Accepted Manuscript

Metagenomic identification, genetic characterization and genotyping of porcine sapoviruses

Jingjiao Li, Wen Zhang, Li Cui, Quan Shen, Xiuguo Hua

PII: S1567-1348(18)30213-2
DOI: doi:10.1016/j.meegid.2018.04.034
Reference: MEEGID 3500

To appear in: *Infection, Genetics and Evolution*

Received date: 13 October 2017
Revised date: 30 March 2018
Accepted date: 25 April 2018

Please cite this article as: Jingjiao Li, Wen Zhang, Li Cui, Quan Shen, Xiuguo Hua , Metagenomic identification, genetic characterization and genotyping of porcine sapoviruses. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Meegid(2018), doi:10.1016/j.meegid.2018.04.034

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Metagenomic identification, genetic characterization and genotyping of porcine sapoviruses

Jingjiao Li[1], Wen Zhang[2], Li Cui[1], Quan Shen[2*], Xiuguo Hua[1 *]

1. School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China
2. School of Medicine, Jiangsu University, Zhenjiang 212013, China

*Corresponding author

E-mail addresses:

JJL: lijingjiao2015SJTU@sjtu.edu.cn;

WZ: z0216wen@yahoo.com;

LC: lcui@sjtu.edu.cn;

QS: shenquanfly@yahoo.com;

XGH: hxg@sjtu.edu.cn

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Sapoviruses (SaVs), belonging to the genus *Sapovirus* of the family *Caliciviridae,* were known as the enteric pathogen causing acute gastroenteritis. SaVs have been detected in humans and several animals including pigs and some porcine SaVs showed close sequence relationship with human strains suggesting the possibility of interspecies transmission. Here, we sequenced the genomes of two porcine SaVs (with strain names of p38 and SH1703) using the metagenomic analyses and traditional RT-PCR methods. Phylogenetic trees were constructed based on the complete genome, the full-length VP 1 nucleotide and amino acid sequences to group those two strains. The two porcine SaV strains, p38 and SH1703, detected in this study, were classified as genogroup III and genogroup VII, respectively. These two strains showed similar genomic organization with that of other SaVs. We firstly divided SaVs into 51 genotypes within 19 genogroups. Our data are helpful for genetic characterization and classification of newly detected SaVs worldwide.

Key words: sapovirus; viral metagenomics; genotype; phylogenetic tree; complete genome; interspecies transmission

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Introduction

Sapoviruses (SaVs) belong to the genus *Sapovirus* within the family *Caliciviridae*. The SaV genome is a linear, positive sense and single stranded RNA of 7.1-7.7 kb with a polyadenylated tail at the 3' terminus. The SaVs genome contains two or three open reading frames (ORFs). ORF 1 is predicted to encode seven nonstructural proteins (NSs) (NS1, NS2, NS3 (2C-like NTPase), NS4, NS5 (VPg) and NS6-NS7 (fused 3C-like protease-3D-like RNA-dependent RNA polymerase))[1-3] and the major capsid protein VP1. ORF2 encodes the minor structural protein VP2.

Based on the complete VP1 amino acid sequences, SaVs could be divided into 19 genogroups (GI- GXIX)[4-7] and some porcine SaVs were clustered into genogroup V and showed highly genetically homologous to human ones[5]. The porcine SaV was firstly detected in the United States by electron microscopy in 1980 and the complete genome of this strain was sequenced and genetically characterized as a sapovirus in 1999[8, 9]. At the beginning of the study, there were 49 complete genomes of SaVs available in GenBank database, including 29 human SaVs, 15 porcine SaVs, 3 bat SaVs, 1 canine SaV and 1 sea lion SaV, however, only porcine GIII isolates with entire genome were reported in China.

In the present study, based on the metagenomic analyses and traditional RT-PCR method, we characterized the nearly full genome of a porcine GIII SaV which is divergent from the other reported GIII strains and the complete genome of a porcine GVII SaV which is firstly reported in China.

Methods

Fecal Specimens

Two fecal samples were collected from diarrheic pigs and were detected in different farms and different years, i.e., sample 1 from a 42-day-old pig on 2017 Mar and sample 2 was collected from a 75-day-old pig on 2015 NoV. Stool samples were suspended in 500-1000 µl Dulbecco's Phosphate Buffered Saline without Mg^{2+} and Ca^{2+} (Gibco, USA) vigorously vortexed and centrifuged at 10,000 g at 4 ℃ for 5 minutes. The supernatant was treated with Antibiotics (Gibco, USA) for 1 hour and

ACCEPTED MANUSCRIPT

filtered through 0.45-um filter (Millipore, USA) to remove bacteria and eukaryote.

Metagenomic analyses

The supernatant was then treated with a cocktail of DNase (Turbo DNase(Ambion), Baseline-ZERO (Epicetre) and benzonase (Novagen)) and RNase (Fermentas) to digest unprotected nucleotide at 37 ℃ for 90 minutes. Viral nucleic acids were extracted using QIAamp viral RNA extraction kit (Qiagen, USA) according to the manufacture's instruction. Then cDNA synthesis was performed as described previously[10]. Briefly, each viral nucleic acid pool was subjected to RT reaction using SuperScript III reverse transcriptase (Invitrogen, USA) and random hexamer primers followed by a single round of DNA synthesis with Klenow fragment polymerase (New England BioLabs). Then two libraries were constructed using the Nextera XT DNA Sample Preparation kit (Illumina) and sequenced using the MiSeq Illumina platform with 250 bp paired ends with a different tag for each pool and performed for bioinformatics analyses.

RT-PCR

To amplify the fragments (Fig. 1C) of strain p38, cDNA synthesis was exactly followed the protocols of Thermo Scientific RevertAid First Strand cDNA Synthesis Kit (Thermo, USA) with 1 μl random hexamer primer supplied by the kit. Then, semi-nested PCRs were performed using the PrimeSTAR HS or GXL DNA polymerase (Takara, Japan). Specific primers were designed based on the assembled contigs of p38 from the library except for 38-1F which was designed based on the reference strain ah-1(GenBank accession number: JX678943) (Table 1). A final volume of 50 μl of the reaction mixture contained 10 μl of 5x PrimeSTAR GXL buffer, 4 μl of dNTPs (2.5 mM each), 1 μl of forward primer (10 pmol/μl), 1 ul of reverse primer (10 pmol/μl), 1 ul PrimeSTAR GXL DNA polymerase (1.25 U/μl) and 2 μl of cDNA or the first round of PCR products. PCR were performed at 98℃ for 1 min followed by 30 cycles of 98 ℃ for 10 sec, 55 ℃ for 15 sec and 68 ℃ for 10 sec/kb, and a final extension at 68 ℃ for 7 min.

To amplify the 3'end of the two SaVs genome (indicated as 3' end in Fig. 1B and Fig. 1D), p38 and SH1703 cDNA was synthetized using the modified oligo dT primer, Q_T[11] (Table 1). Then, nested PCRs were performed with the gene specific forward

ACCEPTED MANUSCRIPT

primers designed based on the assembled region from the corresponding library and Q_O (Table 1) for the primary PCR and Q_I (Table 1) for the secondary PCR, using the PrimeSTAR HS or GXL polymerase (Takara, Japan) and the conditions were identical as described above.

