

# SCIENTIFIC REPORTS



OPEN

## Molecular Evolution of the Capsid Gene in Norovirus Genogroup I

Miho Kobayashi<sup>1</sup>, Shima Yoshizumi<sup>2</sup>, Sayaka Kogawa<sup>3</sup>, Tomoko Takahashi<sup>4</sup>, Yo Ueki<sup>5</sup>, Michiyo Shinohara<sup>6</sup>, Fuminori Mizukoshi<sup>7</sup>, Hiroyuki Tsukagoshi<sup>1</sup>, Yoshiko Sasaki<sup>1</sup>, Rieko Suzuki<sup>8</sup>, Hideaki Shimizu<sup>9</sup>, Akira Iwakiri<sup>10</sup>, Nobuhiko Okabe<sup>9</sup>, Komei Shirabe<sup>11</sup>, Hiroto Shinomiya<sup>12</sup>, Kunihiisa Kozawa<sup>1</sup>, Hideki Kusunoki<sup>13</sup>, Akihideo Ryo<sup>14</sup>, Makoto Kuroda<sup>15</sup>, Kazuhiko Katayama<sup>16</sup> & Hirokazu Kimura<sup>14,17</sup>

Received: 11 February 2015

Accepted: 06 August 2015

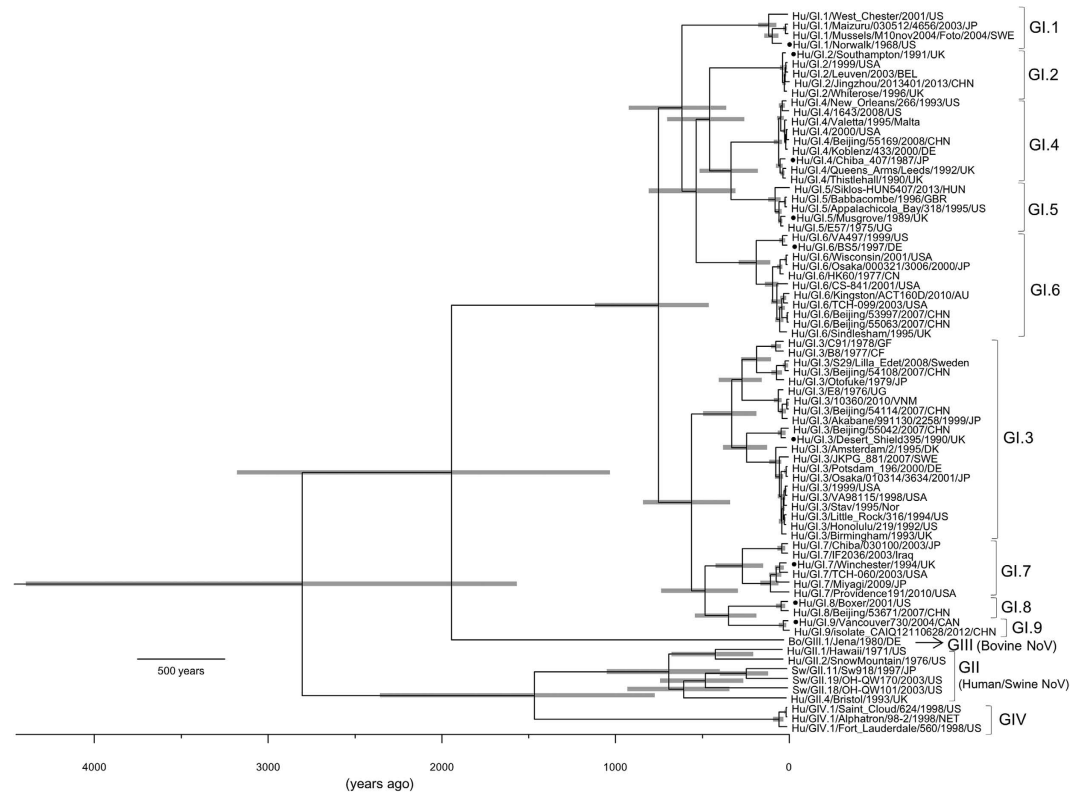
Published: 04 September 2015

We studied the molecular evolution of the capsid gene in all genotypes (genotypes 1–9) of human norovirus (NoV) genogroup I. The evolutionary time scale and rate were estimated by the Bayesian Markov chain Monte Carlo (MCMC) method. We also performed selective pressure analysis and B-cell linear epitope prediction in the deduced NoV GI capsid protein. Furthermore, we analysed the effective population size of the virus using Bayesian skyline plot (BSP) analysis. A phylogenetic tree by MCMC showed that NoV GI diverged from the common ancestor of NoV GII, GIII, and GIV approximately 2,800 years ago with rapid evolution (about  $10^{-3}$  substitutions/site/year). Some positive selection sites and over 400 negative selection sites were estimated in the deduced capsid protein. Many epitopes were estimated in the deduced virus capsid proteins. An epitope of GI.1 may be associated with histo-blood group antigen binding sites (Ser377, Pro378, and Ser380). Moreover, BSP suggested that the adaptation of NoV GI strains to humans was affected by natural selection. The results suggested that NoV GI strains evolved rapidly and date back to many years ago. Additionally, the virus may have undergone locally affected natural selection in the host resulting in its adaptation to humans.

