

A divergent clade of circular single-stranded DNA viruses from pig feces

Andrew K. Cheung · Terry F. Ng · Kelly M. Lager ·
Darrell O. Bayles · David P. Alt · Eric L. Delwart ·
Roman M. Pogranichniy · Marcus E. Kehrli Jr.

Received: 20 February 2013 / Accepted: 19 March 2013 / Published online: 24 April 2013
© Springer-Verlag Wien (outside the USA) 2013

Abstract Using metagenomics and molecular cloning methods, we characterized five novel small, circular viral genomes from pig feces that are distantly related to chimpanzee and porcine stool-associated circular viruses, (ChiSCV and PoSCV1). Phylogenetic analysis placed these viruses into a highly divergent clade of this rapidly growing new viral family. This new clade of viruses, provisionally named porcine stool-associated circular virus 2 and 3 (PoSCV2 and PoSCV3), encodes a stem-loop structure (presumably the origin of DNA replication) in the small intergenic region and a replication initiator protein commonly found in other biological systems that replicate their genomes via the rolling-circle mechanism.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

A. K. Cheung (✉) · K. M. Lager · M. E. Kehrli Jr.
Virus and Prion Diseases, Research Unit, National Animal
Disease Center, USDA, Agricultural Research Service,
PO Box 70, Ames, IA 50010, USA
e-mail: Andrew.Cheung@ars.usda.gov

T. F. Ng · E. L. Delwart
Department of Laboratory Medicine, Blood Systems Research
Institute, University of California at San Francisco,
San Francisco, CA, USA

D. O. Bayles · D. P. Alt
Infectious Bacterial Diseases, Research Unit, National Animal
Disease Center, USDA, Agricultural Research Service, Ames,
IA 50010, USA

R. M. Pogranichniy
Indiana Animal Disease Diagnostic Laboratory
and Department of Comparative Pathobiology,
Purdue University, West Lafayette, Indiana

Furthermore, these viruses also exhibit three additional overlapping open reading frames in the large intergenic region between the capsid and replication initiator protein genes.

Keywords Deep sequencing · Porcine stool-associated circular virus

The application of pyrosequencing technology to study the pig virome has led to the discovery of unique viruses that may or may not necessarily play a role in disease [8]. Recently, previously unknown single-stranded (ss) circular DNA viruses, similar to chimpanzee stool-associated circular virus (ChiSCV) [1], were identified in fecal samples collected from sick or healthy pigs. The genomes of these pig- or porcine-stool-associated viruses (PigSCV and PoSCV1) [7, 9] are about 2.5 kilobases (kb) in size and contain two major open reading frames (ORFs) that encode the capsid protein (Cap) and replication initiator protein (Rep). Both PigSCV and PoSCV1 contain a palindromic sequence capable of forming a stem-loop structure in the small intergenic region (SIR), which suggests they may synthesize their respective genomes by the rolling-circle replication mechanism. Whereas the Rep and Cap of PigSCV are encoded by the same DNA strand, the Rep and Cap of PoSCV1 are transcribed bidirectionally from the large intergenic region (LIR) in opposite orientations.

In this study, diarrheal fecal materials were collected from sick pigs (1 day to 6 weeks of age) with no common background from various Midwest farms in the United States that had been submitted to the Indiana Animal Disease Diagnostic Laboratory, Purdue University, West Lafayette, Indiana, between December, 2009 and June, 2010. Following routine diagnostic analysis, rotaviruses,

coronaviruses and enteroviruses were detected in these samples that were subsequently submitted to National Animal Disease Center for additional study. Fecal samples from six pigs were pooled and processed to prepare a viral nucleic acid library. Briefly, viral particles were first purified using size filtration and nucleases [8]. The extracted viral nucleic acids were amplified by random PCR with a specific nucleotide sequence tag for identification. Several libraries, each prepared with a different sequence tag for identification, were combined and subjected to 454 pyrosequencing and analyzed as described previously [4]. Sequencing was performed on a Roche FLX sequencer using Titanium chemistry (Roche, Branford, CT). For comparative purposes, the best BLASTx results were used to categorize the sequences (contigs and singletons) into virus family and genus.

