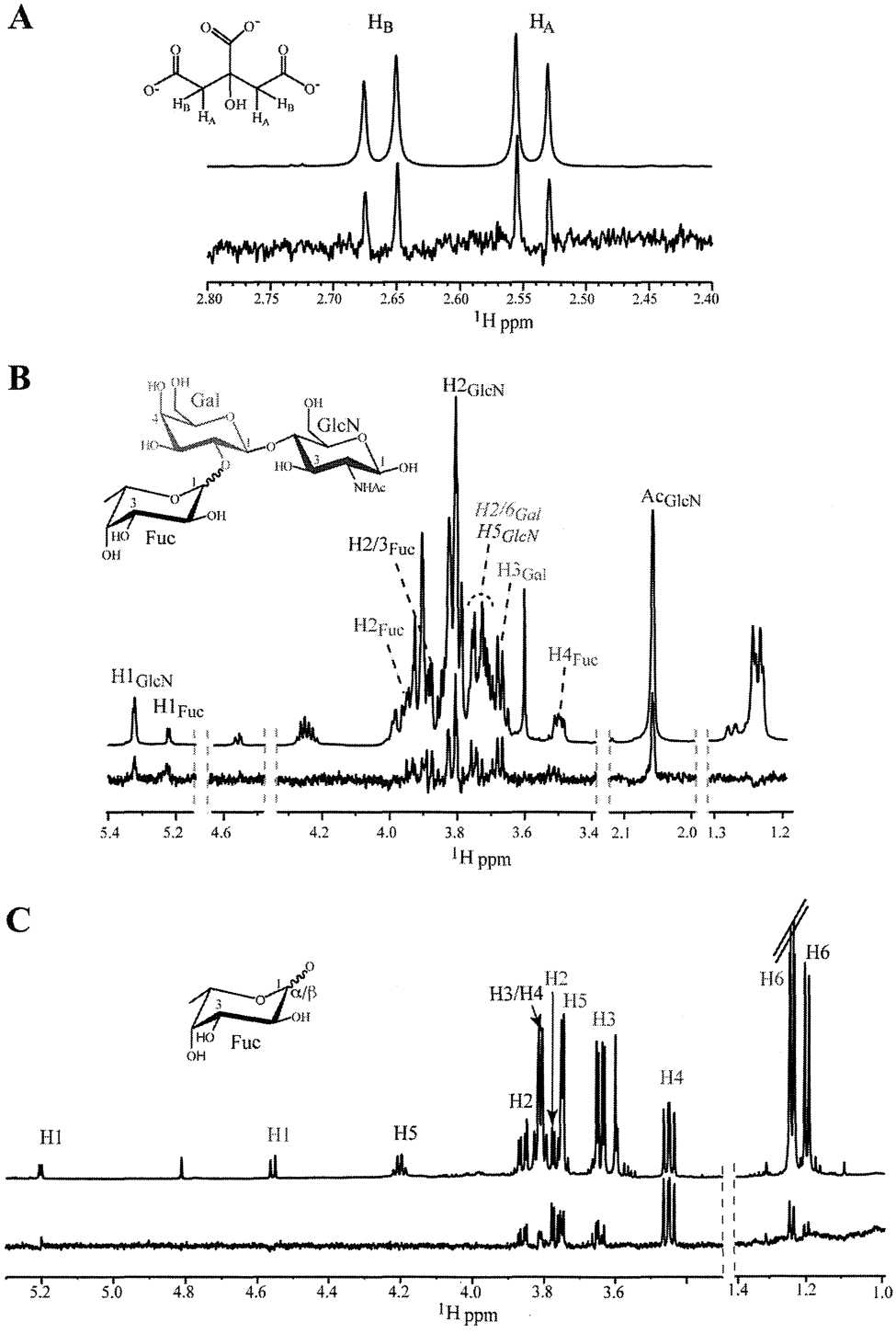

FIG 3 STD NMR spectra for citrate, H type 2 trisaccharide, and L-fucose bound to the GII.10 P domain. STD (lower) and reference (upper) spectra of (A) citrate (1.2 mM), (B) H type 2 trisaccharide (1.2 mM), and (C) L-fucose (mixture of α and β anomers) (1.2 mM) in the presence of the GII.10 P domain (15 μM). Nonoverlapping protons that exhibit STD enhancements are labeled and color coded by sugar residue, and signals for β-Fuc are red. One group of overlapping signals appears in italics.

world and as such the most well studied. Most studies agreed that a dominant GII.4 norovirus was replaced the following year or next by a new GII.4 "variant" norovirus that had ~5% amino acid change in the capsid gene (6, 9, 10, 30, 31, 47). The reason that the GII.4 variants dominated and not some other genotype was unknown, but studies have shown specific mutations at or surrounding the HBGA binding site were capable of altering the HBGA binding patterns (15, 30, 31, 52). These small changes were thought to lead to new GII.4 variants capable of causing pandemics, analogous to influenza A virus evolution (14, 29). Despite

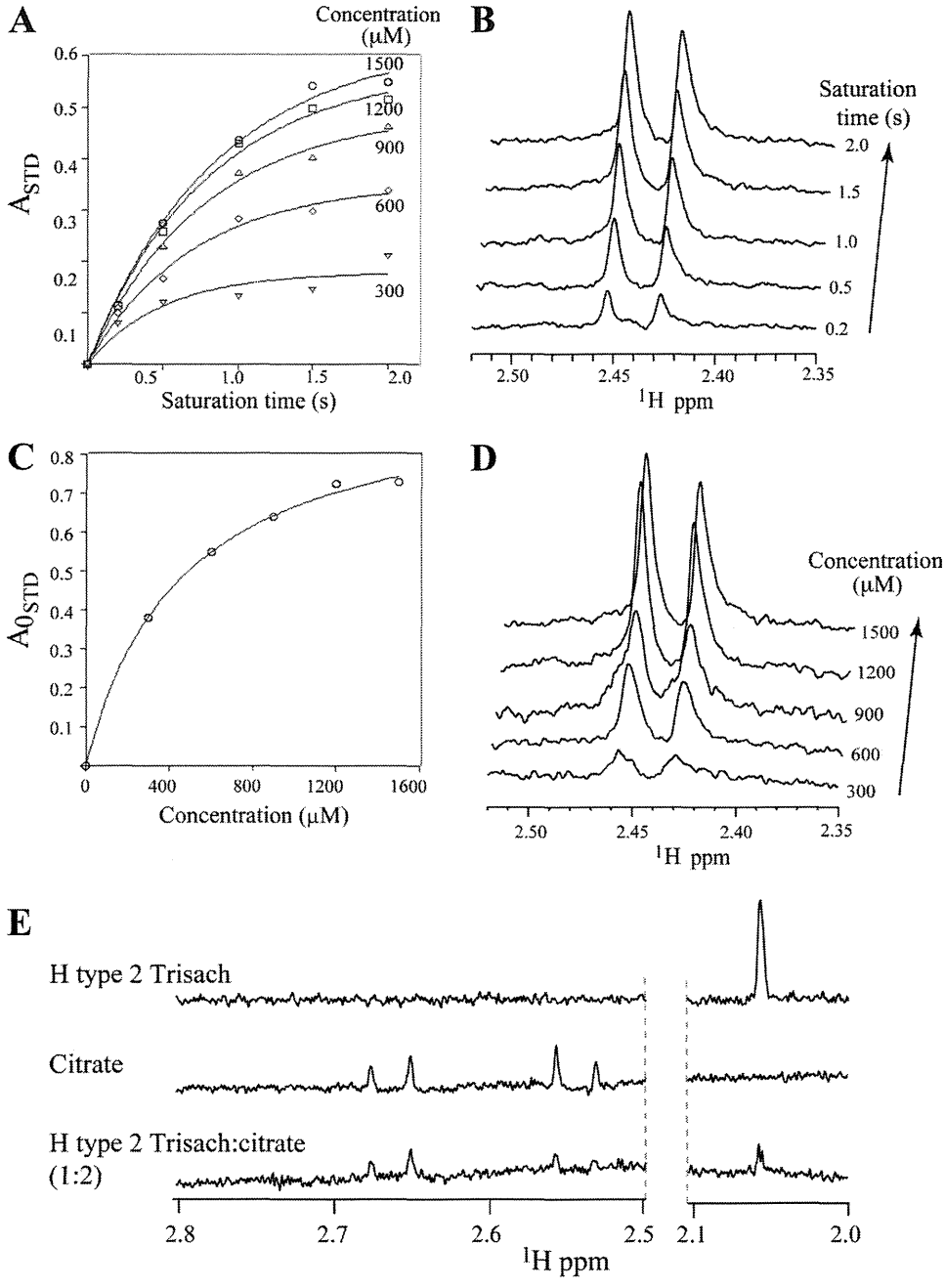

FIG 4 Binding affinity of citrate and HBGAs to GII.10 P domain by STD NMR. Data were used to obtain the K_D for citrate and H type 2 trisaccharide (Trisach) binding to GII.10 P by single-ligand titration STD NMR experiments (2). (A) Effect on STD enhancement (expressed as A_{STD}) (34, 37) as a function of saturation time (t_{sat}) and ligand concentration; (B) stacked plots of spectra for 1.5 mM citrate as a function of t_{sat} (y axis); (C) Langmuir binding curve used to obtain the K_D from the initial slope of A_{STD} as a function of citrate concentration. (D) Stacked plot of various citrate concentrations (t_{sat}, 2 s, 15 μM protein); (E) competition STD spectra of H type 2 trisaccharide (top), citrate (middle), and 1:2 H type 2 trisaccharide-citrate (0.75:1.5 mM; bottom) used to calculate the K_i of citrate (36).

