Journal of Microbiology (2014) Vol. 52, No. 5, pp. 427–434
DOI 10.1007/s12275-014-3538-x

Genetic Analysis of the Capsid Region of Norovirus GII.4 Variants Isolated in South Korea

Ju-Eun Kim[1], Sung-Geun Lee[2], Han-Gil Cho[3], Sang-Ha Han[1], Lae-Hyung Kang[1], Youn-Mi Lee[4], Chul-Jong Park[4], and Soon-Young Paik[1]*

[1]Department of Microbiology, College of Medicine, the Catholic University of Korea, Seoul 137-701, Republic of Korea
[2]Korea Zoonosis Research Institute, Chonbuk National University, Jeonju 561-756, Republic of Korea
[3]Division of Virology, Gyeonggi Provincial Research Institute of Public Health and Environment, Suwon 440-290, Republic of Korea
[4]Department of Dermatology, College of Medicine, the Catholic University of Korea, Seoul 137-701, Republic of Korea

(Received Oct 14, 2013 / Revised Nov 8, 2013 / Accepted Nov 11, 2013)

Norovirus is one of the major causes of non-bacterial gastroenteritis in humans. The aim of this study was to analyze the amino acid variation of open reading frame 2 of GII.4 variants in South Korea during the period from November 2006 to December 2012. Sixty-nine complete nucleotide sequences of open reading frame 2 were obtained from 113 GII.4 strains. The GII.4 2006b variants were detected predominantly between 2006 and 2009; however, new GII.4 variants, which were termed the 2010 variant and the 2012 variant, emerged in 2010 and 2012, respectively. The number of GII.4 2006b variants steadily decreased until 2012, whereas the number of gastroenteritis cases caused by the new variants increased between 2010 and 2012. The amino acid sequence in the ORF2 region obtained in this study was compared with other GII.4 variants isolated in various countries. Amino acid variations were observed primarily at epitope sites and the surrounding regions. Amino acids 294, 359, 393, and 413 of the P2 subdomain were the most variable sites among the GII.4 variants. The information in this study can be useful in basic research to predict the emergence and determine the genetic functions of new GII.4 variants.

Keywords: norovirus, GII.4 variant, sequence analysis

Introduction

Norovirus (NoV) is one of the important viruses that causes non-bacterial acute gastroenteritis in humans worldwide (Park *et al.*, 2011). NoV can infect humans through contaminated food and water, direct person to person contact, and contaminated aerosols (Mathijs *et al.*, 2011). Clinically, patients infected with NoV commonly present with diarrhea, vomiting, abdominal pain, and mild fever (Vega *et al.*, 2011; Park *et al.*, 2012).

NoV is a non-enveloped virus that belongs to the family *Caliciviridae* (Ando *et al.*, 2000; Mathijs *et al.*, 2011). Its viral genome is composed of a single-stranded, positive-sense RNA of 7.0–7.5 kb (Donaldson *et al.*, 2010). The viral RNA genome is divided into three open reading frames (ORFs): ORF1, ORF2, and ORF3. ORF1 encodes nonstructural proteins, including NTPase, 3C-like protease, and the highly conserved RNA-dependent RNA polymerase (RdRp) (Siebenga *et al.*, 2007; Bill *et al.*, 2010). ORF2 encodes a major viral capsid protein (VP1), which consists of a shell (S) domain and two protruding (P) domains, which includes the P1 and P2 subdomains (Ando *et al.*, 2000). The P1 subdomain is a relatively conserved sequence and reinforces the stability of virus particles (Bill *et al.*, 2010; Bull and White, 2011), and the P2 subdomain, which contains host-cell binding receptors such as histo-blood group antigens (HBGA) is the most variable region (Tan and Jiang, 2005; Shanker *et al.*, 2011). ORF3 encodes a minor capsid protein (VP2) (Siebenga *et al.*, 2007).

NoV is divided into five distinct genogroups (GI, GII, GIII, GIV, and GV) on the basis of capsid protein sequence analysis, and the genogroups are further subdivided into different genotypes (Zheng *et al.*, 2006, 2010). The GI and GII genogroups are known to have caused the major outbreaks of NoV gastroenteritis in humans (Park *et al.*, 2012). Among the outbreaks of NoV gastroenteritis, GII.4 was the most frequently detected genotype (Zakikhany *et al.*, 2012), and this genotype has been identified in several subgenotypes (Han *et al.*, 2011).

Over the past 20 years, the NoV GII.4 genotype has evolved through genetic variation (Lam *et al.*, 2012). Additionally, since 2002, new epidemic variants have emerged every 2–3 years (Han *et al.*, 2011; Beek *et al.*, 2013). On the basis of phylogenetic analyses, several major GII.4 variants have been discovered: pre-1996 (Camberwell), 1996 (Grimsby), 2002 (Farmington Hill), 2004 (Hunter), 2006a (Yerseke), 2006b (DenHaag89), and 2008 (Apeldoorn) (Siebenga *et al.*, 2007; Bull and White, 2011; Shanker *et al.*, 2011; Lam *et al.*, 2012).

