수의학석사학위논문

Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea

국내 돼지 분변에서의 원형 단일가닥 DNA 바이러스의 게놈 분석

2015년 2월

서울대학교 대학원
수의학과 수의미생물학 전공
김 아 름

Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea

국내 돼지 분변에서의 원형 단일가닥 DNA 바이러스의 게놈 분석

By

A Reum Kim

February, 2015

Department of Veterinary Medicine

The Graduate School

Seoul National University

Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea

By

A Reum Kim

Supervisor: Professor Bong Kyun Park, D.V.M., M.Sc., Ph.D.

A dissertation submitted to the faculty of the Graduate School of Seoul National University in partial fulfillment of the requirement for the degree of Master in Veterinary Microbiology

February, 2015

Department of Veterinary Medicine

The Graduate School

Seoul National University

수의학석사학위논문

Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea

국내 돼지 분변에서의 원형 단일가닥 DNA 바이러스의 게놈 분석

지도교수: 박 봉 균

이 논문을 수의학석사 학위논문으로 제출함

2014년 11월

서울대학교 대학원
수의학과 수의미생물학 전공
김 아 름

김아름의 석사 학위논문을 인준함

2014년 12월

위 원 장: 김 재 홍 (인)

부위원장: 박 봉 균 (인)

위 원: 유 한 상 (인)

Abstract

Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea

A Reum Kim
Department of Veterinary Medicine
The Graduate School of
Seoul National University

Porcine circular single-stranded DNA viruses have been just identified from swine feces in Korea. This virus was mentioned as bovine stool-associated circular DNA virus (BoSCV)-like virus discovered from porcine stools. However, the thorough characteristics of the virus were not identified. Therefore, this research focuses on finding a full genome sequence and analyzing the genetic features of the virus. The virus, now called porcine stool-associated circular DNA virus in Korea (PoSCV Kor), consists of 2,589 bases forming circular structure. It has two major ORFs inversely encoding replicase and capsid protein, with each stem-loop structure between 5' ends and 3' ends of the two putative ORFs. This characteristics is the same as PoSCV in New Zealand, but different from chimpanzee stool-associated circular virus (ChiSCVs) and BoSCV, which have one stem-loop structure. Therefore, it would be sure that PoSCV Kor is very similar to

PoSCV in respect to the genetic aspect; the same number of nucleotide bases and high amino acid identity of replicase and capsid protein (96 and 93%, respectively). This fact could be certified through the finding that PoSCV Kor and PoSCV are in the same cluster by phylogenetic analysis based on the comparison with full-sequences of other circular ssDNA viruses.

Keywords: PoSCV, Korea, Genetic analysis, Circular DNA virus, Complete genome

Student ID: 2013-21545

Contents

List of Tables

List of Figures

INTRODUCTION

Viruses with small circular DNA genome have been known to infect human, animals, even plants. They have been proposed to be classified into four types by phylogenetic relationships. Among them, *Circovirus*, *Geminivirus*, and *Nanovirus*, which are included into type 1, have a small circular DNA genome with the replicase protein and the major open reading frame (ORF) in opposite orientations (Ilyna & Koonin, 1992; Orozco & Hanley-Bowdoin, 1998). However, unlike *Geminivirus* and *Nanovirus* infecting plants (Blinkova *et al*., 2010), *Circovirus* has infected vertebrates including avian, porcine, even fish (Lorincz *et al*., 2011; Neumann *et al*., 2007; Segales & Domingo, 2002; Todd *et al*., 2001). *Cyclovirus*, which has been supposed to be a newly emerging circular DNA virus, has also been known to infect chimpanzees, humans, and farm animals; others have been reported in dragonflies and bats (Dayaram *et al*., 2013; Ge *et al*., 2011; Li *et al*., 2011; Rosario *et al*., 2009; Thompson *et al*., 1994). From this fact of a wide range of hosts, the potential of cross-species transmission of these circular viruses has been carefully mentioned (Li *et al*., 2010; Li *et al*., 2011).

Novel circular single-stranded DNA viruses have been identified for recent years through a metagenomic method, which has been considered a breakthrough way for identifying several unknown viruses, not interfering hosts or animals for sampling (Blinkova *et al*., 2010; Delwart, 2007; Edwards & Rohwer, 2005; Svraka *et al*., 2010). Studies of fecal samples from a broad range of animals such as chimpanzees, rodents, bovines, bats, badgers, and pine martens, and pigs have contributed to the discoveries of new circular single-stranded DNA viruses (Blinkova *et al*., 2010; Breitbart *et al*., 2008; Breitbart *et al*., 2003; Ge *et al*., 2011; Kim *et al*., 2012; Li *et al*., 2010; Li *et al*., 2011; Nakamura *et al*., 2009; Rosario *et al*., 2009; Sachsenroder *et al*., 2012; Sikorski *et al*., 2013; Zhang *et al*., 2006). Some of them might provide clues to the possibility of cross-infection among animals (Sikorski *et al*., 2013).

