Whole-Genome Sequencing Analysis of Sapovirus Detected in South Korea

Hye Lim Choi¹, Chang-II Suh², Seung-Won Park³, Ji-Young Jin¹, Han-Gil Cho⁴, Soon-Young Paik¹*

¹ Department of Microbiology, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul, 137–701, Republic of Korea, ² Department of Medical Consilience, 152, Dankook University, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 448–701, Republic of Korea, ³ Division of Biotechnology, Catholic University of Daegu, Daegu, 712–702, Republic of Korea, ⁴ Division of Public Health Research, Gyeonggi Province Institute of Health and Environment, Suwon, Republic of Korea

* paik@catholic.ac.kr

Abstract

Sapovirus (SaV), a virus residing in the intestines, is one of the important causes of gastroenteritis in human beings. Human SaV genomes are classified into various genogroups and genotypes. Whole-genome analysis and phylogenetic analysis of ROK62, the SaV isolated in South Korea, were carried out. The ROK62 genome of 7429 nucleotides contains 3 open-reading frames (ORF). The genotype of ROK62 is SaV GI-1, and 94% of its nucleotide sequence is identical with other SaVs, namely Manchester and Mc114. Recently, SaV infection has been on the rise throughout the world, particularly in countries neighboring South Korea; however, very few academic studies have been done nationally. As the first whole-genome sequence analysis of SaV in South Korea, this research will help provide reference for the detection of recombination, tracking of epidemic spread, and development of diagnosis methods for SaV.

Introduction

Sapovirus (SaV) is the one of the etiological agents of human gastroenteritis and is named after the Japanese city Sapporo, where it was first discovered [1]. It is an important cause of gastroenteritis in young children and adults, and can induce symptoms such as diarrhea, vomiting, and fever [2, 3]. Its transmission routes are person-to-person (fecal–oral), through aerosols, or through contaminated water or foods [4].

SaV is an RNA virus with a non-segmented, positive-sense, single-stranded RNA molecule of approximately 7.3–7.5 kb. It belongs to the family Caliciviridae, which also includes norovirus [5, 6]. Phylogenetic analysis based on capsid protein (VP1) nucleotide sequences can divide this genus into 5 genogroups (GI–GV). Further analysis of 4 human SaV genogroups has led to their subdivision into 16 genotypes (GI.1–GI.7, GII.1–GII.7, GIV, and GV) [5–7]. Genogroups GI, GII, GIV, and GV can cause severe infection in humans, while GIII infects pigs [8]. GII and GIII genogroups have 2 ORFs, and the others have 3 each [9–11]. ORF1 encodes nonstructural proteins and the capsid protein VP1, but the roles of ORF2- and ORF3-encoded proteins have
not been clearly defined [6, 12]. For human SaV strains which were not cultivable through cell culture, molecular studies including characterization of the infectious cycle of the virus were limited. The detection system for SaV with reverse transcription-polymerase chain reaction (RT-PCR) analysis needs to be highly sensitive and accurate [13–15]. The purpose of this study was to analyze and present, for the first time, the full-length genome sequence of a SaV in South Korea. Phylogenetic analysis was performed for comparison with genotypes which have already been reported. We expect the data acquired from whole-genome sequencing to be useful not only for research in molecular biology, but also for basic epidemiologic analyses such as tracking of international spread.

Materials and Methods

Ethics statement

The stool sample was provided by Waterborne Virus Bank (WAVA). Due to issues concerning difficulties in tracking the exact records of the patient from the donor hospital, informed consent from the parent of the child participant could not be acquired. The Institutional Review Board reviewed and approved the use of this sample for the purpose of research as this study does not affect the patient. All of the experimental work and sample collections were supervised by the Catholic Medical Center Office of Human Research Protection Program (CMC OHRP) of South Korea (approval no. MC14SISI0096).

Sample preparation and viral RNA extraction

A SaV-positive stool sample, obtained from a female infant who presented with fever and diarrhea, was obtained from the Waterborne Virus Bank (WAVA, Seoul, South Korea). The stool sample was stored at −70°C until RNA extraction. The frozen stool sample was thawed and diluted with 10% with phosphate-buffered saline (PBS), after which it was centrifuged. Viral RNA of SaV was extracted from 140 μL of supernatant using a QIAamp Viral RNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Isolated RNA was stored at −70°C until further use.