Of the 1,296,370 total keypass reads generated, 125,282 reads contained sequence tags belonging to the six pooled fecal samples. Positive sequence reads for a taxonomic group were identified based on deduced protein sequence similarity using a stringent expectation value, best BLASTx expectation scores of $\leq 10^{-10}$, as the cutoff. Sequence reads at this cutoff level exhibited highly significant protein sequence similarities with known viruses in the database. Viral sequences (coronavirus, enterovirus, rotavirus) corresponding to the viruses identified by the Indiana Animal Disease Diagnostic Laboratory (West Lafayette, IN) were detected. Other viral sequences belonging to the RNA virus families (astrovirus, picobirnavirus, teschovirus, torovirus and sapelovirus) and DNA virus families (anellovirus, circovirus, and parvovirus) were also observed. Several sequences encoding amino acid sequences related to Rep of ChiSCV and PoSCV1 were identified.

The ChiSCV- and PoSCV1-related nucleotide sequences detected by deep sequencing (designated Tp1 and Tp2) were used to design primers for PCR. DNA amplification employing converging primers (conventional PCR) was used to confirm the presence of contig sequences in the sample, and diverging primers (inverse PCR) were used to amplify and clone the complete circular viral genomes. Nucleic acids were extracted directly from fecal samples using a QIAamp MinEluteVirus Vacuum Kit (QIAGEN, Valencia, CA) and subjected to rolling-circle amplification to amplify circular DNA molecules (Illustra GenomiPhi V2 DNA Amplification Kit, GE Healthcare Biosciences, Piscataway, NJ). The amplified DNA was used as a template for PCR using converging or diverging primers based on 454 pyrosequencing results. The amplicons were resolved and excised from agarose gels, cloned into plasmid TOPO-CLX104 and introduced into *Escherichia coli* TOP10 (Invitrogen, Carlsbad, CA) by transformation. Multiple clones were picked and used for sequence

determination using Sanger methods. From the Tp1 PCR product, three clones were analyzed, and they all yielded identical sequences. This viral genome was designated porcine stool-associated circular virus 2 (PoSCV2; GenBank accession number KC545226). From the Tp2 PCR product, four variant genomes were obtained, and the individual genomes were designated PoSCV3-4L5, -3L7, -LT2 and -4L13 with GenBank accession numbers KC545229, KC545227, KC545230 and KC545228, respectively.

Similar to the genome organization of other SCVs, the Tp clones (PoSCV2 and all four PoSCV3 clones) were about 2.5 kb in length (Fig. 1a). The viral genomes can be divided into four regions: two large ORFs with deduced amino acid sequences exhibiting homology to the Rep and Cap of ChiSCV, a LIR that encodes multiple overlapping ORFs, and an SIR that contains a palindromic sequence capable of forming a stem-loop structure. The Rep ORF and Cap ORF are transcribed divergently from the LIR and converge at the SIR. In contrast to the LIR of PoSCV1, which encodes two small ORFs (ORF3 and ORF4) in the same orientation as the Cap gene, the LIRs of PoSCV2 and PoSCV3 also contain an additional ORF (ORF5) in the reverse orientation as the Cap gene.

The four PoSCV3 genomes were aligned, and a schematic representation is shown in Fig. 2a. The LIR, Cap region and 5' portion of the Rep region exhibited few to no nucleotide differences. Genetic differences were concentrated around the stem-loop structure in the SIR and the 3' portion of the Rep ORF. The four genome regions (*SIR*, *Rep-ORF*, *Cap-ORF* and *LIR*) are described individually in greater detail below.

SIR: The SIR sequences of PoSCV2 and PoSCV3 are shown in Fig. 2b. Whereas the Rep ORF of PoSCV3-4L5 overlaps the stem-loop structure, the other four PoSCV3 genomes do not. All five genomes contain a palindromic sequence in the SIR that is capable of forming a stem-loop structure whose nucleotide sequence is well conserved. This stem-loop structure may be part of the origin of DNA replication. Among the PoSCV3 genomes, the SIR sequences on the Cap-gene side are more conserved, while sequences on the Rep-gene side exhibit the greatest differences.

