



# Evolutionary Constraints on the Norovirus Pandemic Variant GII.4\_2006b over the Five-Year Persistence in Japan

Hironori Sato<sup>1\*</sup>, Masaru Yokoyama<sup>1</sup>, Hiromi Nakamura<sup>1</sup>, Tomoichiro Oka<sup>2</sup>, Kazuhiko Katayama<sup>2,3</sup>, Naokazu Takeda<sup>4,5</sup>, Mamoru Noda<sup>6</sup>, Tomoyuki Tanaka<sup>7</sup> and Kazushi Motomura<sup>1,4,5</sup>

<sup>1</sup> Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan, <sup>2</sup> Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan, <sup>3</sup> Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan, <sup>4</sup> Research Institute for Microbial Diseases, Osaka University, Osaka, Japan, <sup>5</sup> Thailand-Japan Research Collaboration Center on Emerging and Re-emerging Infections, Nonthaburi, Thailand, <sup>6</sup> National Institute of Health Sciences, Tokyo, Japan, <sup>7</sup> Sakai City Institute of Public Health, Osaka, Japan

## OPEN ACCESS

### Edited by:

Stefan Taube,  
University of Lübeck, Germany

### Reviewed by:

Hirota Ode,  
National Hospital Organization  
Nagoya Medical Center, Japan  
Janet Mans,  
University of Pretoria, South Africa

### \*Correspondence:

Hironori Sato  
hirosato@nih.go.jp

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 06 January 2017

**Accepted:** 27 February 2017

**Published:** 13 March 2017

### Citation:

Sato H, Yokoyama M, Nakamura H, Oka T, Katayama K, Takeda N, Noda M, Tanaka T and Motomura K (2017) Evolutionary Constraints on the Norovirus Pandemic Variant GII.4\_2006b over the Five-Year Persistence in Japan. *Front. Microbiol.* 8:410. doi: 10.3389/fmicb.2017.00410

Norovirus GII.4 is a major cause of global outbreaks of viral gastroenteritis in humans, and has evolved by antigenic changes under the constantly changing human herd immunity. Major shift in the pandemic GII.4 strain periodically occurs concomitant with changes in the antigenic capsid protein VP1. However, how the newly emerged strain evolves after the onset of pandemic remains unclear. To address this issue, we examined molecular evolution of a pandemic lineage, termed the GII.4\_2006b, by using the full-length viral genome and VP1 sequences ( $n = 317$ ) from stools collected at 20 sites in Japan between 2006 and 2011. Phylogenetic tree showed a radial diversification of the genome sequences of GII.4\_2006b, suggesting a rapid genetic diversification of the GII.4\_2006b population from a few ancestral variants. Impressively, amino acid sequences of the variable VP1 in given seasons remained as homogeneous as those of viral enzymes under annual increase in the nucleotide diversity in the VP1 coding region. The Hamming distances between the earliest and subsequent variants indicate strong constraints on amino acid changes even for the highly variable P2 subdomain. These results show the presence of evolutionary constraints on the VP1 protein and viral enzymes, and suggest that these proteins gain near maximal levels of fitness benefits in humans around the onset of the outbreaks. These findings have implications for our understanding of molecular evolution, mechanisms of the periodic shifts in the pandemic NoV GII.4 strains, and control of the NoV GII.4 pandemic strain.

**Keywords:** norovirus, pandemic strain, GII.4, capsid, molecular evolution, evolutionary constraints

## INTRODUCTION

Norovirus (NoV) is a non-enveloped RNA virus that belongs to the family *Caliciviridae*. NoV is divided into multiple genogroups, which are further subgrouped into more than 40 genotypes (Zheng et al., 2006; Donaldson et al., 2010; Robilotti et al., 2015). Among them, genogroup II genotype 4 (GII.4) is especially important in public health, because it is the leading cause of NoV-associated acute gastroenteritis in humans (Noel et al., 1999; Siebenga et al., 2007;

Zheng et al., 2010). NoV is highly variable, and emergence of a novel GII.4 variant is in some cases associated with global outbreaks of gastroenteritis (Lopman et al., 2004; Vipond et al., 2004; Reuter et al., 2005; Bull et al., 2006; Gallimore et al., 2007; Motomura et al., 2008). A major shift in the pandemic strain occurs with antigenic mutations (Lindesmith et al., 2008, 2011, 2012a, 2013) and viral genome recombination at the ORF1/2 boundary region (Motomura et al., 2010; Eden et al., 2013).

The VP1 protein is the major structural protein of the mature virion, which protrudes from the virion surface and plays pivotal roles in the viral interactions with hosts. The VP1 protein is composed of two domains, protruding (P) and shell (S) (Prasad et al., 1999). The P domain is further divided into two subdomains, P1 and P2 (Prasad et al., 1999). The P2 subdomain is placed on the tip of the VP1 protein and constitutes the major antigenic site around the binding site to the putative receptor(s) for infection (Donaldson et al., 2010). This structural feature causes sequence variation (Lindesmith et al., 2008, 2011, 2012a, 2013; Bok et al., 2009; Debbink et al., 2012) and structural diversity (Chen et al., 2004, 2006; Donaldson et al., 2010), particularly in the P2 subdomain. Meanwhile, the functional importance of the P2 subdomain can cause suppression of deleterious changes and/or changes that reduce viral replication fitness. However, very little is known about evolution of the VP1 protein during viral maintenance in human populations.

To address this issue, we examined here molecular evolution of the VP1 protein of a pandemic lineage, termed GII.4\_2006b, which is also known as the GII.4 Den Haag 2006b. In the autumn/winter of 2006, the national epidemiological surveillance of infectious diseases in Japan reported an unusual increase in the number of outbreaks of NoV infections (Infectious Disease Surveillance Center<sup>1</sup>). This augmentation was associated with the nationwide spread of a newly emerging GII.4 variant (Motomura et al., 2008), termed GII.4\_2006b. The GII.4\_2006b initially coexisted as a minority strain among various other NoV lineages in Japan, but starting in October of 2006 it spread extremely rapidly and remained as the major epidemic variant across Japan between 2006 and 2009 (Motomura et al., 2008, 2010). In this study, we characterized nucleotide and amino acid diversities of the VP1 proteins, using serially collected full-length 317 NoV genome and VP1 sequences from infections in Japan between 2006 and 2011. The obtained results show long-term persistence of GII.4\_2006b in human populations in Japan as a dominant GII.4 subpopulation. Interestingly, both the VP1 protein and viral enzymes had remained as highly homogeneous populations, indicating strong evolutionary constraints on changes in these proteins following the onset of the outbreaks.