In Korea, a novel circular single-stranded DNA virus was first introduced from bovine stool samples, which has been now named as bovine stool-associated circular ssDNA virus (BoSCV) (Kim *et al*., 2012). Besides, pig and porcine stool-associated circular single-stranded viruses (PigSCV and PoSCV) have been identified in Germany (Sachsenroder *et al*., 2012) and New Zealand (Sikorski *et al*., 2013), respectively. These were all disclosed through random PCR and deep sequencing using specially designed primers. BoSCV contained 2,600 bases with two major ORFs, one of which has been anticipated to encode capsid protein and the other might be replicase-associtaed protein. The each genome of PigSCV and PoSCV has a length of 2,459 and 2,589 bases, respectively, containing two major ORFs of the putative capsid protein and replicase. However, the diverse patterns of genomic constitution demonstrate that they should be phylogenetically different viruses.

In this study, a porcine stool-associated circular single-stranded virus was detected in Korean swine farms, and analyzed in terms of its genomic structure and phylogeny compared with other circular single-stranded viruses.

MATERIALS AND METHODS

Porcine Samples

Porcine samples were collected from July 2012 to March 2013 (n=388) for 9 months from commercial farms in 9 provinces of Korea, and sampling was performed by using porcine stool.

DNA extraction

All these collections of samples were conducted for the extraction of DNAs in feces as follow; Proteinase K solution 8μl, DNA lysis buffer 500μl, and a stool sample 200μl were mixed. And, the mixture was thoroughly vortexed and incubated for 1h. And then, 200μl of phenol-chloroform-isoamyl alcohol (25:24:1) was added into the preparation, thoroughly vortexed, and centrifuged 10min, 13,000 rpm. DNA in aqueous phase was precipitated with an equal volume of isopropanol, and centrifuged again. The resulting DNA pellet was washed with 1ml of 70% ethanol, centrifuged, dried, and resuspended in 30μl TE buffer.

Nested PCR and detection of the existence of the virus

To assess the prevalence of PoSCV, these extracted DNA samples followed by being quantified into 100ng/μl were analyzed using a nested PCR assay with specific primers. Using a Maxime PCR PreMix kit (iNtRON), first round of PCR was done with 1F(forward)/1R(reverse) primers (Table 1), containing 1μl of 1F and 1R primers, respectively, 16μl of iStarMax, and 2μl of sample DNA template as follows: 95℃ for 5min; 38 cycles of 95℃ for 30s, 54℃ for 30s, 72℃ for 30s. After then, at the second round, the products were 20-time diluted for the detection of more partial specific sequences of PoSCV using 2F/2R primers, containing 1μl of 2F and 2R primers, respectively, 17μl of iStarMax, and 1μl of diluted products of first round as follows: 9

5℃ for 3min; 30 cycles of 95℃ for 20s, 58℃ for 30s, 72℃ for 1min 30s. Products from a nested PCR were visualized by electrophoresis on a 1.5% agarose gel. PCR products of the expected size were 253 bases and directly asked for sequencing (Fig. 1).

Table 1. List of primers used for nested PCR in this study

Primer	Sequence(5'-3')	Size
1F	CTGATGGTTGCTCCGGTTAT	578 bp
1R	CTCTCCAGGTTTCGTTCGAG	
2F	CCGGTTCTTCCATTCTCAAC	253 bp
2R	TCTCAGTATCGTGCTGGTGC	

**This primer sets were designed based on PoSCV genome sequence in this own study.*

PCR for full-sequencing

Positive DNA samples ascertained from a nested PCR were used for this stage. The two sets of PCR primers were designed based on the full-sequence of PoSCV. One, which is regarded as the first primer set, consisted of 1F' (5'-TTCGGCATGGTGT TTGTG-3') / 1R' (5'-ACGAGCACCATACAGAGCACTA-3') and the other indicated as the second set was 2F' (5'-GAAGATATTGGGCCGATTACTACA-3') / 2R' (5'-TAACTC CGCCGACAATCG-3') (Table 2). Using a Maxime PCR PreMix kit (iNtRON), PCR containing 1μl of 1F' and 1R' primers, respectively, 16μl of iStarMax, and 2μl of sample DNA was performed as follows: with the first primer set, 95℃ for 3min; 41 cycles of 95℃ for 20s, 53℃ for 30s, 72℃ for 1min 10s, and 72℃ for 5min. Next, making use of reaction agents prepared as above with the second set, PCR was processed on two rounds. At the first round, 95℃ for 3min; 28 cycles of 95℃ for 20s, 57℃ for 30s, 72℃ for 1min 20s, and 72℃ for 4min. Then, at the second round, the first products were 20-time diluted and PCR was performed under identical conditions above. Products from both the first set and second round of the second set were visualized by electrophoresis on a 2.0% agarose gel. The first set and second round of the second set could make up about 1,000 and 1,800 bases, respectively. After identifying samples of these sizes, they were directly asked for sequencing.