Reverse transcription (RT) polymerase chain reaction

For the detection of SaV, RT-PCR was performed with the OneStep RT-PCR Kit (Qiagen) using SV-F11 and SV-R1 primers (Table 1). To analyze the whole-genome sequence of SaV, 10 more primer pairs were newly designed based on the Manchester strain (GenBank accession no. X86560). RT-PCR was performed with a S1000 thermal cycler (Bio-Rad, Hercules, CA, USA), and the steps comprised RT (50°C for 30 min), initial PCR activation (95°C for 15 min), 39 cycles of 3-step cycling (94°C for 30 s, 52°C–55°C for 30 s, and 72°C for 1 min), and final extension (72°C for 10 min). All RT-PCR products were examined by electrophoresis in ethidium bromide-stained 2% agarose gels.

Determination of the 5′ and 3′–ends of the SaV genomic RNA

To determine the 5′-end of the SaV genomic RNA, RACE (Rapid-Amplification of cDNA Ends) was performed with the 5′-Full RACE Core Set Kit (Takara Bio Inc., Ohtsu, Japan). The first cDNA strand was synthesized through reverse transcription from target mRNA using 5′-end-phosphorylated RT Primer (5′-SV-PR, Table 1), after which it was treated with RNase H to remove hybrid RNA and with RNA Ligase to form circularized single-strand cDNA or concatemers. To amplify the product, the first PCR reaction was performed using 5′-SV-F1 and 5′-SV-R1 primers under the following conditions: 94°C for 3 min, followed by 25 cycles each of...
94°C for 30 sec, 56°C for 30 sec and 72°C for 5 min. Then, the second PCR reaction was conducted with 5'-SV-F2 and 5'-SV-R2 primers through 30 cycles of 3-step cycling (94°C for 30 sec, 56°C for 30 sec and 72°C for 5 min).

To attain the exact sequence for the 3'-end of the SaV genomic RNA, cDNA was synthesized using RT reaction performed with 3'-end poly A tail-based 3'-Oligo (dT)-anchor primer (Table 1). The second PCR reaction was conducted using the SV-10F and 3'-anchor-R primers (Table 1) under the following conditions: 30 cycles of 3-step cycling (98°C for 10 sec, 56°C for 30 sec and 72°C for 1 min) and 72°C for 7 min.

### Cloning and sequencing of the complete genome

All PCR products obtained using 13 primer pairs were extracted from 2% agarose gels using HiYield Gel/PCR DNA Fragments Extraction Kit (RBC, Taipei, Taiwan) and were cloned into pGEM-T easy vectors (Promega, Madison, WI, USA). Transformed *Escherichia coli* DH5α-