these amino acid changes, few if any occurred at the fucose-binding site, thus highlighting the common site of vulnerability for GII noroviruses, especially for the pandemic GII.4 variant noroviruses. It is not known if the GI noroviruses will bind citrate given that the GI and GII P domain interactions with HBGAs were different, but since GI.1 P domain interacted with α-fucose1-2 and it was reported that the HBGA binding site was conserved among GI noroviruses (12), we suspect that GI noroviruses may also bind citrate, although further structural studies are needed.

Our unexpected finding that citrate and fucose have similar binding modes to the norovirus GII.10 P domain raises the question of whether such citrate mimicry of monosaccharide binding could be a general phenomenon or whether it is spe-

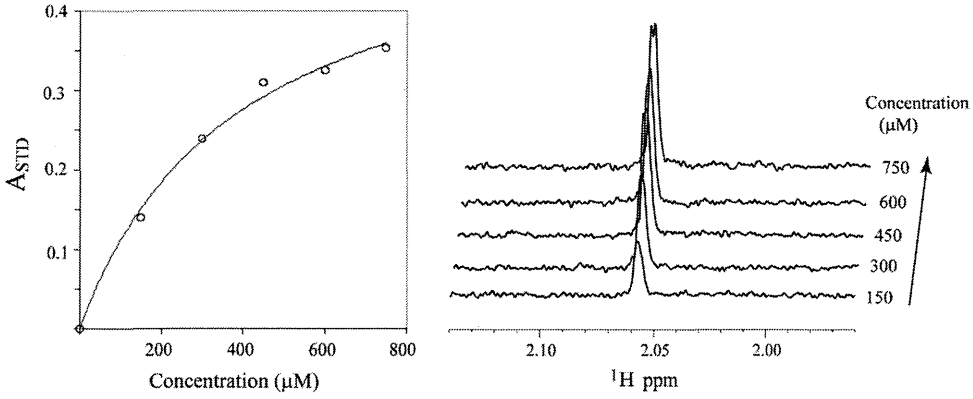

FIG 5 Binding affinity of H type 2 trisaccharide to the GII.10 P domain (left) effect on STD enhancement, expressed as A_{STD}, as a function of trisaccharide concentration in the presence of 15 μM GII.10 P domain. t_{sat} = 0.1 s. Curve fitting (described in the text) provides the K_D value. (Right) Stacked STD NMR spectra showing the change in enhancement of the nonoverlapped *N*-acetyl proton signals as a function of increasing concentration of H type 2 trisaccharide (40).

cific to norovirus and other caliciviruses. To investigate this, we performed *in silico* docking studies of citrate against four different fucose-binding proteins (*Anguilla anguilla* agglutinin, *Aleuria aurantia* lectin, *Streptococcus pneumoniae* virulence factor SpGH98, and *Pseudomonas aeruginosa* PA-IIL lectin) and two other saccharide-binding proteins (parainfluenza virus 5 hemagglutinin-neuraminidase and porcine adenovirus type 4 galectin domain), for which fucose or other saccharide-bound crystal structures were available (see Table S1 in the supplemental material). Computational docking analyses reveal different levels of citrate mimicry of monosaccharide binding for other saccharide-binding proteins. For *Anguilla anguilla* agglutinin, citrate, in its predicted binding pose, overlapped with the C-5, C-4, C-3, O-5, O-4, and O-3 atoms of fucose in a similar way to what was observed in the GII.10 P domain (Table S1), while forming hydrogen bonds with the same sets of protein residues as fucose (see Fig. S4 in the supplemental material). Citrate was thus predicted to show a high degree of mimicry to fucose, similarly to our experimental findings for the GII.10 P domain. For the other three fucose-binding proteins, citrate, in its predicted binding poses, did not overlap with the cocrystallized fucose, although it still formed the same sets of polar interactions as the cocrystallized fucose (see Fig. S5 to S7 in the supplemental material). Hence, our docking studies suggest that the mimicry between citrate and fucose binding observed for the GII.10 P domain could be a common, although not universal, phenomenon across other fucose-binding proteins. For all six fucose- and other saccharide-bound proteins for which docking was performed, the predicted citrate binding poses were able to form polar interactions with the same sets of protein residues as the cocrystallized ligand see (Fig. S4 to S9 in the supplemental material), indicating that citrate might be generally useful as a scaffold for designing glycomimetic inhibitors against these and other saccharide-interacting pathogens. Furthermore, a search of the ZINC database (4) revealed that there are more than three thousand compounds with at least 50% similarity to citrate. Thus, *in silico* screening of this database may present a promising approach for identifying small molecules that bind to saccharide-binding proteins. We note, however, that the predicted binding pose of citrate docked to fucose-bound GII.10 P domain had a root mean square deviation (RMSD) of 3.60 Å, while the predicted binding pose of citrate docked to citrate-bound GII.10 P domain with the cocrystallized water molecule had an RMSD of 1.87 Å. This indicates that the resulting docking modes could be error prone. Given that calculating small molecule-receptor binding energies is a difficult and error-prone task (24, 46), ultimately experimental validation would be necessary to confirm the generality of the citrate-saccharide mimicry predicted here.

The STD NMR data provided strong evidence that the integrity of the GII.10 P domain remained unchanged in the presence of different concentrations of citrate buffer and since the pH of the citrate buffer remained more or less the same during the titration, a specific effect of citrate was responsible for the reduction in HBGA attachment. Although the K_D values of citrate and H type 2 trisaccharide for the GII.10 P domain are in the range of 360 to 490 μM, these relatively weak affinities are typical for univalent protein-carbohydrate interactions (17, 28). Given that 90 copies of dimeric P domains are present on norovirus capsid, it is plausible that a multivalent version of citrate- or fucose-like ligands would greatly enhance affinities and provide a starting point for norovirus inhibitors. Indeed, Rademacher et al. show that multivalent fucose-like compounds have increased avidity over their univalent counterparts (44).

In conclusion, we have described the structural basis by which citrate binds to the HBGA binding site of the norovirus GII.10 P domain and can in turn inhibit HBGA binding. Natural compounds, such as juice from lemons and limes, which contain ~300 mM citric acid (42), may already reduce or inhibit norovirus infections, as suggested by a number of recent studies (23, 48–50, 54). In regard to this, it is tempting to speculate that a few drops of lemon juice with one's oysters might reduce norovirus infection. Epidemiological studies on the ingestion of foods high in citrate and norovirus infection may be illuminating, as may be correlations with related glycomimetics—e.g., with ascorbic acid (vitamin C). Controlled possibly volunteer studies should also provide an accurate assessment of norovirus inhibition. Additional compound screening will likely be required to identify a universal norovirus inhibitor with high potency and broad reactivity, and the structural basis for norovirus interaction with citrate as revealed here may be helpful in such efforts.

ACKNOWLEDGMENTS

We thank members of the Structural Biology Section and Structural Bioinformatics Core at the NIH Vaccine Research Center for comments on the manuscript and J. Stuckey for assistance with figures.

G.S.H., S.H., K.K., C.A.B., and P.D.K. designed the research, G.S.H., S.H., J.S.M., and T.S. performed the research, G.S.H., S.H., J.S.M., G.-Y.C., I.G., C.A.B., and P.D.K. analyzed the data, and G.S.H., S.H., G.-Y.C., I.G., C.A.B., and P.D.K. wrote the paper, on which all authors commented.

Support for this work was provided by the Intramural Research Program of the National Institutes of Health (NIDDK, C.A.B; NIAID, P.D.K); the Intramural AIDS Targeted Antiviral Program, Office of the Director, NIH (C.A.B.); a grant from The Japan Health Science Foundation; and by grants from the Ministry of Health, Labor, and Welfare of Japan. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-Eng-38.

