Highly diverse posaviruses in swine faeces are aquatic in origin

Ben M. Hause,1,2 Rachel Palinski,1 Richard Hesse1,2 and Gary Anderson1,2

1Department of Diagnostic Medicine and Pathobiology, Kansas State University, Manhattan, KS 66506, USA
2Kansas State Veterinary Diagnostic Laboratory, Kansas State University, Manhattan, KS 66506, USA

Posaviruses are a group of highly divergent viruses identified in swine faeces that are distantly related to other members of the order *Picornavirales*. Eighteen posavirus genomes were assembled from 10 out of 25 (40 %) faecal-swab pools collected from healthy adult swine. Phylogenetic analysis of the conserved RNA-dependent RNA polymerase (Pol) domain found that posaviruses form a large, highly diverse, monophyletic clade, which includes similar viruses identified in human (husavirus) and fish (fisavirus) faeces or intestinal contents, respectively. Quantitative reverse transcription PCR analysis of water samples collected from commercial swine barns identified four out of 19 (21 %) samples were positive using a 5′-nuclease assay targeting the Pol region of posavirus 1. ICPD (immunoprecipitation coupled to PCR detection) assays to explore serological evidence of posavirus infection found only a single positive sample, suggesting posaviruses do not commonly infect swine, and together these results suggests a likely aquatic host.

Picornavirus-like viruses comprise an extremely diverse order (*Picornavirales*), whose members have been identified in vertebrates, plants, insects and algae (Le Gall et al., 2008). Metagenomic sequencing of environmental samples such as seawater have also found evidence for hundreds of unique viruses and numerous phylotypes distantly related to known members of *Picornavirales* (Culley et al., 2014). The order is subdivided into five families of which only *Picornaviridae* principally infects vertebrates (Sanfaçon et al., 2012). Two families (*Dicistroviridae* and *Iflaviridae*) infect invertebrates, one family infects plants (*Secoviridae*), while the sole member of *Marnaviridae* was identified in algae (Le Gall et al., 2008; Sanfaçon et al., 2012).

The genomes of most *Picornavirales* consist of a single RNA molecule; however, some members of *Secoviridae* are bipartite (Sanfaçon et al., 2012). Members of *Picornaviridae* are positive sense, non-enveloped RNA viruses encoding one ( *Iflaviridae*, *Secoviridae*, *Marnaviridae* and *Picornaviridae*) or two ( *Dicistroviridae* and *Secoviridae*) polyproteins, although exceptions have been noted (Sanfaçon et al., 2012; Woo et al., 2011). The polyproteins are autocatalytically processed and contain a conserved three domain replication block composed of a type III helicase (Hel), followed by a chymotrypsin-like proteinase (Pro) and a RNA-dependent RNA polymerase (Pol) (Le Gall et al., 2008). For picornaviruses with monocistronic genomes, the capsid-encoding regions of the genome are located near the 5′ end with the exception of *Marnaviridae*, which has structural proteins encoded near the 3′ end of the genome (Sanfaçon et al., 2012).

Metagenomic sequencing of swine faeces identified two viruses with predicted Pol domains similar to members of *Picornavirales* (Shan et al., 2011). Interestingly, the complete polyprotein sequences were most similar to picornavirus-like sequences identified in the nematode *Ascaris suum* isolated from a pig. Besides Pol, a type III helicase domain was identified upstream of the Pol domain; however, a centrally located Pro domain could not be identified. As posavirus cDNA could not be detected in *A. suum* DNA by PCR, and the nucleotide composition of both posavirus and *A. suum* were more similar to arthropods than plants or animals, the authors hypothesized that posaviruses infect nematodes parasitizing pigs (Shan et al., 2011). More recently, a novel posavirus with a Pol with only approximately 25 % amino acid similarity to posivirus 1 and 2 was also identified in swine faeces (Hause et al., 2015a). Similarly to posavirus 1 and 2, a Pro domain could not be identified; similar, however, in contrast, neither could a Hel domain. Phylogenetic analysis of the Pol and identification of two picornavirus capsid
domains near the 3′ terminus suggested inclusion of posavirus 3 in a new, unassigned family of Picornavirales (Hause et al., 2015a). Posavirus was also identified in faecal samples from pigs in China (Zhang et al., 2014).