| Primer | Sequence (5'→3') | Location | Size (bp) | Polarity |
|--------|-----------------|----------|-----------|----------|
| KSV-F1 | GTG ATT GGT TRG ATG GYT TCC | 1–21 | 847 | + |
| KSV-R1 | CYA CDA GTG CTG TCA TGG TG | 828–847 | - |
| KSV-F2 | ACC ATC ATC TCT CAA CAA CAC AA | 784–811 | 878 | + |
| KSV-R2 | TCA CAA TTD ART GGG AAS GGT G | 1640–1661 | - |
| KSV-F3 | TGC ATA TGG GAT GAR TTT GAT GTS | 1567–1590 | 797 | + |
| KSV-R3 | GTG TGS TKG TGG WAA ACA TGG A | 2342–2363 | - |
| KSV-F4 | CMA TGT TYA CCA CCA SYA CAC | 2267–2287 | 842 | + |
| KSV-R4 | ATC YCG CAC RCC ACC AGC A | 3090–3108 | - |
| KSV-F5 | ACA TGA SWG TCA ATG ACT TCC TMA | 2945–2968 | 706 | + |
| KSV-R5 | GSA RRC CCT TCC AKT GAA ATT CA | 3628–3650 | - |
| KSV-F6 | TGA ATT TCA MTG GAA GGG YTT SC | 3633–3655 | 748 | + |
| KSV-R6 | TTT TGT GGA RRC TCT GGY TTG DG | 4359–4380 | - |
| KSV-F7 | YTC HAA ARG GGA YTC CAC ACA A | 4359–4380 | 821 | + |
| KSV-R7 | TGC CCT CCA TCT CAA ACA CTA | 5159–5179 | - |
| KSV-F8 | ATG GAM AAT GGG GTK TCA CCM G | 5857–5878 | 632 | + |
| KSV-R8 | AGC CAG TGT GGC TGT GA | 6473–6488 | - |
| KSV-F9 | GAC TTT GAC ACY AGT GYT TTT GC | 6379–6401 | 670 | + |
| KSV-R9 | CCA TTR TGA ATG GAG AGG TCG CY | 7029–7048 | - |
| KSV-F10 | ATA GTR TAC CAG CAG CGY CAG | 6921–6941 | 511 | + |
| KSV-R10 | GSR GGR ACG GYG ACA ATC | 7414–7431 | - |
| *SV-F11 | GGY TTG TTY ATA GGT GGT AC | 5098–5117 | 781 | + |
| *SV-R1 | CWG GTG AMA CMC CAT TKT CCA T | 5857–5878 | - |
| 5'-SV-PR | CAG ATT TGA ATA CAC CTC CAC TA | 842–864 | - |
| 5'-SV-F1 | CCA TTA CAG AAC CCA TTA | 659–676 | 483 | + |
| 5'-SV-R1 | GAA ACA ATG CAC CTT CTT | 260–277 | - |
| 5'-SV-F2 | CAG TAC AAC AAG AAA TGG | 704–721 | 385 | + |
| 5'-SV-R2 | AGG AGA CCC TCC TCG ACG AA | 205–224 | - |
| 3'-Oligo (dT)-anchor | CAA TGA GGT TAT GGC TTT GGA ACT TTTTTT TTT TTT TT | - |
| 3'-Anchor-R | CAA TGA GGT TAT GGC TTT GGA AC | - |

* Molecular epidemiology and phylogenetic analysis of Sapporo-like viruses. The primers were based on the Manchester strain (GenBank accession no. X86560)[13].

doi:10.1371/journal.pone.0132328.t001
competent cells (RBC) were selected from Luria-Bertani (LB) agar plate (Duchefa, Haarlem, Netherlands) containing 40 mg/mL X-gal, 0.1 mM isopropyl-β-D-thio-galactoside, and 50 mg/mL ampicillin at 37°C for 16–18 h. Selected clones were inoculated in LB broth and incubated overnight in a shaking incubator (37°C, 200 rpm, IS-971R, Jeiotech, Daejeon, South Korea). Plasmid DNA was purified using the HiYield Plasmid Mini Kit (RBC) and sequenced (Cosmo Genetech, Seoul, South Korea). The sequencing results were analyzed using NCBI’s BLAST.

Phylogenetic analysis
Comparative sequence analysis, including sequence alignments and estimation of genetic distances, was performed with Clustal W using the Molecular Evolutionary Genetic Analysis software (MEGA soft version 6.0) [16]. Phylogenetic trees were constructed using the neighbor-joining method in MEGA 6 [17].

Results
The SaV RNA was extracted from a stool sample collected and provided by the Waterborne Virus Bank (WAVA, Seoul, South Korea). The isolated SaV strain, designated as ROK62, had a total length of 7429 nucleotides (nt). The complete genome sequence of ROK62 was deposited in GenBank under accession no. KP298674. Its 5'-UTR was 12 nt long, and 3'-UTR was 81 nt long. Its total length was found to be the same as that of the Mc114 virus and was 2 nt shorter than that of the Manchester virus. The location and length of ORFs and VP regions were found to be as follows: ORF1, 13–6855 (6843 nt); ORF2, 6852–7349 (498 nt); ORF3, 5180–5665 (486 nt); and VP1, 5170–6852 (1683 nt).

