

Identification of an enterovirus recombinant with a torovirus-like gene insertion during a diarrhea outbreak in fattening pigs

Nádia Conceição-Neto,^{1,2,†,‡} Sebastiaan Theuns,^{3,†} Tingting Cui,³ Mark Zeller,¹ Claude Kwe Yinda,^{1,2} Isaura Christiaens,³ Elisabeth Heylen,¹ Marc Van Ranst,² Sebastien Carpentier,^{4,5} Hans J. Nauwynck,³ and Jelle Matthijnssens^{1,*}

¹Laboratory of Viral Metagenomics, Department of Microbiology and Immunology, Rega Institute for Medical Research, KU Leuven – University of Leuven, B-3000 Leuven, Belgium, ²Laboratory of Clinical Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, KU Leuven – University of Leuven, B-3000 Leuven, Belgium, ³Laboratory of Virology, Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, B-9820 Merelbeke, Belgium, ⁴Laboratory of Tropical Crop Improvement, Division of Crop Biotechnics, KU Leuven – University of Leuven, B-3000 Leuven, Belgium and ⁵Facility for Systems Biology Based Mass Spectrometry (SYBIOMA), KU Leuven – University of Leuven, B-3000 Leuven, Belgium

*Corresponding author: E-mail: jelle.matthijnssens@kuleuven.be

†These authors contributed equally to this work.

‡<http://orcid.org/0000-0002-9659-6835>

Abstract

Diarrhea outbreaks in pig farms have raised major concerns in Europe and USA, as they can lead to dramatic pig losses. During a suspected outbreak in Belgium of porcine epidemic diarrhea virus (PEDV), we performed viral metagenomics to assess other potential viral pathogens. Although PEDV was detected, its low abundance indicated that other viruses were involved in the outbreak. Interestingly, a porcine bocavirus and several enteroviruses were most abundant in the sample. We also observed the presence of a porcine enterovirus genome with a gene insertion, resembling a C28 peptidase gene found in toroviruses, which was confirmed using re-sequencing, bioinformatics, and proteomics approaches. Moreover, the predicted cleavage sites for the insertion suggest that this gene was being expressed as a single protein, rather than a fused protein. Recombination in enteroviruses has been reported as a major mechanism to generate genetic diversity, but gene insertions across viral families are rather uncommon. Although such inter-family recombinations are rare, our finding suggests that these events may significantly contribute to viral evolution.

Key words: recombinant virus; enterovirus; viral metagenomics; virome; virus evolution; bocavirus.

1. Introduction

Diarrhea is an important health problem affecting piglets, as well as an important cause of production losses in fattening pigs. Typical causes of clinical and subclinical enteric problems in the latter age group are bacteria such as *Brachyspira* spp., *Lawsonia intracellularis* and/or *Salmonella* spp., whereas viral causes at this age are generally less common (Ståhl et al. 2011; Collins and Barchia 2014; De Ridder et al. 2014). In a minority of cases, fattening pigs and sows can be affected by porcine epidemic diarrhea virus (PEDV), a re-emerging enteric coronavirus (Stevenson et al. 2013; Pensaert and Martelli 2016). Classical diagnostic methods such as specific antigen-, gene-, or antibody-detection assays are currently in use in veterinary practice in order to reach an etiological diagnosis and to put into practice targeted prophylactic and therapeutic measures. However, in recent years, a considerable progress has been made in the field of viral metagenomics, which makes it more affordable to apply next-generation sequencing (NGS) technologies to analyze the entire fecal viral content (the fecal virome) of a sample. Using this approach, the most frequently detected mammalian viruses in viral metagenomics studies were kobuviruses, rotaviruses, pig stool-associated ssDNA viruses, astroviruses, sapoviruses, and enteroviruses (Shan et al. 2011; Sachsenröder et al. 2012; Zhang et al. 2013).

In January 2015, a case of diarrhea in fattening pigs occurred on a Belgian pig farm. Diarrhea emerged 2 days upon arrival after transport and clinical signs in the herd (1,000 fattening pigs) lasted for 21 days, affecting a total of 20 fattening pigs. The veterinarian suspected an infection with the re-emerging PEDV. This virus belongs to the genus *Alphacoronavirus* within the family of *Coronaviridae*. The virus replicates in enterocytes and leads to sloughing of the gut villi, which causes diarrhea. It was widely spread across Europe between the 1970s and 1990s, causing epidemics of diarrhea on pig farms. However, since then it has only sporadically been detected (Pensaert and de Bouck 1978). In 2010, this virus was detected during severe outbreaks in Asia, and for the first time in the US swine population during spring 2013, causing mortality and severe economic losses (Song et al. 2015). In 2014, a milder variant of the virus (OH 855) was detected in the US swine herds (Wang et al. 2014). These strains were later also detected in German swineherds in 2014, as well as in other European countries (Grasland et al. 2015; Mesquita et al. 2015; Stadler et al. 2015). However, it is unclear if these genomic changes are correlating with an increased virulence of PEDV (Pensaert and Martelli 2016). At the beginning of 2015 and following the outbreaks of PEDV in USA, there was an increased awareness in Europe, including Belgium, as this virus could also emerge in the local swine population at any given moment. Therefore, when the outbreak of diarrhea occurred on the above-mentioned farm, a mixed fecal sample ($n = 12$ pigs) was sent to the Laboratory of Virology at Ghent University. PEDV RNA was detected using an in-house RT-qPCR, but high Cq values ($Cq > 30$) suggested a rather low viral load, whereas typical clinical infections in gnotobiotic piglets results in shedding of the virus at loads $>10^{10}$ copies per milliliter of feces (Jung et al. 2014). Still, a full genome characterization was performed and this indicated that the strain was genetically highly similar to the INDEL strains circulating in Germany and USA, as reported elsewhere (Theuns et al. 2015). Given the low abundance of PEDV in the fecal sample, and the large number of other viruses which can also cause gastro-intestinal disease in pigs, it was questioned in the present study if other viruses might have been present in the pig's fecal virome and if they

could have contributed to the emergence of diarrhea on this farm.

A well-established virus often found in the feces of pigs is the porcine enterovirus (also known as Enterovirus G), belonging to the *Picornaviridae* family and genus *Enterovirus*. Enterovirus G encloses 16 viral types and has been detected in North America, Europe, and Asia (Boros et al. 2012; Anbalagan et al. 2014; Van Dung et al. 2014, 2016). Picornaviruses are positive-sense single-stranded RNA viruses, which have one open reading frame (ORF) that encodes for a single polyprotein. This polyprotein yields typically a Leader (L) protein, four structural (VP1-VP4), and seven non-structural proteins (2A^{pro}, 2B, 2C, 3A, 3B, 3C^{pro}, and 3D^{pol}) upon cleavage by proteases (Ehrenfeld et al. 2004).

A chymotrypsin-like cysteine proteinase, 3C^{pro}, is the main cleavage protein found in all picornaviruses and contains a conserved cleavage site aiding in the identification of cleavage sites of newly discovered viruses (Lendeckel and Hooper 2009). The 2A^{pro} is another protease encoded by entero- and rhinoviruses, that folds spontaneously into an active form, performing a primary cleavage (Ehrenfeld et al. 2004; Lendeckel and Hooper 2009).

Picornavirus evolution is determined by their high mutation rate, which is predicted to range between 10^{-3} and 10^{-5} mutations per nucleotide, per genomic replication (Domingo and Holland 1997). In addition, the often reported recombination events are crucial to shape their genomic architecture (Simmonds 2006).