## MATERIALS AND METHODS

### NoV Genome Sequencing

Stool specimens were collected from individuals with acute gastroenteritis at 20 regional public health institutes in Japan

<sup>1</sup><http://idsc.nih.go.jp/iasr/prompt/graph-ke.html>

between May 2006 and March 2011 in compliance with the Food Sanitation Law of Japan, according to the methods for the protection of personal information (including methods for anonymization in an unlinkable fashion). The research was approved by research and ethical committee in National Institute of Infectious Diseases. Three to five stool specimens were collected at each site in each year. NoV genome sequences were obtained from the stool specimens as described previously (Motomura et al., 2008, 2010).

### Genotype Determination

Norovirus genotype was determined by construction of phylogenetic trees of viral genome sequences. Multiple sequence alignments were done as described previously (Motomura et al., 2008, 2010) using the MAFFT (Katoh et al., 2009) and alignment tools implemented in the MEGA software suite (Tamura et al., 2011). Phylogenetic trees were constructed as described previously (Motomura et al., 2008, 2010) using MEGA software (Tamura et al., 2011). The reliability of interior branches in the tree was assessed by the bootstrap method with 1,000 resamplings.

### Analysis of Diversity of Sequence Population

Mean diversity in the entire sequence population was computed with the “Sequence Diversity” menu in MEGA software suite (Tamura et al., 2011). The overall pairwise mean distance between the sequences was computed with the “Distances” menu in MEGA. As substitution models, a maximum composite likelihood and a Poisson model were used for nucleotide and amino acid sequences, respectively. Variance was estimated by the bootstrap method with 100 to 500 bootstrap replications.

### Analysis of Individual Amino Acid Variation

Amino acid variations at each position of the VP1 (1–530) were calculated as previously described with a multiple sequence alignment as described previously for other viral proteins (Naganawa et al., 2008; Oka et al., 2009; Takahata et al., 2017) on the basis of Shannon’s equation (Shannon, 1997):

$$H(i) = - \sum_{x_i} p(x_i) \log_2 p(x_i) \quad (1)$$

( $x_i = G, A, I, V, \dots$ ),

where  $H(i)$ ,  $p(x_i)$ , and  $i$  indicate the amino acid entropy score of a given position, the probability of occurrence of a given amino acid at the position, and the number of positions, respectively. An  $H(i)$  score of zero indicates absolute conservation, whereas 4.4 bits for amino acids or 2.0 bits for nucleic acids indicates complete randomness.

### Analysis of Amino Acid Substitutions by Hamming Distance

We used the Hamming distance to assess the changeability of the earliest NoV GII.4\_2006b variant in Japan. In information

theory, the Hamming distance between two sequences indicates the minimum number of “amino acid substitutions” required to change one sequence into the other. Because the length of amino acid sequences of the VP1 proteins of the NoV GII.4\_2006b subpopulations were identical (540 amino acid residues), the Hamming distance measured in this study means the number of different amino acid residues between the earliest and subsequently emerged variants in two aligned sequences. Python was used as the programming language to compute Hamming distances. Hamming distances between the sequence in May 2006 (accession number AB447443; earliest GII.4\_2006b sequence in our NoV genome dataset) and the later GII.4\_2006b sequences ( $n = 249$ ) were computed by creating a sequence that assigns mismatches and matches at corresponding positions in the two sequences, and then by counting the numbers of the mismatches.