Table 2. List of primers used for full-sequencing in this study

Primer		Sequence(5'-3')	Size
#1			
	1F'	TTCGGCATGGTGTTTGTC	1,096 bp
	1R'	ACGAGCACCATACAGAGCACTA	
#2			
	2F'	GAAGATATTGGGCCGATTACTACA	1,835 bp
	2R'	TAACTCCGCCGACAATCG	

**This primer sets were designed based on PoSCV genome sequence in this own study.*

TA cloning for sequencing

After identifying DNA fragments corresponding to 1,000 and 1,800bp, respectively, these DNA samples were used for TA cloning, utilizing commercial TA cloning kit (Topcloner TA kit; Enzynomics) and subsequently transformed into competent *Escherichia coli* cells (DH5α). After incubating transformed cells overnight, colonies were observed and collected a little into PCR tube containing DEPC-DW. The selected recombinant plasmids were sequenced using M13F and M13R universal primers following protocols: 95℃ for 2min; 38 cycles of 95℃ for 30s, 55℃ for 30s, 72℃ for 1min 20s, and 72℃ for 5min. The results of sequencing were asked to MACROGEN. (Kim *et al*., 2012).

Phylogenetic analysis and protein function prediction

From the results of sequencing, several fragments of sequences were edited and trimmed compared to the reference sequence of PoSCV using BioEdit Sequence Editor 7.0 software. This PoSCV Kor J481 sequence and reference sequences of other circular ssDNA viruses, which were obtained from NCBI, were aligned by using the CLUSTAL W (Thompson *et al*., 1994). This could be analyzed for drawing phylogenetic tree, utilizing neighbor-joining model with 1,000 bootstrap values by Molecular Evolutionary Genetics Analysis 5 (Anisimova & Gascuel, 2006; Hall, 2013; Tamura *et al*., 2011). Putative ORFs in the genomes from this study were predicted using GLIMMER with the condition that minimum ORF size was 200 nucleotides (Delcher *et al*., 1999). The stem-loop structure with a nonamer sequence was predicted using the Mfold webserver (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/dna-forml.cgi) (Zuker, 2003). Next, putative protein functions were predicted utilizing BLASTp. The predicted ORF regions and stem-loop structures were drawn by pDRAW32 program(http://www.acaclone.com). Estimates of evolutionary divergence between partial replicase sequences of 4 samples including PoSCV Kor J481 and those of others, which were referred from NCBI, were calculated using *p*-distance model of pairwise distances in MEGA 5 (Blinkova *et al*., 2010)

To get the nucleotide and deduced amino acid identities of partial replicase protein, sequences of samples in this study and other circular ssDNA viruses were aligned using CLUSTAL W in BioEdit and used for Megalign in DNAstar program. Percentage identities of the aligned partial nucleotides and amino acids of replicase protein were saved and transformed into MS Exel file format, showing in the upper and lower triangles, respectivley.

Nucleotide sequence accession number

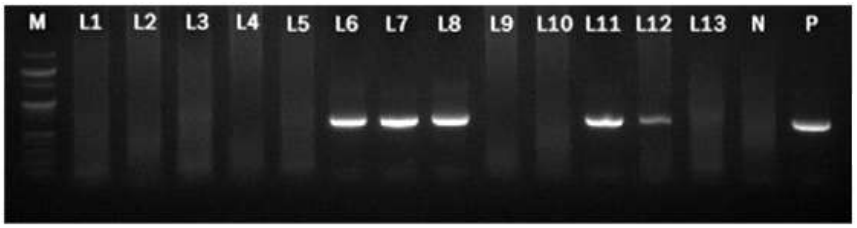
The genome sequence of PoSCV Kor J481 was deposited in GenBank under the Accession No. KF193403.

RESULTS

Detection and sequence analysis of circular DNA genome

Porcine samples were collected from July 2012 to March 2013 (n=388) for 9 months from commercial farms in 9 provinces of Korea. DNAs were extracted from them, processed according to the manufacturer's instructions. The samples showing the positive from second round of a nested PCR were checked and directly asked for sequencing (Fig. 1). Besides, to get the complete sequence from positive samples, another two sets of primer were set up. Each primer set generated ~1,000 and 1,800 bp in length, respectively. The acquired sequences were edited and trimmed based on the genome of PoSCV using BioEdit. The genome contained the two predicted major ORFs. After analyzing by BLASTp, it was found that the ORF 1 was similar to the putative capsid protein of PoSCV (93% amino acid identity) and ChiSCVs (26-50% amino acid identity) and the ORF 2 was similar to the putative replicase protein compared with that of PoSCV (96% identity) and ChiSCVs (44-45% identity).

(a)



(b)

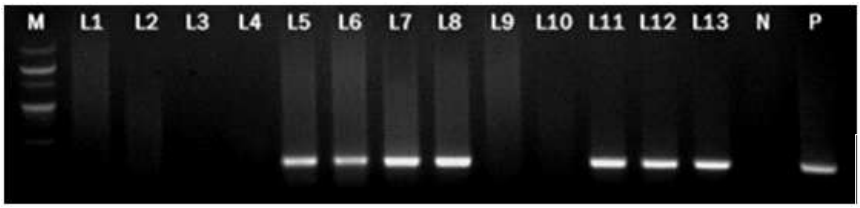


Figure 1. Examples of nested PCR result for PoSCV detection.