Another example highlighting the importance of proteases for viral pathogenesis can be found in the *Coronaviridae* family. Specially studied for severe acute respiratory syndrome coronavirus, its PL^{pro} is also a deubiquitinating enzyme (Ratia et al. 2006). Therefore, it is important in disrupting the host cellular ubiquitination machinery, which leads to enhanced viral replication (Ratia et al. 2006).

In this study, we focus initially on unraveling the gut virome of the diarrhea outbreak in a fattening pig farm, and as a consequence of our findings we also further characterized an unusual recombinant enterovirus genome.

2. Material and Methods

2.1 Collection and processing of fecal sample

A mixed diagnostic fecal sample of fattening pigs ($n = 12$) from Belgium was sent to the Laboratory of Virology (Ghent University, Merelbeke, Belgium) in January 2015. No healthy controls were collected at the moment of the outbreak. The sample was prepared using a slightly modified version of the NetoVIR protocol (Conceição-Neto et al. 2015). A 10% weight/volume fecal suspension in viral transport medium (DMEM, 10% P/S, 5% gentamycin, and 0.01% fungizone) was prepared from sample 15V010 and filtered through 0.45 μm membrane filters (Millipore). The filtrate was treated with a cocktail of Benzonase (Novagen) and Micrococcal Nuclease (New England Biolabs) at 37 °C for 2 h to digest free-floating nucleic acids, in homemade buffer (1M Tris, 100mM CaCl₂, and 30mM MgCl₂, pH = 8). RNA and DNA were extracted using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions but without addition of carrier RNA to the lysis buffer.

2.2 Random amplification, library preparation, and Illumina sequencing

First- and second-strand synthesis and random PCR amplification were performed for 17 cycles using a modified Whole

Transcriptome Amplification (WTA2) kit procedure (Sigma-Aldrich). Denaturation temperature was increased to 95 °C to allow for the denaturation of dsDNA and dsRNA. WTA2 products were purified with MSB Spin PCRapace spin columns (Stratec) and were prepared for Illumina sequencing using the Nextera XT library preparation kit (Illumina). Libraries were quantified with the KAPA Library Quantification kit (Kapa Biosystems) and sequencing of the samples was performed on a HiSeq™ 2500 platform (Illumina) for 300 cycles (150bp paired ends), generating 67,251,870 reads.

2.3 Sequence and phylogenetic analysis

Raw reads were filtered and trimmed for quality and adapters using Trimmomatic (Bolger et al. 2014) and assembled using SPAdes assembler version 3.5.0 (Bankevich et al. 2012). Scaffolds were taxonomically classified using DIAMOND (sensitive option) (Buchfink et al. 2015). ORFs were identified with ORF Finder analysis tools, Pfam was used to help predict enterovirus proteins and HMMER to infer insertion similarities (Zhang and Wood 2003; Finn et al. 2011). Amino-acid alignments of the viral sequences were performed with MUSCLE implemented in MEGA6.0 (Edgar 2004). Substitution models for maximum likelihood phylogenetic trees were calculated using MEGA6.0 (Tamura et al. 2013), and the appropriate best substitution model (with the lowest AIC) was used to build phylogenetic trees with 500 bootstrap replicates.

2.4 RT-PCR and Sanger sequencing to confirm the presence of the torovirus insertion

A reverse-transcription polymerase chain reaction (RT-PCR) was performed on the original sample using the QIAGEN OneStep RT-PCR kit (Qiagen) using primer sequences shown in Table 1 (primers were designed to cover the breakpoint between the enterovirus and the torovirus-like sequence). The reaction was performed as follows: 50 °C for 30 min followed by a PCR activation step at 95 °C for 15 min, 40 cycles of amplification: 30 s at 94 °C, 30 s at 55 °C, and 2 min at 72 °C, and a final extension step for 10 min at 72 °C in a Biometra T3000 thermocycler (Biometra). PCR products were run on a polyacrylamide gel, stained with ethidium bromide, and visualized under UV light. Samples were then purified with ExoSAP-IT (Affymetrix), and positive products were Sanger sequenced with the ABI PRISM BigDye Terminator cycle sequencing reaction kit (Applied Biosystems). The raw chromatograms are provided in Supplementary Files S1 and S2, and the alignments used to generate Fig. 1 are also provided as Supplementary Data.

2.5 Design of synthetic peptides to perform targeted proteomics on the enterovirus–torovirus breakpoint and insertion

To be able to perform targeted analysis of the proteins, synthetic peptides were designed to determine the preferred *m/z* and retention time. Synthetic peptides covering the enterovirus–torovirus breakpoint, the torovirus insertion and the enterovirus were designed using Skyline 3.5.0 (MacLean et al. 2010). Predicted peptides containing Methionine and Cysteine were excluded.

2.6 Protein extraction

Extraction of proteins from the fecal sample was carried out as previously described by Carpentier et al. (2005) and Buts et al.

(2014). In short, 350 µl of fecal suspension were resuspended in 350 µl of ice-cold extraction buffer [50 mM Tris–HCl pH 8.5, 5 mM EDTA, 100 mM KCl, 1% w/v DTT, 30% w/v sucrose; complete protease inhibitor cocktail (Roche Applied Science)] and vortexed for 30 s. Seven hundred microliters of ice-cold Tris buffered phenol (pH 8.0) were added and the sample was vortexed for 10 min at 4 °C. After centrifugation (10 min, 12,000×*g*, 4 °C), the phenolic phase was collected, re-extracted with 350 µl of extraction buffer and vortexed for 30 s. After centrifugation (5 min, 12,000×*g*, 4 °C), the phenolic phase was collected and precipitated overnight with five volumes 100 mM ammonium acetate in methanol at –20 °C. After centrifugation at 16,000×*g* for 30 min at 4 °C, the supernatant was removed and the pellet was rinsed twice in ice-cold acetone/0.2% DTT. Between the two rinsing steps, the sample was incubated for 60 min at –20 °C. The pellet was air-dried, resuspended in 75 µl of lysis buffer (8 M urea, 5 mM DDT, 30 mM Tris DTT), and vortexed for 5 min at room temperature. Then the protein concentration was determined using the 2-Dquant kit from Amersham Biosciences.

2.7 Protein digestion

DTT was added to 20 µg of protein extract up to 20 mM of final concentration and incubated for 15 min. Then, iodoacetamide was added to the mixture up to a final concentration of 50 mM and incubated for 30 min in the dark. Then the sample was diluted 3 times in 150 mM ammonium bicarbonate. For protein digestion, 0.2 µg/µl trypsin was added and incubated overnight at 37 °C. Samples were acidified with trifluoroacetic acid (0.1% final concentration) and purified with Pierce C18 spin columns (Thermo Scientific) according to the manufacturers' instructions. Peptides were eluted with 40 µl and then evaporated using a speedvac.

2.8 Liquid chromatography and mass-spectrometry (MS/MS) analysis

Lyophilized peptide samples from the speedvac were then dissolved in a 0.1% v/v formic acid (FA) and 5% v/v ACN solution. This was followed by liquid chromatography (LC) coupled to a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific) in positive ion mode through a nanoelectrospray ion source (Thermo Scientific).