To amplify the 5' terminal nt sequences of p38 and SH1703 genomes (marked as 5' end in Fig. 1 B and Fig. 1D), cDNA were synthesized using TaKaRa SMARTer RACE 5'/3' kit (TaKaRa, Japan) with 5'-CDS Primer A supplied by the kit and extended with the SMARTer A Oligonucleotide tail at the 3' end by the SMARTScribe Reverse Transcriptase. 5' RACE were performed by nested PCR: Long primer (forward primer, complementary to the 3' cDNA tail) and corresponding reverse primers 5UTR-W (Fig. 1B and Fig. 1D and Table 1) for the first PCR and Short primer (22 nt of 5' end of Long primer) and corresponding reverse primers 5UTR-N for the second PCR. PCR were performed with SeqAmp DNA polymerase according to the instruction of the kit.

Cloning and sequencing

All the PCR products were separated by agarose gel electrophoresis, purified using OMEGA Gel kit (OMEGA, USA) and cloned into the pMD19-T vector (Takara, Japan) or pEASY-Blunt Zero vector (Transgen, China) for sequencing by primer walking using universal M13+/- primers. For each product, at least three clones should be selected for sequencing. Sequences editing and assembly were performed by Lasergene v 8.1 (DNASTAR, Inc.).

Phylogenetic analyses and genotyping

Homologous hits of every sequence were carried out using the Basic Local Alignment Search Tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi). ClustalW v 2.1 was employed to perform the pairwise and multiple alignment for all the nucleotide and amino acid sequences. The construction of Maximum Likelihood trees of SaVs with 1000 bootstrap replicates and calculation of pairwise distances of complete VP1 nucleotide sequences were performed in MEGA version 5[5].

Nucleotide sequence accession numbers

The SaV nucleotide sequences determined in this study have been deposited in GenBank under accession numbers MF766258 for SH1703/GVII and MF766259 for

ACCEPTED MANUSCRIPT

p38/GIII.

Results

Genomic analyses of strain SH1703

Viral metagenomic analysis indicated that sample 1 was positive for porcine SaV and included 12642 SaV (designed SH1703) sequences reads generating a long contig of 6872 bp lacking the UTR sequences of the genome (Fig. 1A and 1B and Table 2). And the co-infection viruses and their reads percentages were listed in Table 2.

Then we successfully got the complete genomic sequences of strain SH1703 (Fig. 1A and Fig. 2). The entire RNA genome of strain SH1703 was 7184 bp excluding its poly (A) tail. The nucleotide composition of the full genome is A (24.33%), C (25.90%), U (23.82%), G (25.95%) with an overall C+G content of 51.85% (which was lower than that of 53.11% of strain K7). The 5'-UTR has 9 bases long starting with "GTG". Similar to the genomes of other SaVs, the SH1703 genome was predicted to have two ORFs. ORF 1 contained 6612 bases (nt 10-6621) and was translated a 2203-aa polyprotein included the putative capsid protein VP 1 (56 kDa). ORF 2, consisting of 507 bases (nt 6618-7124), has 4 bases overlapping with ORF 1(Fig. 1A) and was predicted to encode a 168-aa protein, VP 2 (Fig. 1A).

The phylogenetic tree based on the multiple alignment of the complete genome of 50 human and animal SaVs showed that SH1703 was clustered closely to GVII (Fig. 2 and Table 4). SH1703 shared 63.6%-64.1% nt identity with the complete genome sequence of available GVII strains RV0042 and K7. In order to determine whether SH1703 formed a separate cluster, the Maximum Likelihood tree was constructed based on the complete VP1 nucleotide sequences. SH1703 shared 62.4%-82.2% nt identities compared with the available GVII strains and it was closely related to the Korean strain DO19 (88.1% aa identities) (Fig. 3).

The first five amino acids of the putative NS 1 of SH1703 were "MAAMC", with the forth M clearly different from the available GVII strains (K7 and RV0042, MAA(V/T)C). The five amino acids of the putative cleavage site of RdRp and VP 1 were "YVME/G" (slash indicating the putative cleavage site) and this region was same with one GVII strain WGP247/KC309421 and one GIX strain

ACCEPTED MANUSCRIPT

WG214C/KC309418[5]. The characteristic aa motifs conserved among SaVs[5, 7] were shown in Table 3 and only minor deviations were observed in strain SH1703 compared with p38: in the VPg motif (DDEYDE), the second and third D was replaced with E and Q, respectively (Table 3), in the RdRp motif (WKGL), the G was replaced by D which was same with GXI porcine SaVs (Table 3) and motif (DYSKWDST), the K was substituted by Q. There were also seven highly conserved regions among all the reported available porcine SaVs: NTPase motif DH(Y/H)DXYTG, WDE(F/Y)D, LNCD, AFXRR and DXSHL; RdRp motif RLLWG(C/A)D and FLKR (data not shown).

Genomic analyses of strain p38

Viral metagenomic analysis indicated that sample 2 was positive for porcine SaV and contained 21 SaV sequences reads (named p38) covering 8 different contigs (Table 2 and Fig. 1D). We got the nearly complete genome of strain p38 (7305nt), lacking the 18 bp of the 5' end of the genome (aligned to other GIII strains available in GenBank database). We got the 5' UTR of strain SH1703 and p2 (GenBank no. KX688107) but we failed 5' RACE of p38 which may due to the low titer of virus. Same as the reported SaV strains, ORF2 of strain p38 overlapped its first four bases, ATGA, with the 3' end of ORF1.

Strain p38 was clustered into a separate cluster based on the analyses of entire VP1 nt sequences and showed genetically divergent from the other GIII strains (Fig. 3). The VP2 region (164 aa) was the same in length as in two USA strains (Cowden and LL14) and slightly shorter than other five Chinese strains (171-172 aa).

Aligned with the available reported SaV GIII strains, the five amino acids (FVMEA) surrounding the putative cleavage site between RdRp and VP1 were strongly conserved among the reported seven GIII strains[1, 5]. This new determined strain p38 had typical identified conserved amino acid motifs of NS3 (GPPGIGKT), NS5 (KGKNK and DDEYDE), NS6 (GDCG), NS7 (WKGL, KDEL, DYSKWDST, GLPSG and YGDD) and VP1 (PPG and GWS) (Table 3), which were considered as highly conserved among SaVs[5, 12].

Genotyping of SaVs

The Maximum Likelihood tree was constructed based on the entire VP1 nucleotide

ACCEPTED MANUSCRIPT

sequences of 137 SaV representative strains covering genogroup GI – GXIX (Fig. 3)[5, 7, 13] and the genetic pairwise distances were calculated using the Tamura-Nei method (Fig. 4). Genotyping numbers are updated based on the VP1 nt pairwise distances compared to those of already established genotypes.

SaVs were divided into 19 genogroups and 51 genotypes, among which pig SaVs were classified into 8 genogroups (GIII, GV-GXI) (Table 4). The genotypes numbers were assigned consecutively and showed definitely in Fig. 3.

Phylogenetic distances based on the nucleotide alignments showed that distances between p38 and other GIII strains were 0.157-0.222 and the intra-cluster distances of genogroup III were 0.000-0.234 (Fig. 4). Although strain p38 showed a little divergent from the other reported GIII strains and formed a separate cluster, we proposed that GIII strains were placed into a single genotype.