(a) First round of a nested PCR. (b) Second round of a nested PCR; Lines from L1 to L13 mean samples researched in this study. *M* ladder of DNA marker, *N* negative control, *P* positive control

Screening of stool DNA samples for PoSCV Kor

From the results of a nested PCR, the prevalence rates of the virus according to ages and months were arranged into a table (Table 3). The numbers of swine farms of the positive for PoSCV were presented on Korean map according to provinces (Fig. 2). Total stool samples were 388 and the positive were 178 among them, resulting in 45% of total positive rate. In this findings, there were no significant differences in ages and provinces showing average positive rate; based on ages, 53% positive rate in gilt, 50% in sow, 42% in suckling pigs, 55% in weaner, 51% in grower, 48% in finisher, and 21% in unknown indivisuals; according to provinces, positive rates ranging 31-55% appeared in 9 provinces. However, the monthly positive rates presented that it tends to occur with higher frequency in summer season than in winter season; through July to September, the positive rate of PoSCV Kor was ~80%, but in January and Fabruary, the rates was equal or < 10%.

Table 3. Numbers of samples and the positive from nested PCR conducted according to ages and months

Age	Percentage (positive / No. of samples)	Month	Percentage (positive / No. of samples)
Gilt	53 (16 / 30)	July	81 (44 / 54)
Sow	50 (18 / 36)	August	82 (42 / 51)
Suckling pigs	42 (40 / 95)	September	80 (4 / 5)
Post-weaning	55 (46 / 83)	October	53 (8 / 15)
Grower	51 (22 / 43)	November	41 (24 / 58)
Finisher	48 (26 / 54)	December	45 (19 / 42)
Unknown	21 (10 / 47)	January	40 (6 / 15)
		Fabruary	8 (4 / 45)
		March	26 (27 / 103)

Unknown means that profiles of samples could no be accurately identified.

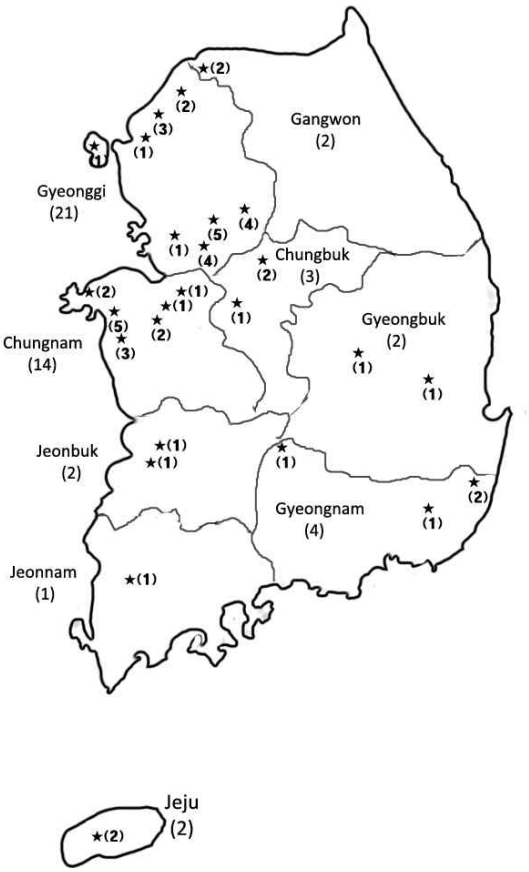
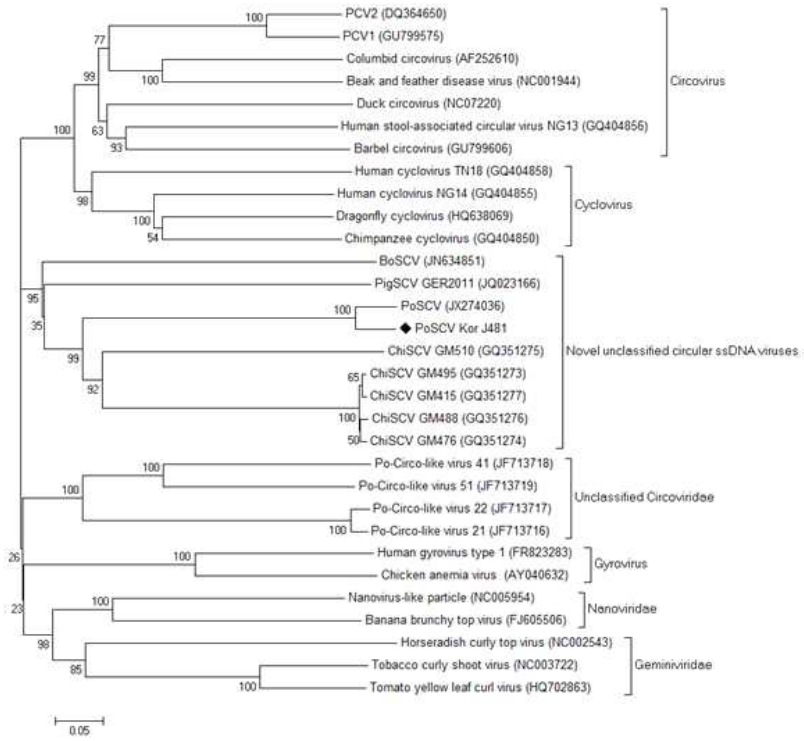


Figure 2. Swine farms representing the positive for PoSCV in nine provinces. The number of farms showing the positive for PoSCV in each province is presented in *parentheses* below the name of province. *Asterisks* represent cities where positive farms are located and the numbers of farms is shown in *parentheses*