Norovirus (NoV) of the genus *Norovirus* and the family *Caliciviridae* causes acute gastroenteritis in humans<sup>1</sup>. NoV shows strong infectivity leading to large epidemics of acute gastroenteritis in various countries including Japan<sup>2–4</sup>. Accumulating evidence suggests that approximately 50% of patients with

<sup>1</sup>Gunma Prefectural Institute of Public Health and Environmental Science, Maebashi-shi, Gunma 371-0052, Japan.

<sup>2</sup>Hokkaido Institute of Public Health, Sapporo-shi, Hokkaido 060-0819, Japan. <sup>3</sup>Aomori Prefectural Public Health and Environment Center, Aomori-shi, Aomori 030-8566, Japan. <sup>4</sup>Iwate Prefectural Meat Inspection Center, Shiwa-cho, Iwate 020-3311, Japan. <sup>5</sup>Miyagi Prefectural Institute of Public Health and Environment, Sendai-shi, Miyagi 983-0836, Japan. <sup>6</sup>Saitama Institute of Public Health, Yoshimi-machi, Saitama 355-0133, Japan. <sup>7</sup>Tochigi Prefectural Institute of Public Health and Environmental Science, Utsunomiya-shi, Tochigi 329-1196, Japan. <sup>8</sup>Kanagawa Prefectural Institute of Public Health, Chigasaki-shi, Kanagawa 253-0087, Japan. <sup>9</sup>Kawasaki City Institute for Public Health, Kawasaki-shi, Kanagawa 210-0821, Japan. <sup>10</sup>Miyazaki Prefecture Kobayashi Meat Inspection Center, Kobayashi-shi, Miyazaki 886-0004, Japan. <sup>11</sup>Yamaguchi Prefectural Institute of Public Health and Environment, Yamaguchi-shi, Yamaguchi 753-0821, Japan. <sup>12</sup>Ehime Prefectural Institute of Public Health and Environmental Science, Matsuyama-shi, Ehime 790-0003, Japan. <sup>13</sup>Department of Safety Research on Blood and Biological Products, Yokohama City University Graduate School of Medicine, Yokohama-shi, Kanagawa 236-0004, Japan. <sup>14</sup>Department of Molecular Biodefence Research, Yokohama City University Graduate School of Medicine, Yokohama-shi, Kanagawa 236-0004, Japan. <sup>15</sup>Pathogen Genomics Center, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan. <sup>16</sup>Department of Virology II, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan. <sup>17</sup>Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan. Correspondence and requests for materials should be addressed to K.K. (email: katayama@nih.go.jp) or H.K. (email: kimhiro@nih.go.jp)



**Figure 1. Phylogenetic tree of ORF2 constructed by the Bayesian Markov Chain Monte Carlo method.** The phylogenetic tree was based on the whole nucleotide sequence of ORF2 (1593 nt corresponding to GI.1/Norwalk/1968/US). We analysed 65 strains of GI, 6 strains of GII, 1 strain of GIII, and 3 strains of GIV. Each node represents mean root height. The scale bar represents the unit of time (years). The grey bars indicate the 95% HPDs for the estimated year. The reference strains of each genotype are indicated by solid circles.

acute gastroenteritis in the winter season in Japan may be due to NoV infection<sup>5,6</sup>. In addition, large outbreaks of food poisoning due to the virus have been reported<sup>7–9</sup>. Thus, NoV is a major causative agent of acute viral gastroenteritis in industrial countries as well as other major viral agents such as rotaviruses<sup>7–9</sup>.

NoV is classified into 5 genogroups (genogroups I–V)<sup>1</sup>. Among them, genogroups I and II are detected mainly in humans<sup>1</sup>. The NoV genome encodes 3 open reading frames (ORF), and ORF2 encodes the NoV capsid protein<sup>10</sup>. On the basis of detailed genetic analysis, Kroneman *et al.* showed that NoV GI and GII strains can be classified into 9 and 22 genotypes, respectively<sup>11</sup>.

In general, the capsid protein may be an essential determinant of the antigenicity of the non-enveloped virus<sup>12</sup>. For example, it plays pivotal roles in not only viral adsorption/entry but also leads to the generation of neutralising antibodies<sup>13–16</sup>. Thus, to control NoV infection, it is important to understand their antigenic variation<sup>13–16</sup>. NoV evolution has been investigated considerably, but most studies have focused on NoV GII<sup>17–19</sup>.

Recent advances of genetic analysis algorithms enable us to obtain the evolutionary information of various viruses. For example, we can assess the evolutionary time scale of viral genes using the Bayesian Markov chain Monte Carlo (MCMC) method<sup>20</sup>. In addition, maximum likelihood approaches may enable us to analyse the determinants of adaptation in viral proteins such as NoV capsid protein<sup>17,18</sup>. In the present study, we utilise these methods to analyse comprehensively the molecular evolution of the NoV GI capsid gene.