Rep ORF: Phylogenetic and pairwise identity analyses were conducted to determine the relationship of Tp clones to other viruses. A phylogram was created based on the deduced amino acid sequences encoded by the Rep gene (Fig. 2c). The amino acid sequences were aligned using Mafft 5.8 [2] with the E-INS-I alignment strategy and previously described parameters [5, 6]. A maximum-likelihood tree was created using RaxML based on the Mafft alignment with previously described parameters [6, 10]. The resulting tree was midpoint rooted using MEGA4 [11]. Pairwise identity analysis of the PoSCV genes and ORFs

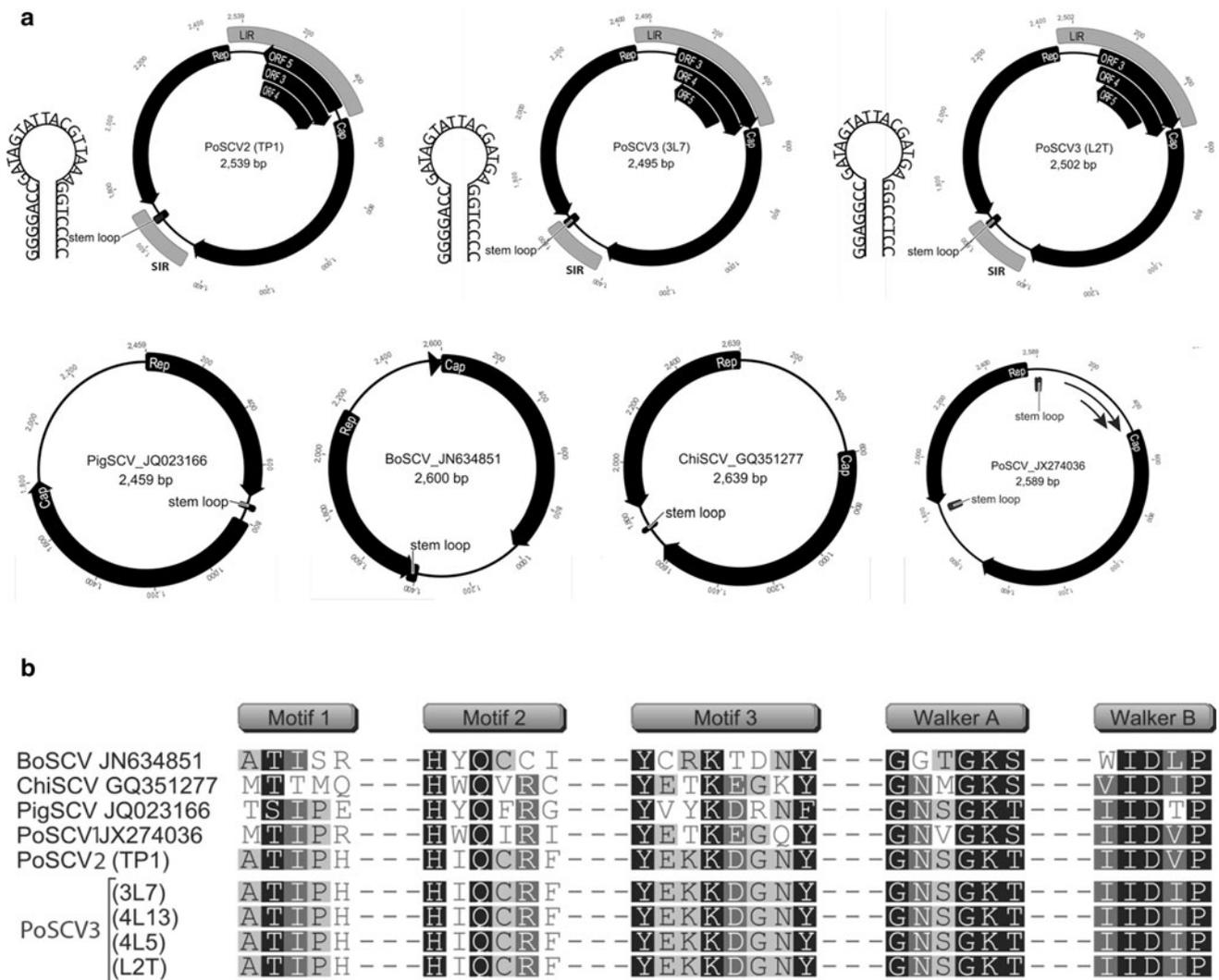


Fig. 1 Genome organization of selected SCVs. (a) The circular genomes of PoSCV2 and PoSCV3 (clones 3L7 and L2T) illustrated alongside several distantly related reference genomes (BoSCV–