REFERENCES

1. **Adams PD, et al.** 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. Biol. Crystallogr. **66**:213–221.
2. **Angulo J, Enriquez-Navas PM, Nieto PM.** 2010. Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. Chemistry **16**:7803–7812.
3. **Ausar SF, Foubert TR, Hudson MH, Vedvick TS, Middaugh CR.** 2006. Conformational stability and disassembly of Norwalk virus-like particles. Effect of pH and temperature. J. Biol. Chem. **281**:19478–19488.
4. **Berman HM, et al.** 2000. The Protein Data Bank. Nucleic Acids Res. **28**:235–242.
5. **Bianchet MA, Odom EW, Vasta GR, Amzel LM.** 2002. A novel fucose recognition fold involved in innate immunity. Nat. Struct. Biol. **9**:628–634.
6. **Bok K, et al.** 2009. Evolutionary dynamics of GII. 4 noroviruses over a 34-year period. J. Virol. **83**:11890–11901.
7. **Boraston AB, Wang D, Burke RD.** 2006. Blood group antigen recognition by a Streptococcus pneumoniae virulence factor. J. Biol. Chem. **281**: 35263–35271.
8. **Bu W, et al.** 2008. Structural basis for the receptor binding specificity of Norwalk virus. J. Virol. **82**:5340–5347.
9. **Bull RA, Tu ET, McIver CJ, Rawlinson WD, White PA.** 2006. Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis. J. Clin. Microbiol. **44**:327–333.
10. **Bull RA, White PA.** 2011. Mechanisms of GII.4 norovirus evolution. Trends Microbiol. **19**:233–240.
11. **Cao S, et al.** 2007. Structural basis for the recognition of blood group trisaccharides by norovirus. J. Virol. **81**:5949–5957.
12. **Choi JM, Hutson AM, Estes MK, Prasad BV.** 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.
13. **Collaborative Computational Project N.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. **50**:760–763.
14. **De Jong JC, Rimmelzwaan GF, Fouchier RA, Osterhaus AD.** 2000. Influenza virus: a master of metamorphosis. J. Infect. **40**:218–228.
15. **de Rougemont A, et al.** 2011. Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. J. Virol. **85**:4057–4070.
16. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. **66**:486–501.
17. **Fang M, Agha S, Lee R, Culpepper-Morgan J, D'Souza A.** 2000. Perforation of jejunal diverticulum: case report and review of literature. Conn. Med. **64**:7–10.
18. **Feng X, Jiang X.** 2006. Library screen for inhibitors targeting norovirus binding to their histo-blood group antigen receptors. Antimicrob. Agents Chemother. **51**:324–331.
19. **Fujihashi M, Peapus DH, Kamiya N, Nagata Y, Miki K.** 2003. Crystal structure of fucose-specific lectin from Aleuria aurantia binding ligands at three of its five sugar recognition sites. Biochemistry **42**:11093–11099.
20. **Guardado-Calvo P, et al.** 2010. Crystallographic structure of porcine adenovirus type 4 fiber head and galectin domains. J. Virol. **84**: 10558–10568.
21. **Hansman GS, et al.** 2011. Crystal structures of GII.10 and GII.12 norovirus protruding domains in complex with histo-blood group antigens reveal details for a potential site of vulnerability. J. Virol. **85**:6687–6701.
22. **Hansman GS, et al.** 2004. Detection of norovirus and sapovirus infection among children with gastroenteritis in Ho Chi Minh City, Vietnam. Arch. Virol. **149**:1673–1688.
23. **Horm KM, D'Souza DH.** 2011. Survival of human norovirus surrogates in milk, orange, and pomegranate juice, and juice blends at refrigeration (4 degrees C). Food Microbiol. **28**:1054–1061.
24. **Huey R, Morris GM, Olson AJ, Goodsell DS.** 2007. A semiempirical free energy force field with charge-based desolvation. J. Comput. Chem. **28**: 1145–1152.
25. **Jiang X, Wang M, Graham DY, Estes MK.** 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J. Virol. **66**:6527–6532.
26. **Kageyama T, et al.** 2004. Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to norovirus in Japan. J. Clin. Microbiol. **42**:2988–2995.
27. **Kapikian AZ, et al.** 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J. Virol. **10**:1075–1081.
28. **Kiessling LL, Young T, Gruber TD, Mortell KH.** 2008. Glycoscience 12. **4**:2483–2523.
29. **Levin SA, Dushoff J, Plotkin JB.** 2004. Evolution and persistence of influenza A and other diseases. Math. Biosci. **188**:17–28.
30. **Lindesmith LC, Donaldson EF, Baric RS.** 2011. Norovirus GII.4 strain antigenic variation. J. Virol. **85**:231–242.
31. **Lindesmith LC, et al.** 2008. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med. **5**:e31.
32. **Marionneau S, et al.** 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. Biochimie **83**:565–573.
33. **Marotte K, et al.** 2007. X-ray structures and thermodynamics of the interaction of PA-IIL from Pseudomonas aeruginosa with disaccharide derivatives. Chem. Med. Chem. **2**:1328–1338.
34. **Mayer M, Meyer B.** 2001. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J. Am. Chem. Soc. **123**:6108–6117.
35. **McCoy AJ, et al.** 2007. Phaser crystallographic software. J. Appl. Crystallogr. **40**:658–674.
36. **Meinecke R, Meyer B.** 2001. Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin alphaIIbbeta3. J. Med. Chem. **44**:3059–3065.
37. **Meyer B, Peters T.** 2003. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem. Int. ed. Engl. **42**:864–890.
38. **Mitchell EP, et al.** 2005. High affinity fucose binding of Pseudomonas aeruginosa lectin PA-IIL: 1.0 A resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. Proteins **58**:735–746.
39. **Morris GM, et al.** 1998. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J. Comput. Chem. **19**:1639–1662.
40. **Neffe AT, Bilang M, Meyer B.** 2006. Synthesis and optimization of peptidomimetics as HIV entry inhibitors against the receptor protein CD4 using STD NMR and ligand docking. Org. Biomol. Chem. **4**:3259–3267.
41. **Otwinowski Z, Minor W.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**:307–326.
42. **Penniston KL, Nakada SY, Holmes RP, Assimos DG.** 2008. Quantitative assessment of citric acid in lemon juice, lime juice, and commercially-available fruit juice products. J. Endourol. **22**:567–570.
43. **Prasad BV, et al.** 1999. X-ray crystallographic structure of the Norwalk virus capsid. Science **286**:287–290.
44. **Rademacher C, et al.** 2011. Targeting norovirus infection-multivalent entry inhibitor design based on NMR experiments. Chemistry **17**: 7442–7453.
45. **Rademacher C, Krishna NR, Palcic M, Parra F, Peters T.** 2008. NMR experiments reveal the molecular basis of receptor recognition by a calicivirus. J. Am. Chem. Soc. **130**:3669–3675.

46. **Shoichet BK.** 2004. Virtual screening of chemical libraries. Nature **432**: 862–865.
47. **Siebenga JJ, et al.** 2010. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. PLoS Pathog. **6**:e1000884.
48. **Su X, Howell AB, D'Souza DH.** 2010. Antiviral effects of cranberry juice and cranberry proanthocyanidins on foodborne viral surrogates—a time dependence study in vitro. Food Microbiol. **27**:985–991.
49. **Su X, Howell AB, D'Souza DH.** 2010. The effect of cranberry juice and cranberry proanthocyanidins on the infectivity of human enteric viral surrogates. Food Microbiol. **27**:535–540.
50. **Su X, Sangster MY, D'Souza DH.** 2010. In vitro effects of pomegranate juice and pomegranate polyphenols on foodborne viral surrogates. Foodborne Pathog. Dis. **7**:1473–1479.
51. **Tan M, Hegde RS, Jiang X.** 2004. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. J. Virol. **78**:6233–6242.
52. **Tan M, et al.** 2009. Conservation of carbohydrate binding interfaces: evidence of human HBGA selection in norovirus evolution. PLoS One **4**:e5058.
53. Reference deleted.
54. **Whitehead K, McCue KA.** 2010. Virucidal efficacy of disinfectant actives against feline calicivirus, a surrogate for norovirus, in a short contact time. Am. J. Infect. Control. **38**:26–30.
55. Reference deleted.
56. **Yuan P, et al.** 2005. Structural studies of the parainfluenza virus 5 hemagglutinin-neuraminidase tetramer in complex with its receptor, sialyllactose. Structure **13**:803–815.
57. **Zheng DP, et al.** 2006. Norovirus classification and proposed strain nomenclature. Virology **346**:312–323.