## Results

**Phylogenetic analysis and evolutionary rates of the NoV capsid gene by the Bayesian MCMC method.** We constructed a phylogenetic tree with an evolutionary time scale by the Bayesian MCMC method. The 95% highest posterior densities (HPDs) for each node of the phylogenetic tree are indicated by grey bars in Fig. 1. In the present phylogenetic tree, the NoV GI strains divided into 2 lineages about 750 years ago. These lineages are subdivided into 9 genotypes (genotypes 1–9). Lineage 1 contains genotypes 1, 2, 4, 5, and 6, while lineage 2 contains genotypes 3, 7, 8, and 9. Furthermore, genotype 2, 4, 5, 6, and genotypes 7–9 subdivided from the same ancestor virus, while genotype 1/genotype 3 evolved uniquely. The mean evolutionary rate of the present strains was estimated as  $1.26 \times 10^{-3}$  substitutions/

aa change	SLAC	FEL	IFEL	MEME
Met2Lys				○
Asp7Gly				○
Pro10Thr, Gln, Thr, Ser Thr10Ser Ser10Pro, Thr		○	○	○
Ser16Thr, Asn				○
Ala18Arg				○
Gly19Arg				○
Ile354Val				○
Asn358Asp, Leu Asp358Ser, Asn Ser358Asp				○
Val359Ile, Met Ile359Val				○
Ala367Val				○
Ile390Thr, Leu, Val, Met				○
Leu393Ile				○
Ser397Cys, Leu				○
Ser400Thr, Arg Thr400Ser, Ala, Glu				○
Asp401Asn, Ser, Gly, Ile Asn401Ser				○
Ala402Pro, Ser, Val, Gly, Asp Ser402Asn				○
Val409Asp, Ile, Ala Asp409Asn				○
Ala545Thr, Arg				○
Ser557Leu Leu557Ile Val557Ile		○	○	○

**Table 1. Positive selection sites in ORF2 of NoV GI.** Cut-off value,  $p < 0.05$ .

site/year (95% highest posterior density [HPD]  $7.22 \times 10^{-4}$ – $1.79 \times 10^{-3}$ ). In addition, we obtained the evolutionary rate of 5 genotypes (GI.2–GI.6), while the rate could not be obtained for the other 4 genotypes due to the small number of strains analysed (Supplemental Table S1). As a result, the evolutionary rate of them was significantly different ( $p < 0.05$ , Kruskal-Wallis test). These results suggested that an ancestor NoV GI strain diverged from the ancestor of NoV GII, GIII, and GIV strains and it could be dated back to 1570–4390 years ago, corresponding to 95% HPD (mean diverged year, 2803 years ago) (Fig. 1). Furthermore, the present NoV GI strains diverged about 750 years ago and the virus formed 9 genotypes with wide genetic divergence and rapid evolution.

**Selection pressure analysis.** To estimate comprehensively the positive selection sites in the capsid protein of NoV, we used 4 methods: conservative single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), internal fixed effects likelihood (IFEL), and mixed effects model of evolution (MEME) (Table 1). Only 2 positive selection sites were estimated by the FEL and IFEL methods, while 19 sites were estimated using MEME. Notably, the amino acid (aa) substitutions of aa10 consisted of a variety of amino acids. In addition, the substitutions of aa557 were of a single amino acid. However, these substitution sites were not located in the protruding 2 (P2) domain, which is associated with cellular binding site of the capsid protein for NoV infection. Furthermore, over 400 negative selection sites were found in the capsid gene (Table 2). These results suggested that the positive selection sites in the NoV GI capsid protein are located mainly near the N- and C-terminal regions.

**Predicted epitopes in reference strains.** Using multiple methods such as LEPS<sup>21</sup>, BCPRED<sup>22</sup>, FBCPRED<sup>22</sup>, BepiPred<sup>23</sup>, Antigenic<sup>24</sup>, and LBtope<sup>25</sup>, we predicted the B-cell linear epitopes in the deduced amino acid sequences of the NoV capsid protein in the reference strains. In the present study, we accepted the epitopes as those identified with 4 or more methods and with >10 consecutive amino acids<sup>26</sup>. The detailed data are shown in Table 3. Many epitopes were estimated in the capsid protein of each NoV GI genotype. Of them, an epitope of GI.1 (aa377–388) may be associated with the histo-blood

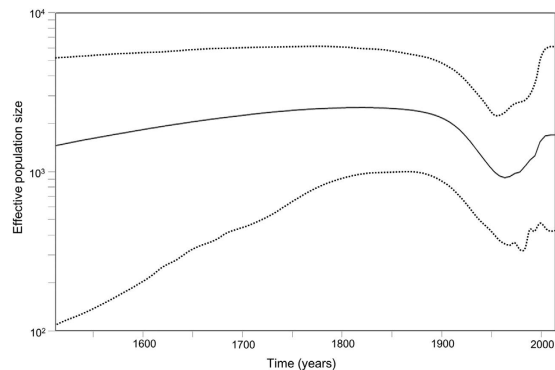
	SLAC	FEL	IFEL
No. of negative selection sites	437	469	428

