

入谷展弘 改田 厚, 田中智之, 野田 衛: カキ喫食を伴う食中毒疑い事例からのウイルスの検出, 第 53 回日本臨床ウイルス学会, 豊中市, 6/16 (2012)

入谷展弘, 改田厚, 阿部仁一郎, 久保英幸, 山元誠司, 後藤 薫, 長谷 篤: 2010-11~2011-12 シーズンに大阪市で発生した非細菌性集団胃腸炎事例から検出されたノロウイルスの分子疫学, 平成 24 年度地方衛生研究所全国協議会近畿支部ウイルス部会総会, 大津 (2012. 9. 21)

入谷展弘, 改田厚, 久保英幸, 阿部仁一郎, 西尾治, 後藤薫, 長谷篤: 市販生食用カキにおけるノロウイルスおよび A 型肝炎ウイルス汚染調査 (2006-2007~2009-2010 シーズン), 平成 22 年度地方衛生研究所全国協議会近畿支部ウイルス部会総会, 奈良 (2010. 9. 17)

入谷展弘, 改田厚, 久保英幸, 関口純一郎, 小倉壽, 勢戸祥介: 2009/10 シーズンに大阪市で認められた GII.2 型ノロウイルス流行, 第 58 回日本ウイルス学会, 徳島 (2010. 11. 7-9)

入谷展弘, 改田厚, 山元誠司, 久保英幸: 2012 年 4~6 月に大阪市内保育所で多発した集団胃腸炎事例のウイルス学的調査, 第 60 回日本ウイルス学会, 大阪 (2012. 11. 12-15)

白土東子, 熊谷安希子, 伊藤浩美, 古川早苗, 成松 久, 石井孝司, 染谷雄一, 脇田隆宇, 久保田智巳: X 線結晶構造解析によるノロウイルスと血液型抗原の結合解析, 第 58 回日本ウイルス学会, 平成 22 年 11 月, 徳島

飯塚節子, 斎藤博之, 田中智之, 野田衛: パンソルビン・トラップ法による食品からのノロウイルス遺伝子の検出-弁当屋を原因施設としたノロウイルス集団食中毒事例から-, 第 60 回日本ウイルス学会学術総会, 大阪市, 11/15 (2012)

北元憲利, 岡 智一郎, 片山和彦, 三好龍也, 田中智之. サポウイルス genogroup に特異的な単クローン抗体の作製とその解析. 第 60 回日本ウイルス学会学術集会, 大阪市, 2012 年 11 月 13-15 日,

北元憲利, 岡智一郎, 片山和彦, Hansman GS, 三好龍也, 田中智之. サポウイルスに対する単クローン抗体の解析. 第 58 回日本ウイルス学会学術集会, 2010 年 11 月 7 日~9 日, 徳島

北元憲利, 三好龍也, 田中智之: サポウイルス単クローン抗体による迅速診断法の可能性. 第 53 回日本臨床ウイルス学会. 2012 年 6 月. 大阪.

北島正章, 岡智一郎, 原本英司, 武田直和, 片山和彦, 片山浩之. 国内の下水および河川水からの Genogroup IV ノロウイルスの検出および遺伝子解析. 第 58 回日本ウイルス学会学術集会, 2010 年 11 月 7 日~9 日, 徳島

本村和嗣, 横山 勝, 岡 智一郎, 片山和彦, 野田 衛, 田中智之, 佐藤裕徳. ゲノミクスと計算科学の手法によるノロウイルス GII.4 進化様式の解析. 第 85 回日本感染症学会総会

ワークショップ 2011年4月21-22日, 東京

本村和嗣, 中村浩美, 佐藤彩, 大出裕高, 佐藤裕徳: 次世代シーケンサーを用いた家族内感染例におけるノロウイルス準種解析. 第60回日本ウイルス学会学術集会. 2012年11月13-15日(火-木), 大阪.

名古屋(小原)真弓, 板持(岩井)雅恵, 稲崎倫子, 堀元栄詞, 小渕正次, 佐多徹太郎, 滝澤剛則: ノロウイルスにおけるキメラウイルスが感染性胃腸炎の流行に与える影響. 第60回日本ウイルス学会学術集会, 大阪市, 平成24年11月13~15日

野田 衛, 上間 匡, 多田有希, 中島一敏, 島田智恵, 中村奈緒美, 清原知子, 田中智之, 石井孝司: 2010年のA型肝炎の分子疫学と食品衛生上の原因究明, 第32回日本食品微生物学会学術総会, 江戸川区, 10/6(2011)

野田 衛, 入谷展弘, 中田恵子, 斎藤博之, 田中 忍, 西川 篤, 北堀吉映, 三谷亜里子, 三瀬敬治, 山下 和予, 岡 智一郎, 片山和彦, 岡部信彦: 関西で同時多発的に発生したノロウイルス食中毒事例の解析, 第58回日本ウイルス学会学術集会, 徳島市, 11/7(2010)

野田 衛, 片山和彦, 石井孝司, 岡智一郎, 多田有希, 山下和予, 三瀬敬治, 吉澄志磨, 植木 洋, 林 志直, 山崎匠子, 小原真弓, 吉田徹也, 小林慎一, 中田恵子, 入谷展弘, 三好龍也, 阿部勝彦, 山下育孝, 糸数清正, 中村奈緒美, 島田智恵, 岡部信彦: 塩基配列情報共有化の食品媒介ウイルス感染症の疫学調査への応用, 第31回日本食品微生物学会学術総会, 大津市, 11/11(2010)

野田 衛, 片山和彦, 石井孝司, 岡智一郎, 多田有希, 山下和予, 中村奈緒美, 島田智恵, 岡部信彦, 吉澄志磨, 植木 洋, 林 志直, 山崎匠子, 小原真弓, 吉田徹也, 小林慎一, 中田恵子, 入谷展弘, 三好龍也, 阿部勝彦, 山下育孝, 糸数清正: 食品媒介ウイルス感染症の疫学調査への塩基配列情報共有化の応用(楽しカリシとV-Nus Net Japan), 第22回ウイルス性下痢症研究会, 徳島市, 11/6(2010)

野田 衛, 多田 有希, 田中智之, 清原 知子, 石井 孝司: 2010年のA型肝炎の分子疫学的解析とA型肝炎サーベイランスシステムの構築, 衛生微生物技術協議会第32回研究会, 江戸川区, 6/30(2011)

野田 衛, 多田有希, 田中智之, 清原知子, 石井孝司: 2010年のA型肝炎の分子疫学的解析とA型肝炎サーベイランスシステムの構築. 衛生微生物協議会第32回研究会, 江戸川区, 6月30日, 2011

野田 衛, 多田有希, 田中智之, 石井孝司: 2010年のA型肝炎の分子疫学と食品衛生上の原因究明, 第32回日本食品微生物学会, 平成23年10月, 東京

野田衛, 入谷展弘, 中田恵子, 斎藤博之, 田中忍, 西川篤, 北堀吉映, 三谷亜里子, 三瀬敬治, 山下和予, 岡智一郎, 片山和彦, 岡部信彦. 関西で同時多発的に発生したノロウイルス

食中毒事例の解析. 第 58 回日本ウイルス学会学術集会, 2010 年 11 月 7 日～9 日, 徳島

野田衛, 片山和彦, 石井孝司, 岡智一郎, 多田有希, 山下和予, 三瀬敬治, 地方衛生研究所, 中村奈緒美, 島田智恵, 岡部信彦. 「塩基配列情報共有化の食品媒介ウイルス感染症の疫学調査への応用」第 31 回日本食品微生物学会総会, 2010 年 11 月 11 日～12 日, 滋賀

李 天成, 安田 俊平, 吉松 組子, 片岡 紀代, 吉崎 佐矢香, 網 康至, 須崎百合子, 有川 二郎, 武田 直和, 脇田隆字. E 型肝炎ウイルスに対するラットの感受性. 第 154 回日本獣医学会学術集会 2012 年 9 月 岩手.

李 天成, 片岡 紀代, 網 康至, 須崎 百合子, 安田 俊平, 吉松 組子, 有川 二郎, 武田 直和, 脇田 隆字. ラット E 型肝炎ウイルス様粒子の作製および粒子形成に必要な領域の同定. 第 60 回日本ウイルス学会学術集会 2012 年 11 月 大阪.

李天成, 方 苓, 王 澤均, 宋士利 片岡 紀代, 鈴木 哲朗, 脇田 隆字. ヒトボカウイルス様粒子の作製およびその応用. 日本ウイルス学会, 第 58 回学術集会 2010 年 11 月 徳島

李天成, 方苓, 片岡 紀代, 宮村 達男 脇田 隆字. 日本のブタから分離したブタエンテロウイルス 8 型の解析. 日本ウイルス学会, 第 58 回学術集会 2010 年 10 月 徳島

李天成, 方苓, 網康至, 須崎百合子, 武田直和, 脇田隆字. 不活化 E 型肝炎ワクチンの検討. 日本ウイルス学会, 第 58 回学術集会 2010 年 11 月 徳島

鈴木亮介, 斎藤憲司, 赤澤大輔, 石井孝司, 松浦善治, 脇田隆字, 鈴木哲朗: C 型肝炎ウイルスの trans-packaging 型粒子を用いた感染機構の解析, 第 58 回日本ウイルス学会, 平成 22 年 11 月, 徳島

實方剛, 中野俊也, 谷口孝喜, 油井晶子, 入谷展弘, H. Nyamdavaa, B. Gunchin, B. Choiyants, G. Choiyants: モンゴル国の急性胃腸炎患者から検出された胃腸炎ウイルス, 第 58 回日本ウイルス学会, 徳島 (2010. 11. 7-9)

實方剛, 入谷展弘, 改田厚, 中野俊也, 谷口孝喜, 油井晶子, Batbaatar Gunchin, Gotov Choiyants: モンゴル国における急性胃腸炎患者からのパレコウイルス, ボカウイルス, アイチウイルスの検出状況, 第 60 回日本ウイルス学会, 大阪 (2012. 11. 12-15)

#### IV 研究成果の刊行物・別刷

## Human sapovirus classification based on complete capsid nucleotide sequences

Tomoichiro Oka · Kohji Mori · Nobuhiro Iritani · Seiya Harada ·  
You Ueki · Setsuko Iizuka · Keiji Mise · Kosuke Murakami ·  
Takaji Wakita · Kazuhiko Katayama

Received: 19 August 2011 / Accepted: 7 October 2011 / Published online: 11 November 2011  
© Springer-Verlag 2011

**Abstract** The genetically diverse sapoviruses (SaVs) are a significant cause of acute human gastroenteritis. Human SaV surveillance is becoming more critical, and a better understanding of the diversity and distribution of the viral genotypes is needed. In this study, we analyzed 106 complete human SaV capsid nucleotide sequences to provide a better understanding of their diversity. Based on those results, we propose a novel standardized classification scheme that meets the requirements of the International Calicivirus Scientific Committee. We believe the classification scheme and strains described here will be of value for the molecular characterization and classification of newly detected SaV genotypes and for comparing data worldwide.

Sapoviruses (SaVs) cause gastroenteritis in humans and are a significant public-health problem. Numerous studies have documented their importance in outbreaks of sporadic

gastroenteritis and in contaminated food destined for human consumption [1, 3, 7, 8, 10–12, 16–19, 21, 23].

The SaV genome is a positive-sense, nonsegmented single-strand RNA molecule of approximately 7.5 kb that is polyadenylated at the 3' terminus. The SaV genome contains two or three open reading frames (ORFs). ORF1 encodes non-structural proteins (i.e., VPg, protease, and RNA-dependent RNA polymerase [RdRp]) and a capsid protein (VP1). ORF2 and ORF3 encode proteins of unknown function. The capsid protein is thought to contain all of the determinants for viral attachment and antigenicity [2, 5].

The most widely used method of human SaV detection is reverse transcription–PCR (RT-PCR). It has high sensitivity and can be used to analyze the virus genetically [9, 16, 17, 20, 22]. Real-time RT-PCR is also used to detect the virus. It is highly sensitive and useful for quantitative analysis [14]. Direct serotyping based on neutralization is not possible for human SaVs because no cell-culture system supporting SaV replication has been established.