Exp. Anim. 61(1), 35–40, 2012

—Original—

Evaluation of Four Antiseptics Using a Novel Murine Norovirus

Takashi MATSUHIRA[1)], Chizuko KAJI[1)], Shoichi MURAKAMI[1)], Kazunori MAEBASHI[1)], Tomoichiro OKA[2)], Naokazu TAKEDA[3)], and Kazuhiko KATAYAMA[2)]

[1)]Pharmaceutical Research Center, Meiji Seika Pharma Co., Ltd., 760 Morooka-cho, Kohoku-ku, Yokohama 222-8567, [2)]Department of Virology II, National Institute of Infectious Diseases, 4–7–1 Gakuen, Musashi-murayama, Tokyo 208-0011, Japan, and [3)]Section of Viral Infections, Thailand-Japan Research Collaboration Center on Emerging and Re-Emerging Infections (RCC-ERI), 6th Fl., Building 10, Department of Medical Sciences, Ministry of Publlic Health, Tiwanon Rd., Muang, Nonthaburi 11000, Thailand

Abstract: We isolated a novel murine norovirus (MNV), MT30-2 strain, from feces of conventional mice in Japan to evaluate the virucidal activity of four antiseptics. The MNV MT30-2 strain was inactivated by as little as 0.2% (w/v) povidone-iodine (PVP-I) and 0.1% (w/v) sodium hypochlorite (NaOCl) treatment as determined by a novel plaque assay. Importantly, PVP-I reduced the MNV titer by 4 $\log_{10}$ within 15 s of exposure. The other two antiseptics, benzethonium chloride (BEC) and chlorhexidine gluconate (CHG), did not reduce the MNV titer even when treatment lasted for 60 s. When the virus titer was reduced by PVP-I or NaOCl treatment, the amount of MNV RNA was not reduced, indicating that the presence of viral RNA was not related to the virucidal activity of the antiseptics. PVP-I and NaOCl will be useful in controlling the spread of MNV, which is a common problem in mice colonies. In this study, we isolated a novel MNV and newly revealed that two antiseptics (PVP-I and NaOCl) were able to inactivate MNV at low concentrations and in a short contact time.

Key words: antiseptics, murine norovirus, plaque assay, quantitative RT-PCR

Introduction

Noroviruses belong to a group of single-stranded RNA, nonenveloped viruses in the Caliciviridae family that are widely transmitted among humans or animals. At present, five major genogroups (GI-GV) of noroviruses have been designated, with GI, GII, and GIV infecting humans, GIII infecting bovines, and GV infecting murines [17]. Murine norovirus (MNV), prototype MNV-1, was first identified in 2003 in immunocompromised mice lacking recombination-activating gene 2 (RAG2), signal transducer, and activator of transcription 1 (STAT1) (RAG2/STAT1$^{-/-}$) [9]. Human norovirus (HuNoV) cannot grow in cell culture [5]. On the other hand, MNV is the only norovirus capable of replicating in both cell culture and small animals [17, 18]. Infection of MNV in a normal mouse shows no visible signs, but MNV can induce a clinical disease with a high mortality in congenitally immunodeficient mice [13]. The major infection route of MNV is a fecal-oral route, and shedding of MNV continues for at least eight weeks post infection [6–8]. Accordingly, MNV is one of the most

(Received 27 April 2011 / Accepted 25 July 2011)

Address corresponding: T. Matsuhira, Pharmaceutical Research Center, Meiji Seika Pharma Co., Ltd., 760 Morooka-cho, Kohoku-ku, Yokohama 222-8567, Japan

transmittable and prevalent pathogens in laboratory mice today [6, 8, 12], making the prevention of the spread of MNV an important issue.

In a previous study using the MNV CW1 strain [18], the antiseptic effect of ethanol (EtOH) against MNV was examined [1], and 60% (v/v) EtOH was shown to reduce the MNV infectious titer by 4 $\log_{10}$ within 30 s of exposure; however, 30% (v/v) EtOH did not reduce the titer. Moreover, 1% (w/v) povidone-iodine (PVP-I) and 0.26% (w/v) sodium hypochlorite (NaOCl) have been shown to inactivate the MNV CW1 strain, although the efficacies of these antiseptics against other MNV strains are unknown.

In this study, to clarify the virucidal activity of antiseptics against another MNV strain, we isolated a novel MNV, strain MT30-2, from conventional mice in Japan. Using the MT30-2 strain, the virucidal activity of PVP-I was examined. PVP-I has been used as an external treatment for humans and has a broad spectrum of virucidal activity against enveloped and nonenveloped DNA and RNA viruses [10]. We compared the virucidal activity of PVP-I with those of NaOCl, benzethonium chloride (BEC), and chlorhexidine gluconate (CHG) using plaque assays. Furthermore, we quantified MNV MT30-2 RNA copies by quantitative reverse transcription polymerase chain reaction (RT-PCR) and compared the results with virus titers determined in the plaque assay.

Materials and Methods

Mice

Six-week-old male conventional ICR mice were purchased from a conventional Japanese mouse breeder (Chiba, Japan). The mice were kept in a temperature-and light-controlled environment with standard food and water given *ad libitum*. All experiments were approved by the Animal Care and Use Committee of Meiji Seika Pharma Pharmaceutical Research Center.

Cells and virus infection

RAW264.7 cells, a mouse macrophage cell line, were purchased from DS-Pharma (Osaka, Japan) and cultured in high-glucose Dulbecco's modified Eagle medium + GlutaMAX™ -I (DMEM; Invitrogen, Carlsbad, CA, USA) supplemented with 25 mM HEPES and 10% fetal bovine serum (FBS) at 37°C in 5% CO_2. MNV was inoculated onto a monolayer of RAW264.7 cells, and cultured in DMEM supplemented with 25 mM HEPES and 2% FBS (DMEM-2% FBS) at 37°C in 5% CO_2. Three to 5 days later, the cultured medium was centrifuged at 15,000 *g* for 5 min at 4°C, and the supernatant was used as the virus stock.

Isolation of MNV

Two stool samples per mouse were collected, homogenized in PBS, and then centrifuged at 15,000 *g* for 15 min at room temperature. The supernatant was passed through a 0.22-μm-pore-size filter, diluted in DMEM-2% FBS, and used to inoculate RAW264.7 cells. Four days later, the culture medium and RAW264.7 cells were frozen at –80°C. After thawing, the culture medium and cells were centrifuged at 8,000 *g* for 5 min at 4°C, and the supernatant was used for further passage. A total of nine passages followed by two successive plaque purifications were performed. Plaque assays were performed as previously described [18] with RAW264.7 cells.

MNV inactivation experiment

Different antiseptics were used to inactivate the MNV. PVP-I (ISODINE solution 10%) was obtained from Meiji Seika Pharma (Tokyo, Japan). NaOCl (Purelox-S) was purchased from Oyalox (Tokyo, Japan). BEC (Hyamine solution 10%) was purchased from Daiichi Sankyo (Tokyo, Japan). CHG (Hibitane gluconate 20%) was purchased from Dainippon Sumitomo Pharma (Osaka, Japan). Inactivation was induced by adding 100 μl of the diluted antiseptic [0.4 or 2% (w/v) PVP-I, 0.2 or 0.4% (w/v) NaOCl, 0.2% (w/v) BEC, 1% (w/v) CHG] to 100 μl of virus stock solution. The final volume of 200 μl reaction mixture was incubated for 15, 30, 45 or 60 s at room temperature prior to neutralization. Neutralization control samples were used as 0-s incubation samples. To stop the effect of PVP-I, 2 μl of 1 M sodium thiosulfate was added to the reaction mixture. To stop the effect of NaOCl, 10 μl of 1 M sodium thiosulfate was added to the reaction mixture, followed by 10 μl of 1 M HEPES (pH 7.4). The neutralized solutions obtained from the above treatment were then 10-fold diluted with DMEM-2% FBS. For the neutralization control, PVP-I and NaOCl were first neutralized with sodium thiosulfate

and diluted with DMEM-2% FBS before being added to the virus stock. To stop the effect of BEC and CHG, the reaction mixtures were 100-fold diluted with DMEM-2% FBS. The neutralization controls for BEC and CHG were prepared by diluting the virus stock with DMEM-2% FBS without adding antiseptics. The virus titer and copy numbers of MNV RNA in the reaction mixture were measured by plaque assays and quantitative RT-PCR, respectively.

RT-PCR

Viral RNA was extracted from the fecal suspension or supernatant using QIAamp Viral RNA Mini Kits (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. An RT-PCR primer pair (forward, 5'-TTTGGAACAATGGATGCTGA-3'; reverse, 5'-TAGGGTGGTACAAGGGCAAC-3') was designed against the conserved capsid sequences of MNV-1 (GenBank accession No. AY228235), MNV-2 (DQ23041), MNV-3 (DQ223042), and MNV-4 (DQ223043). RT-PCR was performed as previously described [7].