To investigate whether genogroup VII was classified into different genotypes, we compared the intra- and inter-cluster distances between the 137 published SaV VP1 nucleotide sequences. The distance between DO19 and SH1703 was 0.203 and the distances between SH1703 with any other strains of GVII were 0.333-0.543 which were significantly longer than any of the inter-cluster distances (0.031-0.270) of the already established 7 clusters of GI SaVs. Based on the calculated distances (0.333-0.588) among the inter-cluster strains of genogroup VII, they were classified into six sub-clusters. So we deduced that genogroup GVII could be divided into 6 genotypes and SH1703 represented a separate genetic cluster with strain DO19 and was tentatively named as GVII.5 (Fig. 3).

Besides, based on the critical nomenclature system of SaVs and comparison of the 137 SaVs, we tentatively proposed that the rat SaV isolation NYC-B2 clustered into GII.9 because the distances between NYC-B2 and all the other GII strains were 0.511-0.580. Considered that the distances between WG194D-1 with the other reported four genotypes of GV were 0.350-0.582 which is significantly longer than the intra- cluster distances (0.013-0.077) of GV, WG194D-1 may form a new cluster of genogroup V, GV.5. Ishi-lm9 grouped significantly further from other GVI strains, with a distance of 0.396-0.417, so we proposed that Ishi-lm9 represented a new genotype of GVI, GVI.2 (Fig. 3). WG214D formed separate from other GVIII strains

ACCEPTED MANUSCRIPT

(with distances of 0.499-0.549) and represented as a new genetic cluster within genogroup VIII of porcine SaV. There were also two separate clusters of GX (Fig. 3) and the inter-cluster distances were 0.327-0.339 which was significantly longer than the distances of established GI.1 intra-cluster distances (0.003-0.084) (Fig. 3 and Fig. 4). So we recommended that GX could be divided into two genotypes, GX.1 and GX.2(Fig.3). Distances based on nt alignments between 2053P4 or HgYa1-2 with other GXI strains were 0.426-475 and 0.426-467, respectively. The intra-cluster distances of the three strains (HgTa2-2-2, Ishi-lm7-3 and HgYa1-3) were 0.177-0.233. These data supported that GXI were clustered into three genotypes (Fig. 3). Although there were only two strains or two clusters of genogroups GIX, GXVII, GXVIII and GXIX, the distances of these two strains/clusters (0.504, 0.399, 0.312 and 0.374) were significantly longer than intra-cluster distances (0.011-0.030) of formed GII.8 SaVs, we assigned the four genogroups classified into two clusters (Fig.3 and Table 4).

Discussion

Sapoviruses were known as an important etiological agent associated with human and pig gastroenteritis[14-17]. Identification of some animal SaVs genetically related to human strains has raised awareness of the possibility of SaV interspecies transmission[7, 18, 19]. Here, we reported the nearly full genome of a porcine GIII strain and the complete genome of GVII strain detected in China and analyzed the genetic characterization of these two strains with other database available SaVs. Both of the two strains contain two ORFs and there were 4 bases of ORF 2 overlapping with ORF 1. ORF 1 encoded a polyprotein putatively processed into seven nonstructural (NS) proteins (NS1, NS2, NS3, NS4, NS5 and NS6-NS7) and the major structural protein VP1[1, 3, 20, 21]. The functions of protease and polymerase have been shown in vitro, but the biological functions of the other SaVs NS proteins have not been experimentally determined[22, 23]. However, these two strains have the identified typical conserved SaV motifs: NTPase motif (GPPGIGKT), VPg motif (GDCG), RdRp motif (KDEL, GLPSG, YGDD) and VP1motif (GWS) (Table 3). Besides, the G+C content of the newly described porcine SaVs falls in the typical G+C range of the genus of *Sapovirus* (49.0%-60.1%).

SaVs have been analyzed for classification based on partial RNA-dependent RNA

ACCEPTED MANUSCRIPT

polymerase (RdRp) or partial VP1 region[4, 6, 24-26]. Oka *et al.* proposed that sapovirus classification based on the complete capsid nucleotide sequences and human strains were clustered into 4 genogroups(GI, GII, GIV and GV)[13]. Lastly, Oka *et al.* suggested that human and animals SaVs could be divided into 15 genogroups based on available VP 1 amino acid sequences[5]. Then, Yinda *et al.* proposed additional four genogroups (in total 19 genogroups)[7]. In the current study, SaVs were analyzed based on the complete VP 1 nucleotide sequences and classified these 19 genogroups into 51 genotypes. SaV GI were sub-clustered into seven genotypes (GI.1 to GI.7) consistent with the previous study[12, 13, 27]. We proposed rat SaV strain NYC-B2 was designed in genotype GII.9 and GII could be divided into nine genotypes (GII.1 to GII.9) and based on the same criteria, we also named a new genotype of GV: GV.5[5, 25, 28] (the representative strain WG194D-1 shown in Fig. 3). In our study, based on the distances aligned among the 137 SaV strains, we firstly designed that GVII were clustered into six genotypes and SH1703 belonged to genotype GVII.5 (Fig. 3 and Table 4). and GXI could be divided into three genotypes based on the VP 1 nt pairwise distances. In addition, these genogroups, GVI, GVIII-GX and GXVII-GXIX, were clustered into two genotypes (Fig. 3 and Table 4). The other genogroups (GIII, GIV, GXII-GXVI) were placed into a single genotype (Fig. 3 and Table 4). Distances for genotyping nomenclature in our study correlated well with the SaV genotyping criteria (the statistically defined cutoff values for genotyping were 0.201-0.560) proposed by Oka *et al.*[13]. More data will be needed to test and prove the classification criteria of SaVs.

SaVs have been detected from many mammalian species, including human and seven kinds of animals (pigs, mink, dogs, sea lions, bats, chimpanzees and rats)[5, 7, 25, 29].However, the ability of SaVs transmitted between humans and animals is still unknown[5, 7, 12]. Until now, human SaVs were clustered into genogroup I, II, IV and V based on the analyses of complete VP1 amino acid sequences, compared with the available data from the GenBank database[5, 13] (Fig. 3). Whereas porcine SaVs showed highly phylogenetically divergent and were clustered into almost eight genoroups (GIII, GV-GXI)[5] (Fig. 3). Especially, some porcine SaVs shared 66-78% VP1 aa identities with human GV strains and were clustered into the same genogroup

ACCEPTED MANUSCRIPT

GV[5, 24] (Fig. 2 and Fig.3). Given that pigs were living in close to human, the authors proposed the possibility of pig as the reservoir for human SaVs or vice versa and suggested that there might have a cross-species transmission of these viruses. It is urgent to understand the replication and pathogenicity of SaVs to prevent their harm to humans, especially younger ones.

Since the discovery of porcine SaVs identified in 1980 in United States[8], porcine SaVs infection has been reported around the world in recent years, including China[15, 16, 30, 31]. However, at the beginning of the study, only genogroup III was detected in China[15, 32]. In the present study, we reported two strains detected from diarrheic pigs and phylogenetic analyses indicated that they were clustered into GIII and GVII, respectively. The GIII strain p38 showed divergent from previous reported Chinese strains and was closely clustered with two Japanese strains (85% VP 1 aa identities). To the best of our knowledge, this is the first report that GVII was detected in China and formed a separate cluster with one Korea strain (Fig. 3).

