Virus Genes (2020) 56:696–704
https://doi.org/10.1007/s11262-020-01791-z

ORIGINAL PAPER

Increased viral read counts and metagenomic full genome characterization of porcine astrovirus 4 and *Posavirus* 1 in sows in a swine farm with unexplained neonatal piglet diarrhea

Steven Van Borm[1] · Kevin Vanneste[2] · Qiang Fu[2] · Dominiek Maes[3] · Alexandra Schoos[3] · Eline Vallaey[4] · Frank Vandenbussche[1]

Received: 19 June 2020 / Accepted: 21 August 2020 / Published online: 3 September 2020

Abstract

Neonatal diarrhea in piglets may cause major losses in affected pig herds. The present study used random high-throughput RNA sequencing (metagenomic next generation sequencing, mNGS) to investigate the virome of sows from a farm with persistent neonatal piglet diarrhea in comparison to two control farms without diarrhea problems. A variety of known swine gastrointestinal viruses was detected in the control farms as well as in the problem farm (*Mamastrovirus*, *Enterovirus*, *Picobirnavirus*, *Posavirus* 1, *Kobuvirus*, *Proprismacovirus*). A substantial increase in normalized viral read counts was observed in the affected farm compared to the control farms. The increase was attributable to a single viral species in each of the sampled sows (porcine astrovirus 4 and *Posavirus* 1). The complete genomes of a porcine astrovirus 4 and two co-infecting *Posavirus* 1 were *de novo* assembled and characterized. The 6734 nt single-stranded RNA genome of porcine astrovirus 4 (*PoAstV-4*) strain *Belgium/2019* contains three overlapping open reading frames (nonstructural protein 1ab, nonstructural protein 1a, capsid protein). Posavirus 1 strains *Belgium/01/2019* and *Belgium/02/2019* have a 9814 nt single-stranded positive-sense RNA genome encoding a single open reading frame (polyprotein precursor) containing the five expected Picornavirales-conserved protein domains. The study highlights the potential of mNGS workflows to study unexplained neonatal diarrhea in piglets and contributes to the scarce availability of both PoAstV-4 and *Posavirus*-1 whole genome sequences from Western Europe.

Keywords Metagenomics · Sus scrofa · Posavirus · Astrovirus · Next generation sequencing

Introduction

Diarrhea is one of the leading causes of morbidity and mortality in piglets and constitutes one of the major economic infectious disease burdens for pig producers. Diarrheal disease in neonates causes a rapid decline in clinical condition as the piglet becomes dehydrated and develops metabolic acidosis [1]. Both bacterial and viral factors have been associated with diarrhea problems in young suckling piglets [2–7].

The increased availability of high-throughput sequencing platforms to research and diagnostic laboratories has highlighted the added value of studies sequencing the complete nucleic acid content of samples (metagenomics) for the characterization of pathogens, the study of multifactorial disease [8, 9], and the identification of novel infectious agents both in human and animals [10–12].

Edited by William Dundon.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11262-020-01791-z) contains supplementary material, which is available to authorized users.

✉ Steven Van Borm
Steven.Vanborm@sciensano.be

[1] Department of Animal Infectious Diseases, Sciensano, Brussels, Belgium

[2] Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium

[3] Department of Reproduction, Obstetrics and Herd Health, Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium

[4] VetWorks BVBA, Poeke, Belgium

A recent systematic review of metagenomics studies in farm animals identified the families *Astroviridae*, *Caliciviridae*, *Coronaviridae*, and *Picornaviridae* as the most frequently documented intestinal virome constituents of pigs with gastrointestinal symptoms, followed by *Anelloviridae*, *Circoviridae*, and *Herpesviridae* [13]. Examples of frequent metagenomically documented virus genera in the swine intestinal tract of healthy and diseased animals include *Bocaparvovirus* (*Parvoviridae*), *porcine circovirus* (*Circoviridae*), *Deltacoronavirus,* and porcine epidemic diarrhea virus or *Alphacoronavirus* (PEDV, *Coronaviridae*), *Rotavirus* (*Reoviridae*), *Sapelovirus, Enterovirus, Kobuvirus, Teschovirus and Pasivirus* (*Picornaviridae*), *Sapovirus* (*Caliciviridae*), *Mamastrovirus* (*Astroviridae*), and Torque teno sus virus (*Anelloviridae*) [13–16]. In addition, high-throughput sequencing-based viral metagenomics studies have identified novel porcine stool-associated Picornaviruses (*Posavirus*, porcine stool-associated virus) with a global distribution [17–20] that are likely associated to the presence of gut commensals or parasites [18]. Metagenomic and virus characterization studies in Belgian swine farms have identified *Kobuvirus* (*Picornaviridae*), *Mamastrovirus* (*Astroviridae*) [21], *Rotavirus* (*Reoviridae*) [22, 23], *Enterovirus* (*Picornaviridae*) [24], and PEDV (*Coronaviridae*) [25] as constituents of the swine fecal virome, mostly focusing on fecal material sourced from fattening pigs and piglets. Targeted screening has identified bacterial species including *Escherichia coli*, *Clostridium perfringens*, *Salmonella* sp., and/or *Brachyspira* sp. as co-infection in *Rotavirus* positive diarrheic samples from Belgian swine herds [26].

The present study investigates the virome of sows (as potential reservoirs from which viruses may infect sensitive piglets) in a Belgian swine production farm with unexplained persisting neonatal diarrhea in comparison with sows from two farms without diarrhea problems. The complete genome of a porcine astrovirus 4 and a *Posavirus 1* is described, the latter consisting of a mixed infection of two viruses.