Quantitative RT-PCR

Viral RNA was prepared from the reaction mixture using QIAamp Viral RNA Mini Kits. A forward primer (5'-CAGATCACATGCTTCCCACAT-3'), reverse primer (5'-CCAGAGACCACAAAAGACTCATCA-3') and probe (5'FAM -CCCATTCAACTCCCTCTTCTTGA-3'TAMRA) were designed against the conserved sequences of MNV. Quantitative RT-PCR was performed using an EZ RT-PCR Kit (Applied Biosystems, Foster City, CA, USA) and ABI Prism 7700 Sequence Detection System (Applied Biosystems). The following RT-PCR parameters were used: 50°C for 2 min, 60°C for 30 min, 95°C for 5 min, and then 50 cycles of denaturation (94°C, 15 s), annealing (60°C, 30 s), and extension (72°C, 30 s). The standard curve was generated with serial dilutions of MNV-1 RNA encoding the capsid region synthesized by an AmpliScribe T7 Transcription Kit (Epicentre, Madison, WI, USA).

Sequence analysis of viral RNA

Viral RNA was purified from MNV-infected RAW cells with a Purelink Viral RNA/DNA Kit (Invitrogen), and cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen) and an oligo (dT) primer. Genome-specific sequences were amplified with an Expand Long kit (Roche Applied Science, Indianapolis, IN, USA), and PCR products were sequenced directly with a Taq DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems) on an ABI 3730XL DNA analyzer using MNV1-specific primers. When these primers failed, additional sequence-specific primers were designed and used for sequencing. The termini of the MNV genomes were obtained using GeneRacer (Invitrogen), and sequencing was carried out with sequence-specific primers. The nucleotide and protein sequences of ORF1, ORF2, or ORF3 were aligned using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) and employed for subsequent phylogenetic analyses. A phylogenetic tree with 1,000 bootstrap replications was constructed by the neighbor-joining method. The distance of nucleotide substitutions per site was calculated by Kimura's two-parameter method and was visualized using the NJplot software (http://pbil.univ-lyon1.fr/software/njplot.html). Nucleotide and amino acid sequences were analyzed with GENETYX-MAC version 12.2.6 (Genetyx, Tokyo, Japan).

Nucleotide sequence accession number

The MNV nucleotide sequence determined in this study was deposited in DDBJ under the accession number AB601769.

Results

Isolation and genetic characterization of a novel MNV

To identify MNV-infected mice, fecal samples were collected from eight mice, and RT-PCR was performed using a novel set of primers designed from conserved sites on MNV genomes. The MNV gene was detected in all eight samples (Fig. 1A). A plaque became visible after nine passages with RAW264.7 cells (Fig. 1B). After two rounds of plaque purification, MNV was isolated as a single plaque and designated as MNV MT30-2. The complete nucleotide sequence was determined, and homology searches with other MNV strains showed the MNV MT30-2 strain to be genetically similar to MNV CR3/2005/USA [15] (Fig. 2).

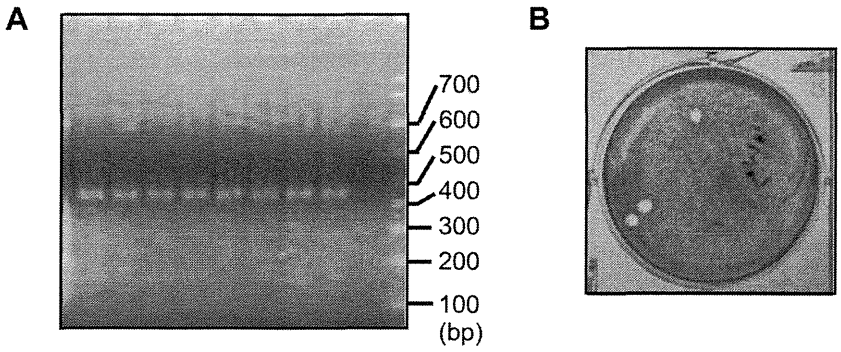

Fig. 1. Detection of MNV gene and MNV plaques. (A) Agarose gel electrophoresis of MNV-specific RT-PCR products (421 bp). (B) Plaques of RAW264.7 cell-adapted MNV.

Table 1. Virucidal effect of PVP-I and NaOCl compared by real-time RT-PCR and plaque assay

Disinfectant	Treatment time (s)	RNA copy no. (log copy no./ml)	Plaque no. (log PFU/ml)
1% (w/v) PVP-I	0	6.69	6.38
	60	6.65	<2.0
0.2% (w/v) NaOCl	0	6.66	6.13
	60	6.30	3.13

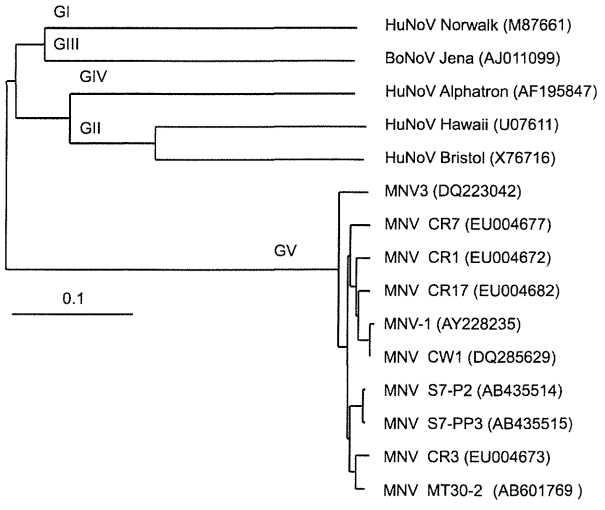

Fig. 2. Phylogenic tree of the Noroviruses. The phylogenic tree was generated using the neighbor-joining method on the basis of an alignment of the entire amino acid sequence of the capsid genes. Numbers in parentheses indicate GenBank accession Nos.

Virucidal activity of antiseptics against MNV

To determine the efficacy of different antiseptics (PVP-I, NaOCl, BEC, and CHG) against the MT30-2 strain, we determined the virus titer using a plaque assay after each antiseptic was added to the virus. Both 0.2 and 1% (w/v) PVP-I reduced the titer by 4 $\log_{10}$ within 15 s (Fig. 3A). Both 0.1 and 0.2% (w/v) NaOCl reduced the titer by 1.6 and 3 $\log_{10}$ in 60 s, respectively (Fig. 3B). In contrast, 0.1% (w/v) BEC and 0.5% (w/v) CHG resulted in only a slight reduction in the titer after 60 s of treatment (Fig. 3C). These results indicated that PVP-I and NaOCl efficiently and rapidly inactivated MNV, whereas both BEC and CHG were ineffective against MNV.

Relationship between plaque assay and quantitative RT-PCR

The effectiveness of antiseptics against a noncultivatable virus is often measured by quantitative RT-PCR. Thus, we quantified MNV MT30-2 RNA copies in samples treated for 60 s with 1% (w/v) PVP-I or 0.2% (w/v) NaOCl and compared the values with the virus titers. We found the number of MNV MT30-2 RNA copies was only slightly reduced when the virus titer was greatly reduced (Table 1), i.e., the quantitative RT-PCR results did not necessarily reflect the plaque assay results.