JN634851, ChiSCV–GQ351277, PigSCV–JQ23166 and PoSCV1–JQ274036). The Rep, Cap, LIR and SIR (with stem–loop) regions are indicated. (b) RCR motifs found in the Rep proteins of selected SCVs

was also performed using MEGA4 [11]. The results showed that the Tp clones were most closely related to ChiSCV or PoSCV1, and they clustered into a distinct clade with PoSCV2 and PoSCV3, separated into two different sub–groups. There is limited amino acid sequence identity (23–32 %) between the Tp clones and bovine SCV (BoSCV) [3] or PigSCV.

The amino acid sequence identities between Tp:PoSCV1 and Tp:ChiSCV were approximately 50 % and 40 %, respectively (Table 1a). The nucleotide or amino acid sequence identity between PoSCV2 and PoSCV3 was approximately 87 %, and the sequence identity among the PoSCV3 variants was 93–100 %. In addition, rolling-circle replication (RCR) amino acid sequence motifs (RCR–I, RCR–II, RCR–III, walker A and walker B) commonly found among the Rep proteins involved in RCR were

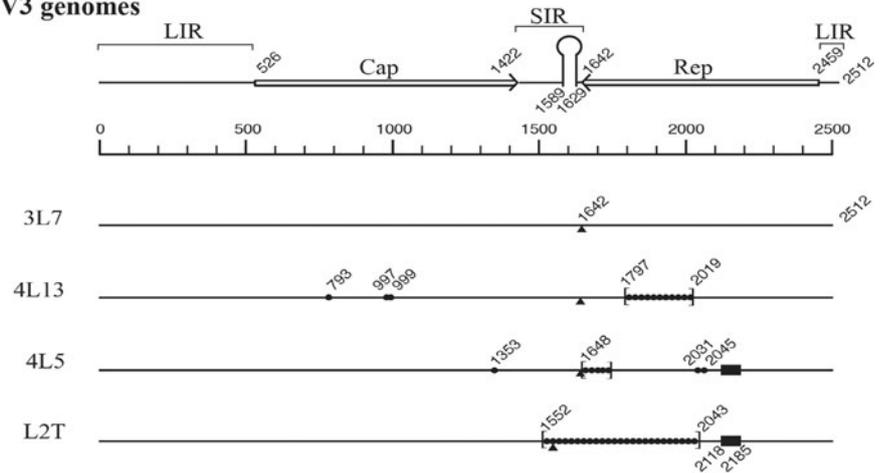
detected [9] (Fig. 1b). These motifs were conserved among members of this new clade.

Cap ORF: The deduced Cap protein sequences of selected SCV were compared (Table 1b). There is limited amino acid sequence homology (17–26 %) between Tp clones and BoSCV, ChiSCV, PigSCV or PoSCV1. The nucleotide sequence identity of the Cap gene (46–48 %) was lower than the amino acid sequence identity (60–62 %) between PoSCV2 and PoSCV3. In general, the Rep gene is more conserved than the Cap gene across the ssDNA viruses. Therefore, it was unusual to find that the nucleotide and amino acid sequence identities among the PoSCV3 Cap genes (99–100 %) were higher than those of the Rep genes (94–100 %).

LIR: The LIR nucleotide sequence identity between PoSCV2 and PoSCV3 was 70.2 %, and the sequences of

Fig. 2 Comparative analysis of the PoSCV3 variant genomes. **(a)** Schematic representations of the circular genome. The 3L7 genome (2495 nt) was used as a reference, and gaps were introduced for sequence alignment. Termination sites of the Rep ORF are indicated by black triangles. Point mutations are indicated by black dots, and the long stretches of nucleotide differences are indicated in brackets. The areas showing identical mutations in 4L5 and L2T are indicated by black boxes. **(b)** Sequence analysis showing the SIR. Nucleotide differences are colored. Dashes were introduced to align the sequences. The stem-loop sequences are indicated by shaded boxes, and C-terminal coding sequences of Rep and Cap are indicated by black arrows. **(c)** A phylogenetic tree derived from the deduced amino acid sequences of the Rep genes from selected single-stranded DNA viruses