Material and methods

Study population

Fresh rectal fecal samples were obtained from sows in sterile containers, transported at 4 °C to the laboratory and stored at − 80 °C until nucleic acid extraction. Samples were obtained from two separately housed sows from two swine farms showing no clinical symptoms of neonatal diarrhea ("control farm 1" using a 4-week batch production system and housing 230 sows; "control farm 2" using a 3-week batch production system and housing 350 sows). A sample was obtained from two separately housed sows from a farm ("problem farm", using a 5-week batch production system and housing 500 sows) with increased (> 10%) incidence of neonatal diarrhea issues not due to infections with pathogens well known to potentially cause diarrhea in young piglets such as *E. coli*, *C. perfringens*, or *Rotavirus* infections. Former analyses proved only sporadic presence of hemolytic *E. coli*, *C. perfringens* Type A, and *Rotavirus*. Furthermore, administration of antimicrobials active against *E. coli* (e.g., gentamycin, colistin) or Clostridium (e.g., amoxycilline) (within 24 h of birth) had little effect. The implementation of an autogenous vaccine against *E.coli* F4(K88)/O8:K87 and *C. perfringens* Type A did not confer the expected results. Additional bacteriological and virological analyses (RT-PCR for *Enterobacteriaceae* of feces from affected and non-affected piglets) could not identify the pathogens causing the problem.

Sequencing

Feces were homogenized in sterile phosphate-buffered saline (10% [wt/vol]). Ten µL of commercial quality control reagent (Mengovirus extraction control kit, bioMérieux) was added as a sample-level control for the efficiency of RNA extraction, cDNA synthesis, library preparation, and sequencing [27]. Homogenates were centrifuged at 10000 g at 5 °C for 5 min prior to collection of the supernatant. RNA was extracted from the supernatant of the homogenate using a combination of the TRIzol reagent (ThermoFisher) and RNeasy mini kit (Qiagen), including on column DNase treatment as previously described (Wylezich et al. 2018). Viral nucleic acid extraction efficiency was evaluated in individual samples using a *Mengovirus*-specific qRT-PCR assay [27]. *cDNA* was synthesized using SuperScript IV reverse transcriptase (Thermo Fisher Scientific) and random hexamer primers, followed by doublestrand cDNA synthesis using the NEBNext mRNA second-strand synthesis module (New England BioLabs). Sequencing libraries were prepared using the Nextera XT kit (Illumina) and standard Nextera XT index adapters (Illumina) and sequenced using a MiSeq reagent kit version 3 (Illumina) with 2 × 300-bp paired-end sequencing aiming for a minimum of 3 million read pairs per sample. Metagenomic NGS data were generated for 6 fecal samples. The resulting fastq raw metagenomics datasets are publicly available in the Sequence Read Archive (SRA) under BioSample accession numbers SAMN14450565, SAMN1450544 (problem farm); SAMN14450538, SAMN14450540 (control farm 1); and SAMN14450542, SAMN14450570 (control farm 2).

Bioinformatic analysis

Metagenomic read classification: A two-step trimming strategy including Trimmomatic v0.38 [28] to remove adapter

sequences and low-quality bases (setting the 'ILLUMINACLIP 2:30:10′, 'LEADING:5′, 'TRAILING:5′, 'SLIDINGWINDOW:4:10′, and 'MINLEN:20′ options), followed by low-complexity sequence information removal using PRINSEQ v0.20.4 [29] (setting the '-trim_tail_right 10′ and '-trim_tail_left 10′ options), was implemented. Only paired reads were retained for further analysis. Classification of trimmed reads was performed with Kraken v1.1 (Wood and Salzberg, 2014) as previously described [27]. A customized Kraken database was built using all available RefSeq "Complete Genome" sequences of six targeted taxonomic groups (archaea, bacteria, fungi, human, protozoa, and viral) downloaded from RefSeq Genome (O'Leary et al., 2016) (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/) on 18/02/2019. Read counts per taxonomic level classified by Kraken were further normalized as reads per million (RPM) total (trimmed) reads [27] to remove technical bias introduced by sequencing depth variation between samples. An arbitrary acceptance criterion of RPM > 1 for a significant mNGS finding was used. *Mengovirus*-normalized read counts were used to validate the complete workflow for the generation of metagenomic data from RNA virus genomes, referring to the range previously documented for swine fecal samples [27].

De novo assembly and complete genome characterization: Raw sequence data from samples yielding increased Astroviral- and Posaviral-normalized read counts in the Kraken analysis were trimmed using Trim Galore! V0.5.0 (q = 30, l = 50, paired; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). A random paired subset (2 × 50,000 reads) was used for de novo assembly using SPAdes v3.9.0 [30] and IVA v1.0.0.[31]. Relevant contigs (*Astrovirus* and *Posavirus*, as determined using a Blastn search of all contigs longer than 500 nt against the NCBI Nucleotide database) were joined using CAP3 [32]. The full trimmed dataset was mapped to the finished full genome length contigs using BWA-0.6.2 [33]. To identify conserved protein domains, a RPS-BLAST search (Marchler-Bauer et al. 2015) against the Conserved Domain Database (CDD) was performed, in the case of *Posavirus* 1 strain Belgium/01/2019 adding a modified 3C protease domain profile for the *Posavirus* 1 sequence as previously described [18]. Due to high variation in the *Posavirus* 1 strain Belgium/01/2019 mapping, variants were called using LoFreq v2.1 [34]. A genomic region with frequency-linked sequence variants was observed (5297–5486), and a frequency cutoff of 18% (based on the lower limit of the 95% confidence interval of allele frequencies in the region) was applied, as well as a minimum positional coverage of 2500 to include a variation in the genome of the co-infecting virus. These cutoffs did not validate any variation beyond the 5297–5486 region. In addition, visual inspection of the alignment files (Tablet, [35]) demonstrated linkage of these variants on reads and read pairs. Additional low-frequency (< 18%) sequence variation was not taken into account. An ambiguous nucleotide Y was included at position 16 (frequency 48% but coverage < 1000) in the genome of the co-infecting variant *Posavirus* 1 strain Belgium/02/2019 genome.