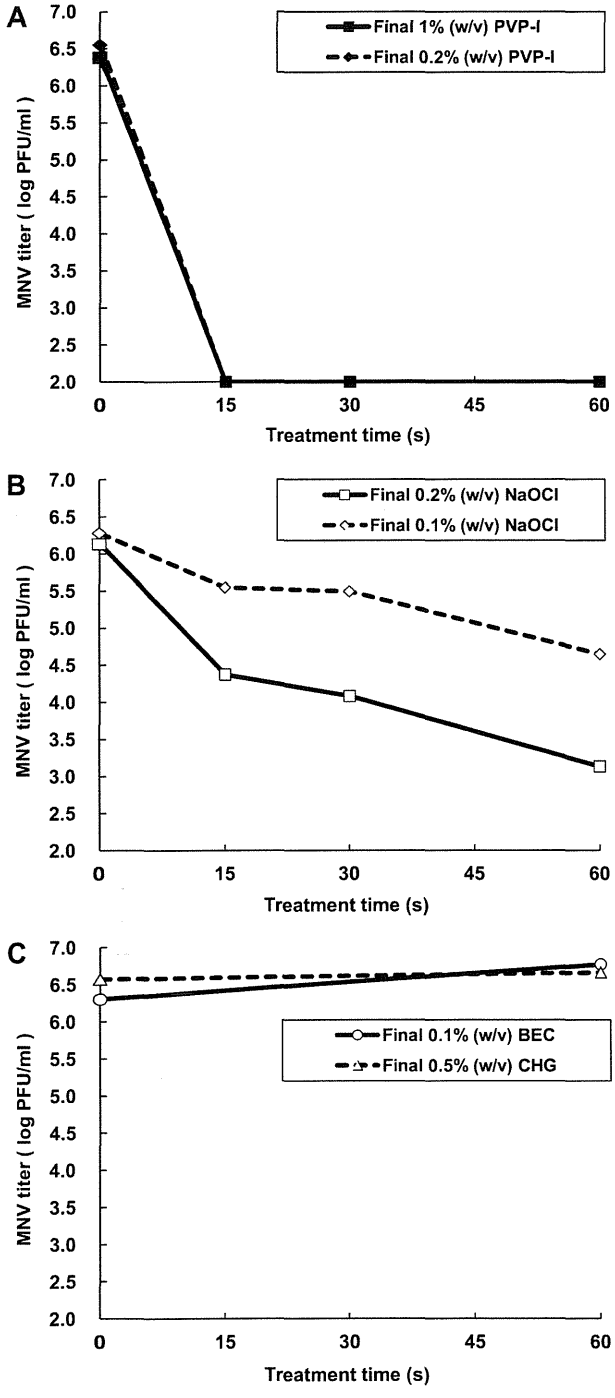

Fig. 3. Virucidal activity of antiseptics against MNV. Each point represents the geometric mean titer (log PFU/ml) of two samples. (A) Effect of PVP-I. Solid squares and solid diamonds indicate the 1% (w/v) and 0.2% (w/v) PVP-I-treated groups, respectively. (B) Effect of NaOCl. Open squares and open diamonds indicate the 0.2% (w/v) and 0.1% (w/v) NaOCl-treated groups, respectively. (C) Effect of BEC and CHG. Open circles and open triangles indicate the 0.1% (w/v) BEC and 0.5% (w/v) CHG-treated groups, respectively.

These results coincided with those in a previous report using MNV CW1 [1].

Discussion

In the present study, we described the isolation of a novel MNV, strain MT30-2. Sequence analysis showed that MT30-2 was genetically similar to MNV CR3/2005/USA (Fig. 2). Using the MT30-2 strain, we showed that two antiseptics (PVP-I and NaOCl) were able to inactivate the virus at low concentrations and in a short contact time (Fig. 3A and 3B). The other two antiseptics, BEC and CHG, which showed virucidal activity against several human viruses [10], appeared to be ineffective against MNV (Fig. 3C). PVP-I and NaOCl also showed strong virucidal activity against MNV CW1 [1], a genetically separated strain from our work (Fig. 2), suggesting that these two antiseptics potentially possess virucidal activity against various MNV strains. EtOH also reduced the MNV infectious titer [1]. However, previous reports demonstrated that EtOH did not show virucidal activity over a short time frame against Feline calicivirus (FCV), another antiseptics-resistant virus that, like MNV, belongs to the *Caliciviridae* family [4, 11]. On the other hand, PVP-I and NaOCl are effective for FCV [3, 4, 11, 16]. Accordingly, PVP-I and NaOCl might be effective against not only other MNV strains but also other related noroviruses including HuNoV.

Since the infectivity of HuNoV is very strong, only a few viral particles induce acute gastroenteritis [2, 14]. Because it is thought that MNV also has strong infectivity the virus needs to be inactivated as completely possible to prevent the spread of MNV. PVP-I, which showed strong virucidal activity in a short time frame against MNV, would be useful for preventing MNV transmissions by human hand washing in laboratories.

In this study, we found that the results of quantitative RT-PCR were not consistent with those of the virucidal assay (Table 1), and the detection of viral RNA was not necessarily related to the virus titers, demonstrating the importance of conducting infectivity assays when evaluating the efficacy of antiseptics. The efficacy of antiseptics may be underestimated if quantitative RT-PCR is used for evaluation, as reported in a previous paper [1]. Thus, interpretation of the results of quantitative

RT-PCR needs further consideration.

In conclusion, our data demonstrate that a novel MNV isolate, MT30-2, is a useful tool to evaluate the effectiveness of antiseptics and that PVP-I as well as NaOCl is the most effective antiseptics for disinfection and infection control of MNV.

References

1. Belliot, G., Lavaux, A., Souihel, D., Agnello, D., and Pothier, P. 2008. Use of murine norovirus as a surrogate to evaluate resistance of human norovirus to disinfectants. 2008. *Appl. Environ. Microbiol.* 74: 3315–3318.
2. Cremon, C., De Giorgio, R., and Barbara, G. 2010. Norovirus gastroenteritis. *N. Engl. J. Med.* 361: 1776–1785.
3. Doultree, J.C., Druce, J.D., Birch, C.J., Bowden, D.S., and Marshall, J.A. 1999. Inactivation of feline calicivirus, a Norwalk virus surrogate. *J. Hosp. Infect.* 41: 51–57.
4. Duizer, E., Bijkerk, P., Rockx, B., De Groot, A., Twisk, F., and Koopmans, M. 2004. Inactivation of caliciviruses. *Appl. Environ. Microbiol.* 70: 4538–4543.
5. Duizer, E., Schwab, K.J., Neill, F.H., Atmar, R.L., Koopmans, M.P., and Estes, M.K. 2004. Laboratory efforts to cultivate noroviruses. *J. Gen. Virol.* 85: 79–87.
6. Goto, K., Hayashimoto, N., Yasuda, M., Ishida, T., Kameda, S., Takakura, A., and Itoh, T. 2009. Molecular detection of murine norovirus from experimentally and spontaneously infection mice. *Exp. Anim.* 58: 135–140.
7. Hsu, C.C., Riley, L.K., Wills, H.M., and Livingston, R.S. 2006. Persistent infection with and serologic cross-reactivity of three novel murine noroviruses. *Comp. Med.* 56: 247–251.
8. Hsu, C.C., Wobus, C.E., Steffen, E.K., Riley, L.K., and Livingston, R.S. 2005. Development of a microsphere-based serologic multiplexed fluorescent immunoassay and a reverse transcriptase PCR assay to detect murine norovirus 1 infection in mice. *Clin. Diagn. Lab. Immunol.* 10: 1145–1151.
9. Karst, S.M., Wobus, C.E., Lay, M., Davidson, M., and Virgin, H.W. 2003. STAT 1-dependent innate immunity to a Norwalk-like virus. *Science* 299: 1575–1578.
10. Kawana, R., Kitamura, T., Nakagomi, O., Matsumoto, I., Arita, M., Yoshihara, N., Yanagi, K., Yamada, A., Morita, O., Yoshida, Y., Furuya, Y., and Chiba, S. 1997. Inactivation of human viruses by povidone-iodine in comparison with other antiseptics. *Dermatology* 195: 29–35.
11. Lages, S.L., Ramakrishnan, M.A., and Goyal, S.M. 2008. In-vivo efficacy of hand sanitisers against feline calicivirus: a surrogate for norovirus. *J. Hosp. Infect.* 68: 159–163.
12. Maler, M. and Kol, W. 2009. A serological survey to evaluate contemporary prevalence of viral agents and Mycoplasma pulmonis in laboratory mice and rats in western Europe. *Lab. Anim. (NY)* 38: 161–165.
13. Mumphrey, S.M., Changotra, H., Moore, T.N., Heimann-Nichols, E.R., Wobus, C.E., Reilly, M.J., Moghadamfalahi, M., Shukla, D., and Karst, S.M. 2007. Murine norovirus 1 infection is associated with histopathological changes in immunocompetent hosts, but clinical disease is prevented by STAT1-dependent interferon responses. *J. Virol.* 81: 3251–3263.
14. Teunis, P.F., Moe, C.L., Liu, P., Miller, S.E., Lindesmith, L., Baric, R.S., Le Pendu, J., and Calderon, R.L. 2008. Norwalk virus: how infectious is it? *J. Med. Virol.* 80: 1468–1476.
15. Thackray, L.B., Wobus, C.E., Chachu, K.A., Liu, B., Alegre, E.R., Henderson, K.S., Kelley, S.T., and Virgin, H.W. 2007. Murine noroviruses comprising a single genogroup exhibit biological diversity despite limited sequence divergence. *J. Virol.* 81: 10460–10473.
16. Urakami, H., Ikarashi, K., Okamoto, K., Abe, Y., Ikarashi, T., Kono, T., Konagaya, Y., and Tanaka, N. 2007. Chlorine sensitivity of feline calicivirus, a norovirus surrogate. *Appl. Environ. Microbiol.* 73: 5679–5682.
17. Wobus, C.E., Thackray, L.B., and Virgin, H.W. 2006. Murine norovirus: a model system to study norovirus biology and pathogenesis. *J. Virol.* 80: 5104–5112.
18. Wobus, C.E., Karst, S.M., Thackray, L.B., Chang, K.O., Sosnovtsev, S.V., Belliot, G., Krug, A., Mackenzie, J.M., Green, K.Y., and Virgin, H.W. 2004. Replication of norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol.* 2: 2076–2084.