In conclusion, based on the metagenomic analyses and traditional RT-PCR, we determined one GIII porcine SaV which is divergent from the other reported GIII strains and one GVII porcine SaV which is firstly reported in China, to enrich the epidemiological data of SaVs. We also proposed that SaVs could be classified into 51 genotypes within 19 genogroups based on the analyses of complete VP1 nt sequences , which will be useful for genetic features and molecular classification of SaV data comparison worldwide.

Funding

This study was funded by the National Key Research and Development Program of China (2017YFC1200203), the National Natural Science Fund of China (Grant No.31270186, 31572525 and 31402211) and the Natural Science Fund of Jiangsu Province (No. BK20140578).

List of abbreviations

SaV: Sapovirus; NSs: Nonstructural proteins; GIII: genogroup III; RdRp: RNA-dependent RNA polymerase; ORF: Open reading frame; UTR: Untranslated

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

region; nt: nucleotides; aa: amino acid.

Competing interests

The authors declare that they have no competing interests.

Author Contributions

JJL and XGH conceived the study and designed the experiments. JJL and WZ performed the laboratory assays. JJL analyzed the data, performed the analysis and drafted the manuscript with the help of WZ and QS. All authors read and approved the final manuscript.

Figure 1. Diagram of putative genomic organization of the newly determined porcine SaVs SH1703 (A) and p38 (C) strains. Both of the two stains were a linear, positive sense and single stranded RNA with a poly A tail at the 3' end, encoding two ORFs. Genome sequences of SaVs, SH1703(B) and p38(D) were determined by *de novo* assembly of Illumina MiSeq sequences (indicated by black line), 5' RACE (indicated by green square), 3'RACE (indicated by green square) and primer-walking RT-PCR (indicated by blue square) as described in the text.

Figure 2. Phylogenetic tree of the complete genome of 50 SaVs was constructed using MEGA 5. All the reported porcine SaVs were highlighted, the newly determined in this study marked in yellow and the other porcine SaVs from GenBank database marked in blue, covering five genogroups: GIII, GV-GVIII. Each SaV strain is indicated in the following format: GenBank accession number - strain name - species - area.

Figure 3. Phylogenetic tree with 1000 bootstrap replicates based on the complete VP1 nucleotide sequences of 137 SaVs strains. The two newly identified SaVs strains (SH1703/MF766258 and p38/MF766259) are marked by yellow highlight and the newly proposed genotypes of each genogroup is indicated by red square (named based on the order of their discovery). Each SaV strain is presented as the following format:

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

GenBank accession number - strain name - species.

Figure 4 Pairwise distances of complete VP1 nucleotide sequences between intra- (yellow marked) and inter-genogroup (green marked) of SaVs

Table 1 Primers for amplifying the complete genome of porcine SaV SH1703 and p38

Table 2 Summary of samples in this study and SaV contigs identified in each sample obtained from deep sequencing

Table 3 Conversed motifs of predicted proteins of porcine SaVs

Table 4 Summary of relationship of genogroup, genotype, and no. of complete genome from different species

References

1. Oka T, Yamamoto M, Katayama K, Hansman GS, Ogawa S, Miyamura T, Takeda N. Identification of the cleavage sites of sapovirus open reading frame 1 polyprotein. J Gen Virol. 2006; 87(Pt 11): 3329-38.
2. Oka T, Yamamoto M, Yokoyama M, Ogawa S, Hansman GS, Katayama K, Miyashita K, Takagi H, Tohya Y, Sato H, Takeda N. Highly conserved configuration of catalytic amino acid residues among calicivirus-encoded proteases. Journal of Virology. 2007; 81(13): 6798-806.
3. Oka T, Murakami K, Wakita T, Katayama K. Comparative site-directed mutagenesis in the catalytic amino acid triad in calicivirus proteases. Microbiol Immunol. 2011; 55(2): 108-14.
4. Tse H, Chan WM, Li KS, Lau SK, Woo PC, Yuen KY. Discovery and genomic characterization of a novel bat sapovirus with unusual genomic features and phylogenetic position. PLoS One. 2012; 7(4): e34987.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5. Oka T, Lu Z, Phan T, Delwart EL, Saif LJ , Wang Q. Genetic Characterization and Classification of Human and Animal Sapoviruses. PLoS One. 2016; 11(5): e0156373.

6. L'Homme Y, Brassard J, Ouardani M , Gagne MJ. Characterization of novel porcine sapoviruses. Arch Virol. 2010; 155(6): 839-46.

7. Yinda CK, Conceicao-Neto N, Zeller M, Heylen E, Maes P, Ghogomu SM, Van Ranst M , Matthijnssens J. Novel highly divergent sapoviruses detected by metagenomics analysis in straw-colored fruit bats in Cameroon. Emerg Microbes Infect. 2017; 6(5): e38.

8. Saif LJ, Bohl EH, Theil KW, Cross RF , House JA. Rotavirus-like, calicivirus-like, and 23-nm virus-like particles associated with diarrhoea in young pigs.pdf. J CLIN MICROBIOL. 1980; 12(1): 105-11.

9. Guo M, Chang KO, Hardy ME, Zhang Q, Parwani AV , Saif LJ. Molecular characterization of a porcine enteric calicivirus genetically related to Sapporo-like human caliciviruses. J Virol. 1999; 73(11): 9625-31.

10. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S , Delwart E. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol. 2009; 83(9): 4642-51.

11. Scotto-Lavino E, Du GW , Frohman MA. 3 ' End cDNA amplification using classic RACE. Nat Protoc. 2006; 1(6): 2742-5.

12. Oka T, Wang Q, Katayama K , Saif LJ. Comprehensive review of human sapoviruses. Clin Microbiol Rev. 2015; 28(1): 32-53.

13. Oka T, Mori K, Iritani N, Harada S, Ueki Y, Iizuka S, Mise K, Murakami K, Wakita T , Katayama K. Human sapovirus classification based on complete capsid nucleotide sequences. Arch Virol. 2012; 157(2): 349-52.

14. Roos-Weil D, Ambert-Balay K, Lanternier F, Mamzer-Bruneel MF, Nochy D, Pothier P, Avettand-Fenoel V, Anglicheau D, Snanoudj R, Bererhi L, Thervet E, Lecuit M, Legendre C, Lortholary O , Zuber J. Impact of norovirus/sapovirus-related diarrhea in renal transplant recipients hospitalized for diarrhea. Transplantation. 2011; 92(1): 61-9.

15. Li J, Shen Q, Zhang W, Zhao T, Li Y, Jiang J, Yu X, Guo Z, Cui L , Hua X.

ACCEPTED MANUSCRIPT

Genomic organization and recombination analysis of a porcine sapovirus identified from a piglet with diarrhea in China. Virol J. 2017; 14(1): 57.

16. Valente CS, Alfieri AF, Barry AF, Leme RA, Lorenzetti E , Alfieri AA. Age distribution of porcine sapovirus asymptomatic infection and molecular evidence of genogroups GIII and GIX? circulation in distinct Brazilian pig production systems. Trop Anim Health Prod. 2016; 48(1): 21-7.

17. de Oliveira DMM, Souza M, Fiaccadori FS, Santos HCP , Cardoso DDD. Monitoring of Calicivirus Among Day-Care Children: Evidence of Asymptomatic Viral Excretion and First Report of GI.7 Norovirus and GI.3 Sapovirus in Brazil. Journal of Medical Virology. 2014; 86(9): 1569-75.