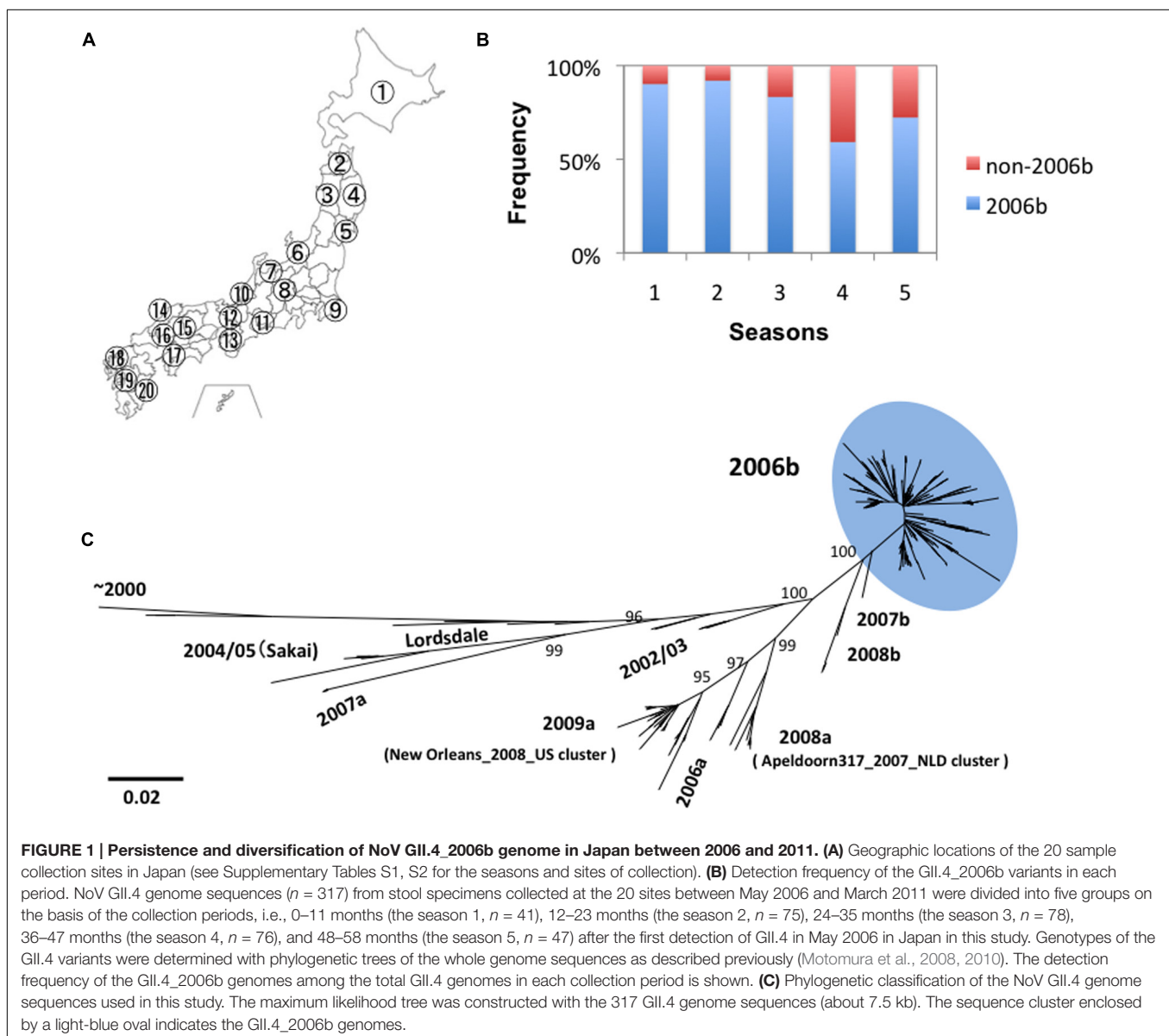
## Nucleotide Accession Numbers

The DDBJ database accession numbers of the 317 NoV GII.4 genome sequences used in this study are provided in Supplementary Table S1 ( $n = 250$ , GII.4\_2006b) and Supplementary Table S2 ( $n = 67$ , GII.4 non-2006b).

## RESULTS

### Persistence and Diversification of NoV GII.4\_2006b Genome in Japan between 2006 and 2011

We obtained 317 genome sequences of NoV GII.4 from the stool specimens collected at 20 sites in Japan between 2006 and 2011 (Figure 1A). Eight distinct lineages of NoV GII.4 were



identified in this period. These lineages include “2004/05” related to Sakai/04-179/2005/JP cluster, “2006a” related to Yerseke 2006a cluster, “2006b” related to Den Haag 2006b cluster, “2007a,” “2007b,” “2008a” related to Apeldoorn317/2007/NL cluster, “2008b,” and “2009a” related to New Orleans 1805/2009/USA cluster (Motomura et al., 2008, 2010; Supplementary Tables S1, S2). Among the newly emerged eight GII.4 lineages, only the pandemic variant GII.4\_2006b had been detected dominantly and continually throughout Japan (Figure 1B). The GII.4\_2006b represented about 79% ( $n = 250$ ) of the total GII.4 genomes detected during the 5 years. Phylogenetic analysis shows that the GII.4\_2006b genome sequences diverged radially from a few roots (Figure 1C). These data suggest genetic bottlenecks followed by a rapid genome diversification of GII.4\_2006b between 2006 and 2011.