JOURNAL OF VIROLOGY, Aug. 2010, p. 8085–8097 Vol. 84, No. 16
0022-538X/10/$12.00 doi:10.1128/JVI.02125-09

Divergent Evolution of Norovirus GII/4 by Genome Recombination from May 2006 to February 2009 in Japan▽‡

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

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

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 ($3C^{pro}$) 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 4.3×10^{-3} nucleotide substitutions/site/year (7), which is comparable 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/.

▽ 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 $3D^{pol}$ 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 5.0×10^4 to 1.9×10^{11} copies/g stool (average, 6.1×10^9 copies/g stool) as judged by the real-time quantitative reverse transcription-PCR assay (23). All stool specimens were stored at −80°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′-GACCCAGCTGGTTGGTT TGGAAAA-3′)/GII4r7516 (5′-ATAGTTTAGCGGCCGCATTCTTATCACA TTACACCCGTGACTCCCCTCG-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. Bootscanning-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/SCRoftware/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 bootscanning 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 χ^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 ~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

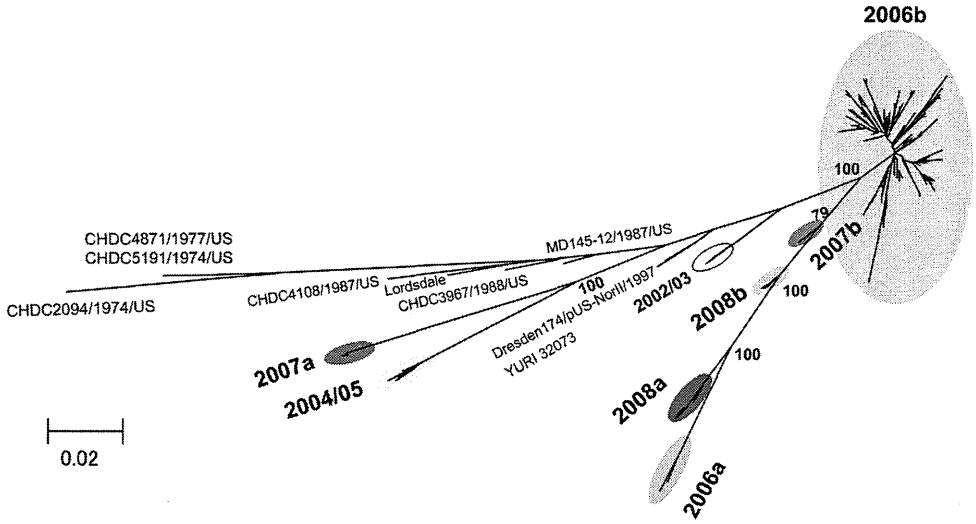

FIG. 1. Phylogenetic classification of the NoV GII/4 subtypes in Japan during 2006 and 2009. The maximum-likelihood tree was constructed with the near-full-length genome sequences (about 7.5 kb) obtained from stool specimens collected at 19 sites in Japan between May 2006 and February 2009 in this study (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

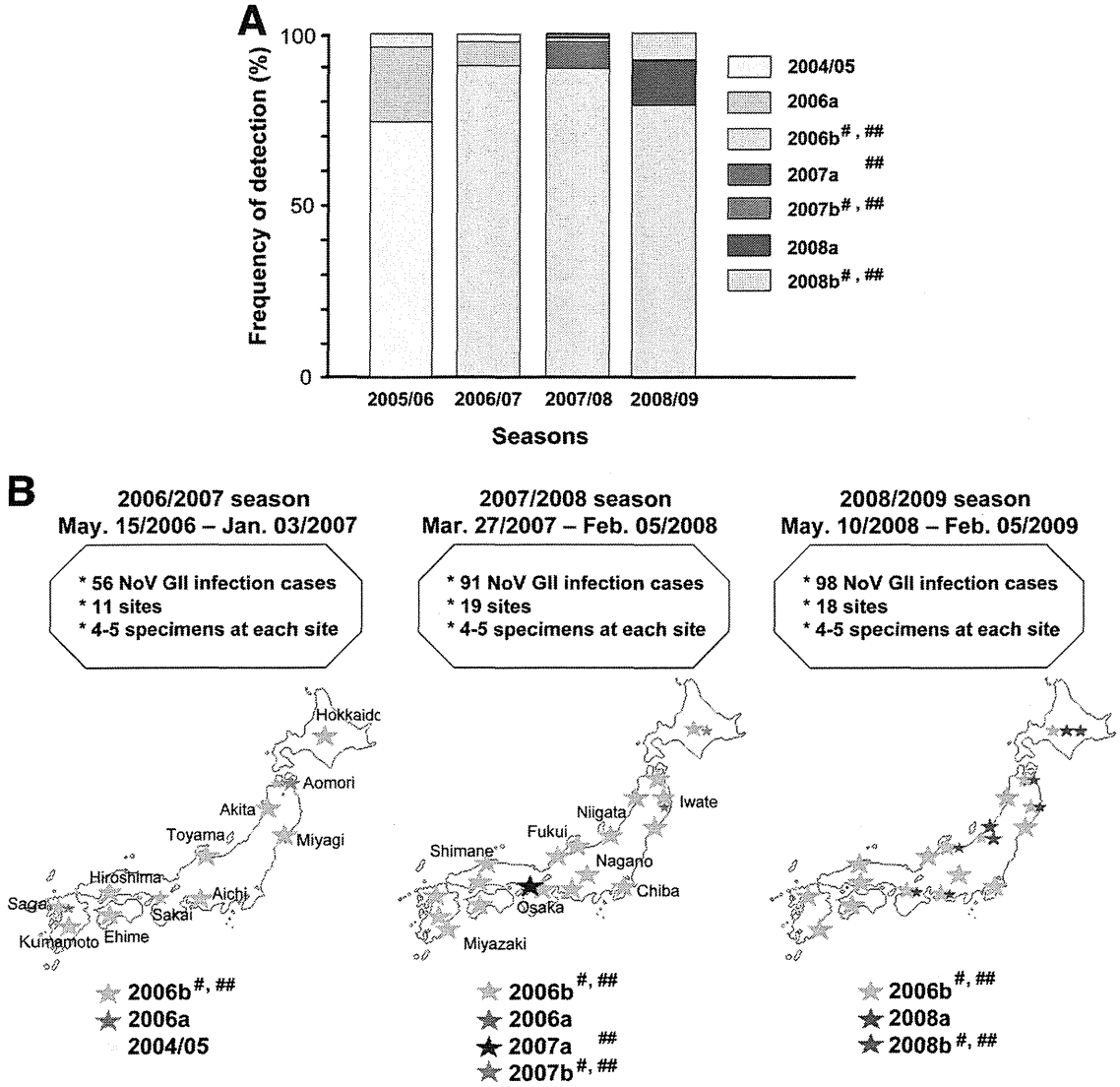

FIG. 2. Temporal and geographical distribution of the NoV GII/4 subtypes in Japan. The 199 near-full-length genome sequences were divided into 3 subgroups according to the collection periods: the 2006/2007 (May 2006 to January 2007) (n = 39), 2007/2008 (March 2007 to February 2008) (n = 78), and 2008/2009 (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