18. Firth C, Bhat M, Firth MA, Williams SH, Frye MJ, Simmonds P, Conte JM, Ng J, Garcia J, Bhuva NP, Lee B, Che X, Quan PL , Lipkin WI. Detection of zoonotic pathogens and characterization of novel viruses carried by commensal Rattus norvegicus in New York City. MBio. 2014; 5(5): e01933-14.

19. Bank-Wolf BR, Konig M , Thiel HJ. Zoonotic aspects of infections with noroviruses and sapoviruses. Vet Microbiol. 2010; 140(3-4): 204-12.

20. Oka T, Katayama K, Ogawa S, Hansman GS, Kageyama T, Ushijima H, Miyamura T , Takeda N. Proteolytic processing of sapovirus ORF1 polyprotein. J Virol. 2005; 79(12): 7283-90.

21. Oka T, Yokoyama M, Katayama K, Tsunemitsu H, Yamamoto M, Miyashita K, Ogawa S, Motomura K, Mori H, Nakamura H, Wakita T, Takeda N , Sato H. Structural and biological constraints on diversity of regions immediately upstream of cleavage sites in calicivirus precursor proteins. Virology. 2009; 394(1): 119-29.

22. Robel I, Gebhardt J, Mesters JR, Gorbalenya A, Coutard B, Canard B, Hilgenfeld R , Rohayem J. Functional characterization of the cleavage specificity of the sapovirus chymotrypsin-like protease. J Virol. 2008; 82(16): 8085-93.

23. Yokoyama M, Oka T, Kojima H, Nagano T, Okabe T, Katayama K, Wakita T, Kanda T , Sato H. Structural basis for specific recognition of substrates by

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

sapovirus protease. Front Microbiol. 2012; 3(312).

24. Oka T, Doan YH, Shimoike T, Haga K , Takizawa T. First complete genome sequences of genogroup V, genotype 3 porcine sapoviruses: common 5'-terminal genomic feature of sapoviruses. Virus Genes. 2017.

25. Kuroda M, Masuda T, Ito M, Naoi Y, Doan YH, Haga K, Tsuchiaka S, Kishimoto M, Sano K, Omatsu T, Katayama Y, Oba M, Aoki H, Ichimaru T, Sunaga F, Mukono I, Yamasato H, Shirai J, Katayama K, Mizutani T, Oka T , Nagai M. Genetic diversity and intergenogroup recombination events of sapoviruses detected from feces of pigs in Japan. Infect Genet Evol. 2017; 55(209-17).

26. Dey SK, Mizuguchi M, Okitsu S , Ushijima H. Novel recombinant sapovirus in Bangladesh. Clin Lab. 2011; 57(1-2): 91-4.

27. Kagning Tsinda E, Malasao R, Furuse Y, Gilman RH, Liu X, Apaza S, Espetia S, Cama V, Oshitani H , Saito M. Complete Coding Genome Sequences of Uncommon GII.8 Sapovirus Strains Identified in Diarrhea Samples Collected from Peruvian Children. Genome Announc. 2017; 5(43): e01137-17.

28. Phan TG, Trinh QD, Yagyu F, Okitsu S , Ushijima H. Emergence of rare sapovirus genotype among infants and children with acute gastroenteritis in Japan. Eur J Clin Microbiol. 2007; 26(1): 21-7.

29. Sisay Z, Djikeng A, Berhe N, Belay G, Gebreyes W, Abegaz WE, Njahira MN, Wang QH , Saif LJ. Prevalence and molecular characterization of human noroviruses and sapoviruses in Ethiopia. Arch Virol. 2016; 161(8): 2169-82.

30. Scheuer KA, Oka T, Hoet AE, Gebreyes WA, Molla BZ, Saif LJ , Wang Q. Prevalence of porcine noroviruses, molecular characterization of emerging porcine sapoviruses from finisher swine in the United States, and unified classification scheme for sapoviruses. J Clin Microbiol. 2013; 51(7): 2344-53.

31. Nakamura K, Saga Y, Iwai M, Obara M, Horimoto E, Hasegawa S, Kurata T, Okumura H, Nagoshi M , Takizawa T. Frequent detection of noroviruses and sapoviruses in swine and high genetic diversity of porcine sapovirus in Japan during Fiscal Year 2008. J Clin Microbiol. 2010; 48(4): 1215-22.

32. Zhang W, Shen Q, Hua XG, Cui L, Liu JF , Yang SX. The first Chinese

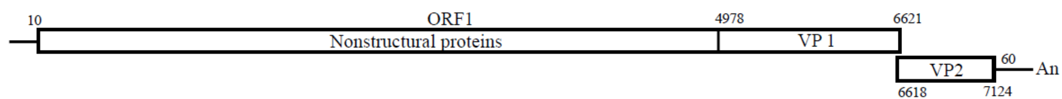
ACCEPTED MANUSCRIPT

porcine sapovirus strain that contributed to an outbreak of gastroenteritis in piglets. Journal of Virology. 2008; 82(16): 8239-40.

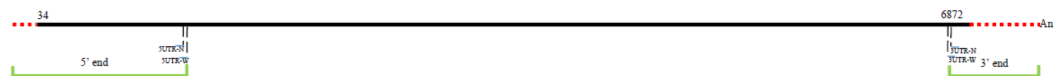
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