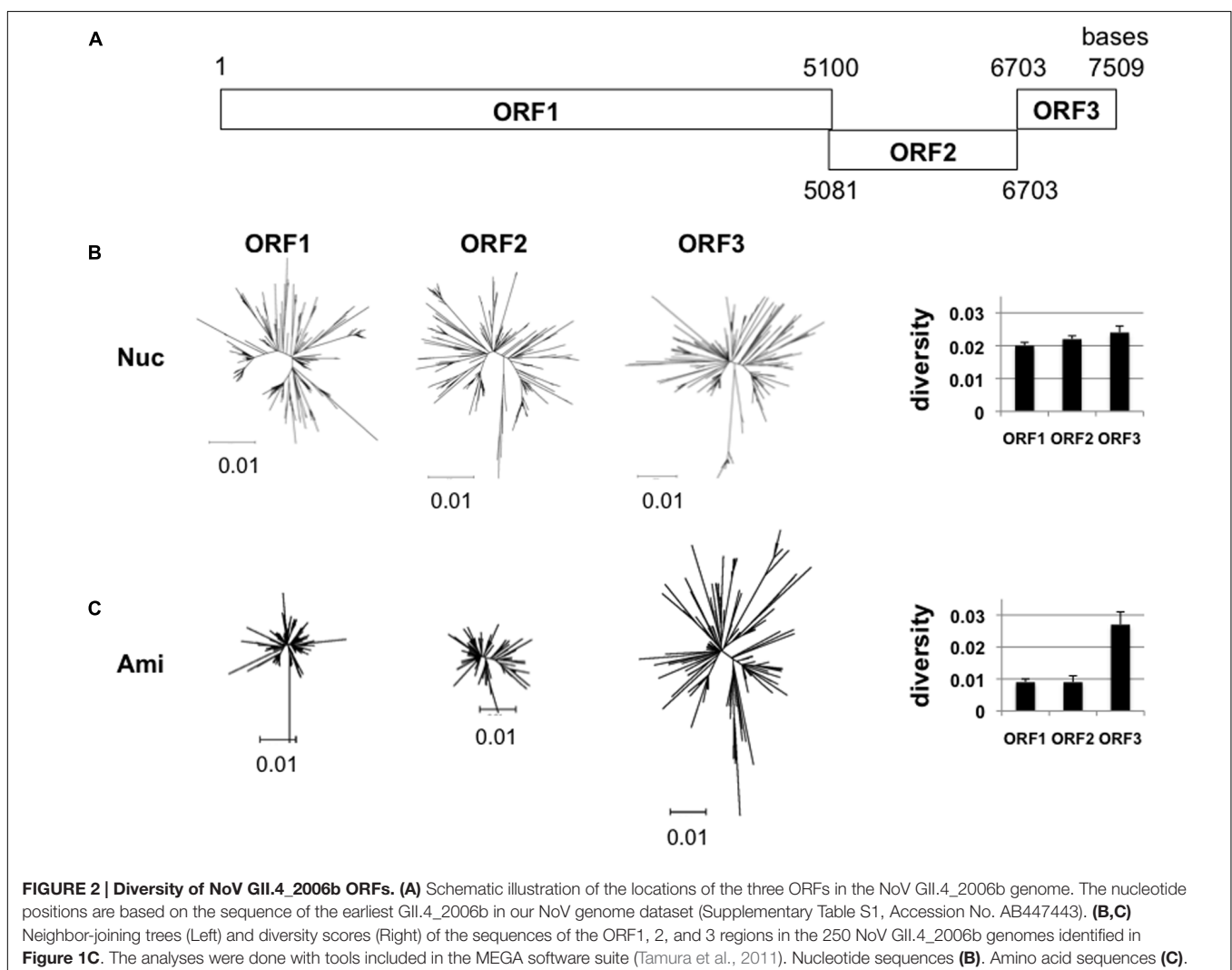
### Diversity of NoV GII.4\_2006b ORFs

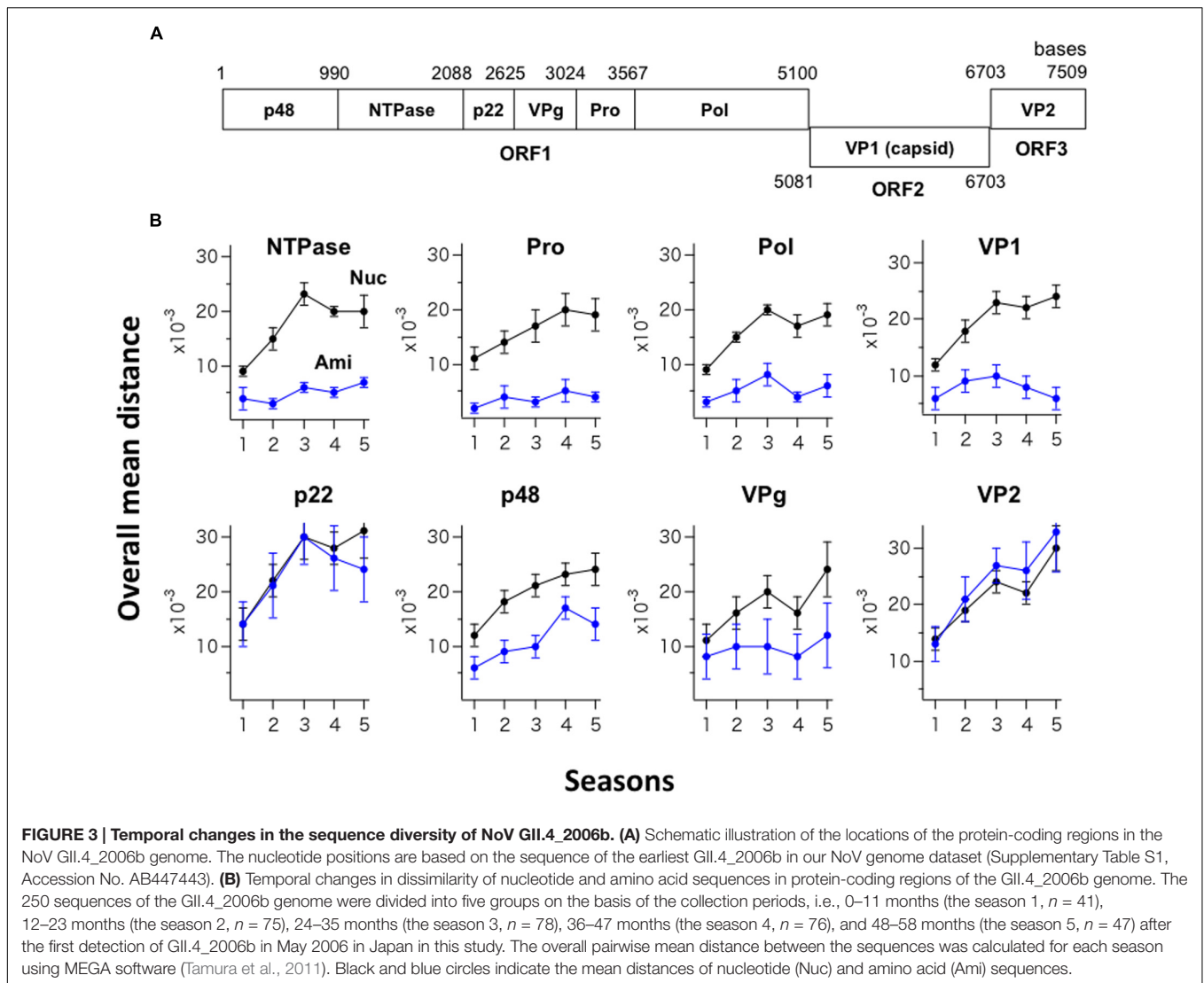
The GII.4\_2006b RNA genome encodes three open reading frames, ORF1, ORF2, and ORF3 (Figure 2A). ORF1 encodes viral enzymes and non-structural proteins. ORF2 and ORF3 encodes

structural proteins, VP1 and VP2, respectively. We first examined whether the sequence diversity is different among the three ORFs by using the 250 GII.4\_2006b genome sequences obtained in this study. The phylogenetic tree and mean diversity in the entire sequence population show that the nucleotide diversity was similar among the three ORFs (Figure 2B). In contrast, a marked difference was observed in the diversity of amino acid sequences: the ORF1 and ORF2 amino acid sequences remained significantly less diversified than that of ORF3 (Figure 2C). The data suggest the presence of constraints on the amino acid changes of proteins encoded by ORF1 and ORF2.

### Temporal Changes in the Sequence Diversity of NoV GII.4\_2006b

The GII.4\_2006b RNA genome encodes eight viral proteins (Figure 3A). Shannon entropy of amino acid sequences of the 2006b ORF1 in the present genome dataset indicates that the potential sites for the internal cleavage of the ORF1 precursor protein were perfectly conserved in amino acid levels [ $H(i) = 0/0$ ] for p48/NTPase (Q/G), NTPase/p22 (Q/G), and VPg/Pro (E/A),