Peptides were separated on an Ultimate 3000 UPLC system (Dionex, ThermoScientific) equipped with an Acclaim PepMap100 pre column (C18 3 µm–100 Å, Thermo Scientific) and an C18 PepMap RSLC (2 µm, 50 µm_15 cm, ThermoScientific) using a linear gradient (300 µl/min) of 0–4% buffer B (80% ACN, 0.08% FA) in 3 min, 4–10% in 12 min, 10–35% in 20 min, 35–65% in 5 min, 65–95% in 1 min 95% for 10 min, 95–5% in 1 min, 5% 10 min. Then the separated peptides were analyzed in the Orbitrap QE operated in positive ion mode (nanospray voltage 1.5 kV, source temperature 250 °C). The instrument was first operated in data-dependent acquisition (DDA) mode on the pool of synthetic peptides ordered (PEPotec SRM Custom Peptide Libraries, Thermo Scientific) and an inclusion list based on the *m/z* and RT of the synthetic peptides detected was created (Supplementary Tables S1 and S2). For DDA, MS scans were performed at a resolution of 70,000 fwhm for the mass range of *m/z* 400–1,600 for precursor ions, followed by MS/MS scans of the top 10 most intense peaks with +2, +3, and +4 charged ions above a threshold ion count of 16,000 at 35,000 of resolution. MS/MS was performed using normalized collision energy (NCE) of 25% with an isolation window of 3.0 *m/z*, an

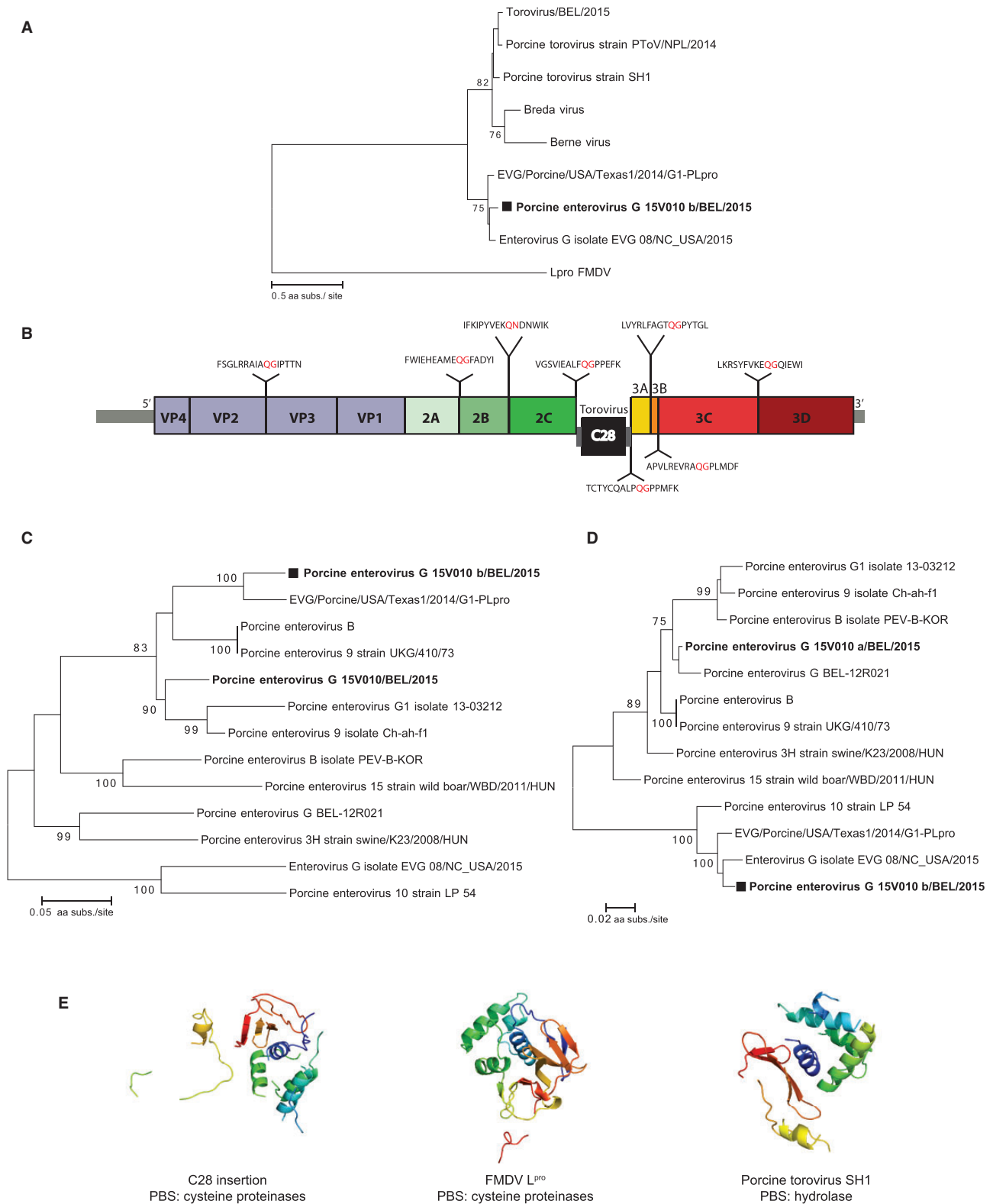


Figure 1. (A) Maximum likelihood phylogenetic tree with 500 bootstraps of the C28 gene insertion. (B) Genome organization of the enterovirus-torovirus recombinant BEL/15V010 and cleavage site prediction. (C and D) Maximum likelihood phylogenetic trees with 500 bootstraps of enteroviruses before and after the gene insertion. Only bootstrap values even or > 70 are shown. The two complete enteroviruses found in this study are shown in bold and the strain with the torovirus insertion is depicted with a square. (E) Structure prediction of the insertion protein, FMDV L^{pro}, and porcine torovirus C28.

Table 1. Primers used to confirm the breakpoints in the enterovirus–torovirus recombinant virus.

Name of the primer	Sequence	Position	Targeted sequence
ET_4062_Fw	5'-GTTGCAGCAACCCAGCC-3'	4062	Enterovirus
ET_447_Rv	5'-TTACACACCCACTTGACAGG-3'	4447	Torovirus
ET_4369_Fw	5'-AGCTTATCCACCAAGGAGCC-3'	4369	Torovirus
ET_5016_Rv	5'-AGTCTTCTCTCATCTACTGGG-3'	5016	Enterovirus

apex trigger 5–15 s, and a dynamic exclusion of 10 s. Data were acquired with Xcalibur 2.2 software (Thermo Scientific). Then the sample and synthetic peptides were re-ran in MRM/MSMS mode (Supplementary Table S2).

2.9 Peptide identification data analysis

For peptide identification, raw MS files were converted into mgf files by Proteome Discover version 1.4 (Thermo Scientific) and processed using Sequest (Eng et al. 1994) (HT version 1.3) against a customized in-house built database with sequences from all enterovirus sequences in UniProt and the recombinant enterovirus sequence. A parent mass tolerance of 10 ppm, a fragment tolerance of 0.02 Da, a variable modification oxidation of M, a fixed modification with carbamidomethyl C and up to one missed cleavage for trypsin were used. Common MS contaminants, such as human keratin and pig trypsin were used as decoy.

2.10 Structure prediction

The amino acid sequence of the novel inserted torovirus-like protein was analyzed using Phyre2 by comparing it against the homologous C28 protein of FMDV and the torovirus SH1 protein (Kelley et al. 2015). The three-dimensional predicted models of the latter two proteins were derived from the Protein Data Bank (PDB) and confidence of prediction was retrieved.

