQ-1</td>
<td>3326</td>
<td></td>
</tr>
<tr>
<td>PBoV G2</td>
<td>F: GGAGGARACCCTGACTT</td>
<td>2559</td>
<td>HM053693</td>
</tr>
<tr>
<td></td>
<td>R: CCAGTTGAGAGGACACG</td>
<td>2698</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: FAM-AACCCCTTACAGGATTTCAGGCG-TAGGCCC-BHQ-1</td>
<td>2627</td>
<td></td>
</tr>
<tr>
<td>PBoV G3</td>
<td>F: GTACCAGCTCTATGATCTT</td>
<td>2959</td>
<td>NC_016031</td>
</tr>
<tr>
<td></td>
<td>R: AAAGGACCCCAARTAAATTAT</td>
<td>3190</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: Quasar 670-CGGAAYTGACCGCTCCATTA-BHQ-2</td>
<td>3134</td>
<td></td>
</tr>
<tr>
<td>PBoV G3A</td>
<td>F: CTCTGTGAGGCTGGA</td>
<td>616</td>
<td>NC_016031</td>
</tr>
<tr>
<td></td>
<td>R: GACGTTGATCTGATATTG</td>
<td>879</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: FAM-AACCCTTACAGGACGMRCA</td>
<td>753</td>
<td></td>
</tr>
<tr>
<td>PBoV G3B</td>
<td>F: AGGCTCTGACGACGCTG</td>
<td>931</td>
<td>JN681175</td>
</tr>
<tr>
<td></td>
<td>R: CCCAGCCAGTGTCGCA</td>
<td>1069</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: Quasar 670-CTCTTGGCAGGAAATAGGATGCTG-BHQ-2</td>
<td>1032</td>
<td></td>
</tr>
<tr>
<td>PBoV G3C</td>
<td>F: GCTCCGTCAGACCTTTGGT</td>
<td>471</td>
<td>JN621325</td>
</tr>
<tr>
<td></td>
<td>R: AGGCCTGTCGAGACGCGA</td>
<td>573</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: CAL Fluor Orange 560-TMGAACCGCTGATGCTGCA-BHQ-1</td>
<td>513</td>
<td></td>
</tr>
<tr>
<td>PBoV genome sequencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBoV G1</td>
<td>F1-1: TGCTGAAACCGCGTCAA</td>
<td>737</td>
<td>HQ291308</td>
</tr>
<tr>
<td></td>
<td>F1-2: TGGCAAGTGTGCCATAA</td>
<td>737</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1: AGCCATGCTGCGCCT</td>
<td>2167</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2: ATACGGACTGCTGTAAC</td>
<td>1906</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2: GCACCCTGTACCTCCCTCA</td>
<td>3550</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F3: CAAAAGGAGGACGGGGA</td>
<td>3334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3: TTGGTTGATTTGTCTTCTTGCC</td>
<td>5153</td>
<td></td>
</tr>
<tr>
<td>PBoV G2</td>
<td>F1: GCCATAWAGATGGAAATTATTTA</td>
<td>7</td>
<td>HM053693</td>
</tr>
<tr>
<td></td>
<td>R1: AACAGTTTTGGCGTAC</td>
<td>1935</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2: TGCAAGCTMTACGGCTG</td>
<td>1496</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2: TTTTCTTCTGCTGGGTTG</td>
<td>3023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F3: AAAAGGAGCTATCCCGG</td>
<td>2438</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3: TTYCCGAGGGAACAG</td>
<td>4120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F4: ACTGCTATCGTGCGCCTT</td>
<td>3742</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R4: TCGTGTCTCCGGGCTKT</td>
<td>5051</td>
<td></td>
</tr>
<tr>
<td>PBoV G3</td>
<td>F1-1: TAGGGGATACATGGCGGAGAT</td>
<td>30</td>
<td>NC_016031</td>
</tr>
<tr>
<td></td>
<td>F1-2: CGGGAGATTAATGATGACAC</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F1-3: TTTTCGGGACACGCCAATC</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2: TGGCAGGAGGCGATTTAC</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2: GGCTGACACCCTTATCTC</td>
<td>1549</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F3: CAGCCGATCGTTCATAT</td>
<td>1126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3-1: AAATCATCCAGAATCCACAT</td>
<td>3024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R3-2: AAATCATCCAGAATCCACAT</td>
<td>3024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F4: GATTTTACTTACCTCCCAAG</td>
<td>2802</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R4: SGTGATGTTAGGGGTATG</td>
<td>4133</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F5: GGTCTTTCACGTGTGCTGT</td>
<td>4011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R5: GGTAGGATATGGTCTTCT</td>
<td>5150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R6-1: TAGTCAACATGGCAGCTTT</td>
<td>5251</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R6-2: TTTGAAATCTGCTATAGAG</td>
<td>5233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R6-3: TTAAATAGTCAACATGAG</td>
<td>5256</td>
<td></td>
</tr>
</tbody>
</table>

*F, Forward; R, reverse; P, probe.
and quantification were performed using the ABI 7500 Fast Real-time PCR System (Applied Biosystems) under the following conditions: 2 min at 50 °C, 15 min at 95 °C, followed by 45 cycles of 15 s at 95 °C, 20 s at 43 °C and 1 min at 59 °C. A sample was considered negative if no cycle threshold (Ct) was detected in 40 amplification cycles. The PBoV genomic loads (g faecal/tissue sample)−1 of faecal or tissue samples were calculated as described previously (Xiao et al., 2013a).