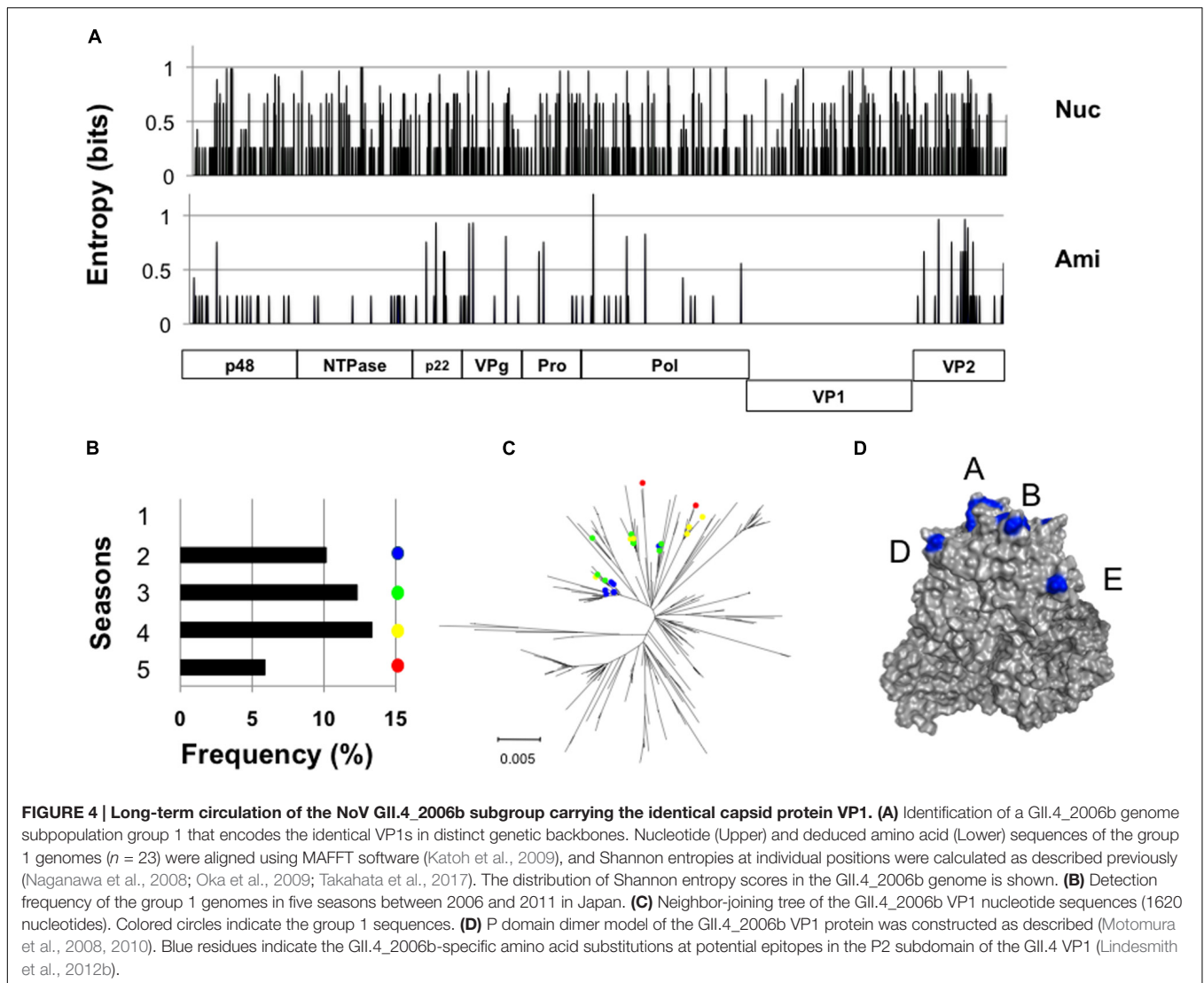
and highly conserved for p22/VPg (E/G) [ $H(i) = 0.03/0.08$ ] and Pro/Pol (E/G) [ $H(i) = 0.03/0.03$ ]. To assess the changeability of individual viral proteins, we examined temporal changes in the sequence diversity of the eight protein-coding regions using the 250 GII.4\_2006b genomes. The genome sequences were divided into five groups based on the collected seasons, and the overall mean distance of the sequences in a season was calculated using MEGA. The nucleotide mean distance sequentially increased for the eight protein-coding regions (Figure 3B, Nuc), indicating a continuous increase in the dissimilarity of every gene segment in the GII.4\_2006b variant population in Japan.

In contrast, the temporal change in amino acid sequence diversity was very different among the eight proteins (Figure 3B, Ami). Interestingly, the amino acid mean distance of the generally hypervariable VP1 protein sequences remained comparable to that of three viral enzymes (NTPase, Pro, Pol) for 5 years, with the mean distance remaining at less than 0.01 with small variances (Figure 3B, Upper). After the 3rd epidemic season, the VP1 amino acid distance even decreased.

Meanwhile, the amino acid mean distances of the p22 and VP2 proteins sharply increased in parallel with an increase in the nucleotide distances (Figure 3B, p22 and VP2), suggesting the continuous diversification of these proteins in association with nucleotide diversification. The mean amino acid distance for the p48 protein increased with time yet less extensively than those for the p22 and VP2 proteins (Figure 3B, p48). The mean amino acid distance for the VPg protein stayed at relatively low levels with large variances (Figure 3B, VPg). In sum, these data suggest the presence of strong constraints on amino acid changes in the capsid protein VP1 and enzymes (NTPase, Pro, Pol) of the GII.4\_2006b under the diversification of nucleotide sequences.

### Long-term Circulation of the NoV GII.4\_2006b Subgroup Carrying the Identical Capsid Protein VP1

We identified a GII.4\_2006b subpopulation ( $n = 23$ ) whose nucleotide sequences differed from each other, yet encoded