3. Results

3.1 Overview of the virome content of fecal sample 15V010

At the end of January 2015, PEDV was identified in a pooled diarrheic fecal sample of fattening pigs from a Belgian pig farm. Using an in house RT-qPCR assay, a low viral load of PEDV was detected, indicating that other viruses may have played a more important role in the pathogenesis of diarrhea in these pigs. Therefore, it was aimed to unravel the complete fecal virome of this pooled sample using an NGS approach. After trimming, a total of 57,164,223 NGS reads were assembled *de novo* into contigs (Bankevich et al. 2012; Bolger et al. 2014). A total of 38,118,443 reads were assigned with DIAMOND (Buchfink et al. 2015), of which 5,937,747 were assigned as viruses (16%). From these reads, 2,454,031 could be assigned to the order of the *Caudovirales* (41%) and 537,279 (9%) to the family *Microviridae*, which both contain exclusively bacteriophages. Only 2,780 NGS reads (0.1%) were obtained for PEDV, which corroborates the low viral load (Cq > 30) found by RT-qPCR. A total of 2,246,215 reads could be attributed to (near) complete mammalian porcine virus genomes (Table 2). Of these, reads matching the porcine bocavirus (2,110,362 reads, 34%) were far most dominant, followed by 63,317 reads matching porcine enteroviruses. All viruses described hereafter were deposited in GenBank and accession numbers are provided in Table 2.

3.2 Discovery and characterization of a chimeric enterovirus with a torovirus-like gene insertion

In this sample, 71,088 reads were assigned to the genus *Enterovirus*, of which 2 complete genomes could be retrieved. Interestingly, a contig of 8,043 nt in length was identified and assigned to enterovirus, with a partial in-frame torovirus-like gene insertion. Remapping of curated reads against this recombinant sequence using BWA (Li and Durbin 2009) as well as re-sequencing the insertion and breakpoints using Sanger sequencing on the original sample, confirmed that the obtained sequence was genuine and neither an artifact of the *de novo* assembly nor a result of random amplification. In Fig. 1B, a description of the genome organization of this novel recombinant enterovirus genome (Porcine enterovirus b 15V010/BEL/2015) is shown. As depicted in Fig. 1B, the torovirus-like gene insertion of 636 nt was present between the enterovirus 2C and 3A non-structural proteins. The inserted region formed a phylogenetic outgroup most closely related to several toroviruses. In the sample, we observed the presence of a small Torovirus contig of 256 bp, which was included in the phylogenetic tree (Fig. 1A, Torovirus/BEL/2015), showing very low similarity with the gene insertion found in the recombinant virus. Initial BLASTp searches attributed the highest aa similarity of this torovirus-like insertion to a porcine torovirus (66% aa similarity with NC_022787), but with the most recent reports of 2 similar recombinant viruses, it now shares 92.0% aa identity with the insertion of the recombinant enterovirus G from North Carolina (08/NC_USA/2015) (Shang et al. 2017) and 90.7% aa identity with the strain from Texas (EVG/Porcine/USA/Texas1/2014) (Knutson et al. 2017). To infer whether the insertion was produced as a separate moiety or fused with the upstream or downstream enterovirus protein, we identified potential polyprotein cleavage sites, which were present up- and downstream of the insertion (Fig. 1B). Pfam searches for motifs using hidden Markov models, revealed that the insertion seems to enclose an L-protease from the C28 family (e-value = $1.3e^{-05}$). We also attempted to infer the function of the insertion using Phyre2 (Kelley et al. 2015), which predicted a cysteine proteinase function (99.1% confidence, Fig. 1E). The same function was also predicted for the FMDV L^{pro}, but not for the porcine torovirus that showed the highest sequence homology with the insertion (Fig. 1E).

The enterovirus genome region before the torovirus insertion (VP1–VP4, 2A, 2B, and 2C) showed its highest similarity on the amino-acid level (94.5%) with EVG/Porcine/USA/Texas1/2014 (Fig. 1C). Downstream of the torovirus insertion (3A, 3B, 3Cpro, and 3Dpol), the virus showed its highest similarity (98.0%) on the amino-acid level with EVG 08/NC_USA/2015 (Fig. 1D). Even though enteroviruses and toroviruses possess linear ssRNA(+) genomes, they belong to two different viral orders, namely *Picornavirales* and *Nidovirales*, respectively.

3.3 Confirmation of the presence of a torovirus-like insert in the fecal sample using targeted proteomics

To further investigate the presence of this highly unusual recombinant virus, we designed synthetic peptides from the

Table 2. Number of reads attributed to the complete genomes described in this study.

Virus	Number of reads mapped	GenBank accession number
Porcine bocavirus BEL/15V010	2,110,362	KY426738
Porcine astrovirus 2 BEL/15V010	22,585	KY214438
Porcine astrovirus 4 BEL/15V010	8,996	KY214437
Porcine enterovirus b BEL/15V010 (Enterovirus–torovirus recombinant)	58,842	KY214435
Porcine enterovirus a BEL/15V010	4,475	KY214436
Porcine feces-associated circular virus BEL/15V010	3,572	KY214434
Porcine stool associated gemycircularvirus BEL/15V010	421	KY214433
Porcine picobirnavirus BEL/15V010/1966	7,948	KY214429
Porcine picobirnavirus BEL/15V010/2296	1,733	KY214426
Porcine picobirnavirus BEL/15V010/2605	1,043	KY214427
Porcine picobirnavirus BEL/15V010/3161	1,908	KY214428
Porcine picobirnavirus BEL/15V010/4445	8,437	KY214430
Porcine picobirnavirus BEL/15V010/4815	2,412	KY214431
Porcine picobirnavirus BEL/15V010/5682	2,753	KY214432
Porcine epidemic diarrhea virus BEL/15V010 (partial)	2,780	KR003452

Table 3. Synthetic and natural peptides identified from the sample SRM on the Enterovirus–torovirus.

Sequence	m/z ^a of the synthetic peptide	m/z ^a of the sample peptide	Retention time of the synthetic peptide	Retention time of the sample peptide
ALNYAEEQQVPR	709.35645	709.36096	28.15 min	28.15 min
EPSVQGLIYR	581.31647	581.31403	30.51 min	29.11 min

^aPeptides are detected as peaks which are characterized by mass-over-charge ratio (m/z) and retention time (rt).

recombinant virus, which we then used for selected reaction monitoring (SRM). These synthetic peptides were predicted to be generated after trypsin digestion. This frequently used approach in mass-spectrometry, allows to focus on the detection of a pre-selected group of peptides. The instrument was run on SRM mode on synthetic peptides of the insertion and on the peptide extracted fecal sample. Using SEQUEST, we were able to identify two peptides from the insertion region of the enterovirus–torovirus recombinant (Table 3 and Supplementary Tables). Table 3 shows the retention time and mass-over-charge ratio (m/z) of both the synthetic peptides and the homologues found in the original sample.

3.4 Characterization of bocavirus

Bocaviruses are ssDNA viruses belonging to the family *Parvoviridae*, which encode an additional ORF, named NP1, which is absent in the genome of other parvoviruses. In this study, a porcine bocavirus (BEL/15V010) that showed 99% identity to the South Korean strain PBoV-KU14 was identified (Fig. 2). The latter porcine bocavirus was identified in pigs with respiratory problems, which revealed a truncated NP1 gene, resulting in the shortest described porcine bocavirus genome thus far (Yoo et al. 2015). The novel bocavirus identified in this study also presented a truncated NP1 gene, thought to be caused by cross-over recombination (Yoo et al. 2015). This suggests that this virus strain might be more widespread than initially thought. Furthermore, the high number of reads matching to this bocavirus in the sample (≈ 2.1 M) suggests an acute and active replication.