In the phylogenetic analysis, ROK62 sequences were aligned and compared with other reported SaV sequences. In the phylogenetic tree, ROK62 was classified under the GI genogroup, closely resembling the Manchester virus and the Mc114 virus, which are SaV GI-1 members (Fig 1). Similarities with the Manchester strain (GenBank accession number X868560) was confirmed using Basic Local Alignment Search Tool (BLAST) analysis, which revealed an identity of 94% (highest similarity), Max scores, total scores, and query coverage values were also determined, and their values were 11271, 11271, and 100%, respectively, for ROK62 and the Manchester virus. ROK62 showed 94% identity with the Mc114. The identity was determined using whole-genome sequence BLAST (Table 2). All identity results were obtained at a query coverage rate greater than 99%.

Even though phylogenetic analysis of ORF3 showed that ROK62 is close to the Sapporo strain, sequence identities with ORF1 (99.93%), ORF2 (99.93%), ORF3 (99.99%) and VP1 (99.95%) showed the closest relationships with the Manchester strain (Fig 2).

Discussion
SaV is one of the important causal agents of acute gastroenteritis worldwide. It mostly infects children but can also infect adults [18], and can occur in during any season [19, 20].

Although the occurrence rate of SaV in South Korea reported in 2012 was not high (0.1%) [21], it has been increasing globally. For example, the rate of SaV-positive gastroenteritis outbreaks was reported to be as high as 8% according to studies in 2000–2012 in Japan [22]. Moreover, there have been steady occurrences of SaV infections in other Asian countries, including China, Thailand, Taiwan, and Hong Kong. SaV infections have also been reported in European countries such as Germany, Sweden, and the Netherlands, where the rates of SaV-positive gastroenteritis outbreaks were in the range of 1.3–4% [23].

This is the first study to determine the whole genome sequence of SaV from a patient with acute gastroenteritis in South Korea. The SaV strain, ROK62, which was detected in South
Fig 1. Phylogenetic tree of sapoviruses based on whole-genome sequences. The numbers associated with each branch indicate the bootstrap values for the genotype. The neighbor-joining method in MEGA was used to construct the trees. Statistical significance for the grouping was obtained when bootstrap values were greater than 95. The scale shows nucleotide substitution per site. The genogroup and genotype of each strain was indicated with strain name, inside ( ). The GenBank accession numbers of the reference strains are as follows: Bristol, AJ249939; BD/697, GQ261222; C12, AY603425; Cowden, AF182760; Dresden, AY694184; Ehime475, DQ366344; Ehime643, DQ366345; Ehime1107, DQ058829; Ehime1596, DQ366346; Manchester, X86560; Mc114, AY237422; Mc2, AY237419; Mc10, NC010624; NK24, AY646856; SK15, AY646855; SW278, DQ125333; SW314, DQ125334; Sapporo, HM002617.

doi:10.1371/journal.pone.0132328.g001

Table 2. SaV, ROK62, BLAST results.

| Description                                                                 | Max score | Total score | Query cover | E value | Identity | Accession number |
|----------------------------------------------------------------------------|-----------|-------------|-------------|---------|----------|-----------------|
| Sapporo virus-Manchester polyprotein and hypothetical proteins, genomic RNA | 11271     | 112713      | 100%        | 0.0     | 94%      | X86560          |
| Sapovirus Mc114, complete genome                                           | 11099     | 11099       | 99%         | 0.0     | 94%      | AY237422        |
| Sapovirus N21, complete genome                                            | 11021     | 11021       | 100%        | 0.0     | 93%      | AY237423        |
| Sapovirus Hu/GI/Sapporo/MT-2010/1982, complete genome                     | 9806      | 9806        | 100%        | 0.0     | 90%      | HM002617        |
| Sapovirus Hu/Dresden/pJG-Sap01/DE, complete genome                        | 9607      | 9607        | 99%         | 0.0     | 90%      | AY694184        |
| Sapovirus NongKhai-50/Thailand, complete genome                           | 9581      | 9581        | 100%        | 0.0     | 90%      | AY646853        |
| Sapovirus Chanthaburi-74/Thailand, complete genome                        | 9437      | 9437        | 100%        | 0.0     | 90%      | AY646854        |