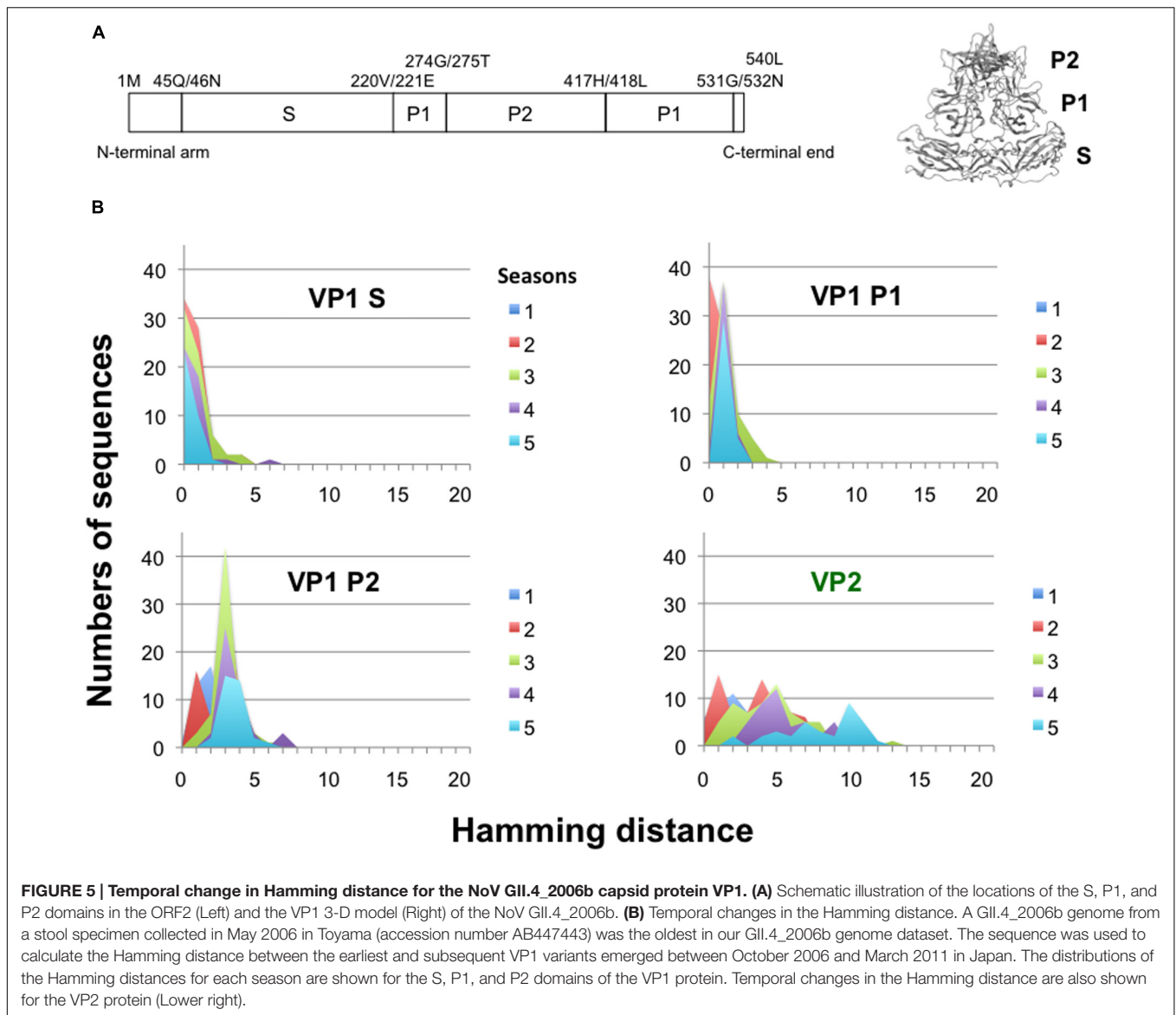


exactly the same VP1 amino acid sequences (Figure 4A). The members of this population, tentatively termed group 1, were detected at distantly located 11 sample collection sites in Japan during the study period (Supplementary Table S1 and Figure S1). They emerged in the second epidemic season in 2007 and continuously circulated without no changes in the VP1 amino acid residues, representing about 6–13% of the GII.4\_2006b genomes in each season (Figure 4B). The group 1 genomes continuously accumulated nucleotide substitutions in the VP1-coding region (Figure 4C), but only synonymous substitutions (Figure 4A). The GII.4\_2006b variants at the onset of epidemics generally had 10 substitutions at the potential epitopes A, B, D, and E (Lindesmith et al., 2012a) of the VP1 P2 subdomain (P294A, T296S, T298N, A368S, D372E, M333V, R382K, S394T, D407S, and T412N), as compared with the VP1 sequence of the past epidemic variant in 2004/2005 in Japan (Sakai/04-179/2005/JP: accession number BAE98194) (Figure 4D). The group 1 VP1 protein had an additional substitution (S393G) on the epitope D.

### Temporal Change in Hamming Distance for the NoV GII.4\_2006b Capsid Protein VP1

The NoV VP1 protein has an architecture similar to that of the VP1 proteins of other single-stranded RNA viruses (Prasad et al., 1994, 1999; Figure 5A). The S domain is highly conserved, whereas the P2 domain is hypervariable among GII.4 variants. To assess the changeability of the P2 domain of the GII.4\_2006b, we examined the temporal accumulations of amino acid substitutions in the S, P1, and P2 regions of the GII.4\_2006b VP1 using the Hamming distance between the earliest and subsequent VP1 variants. As the earliest VP1 variant of the GII.4\_2006b, we used a sequence from a May 2006 sample, which was collected in spring about 5 months before the onset of the nationwide epidemics of the GII.4\_2006b in October of 2006 in Japan (Motomura et al., 2008).

For the S domain, the Hamming distances of the variants in given seasons were at a constant peak of 0 for 5 years (Figure 5B,



VP1 Shell). The data suggest that the GII.4\_2006b variants having amino acid substitutions in the S domain were mostly cleared during epidemics. For the P1 and P2 subdomains, the peaks of Hamming distances were fixed at 1 and 3 after the second and first epidemic seasons, respectively (Figure 5B, VP1 P1 and P2). The data indicate that most of the GII.4\_2006b variants in the early epidemics had a few amino acid substitutions in the P domain but they could not accumulate more mutations after the second epidemic season. Thus the P domain was more variable than the S domain in the GII.4\_2006b variants, as has generally been documented for other NoVs. However, the accumulation of amino acid substitutions was strictly constrained in the P domain of the GII.4\_2006b variants during epidemics. In contrast, the Hamming distances of VP2, a minor structural protein in virion (Glass et al., 2000), continuously increased and showed no evidence of fixation of the peak distance during the study period (Figure 5B, VP2).

## DISCUSSION

In this report, we studied molecular evolution of the NoV capsid protein of a pandemic lineage, GII.4\_2006b. This NoV subpopulation predominated over other coexisting NoV GII.4 subpopulations between the 2006 and 2011 in Japan (Figure 1). Notably, the amino acid sequences of variable VP1 protein of the GII.4\_2006b populations remained as homogeneous as that of the viral enzymes for the 5 years under an increase in nucleotide diversity (Figures 2, 3). Even the GII.4\_2006b population possessing the identical amino acid sequence in the VP1 protein had persisted in the study period (Figure 4). Even the hypervariable antigenic P2 subdomain of the VP1 protein had resisted sequential accumulations of amino acid substitutions (Figure 5). These results suggest the presence of strong evolutionary constraints on the VP1 protein of the NoV pandemic strain. The finding has implications for

our understanding of molecular evolution, mechanisms of the periodic shifts in the pandemic NoV GII.4 strains, and control of the NoV GII.4 pandemic strain.