**Table 2. Negative selection sites in ORF2 of NoV GI.** Cut-off value,  $p < 0.05$ .

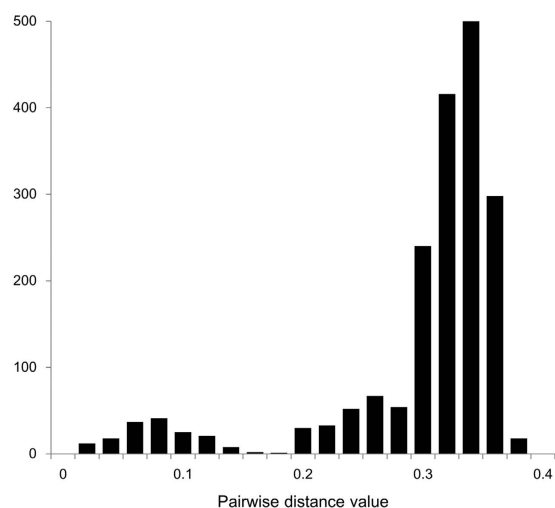
Genotype	Strain (GenBank accession No.)	Position	Predicted epitopes
GI.1	Norwalk/1968/US (M87661)	32–43	AMDPVAGSSTAV
		316–325	<b><u>APIGFPDLGG</u></b>
		377–388	<b><u>SPPSHPSGSQVD</u></b>
		480–496	FLTCVPNGASSGPQQLP
GI.2	Southampton/1991/UK (L07418)	33–45	MEPVAGPTTAVAT
		413–423	<b><u>AANLAPPVFPP</u></b>
		437–448	PGPNNRSAPNDV
		501–511	NGVGAGPQQLP
GI.3	Desert Shield395/1990/UK (U04469)	314–326	<b><u>YHAFESPAPIGFP</u></b>
GI.4	Chiba 407/1987/JP (AB022679)	29–43	DPIPIDPVAGSSTAL
		157–167	PVEVPLEDVRN
		317–327	<b><u>APAPAGFPDLG</u></b>
		386–397	<b><u>TSPPSDSGGANT</u></b>
		435–445	IPGPNQSGSPN
		501–510	SSSTGPQQLP
GI.5	Musgrove/1989/UK (AJ277614)	9–19	TPSADGANGAG
		29–41	EPLPLDPVAGAST
		319–329	<b><u>APTGFDDLGTS</u></b>
		436–447	IPGNNTAHKPNL
GI.6	BS5/1997/DE (AF093797)	312–325	<b><u>PFVPLESPAPVGF</u></b>
GI.7	Winchester/1994/UK (AJ277609)	28–41	AEPLPLEPVGAAT
		189–200	LRAGGASSGTDP
		314–326	<b><u>YHAFESPAPLGF</u></b>
		392–402	<b><u>GARVDPWKIPS</u></b>
GI.8	Boxer/2001/US (AF538679)	495–506	PNTGGGPQNLPT
		189–200	LRSKGASSGTDP
		342–353	<b><u>PTELSTGDPSGK</u></b>
GI.9	Vancouver730/2004/CAN (HQ637267)	441–453	TVSNPKVPCITLPQ
		499–509	PNAGGGPQTLT
		20–33	QLVPENNNTSEPIN
		318–334	<b><u>HAFESPAPLGFDFGDG</u></b>
		351–366	<b><u>NDPVVVGNVQPYNPQF</u></b>
		374–385	<b><u>VVENPTDQVAT</u></b>

**Table 3. Predicted epitopes of the reference strains for each genotype.** The predicted epitopes in the P2 domain are indicated by bold type and underlined. Common epitopes are indicated by grey shading.

group antigen (HBGA) binding sites (Ser377, Pro378, and Ser380)<sup>27</sup> (Table 3). In addition, 1–3 epitopes in each genotype were found in the P2 domain. In the present GI strains, a consensus epitope motif, PAPxGFP, was predicted in the P2 domain in 7 of the 9 genotypes (GI.1, 3–7, and 9). These results suggested that a few viral binding sites of host cells are linked to the epitope sites in the capsid protein of the viruses.



**Figure 2. Bayesian skyline plot of ORF2 in NoV GI.** The Bayesian skyline plot was estimated under the GTR- $\Gamma_4$  model. The MCMC chains were run for 65,000,000 steps. The Y-axis represents the effective population size and the X-axis represents generation time (year). The solid black line represents the mean value over time. The 95% HPD intervals are shown in dotted lines.



**Figure 3. Distributions of the pairwise distance values of ORF2.** The distributions of the  $p$ -distance values based on the nucleotide sequences of NoV GI. A total of 65 strains were analysed.