3.5 Other novel viruses present in the fecal sample

Apart from the recombinant enterovirus found in the sample, another 12,246 viral reads could be attributed to other enteroviruses. From these, 4,475 reads could be used to assemble a

complete porcine enterovirus genome (Porcine enterovirus a 15V010/BEL/2015). The complete viral polyprotein showed its highest similarity on the amino-acid level (94.8%) with the porcine enterovirus 9 isolate Ch-ah-f1 (Zhang et al. 2012), which was found in 8.3% of screened pigs in China from 2007 to 2009. Using recombination detection approaches (program Rdp v4) (Martin et al. 2015), no recombination event was detected for this strain, when comparing to all complete sequences of porcine enteroviruses (data not shown). However, the previously described Belgian strain 12R021 (Theuns et al. 2016) seems to have gone through a recombination event, which could explain the distinct clustering in Fig. 1C.

In this study, we also report the presence of one porcine astrovirus type 2 and one astrovirus type 4 (Fig. 3A). These single-stranded positive sense RNA viruses, encode for a capsid polyprotein and a non-structural polyprotein. Association of porcine astroviruses with gastrointestinal disease is yet to be made, as they have been often reported as coinfections with rotavirus, coronavirus, and calicivirus (Laurin et al. 2011). The porcine astrovirus 2 (BEL/15V010) RNA-dependent RNA polymerase (RdRp) showed its highest similarity (94.2%) on the amino acid level with a recently described astrovirus in a non-diarrheic Belgian piglet co-infected with rotavirus (Theuns et al. 2016). For the porcine astrovirus 4 (BEL/15V010), the highest similarity (96.0%) was observed with a Hungarian wild boar astrovirus (Reuter et al. 2012), which also forms a common genetic lineage with other porcine astrovirus type 4 strains.

Furthermore, we identified 4 picobirnavirus capsid segments (Fig. 3E) and 3 RdRp segments (Fig. 3D). Picobirnaviruses are bisegmented double-stranded RNA viruses, belonging to the *Picobirnaviridae* family. Segment 1 encodes for a capsid protein and ORF1 with unknown function, while segment 2 encodes for the RdRp gene (Ganesh et al. 2012). Up to date, only seven complete sequences for porcine picobirnaviruses are available in

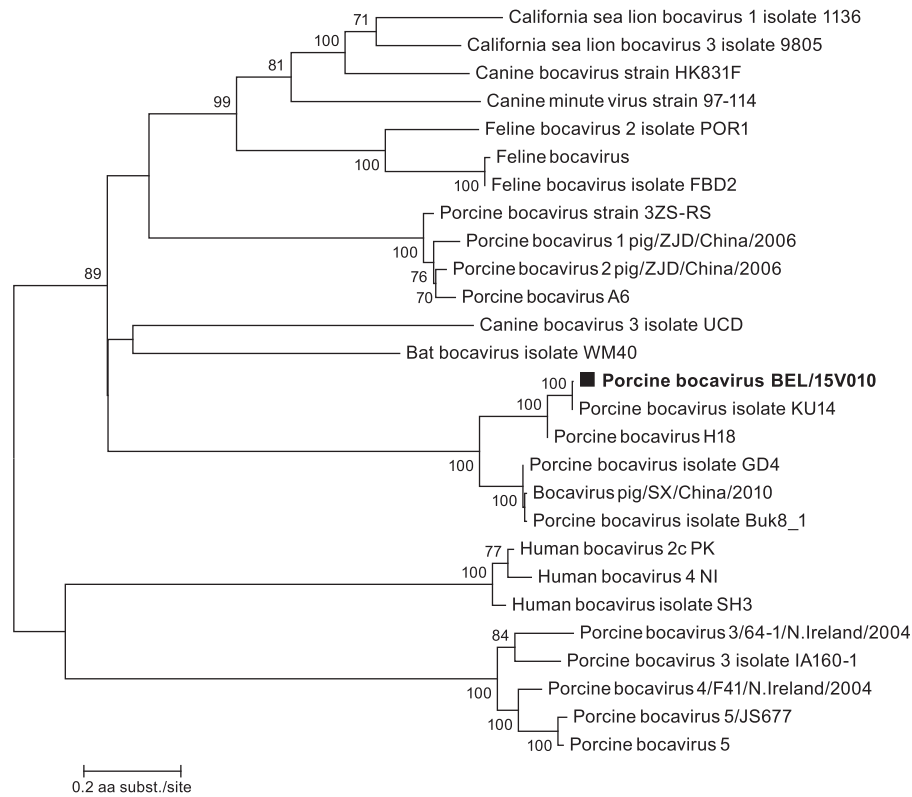


Figure 2. Maximum likelihood phylogenetic tree of the NS1 protein of bocaviruses. Bootstrap values even or > 70 are shown. The sequence identified in this study is shown in bold with a black square.

GenBank. The identified RdRp segments in the pig sample were shown to be rather divergent, as porcine picobirnavirus BEL/15V010/5682 and porcine picobirnavirus BEL/15V010/4815 showed their highest similarities of 75.9% and 70.9% on the amino-acid level, with a picobirnavirus found in a Californian sea lion. The porcine picobirnavirus BEL/15V010/4445 RdRp showed 93.8% similarity with dromedary picobirnavirus c3512 on the amino-acid level. Regarding the capsid genes even lower similarities were observed: picobirnavirus BEL/15V010/2296 shared its highest similarity (34.4%) with dromedary picobirnavirus c1591; porcine picobirnavirus BEL/15V010/2605 only shared 29.8% with dromedary picobirnavirus c5908; porcine picobirnavirus BEL/15V010/1966 was most similar (29.8%) with dromedary picobirnavirus c2033 and porcine picobirnavirus BEL/15V010/3161 was 33.2% similar to picobirnavirus GI/turkey.

Finally, two complete novel ssDNA circular viruses were detected: a gemycircularvirus and a circularvirus (Fig. 3B). Figure 3B shows the genome organization of the novel gemycircularvirus and the phylogenetic analysis of known gemycircularviruses based on the amino acid level of the Rep gene. The porcine gemycircularvirus BEL15V015 presented the typical nonanucleotide stem loop motif (TATAAATAG) and rolling circle replication motifs I (LFTYS), II (HLHVFAD), III (YATKD), GRS (RKFDVEGFHPNIVPSL) and helicase motifs Walker-A (GRSRTGKT) and Walker-B (VFDDI). The novel porcine gemycircularvirus replicase shares its highest similarity on the amino-acid level (76%) with Mongoose feces-associated gemycircularvirus b (Conceição-Neto et al. 2015) (Fig. 3B). The divergent ssDNA circular virus (Fig. 3C) encodes for two ORFs, bidirectionally transcribed: porcine stool associated circular virus BEL/15V010. The complete genome of the virus identified showed its highest similarity with porcine stool associated circular virus 5 isolate CP3 (82.9% nt). Interestingly, the capsid gene

of the virus shows 91.3% nt similarity with isolate CP3 but only 65.2% similarity with the replicase. However, these two viruses did not cluster closely with other circular single-stranded DNA viruses and seem to belong to a new isolated clade (Fig. 3C) (Cheung et al. 2014). Even though only two viruses from this clade were described so far, both were isolated from diarrheic porcine feces.