Several studies have reported a variety of NoV GII.4 variants worldwide. Zheng *et al.* (2010) identified the 1996, 2002, and 2006a and 2006b variants in the United States (US) in 1996, 2002, and 2006, respectively, and Vaga *et al.* (2011) identified the 2010 variant in the US in 2009. After the emergence of new variants, outbreaks of NoV in the US increased (Zheng *et al.*, 2010; Vega *et al.*, 2011). Another study in Canada reported the emergence of new GII.4 vari-

*For correspondence. E-mail: paik@catholic.ac.kr; Tel.: +82-2-2258-7342; Fax: +82-2-535-6473

ants 2002, 2004, and 2006a and 2006b in 2002, 2004, and 2006, respectively (Pang *et al.*, 2010). After the identification of these variants, the number of cases of NoV-induced gastroenteritis dramatically increased by 191, 201, and 313 in the July2002/July2003, July2004/July2005, and July2006/July2007 seasons, respectively (Pang *et al.*, 2010). The GII.4 variants have also been studied in South Korea. Le *et al.* (2010) found that the 2006b variant was predominant in Jeju, South Korea, in September 2007 and July 2008. In 2008, Park *et al.* (2012) identified the 2008-Korea_a and 2008-Korea_b variants in Chungnam, South Korea, and they detected the 2008-Korea_b rather than the 2008-Korea_a variant. Park *et al.* (2010) also reported that the GII.4 2006b variant and the 2008-Korea_b variant co-existed in Chungnam from 2008 to 2010. More recently, a 2010 variant (New Orleans) and a 2012 variant (Sydney) were identified (Mathijs *et al.*, 2011; Beek *et al.*, 2013; Haslng *et al.*, 2013) in the US in late 2009 (Vega *et al.*, 2011) and in Australia in March 2012, respectively (Beek *et al.*, 2013).

The emergence of new GII.4 variants was closely associated with an increase in NoV outbreaks (Haslng *et al.*, 2013), indicating that the NoV outbreaks increased because of the emergence of new variants, and the majority of the outbreaks occurred due to these new variants. Therefore, it is important to monitor the outbreak frequency of NoV and conduct emergence surveillance of new GII.4 variants.

In this study, we investigated the prevalence of GII.4 subgenotypes from stool samples obtained in Gyeonggi-do, South Korea between November 2006 and December 2012. Additionally, the isolated GII.4 strains were compared with other previously identified reference strains by nucleotide sequencing to confirm the amino acid sequence difference of the capsid region.

Materials and Methods

Collection of fecal specimens

A total of 113 clinical stool RNA specimens were obtained from patients with acute gastroenteritis in Gyeonggi, South Korea, during September 2006 to December 2012. At the Gyeonggi Institute of Health and Environment, all samples were pre-identified as the NoV GII.4 genotype and then randomly selected based on the season. The samples were then transported to the Waterborne Virus Bank (Seoul, Korea), and were tested to the Department of Microbiology, Catholic University, for analysis of the complete ORF2 region. The stool samples were stored at -70°C until further analysis.

Reference strain collection

The reference strains were selected from previous studies conducted across various countries (Zheng *et al.*, 2006; Bull and White, 2011; Vega *et al.*, 2011; Beek *et al.*, 2013). The nucleotide sequence of the ORF2 region was obtained from NCBI.

Viral RNA extraction

The fecal specimens were diluted to 10% suspensions with phosphate-buffered saline (PBS) and centrifuged at 800 × g for 15 min. Then, the viral RNA of NoV was extracted from 140 μl of the 10% diluted stool sample with the QIAamp® Viral RNA Mini Kit (Qiagen, Germany) in accordance with the manufacturer's instructions. The extracted RNA was dissolved in 60 μl of RNase-free water and stored at -70°C until analysis.

NoV GII genogroup detection

For the detection of NoV, we employed GII-F1M and GII-R1M primers (Park *et al.*, 2011) that target the ORF1-ORF2 junction region. RT-PCR was performed with the One Step RT-PCR Kit (Qiagen, Germany) and PCR System S1000™ thermal cycler (Bio-Rad, USA). The PCR reaction mixture (final volume = 25 μl) contained 1 μmol/L of each primer and 5 μl of viral RNA, in accordance with manufacturer's protocol. The PCR conditions used in this study were as follows: reverse transcription at 50°C for 30 min and PCR activation at 95°C for 15 min, followed by 40 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR products were analyzed by electrophoresis on a 2% ethidium bromide-stained agarose gel.

Complete ORF2 region amplification and cloning

GII.4 strains identified by amplification of the ORF1-ORF2 junction region were further evaluated by amplifying the complete sequence of the ORF2 region to analyze sequence differences. To obtain the complete nucleotide sequence of the ORF2 region, the ORF2-1F/ORF2-1R and ORF2-2F/ORF2-2R primer sets were used (Table 1), which were designed on the basis of the sequence of the ORF2 region. The complete sequence of the ORF2 region was amplified using the One Step RT-PCR kit (Qiagen). The PCR reaction mixture (final volume = 25 μl) consisted of 5 μl of template viral RNA and 0.8 μmol/L of the ORF2-F and ORF2-R primers, according to the manufacturer's manual. The PCR conditions

Table 1. Primers used for the amplification of the ORF1-ORF2 junction and ORF2

Primer	Sequence (5′→3′)	RT-PCR target	Reference
GII-F1M	GGG AGG GCG ATC GCA ATC T	ORF1-ORF2 junction (5058-5077)	Park *et al.* (2011)
GII-R1M	CCR CCI GCA TRI CCR TTR TAC AT	ORF1-ORF2 junction (5401-5424)	Park *et al.* (2011)
ORF2-F1	AAGAGCCAATGTTCAGATGG	ORF2 (5004-5023)[a]	This study
ORF2-R1	CTCTGAAGGTGCAGATGTT	ORF2(5928-5946)[a]	This study
ORF2-F2	AACATCTGCACCTTCAGAG	ORF2(5928-5946)[a]	This study
ORF2-R2	GAAGCCTGTTGTAGATTGCT	ORF2(6854-6873)[a]	This study

[a] Location of the Farmington Hills strain

used in this study were as follows: reverse transcription at 50°C for 30 min and PCR activation at 95°C for 15 min; followed by 40 cycles of denaturation at 95°C for 1 min, annealing at 52°C for 1 min, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. The PCR products were analyzed using the same method as described in section 2.3.