Phylogenetic analysis of PoSCV Kor

Phylogenetic analysis (Fig. 3a) of the complete sequence showed that PoSCV Kor was nearly perfectively homologous to PoSCV detected from New Zealand, confirming that it might be also a novel single-stranded circular DNA virus. In contrast, despite the genome obtained from porcine feces in the same way to PigSCV, PoSCV Kor and PigSCV showed the significant divergence, aligned in different clades. Another phylogenetic analysis (Fig. 3b) was performed based on sequences from PCR products of other samples, J284, J322, and J490, which were supposed to include partial genes encoding replicase proteins. They were acquired by using a primer set (1F'; forward and 1R'; reverse) ranging 1,867-373nt. These partial replicase sequences and PoSCV Kor were compared to other replicase sequences of single-stranded circular viruses, which were acquired from NCBI. They showed high close relation with PoSCV and also did with ChiSCVs than others.

(a)



(b)

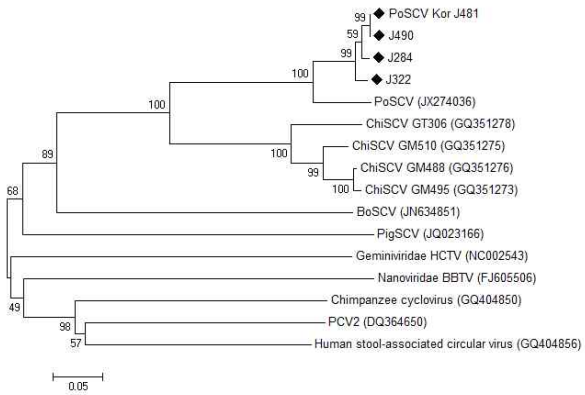


Figure 3. Phylogenetic analysis of the complete sequence and partial replicase sequences of PoSCV. (a) Neighbor-joining phylogenetic tree of PoSCV Kor (Accession No. KF193403) and other circular single-stranded DNA viruses, based on the complete genome sequences using *p*-distances model in MEGA 5. (b) Phylogenetic relationships among partial replicase sequences of PoSCV Kor and searched samples. *Filled diamond* indicates samples detected in this study

Genome organization of PoSCV Kor

PCR process using two specific primer sets, which were designed based on the reference genome sequence of PoSCV, produced ~1,000 and 1,800bp in length, respectively. Each fragment was sequenced by MACROGEN (Seoul, Korea) and both were fully aligned using CLUSTAL W in BioEdit. The length of the completed genome, which has been named as PoSCV Kor, was 2,589bp, demonstrating the same size as that of PoSCV. Analysis of the PoSCV Kor genome revealed that the genome encoded two primary ORFs reversely. ORF 1 encoded 349 amino acids in length ranging from 484 to 1,531 bp of nucleotide. Comparing PoSCV Kor with PoSCV based on BLASTp, the amino acid identity of putative capsid protein was 93%; decreased identities (26-50 and 30%) were obtained for ChiSCVs and PigSCV, respectively. The putative replicase gene, ORF 2, encoding a protein of 243 amino acids, showed 96% of the amino acid identity compared to that of PoSCV; much lower identities (2, 44-45 and 30%) were observed in comparison to ChiSCVs and BoSCV, respectively (Table 4). This genome encoded six additional ORFs without any significant similarities to previously reported proteins. PoSCV Kor included two putative intergenic regions between ORF 5' ends and ORF 3' ends, respectively, with the similarity to PoSCV and BoSCV. Two stem-loop structures with the conserved nonanucleotide, which was perfectively the same as PoSCV, were observed in ORF 5' and 3' ends downstream intergenic region, respectively (Fig. 4).

Table 4. Summary of nucleotide and amino acid sequence identities in the full replicase region between PoSCV Kor identified in this study and references of other circular ssDNA viruses

Strains	Percentage identity (%)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) PoSCV Kor†	***	**94.5**	63.3	62.9	63	53.4	47.7	52.6	30.9	32.6	40.1	34.4	22.4
(2) PoSCV	**96.3**	***	63.3	63.1	63.4	54.3	47.9	53.2	31.1	32.9	40.1	33.8	22.9
(3) GM488‡	45.7	46.9	***	98.2	96.5	62.3	45.9	51.2	32.3	34.8	41.3	35.4	25
(4) GM495‡	45.7	46.9	100	***	95.1	62.4	45.7	51.1	32.3	34.6	41.2	35	24.9
(5) GM510‡	44.9	46.1	97	97	***	62.1	45.6	51.2	32.3	34.6	40.7	34.5	24.8
(6) GT306‡	2.1	2.1	2.6	2.6	3	***	44	46.1	24.2	25.2	31.1	26.9	15.5
(7) PigSCV	18.5	19.3	16.6	16.6	16.6	3.6	***	47.1	28	31.9	37.4	34.3	21
(8) BoSCV	30.9	31.7	23.6	23.6	22.9	8.1	19.8	***	29.5	29.8	35.6	34.7	23.1
(9) PCV2	15.6	16	12.2	12.2	12.2	4.5	14.3	14.1	***	52.5	47.5	33.8	30.3
(10) Human circular virus[a]	16.5	16.9	11.4	11.4	10.7	4.5	11.8	13.7	43.9	***	53.1	37.5	31.8
(11) Chi-cyclovirus[b]	15.2	16.5	12.2	12.2	11.8	4.5	11.4	11.8	35	41.4	***	45.1	29.1
(12) BBTV[c]	22.2	23	18.1	18.1	17.7	1.8	18.6	20	14	16.3	17.4	***	27.7
(13) HCTV[d]	14.8	14.8	12.9	12.9	12.2	9.9	14.3	18	11.5	13	11.2	13.3	***

Based on the sample from this study, the highest similarity scores among circular ssDNA viruses are in boldface. Percentage identities of the aligned full nucleotides and amino acid of putative replicase protein are shown in the upper and lower triangles, respectively.