Genomes similar to posaviruses were recently identified in faecal and intestinal samples of humans and fish, respectively (Oude Munnink et al., 2015; Reuter et al., 2015). In contrast to swine posaviruses, both fish (fisavirus) and human (husavirus) posaviruses-like genomes contained complete Hel–Pro–Pol replication blocks. Phylogenetic analysis of the Pol protein found that both fisavirus and husavirus were more closely related to posavirus 1 than posaviruses 2 and 3 (Oude Munnink et al., 2015).

To investigate viral flora present in pigs at areas of commingling, nasal and faecal swabs were each collected from 125 pigs. Samples were collected at four different slaughterhouses and one sale barn where pigs were bought and resold. Swabs were collected by a veterinarian and all animals were fit for slaughter. The swabs collected from each site were from pigs derived from five separate producers, which were commingled with pigs from other producers for less than 12 h. A total of five pigs were sampled per producer (5 sites × 5 producers per site × 5 pigs per producer = 125 nasal and 125 faecal swabs). Nasal- and faecal-swab pools were assembled from the five pigs from a single producer. A total of 25 nasal-swab and 25 faecal-swab pools were assembled and sequenced separately. Swab pools were clarified by centrifugation at 14 000 g and supernatants were treated with a nuclease cocktail to degrade unprotected nucleic acids (Neill et al., 2014). Viral RNA was extracted and then subjected to reverse transcription, second-strand synthesis and amplified by PCR as previously described (Hause et al., 2015b). Sequencing libraries were prepared with the Nextera XT kit and pooled barcoded libraries were sequenced on an Illumina MiSeq using paired 300 bp reads. Sequences were parsed by barcodes using CLC genomics. Approximately one million reads were determined for each barcoded library. Sequences were *de novo* assembled following subtraction of sequences mapping to *Sus scrofa*. Contigs greater than 6000 bp were analysed for ORFs encoding predicted proteins greater than 2000 aa in length. Putative posaviruses were identified by BLASTP and designated strains using their contig length as identifiers.

Contigs encoding proteins most similar to posavirus and fisavirus were identified in ten faecal-swab pools (Table 1). Four swab pools contained multiple putative posavirus genomes. A total of 18 putative posavirus genomes were identified and deposited in GenBank under accession numbers KT833062–KT833079. Four of the contigs encoded complete coding DNA sequences and portions of the UTRs were determined. BLASTP analysis found that only six of the posavirus genomes had high similarity (>68 % identity) to characterized posaviruses. The remaining genomes had only 28–35 % identity limited to only portions of the posavirus genome.

Conserved protein domains were identified by BLASTP (Marchler-Bauer et al., 2015). With the exception of strain 10611, a Hel domain was identified for all viruses near the 5′ terminus of the polyprotein (Table 2). For strains with sequence information for the initiation codon,
the Hel domain was located between aa 438 and 1260, and encompassed approximately 100 aa. Strain 10611 was 93% identical to a previously described posavirus 3 that also lacked an identifiable Hel domain (Hause et al., 2015a). Hel domain expectation (e) values ranged from 5.6 e^{-2} to 8.3 e^{-20}. As the Hel domain is part of the three domain replication module found in all Picornavirales, the absence of the Hel domain for 10611 is likely due to sequence divergence as opposed to its absence from the genome (Le Gall et al., 2008).

Previous characterization of posaviruses failed to identify a protease domain conserved amongst Picornavirales (Shan et al., 2011; Hause et al., 2015a). Seven of the posavirus genomes determined here had domains with homology to Picornavirales 3C proteases located downstream of the Hel domain. The Pro domain was approximately located between aa 1462 and 1704, and spanned 150 aa. Interestingly, Pro was only identified for the strains most closely related to husavirus or sequences derived from A. suum, which presumably harboured a posavirus-like virus, both of which had Pro domains (Fig. 1). It is likely that failure to detect the Pro domain in some strains is due to sequence divergence.