doi:10.1371/journal.pone.0132328.t002
Korea, belongs to GI-1 and showed no intra- or inter-genogroup recombination of the non-structural protein-encoding region and the VP1-encoding region. ROK62 is very similar to the Sapporo (Hu/GI/Sapporo/MT-2010/1982, HM002617) strain, the first prototype of which was reported from an outbreak in Sapporo, Japan, in 1982 [24–27]. Phylogenetic analysis showed that the strain which shows the most resemblance is the Manchester strain (Sapovirus Manchester/UK, X86560), which was detected in the United Kingdom in 1993 and was the first SaV to have its complete genome sequenced [28, 29]. The genomic organization of ROK62, including the location and length of ORFs, VP1, and VP2, was the same as that of the Manchester strain. The Mc114 (Sapovirus Mc114/JPN, AY237422) and N21 (Sapovirus N21/
THA, AY237423) strains were also very similar. Periodic monitoring of SaV is needed to keep track of the dynamic changes of genogroups and genotypes, as predominant genogroups and genotypes of vast diversity have been reported in the same geographical area [30–32].

Phylogenetic analysis of the currently circulating SaVs is necessary in order to remain updated regarding the rapid evolution of SaV strains. Around 2007, GIV.1 was the predominant SaV strain detected in Japan, Canada, the United States, and Europe, and therefore surveillance was considered important not only at the national level but also at the international level [32–36]. This underscores the importance of international cooperation in the form of information exchange among nations, in addition to national surveillance, for the prevention of epidemics. Through comparative using more data concerning whole-genome sequencing from South Korea and neighboring countries, the development of detection kits for discovering the current predominant strains and for the prediction of future predominant strains can be developed. Therefore, we surmise that this study will not only prove valuable for basic epidemiological research but also for the promotion of public health.

Acknowledgments

This study was supported by the National Research Foundation of Korea (NRF-2012R1A2A2A01045078).

Author Contributions

Conceived and designed the experiments: HLC JYJ. Performed the experiments: HLC. Analyzed the data: HLC CIS. Contributed reagents/materials/analysis tools: SWP HGC SYP. Wrote the paper: HLC CIS SYP.

References

1. Chiba S, Sakuma Y, Kogasaka R, Akihara M, Horino K, Nakao T, et al. An outbreak of gastroenteritis associated with calicivirus in an infant home. J Med Virol. 1979; 4:249–254. PMID:232140
2. Bank-Wolf BR, Konig M, Thiel HJ. Zoonotic aspects of infections with noroviruses and sapoviruses. Vet Microbiol. 2010; 140:204–212. doi:10.1016/j.vetmic.2009.08.021 PMID: 19773133
3. Hansman GS, Katayama K, Maneekarn N, Peerakome S, Khamrin P, Tonusin S, et al. Genetic diversity of norovirus and sapovirus in hospitalized infants with sporadic cases of acute gastroenteritis in Chiang Mai, Thailand. J Clin Microbiol. 2004; 42:1305–1307. PMID: 15004104
4. Koopmans M, Vinje J, Duizer E, de Wilt M, van Duijnhoven Y. Molecular epidemiology of human enteric caliciviruses in The Netherlands. Novartis Found Symp. 2001; 238:197–214; discussion 214–198. PMID: 11444027
5. Hansman GS, Oka T, Katayama K, Takeda N. Human sapoviruses: genetic diversity, recombination, and classification. Rev Med Virol. 2007; 17:133–141. doi:10.1002/rmv.533 PMID: 17340567
6. Clarke IN, Lambden PR. Organization and expression of calicivirus genes. J Infect Dis. 2000; 181 Suppl 2:S309–316. doi: 10.1086/315575 PMID: 10804143
7. Oka T, Mori K, Iritani N, Harada S, Ueki Y, Iizuka S, et al. Human sapovirus classification based on complete capsid nucleotide sequences. Arch Virol. 2012; 157:349–352. doi: 10.1007/s00705-011-1161-2 PMID: 22075918
8. Farkas T, Zhong WM, Jing Y, Huang PW, Espinosa SM, Martinez N, et al. Genetic diversity among sapoviruses. Arch Virol. 2004; 149:1309–1323. doi: 10.1007/s00705-004-0296-9 PMID: 15221533
9. Robinson S, Clarke IN, Vipond IB, Caul EO, Lambden PR. Epidemiology of human Sapporo-like caliciviruses in the South West of England: molecular characterisation of a genetically distinct isolate. J Med Virol. 2002; 67:282–288. doi: 10.1002/jmv.2219 PMID: 11992691
10. 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; 73:9625–9631. PMID: 10516074
11. Jiang X, Cubitt WD, Berke T, Zhong W, Dai X, Nakata S, et al. Sapporo-like human caliciviruses are genetically and antigenically diverse. Arch Virol. 1997; 142:1819–1827. PMID: 9672639
Hansman GS, Takeda N, Oka T, Oseto M, Hedlund KO, Katayama K. Intergenogroup recombination in sapoviruses. Emerg Infect Dis. 2005; 11:1916–1920. doi: 10.3201/eid1112.050722 PMID: 16485479