Phylogenetic analysis

MAFFT v7.310 [36, 37] was used to align the complete genome sequences of *Posavirus* 1 strains Belgium/01/2018 and Belgium/02/2018 with all available *Posavirus* 1 complete genomes (n = 24, sourced from NCBI:txid1105380, as well as Unclassified Viruses NCBI:txid12429). After selection of the most suitable evolutionary model (lowest Bayesian Information Criterion score), a maximum likelihood phylogenetic tree was calculated in Mega X [38] (GTR + G + I [39]; partial deletion of missing data and gaps; 500 bootstrap replicates). In addition, due to low nucleotide and amino acid sequence similarity of *Posavirus* 1 to other *Posavirus* species, the amino acid sequences of the conserved RdRp domain of all available *Posavirus* genomes (n = 44, NCBI:txid2219055) and *Posavirus* 1 Belgium/01/2019 were aligned using MAFFT v7.310, and an amino acid Maximum Likelihood tree was calculated in Mega X using the most optimal model (LG + G + I [40], partial deletion of missing data and gaps; 500 bootstrap replicates).

A similar approach was used for PoAstV-4/Belgium/2019, using all porcine astrovirus 4 whole sequences available in the NCBI nt database (n = 35) with the addition of three porcine astrovirus 2 outgroup sequences and an identical optimal maximum likelihood model as selected by Mega X (GTR + G + I; 500 bootstrap replicates). In addition, due to low nucleotide sequence similarity with other porcine astrovirus genotypes, the amino acid sequence of the conserved RdRp domain was extracted from all *Mamastrovirus* complete genome sequences in the NCBI nucleotide database (n = 245, NCBI:txid249588) and aligned with the RdRp region of PoAstV-4/Belgium/2019 using MAFFT v7.310, followed by maximum likelihood phylogenetic analysis using Mega X (LG + G + I model; 500 bootstrap replicates).

Results

Fecal virome of sows in farms without and with problems of neonatal diarrhea

The efficient extraction of viral nucleic acids was validated for all samples by *Mengovirus* exogenous internal process control Cq values ranging from 32.83 to 34.91 (99% CI of *Mengovirus* Cq values previously documented 31.94–36.70, [27]). The reproducible *Mengovirus* detection, with normalized read counts ranging from 5.26 to 141.32 RPM, validated the complete workflow for all tested samples

[27]. As our mNGS workflow purposely avoided the use of viral enrichment steps, a relatively low percentage of viral reads (0.008–4%) was observed. However, due to the high sequencing effort per sample ($2.87 \times 10^6 - 4.46 \times 10^6$ read pairs), the absolute number of viral reads was sufficiently high to allow confident detection of viral taxa, even using relatively conservative cutoff criteria for significance (RPM > 1). The majority of reads was bacterial (78–94%).

Although there was evidence of presence of known porcine enteric viruses (*Mamastrovirus*, *Enterovirus*, *Picobirnavirus*, *Posavirus 1*, *Kobuvirus*, *Proprismacovirus*) in the farms without problems, their normalized read counts were relatively low (< 218 RPM of raw read pairs in the dataset, Table 1). In addition, low read counts showed similarities to unclassified viruses previously documented in fecal samples from seals, ducks, and humans. A final category of virome constituents in the healthy farms were reads classifying as diet-associated plant viruses or insect viruses (Table 1).

Both samples from the farm with neonatal diarrhea problems had dramatically increased viral normalized read counts (Table 1). In addition to the abovementioned virome categories with low normalized read numbers, the samples from the problem farm showed dramatically increased (> 38,000.00 RPM) normalized read counts for a single known fecal virome species. In the fecal sample from sow 1, 118,944 reads were classified as *Posavirus* 1. The fecal sample from sow 2 had 123,966 reads classified as *Mamastrovirus* (porcine astrovirus). In addition, low normalized read counts were observed for bacteriophage species (Caudovirales) in the samples from the symptomatic farm, whereas negligible normalized read counts were detected for Caudovirales in the control farms (RPM < 1.06).

Complete genome characterization of a porcine astrovirus 4 directly from fecal material of a sow in a farm with recurring neonatal diarrhea issues

The high *Mamastrovirus* load in the fecal sample from sow 2 in the problem farm allowed the de novo assembly of the complete coding sequence of a *Mamastrovirus* (6734 nt contig containing 442,684 reads, average positional coverage of 15,227, genbank accession number MT642666). A maximum likelihood analysis including the RdRp domain amino acid sequence of all publicly available *Mamastrovirus*

Table 1 Metagenomic virome composition of swine fecal samples

	Control Farm 1		Control Farm 2		Problem Farm	
	Sow 1	Sow 2	Sow 1	Sow 2	Sow 1	Sow 2
Dataset size (raw read pairs)	3,778,854	4,462,389	3,602,745	3,041,003	3,056,801	2,868,104
Detected Taxa						
Mengovirus extraction control	44.99	20.48	17.21	5.26	141.32	105.30
Posavirus 1	217.79	–	96.04	179.28	38,911.27	–
Mamastrovirus	181.27	17.70	8.88	0.99*	380.79	43,222.63
Porcine *Enterovirus*	57.95	–	–	–	–	–
Husavirus (Picornavirales)	–	1.34	–	–	–	30.33
Kobuvirus	–	–	–	1.97	–	16.04
Seal stool circular DNA virus (unclassified *Circoviridae*)	23.29	–	–	–	–	8.02
Duck feces associated circular DNA virus (unclassified viruses)	12.17	10.31	–	–	–	11.51
Picobirnavirus	10.59	–	1.94	–	1.31	1.74
Hudisavirus (unclassified circular ssDNA virus)	–	–	–	–	4.58	–
Po-circo-like virus 21 (unclassified *Circoviridae*)	–	–	1.11	–	–	–
Porprismacovirus (*Smacoviridae*)	–	1.79	0.28	–	–	–
Betabaculovirus (Choristoneura fumiferana granulovirus, Baculoviridae)‡	1.59	–	–	–	–	–
Tobamovirus (Pepper mild mottle virus,Virgaviridae)†	–	–	2.50	–	–	–
Sobemovirus (Lucerne streak transient virus) †	–	*24.65*	–	–	–	*9.76*
Carlavirus†	–	3.36	–	–	–	*3.49*
Caudovirales•	1.06	–	–	–	7.19	2.79