The amplified fragments were purified in the gel by using the HiYield Gel/PCR DNA Extraction Kit (RBC, Taiwan). These products were then cloned into the pGEM-T Easy vector (Promega, USA) according to the manufacturer's recommendations, and the vector was transformed into competent *E. coli* DH5α cells (RBC). The plasmid DNA containing the sequences of the ORF2 region were purified using the HiYield Plasmid Mini Kit (RBC) according to the manufacturer's instructions. The purified plasmid DNA was sent to Cosmo Genetech (Korea) for further sequencing.

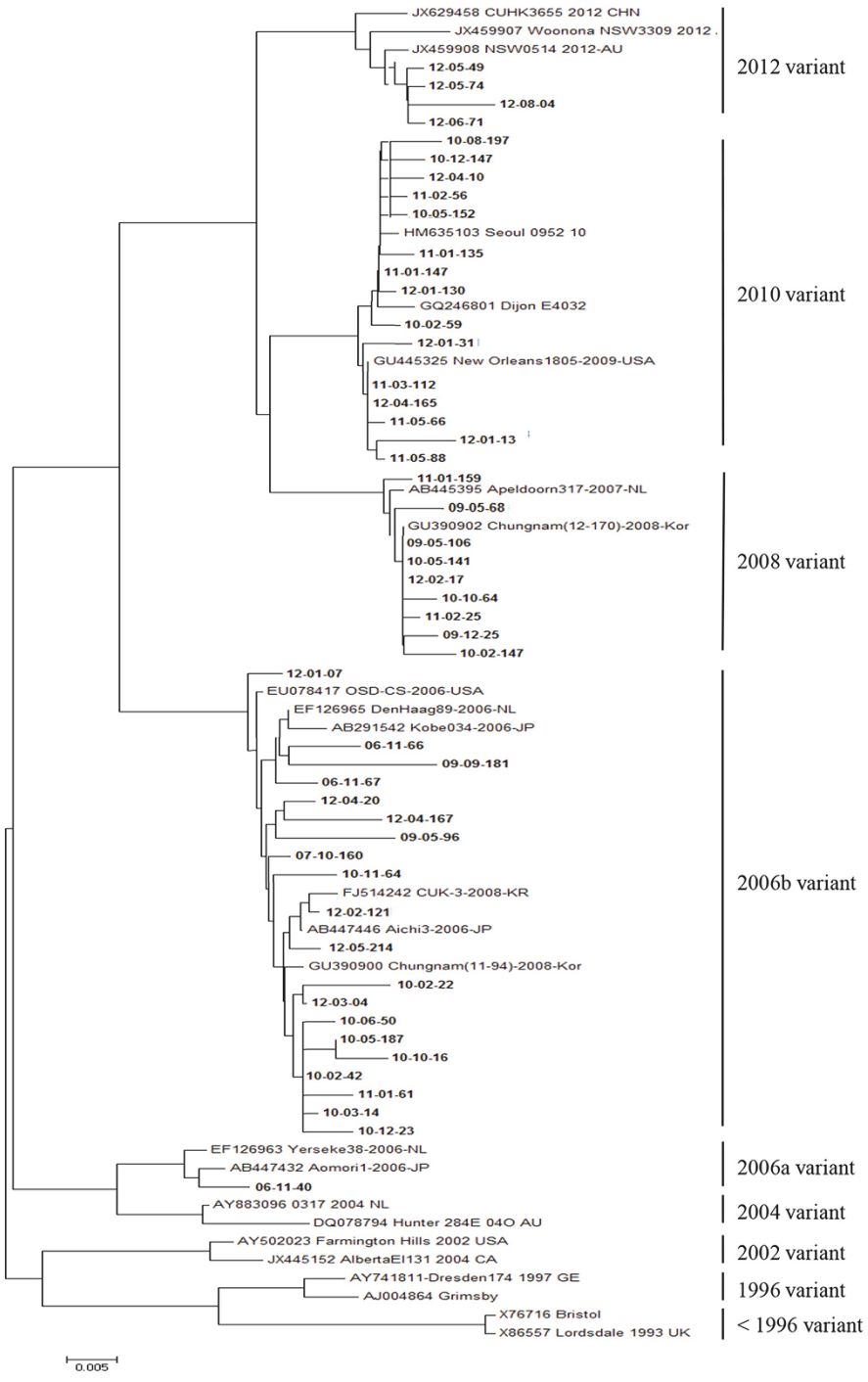

Fig. 1. Phylogenetic analysis of complete capsid protein amino acid sequences of GII.4 variants identified in Korea during 2006–2012.

Nucleotide sequencing and phylogenetic analysis

The amplified sequences from the tested samples were aligned using the Clustal W method with DNAStar software (DNAStar Inc., USA) (Park *et al.*, 2011) and MEGA software (Vega *et al.*, 2011) to characterize the isolated NoV GII genogroup and to analyze the 1,623 bp sequence changes of the complete ORF2 region. We compared GII.4 variants obtained from this study with other reference strains to confirm amino acid differences in the ORF2 region. A phylogenetic tree was constructed using the neighbor-joining method (Saitouand Nei, 1987).

Results

Identification of GII.4 variants

The 113 NoV GII.4 strains in this study were divided into five genetic variants based on the nucleotide sequence of the ORF1-ORF2 junction region and P2 domain. Two GII.4 2006a variants were identified in the 2006/2007 season. The GII.4 2006b variant was the predominant strain from the 2006/2007 season to the 2009/2010 season, and sporadic GII.4 infections and 22 strains were detected. The GII.4 2008 variant was isolated from 38 patients during the 2008/2009 and 2009/2010 seasons. However, this variant was found consistently until December 2012. The GII.4 2010 variant first emerged in January 2010; it became prevalent in the 2010/2011 season and was identified from 36 GII.4 sporadic cases, whereas the GII.4 2006b variant accounted for 20% of the sporadic cases. In 2012, the new GII.4 2012 variant first appeared in May and was identified in four strains during the 2011/2012 season. During a 2-month observation period (November and December 2012), the GII.4 2012 variant was the prevalent strain, which was isolated from 15 of the GII.4 sporadic cases.