First, the finding has implications for understanding fitness landscape and evolution of the VP1 protein of NoV GII.4 pandemic strain. The strong constraints on changes imply that the VP1 protein and enzymes of the GII.4\_2006b variants had already gained near maximal levels of fitness benefits in humans around the onset of the outbreaks and that new mutations in the VP1 protein were mostly cleared from the GII.4\_2006b population, probably due to a reduction in the viral fitness for the spread in humans. In order to predominate over other coexisting GII.4 variants, the pandemic variant should have the VP1 structure that confers the best ability to evade preexisting herd immunity against NoV at that time, while also having affinity to bind to receptor(s) on human cells. Because the antigenic sites are located near the receptor-binding site, new antigenic mutations always have the risk to attenuate VP1 protein function and thereby to cause reduction in the viral replication fitness in humans. Thus, it is possible that the VP1 protein of the pandemic strain had remained conserved in human populations primarily by the necessity to maintain advantageous physical property of the VP1 protein for immune evasion and infectivity simultaneously.

Secondary, the finding has implications in the periodic shifts of the pandemic NoV GII.4 strains. Provided that the VP1 protein sequence of a given pandemic variant remained conserved following the onset of epidemics as seen in the GII.4\_2006b, the human herd immunity against the VP1 protein would become increasingly more effective in association with the spread of the virus in humans. Consequently, niche for the pandemic variant in humans would be reduced, and the pandemic variant eventually be replaced by an alternative variant that has the fittest capsid structure under human herd immunity at that time. Consistently, the numbers of reported NoV infection cases in Japan had decreased annually since the late 2007, and the GII.4\_2006b was replaced by a new global pandemic strain GII.4\_Sydney 2012 in the 2013/2014 season, as reported in other countries (van Beek et al., 2013; Eden et al., 2014).

Finally, the finding has implications in the control of NoV pandemic strains. Although development of vaccines and antiviral agents are of special importance to reduce damages from the NoV infections, structural variations in the viral proteins can be problematic. In this regard, the present study suggests the presence of strong constraints on changes in capsid protein and enzymes of a NoV GII.4 pandemic variant on the course of 5-year persistence across Japan. The finding provides a rationale for developing vaccines and antiviral agents against a pandemic strain. A basic premise of the control is that the sequences of the VP1 protein and viral enzymes of a given pandemic

variant remain highly homogeneous after the onset of pandemic. Therefore, it is important to further accumulate information on the evolution of newly emerged pandemic strains to clarify whether present observations of the amino acid conservation in the VP1 and viral enzymes can be extended to other GII.4 pandemic variants. In parallel, it would be important to study genetic diversity of NoV in nature in order to develop systems to predict a new pandemic variant in advance.

## AUTHOR CONTRIBUTIONS

HS conceived the study. MY prepared the computing environment for information science. KK, NT, MN, and TT organized collection of stool specimen. TO, HN, and KM performed sequencing. HN and HS performed variation analysis. HS prepared the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported by a Grant-in-Aid for Scientific Research on Innovative Areas to HS (Grant Number: 25115519) from the Japan Society for the Promotion of Science, a Grant-in-Aid for the Research Program on Emerging and Re-emerging Infectious Diseases to HS (Grant Number: 11350406 and 10103800) from the Ministry of Health, Labor and Welfare of Japan, and a Grant-in Aid for the Research Program on Re-emerging Infectious Diseases to MY from the Japan Agency for Medical Research and development, AMED.

## ACKNOWLEDGMENTS

We would like to thank the following 20 regional public health institutes for the help of collection of stool specimen: Hokkaido, Aomori Prefecture, Akita Prefecture, Iwate Prefecture, Miyagi Prefecture, Niigata Prefecture, Toyama Prefecture, Fukui Prefecture, Nagano Prefecture, Chiba Prefecture, Aichi Prefecture, Sakai City, Osaka City, Shimane Prefecture, Hiroshima Prefecture, Hiroshima City, Ehime Prefecture, Saga Prefecture, Kumamoto City, and Miyazaki Prefecture.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00410/full#supplementary-material>

## REFERENCES

- Bok, K., Abente, E. J., Realpe-Quintero, M., Mitra, T., Sosnovtsev, S. V., Kapikian, A. Z., et al. (2009). Evolutionary dynamics of GII.4 noroviruses over a 34-year period. *J. Virol.* 83, 11890–11901. doi: 10.1128/JVI.00864-09
- Bull, R. A., Tu, E. T., Mciver, C. J., Rawlinson, W. D., and White, P. A. (2006). Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis. *J. Clin. Microbiol.* 44, 327–333. doi: 10.1128/JCM.44.2.327-333.2006
- Chen, R., Neill, J. D., Estes, M. K., and Prasad, B. V. (2006). X-ray structure of a native calicivirus: structural insights into antigenic diversity and host specificity.