The normalized read count (RPM) of detected taxa is reported for each animal

'- ': not detected

*: swine enteric virus detected below significance threshold (RPM > 1)

Italic font: diet-associated plant virus (†) or insect virus (‡)

•: bacteriophage

genomes (supplementary Fig. 1) confirmed the classification of *PoAstV-4/Belgium/2019* as a porcine astrovirus 4 (*Mamastrovirus* 3). *PoAstV-4/Belgium/2019* shared a whole genome 88.09% nucleotide identity with a bovine *Astrovirus* isolated from fecal material sampled from encephalitic veal calves in Switzerland, and a 87.12% nucleotide identity with porcine astrovirus 4 isolate from mainland China, 2018. A whole genome maximum likelihood analysis (Fig. 1) using all available PoAstV-4 complete genomes confirmed the global circulation of these viruses. The 6734 nt genome contained three overlapping open reading frames encoding the nonstructural protein 1ab (ribosomal slippage combining 2 ORFS, 1332 AA, containing Trypsin-like peptidase and RdPd conserved domains), the nonstructural protein 1a (850 AA, containing a conserved Trypsin-like peptidase domain), and the capsid protein (849 AA, containing an *Astrovirus* capsid conserved domain).

Complete genome characterization of a porcine *Posavirus* 1 directly from fecal material of a sow in a farm with recurring neonatal piglet diarrhea problems

The high *Posavirus* 1 load in the fecal sample from sow 1 in the problem farm allowed the de novo assembly of the complete coding sequence of a *Posavirus* 1 belonging to the Unclassified Picornavirales (9814 nt contig containing 350,910 reads, average positional coverage of 7427, genbank

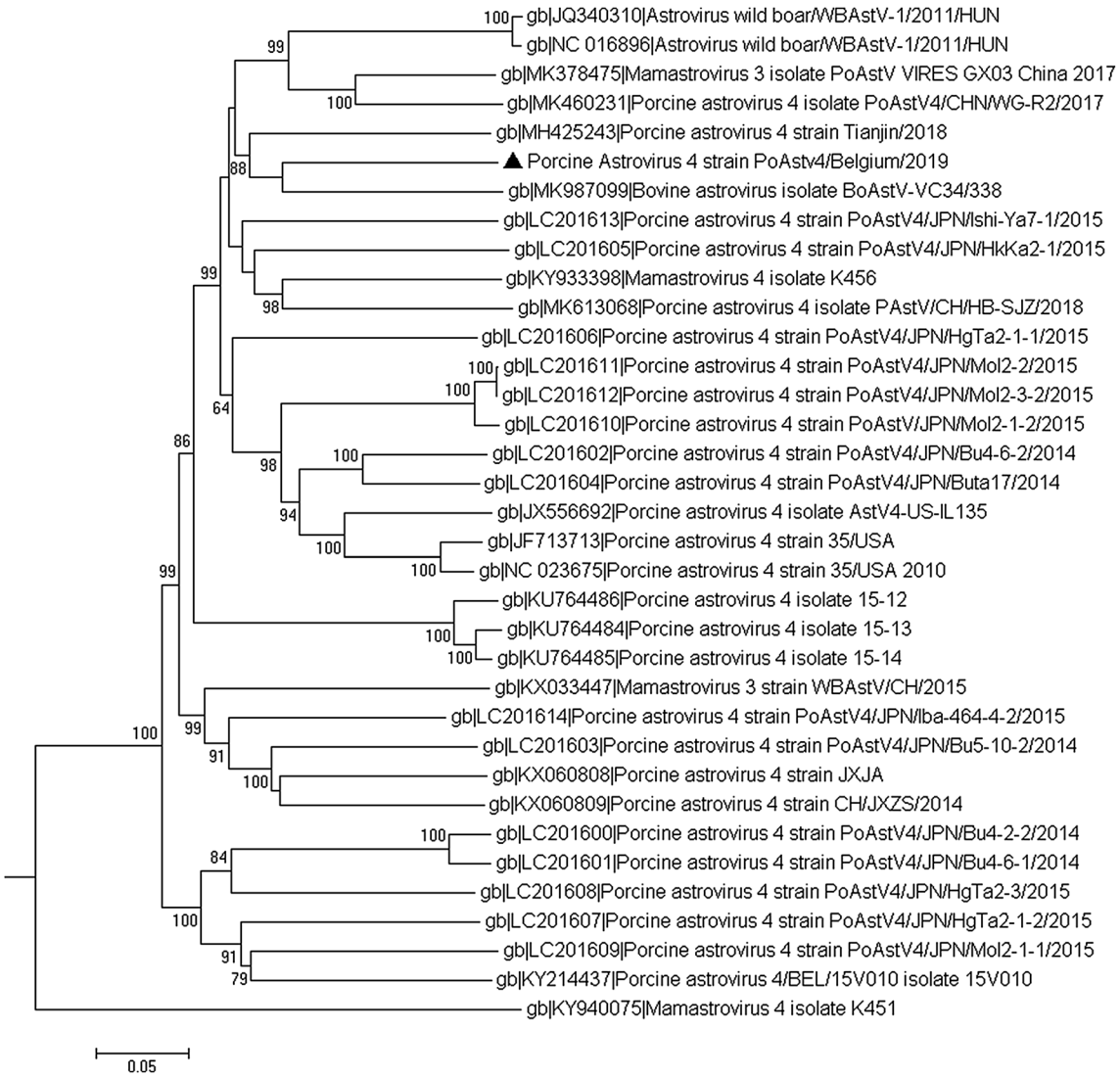

Fig. 1 Maximum likelihood phylogenetic analysis of the complete genome sequence of *PoAstV-4/Belgium/2019* in comparison with 35 publicly available PoAstV-4 complete genome sequences and 3 PoAstV-2 outgroup sequences. For visual simplification, only the subtree showing PoAstV-4 is depicted. Bootstrap support values > 60% are indicated near the nodes. Filled triangle: genome sequence determined in the present study. The scale bar expresses the average number of substitutions per site