Okada M, Shinozaki K, Ogawa T, Kaiho I. Molecular epidemiology and phylogenetic analysis of Sapporo-like viruses. Arch Virol. 2002; 147:1445–1451. doi: 10.1007/s00705-002-0821-7 PMID: 12111418

Honma S, Nakata S, Sakai Y, Tatsumi M, Numata-Kinoshita K, Chiba S. Sensitive detection and differentiation of Sapporo virus, a member of the family Caliciviridae, by standard and booster nested polymerase chain reaction. J Med Virol. 2001; 65:413–417. PMID: 11536253

Vinje J, Deijl H, van der Heide R, Lewis D, Hedlund KO, Svensson L, et al. Molecular detection and epidemiology of Sapporo-like viruses. J Clin Microbiol. 2000; 38:530–536. PMID: 10655340

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30:2725–2729. doi: 10.1093/molbev/mst197 PMID: 24132122

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4:406–425. PMID: 3447015

Johansson PJ, Bergentoft K, Larsson PA, Magnusson G, Thorhagen M, et al. A nosocomial sapovirus-associated outbreak of gastroenteritis in adults. Scand J Infect Dis. 2005; 37:200–204. doi: 10.1080/00365540410020974 PMID: 15849053

Harada S, Tokuoka E, Kiyota N, Katayama K, Oka T. Phylogenetic analysis of the nonstructural and structural protein encoding region sequences, indicating successive appearance of genomically diverse sapovirus strains from gastroenteritis patients. Jpn J Infect Dis. 2013; 66:454–457. PMID: 24047751

Svraka S, Vennema H, van der Veer B, Hedlund KO, Thorhagen M, Siebenga J, et al. Epidemiology and genotype analysis of emerging sapovirus-associated infections across Europe. J Clin Microbiol. 2010; 48:2191–2198. doi: 10.1128/jcm.02427-09 PMID: 20392905
34. Chanit W, Thongprachum A, Khamrin P, Okitsu S, Mizuguchi M, Ushijima H. Intergenogroup recombinant sapovirus in Japan, 2007–2008. Emerg Infect Dis. 2009; 15:1084–1087. doi: 10.3201/eid1507.090153 PMID: 19624925

35. Pang XL, Lee BE, Tyrrell GJ, Preiksaitis JK. Epidemiology and genotype analysis of sapovirus associated with gastroenteritis outbreaks in Alberta, Canada: 2004–2007. J Infect Dis. 2009; 199:547–551. doi: 10.1086/596210 PMID: 19099483

36. Lee LE, Cebelinski EA, Fuller C, Keene WE, Smith K, Vinje J, et al. Sapovirus outbreaks in long-term care facilities, Oregon and Minnesota, USA, 2002–2009. Emerg Infect Dis. 2012; 18:873–876. doi: 10.3201/eid1805.111843 PMID: 22516204