**Phylogenetics of NoV GI strains.** We assessed the phylodynamics of the capsid gene of the NoV GI strains using Bayesian skyline plot (BSP) analysis. As a result, the present strains showed effective population size values over 1000 for a period of 500 years (Fig. 2). In addition, a relatively constant value was seen from 1500–1900 CE, but thereafter the values tended to be low (Fig. 2). These results suggested that NoV GI strains might have adapted to humans over 500 years ago.

**Pairwise distance values of intergenogroup and intergenotypes.** To assess the genetic distance among the present strains, we calculated their pairwise distance ( $p$ -distance) (Fig. 3). The  $p$ -distance value of the intergenogroup was  $0.29 \pm 0.07$  (mean  $\pm$  standard deviation [SD]). The  $p$ -distance value of the intergenotypes was  $0.036 \pm 0.010$ – $0.192 \pm 0.082$  (mean  $\pm$  SD). These results suggested that the NoV GI capsid gene has undergone considerable genetic divergence (intergenogroup  $p$ -distance  $> 0.25$ ).

## Discussion

We studied the molecular evolution of the capsid gene in NoV genogroup I. First, we found that the human NoV GI strains diverged approximately 2,800 years ago from the ancestor of the GII, GIII, and GIV strains, although the mean estimated time of divergence had a large variation (95% HPD, 1570–4390 years ago). NoV GI evolved rapidly (approximately  $10^{-3}$  substitutions/site/year). They also had wide genetic divergence ( $p$ -distance  $> 0.25$ ). In addition, the NoV GI strains diverged and formed 9 genotypes over a period of about 750 years. Some genotypes (genotypes 2, 4, 5, and 6) evolved from the same ancestor. Second, 2–19 positive selection sites and over 400 negative selection sites were estimated in the deduced capsid protein of NoV GI. Third, many epitopes were estimated in the deduced capsid proteins.

However, there were few epitopes at the cellular binding site of the capsid protein. Furthermore, BSP analysis suggested that NoV GI strains adapted to humans a long time ago.

With regard to norovirus, many evolutionary and/or molecular epidemiological studies have been reported<sup>17–19,28</sup>. For example, Rackoff *et al.*, estimated the molecular evolution rate of NoV GI.1 and GI.3 as  $1.37 \times 10^{-3}$  and  $1.25 \times 10^{-3}$  substitutions/site/year, respectively<sup>28</sup>. In the present study, we obtained a mean evolutionary rate of  $1.26 \times 10^{-3}$  substitutions/site/year for all NoV GI genotypes. The evolutionary rate among the genotypes GI.2–GI.6 were significantly different (Supplemental Table S1). These results suggested that the evolutionary rate of the NoV GI capsid gene was variable among NoV GI genotypes, although these data, including our own, are limited, because they were estimated using a small number of strains. However, the present study may be the first report to estimate the evolutionary rate for all genotypes of the NoV GI capsid gene.

The capsid protein of a non-enveloped virus plays pivotal functions such as adsorption and entry of the target cells<sup>12</sup>. Divergence of the capsid protein may be linked to the antigenicity of various viruses<sup>12</sup>. Thus, divergence of the capsid protein may reflect differences in the antigenicity of NoV. Furthermore, host defence mechanisms, including the immune system, may act as a selective pressure to NoV. In general, a viral protein with strong antigenicity may undergo strong selection pressure, resulting in the presence of many positive selection sites in the antigenic protein<sup>29</sup>. Indeed, many positive selection sites were found in the capsid proteins of an enterovirus showing strong antigenicity<sup>30</sup>. To date, some representative studies regarding the relationship between positive selection and antigenicity in NoV have been reported<sup>17,28</sup>. For example, Cotton *et al.* showed some positive selection sites in NoV GII strains, at Glu106Arg and Asn298Asp<sup>31</sup>. Moreover, Siebenga *et al.* confirmed some sites in NoV GII/4: Asn6Ser, Asn9Ser/Thr, Ala15Thr, Ile47Val, and Ala534Thr/Val<sup>17</sup>. The capsid proteins of NoV GI and GII may have undergone selective pressure mainly near the N- and C-terminal regions in the host. In the present study, variations (at 2–19 sites) among the 4 models—SLAC, FEL, IFEL, and MEME—was found. In each method, differences in the number of positive selection sites have been found in other virus genomes<sup>32,33</sup>. This may be due to differences in the principles used in each method to estimate the sites<sup>34,35</sup>.