---

T. Oka (✉) · K. Murakami · T. Wakita · K. Katayama  
Department of Virology II, National Institute of Infectious  
Diseases, Gakuen 4-7-1, Musashimurayama-shi,  
Tokyo 208-0011, Japan  
e-mail: oka-t@nih.go.jp

T. Oka  
Food Animal Health Research Program, Ohio Agricultural  
Research and Development Center, Department of Veterinary  
Preventive Medicine, The Ohio State University, Wooster,  
OH 44691, USA

K. Mori  
Tokyo Metropolitan Institute of Public Health, Tokyo, Japan

N. Iritani  
Osaka City Institute of Public Health and Environmental  
Sciences, Osaka, Japan

S. Harada  
Kumamoto Prefectural Institute of Public Health  
and Environmental Science, Kumamoto, Japan

Y. Ueki  
Miyagi Prefectural Institute of Public Health and Environment,  
Miyagi, Japan

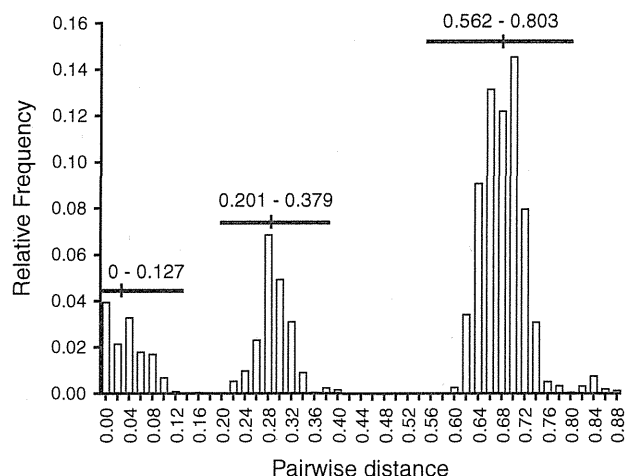
S. Iizuka  
Shimane Prefectural Institute of Public Health and  
Environmental Science, Shimane, Japan

K. Mise  
Sapporo Medical University Center for Medical Education,  
Hokkaido, Japan

From the sequence data for complete capsids, human SaVs are a genetically diverse group [4]. At least four different human SaV genogroups and numerous genotypes have been described, and for the most part, the antigenicity corresponds well to the genetic classification [6, 15]. However, many molecular epidemiological studies have relied on only partially sequenced genes because the broadly reactive RT-PCR primer sets for the SaV capsid-encoding region were directed against highly conserved regions [16, 17, 22]. The International Calicivirus Scientific Committee at the Fourth International Conference on Caliciviruses in Chile (in 2010) proposed that new norovirus and SaV genotypes should only be named after the complete capsid sequence (approximately 1700 nt in length) is determined and compared to other complete capsid sequences. A standardized genotyping and nomenclature system based on appropriate criteria is necessary to facilitate worldwide comparison of data.

In this study, we have analyzed the nucleotide sequences of 107 complete SaV capsids (106 human SaVs and one porcine SaV). Ninety-one human SaV and one porcine SaV sequences were collected from the calicivirus database (<http://teine.cc.sapmed.ac.jp/~calici/ddbj/>) on February 23, 2011. The sequences of 16 additional human SaVs from gastroenteritis patients in five prefectures of Japan were newly determined by PCR-based direct sequencing as described [11, 12] and deposited in the DDBJ/Genbank/EMBL database under the following accession numbers: AB429084, AB522390–AB522392, AB622429, AB622432, AB622435–AB622439, AB622441, AB623037, AB630067, AB630068, and AB630340. The nucleic acid sequences were aligned with ClustalW (version 1.83), and the gaps were removed. The genetic distances were calculated using the Kimura's two-parameter method, and a distance matrix file was created (<http://clustalw.ddbj.nig.ac.jp/top-j.html>). These data were used to generate histograms of the relative frequency distributions of pairwise distance values (Graph Pad Prism version 4.0). The phylogenetic tree with 1000 bootstrap replicates was generated by the neighbor-joining (NJ) method using Clustal W (version 1.83) and was drawn with NJplot software (<http://pbil.univ-lyon1.fr/software/njplot.html>).

The frequency histogram of pairwise distance values with 107 SaV complete capsid nucleotide sequences resulted in three clearly distinct and non-overlapping symmetrical peaks (0–0.112 [mean 0.036], 0.153–0.414 [mean 0.290], and 0.606–0.872 [mean 0.683]) (Fig. 1). These three peaks were considered to represent the strain, genotype, and genogroup, respectively. The mean  $\pm$  3SD of each pairwise distance value for strain, genotype, and genogroup were 0–0.127, 0.201–0.379, and 0.562–0.803, respectively (Fig. 1). The cutoff values for the genotype



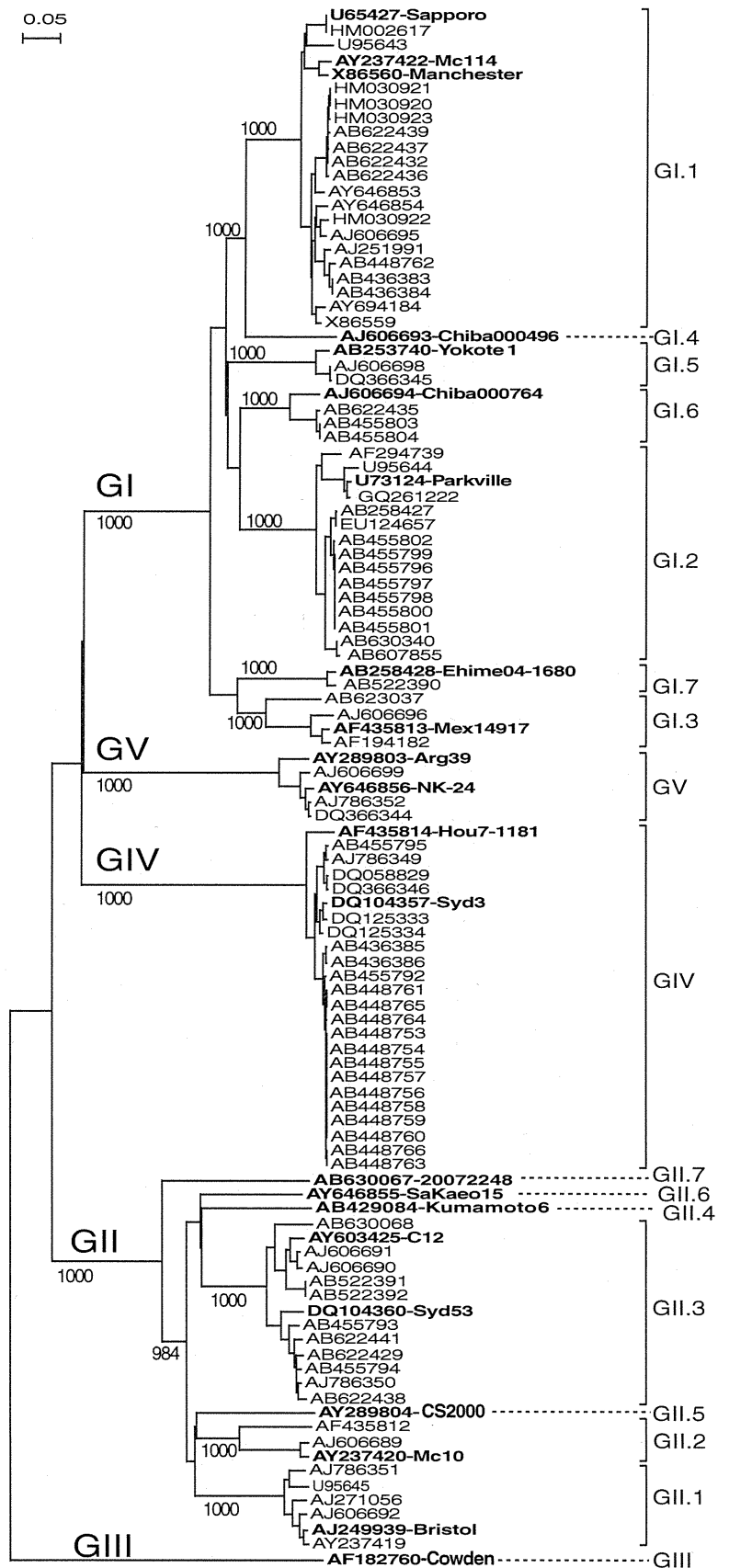
**Fig. 1** Pairwise distance distribution of 107 SaV strains using the entire capsid nucleotide sequences. The horizontal bars indicates the mean value  $\pm$  3 SD for each distribution peaks of strain, genotype, and genogroup, respectively

and genogroup clusters were  $<0.200$  and  $<0.561$ , respectively. A phylogenetic tree of the 107 SaV complete capsid nucleotide sequences was constructed by the NJ method (Fig. 2). SaV was divided into five genogroups, GI, GII, GIII, GIV, and GV, among which human SaV were classified into GI, GII, GIV, and GV. SaV GI and GII were each subdivided into seven genotypes, and GIV and GV were placed into a single genotype by the distance criteria as described above. The genotype numbers were assigned consecutively as shown in Fig. 2.

Recent experiments with virus-like particles (VLPs) have demonstrated the different antigenicities among the GI, GII, GIV, and GV strains [6], and between GI.1 vs GI.5 [6, 7], and GII.2 vs GII.3 strains [15]. Therefore, a classification scheme based on the complete capsid sequence might well reflect the antigenic phenotype of SaV and can also be useful for selecting representative strains for preparing VLPs or VP1 panels for further immunological studies.

In conclusion, we provide a novel human SaV classification based on complete capsid gene sequences. Our classification scheme is based on statistically defined cutoff pairwise distance values. This classification method will aid in the molecular characterization and classification of newly detected SaV genotypes. Recently, a web-based automated genotyping tool for noroviruses was established by the National Institute for Public Health and the Environment and Centers for Disease Control and Prevention [13]. As the number of SaV surveillance laboratories appears to be increasing, a similar web-based genotyping tool with the reference nucleotide sequence sets described in this study will also be valuable for SaV data comparison.

**Fig. 2** Phylogenetic trees constructed using the entire capsid nucleotide sequences of 107 SaV strains. The accession numbers and assigned genogroup and genotypes are indicated. Representative strain(s) of each genogroup or genotype are indicated in boldface type. The bootstrap values correspond to 1000 replications. The number on each branch indicates the bootstrap value, where a value higher than 950 is indicated. The scale represents nucleotide substitutions per site



**Acknowledgments** This work was supported in part by grants for Research on Emerging and Re-emerging Infectious Diseases, as well as Research on Food Safety, from the Ministry of Health, Labour, and Welfare of Japan. We thank Mineyuki Okada for his assistance in sequencing and database submission.

## References

- Akihara S, Phan TG, Nguyen TA, Yagyu F, Okitsu S, Muller WE, Ushijima H (2005) Identification of sapovirus infection among Japanese infants in a day care center. *J Med Virol* 77:595–601
- Clarke IN, Lambden PR (2000) Organization and expression of calicivirus genes. *J Infect Dis* 181(Suppl 2):S309–S316
- Dey SK, Phan TG, Nguyen TA, Nishio O, Salim AF, Yagyu F, Okitsu S, Ushijima H (2007) Prevalence of sapovirus infection among infants and children with acute gastroenteritis in Dhaka City, Bangladesh during 2004–2005. *J Med Virol* 79:633–638
- Farkas T, Zhong WM, Jing Y, Huang PW, Espinosa SM, Martinez N, Morrow AL, Ruiz-Palacios GM, Pickering LK, Jiang X (2004) Genetic diversity among sapoviruses. *Arch Virol* 149:1309–1323
- Green KY (2007) Caliciviridae: the Noroviruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) *Fields Virology*, 5th edn. Lippincott Williams & Wilkins, Philadelphia, pp 949–979
- Hansman GS, Oka T, Sakon N, Takeda N (2007) Antigenic diversity of human sapoviruses. *Emerg Infect Dis* 13:1519–1525
- Hansman GS, Saito H, Shibata C, Ishizuka S, Oseto M, Oka T, Takeda N (2007) Outbreak of gastroenteritis due to sapovirus. *J Clin Microbiol* 45:1347–1349
- Harada S, Okada M, Yahiro S, Nishimura K, Matsuo S, Miyasaka J, Nakashima R, Shimada Y, Ueno T, Ikezawa S, Shinozaki K, Katayama K, Wakita T, Takeda N, Oka T (2009) Surveillance of pathogens in outpatients with gastroenteritis and characterization of sapovirus strains between 2002 and 2007 in Kumamoto Prefecture, Japan. *J Med Virol* 81:1117–1127
- Honma S, Nakata S, Sakai Y, Tatsumi M, Numata-Kinoshita K, Chiba S (2001) Sensitive detection and differentiation of Sapporo virus, a member of the family Caliciviridae, by standard and booster nested polymerase chain reaction. *J Med Virol* 65:413–417
- Iizuka S, Oka T, Tabara K, Omura T, Katayama K, Takeda N, Noda M (2010) Detection of sapoviruses and noroviruses in an outbreak of gastroenteritis linked genetically to shellfish. *J Med Virol* 82:1247–1254
- Ishida S, Yoshizumi S, Miyoshi M, Ikeda T, Okui T, Katayama K, Takeda N, Oka T (2008) Characterization of sapoviruses detected in Hokkaido, Japan. *Jpn J Infect Dis* 61:504–506
- Iwakiri A, Ganmyo H, Yamamoto S, Otao K, Mikasa M, Kizoe S, Katayama K, Wakita T, Takeda N, Oka T (2009) Quantitative analysis of fecal sapovirus shedding: identification of nucleotide substitutions in the capsid protein during prolonged excretion. *Arch Virol* 154:689–693
- Kroneman A, Vennema H, Deforche K, v d Avoort H, Penaranda S, Oberste MS, Vinje J, Koopmans M (2011) An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 51:121–125
- Oka T, Katayama K, Hansman GS, Kageyama T, Ogawa S, Wu FT, White PA, Takeda N (2006) Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. *J Med Virol* 78:1347–1353
- Oka T, Miyashita K, Katayama K, Wakita T, Takeda N (2009) Distinct genotype and antigenicity among genogroup II sapoviruses. *Microbiol Immunol* 53:417–420
- Okada M, Shinozaki K, Ogawa T, Kaiho I (2002) Molecular epidemiology and phylogenetic analysis of Sapporo-like viruses. *Arch Virol* 147:1445–1451
- Okada M, Yamashita Y, Oseto M, Shinozaki K (2006) The detection of human sapoviruses with universal and genogroup-specific primers. *Arch Virol* 151:2503–2509
- Pang XL, Lee BE, Tyrrell GJ, Preiksaitis JK (2009) Epidemiology and genotype analysis of sapovirus associated with gastroenteritis outbreaks in Alberta, Canada: 2004–2007. *J Infect Dis* 199:547–551
- Svraka S, Vennema H, van der Veer B, Hedlund KO, Thorhagen M, Siebenga J, Duizer E, Koopmans M (2010) Epidemiology and genotype analysis of emerging sapovirus-associated infections across Europe. *J Clin Microbiol* 48:2191–2198
- Vinje J, Deijl H, van der Heide R, Lewis D, Hedlund KO, Svensson L, Koopmans MP (2000) Molecular detection and epidemiology of Sapporo-like viruses. *J Clin Microbiol* 38:530–536
- Yamashita Y, Ootsuka Y, Kondo R, Oseto M, Doi M, Miyamoto T, Ueda T, Kondo H, Tanaka T, Wakita T, Katayama K, Takeda N, Oka T (2010) Molecular characterization of Sapovirus detected in a gastroenteritis outbreak at a wedding hall. *J Med Virol* 82:720–726
- Yan H, Yagyu F, Okitsu S, Nishio O, Ushijima H (2003) Detection of norovirus (GI, GII), Sapovirus and astrovirus in fecal samples using reverse transcription single-round multiplex PCR. *J Virol Methods* 114:37–44
- Yoshida T, Kasuo S, Azegami Y, Uchiyama Y, Satsumabayashi K, Shiraishi T, Katayama K, Wakita T, Takeda N, Oka T (2009) Characterization of sapoviruses detected in gastroenteritis outbreaks and identification of asymptomatic adults with high viral load. *J Clin Virol* 45:67–71