Each of [a]Human circular virus, [b]Chi-cyclovirus, [c]BBTV, [d]HCTV means Human stool-associated circular virus NG13, Chimpanzee cyclovirus, Banana bunchy top virus in Nanoviridae, and Horseradish curly top virus in Geminiviridae, repsectibly.

†PoSCV Kor was detected in this study.

‡GT306, ‡GM488, ‡GM495, ‡GM510 mean Chimpanzee stool-associated circular ssDNA viruses detected by Blinkove et al,..

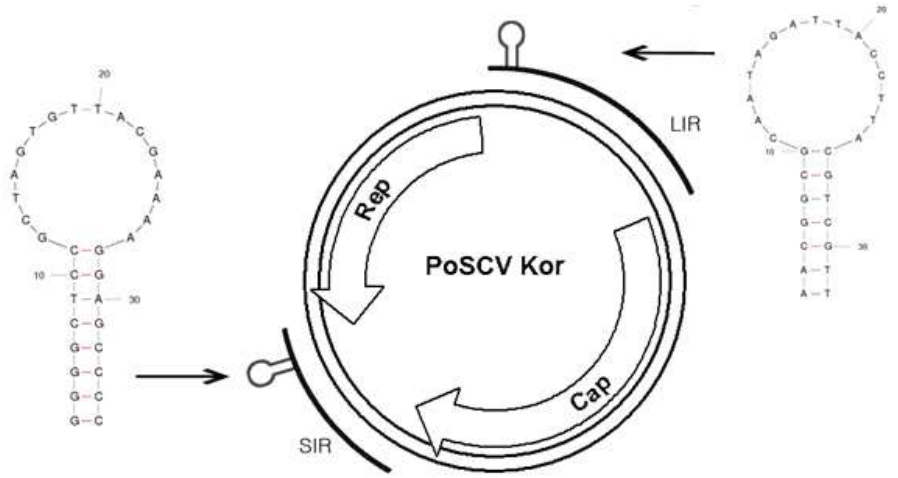


Figure 4. Genomic structure of PoSCV Kor. Two major, inversely originated ORFs encoding putative capsid protein (Cap) and replicase-associated protein (Rep) are indicated on the genome of PoSCV Kor using pDRAW31 program. Two stem-loops are located in 5' and 3' downstream regions, respectively, showing the same nonamer sequence as PoSCV

Estimates of evolutionary divergence between partial replicase sequences

Three sequences (J284, J322, and J490) which were likely to encode partial replicase proteins and PoSCV Kor were computed by using *p*-distance model of pairwise distances in MEGA 5. Among these sequences, there was no considerable difference in evolutionary divergence ratios ranging from 0 to 0.03, meaning that calculated low figures were to be the evolutionary close relation. In addition, there were also tight relations between PoSCV and each sample sequence, suggesting that both samples and PoSCV should be the evolutionary same virus (ranging from 0.09 to 0.12). Compared with ChiSCVs, the divergence figures ranged from 0.38 to 0.41, showing moderate divergence. On the contrary, it might be showed that there were some evolutionary differences between replicase-encoding sequences of samples and other circular ssDNA viruses from NCBI.

Nucleotide and amino acid sequence identity in the partial replicase region

Before making use of Megalign in DNAstar program, the sequences of samples and other circular ssDNA viruses, which might include replicase-encoding region, needed to be aligned by CLUSTAL W in BioEdit. Among samples (J284, J322, and J490) and PoSCV Kor, 93.2, 92.3, and 97.7% of nucleotide identity were observed. Samples and PoSCV had rates ranged from 85.7 to 92.1% of identity. In addition, samples and ChiSCVs showed 42-57%. In respect of amino acid identity, 87.1-95% of identity was shown among samples and PoSCV Kor, especially disclosing the highest similarity between PoSCV Kor and J490. Actually, this identity rate between PoSCV Kor and J490 could be prospected by examining the result of phylogenetic tree. Each Samples and PoSCV showed 70-78.6% of regular identity rates.

DISCUSSION

The discovery of a novel full-length genome of circular single-stranded DNA virus from swine feces in Korea has been reported. It would be named as PoSCV Kor from now on. This sequence had a considerable similarity with PoSCV in respect to genome size and genetic organization, consisting of 2,589bp in length and containing two major opposite ORFs. In addition, PoSCV Kor had two stem-loop structures in the ORF 5' and 3' downstream regions, respectively, which have been known to function as a binding site for the replicase protein in *Circoviruses* (Steinfeldt *et al*., 2001). Phylogenetic analysis based on full-length sequence strongly supported that PoSCV Kor and PoSCV should be the same viruses which are included into one type of novel circular ssDNA viruses. Moreover, this suggestion would be reinforced through analyzing the genetic identity among replicase-encoding sequences of reference viruses and samples from this study.