Two distinct types of epitope, T-cell-recognised and B-cell-recognised epitopes, have been confirmed<sup>36</sup>. B-cell-recognised epitopes may be an important index for the prediction of antibody binding sites against NoV GI. Next, previous reports suggested that the HBGA binding sites of the viral P2 domains are associated with infection of host cells<sup>27</sup>. In the present study, although we only estimated B-cell linear epitopes, we found the following: 1) many predicted epitopes were found in the capsid protein of NoV GI; and 2) a consensus epitope motif (PAPxGFP) was estimated in 7 of the 9 GI genotypes (Table 3). Regarding the HBGA binding sites, many predicted epitopes were found in the capsid protein of NoV GI. Among them, an epitope of GI.1 (aa377–388) was estimated at an HBGA binding site of the P2 domain (Table 3)<sup>27</sup>. Previous reports showed that the host cellular binding sites of NoV may be located in the P2 domain of the capsid protein (corresponding to aa279–405 in ORF2 of GI.1/Norwalk/1968/US)<sup>37–39</sup>. If epitopes are located in the P2 domain, the immune system may react with them, leading to the generation of a neutralising antibody. Furthermore, previous reports estimated some epitopes in NoV GII strains<sup>40</sup>. These epitopes are located in the P2 domain on the surface of the capsid protein of GII.4 strains<sup>40</sup>. Furthermore, some positive selection sites were identified in an area of the P2 domain associated with blockade epitope A by using a monoclonal antibody<sup>31</sup>. An effective neutralising antibody may inhibit NoV infection of the host; however, the majority of epitopes in the other NoV GI genotypes were not detected at the HBGA binding sites of the P2 domain. However, we did not confirm the conformational epitopes in the present NoV GI strains. In various RNA viruses, such as dengue viruses, the conformational epitopes may be associated with the production of neutralising antibodies<sup>41–43</sup>. Together, further studies regarding the relationships among B-cell epitopes, including the conformational epitopes, HBGA binding sites, and the consensus epitope motif (PAPxGFP), are needed to assess whether the human immune system can produce effective neutralising antibodies against most types of NoV GI.

Next, to evaluate the effective population size of NoV GI, we performed BSP analysis. This method enables us to estimate the effective population size over a period of several hundred years, even in there are no sequences from strains aged more than 50 years<sup>20</sup>. The effective population size showed a constant value from 1500 to 1900 CE. The values decreased from 1900 to 1950 CE; however, after that, the values were restored. With regard to NoV GII.4, no relationship was found between epidemics of the virus and its effective population size based on calculations using the capsid gene<sup>17</sup>. Conversely, a relationship was found between epidemics of the virus and its effective population size based on partial sequencing of the polymerase gene<sup>17</sup>. Both previous and present results suggest that NoV GI strains evolved and maintained constant genetic divergence. The virus may be subject to selection pressure in the host, resulting in the lack of a significant change in its effective population size.

In conclusion, NoV GI strains evolved rapidly and their common ancestor dated back to approximately 750 years ago. The virus may be under local positive selection to escape the immune system of the host, resulting in its adaptation to humans. In addition, to understand better the molecular evolution of NoV GI, further studies regarding the evolution of other genes including the RNA dependent RNA polymerase (RdRp) gene may be needed.



## Materials and Methods

**Strains and alignments.** We collected complete capsid gene (ORF2) sequences of all NoV GI strains, excluding ORF1/2 recombinant strains, from GenBank. We confirmed the genotype of each strain using the norovirus typing tool NoroNet<sup>44</sup>. All sequences were aligned using Clustal W<sup>45</sup>. More than 99% of identical sequences were removed from the dataset. A total of 65 strains were collected. The nucleotide sequences correspond to positions 1–1593 in ORF2 of GI.1/Norwalk/1968/US (GenBank accession No. M87661). The detailed data are shown in Supplemental Table S2.

**Phylogenetic analysis by the Bayesian Markov Chain Monte Carlo method.** We estimated time-scaled phylogeny and evolutionary rate of ORF2 using the Bayesian MCMC method in the BEAST program v1.7.5<sup>20</sup>. To estimate the time of divergence from the other genogroups, some sequences of GII, GIII, and GIV were added to the sequences from 65 GI strains (Supplemental Table S2). The KAKUSAN4 program was used to select the best nucleotide substitution model<sup>46</sup>. Four clock models (strict, lognormal, exponential, and random) and 4 demographic models (constant size, exponential growth, expansion growth and lognormal growth) were compared using Akaike's information criterion through MCMC (AICM)<sup>47,48</sup>. The datasets were analysed using the GTR- $\Gamma_4$  model of substitution under a lognormal relaxed clock model with an exponential growth model. The MCMC chains were run for 50,000,000 steps to achieve convergence with sampling every 1000 steps. Convergence was assessed from the effective sample size (ESS) after a 10% burn-in using Tracer v1.6<sup>49</sup>. Only parameters with an ESS above 200 were accepted. Uncertainty in the estimates was indicated by the 95% HPD intervals. The maximum clade credibility tree was generated by Tree Annotator v 1.7.5 after a 10% burn-in. The phylogenetic tree was viewed in FigTree v1.3.1. The evolutionary rates of each genotype were also calculated. In addition, to estimate changes in the effective population size through time of NoV GI, a BSP was constructed using the BEAST program as described above.

**Statistical analyses.** Statistical analyses were performed using the Kruskal-Wallis test or Mann-Whitney U test with Bonferroni correction using EZR<sup>50</sup>. Values of  $p < 0.05$  were considered to be significant.