Phylogenetic analysis of GII.4 variants

The 69 complete sequences of the ORF2 region (1,623 bp) were obtained from 113 NoV GII.4 strains. These sequences were further categorized into GII.4 variants by phylogenetic analysis (Fig. 1).

One of the 69 complete sequences, 06-11-40 strain, was the 2006a variant (Fig. 1A). The 2006a variant had branched from the 2004 variant. The identities of the 06-11-40 strain and other GII.4 variants differed within the range of 91.1–93.3% (nucleotide identity) and 93.9–94.3% (amino acid identity).

The 2006b variant was a descendant of the 2002 variant (Fig. 1B). Twenty-six of the complete sequences identified in this study belonged to the 2006b variant subcluster. Compared with 2006b strains previously isolated in Korea and Japan, the 2006b variants identified in this study were closely related at the nucleotide level (97.3–98.9%) and the amino acid level (98.7–99.6%). The 2006b variants in 2010

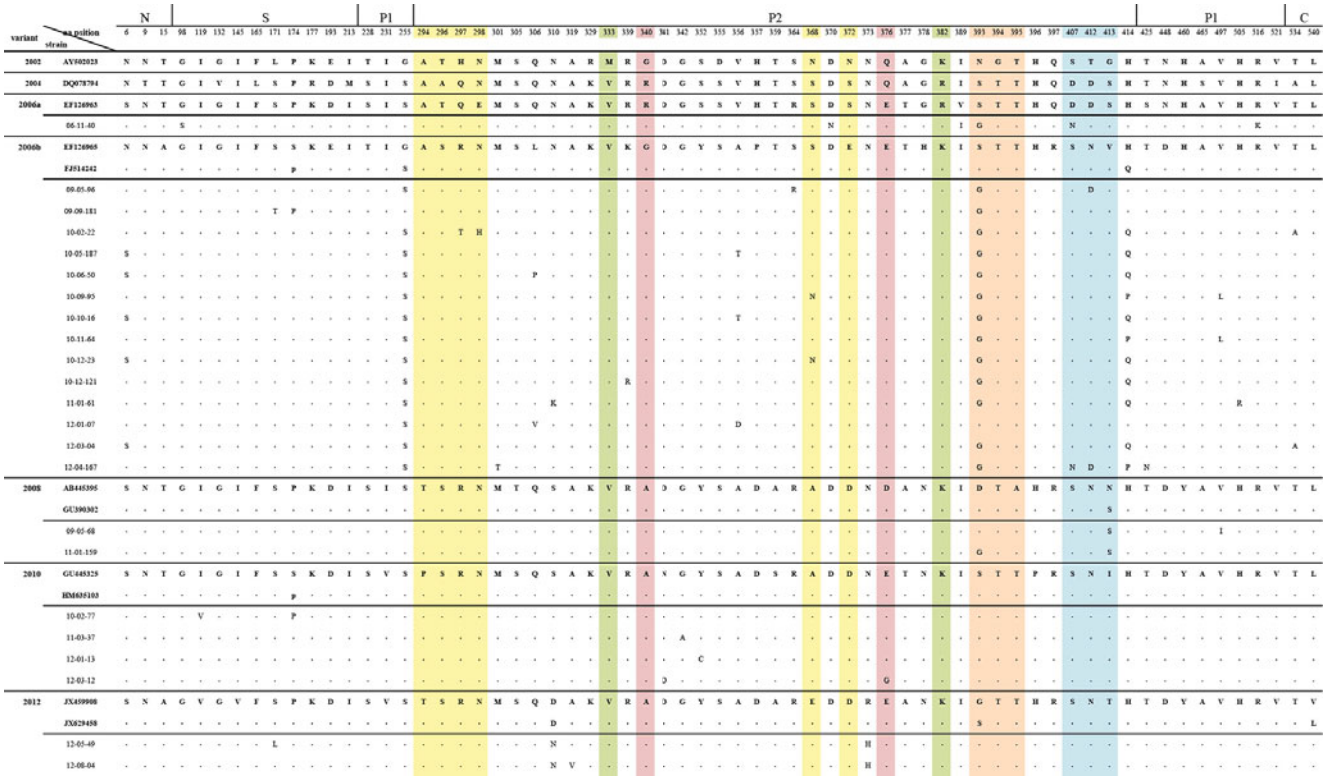

Fig. 2. Amino acid changes in capsid sequences of GII.4 sporadic strains collected between 2006 and 2012. The amino acid sequences of the variants were compared to observe evolutionary patterns. Individual epitope sites are represented in different colors: Yellow, epitope site A; green, epitope site B; red, epitope site C; orange, epitope site D; blue, epitope site E.

and 2012 identified in this study were closely related with the Chungnam (11–94) (98.9–99.6%) and CUK-3 (99.6–99.1%) strains identified in Korea, respectively. Additionally, the 2008, 2010, and 2012 variants had branched from the 2006 variant.

Eleven of the 69 strains were identified as the 2008 variant (Fig. 1C). The 11 strains and other GII.4 variants differed within the range of 91.1–96.4% (nucleotide identity) and 94.5–97.6% (amino acid identity). The 11 strains identified as the 2008 variant were closely related to the Chungnam (12–170) strain from Korea, with 98.1–99.8% identity at the nucleotide level and 99.4–100% at the amino acid level.

The 27 strains identified as the 2010 variant demonstrated <3.2% nucleotide and <2% amino acid level differences compared with the 2010 reference strains (Fig. 1D). The variants in 2010 identified in this study were closely related to the Dijon E0432 (99.86–99.8%) and Seoul0925 (99.3–99.6%) strains from France and Korea, respectively; however, the variants in 2012 identified in this study were closely related to the New Orleans strain (99–100%) isolated from the US.