accession number MT642667). The maximum likelihood analysis of the RdRp domain amino acid sequence of all publicly available *Posavirus* complete genomes confirmed the classification as *Posavirus*-1 (supplementary figure S2). *Posavirus 1/Belgium/01/2019* had a whole genome 93.80% nucleotide identity with *Posavirus* 1 isolate Ishi-Im3/JPN/2015 from Japan and had an intermediate phylogenetic position between *Posavirus* genomes characterized from Japan and North America as indicated in a whole genome maximum likelihood phylogenetic analysis comparing it with all publicly available *Posavirus* 1 genomes (Fig. 2). Of note, the publicly available whole genome sequence information for *Posavirus* 1 is biased by two studies sampling the majority of whole genomes in Vietnam (in 2013) and Japan (2015–2016). The 9814 nt single-stranded positive-sense RNA genome contained one open reading frame encoding a polyprotein precursor protein (3069 AA) containing the five expected Picornavirales-conserved protein domains (helicase, 3C cysteineprotease, RNAdependent RNA polymerase, and two Picornavirus capsid domains). The *Posavirus* RNA population in the fecal sample was highly diverse (supplementary Table 1). Frequency-linked high-frequency variants in the population, that were also physically linked on reads and read pairs upon inspection of the alignment, allowed us to identify a co-infecting *Posavirus* 1 strain Belgium/02/2019 (genbank accession number MT642668), while additional low-frequency variation was observed (as typical for replicating RNA viruses).

Discussion

Neonatal diarrhea in pigs is a common and important problem worldwide and affected farms incur major financial losses due to piglet mortality, decreased performance, and costs for treatment and control. Numerous studies have investigated the swine intestinal virome as a potential contributing factor [13]. Besides known swine intestinal virome constituents like Astroviruses, metagenomics NGS studies have added multiple novel viral species with currently unstudied biological implication in the occurrence of diarrhea [17–20]. Although the present study is too limited in

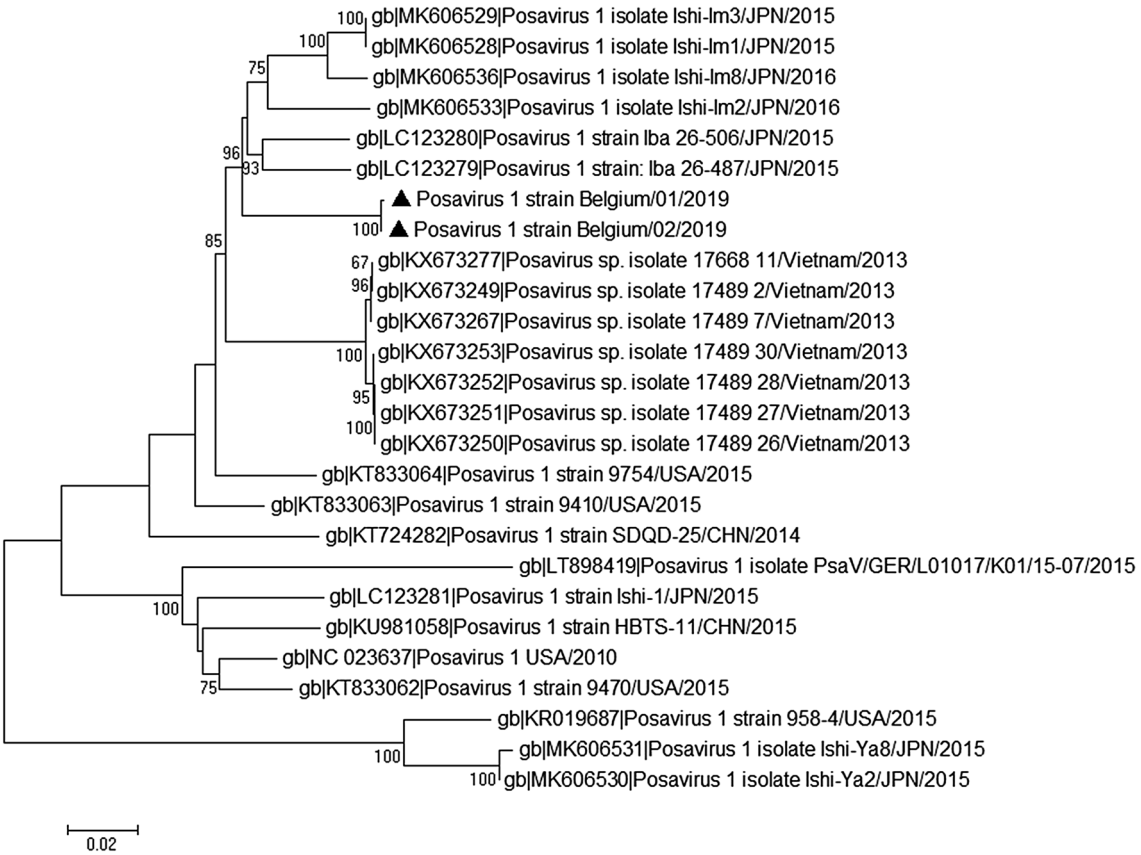

Fig. 2 Maximum likelihood phylogenetic analysis of the complete genome sequence of *Posavirus 1/ Belgium/01/2019* and *Posavirus 1/ Belgium/02/2019* in comparison with 24 publicly available *Posavirus* 1 complete genome sequences. Bootstrap support values > 60% are indicated near the nodes. Filled triangle: genome sequence determined in the present study. The scale bar expresses the average number of substitutions per site