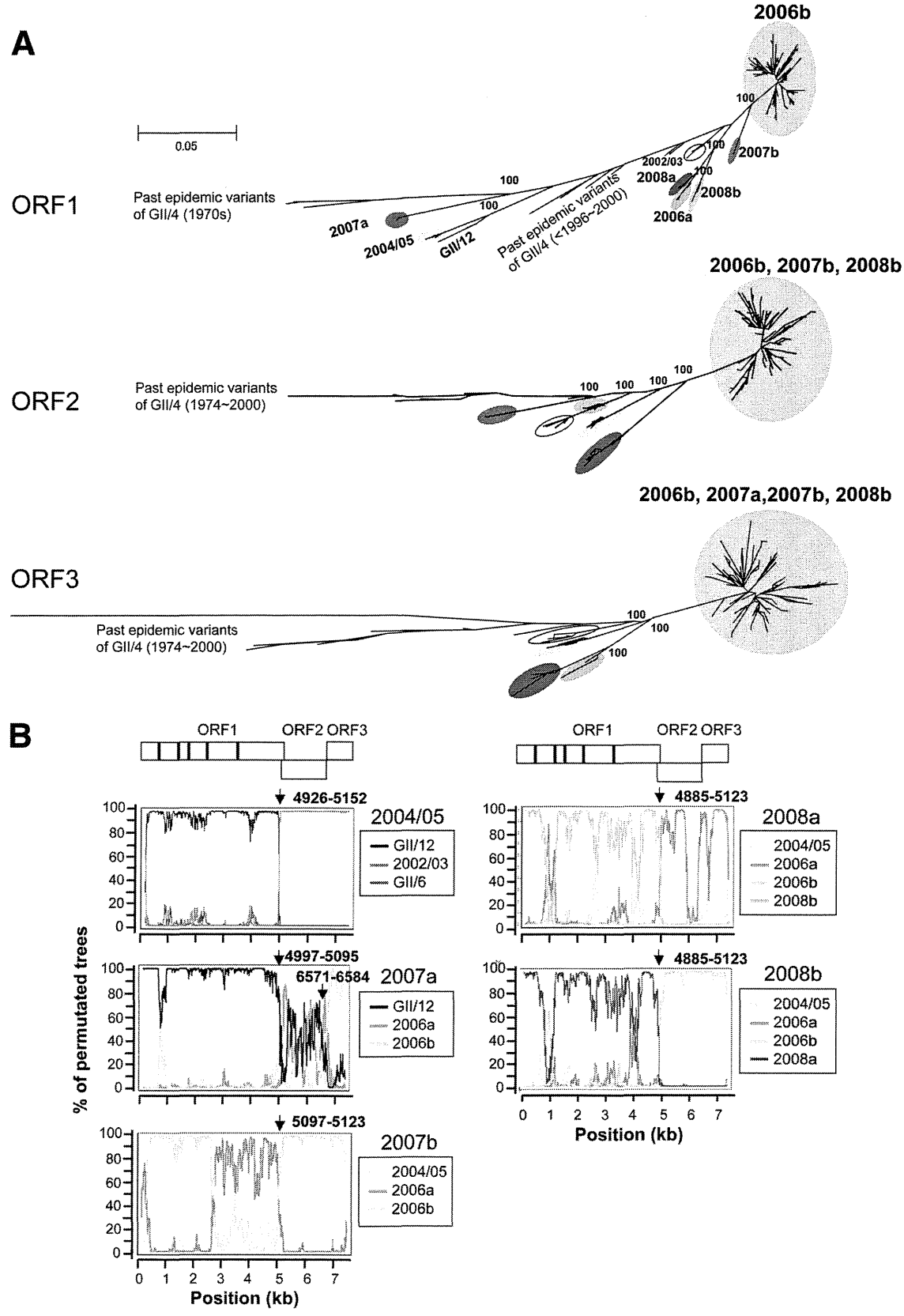

FIG. 3. Phylogenetic evidence for NoV genome mosaicism. (A) Maximum-likelihood trees of the nucleotide sequences of the complete ORF1 (about 5.1 kb), ORF2 (about 1.6 kb), and ORF3 (about 0.8 kb). The trees were constructed with the sequences obtained in previous (38) and present studies (n = 199) and 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 (Hokkaido4/2008/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 χ^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 mosaicisms 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.

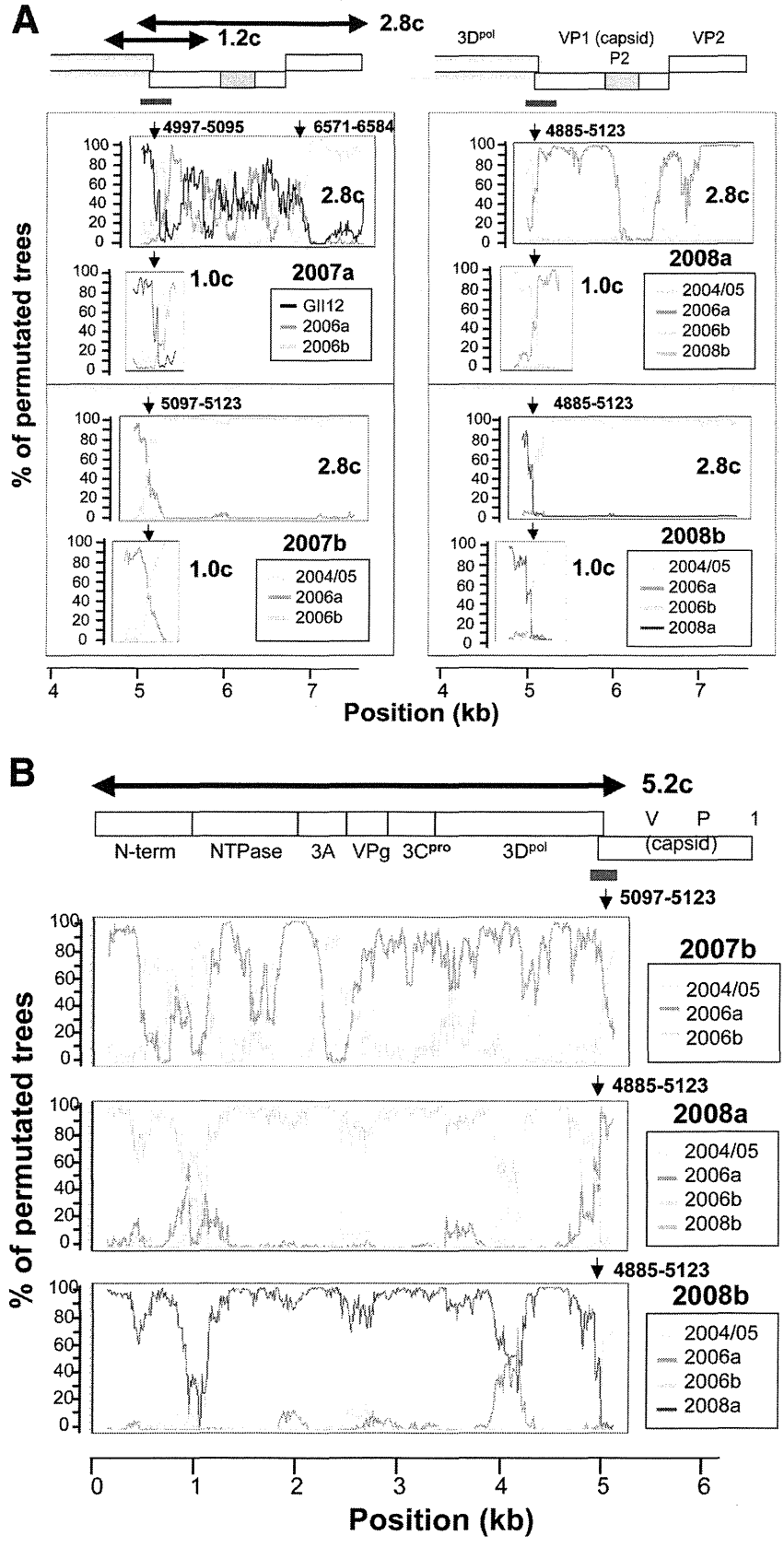

FIG. 4. Isolation of NoV mosaic genome segments. Three genome segments (5.2, 1.0, and 2.8 kb) were amplified from the 2007a, 2007b, 2008a, and 2008b stool specimens, cloned into plasmid vectors, and sequenced. Nucleotide sequences of the cloned segments were subjected to the bootscanning-plot analysis 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.

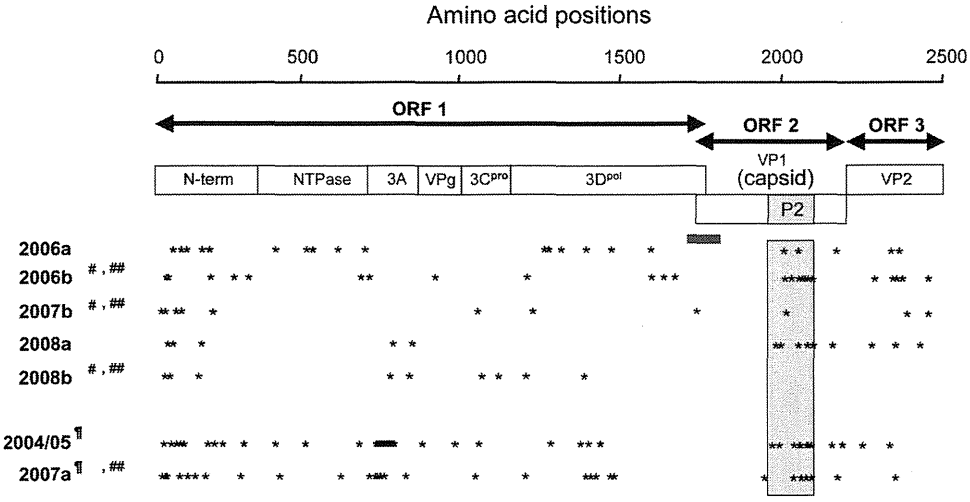

FIG. 5. Amino acid signatures of the NoV GII/4 subtypes. The deduced amino acid sequences of ORF1, ORF2, and ORF3 of a given GII/4 subtype were aligned with the GII/4 sequences identified before the outbreak season of the subtype. Amino acids specific to each subtype at the time of its first outbreak 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 β2 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.