Four of the 69 strains were identified as the 2012 variant. Compared with the other GII.4 variants, these four strains differed within the range 89.9–96.9% at the nucleotide level and 93.3–97.2% at the amino acid level. Additionally, the 2012 variant was related to the 2008 and 2010 variants. The four strains identified in this study were closely related to the NSW 0514 strain isolated from Australia, with 98.9–99.3% nucleotide identity and 98.7–99.4% amino acid identity (Fig. 1E).

Sequence analysis of the capsid protein

The nucleotide and amino acid sequences of the 69 ORF2 regions determined in this study were compared with the following reference strains from previous studies (Fig. 2): 2002 variant AY485642 (Farmington Hill), 2004 variant AY883096 (Hunter), 2006a variant EF126963 (Yerseke), 2006b variants EF126965 (DenHaag) and FJ514242 (CUK-3), 2008b variants AB445395 (Apeldoorn) and GU390302 (Chungnam (12-170)), 2010 variants GU445325 (New Orleans) and HM635103 (Seoul0952), and 2012 variants JX459908 (Sydney) and JX629458 (CUHK3655). The amino acid variations in the complete capsid protein, including the P2 domain, were observed.

The 2004 and 2006a variants differed at 16 amino acids: two substitutions in the N domain, four substitutions in the S domain, four substitutions in the P1 subdomain, and six substitutions in the P2 subdomain. Additionally, the amino acid sequence of the 06-11-40 strain identified in this study was comparable with the amino acid sequence of the 2006a reference strain (EF126963). The amino acid substitution (G98S) within the S domain was identified in the isolated strain. Furthermore, compared with the 2006a reference strain, the strain identified in this study had an additional five amino acid substitutions in the P domain: D370N, V389I, S393G, D407N, and R516K.

Compared with the 2002, 2004, and 2006a variants, the 2006b variant contained 14 amino acid substitutions: T15A, P174S, T296S, H267R, Q306L, S352Y, V356A, H357P, N372E, G378H, Q397R, T412 N, G413V, and N448D. The amino acid sequence of the GII.4 2006b variant was comparable to the amino acid sequence of the 2006b DenHaag strain. Two amino acids were substituted in the P domain of most of the strains identified in this study: G255S (88.4%) in the P1 subdomain and S393G (85.6%) in the P2 subdomain (data not shown). Additionally, an N6S substitution (7/26 strains) in the N domain and an N412D (5/26 strains), H414Q (15/26 strains), and H414P (5/26 strains) substitution in the P2 subdomain were also observed in the strains identified in this study (data not shown). Some strains in this study contained the following amino acid substitutions: S171T and S174P in the S domain; R297T, N298H, M301T, L306P or L306V, N310K, K339R, A356T (n = 2 strains) or A356D, S364R, S368N (n = 2 strains), and S407N in the P2 subdomain; T425V, V497L (n = 2 strains), and H505R in the P1 domain; and T534A (n = 2 strains) in the C domain.

Furthermore, three variants (2008, 2010, and 2012 variant) and the 2006b variant contained seven amino acid substitutions at common loci: T296S, H297R, S352Y, V356A, Q397R, T412N, and N448D. Compared with the previous variants, the 2008 variants differed at 14 amino acid loci. Seven amino acid substitutions were common in the 2008, 2010, and 2012 variants: N6S (N domain) and G340R, H357D, S364R, N372D, G378N, and H460Y (P domain). In addition, the 11 amino acid sequences of the GII.4 2008 variants in this study differed from the amino acid sequence of the 2008 Apeldoorn strain at one amino acid locus (N413S) (data not shown). The 09-05-68 and 11-01-159 strains each contained one amino acid substitution (V497I and D393G, respectively).

Compared with the 2006b and 2008 variants, the 2010 variant differed at five amino acid loci: I231V, D341N, T359S, H396P, and G413I. The I231V substitution was maintained in the 2012 variant. Furthermore, compared with the 2010 New Orleans strain, 14 of the 27 GII.4 2010 variants contained the S174P substitution in the S domain (data not shown). The 10-02-77, 11-03-37, and 12-01-13 strains contained one amino acid substitution (I119V, G342A, and Y352C, respectively). The 12-03-12 strain contained two substitutions in the P2 subdomain, N341D and E376G.

Compared with the 2008 and 2010 variants, the 2012 variant contained amino acid substitutions at seven loci: I119V, I145V, N310D, N373R, S393G, V413T, and L540V. Additionally, compared with the 2012 Sydney variant, the GII.4 2012 variant identified in this study contained two amino acid substitutions in the P2 domain (D310N and R373H). The 12-05-49 and 12-08-04 strains each contained one amino acid change (S171L and A319V, respectively).

Discussion

NoV is an important pathogen of gastroenteritis worldwide. Three NoV genogroups (GI, GII, and GVI) cause human disease, and the GII.4 genotype causes a large proportion of the NoV-associated acute gastroenteritis cases (Mathijs *et al.*, 2011; Park *et al.*, 2011, 2012; Vega *et al.*, 2011). The GII.4 genotype has emerged as several different epidemic variants, including the 1996, 2002, 2004, 2006a, 2006b, 2008, 2010, and 2012 variants.

In this study, the predominant GII.4 variants of NoV-as-

sociated viral gastroenteritis were explored in Gyeonggi, South Korea, between November 2006 and December 2012. Phylogenetic analysis and comparison sequence of the complete ORF2 sequence provided basic evolutionary information on the variant strains (Motomura *et al*., 2008).

Amino acid variations in the complete capsid protein, including the P2 domain, were investigated comprehensively. Additionally, the substitutions were screened at epitope sites within the P2 domain: A (294, 296–298, 368, and 372), B (333 and 382), C (340 and 376), D (393–395), and E (407, 412, and 413) (Lindesmith *et al*., 2012) (Fig. 2).