A Pol domain was identified by BLASTP for all 18 posavirus strains, and was located downstream of both Hel and Pro domains from aa 1695 to 2672, and encompassed approximately 400 aa. The conserved nature of the Pol domain was evident by more significant e values (1.2 e^{-29}-4.3 e^{-32}). For 15 of the 18 strains, two picornavirus capsid domains were located further downstream between Pol and the 3′ terminus (1.1 e^{-3}-9.5 e^{-21}). No capsid domains were identified for strain 6282, likely due to the incomplete genome sequence. Two additional strains (11038 and 9225) had only a single capsid domain and, interestingly, were closely related (Fig. 1). As the complete coding DNA sequence was not determined for either of these strains, it is unclear whether the lack of a second capsid domain is due to incomplete genome sequence or insufficient domain homology. Strain 11038 was the largest genome assembled and was incomplete. To our knowledge, this is the largest Picornavirales genome described.

Predicted Pol amino acid sequences were aligned using ClustaW and phylogenetic analysis was performed using the maximum likelihood algorithm using the best fitting LG +G+I model in MEGA 6.06 (Tamura et al., 2013). The posavirus sequences determined here were monophyletic and clustered with previously characterized posaviruses, fisaviruses, husaviruses and viruses presumed to have infected A. suum. The posavirus Pol sequences were extremely divergent, with 21–99% identity. The percent age identity between posavirus and members of Picornaviridae, Marnaviridae, Iflaviridae and Secoviridae was 13–24%, similar to identities between members of other Picornavirales families.

The posavirus genomes determined here where the Pro domain could be identified, as well as husaviruses and fisaviruses, possessed Hel–Pro–Pol domains followed downstream by multiple capsid domains. The remaining posavirus genomes here lacked identifiable Pro domains, but have similarly located Hel and Pol domains, with the exception of posavirus strain 10611, which lacked an identifiable Hel domain. Genomic organization of conserved Hel–Pro–Pol domains preceding capsid domains is similar to those observed in

<table>
<thead>
<tr>
<th>Strain</th>
<th>ORF (aa)</th>
<th>Helicase (aa)</th>
<th>Proteinase (aa)</th>
<th>Polymerase (aa)</th>
<th>Capsid (aa)</th>
<th>Capsid (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9470</td>
<td>2935</td>
<td>680–800</td>
<td>NI</td>
<td>1851–2255</td>
<td>2392–2560</td>
<td>2671–2845</td>
</tr>
<tr>
<td>9754</td>
<td>3070</td>
<td>650–770</td>
<td>NI</td>
<td>1823–2281</td>
<td>2364–2532</td>
<td>2644–2818</td>
</tr>
<tr>
<td>8295</td>
<td>2765</td>
<td>480–600</td>
<td>NI</td>
<td>1655–2059</td>
<td>2228–2364</td>
<td>2475–2649</td>
</tr>
<tr>
<td>6282</td>
<td>2094</td>
<td>571–682</td>
<td>1376–1513</td>
<td>1589–2044</td>
<td>NI</td>
<td>NI</td>
</tr>
<tr>
<td>8729</td>
<td>2909</td>
<td>1037–1141</td>
<td>1433–1583</td>
<td>1644–2113</td>
<td>2197–2376</td>
<td>2497–2672</td>
</tr>
<tr>
<td>7390</td>
<td>2463</td>
<td>1–100</td>
<td>1148–1181</td>
<td>1242–1712</td>
<td>1759–1938</td>
<td>2210–2391</td>
</tr>
<tr>
<td>10835</td>
<td>3546</td>
<td>438–540</td>
<td>NI</td>
<td>2238–2672</td>
<td>2851–2975</td>
<td>3086–3252</td>
</tr>
<tr>
<td>9225</td>
<td>3075</td>
<td>433–537</td>
<td>NI</td>
<td>2166–2603</td>
<td>2779–2903</td>
<td>NI</td>
</tr>
<tr>
<td>11038</td>
<td>3679</td>
<td>647–751</td>
<td>NI</td>
<td>2381–2818</td>
<td>3229–3395</td>
<td>NI</td>
</tr>
<tr>
<td>8805</td>
<td>2934</td>
<td>865–966</td>
<td>NI</td>
<td>1538–1942</td>
<td>2061–2226</td>
<td>2316–2495</td>
</tr>
<tr>
<td>10611</td>
<td>3100</td>
<td>–</td>
<td>NI</td>
<td>1751–2111</td>
<td>2184–2342</td>
<td>2434–2613</td>
</tr>
</tbody>
</table>

NI, Not identified.
Marnaviridae (Lang et al., 2004). Published work using targeted PCR and metagenomic sequencing found evidence for hundreds of unique, highly diverse Picornavirales phylotypes in a single sample of seawater (Culley et al., 2014). These viruses are thought to infect marine plankton. Interestingly, from the six complete viral genomes determined from seawater samples, three had monopartite genomes and were organized similarly to Marnaviridae, which was isolated from the marine algae Heterosigma akashiwo (Culley et al., 2014; Lang et al., 2004). Similarly to the majority of posaviruses characterized here, a protease domain could not be identified from genomes assembled from seawater (Culley et al., 2014).