## Divergent Evolution of Norovirus GII/4 by Genome Recombination from May 2006 to February 2009 in Japan<sup>∇‡</sup>

Kazushi Motomura,<sup>1</sup> Masaru Yokoyama,<sup>1</sup> Hirotaka Ode,<sup>1</sup> Hiromi Nakamura,<sup>1</sup> Hiromi Mori,<sup>1</sup> Tadahito Kanda,<sup>1</sup> Tomoichiro Oka,<sup>2</sup> Kazuhiko Katayama,<sup>2</sup> Mamoru Noda,<sup>3</sup> Tomoyuki Tanaka,<sup>4</sup> Naokazu Takeda,<sup>2#</sup> Hironori Sato,<sup>1\*</sup> and the Norovirus Surveillance Group of Japan<sup>†</sup>

Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo 208-0011, Japan<sup>1</sup>; Department of Virology II, National Institute of Infectious Diseases, Tokyo 208-0011, Japan<sup>2</sup>; National Institute of Health Sciences, Tokyo 158-8501, Japan<sup>3</sup>; and Sakai City Institute of Public Health, Osaka 590-0953, Japan<sup>4</sup>

Received 7 October 2009/Accepted 29 May 2010

**Norovirus GII/4 is a leading cause of acute viral gastroenteritis in humans. We examined here how the GII/4 virus evolves to generate and sustain new epidemics in humans, using 199 near-full-length GII/4 genome sequences and 11 genome segment clones from human stool specimens collected at 19 sites in Japan between May 2006 and February 2009. Phylogenetic studies demonstrated outbreaks of 7 monophyletic GII/4 subtypes, among which a single subtype, termed 2006b, had continually predominated. Phylogenetic-tree, bootscanning-plot, and informative-site analyses revealed that 4 of the 7 GII/4 subtypes were mosaics of recently prevalent GII/4 subtypes and 1 was made up of the GII/4 and GII/12 genotypes. Notably, single putative recombination breakpoints with the highest statistical significance were constantly located around the border of open reading frame 1 (ORF1) and ORF2 ( $P \leq 0.000001$ ), suggesting outgrowth of specific recombinant viruses in the outbreaks. The GII/4 subtypes had many unique amino acids at the time of their outbreaks, especially in the N-term, 3A-like, and capsid proteins. Unique amino acids in the capsids were preferentially positioned on the outer surface loops of the protruding P2 domain and more abundant in the dominant subtypes. These findings suggest that intersubtype genome recombination at the ORF1/2 boundary region is a common mechanism that realizes independent and concurrent changes on the virion surface and in viral replication proteins for the persistence of norovirus GII/4 in human populations.**

*Norovirus* (NoV) is a nonenveloped RNA virus that belongs to the family *Caliciviridae* and can cause acute gastroenteritis in humans. The NoV genome is a single-stranded, positive-sense, polyadenylated RNA that encodes three open reading frames, ORF1, ORF2, and ORF3 (68). ORF1 encodes a long polypeptide (~200 kDa) that is cleaved in the cells by the viral proteinase (3C<sup>pro</sup>) into six proteins (4). These proteins function in NoV replication in host cells (19). ORF2 encodes a viral capsid protein, VP1. The capsid gene evolved at a rate of  $4.3 \times 10^{-3}$  nucleotide substitutions/site/year (7), which is compara-

ble to the substitution rates of the envelope and capsid genes of human immunodeficiency virus (30). The capsid protein of NoV consists of a shell (S) and two protruding (P) domains: P1 and P2 (47). The S domain is relatively conserved within the same genetic lineages of NoVs (38) and is responsible for the assembly of VP1 (6). The P1 subdomain is also relatively conserved (38) and has a role in enhancing the stability of virus particles (6). The P2 domain is positioned at the most exposed surface of the virus particle (47) and forms binding clefts for putative infection receptors, such as human histo-blood group antigens (HBGA) (8, 13, 14, 60). The P2 domain also contains epitopes for neutralizing antibodies (27, 33) and is consistently highly variable even within the same genetic lineage of NoVs (38). ORF3 encodes a VP2 protein that is suggested to be a

\* Corresponding author. Mailing address: Pathogen Genomics Center, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-Murayama-shi, Tokyo 208-0011, Japan. Phone: 81-42-5610771. Fax: 81-42-5675632. E-mail: hirosato@nih.go.jp.

# Present address: Research Collaboration Center on Emerging and Re-emerging Infections, National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Tivanond 14 Road, Muang, Nonthaburi 11000, Thailand.

† Participants of the Norovirus Surveillance Group of Japan who contributed to this study include Shima Yoshizumi (Hokkaido Institute of Public Health), Toshiyuki Mikami (Aomori Institute of Public Health), Hiroyuki Saito (Akita Prefectural Research Center for Public Health and Environment), You Ueki (Miyagi Prefectural Institute of Public Health and Environment), Akemi Takahashi (Research Institute of Environmental Sciences and Public Health of Iwate Prefecture), Tetuo Hebiguchi (Research Institute of Environmental Sciences and Public Health of Iwate Prefecture), Kuniko Shinozaki (Chiba Prefectural Institute of Public Health), Tetsuya Yoshida (Nagano Environmental Conservation Research Institute), Tsutomu Tamura (Niigata Prefectural Institute of Public Health and Environmental Sciences), Takenori Takizawa (Toyama Institute of Health), Miho Toho

(Fukui Prefectural Institute of Public Health and Environmental Science), Shinichi Kobayashi (Aichi Prefectural Institute of Public Health), Kiyoko Uchino (Sakai City Institute of Public Health), Nobuhiro Iritani (Osaka City Institute of Public Health and Environmental Sciences), Setsuko Iizuka (Shimane Prefectural Institute of Public Health and Environmental Science), Fumiaki Itoh (Hiroshima City Institute of Public Health), Shinji Fukuda (Hiroshima Prefectural Technology Research Institute), Reiko Kondo (Ehime Prefecture Institute of Public Health and Environmental Science), Yasutaka Yamashita (Ehime Prefecture Institute of Public Health and Environmental Science), Sadayuki Funatsumaru (Saga Prefectural Institute of Public Health and Pharmaceutical Research), Yumiko Matsuoka (Kumamoto City Environmental Research Institute), and Akira Iwakiri (Miyazaki Prefectural Institute for Public Health and Environment).

‡ Supplemental material for this article may be found at <http://jvi.asm.org/>.

<sup>∇</sup> Published ahead of print on 9 June 2010.

minor structural component of virus particles (18) and to be responsible for the expression and stabilization of VP1 (5).

Thus far, the NoVs found in nature are classified into five genogroups (GI to GV) and multiple genotypes on the basis of the phylogeny of capsid sequences (71). Among them, genogroup II genotype 4 (GII/4), which was present in humans in the mid-1970s (7), is now the leading cause of NoV-associated acute gastroenteritis in humans (54). The GII/4 is further subclassifiable into phylogenetically distinct subtypes (32, 38, 53). Notably, the emergence and spread of a new GII/4 subtype with multiple amino acid substitutions on the capsid surface are often associated with greater magnitudes of NoV epidemics (53, 54). In 2006 and 2007, a GII/4 subtype, termed 2006b, prevailed globally over preexisting GII/4 subtypes in association with increased numbers of nonbacterial acute gastroenteritis cases in many countries, including Japan (32, 38, 53). The 2006b subtype has multiple unique amino acid substitutions that occur most preferentially in the protruding subdomain of the capsid, the P2 subdomain (32, 38, 53). Together with information on human population immunity against NoV GII/4 subtypes (12, 32), it has been postulated that the accumulation of P2 mutations gives rise to antigenic drift and plays a key role in new epidemics of NoV GII/4 in humans (32, 38, 53).

Genetic recombination is common in RNA viruses (67). In NoV, recombination was first suggested by the phylogenetic analysis of an NoV genome segment clone: a discordant branching order was noted with the trees of the 3D<sup>pol</sup> and capsid coding regions (21). Subsequently, many studies have reported the phylogenetic discordance using sequences from various epidemic sites in different study periods (1, 10, 11, 16, 17, 22, 25, 40, 41, 44–46, 49, 51, 57, 63, 64, 66). These results suggest that genome recombination frequently occurs among distinct lineages of NoV variants *in vivo*. However, the studies were done primarily with direct sequencing data of the short genome portion, and information on the cloned genome segment or full-length genome sequences is very limited (21, 25). Therefore, we lack an overview of the structural and temporal dynamics of viral genomes during NoV epidemics, and it remains unclear whether NoV mosaicism plays a role in these events.

To clarify these issues, we collected 199 near-full-length genome sequences of GII/4 from NoV outbreaks over three recent years in Japan, divided them into monophyletic subtypes, analyzed the temporal and geographical distribution of the subtypes, collected phylogenetic evidence for the viral genome mosaicism of the subtypes, identified putative recombination breakpoints in the genomes, and isolated mosaic genome segments from the stool specimens. We also performed computer-assisted sequence and structural analyses with the identified subtypes to address the relationship between the numbers of P2 domain mutations at the times of the outbreaks and the magnitudes of the epidemics. The obtained data suggest that intersubtype genome recombination at the ORF1/2 boundary region is common in the new GII/4 outbreaks and promotes the effective acquisition of mutation sets of heterogeneous capsid surface and viral replication proteins.

## MATERIALS AND METHODS

**Stool specimens.** The Norovirus Surveillance Group of Japan collected stool specimens from NoV-GII- or GII/4-positive individuals with acute gastroenteritis ( $n = 247$ ). Most of the specimens were from NoV outbreaks around the collection sites. The group collected the specimens in spring, summer, autumn, and winter for 3 years: the 2006/2007 season (May 2006 to January 2007), 2007/2008 season (March 2007 to February 2008), and 2008/2009 season (May 2008 to February 2009). The collection sites were located at 20 different regional public health institutes in Japan (five samples from each institute per year). The genogroup of NoVs was evaluated by real-time reverse transcription-PCR (RT-PCR) (23). In some cases, the genotype of NoVs was evaluated by sequencing of the reverse transcription-PCR products of the ORF1 and ORF2 bordering region (29). Near-full-length genome sequences were obtained with 199 of the 247 specimens. Epidemiological information on 37 of the 199 samples from the 2006/2007 season was described previously (38). Information on the rest ( $n = 162$ ) is described in Tables S1 and S2 in the supplemental material. Briefly, the 162 specimens were from outbreaks ( $n = 90$ ), sporadic infection cases ( $n = 15$ ), and undescribed cases ( $n = 57$ ) during December 2006 to February 2009 in Japan. The major sites of the incidences were a nursing care center ( $n = 19$ ), restaurant ( $n = 17$ ), kindergarten ( $n = 15$ ), hotel ( $n = 8$ ), hospital ( $n = 7$ ), sports event ( $n = 1$ ), self-defense force ( $n = 1$ ), family home ( $n = 1$ ), elementary school ( $n = 1$ ), and bank ( $n = 1$ ), and one was undescribed ( $n = 91$ ). The viral RNA copy numbers in the specimens ranged from  $5.0 \times 10^4$  to  $1.9 \times 10^{11}$  copies/g stool (average,  $6.1 \times 10^9$  copies/g stool) as judged by the real-time quantitative reverse transcription-PCR assay (23). All stool specimens were stored at  $-80^\circ\text{C}$  until use.