The prevalence level of PoSCV Kor from swine feces in Korean commercial farms, which was obtained from performing a nested PCR with PoSCV-specific primer sets, might be moderate, showing 45% of total positive rate. This figure might not be low comparing with the detection rates from bats (6.5-62.1%), bovine (65%), and chimpanzees (18%) (Blinkova *et al*., 2010; Ge *et al*., 2011; Kim *et al*., 2012). Besides, average positive rate was observed, irrespective of ages and provinces. Therefore, this virus might be prevalent in Korean swine farms and have the potency for endemic infection. However, there was no related evidence between the detection rate of PoSCV Kor and seasons. It might be just suspected that the environment would be an influential cause for the significant divergence among seasons. Besides, this study was not analyzed in terms of the relation between symptoms of individual swine and the detection of this viral genome. It was not certain whether this virus not only could induce any gastroentero-pathologic disease but also cause infection.

The metagenomic method might be considered as a striking technique for detecting novel viruses not having been reported so far. Moreover, sampling is

non-invasive and a bulk of feces is thought as a treasure for various virus particles or their genomes. There are, however, some apprehensions about unexpected contamination from exposed environment or uncertainty of virus sequences origin (Ge *et al*., 2011). Nevertheless, particle-associated nucleotide PCR with some specific primers to identify unknown viruses might be able to alleviate these concerns, providing the improved sensitivity and specificity (Kim *et al*., 2012).

A lot of circular single stranded-viruses including *Circovirus* or *Cyclovirus* have been considered as pathogens that might have the possibility for cross-species infection; in particular, PCV-2 is the only agent related to mammalian disease, causing porcine circovirus-associated disease (PCVAD) (Firth *et al*., 2009; Ge *et al*., 2011) and *Cycloviruses* have been found in various farm animals, including goats, cows, sheep, camels, and chickens (Li *et al*., 2011) and detected in a wide range of hosts or countries. These facts indicate that circular viruses researched up to now and newly detected have the potential for cross-species transmission. Therefore, further studies should be performed to prove this suspicion.

In conclusion, a novel circular DNA virus (PoSCV Kor) was identified from swine stools and turned out to be an obviously complete same virus as PoSCV detected from New Zealand. This fact could be demonstrated through analyzing genome organization and phylogenetic relation. Furthermore, the detection rate of this virus in Korean swine farms could not be overlooked because of moderate positive rates. Therefore, further deep investigation is proposed to determine its natures involving the possibility of cross-species infection and the relation with enteropathology.

REFERENCES

Ilyina, T. V., Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria *Nucleic Acids Res* 20, 3279-3285.

Orozco, B. M., Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *J Biol Chem* 273, 24448-24456.

Blinkova, O., Victoria, J., Li, Y., Keele, B. F., Sanz, C., Ndjango, J. B., Peeters, M., Travis, D., Lonsdorf, E. V., Wilson, M. L., Pusey, A. E., Hahn, B. H., Delwart, E. L. (2010). Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* 91, 74-86.

Todd, D., Weston, J., Ball, N. W., Borghmans, B. J., Smyth, J. A., Gelmini, L., Lavazza, A. (2001). Nucleotide sequence-based identification of a novel circovirus of canaries. *Avian Pathol* 30, 321-325.

Segales, J., Domingo, M. (2002). Postweaning multisystemic wasting syndrome (PMWS) in pigs. A review. *Vet Q* 24, 109-124.

Neumann, E. J., Dobbinson, S. S., Welch, E. B., Morris, R. S. (2007). Descriptive summary of an outbreak of porcine post-weaning multisystemic wasting syndrome (PMWS) in New Zealand. *NZ Vet J* 55, 346-352.

Lorincz, M., Csagola, A., Farkas, S. L., Szekely, C., Tuboly, T. (2011). First detection and analysis of a fish circovirus. *J Gen Virol* 92, 1817-1821.

Rosario, K., Duffy, S., Breitbart, M. (2009). Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* 90, 2418-2424.

Li, L., Shan, T, Soji, O. B., Alam, M. M., Kunz, T. H., Zaidi, S. Z., Delwart, E. (2011). Possible cross-species transmission of circoviruses and cyloviruses among

farm animals. *J Gen Virol* 92, 768-772.

Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment throguh sequence weighting, position-specific gap penalies and weight matrix choice. *Nucleic Acids Res* 22, 4573-4580.

Ge, X., Li, J., Peng, C., Wu, L., Yang, X., Wu, Y., Zhang, Y., Shi, Z. (2011) Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol* 92, 2646-2653.

Dayaram, A., Potter, K. A., Moline, A. B., Rosenstein, D. D., Marinov, M., Thomas, J. E., Beitbart, M., Rosario, K., Arguello-Astorga, G. R., Varsani, A. (2013). High global diversity of cycloviruses amongst dragonflies. *J Gen Virol*.