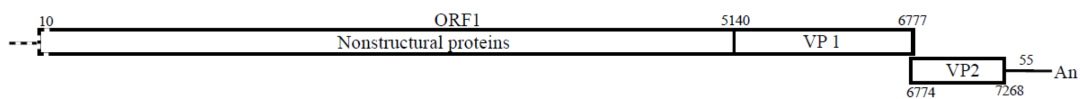
A. porcine SaV-SH1703/GVII



B. Genome determination of SH1703



C. porcine SaV-p38/GIII



D. Genome determination of p38

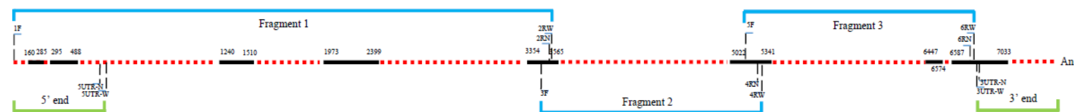


Figure 1

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

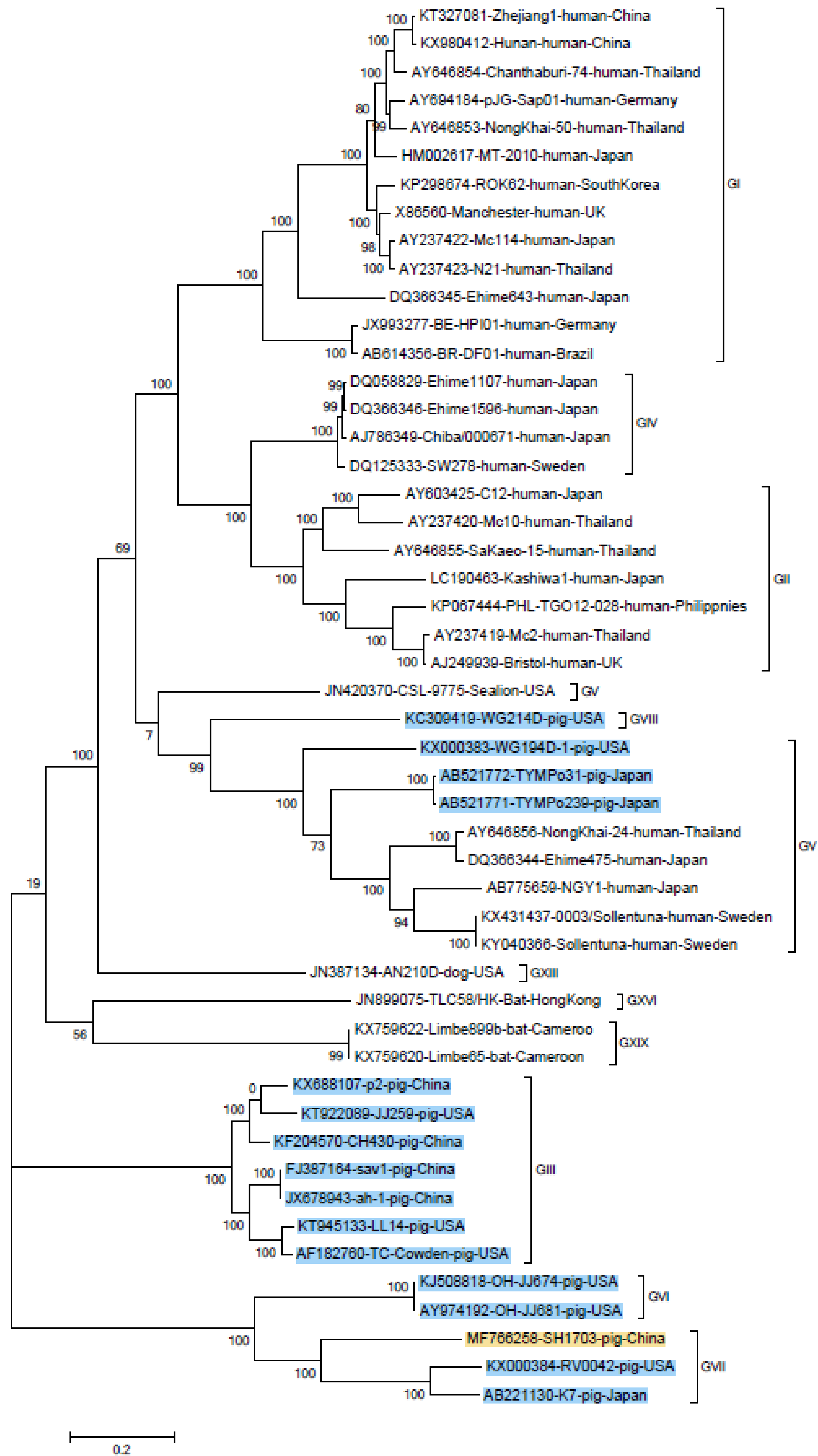


Figure 2

ACCEPTED MANUSCRIPT

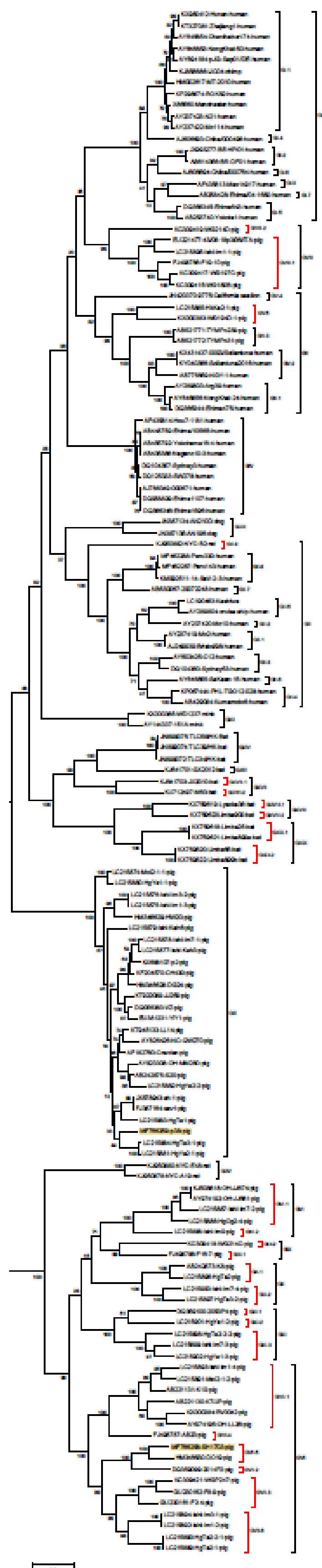


Figure 3

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

	GI	GII	GIII	GIV	GV	GVI	GVII	GVIII	GIX	GX	GXI	GXII	GXIII	GXIV	GXV	GXVI	GXVII	GXVIII	GXIX
GI	0.003-0.323																		
GII	0.642-0.748	0.011-0.580																	
GIII	0.742-0.895	0.766-0.890	0.000-0.234																
GIV	0.582-0.661	0.688-0.757	0.768-0.853	0.003-0.065															
GV	0.589-0.722	0.634-0.761	0.735-0.856	0.603-0.670	0.000-0.582														
GVI	0.915-1.017	0.853-0.995	0.880-0.995	0.855-0.935	0.911-1.032	0.000-0.417													
GVII	0.885-1.023	0.869-1.032	0.877-1.021	0.859-0.987	0.881-1.041	0.624-0.716	0.000-0.588												
GVIII	0.587-0.678	0.649-0.757	0.794-0.913	0.598-0.675	0.569-0.681	0.956-1.073	0.884-1.080	0.049-0.549											
GIX	0.905-1.033	0.902-1.007	0.883-1.046	0.934-1.013	0.906-1.013	0.549-0.656	0.592-0.688	0.940-1.066	0.504										
GX	0.804-0.951	0.904-1.053	0.847-0.982	0.822-0.885	0.849-0.944	0.612-0.691	0.643-0.738	0.856-0.971	0.599-0.6707	0.131-0.358									
GXI	0.855-1.017	0.891-1.032	0.904-1.023	0.897-0.975	0.918-1.037	0.590-0.684	0.602-0.707	0.898-1.061	0.605-0.676	0.622-0.700	0.177-0.475								
GXII	0.760-0.817	0.751-0.850	0.826-0.926	0.744-0.780	0.759-0.806	0.879-1.029	0.855-0.982	0.733-0.793	0.923-0.942	0.859-0.897	0.856-0.973	0.144							
GXIII	0.727-0.827	0.634-0.727	0.779-0.855	0.704-0.741	0.699-0.799	0.871-0.982	0.890-1.024	0.681-0.755	0.901-0.950	0.931-0.973	0.920-1.024	0.736-0.763	0.319						
GXIV	0.857-0.917	0.793-0.906	0.888-0.968	0.814-0.848	0.823-0.945	0.917-1.003	0.893-0.977	0.843-0.909	0.930-0.958	0.925-0.970	0.915-0.948	0.840-0.871	0.927-0.937	0.002-0.007					
GXV	0.916-0.989	0.909-1.012	0.864-0.936	0.862-0.930	0.906-1.048	0.716-0.792	0.722-0.815	0.903-1.066	0.718-0.866	0.790-0.851	0.752-0.833	0.890-0.959	0.960-1.009	0.945-0.948	0.152				
GXVI	0.830-0.912	0.838-0.954	0.846-0.919	0.840-0.870	0.865-0.927	0.957-1.002	0.880-0.972	0.864-0.936	0.985-1.013	0.906-0.951	0.948-1.009	0.897-0.922	0.889-0.890	0.561-0.564	1.040-1.052	0			
GXVII	0.822-0.907	0.811-0.901	0.814-0.916	0.808-0.843	0.837-0.914	0.973-1.054	0.934-1.054	0.802-0.923	0.975-1.007	0.916-0.982	0.957-1.005	0.861-0.966	0.872-0.917	0.665-0.713	0.925-0.992	0.662-0.674	0.399		
GXVIII	0.877-1.027	0.863-0.999	0.950-1.034	0.845-0.867	0.915-1.014	1.067-1.135	0.970-1.164	0.876-0.995	1.004-1.073	0.946-1.054	1.061-1.159	0.954-1.051	0.908-0.991	0.842-0.907	1.035-1.153	0.904-0.905	0.891-0.984	0.321	
GXIX	0.941-1.033	0.926-1.067	0.954-1.088	0.898-0.961	0.923-1.095	1.065-1.100	0.984-1.139	0.949-1.013	1.028-1.075	1.030-1.141	1.072-1.137	0.979-1.044	0.906-0.972	0.860-0.887	1.072-1.132	0.875-0.893	0.860-0.894	0.556-0.641	0.000-0.374