**Viral genome sequencing.** NoV GII/4 genome sequencing was done as described previously (38). Briefly, two overlapping fragments (approximately 5.2 and 2.5 kb) were amplified by RT-PCR from stool specimens. The PCR products were purified and used as a template for sequencing in a 96-well scale using an ABI 3730 xl DNA analyzer (Applied Biosystems, Foster City, CA). The sequences of 5.2-kb and 2.5-kb segments from the same individual were used to reconstruct near-full-length genome sequences (about 7.5 kb) by alignment at an overlapping region using the Staden Package (<http://staden.sourceforge.net>). The 5.2-kb fragment covers the complete ORF1 and the 5' end of ORF2. The 2.5-kb fragment covers the 3' end of ORF1, complete ORF2 and ORF3, and 3'-end noncoding region of the genome. The primers used for reverse transcription and nested PCR for the 5.2-kb fragment were GII4-1F/GII4r5412 (outer primer pair) and GII4-2F/GII4r5295 (inner primer pair) (38). Those for the 2.5-kb fragment were COG-2F/Tx30SXN (outer primer pair) and G2SKF/Tx30SXN (inner primer pair) (38). The initial 22 nucleotides at the 5' ends of the reconstructed genomes were from PCR primers. The final 45 nucleotides at the 3' ends of the genome were excluded from analysis because of the low levels of sequence accuracy. We obtained 199 near-full-length genome sequences from 247 GII-positive specimens. The 199 sequences included 37 GII/4 sequences previously reported between May 2006 and January 2007 (38) and 162 sequences newly obtained between December 2006 and February 2009.

**Molecular cloning and sequencing of genome segments.** The 5.2-kb, 1.0-kb, and 2.8-kb genome segments were amplified by RT-PCR products as described above and cloned into pPCR-XL-TOPO vectors (Invitrogen, Carlsbad, CA). Each of the segments covers a junction of putative recombination breakpoints around the 5' end of ORF2: the 5.2-kb segment contains the near-full-length ORF1 and 5'-end portion of ORF2, the 2.8-kb segment contains the 3'-end portion of ORF1, complete ORF2, and complete ORF3, and the 1.0 kb segment contains the 3'-end portion of ORF1 and 5'-end portion of ORF2. The primers used for the nested PCR of the 5.2-kb segment were the same ones described above: GII4-1F/GIIr5412 (outer primer pair) and GII4-2F/GIIr5295 (inner primer pair) (38). The primers used for the nested PCR of the 2.8-kb fragment were GII4f4117 (5'-CTGACAAAATTTATGGTAAGATCAAGAAGAGG-3')/Tx30 SXN (outer primer pair) and GII4f4762 (5'-GACCCAGCTGGTTGGTTGGAAAA-3')/GII4r7516 (5'-ATAGTTTACGGCCGCATTCTTATCACA TTACACCCGTGACTCCCTCG-3') (inner primer pair). The primers used for the nested PCR of the 1.0-kb fragment were GII4f4117/GII4r5412 (outer primer pair) and GII4f4223 (5'-GGTATGAATATGAATGAGGATG-3')/GII4r5295 (inner primer pair). The single clones of the 5.2-kb, 2.8-kb, and 1.0-kb genome segments of the GII/4 subtypes were randomly chosen and sequenced in 96-well plates using an ABI 3730 xl DNA analyzer as described above.

**Phylogenetic analysis.** Phylogenetic trees were constructed using the neighbor-joining method and maximum-likelihood method. Briefly, the near-full-length genome sequences from this study were aligned with the available GII/4 genome sequences from past NoV epidemics occurring over the past 3 decades, using CLUSTAL W software included in the MEGA software package, version 4.0

(58) (<http://evolgen.biol.metro-u.ac.jp/MEGA/>) and the MAFFT multiple sequence alignment software program, version 6.0 (26) (<http://align.bmr.kyushu-u.ac.jp/mafft/software/>). The neighbor-joining trees were constructed with the nucleotide substitution values estimated with the maximum composite likelihood model (59) using MEGA. The maximum-likelihood trees were inferred on the basis of the general time reversible models (31) using the PHYML software program included in the RDP3 software package (35) (<http://darwin.uvigo.es/rdp/rdp.html>). The reliability of interior branches in the phylogenetic tree was assessed by the bootstrap method with 1,000 resamplings. The GII/4 genome reference sequences were from samples taken before 1990 (<1990) (6 sequences, CHDC591-1974, CHDC2490-1974, CHDC4871-1977, CHDC4108-1987, Lordsdale, MD145-12/US/1987, and CHDC3967-1988), before 2000 (<2000) (2 sequences, and Dresden174/US/1997), in 2002/2003 (6 sequences, Famington Hill, B2S16/2002/UK, B5S22/2002/UK, Langen1061/2002/DE, YURI32073/2002/JPN, and MD-2004/2004/US), and in 2004/2005 (4 sequences, Guangzhou/NVgz01/CHN/2006, Chiba/04-1050/2005/JP, Sakai/04-179/2005/JP, and Ehime/05-30/2005/JP). Accession numbers for the reference genome sequences are given elsewhere (7, 38).

We initially constructed the phylogenetic trees with 199 genome sequences from the Japanese variants from 2006 to 2009 and 6 representative sequences of GII genotypes whose complete genome sequences were available in GenBank in October 2009 (GII/1, GII/3, GII/4, GII/6, GII/10, and GII/12; accession no. U07611, AB067542, X86557, AB039776, AY237415, and AB039775, respectively). The trees showed that the 199 genome sequences reproducibly grouped with the GII/4 reference sequences outside other GII references. The GII/4 cluster was positioned most closely to the GII/12 reference (Saitama U1/JP). Therefore, we used GII/12 as an outgroup in the present study for a better grasp of the relationship of the phylogeny among the Japanese GII/4 variant subgroups and between GII/4 and GII/12 variants.

**Bootscanning-plot analysis.** Bootsclanning-plot analysis was performed as described previously (69). Briefly, each query sequence was aligned with three NoV reference sequences using CLUSTAL W software, version 1.4 (62). The bootstrap values were plotted for a window of 300 bp, moving in increments of 10 bp along the alignment using the software program Simplot (48) (version 3.5.1; <http://sray.med.som.jhmi.edu/SCRSoftware/simplot/>). Thus far, 19 genotypes of the NoV GII variants have been reported on the basis of complete capsid sequences (65, 71). Among them, only 7 genotypes have been fully sequenced at the genome level (GII/1, GII/3, GII/4, GII/6, GII/8, GII/10, and GII/12; accession numbers U07611, AB067542, X86557, AB039776, AB067543, AY237415, and AB039775, respectively). To search for sequences that are phylogenetically relevant to the query sequences, we constructed phylogenetic trees of the complete ORF1, ORF2, and ORF3 sequences using all available representatives of the 19 genotypes in the GenBank database. We also used the automated exploratory analysis tool included in the RDP3 software package (35). The genome sequence set used for the analysis consisted of 7 query sequences (2004/05, 2006a, 2006b, 2007a, 2007b, 2008a, and 2008b), all available GII genotype representatives, and all available GII/4 variant subgroups which caused epidemics over the past 34 years (7, 31). Two putative parent sequences with the best confidence values and a single distantly related sequence were used for the bootsclanning plots with MEGA. The confidence values of the recombination events were also assessed with tools included in the RDP3 software package, such as RDP, GENECONV, Maxchi, Chimera, 3seq, and Siscan. The query sequences used in this study were Sakai2/2006/JP for the 2004/05 subtype (accession no. AB447448) and representative genomes of the 2007a (Osaka1/2007/JP), 2007b (Iwate5/2007/JP), 2008a (Hokkaido5/2008/JP), and 2008b (Hokkaido4/2008/JP) subtypes obtained in this study. The reference sequences were Saitama\_U1/JP (GII/12 genotype [25], accession no. AB039775), B2S16/2002/UK (2002/03 subtype [38], accession no. AY587989), Saitama\_U3/JP (GII/6 genotype [25], accession no. AB039776), Sakai2/2006/JP (2004/05 subtype [38], accession no. AB447448), Aomori1/2006/JP (2006a subtype [38], accession no. AB447432), Aichi3/2006/JP (2006b subtype [38], accession no. AB447446), and Hokkaido5/2008/JP and Hokkaido4/2008/JP (2008a and 2008b subtypes, respectively, obtained in this study).

**Informative-site analysis.** The informative-site analysis was performed as described previously (50). Briefly, each query sequence was aligned with two putative parental sequences and an outgroup sequence. The alignments were used to identify informative sites that support alternative tree topologies between downstream and upstream regions using the Simplot software program (48), version 3.5.1. This information allowed identification of genome regions that were assigned as chimeras of heterologous sequences of distinct evolutionary origins. The statistical significance of the resultant division by the informative sites was evaluated by the maximal  $\chi^2$  test using in-house programs. The programs were designed to execute the calculation algorithms described by Robertson et al. (50, 55).

**Molecular modeling.** Three-dimensional (3-D) structural models of the capsid P-domain dimers were constructed by homology modeling as described previously (38). Briefly, the P-domain monomer models were first constructed using the crystal structure of the NoV capsid P domain of the GII/4 VA387 strain at a resolution of 2.00 Å (PDB code 2OBS [13]) as the template. The P domains of the GII/4 subtypes described in this study have sequence similarities of greater than 90% to that of VA387, high enough to construct models with a root mean square distance (RMSD) of  $\sim 1$  Å for the main chain between the predicted and actual structures (3). The P-domain monomer models were used to construct the P-domain dimer models by superimposing the chains A and B using the crystal structure of the NoV capsid dimer (PDB code 1IHM [47]).

**Nucleotide sequence accession numbers.** The DDBJ database accession numbers for the nucleotide sequences of NoV genomes for the 2006/2007 season ( $n = 37$ ) have been reported elsewhere (38). The DDBJ database accession numbers for the nucleotide sequences of NoV genomes for the 2007/2008 and 2008/2009 seasons ( $n = 162$ ) are AB541201 to AB541362. The DDBJ database accession numbers for the nucleotide sequences of NoV genome segment clones ( $n = 11$ ) are AB541190 to AB541200.

## RESULTS

**Phylogenetic classification of NoV GII/4 subtypes in Japan during 2006 and 2009.** First, we investigated the phylogeny of the NoV near-full-length genome sequences (about 7.5 kb). For this study, we used sequences obtained in this study from 19 sites in Japan between May 2006 and February 2009 ( $n = 199$ ), various reported GII/4 reference sequences of past global or Japanese epidemics, and various reported outgroup sequences of other NoV genotypes. Figure 1 shows a maximum-likelihood tree constructed with the 199 Japanese genome sequences and the 19 GII/4 reference sequences from past NoV epidemics throughout the world during the 1970s and 1980s (7), <2000, in 2002/2003, and in 2004/2005 (38). The tree shows that the 199 Japanese sequences are divisible into 7 distinct lineage groups within a GII/4 cluster with a high bootstrap value (the 7 colored ovals in Fig. 1). The monophyly of the 7 GII/4 groups was reproducible independently of the algorithms to infer the phylogeny and reference sequences used. We tentatively named the 7 monophyletic subtypes of GII/4 variants 2004/05, 2006a, 2006b, 2007a, 2007b, 2008a, and 2008b.

The 2004/05 genome sequences were first obtained in Japan in the winter of 2004-2005 (accession no. AB220921 to AB220923 [42]). The geographic distribution of the 2004/05 sequences seemed to be restricted to East Asia (54). The 2006a and 2006b genome sequences were first obtained in Japan during the winter of 2006-2007 (accession no. AB447427 to AB447463 [38]). The 2006a and 2006b sequences were detected in many countries in Europe, North America, and East Asia during 2006-2007, wherein the 2006b subtype was generally more dominant than the 2006a subtype (54). The 2007a, 2007b, 2008a, and 2008b genome sequences were newly obtained in this study. Phylogenetic tree analyses showed that the nucleotide sequences of ORF2 of the 2008a subtype were genetically closely related to the ORF2 sequence obtained in the Netherlands in 2008 (accession no. AB445395), and together these sequences formed a single monophyletic group with a high bootstrap value (data not shown). These results suggest that at least 4 of the 7 GII/4 subtypes identified in Japan during 2006 and 2009, i.e., the 2004/05, 2006a, 2006b, and 2008a subtypes, caused NoV infections outside Japan.