**Estimation of positive and negative selection sites.** To evaluate the selection pressure on the ORF2 region, synonymous ( $dS$ ) and nonsynonymous ( $dN$ ) substitution rates at every codon were calculated by Datamonkey using the following methods: SLAC, FEL, IFEL, and MEME<sup>51</sup>. SLAC is intensive for large alignments compared to the other methods<sup>35</sup>; however, this method pretends to underrate the substitution rate<sup>35</sup>. In contrast, the FEL and IFEL methods consider both synonymous and nonsynonymous rate variations and may be efficiently parallelised<sup>35</sup>. MEME can consider episodic selective pressure<sup>34</sup>. We employed 4 different methods for accurate calculations. The cut-off  $p$ -value was set at 0.05.

**Epitope prediction.** The B-cell linear epitopes of the standard reference strains were predicted as described previously<sup>26</sup>. We used the following six tools: LEPS<sup>21</sup>, BCPRED<sup>22</sup>, FBCPRED<sup>22</sup>, BepiPred<sup>23</sup>, Antigenic<sup>24</sup>, and LBtope<sup>25</sup>. All tools were used in the default condition. We accepted the common epitopes estimated by 4 or more tools and with  $> 10$  consecutive amino acids<sup>26</sup>.

**Calculation of  $p$ -distance values.** To assess the frequency distribution of NoV GI, the  $p$ -distance values of intergenogroup and intergenotypes were calculated. We analysed the present strains using MEGA 6.0<sup>52</sup>.

## References

- Green, K. Y. in *Fields Virology* 6th edn, Vol. 1 (eds Knipe, D. M. *et al.*) Ch. 20, 582–608 (Lippincott Williams & Wilkins, 2013).
- Hall, A. J. *et al.* Norovirus disease in the United States. *Emerg. Infect. Dis.* **19**, 1198–1205 (2013).
- Ozawa, K., Oka, T., Takeda, N. & Hansman, G. S. Norovirus infections in symptomatic and asymptomatic food handlers in Japan. *J. Clin. Microbiol.* **45**, 3996–4005 (2007).
- Siebenga, J. J. *et al.* Norovirus illness is a global problem: emergence and spread of norovirus GII. 4 variants, 2001–2007. *J. Infect. Dis.* **200**, 802–812 (2009).
- Dey, S. K. *et al.* Seasonal pattern and genotype distribution of norovirus infection in Japan. *Pediatr. Infect. Dis. J.* **29**, e32–e34 (2010).
- Nakanishi, K. *et al.* Detection of enteric viruses in rectal swabs from children with acute gastroenteritis attending the pediatric outpatient clinics in Sapporo, Japan. *J. Clin. Virol.* **46**, 94–97 (2009).
- Bernard, H. *et al.* Outbreak Investigation Team. Large multistate outbreak of norovirus gastroenteritis associated with frozen strawberries, Germany, 2012. *Euro. Surveill.* **19**, 20719 (2014).
- Li, Y. *et al.* An outbreak of norovirus gastroenteritis associated with a secondary water supply system in a factory in south China. *BMC Public Health* **13**, 283 (2013).
- Zomer, T. P. *et al.* A foodborne norovirus outbreak at a manufacturing company. *Epidemiol. Infect.* **138**, 501–506 (2010).
- Xi, J. N., Graham, D. Y., Wang, K. N. & Estes, M. K. Norwalk virus genome cloning and characterization. *Science* **250**, 1580–1583 (1990).
- Kroneman, A. *et al.* Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* **158**, 2059–2068 (2013).
- Harrison, S. C. in *Fields Virology* 6th edn, Vol. 1 (eds Knipe, D. M. *et al.*) Ch. 3, 52–86 (Lippincott Williams & Wilkins, 2013).
- Chakravarty, S., Hutson, A. M., Estes, M. K. & Prasad B. V. Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. *J. Virol.* **79**, 554–568 (2005).