In this study, the 2006a variant was detected in two cases from 2006, but only one complete ORF2 sequence was obtained from the 06-11-40 strain. The variant was detected rarely in Asia (China, Japan, Hong Kong, India, and Malaysia) (Siebenga *et al*., 2009). The 2006a variant was a descendant of the 2004 variant (Fig. 1A) (Bull and White, 2011). The 2004 and 2006a variants were similar at nine amino acid loci in the epitope site; however, the 2006a variant contained two amino acid substitutions (N298E and Q376E). The amino acid substitutions in the epitope site of the 2004 and 2006a variants were similar. Additionally, compared with the Yerseke strain, the 06-11-40 strain was changed at two epitope sties (S393G and D407; Fig. 2). The 06-11-40 strain identified in this study was similar to the Terneuzen (EF126964) strain isolated from the Netherlands, with 99.0% and 99.3% identity at the nucleotide and amino acid sequence levels, respectively (data not shown). Additionally, compared with other 2006a variants identified in various countries, the 06-11-40 strain contained amino acid substitutions (370N, 389I, S393G, and 516K).

The GII.4 2006b variant was first identified in the US in 2006 and then detected worldwide (Siebenga *et al*., 2007; Huh *et al*., 2008; Lindesmith *et al*., 2008; Mathijs *et al*., 2011). In a previous study, the GII.4 2006b variant was predominant in Jeju, South Korea, from September 2007 to July 2008 (Le *et al*., 2010). The 2006b variant was also predominant in Gyeonggi from 2006 to 2008. The 2006b variant was prevalent until 2010 and has decreased since 2010. The 2006b variant evolved from the 2002 variant (Bull and White, 2011), and the variants identified in this study were closely related (98.9.6–99.6% amino acid identity) with the Chungnam (11–94) and CUK-3 strains identified from Korea (data not shown). The 2006b variant differs at six amino acid loci in the epitope site (T296S, Q297R, R340G, S372E, D412N, and S413V) compared with the 2002, 2004, and 2006a variant strains (Fig. 2). Because of many antigenic variations, the 2006b variant might be advantageous to avoid the pre-existing herd immunity as compared to previous GII.4 variants. Additionally, compared with the 2006b Denhaag strain, the variants identified in this study and the CUK-3 strain isolated from Korea varied at two common amino acid loci at the epitope site (G255S and H413Q). Two substitutions (S393G and H414P) in the P2 subdomain were identified in the present study, compared with previously reported 2006b variants. Furthermore, the strains in this study contained several amino acid substitutions within the ORF2 region.

The GII.4 2008 variant was isolated in 2008, and the variant was closely related to the 2006b variant (Fig. 1C) (Belliot *et al*., 2010). A previous study in Korea reported that the 2006b variant and the 2008 variant co-existed in Chungnam from 2008 to 2010 (Park *et al*., 2010, 2012). The two GII.4 variants were detected together, but the GII.4 2006b variant was predominant in Gyeonggi during the study period. The 2008 variants in this study differed from the Chungnam (12-170) strain isolated from Korea (GU390902) by <2.5% nucleotide and <1% amino acid sequence identity (data not shown). Compared with the Apeldoorn strain, the variants identified in this study and the Chungnam (12-107) strain contained a common substitution (N413S) (Fig. 2). Additionally, two of the 11 strains contained the amino acid variations V497I and D393G.

The GII.4 2010 variant was first identified in the US in October 2009 (Vega *et al*., 2011); in the present study, this variant was found in January 2010. The frequency of the 2010 variant gradually increased, whereas those of the 2006 variant decreased until 2011/2012. The 2010 variant emerged from the 2008 variant (Fig. 1D). GII.4 variants were defined as strains showing more than 5% sequence difference (Siebenga *et al*., 2007; Han *et al*., 2011). However, the sequence of the 2010 variant differed from the 2008 variant by as little as 3.7% and as much as 5.1% (data not shown). Compared with the 2008 variant, the 2010 variant contained several amino acid variations in the ORF2 region (Vega *et al*., 2011). The 2010 and 2008 variants differed at 11 amino acid loci. The variations of amino acids mostly appeared at the epitope site or surrounding the P2 subdomain (Lindesmith *et al*., 2013). In addition, the 2010 variants identified' in this study had complete ORF2 sequences with <2% amino acid difference compared with the GII.4 2010 variants (data not shown). Additionally, similar to the previously isolated Seoul 0952 strain from Korea (HM635103), the 14 strains isolated in this study contained a S174P substitution (Fig. 2).

The GII.4 2012 variant was identified in Australia in March 2012 prior to being identified in other countries (Beek *et al*., 2013). In May, we also identified the 2012 variant, and the frequency of the new variant increased from November to December 2012. Additionally, similar to the previously isolated NSW0514 strain from Australia (JX459908) and the CUHK3655 strain (JX629458) from China, the amino acid sequence of the variants in present study differed by as little as 0.6% and as much as 1.3% (data not shown). Unlike the previously reported GII.4 variants, the 2012 variants identified in this study and Australia contained the L540V amino acid substitution (Fig. 2). Additionally, the 2012 variant identified in this study contained two amino acid substitutions in the P2 subdomain (D310N and R373H), compared with the other 2012 variants. Additionally, two strains contained amino acid substitutions (S171L and A319V).