We estimated the genetic divergence within and between the 7 monophyletic groups on the basis of the maximum composite

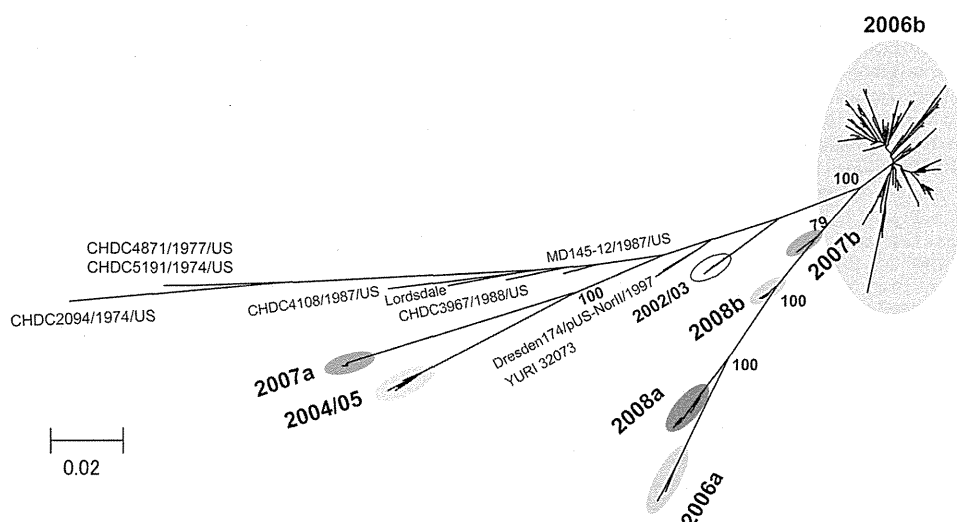


FIG. 1. Phylogenetic classification of the NoV GII/4 subtypes in Japan during 2006 and 2009. The maximum-likelihood tree was constructed with the near-full-length genome sequences (about 7.5 kb) obtained from stool specimens collected at 19 sites in Japan between May 2006 and February 2009 in this study ( $n = 199$ ) and GII/4 reference genome sequences from past epidemics in Japan and other countries in the <2000, 2002/2003, and 2004/2005 winter seasons (7, 38) ( $n = 18$ ). The sequence clusters enclosed by colored ovals indicate the 7 monophyletic GII/4 subtypes identified in Japan in previous (38) and present studies.

likelihood model using MEGA software. The intragroup divergence was comparably high in the 2006b subtype among the 7 groups (see Table S3, diagonal lines, in the supplemental material), suggesting that the diversity of the 2006b genome is higher than that of the other subtypes. This is consistent with the epidemiological data that 2006b had predominated for 3 years in Japan whereas the others emerged only temporally. The intergroup divergence was comparably high between 2004/05 and the other groups and between 2007a and the other groups, and about 12 to 15% sequence divergence existed in the genomes (see Table S3, bottom left portion, in the supplemental material).

**Temporal and geographical distribution of Nov GII/4 subtypes in Japan.** We then analyzed the temporal and geographical distribution of the 7 GII/4 subtypes in Japan. The 199 near-full-length genome sequences were divided into 3 groups according to the collection periods: the 2006/2007 (May 2006 to January 2007) ( $n = 39$ ), 2007/2008 (March 2007 to February 2008) ( $n = 78$ ), and 2008/2009 (May 2008 to February 2009) ( $n = 82$ ) seasons. The frequencies of detection of particular NoV subtypes were obtained for each of the three seasons. We also used published subtyping data for the analysis of the previous winter season in Japan (November 2005 to March 2006) ( $n = 38$ ) (38, 43).

The 2004/05 and 2006a sequences were detected at multiple collection sites and were prevalent in the 2005/2006 season (38, 43) (Fig. 2A and B, 2004/05 and 2006a). However, they became minor in the 2006/2007 season and were hardly detected thereafter. The 2006b sequences were minor in the 2005/2006 season (38) (Fig. 2A, 2006b). However, they rapidly became dominant in the 2006/2007 season and continually predominated in most of the collection sites in Japan, representing 176 of the 199 genome sequences (88.4%) during the study period. This result is consistent with the data of partial capsid sequences obtained during December 2007 to January 2008 in Japan (28).

The 2007a and 2007b sequences were detected only at single collection sites in the 2007/2008 season (Fig. 2A and B, 2007a and 2007b). The 2008a and 2008b sequences were detected most recently at multiple collection sites in the 2008/2009 season (Fig. 2A and B, 2008a and 2008b). These data indicate that the 2006b subtype displaced the 2004/05 subtype in the 2006/2007 season and continued to predominate for the next 2 years in Japan. During the period of the 2006b predominance, however, several GII/4 subtypes caused NoV outbreaks in Japan, and the frequencies and sites of non-2006b outbreaks increased slightly in the 2008/2009 season.

**Phylogenetic evidence for NoV genome mosaicism.** Next, we investigated the possibility of genome mosaicism of the 7 GII/4 subtypes. For this purpose, we first compared the branching orders of the subtype clusters in the maximum-likelihood and neighbor-joining trees of the ORF1, ORF2, and ORF3 sequences using representative sequences of the 19 GII genotypes (GII/1 to GII/19) reported to date in the GenBank database. Figure 3A shows the maximum-likelihood trees, in which most of the non-GII/4 sequences were positioned far from the GII/4 cluster and were therefore excluded for a better grasp of the relationship of the phylogeny among the GII/4 variant subgroups. The exception was the ORF1 tree, in which the GII/12 sequence branched inside the GII/4 cluster. The comparisons of the three trees revealed that there was marked inconsistency in the branching orders of the GII/4 subgroups. The inconsistency was reproducible independently of the algorithms to infer the phylogeny and reference sequences used. First, the ORF1 sequences of the 2006b, 2007a, 2007b, and 2008b subtypes formed independent monophyletic clusters, whereas the ORF2 sequences of the 2006b, 2007a, and 2008b subtypes formed a single cluster and the ORF3 sequences of these four subtypes formed the same cluster (Fig. 3A, light blue circles). Second, the ORF1 sequences of the 2004/05 subtype were clustered near the ORF1 sequence of a GII/12

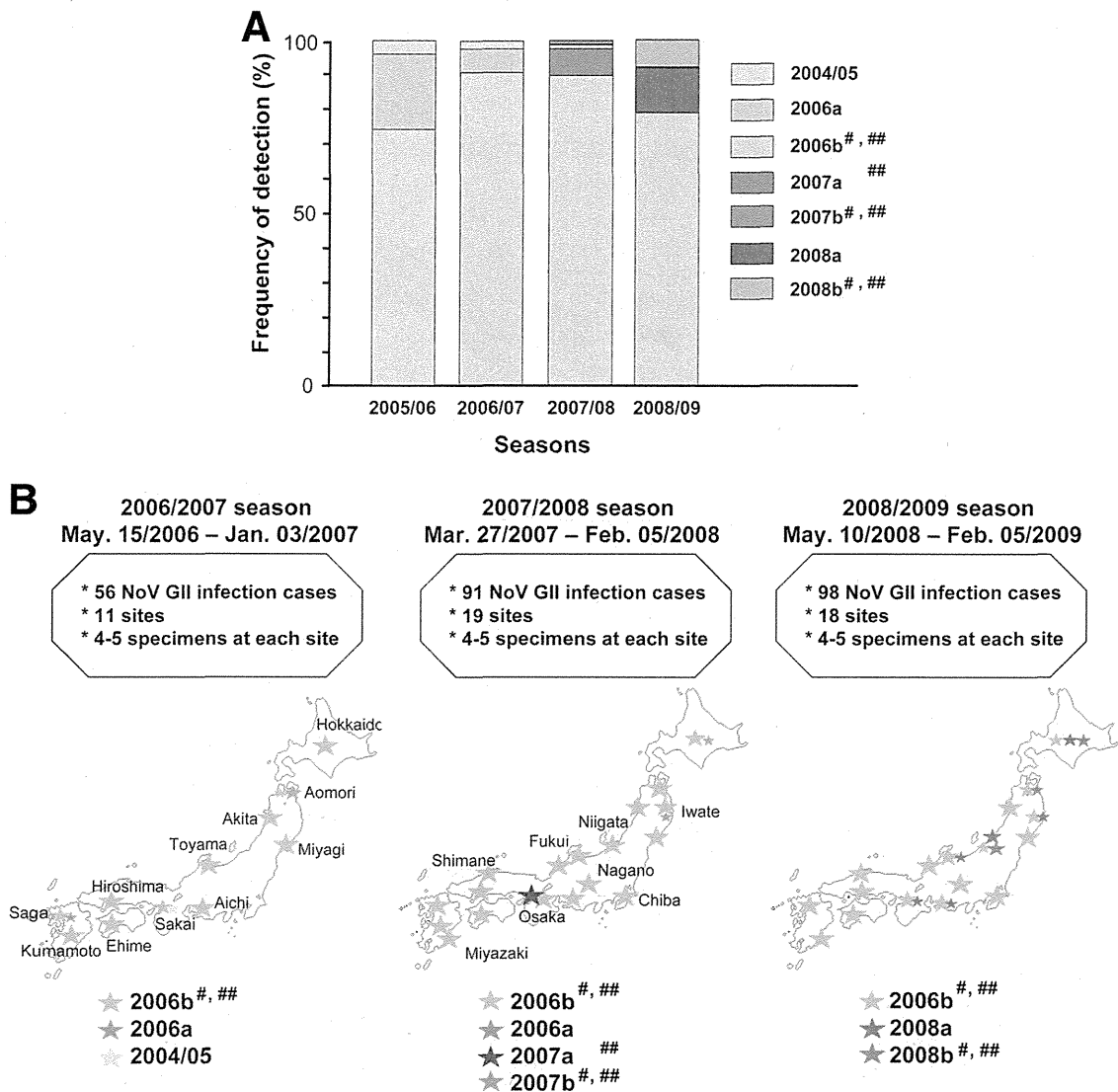


FIG. 2. Temporal and geographical distribution of the NoV GII/4 subtypes in Japan. The 199 near-full-length genome sequences were divided into 3 subgroups according to the collection periods: the 2006/2007 (May 2006 to January 2007) ( $n = 39$ ), 2007/2008 (March 2007 to February 2008) ( $n = 78$ ), and 2008/2009 (May 2008 to February 2009) ( $n = 82$ ) seasons. For the analysis of the 2005/2006 season, published subtyping data (38, 43) were used ( $n = 38$ ). (A) Frequencies of detection of particular NoV GII/4 subtypes in each season in Japan. (B) Geographic locations of the GII/4 subtype outbreaks. Colored stars indicate the locations of sample collection sites. Larger stars indicate the collection sites with greater frequencies of detection. #, ORF2s were classified as the same phylogenetic group (see Fig. 3A, ORF2). ##, ORF3s were classified as the same phylogenetic group (see Fig. 3A, ORF3).

strain (Saitama\_U1/JP [25]) and relatively distant from the reported GII/4 reference sequences, whereas the ORF2 and ORF3 sequences of the 2004/05 subtype were very distantly related to the Saitama\_U1/JP sequence and closely related to the GII/4 reference sequences (Fig. 3A, yellow circles). Third, the branching orders of the 2008a sequences were also different in the ORF1, ORF2, and ORF3 trees (Fig. 3A, red circles). These results suggested that most subtypes identified in this study had mosaic genomes.

To further assess this possibility, we performed bootscanning-plot analyses as described previously (69). For each bootscanning plot, we used a query genome sequence of a given subtype, two to three reference sequences that were positioned relatively closely to the query sequence in the

neighbor-joining trees, and a distantly related outgroup sequence. The analyses showed that the genomes of the 2004/05, 2007a, 2007b, 2008a, and 2008b subtypes were indeed composed of multiple segments from recently prevalent or as-yet-undefined genogroups, genotypes, and subtypes of NoVs in this and previous reports (2, 7, 25, 38, 53, 65, 71) (Fig. 3B; see also Fig. 3A). The 2004/05 genome (Sakai2/2006/JP) was comprised of the ORF1 related to GII/12 (Saitama\_U1/JP) and the ORF2/3 related to GII/4 2002/03 (B2S16/2002/UK). The 2007a genome (Osaka1/2007/JP) was made up of the ORF1 related to GII/12, the ORF2 of as-yet-undefined classes of GII/4, and the ORF3 related to GII/4 2006b (Aichi3/2006/JP). The 2007b genome (Iwate5/2007/JP) was made up of the ORF1 related to GII/4 2006b and 2006a (Aomori1/2006/JP) and the ORF2 and