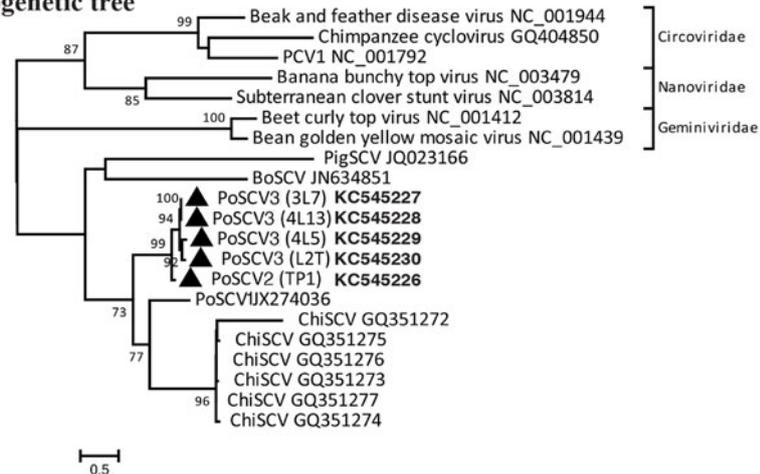
a PoSCV3 genomes



b SIR



c Phylogenetic tree



the PoSCV3s were identical. Both PoSCV1 and the Tp clones exhibit two overlapping ORFs, ORF3 and ORF4, transcribed in the same orientation as the Cap gene. There were no detectable amino acid homologies between the PoSCV1 and the Tp clones. For ORF3, the amino acid sequence identity between PoSCV2 and PoSCV3 was approximately 68 %, and the sequence identity among the PoSCV3s was 99–100 % (Table 1c). For ORF4, the amino acid sequence identity between PoSCV2 and

PoSCV3 was approximately 64 %, and the sequence identity among the PoSCV3 genes was 99–100 % (Table 1c). It is expected that the deduced amino acid sequences of ORF3 and ORF4 would be identical among the PoSCV3 variants since the nucleotide sequences are identical. However, it is surprising that the amino acid identity of these two ORFs between PoSCV2 and PoSCV3 was 64–68 %, which is slightly higher than the capsid protein homology of 61 %. This finding lends

Table 1 Deduced amino acid sequence identities of the (a) Rep proteins, (b) Cap proteins, (c) ORF3s, ORF4s, ORF5s of selected SCV

	BoSCV	ChiSCV	PigSCV	PoSCV1	PoSCV2	PoSCV3								
					TP1	3L7	4L13	4L5						
(a) Rep														
BoSCV_JN634851														
ChiSCV_GQ351272	23.7													
PigSCV_JQ023166	23.9	20.7												
PoSCV1_JX274036	31.9	39.8	23.8											
TP1	31.4	39.1	27.2	52.8										
3L7	31.0	40.4	24.9	51.1	88.1									
4L13	31.0	40.4	24.9	51.1	88.1	100.0								
4L5	31.0	40.9	24.9	50.6	88.1	93.7	93.7							
L2T	30.6	40.4	24.9	50.2	86.9	94.8	94.8	94.1						
(b) Cap														
BoSCV_JN634851														
ChiSCV_GQ351272	36.2													
PigSCV_JQ023166	38.6	50.8												
PoSCV1_JX274036	23.7	25.8	26.3											
TP1	20.0	21.1	21.0	23.3										
3L7	17.8	21.3	21.1	25.9	61.4									
4L13	17.8	21.3	21.1	25.9	61.4	100.0								
4L5	17.8	21.3	21.1	25.9	64.1	100.0	100.0							
L2T	17.8	21.3	21.1	25.9	60.7	99.3	99.3	99.3						
(c) ORF3s, ORF4s, ORF5s														
ORF3	TP1	3L7	4L13	4L5	ORF4	TP1	3L7	4L13	4L5	ORF5	TP1	3L7	4L13	4L5
TP1	–	–	–	–	TP1	–	–	–	–	TP1	–	–	–	–
3L7	68.6	–	–	–	3L7	64.3	–	–	–	3L7	59.2	–	–	–
4L13	68.6	100.0	–	–	4L13	64.3	100.0	–	–	4L13	59.2	100.0	–	–
4L5	68.6	100.0	100.0	–	4L5	64.3	100.0	100.0	–	4L5	59.2	100.0	100.0	–
L2T	68.3	99.8	99.8	99.8	L2T	64.0	99.8	99.8	99.8	L2T	58.9	99.7	99.7	99.7

credence to the speculation that either ORF may code for an important functional domain or protein.