Li, L, Kapoor, A., Slikas, B., Bamidele, O. S., Wang, C., Shaukat, S., Masroor, M. A. Wilson, M. L., Ndjango, J. B., Peeters, M., Gross-Camp, N. D., Muller, M. N., Hahn, B. H., Wolfe, N. D., Triki, H., Bartkus, J., Zaidi, S. Z., Delwart, E. (2010). Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* 84,1674-1682.

Svraka, S., Rosario, K., Duizer, E., van der Avoort, H., Breitbart, M., Koopmans, M. (2010). Metagenomic sequencing for virus identification in a public-health setting. *J Gen Virol* 91, 2846-2856.

Edwards, R. A., Rohwer, F. (2005). Viral metagenomics. *Nat Rev Microbiol* 3, 504-510.

Delwart, E. L. (2007). Viral metagenomics. *Rev Med Virol* 17, 115-131.

Kim, H. K., Park, S. J., Nguyen, V. G., Song, D. S., Moon, H. J., Kang, B. K., Park, B. K. (2012). Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* 93, 635-639.

Breitbart, M., Hewson, Il, Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from

human feces. *J Bacteriol* 185, 6220-6223.

Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R. A., Felts, B., Mahaffy, J. M., Mueller, J., Nulton, J., Rayhawk, S., Rodriguez-Brito, B., Salamon, P., Rohwer, F. (2008). Viral diversity and dynamics in an infant gut. *Res Microbiol* 159, 367-373.

Sachsenroder, J., Twardziok, S., Hammerl, J. A., Janczyk, P., Wrede, P., Hertwig, S., Johne, R. (2012). Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One* 7, e34631.

Sikorski, A., Arguello-Astorga, G. R., Dayaram, A., Dobson, R. C., Varsani, A. (2013). Discovery of a novel circular single-stranded DNA virus from porcine faeces. *Arch Virol* 158, 283-289.

Zhang, T., Breitbart, M., Lee, W. H., Run, J. Q., Wei, C. L., Soh, S. W., Hibberd, M. L., Liu, E. T., Rohwer, F., Ruan, Y. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol* 4, e3.

Nakamura, S., Yang, C. S., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., Goto, N., Takahashi, K., Yasunaga, T., Ikuta, K., Mizutani, T., Okamoto, Y., Tagami, M., Morita, R., Maeda, N., Kawai, J., Hayashizaki, Y., Nagai, Y., Horii, T., Iida, T., Nakaya, T. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4, e4219.

Anisimova, M., Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55, 539-552.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011) MEGA5: molecular evolutionary genetic analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol Biol Evol* 28, 2731-2739.

Hall, B. G. (2013). Building Phylogenetic Trees from Molecular Data with MEGA.

Mol Biol Evol 30, 1229-1235.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.

Steinfeldt, T., Finsterbusch, T., Mankertz, A. (2001). Rep and Rep' protein of porcine circovirus type 1 bind to the origin of replicationin vitro. *Virology* 291, 152-160.

Firth, C., Charleston, M. A., Duffy, S., Shapiro, B., Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *J Virol* 83, 12813-12821.

국문초록

국내 돼지 분변에서의 원형 단일 가닥 DNA 바이러스의 게놈 분석

김 아 름

수의미생물학전공

서울대학교 대학원 수의학과

돼지 원형 단일가닥의 DNA 바이러스가 한국에서 돼지 분변으로부터 발견되었다. 실제로, 이 바이러스는 이전에 돼지 분변으로부터 발견된 소 분변 관련 원형 DNA virus (BoSCV-like virus)로 언급이 되었다. 하지만 이 바이러스에 대한 정확한 특성은 밝혀지지 않았다. 그러므로 본 연구는 앞서 제시된 바이러스의 전체 게놈 서열을 밝히고 이의 유전적 특성을 분석하는데 집중하였다. 이 바이러스는, 한국의 돼지 분변 관련 원형 DNA 바이러스 (PoSCV Kor)로써, 원형의 2,589 bp로 구성된 게놈을 가지고 있다. 이 바이러스는 복제효소와 캡시드 단백을 역방향으로 인코딩하고 있는 두 개의 주요 ORF를 가지고 있다. 그리고 그 잠정적인 ORF의 5'과 3' 말단에는 각각 하나씩의 줄기-루프 구조가 형성되어 있다. 이 특징은 뉴질랜드에서 발견된 돼지 분변 관련 원형 바이러스 (PoSCV)와 같았으나, 하나의 줄기-루프 구조를 가지는 침팬지 분변 관련 원형 바이러스(ChiSCVs)와 소 분변 관련 원형 바이러스(BoSCV)와는 다르다. 그러므로 PoSCV Kor이 PoSCV와 뉴클레오타이드 길이가 같다는 점과 두 바이러스의 복제효소와 캡시드 단백의 높은 아미노산 상동성으

로 보았을 때 (각각 96과 93%의 상동성), PoSCV Kor은 유전적 측면에서 PoSCV와 굉장히 유사함을 알 수 있다. 이 사실은 또한 다른 원형 단일가닥 DNA 바이러스들의 전체 게놈들과의 계통 분석을 통해, PoSCV Kor과 PoSCV가 하나의 클러스터에 분류되는 것을 통해 증명될 수 있다.

주요어: PoSCV, 한국, 원형 DNA 바이러스, 유전적 분석, 전체 게놈
학 번: 2013-21545