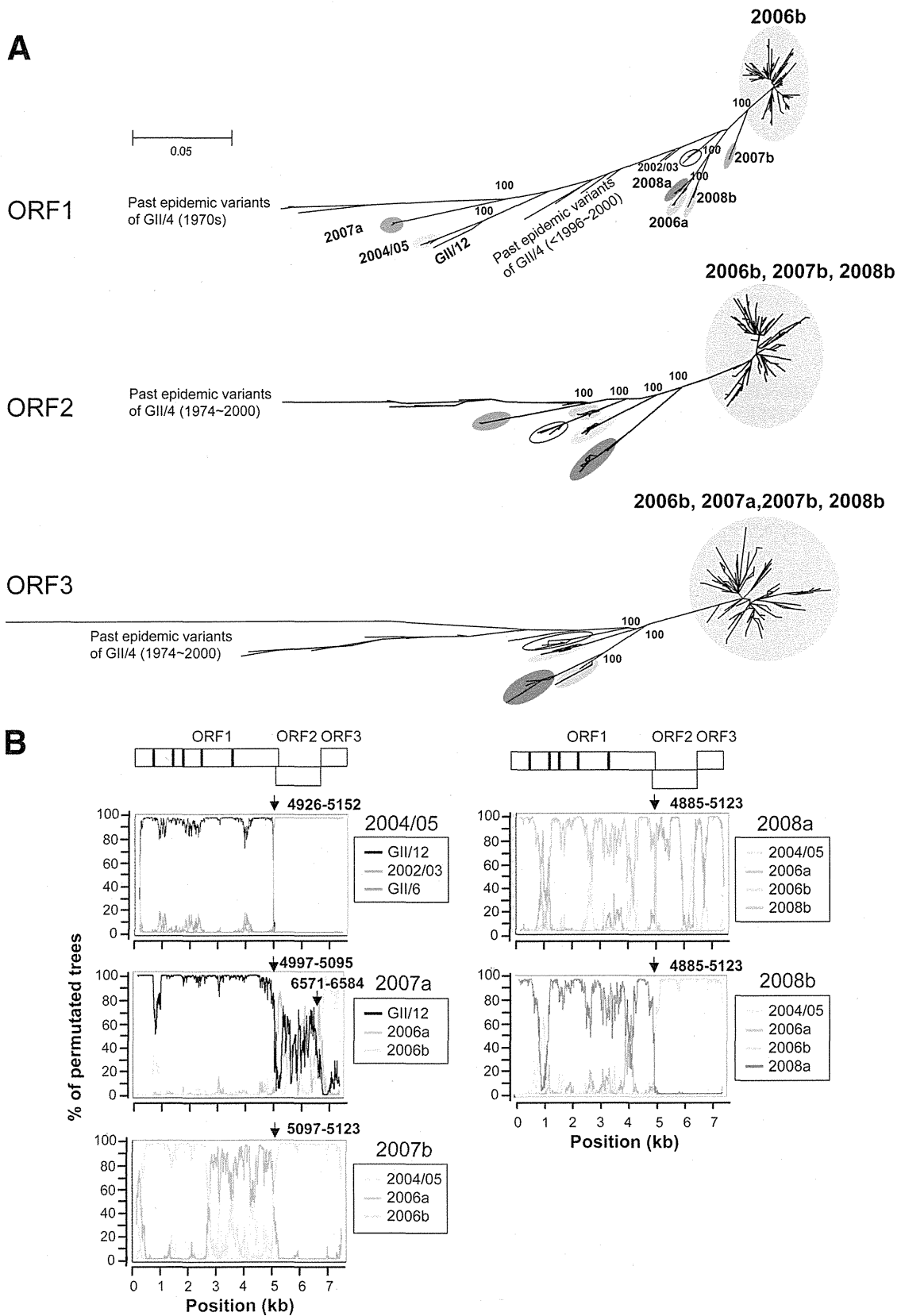


FIG. 3. Phylogenetic evidence for NoV genome mosaicism. (A) Maximum-likelihood trees of the nucleotide sequences of the complete ORF1 (about 5.1 kb), ORF2 (about 1.6 kb), and ORF3 (about 0.8 kb). The trees were constructed with the sequences obtained in previous (38) and present studies ( $n = 199$ ) and the reference sequences described in Fig. 1. The GII/12 sequence (Saitama\_U1/JP [25]) was used as an outgroup

ORF3 related to GII/4 2006b. The 2008a genome (Hokkaido/2008/JP) was comprised of the ORF1 related to GII/4 2008b (Hokkaido5/2008/JP) and as-yet-undefined classes of GII/4 and the ORF2 and ORF3 of GII/4 2006a and as-yet-undefined classes of GII/4. The 2008b genome was made up of the ORF1 of 2008a and the ORF2 and ORF3 related to GII/4 2006b. We also investigated possible genome mosaicism for the 2006a and 2006b subtypes, but we could not identify putative ancestral sequences of ORF1, ORF2, and ORF3 that were genetically closely related to 2006a and 2006b when we used the available NoV sequences in the public database as references.

To define statistically the possible recombination breakpoints of the 2004/05, 2007a, 2007b, 2008a, and 2008b genomes, we performed informative-site analysis (50) using the same reference sequences used in the bootscanning-plot analysis. With this approach, we identified several patches of genome regions that were assigned with statistical significance as putative recombination breakpoints. Notably, a putative breakpoint located around the junction of ORF1 and ORF2 constantly gave the highest statistical significance, i.e., maximum  $\chi^2$  values, in the 2004/05, 2007a, 2007b, 2008a, and 2008b genomes ( $P \leq 0.000001$ ) (Fig. 3B, arrows). The results were in good agreement with the phylogenetic-tree and bootscanning-plot analyses. These data consistently suggest that the new GII/4 subtypes identified in Japan were mostly hybrid viruses composed of viral protein elements from distinct genetic lineages of NoVs.

We further assessed possible genome recombination events using other tools included in the RDP3 software package (30). The analysis again identified single recombination breakpoints with the best or second-best confidence values around the junction of ORF1 and ORF2 in the 2004/05, 2007a, 2007b, 2008a, and 2008b genomes ( $P < 0.001$ ). The analysis also identified additional putative breakpoints around the junction of ORF2 and ORF3 of 2007a. However, we could not obtain evidence for genome mosaicism with 2006a and 2006b using a selected sequence data set of the NoV GII genotypes reported to date (GII/1 to GII/19) (25, 65, 71) and GII/4 subtypes (7, 38, 53). Because information on the entire genome sequences of NoV is very limited, it remains to be determined whether 2006a and 2006b also have mosaic genomes.

**Isolation of NoV mosaic genome segments.** To clarify the presence of the mosaic viral genomes in nature, we cloned and sequenced the genomes of the 2007a, 2007b, 2008a, and 2008b subtypes. For this study, we cloned the genome segments, i.e., the 5.2-kb, 2.8-kb, and 1.0-kb genome segments, that presumably contain a junction of putative recombination breakpoints around the ORF1/ORF2 junction (Fig. 4). The subtype 2004/5 was not included in the cloning analysis because we did not have sufficient amounts of clinical specimens for the cloning.

Likewise, the subtypes 2006a and 2006b were not included because the major parental sequences are not clear in the present study. We successfully obtained the molecular clones except for the 5.2-kb fragment of 2007a. We could amplify but failed to clone the 5.2-kb fragment of 2007a. Although the precise reason for the failure is unclear at present, it might be due to the decreased cloning efficiency of the larger insert by the TA-cloning method. Because the appropriate restriction enzyme sites for the cloning were absent in the 2007a 5.2-kb fragment, we did not include this fragment in further analyses. Nucleotide sequences of the segments were used for the bootscanning-plot analysis using the same sets of reference sequences described in Fig. 3B, and the statistically significant putative recombination breakpoints were assessed by informative-site analysis.

Figure 4 shows representative results of the bootscanning-plot and informative-site analyses with the 5.2-, 1.0-, and 2.8-kb segment clones. Importantly, all 11 clones from the 2007a, 2007b, 2008a, and 2008b stool specimens had the same putative recombination breakpoints, with the highest statistical significance around the ORF1/ORF2 junction region identified with direct sequencing analyses (Fig. 4A and B, arrows). In addition, the patterns of the bootscanning plots were almost identical over the viral genomes examined between the sequences of the uncloned and cloned genome segment except for the 5'-half region of the 2007b ORF1 (Fig. 3B and 4). Although the precise reason for the discrepancy is unclear at present, it might be due to the cloning of the minor population of the 2007b quasispecies in the stool specimens. The overall good agreement of the results by the two sequencing strategies strongly suggests that the genome mosaicism we found by analysis of the direct sequencing data were intrinsic rather than an artifact of the analysis. Taken together, these data indicate that the NoV mosaic genomes were present in the human stool specimens and that the ORF1/ORF2 junction region is the common hot spot for generating the mosaic genomes in GII/4 subtypes in nature.

**Amino acid signatures of the NoV GII/4 subtypes.** We then investigated sequential characteristics of the proteins of the 7 GII/4 subtypes by searching for unique amino acid signatures in viral proteins. The deduced amino acid sequences of ORF1, ORF2, and ORF3 of a given subtype were aligned with reference sequences of the past GII/4 subtypes (38) that were identified before detection of the query subtype. Amino acids specific to the query subtype were extracted and referred to as amino acid signatures of the new epidemic subtype. In the case of the 2006b subtype, we also analyzed the changes in the signatures in the capsid protein VP1 between 2006 and 2009, because information on the structure and function is more abundant for the capsid than for other viral proteins.

sequence in each tree but is shown only in the ORF1 tree. In the ORF2 and ORF3 trees, the GII/12 sequence was located far apart from the GII/4 cluster and is not shown for simplicity. (B) Bootscanning plots of nucleotide sequences of near-full-length NoV genomes. A query genome sequence (2004/05, 2007a, 2007b, 2008a, or 2008b) was aligned with three reference sequences, two sequences that were positioned relatively closely to the query sequence in the neighbor-joining trees and a sequence that was distantly related to the query sequence, using CLUSTAL W software, version 1.4 (62). The bootstrap values are plotted for a window of 300 bp moving in increments of 10 bp along the alignment using the program Simplot (48). Informative-site analyses (50) were performed using the same query and reference sequence set. Arrows indicate putative recombination breakpoints with the highest statistical significance ( $P \leq 0.000001$ ) in the informative-site analysis.

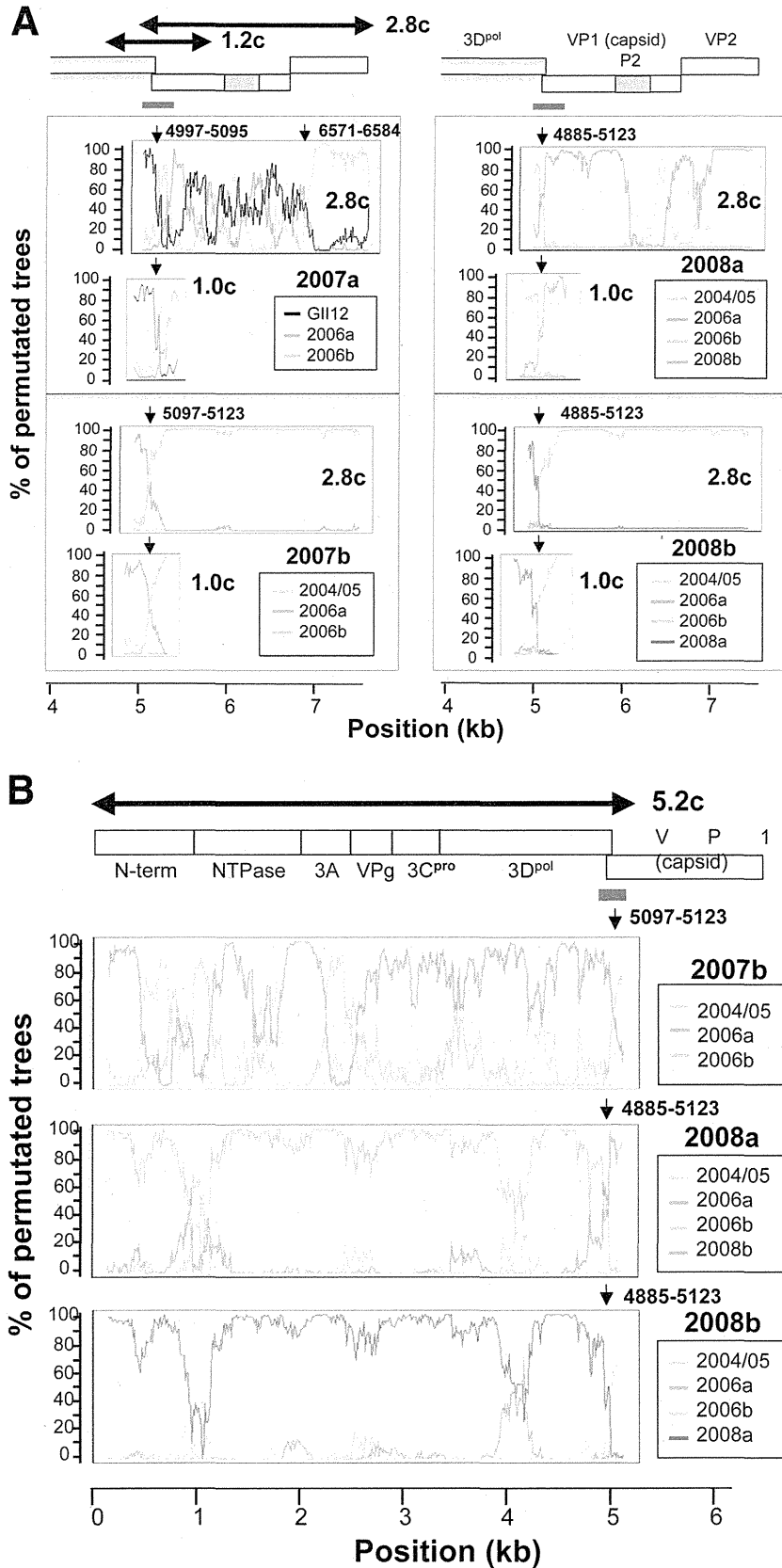


FIG. 4. Isolation of NoV mosaic genome segments. Three genome segments (5.2, 1.0, and 2.8 kb) were amplified from the 2007a, 2007b, 2008a, and 2008b stool specimens, cloned into plasmid vectors, and sequenced. Nucleotide sequences of the cloned segments were subjected to the bootscanning-plot analysis using the same sets of reference sequences described in Fig. 3B, and the putative recombination breakpoints were assessed by informative-site analysis. (A) Results for the 2.8-kb and 1.0-kb genome segment clones (2.8c and 1.0c). (B) Results for the 5.2 kb-genome segment clones (5.2c). Red bars indicate the ORF1/ORF2 bordering region. Arrows indicate the putative recombination breakpoints with the highest statistical significance ( $P \leq 0.000001$ ) in the informative-site analysis.