Figure 4

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1

	Primer name	Primer sequence (5'-3')	Position
SH1703	1703-5UTR-W	GTACAGATGTAATAGAACCCACAAAGG	939-965
	1703-5UTR-N	CAAGATCTGCAGACTGAAGTGAATATC	893-919
	1703-3UTR-W	CCCAATGGGAATTCGAGGGG	6580-6599
	1703-3UTR-N	TGTTTTGAACGGTTTGGGGC	6632-6651
p38	38-1F	GTGATCGTGATGGCTAATTGC	1-22
	38-2RW	GCAGTTGTTGTGCCGTTGTT	3466-3485
	38-2RN	TTGTAGTGTCCCAACTGCCC	3414-3433
	38-3F	ACTACAACCACTCGCGTGAC	3427-3446
	38-4RW	GTCATTGATTTGCCCTGTGGC	5272-5292
	38-4RN	GCATTGTACGTTGCCACTGC	5053-5072
	38-5F	AGTGTTCGTGATGGAGGCAC	5130-5149
	38-6RW	CCACTTGGTCCAATGAGGGG	6724-6743
	38-6RN	CAGAGTTGTTGCCTGTGTGC	6480-6499
	38-5UTR-W	CCAACCCAAGATGTGCGCGAGCCACG	692-717
	38-5UTR-N	GGGCGCTATCCGGCACACATTCCCAC	492-517
	38-3UTR-W	CAGGCCCAGATACAGTTGGC	6888-6907
	38-3UTR-N	GGCCAAAGACCAGTTGAAGC	6905-6924
5RACE	Long primer	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	Supplied by the kit
	Short primer	CTAATACGACTCACTATAGGGC	
3RACE	Q_T	CCAGTGAGCAGAGTGACGAGGAC TCGAGCTCAAGCT $_{17}$	[11]
	Q_O	CCAGTGAGCAGAGTGACG	
	Q_I	GAGGACTCGAGCTCAAGC	

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2

Sample name	Reads and sequences obtained from deep sequencing				SaV genogroup	Co-infected viruses and their reads percentage %
	Total reads	SaV reads	SaV reads %	Number of SaV contigs		
1	1,149,234	12,642	1.1	1	GVII	Astroviridae(1.87), Parvorividae(0.34), Circoviridae(0.30), Herpesviridae(0.01), Picornaviridae(0.006)
2	158,824	21	0.01	8	GIII	Picornaviridae(1.42), Astroviridae(0.77), Circocividae(0.66), Parvoviridae(0.32)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3

Genogroup	Strain	NS3-NTPase GAPGIGKT**	NS5-VPg KGKTK and DDEYDE	NS6-Protease GDCG	NS7-RdRp WKGL, KDEL, DYSKWDST, GLPSG and YGDD	VP1 PPG and GWS
SaV/GIII	**p38***	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1563}YGDD	^{1838}PPG and ^{1982}GWS
	HgYa2-1	^{463}GPPGIGKT	^{935}KGKNK and ^{952}DDEYDE	^{1160}GDCG	^{1204}WKGL, ^{1364}KDEL, ^{1439}DYSKWDST, ^{1494}GLPSG and ^{1542}YGDD	^{1837}PPG and ^{1981}GWS
	HgTa1	^{462}GPPGIGKT	^{934}KGKNK and ^{951}DDEYDE	^{1159}GDCG	^{1203}WKGL, ^{1363}KDEL, ^{1438}DYSKWDST, ^{1493}GLPSG and ^{1541}YGDD	^{1836}PPG and ^{1980}GWS
	sav1	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PPG and ^{1982}GWS
	HgYa2-2	^{460}GPPGIGKT	^{932}KGKNK and ^{949}DDEYDE	^{1157}GDCG	^{1201}WKGL, ^{1361}KDEL, ^{1436}DYSKWDST, ^{1491}GLPSG and ^{1539}YGDD	^{1834}PPG and ^{1978}GWS
	Cowden	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PPG and ^{1982}GWS
	LL14	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PPG and ^{1982}GWS
	JJ259	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PHG and ^{1982}GWS
	CH430	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PPG and ^{1982}GWS
	p2	^{464}GPPGIGKT	^{936}KGKNK and ^{953}DDEYDE	^{1161}GDCG	^{1205}WKGL, ^{1365}KDEL, ^{1440}DYSKWDST, ^{1495}GLPSG and ^{1543}YGDD	^{1838}PPG and ^{1982}GWS
	Ishi-Im7-1	^{463}GPPGIGKT	^{935}KGKNK and ^{952}DDEYDE	^{1160}GDCG	^{1204}WKGL, ^{1364}KDEL, ^{1439}DYSKWDST, ^{1494}GLPSG and ^{1542}YGDD	^{1837}PPG and ^{1981}GWS
	Ishi-Kah6	^{463}GPPGIGKT	^{935}KGKNK and ^{952}DDEYDE	^{1160}GDCG	^{1204}WKGL, ^{1364}KDEL, ^{1439}DYSKWDST, ^{1494}GLPSG and ^{1542}YGDD	^{1837}PPG and ^{1981}GWS
SaV/GV	HkKa2-1	^{475}GPPGIGKT	^{941}KGKTK	^{1176}GDCG	^{1220}WKGL, ^{1381}KDEL,	^{1866}PPG