Amino acid variations in the complete capsid protein, including the P2 domain, were observed. Specifically, the P2 domain was the most variable region in capsid (33 amino acid positions). The N domain and S domain of the ORF2 region displayed high amino acid sequence similarity; however, the P domain of the ORF2 region demonstrated relatively lower similarity, while the P2 subdomain exhibited extremely low similarity (Jeong *et al*., 2009). Among those substitutions, amino acids 294, 359, 393, and 413 in the P2 subdomain were the most variable sites among the GII.4 variants. The amino acid variations of epitope site were predicted to form antibody blockade epitopes by other groups

(Lindesmith *et al*., 2013). Using mutant VLPs designed to vary predicted antigenic epitopes, Lindesmith *et al.* (2013) identified position 294, 296-298, 368, and 372 as important antigenic sites and they confirmed the immunodominance of epitope A since alterations of epitope A. This data explained that the GII.4.2009 New Orleans has evolved a blockade epitope site, possibly enabling for at least partial escape from herd immunity (Lindesmith *et al*., 2013).

In addition, frequent occurrence of intragenotypic recombination of NoVs has been reported in recent studies (Le *et al*., 2010; Mahar and Kirkwood, 2011; Lam *et al*., 2012; Martella *et al*., 2013). Lam *et al*. (2012) has been reported the emergence of the NoV GII.4 2008 variants by recombination with 2006a and 2006b variant. In the study by Martella *et al.* (2013) identified recombinant strain that was recombination with ORF1 of GII.4 New Orleans and complete ORF2 and ORF 3 of GII.4 Sydney. The most of the recombination events were frequently occurred in the ORF1/ORF2 junction region and ORF2 region. Also, these events were associated with emergence of a new GII.4 variant (Eden *et al*., 2013).

Besides, recombination strains of NoV, including NVGII.1/NVGII.5 and GII3/GIIb were characterized by novel RdRp clusters in India and Australia (Mahar and Kirkwood, 2011; Nataraju *et al*., 2011). Thus, a dual nomenclature using both ORF1 and VP1 sequences proposed to recognize recombinant strains by recently study (Kroneman *et al*., 2013).

New GII.4 variant could be emerged by variation of amino acid and recombination. The information obtained from 69 complete GII.4 ORF2 sequences in this study can be useful for generating more accurate diagnoses and creating a foundation for fundamental research to elucidate the genetic functions of GII.4 variants. Additionally, by comparing the strains identified in this study with GII.4 variants worldwide, our data may provide information that is necessary for the identification of emergent strains, newly occurring variants, and a basis for the development of a vaccine.

Ethics statement

The stool specimens were obtained from the business department of Enternet-Korea in the Gyeonggi Institute of Health and Environment. Because personal information for research was not acquired, ethical considerations do not apply. The acquisition of the clinical samples and all procedures to analyze the norovirus genetic sequence received IRB approval from Catholic University (Approval number: MC13SASI0086).

Acknowledgements

This study was supported by the Mid-career Researcher Program (2012R1A2A2A01045078).

References

Ando, T., Noel, J.S., and Fankhauser, R.L. 2000. Genetic classification of "Norwalk-like Virus". *J. Infect. Dis.* **181**, S336–348.

Belliot, G., Kamel, H.A., Estienney, M., Ambert-Balay, K., and Pothier, P. 2010. Evidence of emergence of new GGII.4 norovirus variants from gastroenteritis outbreak survey in France during the 2007-to-2008 and 2008-to-2009 winter season. *J. Clin. Microbiol.* **48**, 994–998.

Bill, A.R., Eden, J.S., Rawlinson, D.W., and White, A.P. 2010. Rapid evolution of pandemic noroviruses of the GII.4 linage. *PLoS Pathog.* **6**, e1000831.

Bull, A.R. and White, A.P. 2011. Mechanisms of GII.4 norovirus evolution. *Trends Microbiol.* **19**, 233–240.

Donaldson, F.E., Lindesmith, C.L., LeBue, D.A., and Baric, S.R. 2010. Viral shape-shifting: norovirus evasion of the human immune system. *Nat. Rev. Microbiol.* **8**, 231–241.

Eden, J.S., Tanaka, M.M., Boni, F.M., Rawlinson, D.W., and White, A.P. 2013. Recombination within the pandemic norovirus GII.4 lineage. *J. Virol.* **87**, 6270–6282.

Han, T.H., Kim, C.H., Chung, J.Y., Park, S.H., and Hwang, E.S. 2011. Emergence of norovirus GII-4/2008 variant and recombinant strains in Seoul, Korea. *Arch. Virol.* **156**, 323–329.

Haslng, E.M., Lee, E.B., Prelksaltls, K.J., Tellier, R., Honish, L., Senthilselven, A., and Pang, L.X. 2013. Emergence of a new norovirus GII.4 variant and changes in the historical biennial pattern of norovirus outbreak activity in Alberta, Canada, from 2008 to 2013. *J. Clin. Microbiol.* **51**, 2204–2211.

Huh, J.W., Kim, W.H., Moon, S.G., Lee, J.B., and Lim, Y.H. 2008. Viral etiology and incidence associated with acute gastroenteritis in a 5-year survey in Gyeonggi province, South Korea. *J. Clin. Virol.* **44**, 152–156.

Jeong, A.Y., Yoon, S.I., Ji, Y.M., Kang, Y.S., and Lee, Y.M. 2009. Molecular characterization of a Korean isolate of human norovirus, the Hu/NLV/Gunpo/2006/KO strain. *Kor. J. Microbiol.* **45**, 105–111.

Lam, T.Y.T., Zhu, H., Smith, K.D., Guan, Y., Holmes, C.E., and Pybus, G.O. 2012. The recombinant origin of emerging human norovirus GII.4/2008: intra-genotypic exchange of the capsid P2 domain. *J. Gen. Virol.* **93**, 817–822.

Le, V.P., Jung, Y.C., Kang, K.S., Lim, I.S., Myung, S.C., and Kim, W.Y. 2010. Genetic characterization of norovirus GII.4 2006b variants from Jeju island, South Korea. *J. Med. Virol.* **82**, 1665–1070.

Lindesmith, C.L., Beltramello, M., Donaldson, F.E., Corti, D., Swanstrom, J., Debbink, K., Lanzavecchia, A., and Baric, S.R. 2012. Imminogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog.* **8**, e1002705.