14. Nilsson, M. *et al.* Evolution of human calicivirus RNA *in vivo*: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J. Virol.* **77**, 13117–13124 (2003).
15. Prasad, B. V. *et al.* X-ray crystallographic structure of the Norwalk virus capsid. *Science.* **286**, 287–290 (1999).
16. Tan, M., Hegde, R. S. & Jiang, X. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *J. Virol.* **78**, 6233–6242 (2004).
17. Siebenga, J. J. *et al.* Phylogenetic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. *PLoS Pathog.* **6**, e1000884 (2010).
18. Eden, J. S., Tanaka, M. M., Boni, M. F., Rawlinson, W. D. & White, P. A. Recombination within the pandemic norovirus GII. 4 lineage. *J. Virol.* **87**, 6270–6282 (2013).
19. Vega, E. *et al.* Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. *J. Clin. Microbiol.* **52**, 147–155 (2014).
20. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
21. Wang, H. W., Lin, Y. C., Pai, T. W. & Chang, H. T. Prediction of B-cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification. *J. Biomed. Biotechnol.* **2011**, 432830 (2011).
22. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21**, 243–255 (2008).
23. Larsen, J. E., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* **2**, 2 (2006).
24. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
25. Singh, H., Ansari, H. R. & Raghava, G. P. Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One* **8**, e62216 (2013).
26. Kim, Y. J. *et al.* Rapid replacement of human respiratory syncytial virus A with the ON1 genotype having 72 nucleotide duplication in G gene. *Infect. Genet. Evol.* **26**, 103–112 (2014).
27. Choi, J. M., Hutson, A. M., Estes, M. K. & Prasad, B. V. Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. *Proc. Natl. Acad. Sci. USA* **105**, 9175–9180 (2008).
28. Rackoff, L. A., Bok, K., Green, K. Y. & Kapikian, A. Z. Epidemiology and Evolution of Rotaviruses and Noroviruses from an Archival WHO Global Study in Children (1976–79) with Implications for Vaccine Design. *PLoS One.* **8**, e59394 (2013).
29. Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
30. Zhang, D., Lu, J. & Lu J. Enterovirus 71 vaccine: close but still far. *Int. J. Infect. Dis.* **14**, e739–743 (2010).
31. Cotton, M. *et al.* Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. *J. Virol.* **88**, 11056–11069 (2014).
32. Boulila, M., Ben Tiba, S. & Jilani, S. Molecular adaptation within the coat protein-encoding gene of Tunisian almond isolates of Prunus necrotic ringspot virus. *J. Genet.* **92**, 11–24 (2013).
33. Botosso, V. F. *et al.* Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. *PLoS Pathog.* **5**, e1000254 (2009).
34. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
35. Pond, S. L. & Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222 (2005).
36. Braciale, T. J., Hahn Y. S. & Burton, D. R. in *Fields Virology 6th edn*, Vol. 1 eds Knipe, D. M. *et al.* Ch. 9, 214–253 (Lippincott Williams & Wilkins, 2013).
37. Katpally, U., Wobus, C. E., Dryden, K., Virgin, H. W. IV & Smith, T. J. Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. *J. Virol.* **82**, 2079–2088 (2008).
38. Lochridge, V. P. & Hardy, M. E. A single-amino-acid substitution in the P2 domain of VP1 of murine norovirus is sufficient for escape from antibody neutralization. *J. Virol.* **81**, 12316–12322 (2007).
39. Motomura, K. *et al.* Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genome-wide tracing of evolutionary history. *J. Virol.* **82**, 11247–11262 (2008).
40. Chen, L. *et al.* Bioinformatics analysis of the epitope regions for norovirus capsid protein. *BMC Bioinformatics.* **14**, S5 (2013).
41. Amat-ur-Rasool, H., Saghir, A. & Idrees, M. Computational Prediction and Analysis of Envelop Glycoprotein Epitopes of DENV-2 and DENV-3 Pakistani Isolates: A First Step towards Dengue Vaccine Development. *PLoS One.* **10**, e0119854 (2015).
42. Keck, Z. Y. *et al.* Human monoclonal antibodies to a novel cluster of conformational epitopes on HCV E2 with resistance to neutralization escape in a genotype 2a isolate. *PLoS Pathog.* **8**, e1002653 (2012).
43. Klein, F. *et al.* Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J. Exp. Med.* **209**, 1469–1479 (2012).
44. Kroneman A. *et al.* An automated genotyping tool for enteroviruses and noroviruses. *J. Clin. Virol.* **51**, 121–125 (2011).
45. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
46. Tanabe, A. S. Kaksusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* **11**, 914–921 (2011).
47. Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol.* **18**, 1001–1013 (2001).
48. Kass, R. E. & Raftery, A. E. Bayes factors. *J Am Stat Assoc.* **90**, 773–795 (1995).
49. Rambaut, A. & Drummond, A. J. *Tracer.* (2013) Available at: <http://tree.bio.ed.ac.uk/software/tracer>. (Accessed: 11th December 2014).
50. Kanda, Y. Investigation of the freely available easy-to-use software 'EZ' for medical statistics. *Bone Marrow Transplant.* **48**, 452–458 (2013).
51. Pond, S. L. & Frost, S. D. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533 (2005).
52. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

## Acknowledgements

This work was partly supported by a commissioned project for Research on Emerging and Re-emerging Infectious Diseases from the Japanese Ministry of Health, Labour and Welfare and Japan Agency for Medical Research and Development.



### Author Contributions

H.K. and K.K. designed the study. M.K., S.Y., S.K., T.T., Y.U., M.S., F.M., H.T., Y.S., R.S., H.S. and A.I. analysed data. N.O., K.S., H.S., K.K. and H.K. contributed analytic tools. H.K., M.K., A.R., H.M. and K.K. wrote the paper.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Kobayashi, M. *et al.* Molecular Evolution of the Capsid Gene in Norovirus Genogroup I. *Sci. Rep.* **5**, 13806; doi: 10.1038/srep13806 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>