scale to suggest an association of viral species to the occurrence of neonatal diarrhea, we observed an increased viral load (as indirectly measured using normalized NGS read counts) in sows in a farm with persisting neonatal diarrhea problems compared to two control farms without neonatal diarrhea problems. The increased viral load was in both samples from the problem farm attributable to high normalized read counts of a single viral genus (*Posavirus* and *Mamastrovirus*, respectively). Importantly, caution should be taken not to interpret metagenomic read numbers as an absolute quantification of the taxa identified. Indeed, multiple factors like differences in nucleic acid extraction efficiency due to different virion structure may bias the generation of sequence data of certain taxa. An earlier study [21] applied MinION nanopore sequencing-based mNGS on a single piglet diarrhea sample from a farm with recurring diarrhea problems and identified low read numbers of porcine *Kobuvirus*, *Enterovirus,* and *Astrovirus*, viruses that we detected in both the problem and control farms. Although no direct comparison is possible between both studies (different sampling time, farm and sampling of one piglet vs. two sows), the lower read numbers per sample attained through MinION sequencing (3×10^4 reads comparted to $> 3 \times 10^6$ read pairs in our study), did not allow the assembly of complete genomes or the identification of circulating virus variants as was possible in our study and other studies using second generation (Illumina) sequencing platforms [24]. In addition to confirming the virus genera detected by MINION sequencing, our study detected the circulation of *Posavirus 1*, *Picobirnavirus*, *Porprismacovirus*, unclassified circular DNA viruses, *Husavirus*, and *Hudisavirus* (Posa-like Picornavirales from human stool samples) with significant normalized read counts (RPM > 1). Our study also detected several feed-associated plant and insect viruses. However, biological differences between the samples in both studies (farm, sow vs. piglet, sampling time) may equally have influenced the amount of virus in the samples and thus the success in whole genome assembly and variant detection. The complementarity between second and third generation sequencing platforms should be stressed. The higher number of sequencing reads produced by second generation sequencing platforms may currently still provide superior mNGS detection sensitivity, sequencing depth, and sequencing breadth (i.e., genome coverage), all of which are important parameters in the confident diagnostic interpretation of mNGS results. However, the long read length provided by nanopore sequencing (in addition to advantages like portability and turnaround time) facilitates genome assembly, especially in the context of mixed infections. In the present study, the high read numbers of *Mamastrovirus* and *Posavirus 1* in animals from the problem farm allowed the assembly of a complete PoAstV-4 genome from one sow, as well as the identification and complete genome characterization of two co-infecting *Posavirus 1* strains directly from a fecal sample from another sow.

Porcine astroviruses (PoAstVs) are well known associates of porcine gastrointestinal disease. Initially identified using electron microscopy in feces from diarrheal pigs in the USA in 1980 [41], they currently show a worldwide distribution. Five genotypes of PoAstV have been identified [42] belonging to seven ICTV (International Committee on Taxonomy of Viruses) recognized clades of the genus *Mamastrovirus* [43]. This classification is testimony to a complex evolutionary history with different origins of PoAstVs including interspecies transmission and recombination events [44, 45]. PoAstVs have been detected, often using metagenomics sequencing methods, in the USA, Canada, Colombia, China, India, Thailand, South Korea, the Czech Republic, Croatia, Slovakia, Hungary, Italy, Spain, Austria, Germany, Belgium and Sweden at varying prevalences and have been detected in both healthy and diseased animals [14–16, 20, 21, 24, 42, 46–61]. In addition to its implication in gastrointestinal disease, PoAstV has been detected in nasal swabs of piglets with respiratory disease [62], and it was detected in the brains of piglets suffering from congenital tremors [63]. Due to its presence in both healthy and diseased animals, and the frequently documented co-infection with other gastrointestinal pathogens, the economic impact of swine *Astrovirus* infections is poorly understood and difficult to assess.

A study in five European countries (targeted screening of PoAstV-4, porcine *Rotavirus* A and C, porcine *Circovirus* 2 and *Kobuvirus*) showed that porcine astrovirus genotype 4 (PoAstV-4) is the most prevalent gastrointestinal virus in swine in Austria, Germany, Hungary, Spain, and Sweden, with an average prevalence of 72.4% in healthy and 68.5% in diarrheal pigs [61]. Similarly, PoAstV-4 is widely distributed in 12 states of the USA, showing a 62–64% prevalence in US farms with a history of diarrhea [42, 54]. In contrast, lower prevalences were found in domestic pigs in South Korea (19.4%) [52]) and in China (3.1—16.1%). [64, 65]. This might reflect differences in swine population density and farming practices. PoAstV-4 was detected in both domestic pigs and wild boar in Hungary [55, 56]. It has been recently detected in Belgium through metagenomics [21, 24]. Until present, only three PoAstV-4 complete genomes from Europe were available in public sequence databases, including a swine isolate from Belgium, 2015 (KY214437, [24]), a wild boar isolate from Hungary in 2011 (NC_016896; JQ340310, [55]), and an isolate from the feces of veal calves in Switzerland genetically classifying as PoAstV-4 (MK987099). In view of the very scarce availability of PoAstV-4 genomes from Europe, the present data represent an important contribution towards better understanding of the characteristics and evolution of these gastrointestinal swine viruses.

Porcine stool-associated virus (*Posavirus*) (Picornavirales) is a typical example of a novel virus genus that has been associated to the swine intestinal virome using metagenomic NGS. It was initially detected in fecal samples from diarrhetic as well as healthy piglets in 2010 [20] and has since been detected worldwide in swine farms [17–20] as well as well water collected from farms [66]. Although posaviruses can be detected at high frequency in pig feces (40%) and drinking water (21%), *Posavirus* antibodies are infrequently detected [67]. High sequence similarity of *Posavirus* sequences to RNA sequences from the parasite *Ascaris suum* [20, 68] suggests that these viruses may not infect pigs, but rather infect gut parasites or commensal organisms or have a dietary or environmental origin. Furthermore, RNA sequencing libraries from the mosquito *Anopheles sinensis* and the fruit fly *Drosophila subobscura* contained sequences with some identity to posaviruses [69]. In addition, a nucleotide composition analysis revealed compositional properties similar to arthropod viruses and suggested a non-mammalian replicative host for all Posaviruses [18]. The increased detection of posaviral reads in our study is thus unlikely to be directly associated with the persisting neonatal diarrhea problems. More likely, it is indirectly associated to the presence of a gut commensal or parasite, as previously suggested.