Lindesmith, C.L., Costantini, V., Swanstrom, J., Debbink, K., Donaldson, F.E., Vinje, J., and Barica, S.R. 2013. Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. *J. Virol.* **87**, 2803–2813.

Lindesmith, C.L., Donaldson, F.E., LoBue, D.A., Cannon, L.J., Zheng, D.P., Vinje, J., and Baric, S.R. 2008. Mechanisms of GII.4 norovirus persistence in human population. *PLoS Med.* **5**, 269–290.

Mahar, J.E. and Kirkwood, C.D. 2011. Characterization of norovirus strains in Australian children from 2006 to 2008: prevalence of recombinant strains. *J. Med. Virol.* **83**, 2213–2219.

Martella, V., Medici, C.M., Grazia, De S., Tommolo, F., Calderaro, A., Bonura, F., Saporito, L., Terio, V., Catella, C., Lanave, G., and *et al.* 2013. Evidence for recomvination berween pandemic GII.4 norovirus strains New Orleans 2009 and Sydney 2012. *J. Clin. Microbiol.* **51**, 3855–3857.

Mathijs, E., Denayer, S., Palmeira, L., Botteldoorn, N., Scipioni, A., Vanderplasschen, A., Thiry, E., and Dierick, K. 2011. Novel norovirus recombinants and GII.4 sub-lineages associated with outbreaks between 2006 and 2010 in Belgium. *Virol. J.* **8**, 310.

Motomura, K., Oka, T., Yokoyama, M., Nakamura, H., Mori, H., Ode, H., Hansman, S.G., Katayama, K., Kanda, T., Tanaka, T., and *et al.* 2008. Identification of monomorphic and divergent

haplotypes in the 2006-2007 norovirus GII/4 epidemic population by genomewide tracing evolutionary history. *J. Viol.* **82**, 11247–11262.

Nataraju, S.M., Pativada, M.S., Kumar, R., Bhattacharya, M.K., Bagchi, S.R., Kobayashi, N., and Krishnan, T. 2011. Emergence of novel norovirus recombinants with NVGII.1/NVGII.5 RdRp gene and NVGII.13 capsid gene among children and adults in Kolkata, India. *Int. J. Mol. Epidemiol. Genet.* **2**, 130–137.

Pang, L.X., Preiksaitis, K.J., Wong, S., Li, V., and Lee, E.B. 2010. Influence of novel norovirus GII.4 variants on gastroenteritis outbreak dynamics in Alberta and the Northern Territories, Canada between 2000 and 2008. *PLoS One* **5**, e11599.

Park, K.S., Jeong, H.S., Baek, K.A., Lee, C.G., Park, S.M., Park, J.S., Choi, Y.J., Choi, H.J., and Cheon, D.S. 2010. Genetic analysis of norovirus GII.4 variants circulating in Korea in 2008. *Arch. Virol.* **155**, 635–641.

Park, J.W., Lee, S.G., Lee, Y.M., Jheong, W.H., Ryu, S.R., and Paik, S.Y. 2011. Full sequence analysis and characterization of the South Korean Norovirus GII-4 variant CUK-3. *Virol. J.* **8**, 167.

Park, K.S., Yeo, S.G., Jeong, H.S., Beak, K.A., Kim, D.U., Shin, M.H., Song, J.H., Lee, S.J., Choi, Y.J., Park, J.S., and *et al.* 2012. Updates on the genetic variations of Norovirus in sporadic gastroenteritis in Chungnam Korea, 2009-2010. *Virol. J.* **9**, 29.

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

Shanker, S., Choi, J.M., Sankaran, B., Atmar, L.R., Estes, K.M., and Prasad, B.V.V. 2011. Structural analysis of histo-blood antigen biding specificity in a norovirus GII.4 epidemic variant: Implications for epochal evolution. *J. Viol.* **85**, 8635–8645.

Siebenga, J.J., Vennema, H., Renchens, B., Bruin, de E., van der Veer, B., Siezen, J.R., and Koopmans, M. 2007. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* **81**, 9932–9941.

Siebenga, J.J., Vennema, H., Zheng, D.P., Vinjé, J., Lee, E.B., Pang, X.L., Ho, M.C.E., Lim, W., Choudekar, A., Broor, S., and *et al.* 2009. Norovirus illness is a global problem: Emergence and spread of norovirus GII.4 variants, 2001–2007. *J. Infect. Dis.* **200**, 802–812.

Tan, M. and Jiang, X. 2005. The P domain of norovirus capsid protein forms a subviral particle that bind to histo-blood group antigen receptor. *J. Viol.* **79**, 14017–14030.

van Beek, J., Ambert-Balay, K., Botteldoorn, N., Eden, J.S., Fonager, J., Hewitt, J., Iritani, N., Kroneman, A., Vennema, H., Vinje, J., and *et al.* 2013. Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro Surveill.* **18**, 8–9.

Vega, E., Barclay, L., Gregoricus, N., Williams, K., Lee, D., and Vinjé, J. 2011. Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerg. Infect. Dis.* **17**, 1989–1395.

Zakikhany, K., Allen, D.J., Brown, D., and Iturriza-Gómara, M. 2012. Molecular evolution of GII-4 norovirus strain. *PLoS One* **7**, e41625.

Zheng, D.P., Ando, T., Fankhauser, L.R., Beard, S.R., Glass, I.R., and Monroe, S.S. 2006. Norovirus classification and proposed strain nomenclature. *Virology* **346**, 312–323.

Zheng, D.P., Widdowson, M.A., Glass, I.R., and Vinjé, J. 2010. Molecular epidemiology of genogroup II-genotype 4 norovirus in the United States between 1994 and 2006. *J. Clin. Microbiol.* **48**, 168–177.