4. Discussion

Diarrhea is a major concern in farms leading to growth impairment. In the last years, major awareness was raised in Europe due to the emergence of PEDV in USA (Stevenson et al. 2013). Another recent study characterized a completely novel mammalian orthoreovirus, which was able to induce 100% mortality in experimentally infected piglets, further raising awareness that other viruses than those from the typical diagnostic lists can cause severe problems (Thimmasandra Narayanappa et al. 2015). In fact, NGS forms a powerful tool for diagnostics. Its application will lead to a better identification of viral enteric disease complexes and will allow investigating the relevance of coinfections of different enteric viruses, which may have remained unnoticed using traditional diagnostic techniques. However, it should be noted that this will lead to an abundance of information generated, which veterinarians and farmers may not be able to cope with without proper guidance.

Here, an outbreak of diarrhea in a Belgian pig farm was reported, 2 days after the arrival of new pigs, suggesting that these novel pigs were the source of infection. Since PEDV was being reported in Europe at the time, it was first diagnostically tested for. Even though the sample was positive for PEDV (RT-qPCR), the viral loads were rather low ($C_q > 30$). Our study does not provide evidence for direct causality of disease by a single

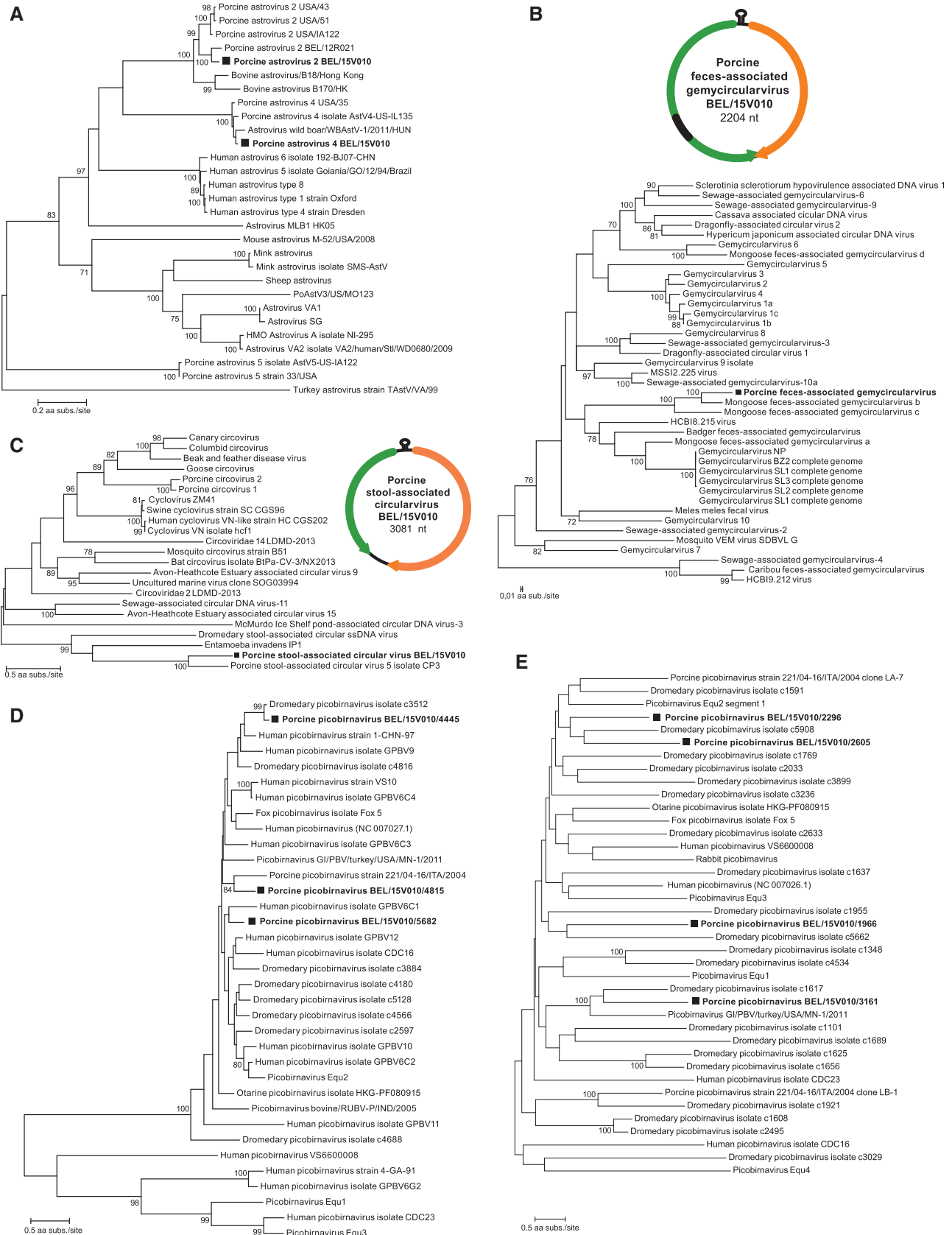


Figure 3. Maximum likelihood phylogenetic tree of the RdRp protein of astroviruses (A), the REP protein of gemycircularviruses (B), REP protein of circularviruses (C), the RdRp protein (D), and capsid (E) sequences of picobirnaviruses. Bootstrap values even or > 70 are shown. Sequences identified in this study are shown in bold with a black square.

viral agent, and the high number of viral reads attributed to a porcine bocavirus, the presence of other possible causative agents of diarrhea such as astroviruses, enteroviruses, and picobirnaviruses, merely indicate that more work needs to be performed to elucidate the role of coinfections in gastrointestinal disease. Moreover, none of these viral species are routinely tested for at diagnostic laboratories in cases of diarrhea on pig farms. Alongside these findings, an enterovirus with an insertion of a torovirus was described. Due to the large number of viral reads attributed to bocavirus and the enterovirus–torovirus recombinant, we attempted to isolate them using primary porcine kidney epithelial cells and ST (swine testis) cells (data not shown). Although CPE was observed 2 days after inoculation on porcine kidney epithelial cells, neither bocavirus nor the enterovirus–torovirus recombinant could be detected using PCR assays on any of the passaged cells. This is in contrast with the recently described recombinant virus from North Carolina, USA, where Shang et al. (2017) could isolate the virus using the ST cell line. We speculate that due to the large amount of additional non-recombinant enteroviruses in the sample, this might have put the recombinant enterovirus in an *in vitro* disadvantage. Using Pfam and Hmmer searches for motifs, the insertion was predicted to enclose an L-protease from the C28 family. This protein has been best studied for foot-and-mouth disease virus (FMDV), a member of the *Picornaviridae* and is known to cleave host cell proteins, namely the p220 subunit of eukaryotic initiation factor 4F (eIF-4F). The cleavage of this initiation factor 4F results in the shutoff of cap-dependent host cell protein synthesis, without affecting viral protein synthesis which can occur in the presence of cleaved p220 (Piccone et al. 1995). Moreover, our structural analysis confirmed that the function prediction from the structure was a cysteine proteinase (for both our insertion and L^{Pro} of FMDV), even though their relatedness in sequence identity is low (Fig. 1E). Structure prediction is often important to determine function when sequence similarities are lower. The predicted cleavage sites for the virus could be identified, suggesting that the inserted gene is producing a separate protein. Even though a small contig of 256 bp of Torovirus from the same insertion region could be detected in our pooled fecal sample, it clustered very distantly from the gene inserted (Fig. 1A). Since the fecal sample resulted of a pool of 12 pig feces, this is not surprising and likely has a different host origin. In addition to confirming the presence of the enterovirus–torovirus recombinant using Sanger sequencing, we used proteomics to infer whether the protein of the insertion could be found in the sample. This is specially challenging since fecal samples are a very complex matrix. In addition, the torovirus insertion codes for a non-structural protein, which are only present inside infected cells, which might explain why only a few of the predicted peptides could be identified. In recent years, the field of proteomics has evolved greatly, in fact, SRM has proven to be a powerful technique to detect and quantify proteins (Picotti and Aebersold 2012). This is especially suitable for detecting low abundant peptides, since the mass spectrometer focuses on detecting a preselected group of peptides. Taking the metagenomics data, the Sanger sequence confirmation and the proteomics results all together we can conclude that this recombinant virus is present in the sample and that the inserted protein is being expressed. Interestingly, two recent studies described porcine enteroviruses with a similar insertion (Fig. 1A) (Knutson et al. 2017; Shang et al. 2017). One of these viruses was isolated and a knockout mutant virus without the insertion yielded impaired growth and higher expression levels of innate immune genes in infected cells (Shang et al. 2017). Our analyses suggests that