The observed high genetic variation in our *Posavirus* datasets is most likely due to the co-infection of multiple Posavirus 1 strains. Indeed, in addition to low-frequency variation expected in an RNA virus population ("quasi-species"), high-frequency variants (> 18% of reads covering a position) were frequency linked suggesting they belonged to the same co-infecting strain. As viral isolation and cloning approaches are currently an untouched domain for these novel viruses, future investigations using long read sequencing platforms may assist in the reconstruction of co-infecting genomes.

Although more extensive studies are needed to investigate the association of virome composition and health status of farms, this study highlights the potential of metagenomic NGS workflows to study clinical problems with unknown etiology. When sufficient sequencing effort is used and/or pathogen load is sufficiently high, it allows complete genome characterization and identification of mixed infections and co-circulating variants. In the particular case of the present study, this contributed to the scarce availability of both PoAstV-4 and Posavirus 1 whole genome sequences from Western Europe.

Acknowledgements This work was supported by funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 773830: One Health European Joint Programme as internal joint research project METASTAVA. Raw sequencing data were professionally generated by the Genomics Core Leuven and Laboratory for NGS and microarray diagnostics (Dr. D. Höper, Friedrich Loeffler Institut, Germany). The authors thank L. Weckx for excellent technical assistance and R. Winand for supporting the bioinformatic analyses.

Author contributions SVB, FV, DM, AS, and EV performed experiments; KV, RW, QF, SVB, and FV performed data analysis; SVB and FV designed the experiment and wrote the manuscript.

Funding This work was supported by funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 773830: One Health European Joint Programme as internal joint research project METASTAVA.

Data Availability Raw NGS metagenomic data are publicly available in the Sequence Read Archive (SRA) under BioSample accession numbers SAMN14450565, SAMN1450544, SAMN14450538, SAMN14450540, SAMN14450542, and SAMN14450570. Viral genome sequences are available under Genbank accession numbers MT642666-MT642668.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval As only animal fecal samples were used, no approval from the ethical committee for animal experiments was needed. Laboratory work and data analysis at Sciensano was not subject to.

Consent for publication All contributing authors have read and approved of the final version of the manuscript.

References

1. Holland RE (1990) Clin Microbiol Rev 3:345
2. Collins AM, Barchia IM (2014) Vet Microbiol 168:455
3. Kongsted H, Pedersen K, Hjulsager CK, Larsen LE, Pedersen KS, Jorsal SE, Bækbo P (2018) Porc Health Manag 4:17
4. S. Mesonero-Escuredo, K. Strutzberg-Minder, C. Casanovas, and J. Segalés, Porc. Health Manag. **4**, (2018).
5. Pensaert MB, Martelli P (2016) Virus Res 226:1
6. G. W. Stevenson, H. Hoang, K. J. Schwartz, E. R. Burrough, D. Sun, D. Madson, V. L. Cooper, A. Pillatzki, P. Gauger, B. J. Schmitt, L. G. Koster, M. L. Killian, and K. J. Yoon, J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc **25**, 649 (2013).
7. J. R. Thomson and R. M. Friendship, in *Diseaes Swine 10th Ed.*, 10th ed. (2012), pp. 199–244.
8. Blomström A-L, Fossum C, Wallgren P, Berg M (2016) PLoS ONE 11:e0166863
9. Qin S, Ruan W, Yue H, Tang C, Zhou K, Zhang B (2018) Sci Rep 8:13341
10. Chiu CY, Miller SA (2019) Nat Rev Genet 20:341
11. Höper D, Wylezich C, Beer M (2017) Adv Virus Res 99:17
12. Van Borm S, Belák S, Freimanis G, Fusaro A, Granberg F, Höper D, King DP, Monne I, Orton R, Rosseel T (2015) Methods Mol Biol Clifton NJ 1247:415
13. K. T. T. Kwok, D. F. Nieuwenhuijse, M. V. T. Phan, and M. P. G. Koopmans, Viruses **12**, (2020).
14. Chen Q, Wang L, Zheng Y, Zhang J, Guo B, Yoon K-J, Gauger PC, Harmon KM, Main RG, Li G (2018) Virol J 15:95