The LIRs of PoSCV2 and PoSCV3 also code for an additional ORF5 that is transcribed in the opposite orientation to the Cap gene and overlaps ORF3 and ORF4. The amino acid sequence identity between PoSCV2 and PoSCV3 was approximately 59 %, and the sequence identity among the PoSCV3 variants was 99–100 % (Table 1c). Thus, the amino acid sequence identity of ORF5 between PoSCV2 and PoSCV3 was almost as high as that of the capsid protein identity of 61 %.

In this work, we report a clade of novel viruses that includes PoSCV2 and PoSCV3, which encode a Rep-like protein and a palindromic sequence capable of forming a stem-loop structure (in the SIR), suggesting that their genomes may replicate via a common RCR mechanism. Interestingly, this clade of viruses encodes three

overlapping “conserved” ORFs (ORF3, ORF4 and ORF5) in the LIR. Whereas the amino acid sequence identities between PoSCV2 and PoSCV3 for these ORFs range from 58.9 % to 68.6 %, the amino acid sequence identities among the capsid proteins range from 60.7 % to 64.1 %. Whether these additional ORFs code for functionally important proteins is not known. Likewise, the role of these viruses in any disease is unknown. The growing diversity of SCV-related genomes currently reported in the stool of chimpanzees, cows, and pigs likely portend further identification in other mammalian species. However, it remains to be seen whether these stool-associated viruses replicate in the host or that they are pass-through viruses present in the diet. Confirmation of their host and organ tropisms will require detection of SCV-specific antibodies or finding virions in animal tissues. A high level of co-infections involving numerous known viruses (coronavirus,

enterovirus, rotavirus, astrovirus, picobirnavirus, teschovirus, torovirus, sapelovirus, anellovirus, circovirus and parvovirus) was detected in just six animals from this study. This report, and the work of others, demonstrates the growing complexity of the pig virome and the challenge to understand the biology, interactions and significance of these newly discovered viruses.

Acknowledgments The authors thank N. Otis, L. Hobbs, D. Michael and M. Woodruff for technical assistance and S. Ohlendorf for manuscript preparation. T.F.N. and E.L.D. were supported by R01 HL105770.

References

1. Blinkova O, Victoria J, Li Y, Keele BF, Sanz C, Ndjanga JB, Peeters M, Travis D, Lonsdorf EV, Wilson ML, Pusey AE, Hahn BH, Delwart EL (2010) Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* 91:74–86
2. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518
3. Kim HK, Park SJ, Nguyen VG, Song DS, Moon HJ, Kang BK, Park BK (2012) Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* 93:635–639
4. Lager KM, Ng TF, Bayles DO, Alt DP, Delwart EL, Cheung AK (2012) Diversity of viruses detected by deep sequencing in pigs from a common background. *J Vet Diagn Invest* 24:1177–1179
5. Ng TF, Wheeler E, Greig D, Waltzek TB, Gulland F, Breitbart M (2011) Metagenomic identification of a novel anellovirus in Pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. *J Gen Virol* 92:1318–1323
6. Ng TF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* 86:12161–12175
7. Sachsenroder J, Twardziok S, Hammerl JA, Janczyk P, Wrede P, Hertwig S, John R (2012) Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One* 7:e34631
8. Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E (2011) The fecal virome of pigs on a high-density farm. *J Virol* 85:11697–11708
9. Sikorski A, Arguello-Astorga GR, Dayaram A, Dobson RC, Varsani A (2012) Discovery of a novel circular single-stranded DNA virus from porcine faeces. *Arch Virol* 158(1):283–289
10. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RA × ML Web servers. *Syst Biol* 57:758–771
11. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599