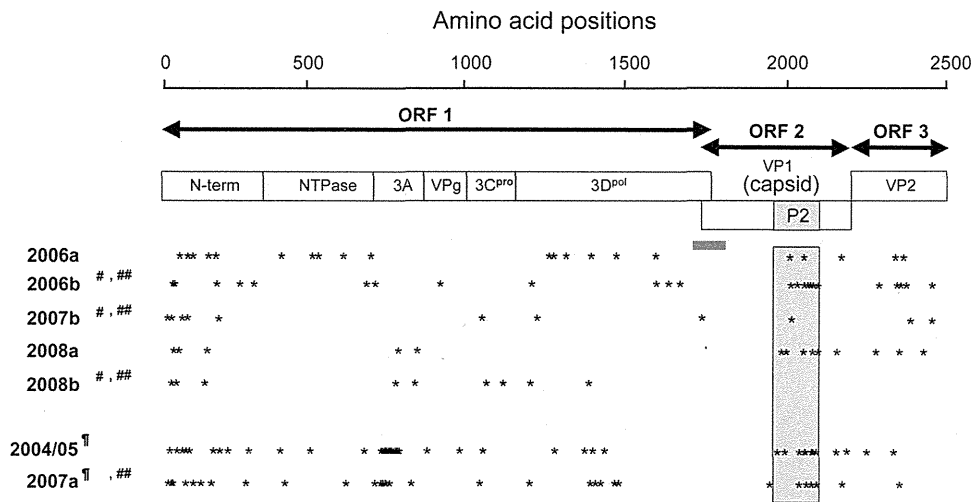


FIG. 5. Amino acid signatures of the NoV GII/4 subtypes. The deduced amino acid sequences of ORF1, ORF2, and ORF3 of a given GII/4 subtype were aligned with the GII/4 sequences identified before the outbreak season of the subtype. Amino acids specific to each subtype at the time of its first outbreak season were extracted and referred to as amino acid signatures of the new epidemic subtype. Asterisks illustrate approximate locations of the amino acid signatures in ORF1, ORF2, and ORF3. A light-blue box denotes approximate locations of the capsid P2 domain in ORF2. A red bar indicates the ORF1/2 boundary region where the single putative recombination breakpoint was assigned for each subtype genome by informative-site analyses (50). †, ORF1s were similar to those for GII/12 (see Fig. 4, ORF2). 2004/05 and 2007a had 27 and 63 amino acid substitutions, respectively, in ORF1s compared to the two available complete ORF1 sequences of GII/12 (accession numbers AB045603 and AB039775). #, ORF2s were classified as the same phylogenetic group (see Fig. 3A, ORF2). ##, ORF3s were classified as the same phylogenetic group (see Fig. 3A, ORF3).

The amino acid signatures of the 7 GII/4 subtypes were distributed throughout the three ORFs (Fig. 5, asterisks). 2004/5 and 2007a had more substitutions in ORF1 than the others because their ORF1s seemed to have originated with the GII/12 relatives (Fig. 3A and 4). When they were compared with the two available complete ORF1 sequences of GII/12, they still had many amino acid substitutions (27 and 63 for 2004/05 and 2007a, respectively). 2007b and 2008b had fewer substitutions in ORF2s and ORF3s than the others because these regions seemed to have originated from the 2006b relatives.

As seen in the 2006b variants in the 2006/2007 season (38), the capsid protein signatures were preferentially distributed on the P2 domain in other GII/4 subtypes (Fig. 5, blue box). All 7 capsid signatures identified in the 2006b variants in the 2006/2007 season were highly conserved during the 2006/2007 season, although two of them (P357 and N412) were gradually lost in the 2006b variant population during 2007 and 2009. Instead, other amino acid substitutions were sporadically accumulated in the P2 domain of the later 2006b variants (data not shown). The 7 signatures in the P2 domain were also well retained in the 2007b and 2008b subtypes, whose genomes had capsid gene segments from the 2006b relatives (Fig. 3B). These data indicate that (i) all of the 7 GII/4 subtypes had unique amino acid substitutions in viral capsid and replication proteins at the time of their outbreaks in Japan, (ii) the dominant 2006b subtype retained the capsid signatures during its persistence between 2006 and 2009, and (iii) some GII/4 subtypes acquired unique mutation sets of the 2006b capsid P2 domain by putative genome recombination events.

**3-D locations of the subtype-specific amino acids in the capsid P domain dimer.** To clarify 3-D locations of the capsid signatures, we constructed structure models of the VP1 P-

domain dimer of the GII/4 subtypes by the homology modeling method as described previously (38). The 2007b and 2008b models were not included for the study because their capsid proteins had no signature or a single signature in the P2 domain due to putative genome recombination with 2006b (Fig. 3, 4, and 5). The thermodynamically and sterically optimized structural models of the P-domain dimer of the 2004/05, 2006a, 2006b, 2007a, and 2008a subtypes showed no major differences in the folding of the main chains (Fig. 6). This result suggests that the capsid amino acid substitutions primarily influenced physicochemical properties around the substitution sites by changing the size and chemical properties of the side chains. These models were then used to map the 3-D locations of the P2 domain mutations.

Importantly, the capsid P2 domain signatures were mostly mapped on the outer surface loops in all of the GII/4 subtypes examined (Fig. 6). These loops form an accessible protein surface with which host proteins, such as a cellular receptor(s) and antibodies, can directly interact. The P2 mutations are often positioned near the putative functional sites for virus entry into the cells: the fucose ring binding sites formed by the P-domain dimer (8, 13) (blue-dotted ovals) and an RGD motif (60) on the  $\beta$ 2 sheet of the P domain (cyan chain). Notably, the P2 mutations were more abundant in the widely prevalent subtypes, 2004/05 and 2006b, than in the others (Fig. 2 and 6). The 2008a subtype, which was detected most recently and caused NoV epidemics at multiple sites in the 2008/2009 season, also had 5 unique substitutions in the P2 domain (Fig. 2 and 6). These data indicate that the GII/4 subtypes that were dominant between 2004 and 2009 in Japan had a greater number of unique amino acids preferentially positioned on their capsid surfaces at the time of their first outbreaks.

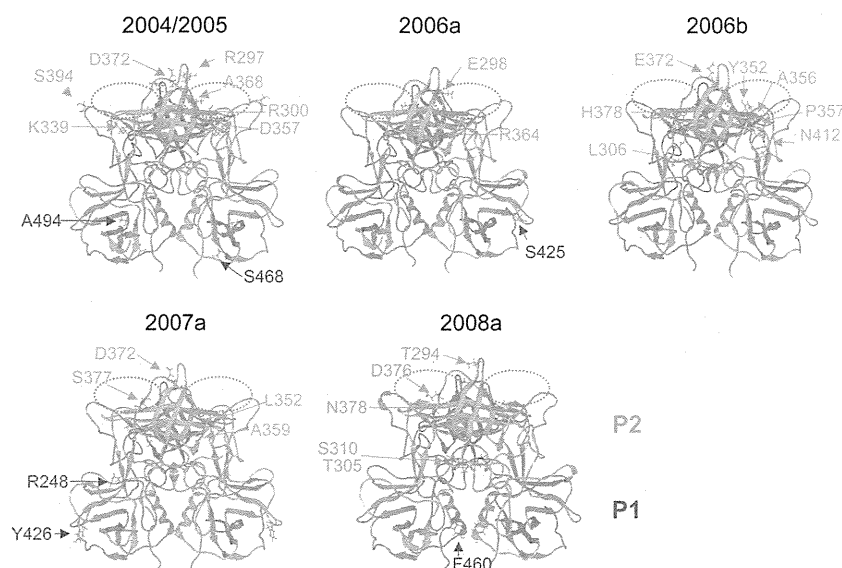


FIG. 6. 3-D locations of the subtype-specific amino acids in the capsid P domain dimer. Structural models of the capsid P domain dimers of recent NoV subtypes were constructed by homology modeling as described previously (38). The 2007b and 2008b capsid models were not included because their ORF2s were classified as belonging to the same phylogenetic group as ORF2 of 2006b due to putative genome recombination (Fig. 3 and 4), and their capsid proteins had no signature or only a single signature in the P2 domain with 2006b (Fig. 5). Orange arrows and letters indicate locations and types of the unique amino acids in each GII/4 subtype at the time of its first outbreak season. Putative functional sites for virus entry into the cells are highlighted. Blue-dotted ovals, the fucose ring binding sites formed by the P domain dimer (8, 13); cyan chain, an RGD motif (60) on the  $\beta 2$  sheet of the P domain.

## DISCUSSION

In this study, we have examined the possible involvement of genome recombination in the generation of new outbreaks of the NoV GII/4 variants. We first analyzed the evolutionary lineage of the GII/4 variants that were present in Japan during 2006 to 2009 and clarified their temporal and geographical distribution. We showed the following: (i) that at least 7 monophyletic GII/4 subtypes were present in humans during the 3-year period, (ii) that 3 to 4 subtypes were cocirculated in each NoV season, and (iii) that the 2006b subtype had spread and persisted more effectively in human communities than the other GII/4 subtypes during the study period (Fig. 1 and 2). These and other findings on the recent GII/4 subtypes (32, 38, 53, 54) consistently support the notion that the 2006b subtype had some selective advantages over the other GII/4 subtypes, which allowed it to quickly spread throughout human communities at the time of its initial appearance in the early winter of 2006. Our study additionally suggests that the possible advantages of the 2006b subtype remained effective over the subsequent 2 years in Japan. However, we could not obtain evidence for genome mosaicism with the 2006b subtype using the available sequence data set of the NoV genogroups, genotypes, and subtypes reported to date (2, 7, 25, 38, 53, 65, 71). Therefore, it is not clear whether genome recombination played a significant role in the generation of new large-scale outbreaks. Because information on the entire genome sequences of NoV is very limited at present, further genome study of NoVs is necessary to clarify this issue.

We then analyzed the GII/4 subtypes to determine whether they showed genome mosaicisms. We demonstrated clearly and for the first time that intersubtype genome recombination is common in the new NoV GII/4 outbreaks: 4 of the 7 new

GII/4 subtypes (2007a, 2007b, 2008a, and 2008b) were mosaics comprised primarily of sequences of the recently prevalent distinct GII/4 subtypes, and 1 (2004/05) was made up of GII/4 and GII/12 genotypes (Fig. 3 and 4). Because of the genome mosaicism, the number of monophyletic clusters of the new variants in the phylogenetic trees differed depending on the region of the genome studied; the numbers of clusters were 7, 7, 5, and 4 for the near-full-length genome, ORF1, ORF2, and ORF3, respectively. Sequences of 11 randomly selected genome segment clones all exhibited bootscanning-plot profiles identical to those obtained with the direct sequencing data except in one case, suggesting dominance of the specific mosaic genomes in the corresponding stool specimens. Phylogenetic-tree, bootscanning-plot, and informative-site analyses consistently provided the same conclusions in regard to genome mosaicism: these virus genomes encoded capsid proteins whose evolutionary lineages were distinct from those of nonstructural proteins. The good agreement of the results by the two sequencing strategies and by three evolutionary methods strongly suggests that the mosaic genomes made from multiple GII/4 subtypes were indeed constantly arising *in vivo* and became the dominant species in infected individuals in some of the NoV GII/4 outbreaks.

We failed to find evidence for the genome mosaicism of the 2006a and 2006b subtypes by using available NoV sequences. These subtypes are unlikely to be the intergenotype or intergenogroup recombinants, because their ORF1, ORF2, and ORF3 sequences constantly showed the strong monophyly within the GII/4 cluster out of the other genotypes and genogroups at their first (38) and successive outbreaks (Fig. 3A). However, the possibility of intersubtype recombination among as-yet-defined classes of GII/4 subtypes remains undetermined for the 2006a and 2006b subtypes.

Interestingly, the mosaic genomes that caused the new NoV GII/4 outbreaks all had the putative recombination breakpoints with the highest statistical significance in the ORF1/2 boundary region ( $P \leq 0.000001$ ) (Fig. 3B and 4). This breakpoint location is consistent with previous reports on intergenogroup and intergenotype recombination (1, 10, 11, 17, 21, 22, 25, 40, 41, 44–46, 49, 57, 63, 64, 66), suggesting the presence of a common hot spot for generation and survival of recombinant NoVs in nature. To a lesser extent, a putative recombination event around the ORF2/3 boundary was identified in 1 of 7 new variant subgroups (Fig. 4A, 2007a). A recombination event around the ORF2/3 junction has also been reported for GII/4 variants circulating in Cairo, Egypt, between 2006 and 2007 (24). The ORF1/2 boundary region is highly conserved in NoV GII/4, as shown previously by very low scores of Shannon entropy within the reported GII/4 sequences (38). This and our present findings on the presence of the putative parent GII/4 sequences of the mosaic genomes suggest that the ORF1/2 mosaic genomes we identified were generated by homologous recombination, as seen in other single-stranded, positive-sense RNA viruses, including poliovirus (20), foot-and-mouth disease virus (36), brome mosaic virus (9, 39), turnip crinkle virus (70), and tomato ringspot virus (52). If this were the case, the intersubtype recombination at the ORF1/2 boundary region would occur and generate variable recombinant viruses *in vivo* more frequently than the intergenotype and intergenogroup recombination would, because the boundary region and neighboring sequences are more similar within the NoV subtype than within the genotype and genogroup. Our results are consistent with this possibility.

The presence of putative recombination at the ORF1/2 boundary region has a direct impact on the modes of NoV subtype evolution *in vivo*. First, the presence of the breakpoint at this region drives independent evolution of ORF1 and ORF2/3 nucleotide sequences and thus of nonstructural and capsid proteins (Fig. 3), leading to divergent evolution of the NoV GII/4 genome (Fig. 1). Second, the presence of the breakpoint allows concurrent acquisition of new mutation sets that arise independently in ORF1 and ORF2/3 among distinct GII/4 subtypes. However, further study is necessary to clarify whether the genome recombination indeed confers any fitness advantage to the virus within a mixed NoV variant population in nature.