15. Karlsson OE, Larsson J, Hayer J, Berg M, Jacobson M (2016) PLoS ONE 11:e0151481
16. Zhang B, Tang C, Yue H, Ren Y, Song Z (2014) J Gen Virol 95:1603
17. Aoki H, Sunaga F, Ochiai H, Masuda T, Ito M, Akagami M, Naoi Y, Sano K, Katayama Y, Omatsu T, Oba M, Sakaguchi S, Furuya T, Ouchi Y, Shirai J, Mizutani T, Oka T, Nagai M (2019) Arch Virol 164:2147
18. B. B. Oude Munnink, M. V. T. Phan, VIZIONS Consortium, P. Simmonds, M. P. G. Koopmans, P. Kellam, L. van der Hoek, and M. Cotten, Virus Evol. **3**, vex022 (2017).
19. Sano K, Naoi Y, Kishimoto M, Masuda T, Tanabe H, Ito M, Niira K, Haga K, Asano K, Tsuchiaka S, Omatsu T, Furuya T, Katayama Y, Oba M, Ouchi Y, Yamasato H, Ishida M, Shirai J, Katayama K, Mizutani T, Nagai M (2016) Arch Virol 161:3541
20. Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E (2011) J Virol 85:11697
21. Theuns S, Vanmechelen B, Bernaert Q, Deboutte W, Vandenhole M, Beller L, Matthijnssens J, Maes P, Nauwynck HJ (2018) Sci Rep 8:9830
22. Theuns S, Vyt P, Desmarets LMB, Roukaerts IDM, Heylen E, Zeller M, Matthijnssens J, Nauwynck HJ (2016) Virus Res 213:172
23. Theuns S, Conceição-Neto N, Zeller M, Heylen E, Roukaerts IDM, Desmarets LMB, Van Ranst M, Nauwynck HJ, Matthijnssens J (2016) Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 43:135
24. N. Conceição-Neto, S. Theuns, T. Cui, M. Zeller, C. K. Yinda, I. Christiaens, E. Heylen, M. Van Ranst, S. Carpentier, H. J. Nauwynck, and J. Matthijnssens, Virus Evol. **3**, vex024 (2017).
25. S. Theuns, N. Conceição-Neto, I. Christiaens, M. Zeller, L. M. B. Desmarets, I. D. M. Roukaerts, D. D. Acar, E. Heylen, J. Matthijnssens, and H. J. Nauwynck, Genome Announc. **3**, (2015).
26. Theuns S, Desmarets LMB, Heylen E, Zeller M, Dedeurwaerder A, Roukaerts IDM, Van Ranst M, Matthijnssens J, Nauwynck HJ (2014) Vet Microbiol 172:23
27. Van Borm S, Fu Q, Winand R, Vanneste K, Hakhverdyan M, Höper D, Vandenbussche F (2020) J Virol Methods 283:113916
28. Bolger AM, Lohse M, Usadel B (2014) Bioinforma Oxf Engl 30:2114
29. Schmieder R, Edwards R (2011) Bioinforma Oxf Engl 27:863
30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) J Comput Biol J Comput Mol Cell Biol 19:455
31. Hunt M, Gall A, Ong SH, Brener J, Ferns B, Goulder P, Nastouli E, Keane JA, Kellam P, Otto TD (2015) Bioinforma Oxf Engl 31:2374
32. Huang X, Madan A (1999) Genome Res 9:868
33. Li H, Durbin R (2009) Bioinforma Oxf Engl 25:1754
34. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N (2012) Nucleic Acids Res 40:11189
35. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D (2013) Brief Bioinform 14:193
36. Katoh K, Misawa K, Kuma K, Miyata T (2002) Nucleic Acids Res 30:3059
37. Katoh K, Standley DM (2013) Mol Biol Evol 30:772
38. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) Mol Biol Evol 35:1547
39. Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York
40. Le SQ, Lartillot N, Gascuel O (2008) Philos Trans R Soc Lond B Biol Sci 363:3965
41. Bridger JC (1980) Vet Rec 107:532
42. Xiao C-T, Giménez-Lirola LG, Gerber PF, Jiang Y-H, Halbur PG, Opriessnig T (2013) J Gen Virol 94:570
43. Boujon CL, Koch MC, Seuberlich T (2017) Adv Virus Res 99:109
44. Cortez V, Meliopoulos VA, Karlsson EA, Hargest V, Johnson C, Schultz-Cherry S (2017) Annu Rev Virol 4:327
45. Zhao C, Chen C, Li Y, Dong S, Tan K, Tian Y, Zhang L, Huang J, Zhang L (2019) Arch Virol 164:1469
46. Brnić D, Prpić J, Keros T, Roić B, Starešina V, Jemeršić L (2013) Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 14:258
47. Indik S, Valícek L, Smíd B, Dvoráková H, Rodák L (2006) Vet Microbiol 117:276
48. Kattoor JJ, Malik YS, Saurabh S, Sircar S, Vinodhkumar OR, Bora DP, Dhama K, Ghosh S, Banyai K, Touil N, Abdel-Moneim AS, Vlasova AN, Kobayashi N, Singh RK (2019) Transbound Emerg Dis 66:47
49. Kumthip K, Khamrin P, Saikruang W, Kongkaew A, Vachirachewin R, Ushijima H, Maneekarn N (2018) Arch Virol 163:1823
50. Lan D, Ji W, Shan T, Cui L, Yang Z, Yuan C, Hua X (2011) Arch Virol 156:1869
51. M.-A. Laurin, M. Dastor, and Y. L'homme, Arch. Virol. **156**, 2095 (2011).
52. Lee M-H, Jeoung H-Y, Park H-R, Lim J-A, Song J-Y, An D-J (2013) Virus Genes 46:175
53. Z. Luo, S. Roi, M. Dastor, E. Gallice, M.-A. Laurin, and Y. L'homme, Vet. Microbiol. **149**, 316 (2011).
54. S. K. Mor, Y. Chander, D. Marthaler, D. P. Patnayak, and S. M. Goyal, J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc **24**, 1064 (2012).
55. Reuter G, Nemes C, Boros A, Kapusinszky B, Delwart E, Pankovics P (2012) Arch Virol 157:1143
56. Reuter G, Pankovics P, Boros A (2011) Arch Virol 156:125
57. Shan T, Wang C, Tong W, Zheng H, Hua X, Yang S, Guo Y, Zhang W, Tong G (2012) J Virol 86:13820
58. Tassoni L, Zamperin G, Schiavon E, Vendramin V, Cavicchio L, Mion M, Tonon FT, Monne I, Beato MS (2019) Vet Ital 55:221
59. Ulloa JC, Gutiérrez MF (2010) Can J Microbiol 56:569
60. Vilcek S, Salamunova S, Jackova A (2019) J Vet Sci 20:91
61. Zhou W, Ullman K, Chowdry V, Reining M, Benyeda Z, Baule C, Juremalm M, Wallgren P, Schwarz L, Zhou E, Pedrero SP, Hennig-Pauka I, Segales J, Liu L (2016) Vet Microbiol 182:75
62. Padmanabhan A, Hause BM (2016) Arch Virol 161:2575
63. Blomström A-L, Ley C, Jacobson M (2014) Acta Vet Scand 56:82
64. Qin Y, Fang Q, Li X, Li F, Liu H, Wei Z, Ouyang K, Chen Y, Huang W, Vet BMC (2019) Res 15:471
65. Xiao C-T, Luo Z, Lv S-L, Opriessnig T, Li R-C, Yu X-L (2017) Arch Virol 162:943
66. Hause BM (2016) J Swine Health Prod 24:304
67. Hause BM, Palinski R, Hesse R, Anderson G (2016) J Gen Virol 97:1362
68. Wang J, Czech B, Crunk A, Wallace A, Mitreva M, Hannon GJ, Davis RE (2011) Genome Res 21:1462
69. Webster CL, Longdon B, Lewis SH, Obbard DJ (2016) Evol Bioinforma Online 12:13