Fig. 1. Phylogenetic analysis of the predicted RNA-dependent RNA polymerase protein domain of Picornavirales. Viral phylogeny was inferred using maximum likelihood analysis with tree topology assessed by 500 bootstrap replicates using MEGA 6.06 software. The bootstrap values are indicated by the branches and GenBank accession numbers are in parentheses.
An outstanding question on posavirus, husavirus and fisavirus is their host. While identified in stool or intestinal contents from pigs, humans and fish, it is unclear whether these viruses utilize these animals as hosts or rather infect microbes present in the gastrointestinal tract. Due to the similarity in genome architecture to Marnaviridae and unassigned ocean Picornavirales, we explored water collected from commercial pig farms (derived from wells on site) as the source of posavirus identified in pig faeces. Nineteen water samples were concentrated 100 x by centrifugal concentrators with 50 kDa filters and RNA was isolated. Quantitative reverse transcription PCR was performed using a 5'-nuclease assay designed based on the Pol region of posavirus 1. Four of the nineteen (21%) were positive with cycle threshold (Ct) values of 32.4-36.6.

To investigate possible direct infection of pigs, an immunoprecipitation-coupled to PCR detection (ICPD) assay was performed. Previously, ICPD analysis was used to determine the seroprevalence of porcine parainfluenza virus 1 in USA pigs with results in good agreement with those obtained with an ELISA using a recombinant fusion protein (Palinski et al., 2016). Swab pools 16 and 2 served as antigen for posavirus 1 and posavirus 3, respectively. Two 5'-nuclease assays were designed using the Pol region from the posavirus 1 and posavirus 3 genomes assembled from swab pools 16 and 2, respectively. ICPD was performed as previously described using sera collected from 10 gilts and 15 multiparous sows (parity ≥3) located in the Southeastern USA (Palinski et al., 2016). Ct values for the posavirus 1 and 3 antigen pools were 25.4 and 20.4, respectively. Only a single gilts serum sample was positive by ICPD with a Ct of 34.2 for posavirus 1. All other samples were negative, as were controls where serum was replaced with PBS. These results suggest that posavirus 1 and 3 do not commonly infect swine; however, additional testing is required to conclusively exclude posavirus infection of pigs.

Conserved genome features and phylogenetic analysis of the Pol domain suggest that posavirus, husavirus, fisavirus and viral sequences derived from A. suum represent a novel linage of Picornavirales. Previous work proposed that fisivirus andposavirus represented novel genera in a new family in the order Picornavirales (Reuter et al., 2015). Likewise, husavirus virus was proposed to be a member of a new viral family (Oude Munnink et al., 2015). Our results further support these proposals, as our phylogenetic analysis found that posavirus, husavirus and fisavirus are members of a monophyletic lineage that is only distantly related to other Picornavirales families. Several well-supported clades shown in Fig. 1 may represent genus distinctions; however, further research is needed to understand the biological significance of the observed genetic diversity and proper taxonomical classification.

This work demonstrates an incredible amount of Picornavirales diversity exists in swine faeces that is likely derived from aquatic viruses ingested by swine; however, further research is needed to establish their host. These results are consistent with surveys of a variety of aquatic environments, which have found that the RNA virome is principally composed of Picornavirales and that RNA virus abundance in these environments can exceed that of DNA viruses (Culley et al., 2006; Steward et al., 2013; Dijkeng et al., 2009; Rosario et al., 2009; López-Bueno et al., 2015). Similar to Marnaviridae, it is likely the host for posavirus replication is aquatic algae.

Acknowledgements

Funding was supported in part by the USDA's (United States Department of Agriculture's) Animal Health and Disease Research Program (under the provisions of section 1433 of subtitle E, title XIV of public law 95–113), the Kansas State Veterinary Diagnostic Laboratory, the State of Kansas National Bio and Agro-Defense Facility fund, and Boehringer Ingelheim.

References