these viruses have a common ancestor (Knutson et al. 2017) even though a significant diversity was noted among them (>90% aa identity) (Fig. 1A). Moreover, the fact that our strain before the insertion clusters with the virus isolated in Texas and after the insertion with the virus isolated from both Texas and North Carolina, can hint that an additional recombination occurred after the insertion event. Even though our recombinant strain is the first reported in Europe, it is very likely that such viruses might be more widely spread in the pig population, as the *in vitro* experiments of Shang et al. (2017) indicate that the virus might induce higher pathogenesis.

While recombinations between enteroviruses are frequently being described (Ren et al. 2012), recombination events between different viral families are more scarcely reported among the virosphere. For example, it is established that coronaviruses encode a gene derived from ancestral influenza C virus (Zhang et al. 1992). With advances and democratization of NGS, these events are more likely to be picked up, as was also recently shown in a study in bats identifying a recombinant bat coronavirus with an inserted reovirus gene (Huang et al. 2016).

Not unexpectedly, a great number of other eukaryotic viruses were identified using viral metagenomics. However, even though the pathogenic role of PEDV has been well demonstrated, for other viruses this link is less clear. For instance, the novel porcine circular virus described in this study is highly similar to another virus also identified in diarrheic pigs (Cheung et al. 2014). Therefore, it still remains a possibility that these viruses infect pigs and might play a role in diarrhea, which needs further elucidation. Gemycircularviruses, on the other hand, have been identified in a great variety of hosts, and no link with diarrhea has been yet demonstrated. It is likely that these viruses have a diet or plant origin, especially because the virus found in the pig sample clusters together with two gemycircularviruses found in the feces of a healthy mongoose (Conceição-Neto et al. 2015). As for the picobirnaviruses, in animals they have been identified in diarrheic and non-diarrheic samples (Ganesh et al. 2014). The main issue with segmented viruses is to link the different segments to the same virus. In our case, we identified four capsid and three RdRp genes, and further research is needed to link the different segments to their respective RdRp. As capsid segments are more divergent than RdRp, it is unexpected that more capsid than RdRp segments were identified. However, it cannot be excluded that there were more picobirnaviruses present in the pool and/or a reassortment event occurred, as previously described (Conceição-Neto et al. 2016).

As a conclusion, this study raises awareness for the presence of many viruses in a porcine diarrheic fecal sample. In fact, it is important to consider the possibility of diarrhea as a result of the replication of a viral intestinal disease complex, where more than one agent might play a role. Given the availability and democratization of next- and third-generation sequencing technologies, this will certainly change the way diagnostics are being performed in the coming decades. However, one should bear in mind that the abundance of information generated with these methods is not easy to interpret and appropriate tools should be developed to help farmers and the pig industry in translating this information into useful management information.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

NCN and ST were supported by Flanders Innovation & Entrepreneurship (VLAIO Vlaanderen). The Laboratory of

Virology of Ghent University is supported by VLAIO Vlaanderen (Agriculatory Trajectory 140955). TC is supported by a doctoral grant of the Chinese Scholarship Council (CSC). KCY was supported by the Interfaculty Council for Development Cooperation (IRO) from the KU Leuven. We would like to thank Kusay Arat for excellent technical assistance.

Conflict of interest: None declared.