Development of the multiplex real-time PCR assay to investigate the prevalence of PBoV G3 subgroups A, B and C in G3 DNA-positive samples. Based on the alignment and phylogenetic analysis of all partial NS1 nucleotide sequences of the 88 PBoVs used in this study, PBoV G3 was further subdivided into three subgroups: G3A (represented by PBoV3 SH20F, GenBank accession no. JF429834), G3B (represented by PBoV3C, GenBank accession no. JN681175) and G3C (represented by PBoV5 HB11, GenBank accession no. JN621325) (Table S3, Fig. 1). Three pairs of degenerate primers and three different TaqMan probes were designed based on the NS1 gene (Table 3). The multiplex real-time PCR was the same as described for the PBoV-G1-G2-G3 multiplex real-time PCR, except that 0.3 (G3A), 0.2 (G3B) and 0.3 (G3C) μM primers were utilized, and 0.2 (G3A), 0.2 (G3B) and 0.4 (G3C) μM of each probe was used. Amplification and quantification were performed under the following conditions: 2 min at 50 °C and 15 min at 95 °C, followed by 45 cycles of 15 s at 95 °C, 15 s at 61 °C and 45 s at 68 °C.

Positive controls, specificity and sensitivity of the multiplex real-time PCR assays. As initially there was no positive control available, five samples positive for PBoV G1, G2, G3A, G3B or G3C were chosen for reamplification. The PCR products were purified with the QIAquick Gel Extraction Kit (Qiagen) and then cloned into the pGEM-T easy vector (Promega). The recombinant plasmids were transformed into TOP10 Escherichia coli bacteria (Invitrogen) and propagated following the manufacturer’s instructions. The plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturers’ instructions, quantified using a spectrophotometer (Nanophotometer; IMPLEN) and sequenced. The confirmed plasmids containing the desired PBoV sequences were further used as standards for the multiplex differential real-time PCR assays to determine the specificity of the assays (Xiao et al., 2012) and to quantify viral loads in samples. The specificity of the probes was confirmed by BLAST analysis and by testing samples positive for other DNA viruses, including PCV1, PCV2, PPV1, PPV2, PPV3, PPV4, PPV5, TTSuV1 and TTSuV2.

Genome sequencing of PBoVs and phylogenetic analysis. To investigate the molecular characteristics of the PBoVs identified in the present study, 13 positive samples were selected randomly for genome sequencing. The DNA was amplified by degenerate primers designed from multiple alignments of the PBoV partial or nearly-full-length genomes available in GenBank (Table 3). Furthermore, additional primers covering the original degenerate primer sites were designed based on the resulting sequences obtained after the first and subsequent rounds of sequencing. The sequence PCRs were performed with high-fidelity LA Taq DNA polymerase (Takara) following the manufacturer’s instructions. The PCR products were purified and cloned into the pGEM-T easy vector as described above. The recombinant plasmids were sequenced in both directions either directly or by primer walking for fragments >1500 bp.

Sequence editing, assembly and comparison were performed to produce final sequences of the viral genomes by using BioEdit (Hall, 1999) and the SeqMan program (DNASTAR). The nucleotide sequences of the genomes and the predicted ORFs were compared to those of other PBoVs available in GenBank. Multiple alignments of the sequences were carried out using Clustal W (DNASTAR), genetic distances between sequences were determined with the P-distance model and phylogenetic analyses were carried out with MEGA5.0 (Tamura et al., 2011). The phylogenetic trees were constructed by using the maximum-likelihood method based on the general time-reversible (GTR) model under a bootstrap test of 1000 replicates (Tamura et al., 2011).

Recombination analysis. Selected PBoV sequences obtained in this study were screened for possible recombination using eight different programs available in the REB program software package with their default parameters, and only potential recombination events detected by three or more of the programs coupled with phylogenetic evidence of recombination were considered significant with a Bonferroni-corrected P value cut-off of 0.05 (Martin et al., 2010).

Nucleotide sequence accession numbers. The PBoV sequences reported here have been deposited in GenBank under the accession numbers KF025378–KF025571.

Statistical analysis. Differences in PBoV prevalence rates among sample types and age groups were investigated using an overall Fisher’s exact test for difference among all groups, followed by a post-hoc pairwise Fisher’s exact test with P values adjusted by the Bonferroni method. The results were considered significant at P<0.05. Viral loads were assessed using analysis of variance (ANOVA). If an ANOVA test was significant (P<0.05), then pairwise tests with Tukey’s adjustment were used to assess specific group differences. Statistical analysis was performed using Prism software version 5.00 (GraphPad).

Ethics statement. All samples utilized were selected arbitrarily and originated from pig case submissions to the Iowa State University Veterinary Diagnostic Laboratory for diagnostic work-up. Sample collection/submission was unrelated to and not part of this study. The protocol for this study was approved by the Iowa State University Institutional Biosafety Committee (permit no. 12-D/I-0038-A).

ACKNOWLEDGEMENTS

We thank the National Pork Board for funding this study.

REFERENCES


Schwarz, D., Green, B., Carmichael, L. E. & Parrish, C. R. (2002). The canine minute virus (minute virus of canines) is a distinct parvovirus that is most similar to bovine parovirus. Virology 302, 219–223.


Y.-H. Jiang and others