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

			and ^{962}DDEYNE		^{1456}DYSKWDST, ^{1511}GLPSG and ^{1559}YGDD	and ^{2012}GWS
	WG194D-1	^{478}GPPGIGKT	^{944}KGKTK and ^{965}DDEYNE	^{1179}GDCG	^{1223}WKGL, ^{1384}KDEL, ^{1459}DYSKWDST, ^{1514}GLPSG and ^{1562}YGDD	^{1869}PPG and ^{2015}GWS
	TYMPo239	^{478}GPPGIGKT	^{944}KGKTK and ^{965}DDEYNE	^{1179}GDCG	^{1223}WKGL, ^{1384}KDEL, ^{1459}DYSKWDST, ^{1514}GLPSG and ^{1562}YGDD	^{1866}PPG and ^{2011}GWS
	TYMPo31	^{478}GPPGIGKT	^{944}KGKTK and ^{965}DDEYNE	^{1179}GDCG	^{1223}WKGL, ^{1384}KDEL, ^{1459}DYSKWDST, ^{1514}GLPSG and ^{1562}YGDD	^{1866}PPG and ^{2011}GWS
SaV/GVI	OH-JJ674	^{456}GPPGIGKT	^{923}KGKNK and ^{937}DDEYQE	^{1122}GDCG	^{1169}WKGL, ^{1326}KDEL, ^{1402}DYSKWDST, ^{1457}GLPSG and ^{1505}YGDD	^{1798}PPR and ^{1941}GWS
	OH-JJ681	^{456}GPPGIGKT	^{923}KGKNK and ^{937}DDEYQE	^{1122}GDCG	^{1169}WKGL, ^{1326}KDEL, ^{1402}DYSKWDST, ^{1457}GLPSG and ^{1505}YGDD	^{1798}PPR and ^{1941}GWS
	HgOg2-4	^{449}GPPGIGKT	^{916}KGKNK and ^{930}DDEYQE	^{1115}GDCG	^{1162}WKGL, ^{1319}KDEL, ^{1395}DYSKWDST, ^{1450}GLPSG and ^{1498}YGDD	^{1791}PPR and ^{1934}GWS
	Ishi-Im9	^{435}GPPGIGKT	^{902}KGKNK and ^{916}DDEYQE	^{1101}GDCG	^{1148}WKGL, ^{1305}KDEL, ^{1381}DYSKWDST, ^{1436}GLPSG and ^{1484}YGDD	^{1777}PPR and ^{1920}GWS
SaV/GVII	**SH1703**	^{451}GPPGIGKT	^{918}KGKTK and ^{931}DEEYQE	^{1116}GDCG	^{1162}WKDL, ^{1318}KDEL, ^{1394}DYSQWDST, ^{1449}GLPSG and ^{1497}YGDD	^{1786}PPG and ^{1928}GWS
	HgTa2-2-1	^{431}GPPGVGKT	^{898}KGKNK and ^{912}DDEYTE	^{1095}GDCG	^{1140}WKGL, ^{1296}KDEL, ^{1372}DYSRWDST, ^{1427}GLPSG and ^{1475}YGDD	^{1764}PPG and ^{1906}GWS
	Ishi-Im3-1	^{451}GPPGVGKT	^{918}KGKNK and ^{932}DDEYTE	^{1115}GDCG	^{1160}WKGL, ^{1316}KDEL, ^{1392}DYSRWDST, ^{1447}GLPSG and ^{1495}YGDD	^{1784}PPG and ^{1926}GWS
	RV0042	^{448}GPPGIGKT	^{915}KGKNK and ^{928}DEEYQE	^{1113}GDCG	^{1159}WKGL, ^{1315}KDEL, ^{1391}DYSQWDST, ^{1446}GLPSG and ^{1494}YGDD	^{1785}PPG and ^{1927}GWS
	K7/JP	^{448}GPPGIGKT	^{915}KGKNK and ^{928}DEEYQE	^{1113}GDCG	^{1159}WKGL, ^{1315}KDEL, ^{1391}DYSKWDST, ^{1446}GLPSG and ^{1494}YGDD	^{1785}PPG and ^{1927}GWS
	Ishi-Im1-4	^{444}GPPGIGKT	^{911}KGKNK and ^{924}DEEYQE	^{1109}GDCG	^{1155}WKGL, ^{1311}KDEL, ^{1387}DYSKWDST, ^{1442}GLPSG and ^{1490}YGDD	^{1781}PPG and ^{1923}GWS
SaV/GVIII	WG214D	^{495}GPPGIGKT	^{957}KGKTK	^{1189}GDCG	^{1233}WKGL, ^{1394}KDEL,	^{1875}PPG

ACCEPTED MANUSCRIPT

			and ^{978}DDEYDE		^{1469}DYSKWDST, ^{1524}GLPSG and ^{1572}YGDD	and ^{2020}GWS
	Ishi-Im1-1	^{474}GPPGIGKT	^{935}KGKTK and ^{956}DDEYNE	^{1168}GDCG	^{1212}WKGL, ^{1373}KDEL, ^{1448}DYSKWDST, ^{1503}GLPSG and ^{1551}YGDD	^{1856}PPG and ^{2001}GWS
SaV/GIX	WG214C	-***	-	-	^{12}WKGL, ^{168}KDEL, ^{244}DYSKWDST, ^{299}GLPSG and ^{347}YGDD	^{640}PPG and ^{738}GWS
	F16-7	-	-	-	- - - ^{48}GLPSG and ^{96}YGDD	^{388}PPG and ^{531}GWS
SaV/GX	HgTa2	^{446}GPPGIGKT	^{913}KGKNK and ^{926}DDEYTE	^{1108}GDCG	^{1153}WKGL, ^{1309}KDEL, ^{1385}DYSKWDST, ^{1440}GLPSG and ^{1488}YGDD	^{1778}PPG and ^{1921}GWS
	HgTa3-2	^{431}GPPGIGKT	^{898}KGKNK and ^{911}DDEYTE	^{1093}GDCG	^{1138}WKGL, ^{1294}KDEL, ^{1370}DYSKWDST, ^{1425}GLPSG and ^{1473}YGDD	^{1763}PPG and ^{1906}GWS
SaV/GXI	HgTa2-2-2	^{447}GPPGIGKT	^{914}KGKNK and ^{927}DEEYQE	^{1112}GDCG	^{1158}WKDL, ^{1314}KDEL, ^{1390}DYSQWDST, ^{1445}GLPSG and ^{1493}YGDD	^{1788}PPG and ^{1930}GWS
	Ishi-Im7-3	^{449}GPPGIGKT	^{916}KGKNK and ^{929}DEEYQE	^{1114}GDCG	^{1160}WKDL, ^{1316}KDEL, ^{1392}DYSQWDST, ^{1447}GLPSG and ^{1495}YGDD	^{1790}PPG and ^{1932}GWS

*The strains were determined in this study are bolded.

**Boxed capitals: deviant amino acid residues of conserved motifs among SaVs and values before motifs: indicate their position in the genome sequence.

***Not available

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 4

	Genogroup	Genotype	No. of complete genome from each species								Total
			human	chimp	rat	pig	sea lion	mink	dog	bat	
	GI	7	13	0							13
	GII	9	7		0						7
	GIII	1				7					7
	GIV	1	4								4
	GV	5	5			3	1				9
	GVI	2				2					2
	GVII	6				3					3
	GVIII	2				1					1
	GIX	2				0					0
	GX	2				0					0
	GXI	3				0					0
	GXII	1						0			0
	GXIII	1							1		1
	GXIV	1								1	1
	GXV	1			0						0
	GXVI	1								0	0
	GXVII	2								0	0
	GXVIII	2								0	0
	GXIX	2								2	2
Total	19	51	29	0	0	16	1	0	1	3	50

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights:

- Two porcine SaVs from diarrheic pigs were identified.
- GVII strain with complete genome was firstly reported in China.
- It is firstly proposed that SaVs were divided into 51 genotypes within 19 genogroups.

ACCEPTED MANUSCRIPT