The high levels of sequence homology of the ORF1/2 boundary region (38) also suggest that the region is functionally and/or structurally very important for NoV replication and receives strong selective constraints against diversity for NoV survival in nature. Consistently, this region is indicated to contain an important functional motif that regulates capsid expression from a full-length genome in bovine NoV (37). Thus, the ORF1/2 boundary region may be a multifunctional region critical for both replication and evolution of NoVs.

The relatively high detection frequency of the ORF1/2 mosaic genomes in the new GII/4 subtypes (5 of 7) was rather unexpected, because multiple factors, such as retention of virion stability, viral infectivity, and viral replication capabilities in human cells, should restrict the generation of viable hybrid viruses. The present findings therefore raise the possibility of large-scale coinfections by distinct lineage groups of NoVs and of natural selection for the particular ORF1/2 hy-

brid viruses. The former possibility remains to be clarified but is feasible (57) if one considers the high stability of the NoV virion outside the host, as well as NoV transmission modes, i.e., ingestion of contaminated food and water, direct person-to-person contact, and exposure to contaminated airborne vomitus droplets in a semiclosed community (15).

The latter possibility of natural selection also remains to be clarified. However, it is possible that some of the unique mutations identified in each ORF1/2 hybrid genome at the time of their outbreaks (Fig. 5) may be involved in the survival of the hybrid viruses. In this regard, it is noteworthy that the hybrid viruses had multiple mutations in the N-term, NTPase, 3A-like, Vpg, 3C<sup>pro</sup>, and 3D<sup>pol</sup> proteins. These proteins are likely to function primarily in NoV replication in host cells (19). Therefore, acquisition of an appropriate mutation set in ORF1 might confer some advantages in replication of the hybrid viruses in particular hosts. It should also be noted that the 2007b and 2008b subtypes encoded the VP1 and VP2 proteins from 2006b (Fig. 3 and 4). VP1 plays critical roles in binding to the putative infection receptors (8, 60, 61) and antibody neutralization (33, 34). The VP2 protein is also essential for the production of infectious virions in caliciviruses (56). Therefore, acquisition of an appropriate mutation set in ORF2 and ORF3 of 2006b might confer some advantages in infection and/or immune escape of the hybrid viruses at some outbreaks.

Computer-assisted modeling studies provide a structural basis for addressing the potential selective advantages of the capsids of the new GII/4 subtypes. We showed that unique capsid amino acids of the 7 GII/4 subtypes identified in this study were preferentially positioned on the outer surface loops of the protruding P2 domain and were more abundant in the dominant subtypes (Fig. 6). This is also a common characteristic of the past epidemic GII/4 subtypes (32, 38, 53). These findings suggest that physicochemical changes in the capsid surface are a prerequisite for effective virus spread of NoV GII/4 in humans. The specific mutations around the outer surface loops of the protruding P2 domain can modulate the local electrostatic environment and shape of the exposed capsid surface by changing the chemical properties and the size of side chains, respectively. Therefore, acquisition of an adequate set of capsid P2 mutations might be able to decrease antibody affinity without decreasing affinity to the infection receptors of GII/4. This would confer an advantage to the variants that would allow them to spread in human communities in the presence of immunity against precirculated variants. To effectively gain such a set of P2 domain mutations, as well as those of nonstructural proteins, genetic recombination around the ORF1/2 boundary region may be an ideal mechanism. Establishment of a tissue culture system to support effective replication of human NoVs, as well as a reverse genetics system to study the roles of mutations in NoV infection and replication, will be critical to clarify each of these possibilities.

It should be noted that despite the prolonged dominance of the 2006b subtype, the magnitudes of NoV epidemics in Japan have gradually declined since 2007: the total numbers of reported NoV infection cases during October and March of 2007–2008 and 2008–2009 showed more than 2- and 5-fold decreases, respectively, compared with the same period in 2006–2007 under the same surveillance system (Infectious Disease Surveillance Center [<http://idsc.nih.gov.jp/iasr/index>]).

.html)). These observations may imply that biological niches within human communities that support replication of the 2006b subtype have gradually been shrinking in Japan since 2007. A possible explanation for this phenomenon is that immunity against the 2006b subtype has been gradually strengthened in human populations due to the persistence of the 2006b infections in Japan. Nevertheless, none of the new GII/4 variant subtypes were able to replace the 2006b epidemic in the 2007/2008 and 2008/2009 seasons. In addition, two of the four new putative recombinants (2007b and 2008b), which appeared in the 2007/2008 and 2008/2009 seasons, gained ORF2/3 of 2006b. These observations may imply that 2006b still had some selective advantages over other GII/4 variant subgroups in the 2008/2009 seasons. Further follow-up study is necessary to address these possibilities.

Our findings on genome mosaicism may have an impact on epidemiological and virological studies of NoVs. For example, mosaicism could influence the validity of NoV classification, which is based on the sequences of parts of the NoV genome. Because hybrid viruses that cause epidemics seem to share a recombination breakpoint around the ORF1/2 boundary region, this junction segment may be useful for monitoring the prevalence of hybrid NoVs in nature. The genome mosaicism could also impact measurement of the mutation rates of NoVs in nature: careful selection of the genome segments that contain no recombination breakpoints would be critical to measure the nucleotide substitution rates. Continual accumulation of information on the complete genome sequences of NoVs in natural and living environments will provide genetic bases for dealing with these issues and illustrate mechanisms by which NoV evolves to generate and sustain new epidemics in human populations.

#### ACKNOWLEDGMENTS

We thank T. Shiino for helpful comments and suggestions.

This work was supported by grants for Research on Food Safety from the Ministry of Health, Labor, and Welfare, Japan, and for Research on Publicly Essential Drugs and Medical Devices from the Japan Health Sciences Foundation.

#### REFERENCES

- Ambert-Balay, K., F. Bon, F. Le Guyader, P. Pothier, and E. Kohli. 2005. Characterization of new recombinant noroviruses. *J. Clin. Microbiol.* **43**: 5179–5186.
- Ando, T., J. S. Noel, and R. L. Fankhauser. 2000. Genetic classification of "Norwalk-like viruses." *J. Infect. Dis.* **181**(Suppl. 2):S336–S348.
- Baker, D., and A. Sali. 2001. Protein structure prediction and structural genomics. *Science* **294**:93–96.
- Belliot, G., S. V. Sosnovtsev, T. Mitra, C. Hammer, M. Garfield, and K. Y. Green. 2003. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. *J. Virol.* **77**:10957–10974.
- Bertolotti-Ciarlet, A., S. E. Crawford, A. M. Hutson, and M. K. Estes. 2003. The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. *J. Virol.* **77**:11603–11615.
- Bertolotti-Ciarlet, A., L. J. White, R. Chen, B. V. Prasad, and M. K. Estes. 2002. Structural requirements for the assembly of Norwalk virus-like particles. *J. Virol.* **76**:4044–4055.
- Bok, K., E. J. Abente, M. Realpe-Quintero, T. Mitra, S. V. Sosnovtsev, A. Z. Kapikian, and K. Y. Green. 2009. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. *J. Virol.* **83**:11890–11901.
- Bu, W., A. Mamedova, M. Tan, M. Xia, X. Jiang, and R. S. Hegde. 2008. Structural basis for the receptor binding specificity of Norwalk virus. *J. Virol.* **82**:5340–5347.
- Bujarski, J. J., and P. Kaesberg. 1986. Genetic recombination between RNA components of a multipartite plant virus. *Nature* **321**:528–531.
- Bull, R. A., G. S. Hansman, L. E. Clancy, M. M. Tanaka, W. D. Rawlinson, and P. A. White. 2005. Norovirus recombination in ORF1/ORF2 overlap. *Emerg. Infect. Dis.* **11**:1079–1085.
- Bull, R. A., M. M. Tanaka, and P. A. White. 2007. Norovirus recombination. *J. Gen. Virol.* **88**:3347–3359.
- Cannon, J. L., L. C. Lindesmith, E. F. Donaldson, L. Saxe, R. S. Baric, and J. Vinje. 2009. Herd immunity to GII.4 noroviruses is supported by outbreak patient sera. *J. Virol.* **83**:5363–5374.
- Cao, S., Z. Lou, M. Tan, Y. Chen, Y. Liu, Z. Zhang, X. C. Zhang, X. Jiang, X. Li, and Z. Rao. 2007. Structural basis for the recognition of blood group trisaccharides by norovirus. *J. Virol.* **81**:5949–5957.
- Choi, J. M., A. M. Hutson, M. K. Estes, and B. V. Prasad. 2008. Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. *Proc. Natl. Acad. Sci. U. S. A.* **105**:9175–9180.
- Estes, M. K., B. V. Prasad, and R. L. Atmar. 2006. Noroviruses everywhere: has something changed? *Curr. Opin. Infect. Dis.* **19**:467–474.
- Etherington, G. J., J. Dicks, and I. N. Roberts. 2006. High throughput sequence analysis reveals hitherto unreported recombination in the genus Norovirus. *Virology* **345**:88–95.
- Fukuda, S., Y. Sasaki, S. Takao, and M. Seno. 2008. Recombinant norovirus implicated in gastroenteritis outbreaks in Hiroshima Prefecture, Japan. *J. Med. Virol.* **80**:921–928.
- Glass, P. J., L. J. White, J. M. Ball, I. Leparco-Goffart, M. E. Hardy, and M. K. Estes. 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* **74**:6581–6591.
- Hyde, J. L., S. V. Sosnovtsev, K. Y. Green, C. Wobus, H. W. Virgin, and J. M. Mackenzie. 2009. Mouse norovirus replication is associated with virus-induced vesicle clusters originating from membranes derived from the secretory pathway. *J. Virol.* **83**:9709–9719.
- Jarvis, T. C., and K. Kirkegaard. 1992. Poliovirus RNA recombination: mechanistic studies in the absence of selection. *EMBO J.* **11**:3135–3145.
- Jiang, X., C. Espul, W. M. Zhong, H. Cuello, and D. O. Matson. 1999. Characterization of a novel human calicivirus that may be a naturally occurring recombinant. *Arch. Virol.* **144**:2377–2387.
- Jin, M., H. P. Xie, Z. J. Duan, N. Liu, Q. Zhang, B. S. Wu, H. Y. Li, W. X. Cheng, S. H. Yang, J. M. Yu, Z. Q. Xu, S. X. Cui, L. Zhu, M. Tan, X. Jiang, and Z. Y. Fang. 2008. Emergence of the GII.4/2006b variant and recombinant noroviruses in China. *J. Med. Virol.* **80**:1997–2004.
- Kageyama, T., S. Kojima, M. Shinohara, K. Uchida, S. Fukushi, F. B. Hoshino, N. Takeda, and K. Katayama. 2003. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *J. Clin. Microbiol.* **41**:1548–1557.
- Kamel, A. H., M. A. Ali, H. G. El-Nady, A. de Rougemont, P. Pothier, and G. Belliot. 2009. Predominance and circulation of enteric viruses in the region of Greater Cairo, Egypt. *J. Clin. Microbiol.* **47**:1037–1045.
- Katayama, K., H. Shirato-Horikoshi, S. Kojima, T. Kageyama, T. Oka, F. Hoshino, S. Fukushi, M. Shinohara, K. Uchida, Y. Suzuki, T. Gojohori, and N. Takeda. 2002. Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**:225–239.
- Katoh, K., G. Asimenos, and H. Toh. 2009. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**:39–64.
- Katpally, U., C. E. Wobus, K. Dryden, H. W. Virgin IV, and T. J. Smith. 2008. Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. *J. Virol.* **82**:2079–2088.
- Khamrin, P., S. Takanashi, W. Chan-It, M. Kobayashi, S. Nishimura, N. Katsumata, S. Okitsu, N. Maneekarn, O. Nishio, and H. Ushijima. 2009. Immunochromatography test for rapid detection of norovirus in fecal specimens. *J. Virol. Methods* **157**:219–222.
- Kojima, S., T. Kageyama, S. Fukushi, F. B. Hoshino, M. Shinohara, K. Uchida, K. Natori, N. Takeda, and K. Katayama. 2002. Genogroup-specific PCR primers for detection of Norwalk-like viruses. *J. Virol. Methods* **100**: 107–114.
- Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**:1789–1796.
- Lanave, C., G. Preparata, C. Saccone, and G. Serio. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**:86–93.
- Lindesmith, L. C., E. F. Donaldson, A. D. Lobue, J. L. Cannon, D. P. Zheng, J. Vinje, and R. S. Baric. 2008. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* **5**:e31.
- Lochridge, V. P., and M. E. Hardy. 2007. A single-amino-acid substitution in the P2 domain of VP1 of murine norovirus is sufficient for escape from antibody neutralization. *J. Virol.* **81**:12316–12322.
- Lochridge, V. P., K. L. Jutila, J. W. Graff, and M. E. Hardy. 2005. Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions. *J. Gen. Virol.* **86**:2799–2806.
- Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**:260–262.
- McCahon, D., and W. R. Slade. 1981. A sensitive method for the detection and isolation of recombinants of foot-and-mouth disease virus. *J. Gen. Virol.* **53**:333–342.
- McCormick, C. J., O. Salim, P. R. Lambden, and I. N. Clarke. 2008. Trans-