References

- Anbalagan, S., Hesse, R. A., and Hause, B. M. (2014) 'First Identification and Characterization of Porcine Enterovirus G in the United States', *PLoS One*, 9: e97517.
- Bankevich, A. et al. (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology*, 19: 455–77.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.
- Boros, Á. et al. (2012) 'Characterization of a Novel Porcine Enterovirus in Wild Boars in Hungary', *Archives of Virology*, 157: 981–6.
- Buchfink, B., Xie, C., and Huson, D. H. (2015) 'Fast and Sensitive Protein Alignment using DIAMOND', *Nature Methods*, 12: 59–60.
- Buts, K. et al. (2014) 'Improving the Identification Rate of Data Independent Label-Free Quantitative Proteomics Experiments on Non-Model Crops: A Case Study on Apple Fruit', *Journal of Proteomics*, 105: 31–45.
- Carpentier, S. C. et al. (2005) 'Preparation of Protein Extracts from Recalcitrant Plant Tissues: An Evaluation of Different Methods for Two-Dimensional Gel Electrophoresis Analysis', *Proteomics*, 5: 2497–507.
- Cheung, A. K. et al. (2014) 'Identification of a Novel Single-Stranded Circular DNA Virus in Pig Feces', *Genome Announcements*, 2: e00347–14.
- Collins, A. M., and Barchia, I. M. (2014) 'The Critical Threshold of *Lawsonia intracellularis* in Pig Faeces That Causes Reduced Average Daily Weight Gains in Experimentally Challenged Pigs', *Veterinary Microbiology*, 168: 455–8.
- Conceição-Neto, N. et al. (2015) 'Fecal Virome Analysis of Three Carnivores Reveals a Novel Nodavirus and Multiple Gemycircularviruses', *Virology Journal*, 12: 79.
- et al. (2015) 'Modular Approach to Customise Sample Preparation Procedures for Viral Metagenomics: A Reproducible Protocol for Virome Analysis', *Scientific Reports*, 5: 16532.
- et al. (2016) 'Reassortment Among Picobirnaviruses Found in Wolves', *Archives of Virology*, 161: 2859–62.
- De Ridder, L. et al. (2014) 'Use of a Live Attenuated *Salmonella enterica* serovar Typhimurium Vaccine on Farrow-To-Finish Pig Farms', *Veterinary Journal*, 202: 303–8.
- Domingo, E., and Holland, J. J. (1997) 'RNA Virus Mutations and Fitness for Survival', *Annual Review of Microbiology*, 51: 151–78.
- Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.
- Ehrenfeld, E. et al. (2010) *The Picornaviruses*. Washington DC, USA: ASM Press.
- Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) 'An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database', *Journal of the American Society for Mass Spectrometry*, 5: 976–89.
- Finn, R. D., Clements, J., and Eddy, S. R. (2011) 'HMMER Web Server: Interactive Sequence Similarity Searching', *Nucleic Acids Research*, 39: W29–W37.
- Ganesh, B. et al. (2012) 'Picobirnavirus Infections: Viral Persistence and Zoonotic Potential', *Reviews in Medical Virology*, 22: 245–56.
- , Masachessi, G., and Mladenova, Z. (2014) 'Animal Picobirnavirus', *VirusDisease*, 25: 223–38.
- Grasland, B. et al. (2015) 'Complete Genome Sequence of a Porcine Epidemic Diarrhea S Gene Indel Strain Isolated in France in December 2014', *Genome Announcements*, 3: e00535–15.
- Huang, C. et al. (2016) 'A Bat-Derived Putative Cross-Family Recombinant Coronavirus with a Reovirus Gene', *PLoS Pathogens*, 12: e1005883.
- Jung, K. et al. (2014) 'Pathology of US Porcine Epidemic Diarrhea Virus Strain PC21A in Gnotobiotic Pigs', *Emerging Infectious Diseases*, 20: 662–5.
- Kelley, L. A. et al. (2015) 'The Phyre2 web Portal for Protein Modeling, Prediction and Analysis', *Nature Protocols*, 10: 845–58.
- Knutson, T. P., Velayudhan, B. T., and Marthaler, D. G. (2017) 'A Porcine Enterovirus G Associated with Enteric Disease Contains a Novel Papain-Like Cysteine Protease', *Journal of General Virology*.
- Laurin, M.-A., Dastor, M., and L'Homme, Y. (2011) 'Detection and Genetic Characterization of a Novel Pig Astrovirus: Relationship to Other Astroviruses', *Archives of Virology*, 156: 2095.
- Lendeckel, U., and Hooper, N. M. (2009) *Viral Proteases and Antiviral Protease Inhibitor Therapy: Proteases in Biology and Disease*. Netherlands: Springer.
- Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform', *Bioinformatics*, 25: 1754–60.
- MacLean, B. et al. (2010) 'Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments', *Bioinformatics*, 26: 966–8.
- Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: 1–5.
- Mesquita, J. R. et al. (2015) 'Outbreak of Porcine Epidemic Diarrhea Virus in Portugal, 2015', *Transboundary and Emerging Diseases*, 62: 586–8.
- Pensaert, M. B., and de Bouck, P. (1978) 'A New Coronavirus-Like Particle Associated with Diarrhea in Swine', *Archives of Virology*, 58: 243–7.
- , and Martelli, P. (2016) 'Porcine Epidemic Diarrhea: A Retrospect from Europe and Matters of Debate', *Virus Research*, 226: 1–6.
- Piccione, M. E. et al. (1995) 'The Foot-and-Mouth Disease Virus Leader Proteinase Gene is Not Required for Viral Replication', *Journal of Virology*, 69: 5376–82.
- Picotti, P., and Aebersold, R. (2012) 'Selected Reaction Monitoring-Based Proteomics: Workflows, Potential, Pitfalls and Future Directions', *Nature Methods*, 9: 555–66.
- Ratia, K. et al. (2006) 'Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease: Structure of a Viral Deubiquitinating Enzyme', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 5717–22.
- Ren, L. et al. (2012) 'Sequencing of a Porcine Enterovirus Strain Prevalent in Swine Groups in China and Recombination Analysis', *Veterinary Microbiology*, 159: 265–8.
- Reuter, G. et al. (2012) 'Astrovirus in Wild Boars (*Sus scrofa*) in Hungary', *Archives of Virology*, 157: 1143–7.

- Sachsenröder, J. et al. (2012) 'Simultaneous Identification of DNA and RNA Viruses Present in Pig Faeces Using Process-Controlled Deep Sequencing', *PLoS One*, 7.
- Shan, T. et al. (2011) 'The Fecal Virome of Pigs on a High-Density Farm', *Journal of Virology*, 85: 11697–708.
- Shang, P. et al. (2017) 'A Naturally Occurring Recombinant Enterovirus Expresses a Torovirus Deubiquitinase', *Journal of Virology*, 91: e00450–17.
- Simmonds, P. (2006) 'Recombination and Selection in the Evolution of Picornaviruses and Other Mammalian Positive-Stranded RNA Viruses', *Journal of Virology*, 80: 11124–40.
- Song, D., Moon, H., and Kang, B. (2015) 'Porcine Epidemic Diarrhea: A Review of Current Epidemiology and Available Vaccines', *Clinical and Experimental Vaccine Research*, 4: 166–76.
- Stadler, J. et al. (2015) 'Emergence of Porcine Epidemic Diarrhea Virus in Southern Germany', *BMC Veterinary Research*, 11: 142.
- Ståhl, M. et al. (2011) 'The Use of Quantitative PCR for Identification and Quantification of *Brachyspira pilosicoli*, *Lawsonia intracellularis* and *Escherichia coli* Fimbrial Types F4 and F18 in Pig Feces', *Veterinary Microbiology*, 151: 307–14.
- Stevenson, G. W. et al. (2013) 'Emergence of Porcine Epidemic Diarrhea Virus in the United States: Clinical Signs, Lesions, and Viral Genomic Sequences', *Journal of Veterinary Diagnostic Investigation*, 25: 649–54.
- Tamura, K. et al. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, 30: 2725–9.
- Theuns, S. et al. (2015) 'Complete Genome Sequence of a Porcine Epidemic Diarrhea Virus from A Novel Outbreak in Belgium, January 2015', *Genome Announcements*, 3.
- et al. (2016) 'Characterization of a Genetically Heterogeneous Porcine Rotavirus C, and Other Viruses Present in the Fecal Virome of a Non-Diarrheic Belgian Piglet', *Infection, Genetics and Evolution*, 43: 135–45.
- Thimmasandra Narayanappa, A. et al. (2015) 'A Novel Pathogenic Mammalian Orthoreovirus from Diarrheic Pigs and Swine Blood Meal in the United States', *mBio*, 6.
- Van Dung, N. et al. (2014) 'Prevalence, Genetic Diversity and Recombination Of Species G Enteroviruses Infecting Pigs in Vietnam', *Journal of General Virology*, 95: 549–56.
- et al. (2016) 'Large-Scale Screening and Characterization of Enteroviruses and Kobuviruses Infecting Pigs in Vietnam', *Journal of General Virology*, 97: 378–88.
- Wang, L., Byrum, B., and Zhang, Y. (2014) 'New Variant of Porcine Epidemic Diarrhea Virus, United States, 2014', *Emerging Infectious Diseases*, 20: 917–9.
- Yoo, S. J. et al. (2015) 'A Novel Porcine Bocavirus Harbors a Variant NP Gene', *Springerplus*, 4: 370.
- Zhang, Q. et al. (2013) 'Occurrence and Investigation of Enteric Viral Infections in Pigs with Diarrhea in China', *Archives of Virology*, 158: 1631–6.
- Zhang, W. et al. (2012) 'Complete Genome Sequence of a Novel Porcine Enterovirus Strain in China', *Journal of Virology*, 86: 7008–9.
- Zhang, X., Kousoulas, K. G., and Storz, J. (1992) 'The Hemagglutinin/Esterase Gene of Human Coronavirus Strain OC43: Phylogenetic Relationships to Bovine and Murine Coronaviruses and Influenza C Virus', *Virology*, 186: 318–23.
- Zhang, Z., and Wood, W. I. (2003) 'A Profile Hidden Markov Model for Signal Peptides Generated by HMMER', *Bioinformatics*, 19: 307–8.