- Proc. Natl. Acad. Sci. U.S.A.* 103, 8048–8053. doi: 10.1073/pnas.0600421103
- Chen, R., Neill, J. D., Noel, J. S., Hutson, A. M., Glass, R. I., Estes, M. K., et al. (2004). Inter- and intragenus structural variations in caliciviruses and their functional implications. *J. Virol.* 78, 6469–6479. doi: 10.1128/JVI.78.12.6469-6479.2004
- Debbink, K., Donaldson, E. F., Lindesmith, L. C., and Baric, R. S. (2012). Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *J. Virol.* 86, 1214–1226. doi: 10.1128/JVI.06189-11
- Donaldson, E. F., Lindesmith, L. C., Lobue, A. D., and Baric, R. S. (2010). Viral shape-shifting: norovirus evasion of the human immune system. *Nat. Rev. Microbiol.* 8, 231–241. doi: 10.1038/nrmicro2296
- Eden, J. S., Hewitt, J., Lim, K. L., Boni, M. F., Merif, J., Greening, G., et al. (2014). The emergence and evolution of the novel epidemic norovirus GII.4 variant Sydney 2012. *Virology* 450–451, 106–113. doi: 10.1016/j.virol.2013.12.005
- Eden, J. S., Tanaka, M. M., Boni, M. F., Rawlinson, W. D., and White, P. A. (2013). Recombination within the pandemic norovirus GII.4 lineage. *J. Virol.* 87, 6270–6282. doi: 10.1128/JVI.03464-12
- Gallimore, C. I., Iturriza-Gomara, M., Xerry, J., Adigwe, J., and Gray, J. J. (2007). Inter-seasonal diversity of norovirus genotypes: emergence and selection of virus variants. *Arch. Virol.* 152, 1295–1303. doi: 10.1007/s00705-007-0954-9
- Glass, P. J., White, L. J., Ball, J. M., Leparco-Goffart, L., Hardy, M. E., and Estes, M. K. (2000). Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* 74, 6581–6591. doi: 10.1128/JVI.74.14.6581-6591.2000
- Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64. doi: 10.1007/978-1-59745-251-9\_3
- Lindesmith, L. C., Beltramello, M., Donaldson, E. F., Corti, D., Swanstrom, J., Debbink, K., et al. (2012a). Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog.* 8:e1002705. doi: 10.1371/journal.ppat.1002705
- Lindesmith, L. C., Costantini, V., Swanstrom, J., Debbink, K., Donaldson, E. F., Vinje, J., et al. (2013). Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. *J. Virol.* 87, 2803–2813. doi: 10.1128/JVI.03106-12
- Lindesmith, L. C., Debbink, K., Swanstrom, J., Vinje, J., Costantini, V., Baric, R. S., et al. (2012b). Monoclonal antibody-based antigenic mapping of norovirus GII.4-2002. *J. Virol.* 86, 873–883. doi: 10.1128/JVI.06200-11
- Lindesmith, L. C., Donaldson, E. F., and Baric, R. S. (2011). Norovirus GII.4 strain antigenic variation. *J. Virol.* 85, 231–242. doi: 10.1128/JVI.01364-10
- Lindesmith, L. C., Donaldson, E. F., Lobue, A. D., Cannon, J. L., Zheng, D. P., Vinje, J., et al. (2008). Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5:e31. doi: 10.1371/journal.pmed.0050031
- Lopman, B., Vennema, H., Kohli, E., Pothier, P., Sanchez, A., Negro, A., et al. (2004). Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* 363, 682–688. doi: 10.1016/S0140-6736(04)15641-9
- Motomura, K., Oka, T., Yokoyama, M., Nakamura, H., Mori, H., Ode, H., et al. (2008). Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. *J. Virol.* 82, 11247–11262. doi: 10.1128/JVI.00897-08
- Motomura, K., Yokoyama, M., Ode, H., Nakamura, H., Mori, H., Kanda, T., et al. (2010). Divergent evolution of norovirus GII/4 by genome recombination from May 2006 to February 2009 in Japan. *J. Virol.* 84, 8085–8097. doi: 10.1128/JVI.02125-09
- Naganawa, S., Yokoyama, M., Shiino, T., Suzuki, T., Ishigatsubo, Y., Ueda, A., et al. (2008). Net positive charge of HIV-1 CRF01\_AE V3 sequence regulates viral sensitivity to humoral immunity. *PLoS ONE* 3:e3206. doi: 10.1371/journal.pone.0003206
- Noel, J. S., Fankhauser, R. L., Ando, T., Monroe, S. S., and Glass, R. I. (1999). Identification of a distinct common strain of “Norwalk-like viruses” having a global distribution. *J. Infect. Dis.* 179, 1334–1344. doi: 10.1086/314783
- Oka, T., Yokoyama, M., Katayama, K., Tsunemitsu, H., Yamamoto, M., Miyashita, K., et al. (2009). Structural and biological constraints on diversity of regions immediately upstream of cleavage sites in calicivirus precursor proteins. *Virology* 394, 119–129. doi: 10.1016/j.virol.2009.08.018
- Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., and Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* 286, 287–290. doi: 10.1126/science.286.5438.287
- Prasad, B. V., Rothnagel, R., Jiang, X., and Estes, M. K. (1994). Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *J. Virol.* 68, 5117–5125.
- Reuter, G., Krisztalovics, K., Vennema, H., Koopmans, M., and Szucs, G. (2005). Evidence of the etiological predominance of norovirus in gastroenteritis outbreaks—emerging new-variant and recombinant noroviruses in Hungary. *J. Med. Virol.* 76, 598–607. doi: 10.1002/jmv.20403
- Robilotti, E., Deresinski, S., and Pinsky, B. A. (2015). Norovirus. *Clin. Microbiol. Rev.* 28, 134–164. doi: 10.1128/CMR.00075-14
- Shannon, C. E. (1997). The mathematical theory of communication. 1963. *MD Comput.* 14, 306–317.
- Siebenga, J. J., Vennema, H., Renckens, B., De Bruin, E., Van Der Veer, B., Siezen, R. J., et al. (2007). Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* 81, 9932–9941. doi: 10.1128/JVI.00674-07
- Takahata, T., Takeda, E., Tobiume, M., Tokunaga, K., Yokoyama, M., Huang, Y. L., et al. (2017). Critical contribution of Tyr15 in the HIV-1 integrase (IN) in facilitating IN assembly and nonenzymatic function through the IN precursor form with reverse transcriptase. *J. Virol.* 91:e02003-16. doi: 10.1128/jvi.02003-16
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- van Beek, J., Ambert-Balay, K., Botteldoorn, N., Eden, J. S., Fonager, J., Hewitt, J., et al. (2013). Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro Surveill.* 18, 8–9.
- Vipond, I. B., Caul, E. O., Hirst, D., Carmen, B., Curry, A., Lopman, B. A., et al. (2004). National epidemic of Lordsdale Norovirus in the UK. *J. Clin. Virol.* 30, 243–247. doi: 10.1016/j.jcv.2003.11.011
- Zheng, D. P., Ando, T., Fankhauser, R. L., Beard, R. S., Glass, R. I., and Monroe, S. S. (2006). Norovirus classification and proposed strain nomenclature. *Virology* 346, 312–323. doi: 10.1016/j.virol.2005.11.015
- Zheng, D. P., Widdowson, M. A., Glass, R. I., and Vinje, J. (2010). Molecular epidemiology of genogroup II-genotype 4 noroviruses in the United States between 1994 and 2006. *J. Clin. Microbiol.* 48, 168–177. doi: 10.1128/JCM.01622-09

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Sato, Yokoyama, Nakamura, Oka, Katayama, Takeda, Noda, Tanaka and Motomura. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.