Microbiology and Immunology

*Microbiol Immunol* 2016; 60: 418–426
doi: 10.1111/1348-0421.12384

ORIGINAL ARTICLE

# Predicting genotype compositions in norovirus seasons in Japan

Yoshiyuki Suzuki[1], Yen Hai Doan[2], Hirokazu Kimura[3], Hiroto Shinomiya[4], Komei Shirabe[5] and Kazuhiko Katayama[2]

[1]Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Nagoya, Aichi 467-8501, [2]Department of Virology II, [3]Infectious Disease Surveillance Center, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama, Tokyo 208-0011, [4]Ehime Prefectural Institute of Public Health and Environmental Science, 8-234 Sanbancho, Matsuyama, Ehime 790-0003 and [5]Yamaguchi Prefectural Institute of Public Health and Environment, 2-5-67 Aoi, Yamaguchi, Yamaguchi 753-0821, Japan

## ABSTRACT

**Noroviruses cause acute gastroenteritis. Since multiple genotypes of norovirus co-circulate in humans changing the genotype composition and eluding host immunity, development of a polyvalent vaccine against norovirus in which the genotypes of vaccine strains match the major strains in circulation in the target season is desirable. However, this would require prediction of changes in the genotype composition of circulating strains. A fitness model that predicts the proportion of a strain in the next season from that in the current season has been developed for influenza A virus. Here, such a fitness model that takes into account the fitness effect of herd immunity was used to predict genotype compositions in norovirus seasons in Japan. In the current study, a model that assumes a decline in the magnitude of cross immunity between norovirus strains according to an increase in the divergence of the major antigenic protein VP1 was found to be appropriate for predicting genotype composition. Although it is difficult to predict the proportions of genotypes accurately, the model is effective in predicting the direction of change in the proportions of genotypes. The model predicted that GII.3 and GII.4 may contract, whereas GII.17 may expand and predominate in the 2015–2016 season. The procedure of predicting genotype compositions in norovirus seasons described in the present study has been implemented in the norovirus forecasting system (NOROCAST).**

**Key words** genotype composition, herd immunity, norovirus, prediction, vaccine.

Noroviruses cause acute gastroenteritis (1). It has been estimated that the overall prevalence of norovirus among all cases of acute gastroenteritis is 18% (2) and that norovirus causes ~200,000 deaths annually worldwide (3, 4). The prevalence of norovirus in low-mortality developing (19%) and developed (20%) countries appeared to be greater than that in high-mortality developing countries (14%) (2), suggesting that improvements in sanitation and hygiene may not be sufficient to control norovirus infection (4). Thus, there is a need to develop vaccines against norovirus.

Norovirus is classified in the genus *Norovirus* of the family *Caliciviridae* (5). The norovirus virion is non-enveloped and ~38 nm in diameter with an icosahedral

**Correspondence**
Yoshiyuki Suzuki, Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Nagoya, Aichi 467-8501, Japan. Tel: +81 52 872 5821; fax: +81 52 872 5821; email: yossuzuk@nsc.nagoya-cu.ac.jp

Received 21 February 2016; revised 25 April 2016; accepted 5 May 2016.

**List of Abbreviations:** $\sigma_b$, standard deviation of patristic distance between strains in a cluster and those in the nearest cluster; $\sigma_w$, standard deviation of patristic distance between strains within a cluster; G, genogroup; HA, hemagglutinin; IASR, Infectious Agents Surveillance Report; INSD, International Nucleotide Sequence Database; $\mu_b$, average patristic distance between strains in a cluster and those in the nearest cluster; $\mu_w$, average patristic distance between strains within a cluster; NIID, National Institute of Infectious Diseases; NOROCAST, norovirus forecasting system; NS, viral non-structural protein; ORF, open reading frame; P, protruding; S, shell; VP, viral structural protein.

symmetry (6). The norovirus genome is a linear, non-segmented, single-stranded, positive-sense RNA ~7.5 kb in length, containing ORF1–ORF3 (7, 8). ORF1 encodes a non-structural polyprotein precursor, which is cleaved into NS1/2 (p48), NS3 (nucleoside triphosphate), NS4 (p22), NS5 (VPg), NS6 (protease) and NS7 (RNA-dependent RNA polymerase). ORF2 and ORF3 encode the major (VP1) and minor (VP2) capsid proteins, respectively. VP1 consists of shell (S) and protruding (P) domains and determines the antigenicity of norovirus (9). The VP1-derived virus-like particle and the P domain-derived P particle are considered to be candidate vaccines against norovirus (10).

Based on similarities in the amino acid sequence of VP1, noroviruses are divided into genogroups GI–GVII, among which GI, GII and GIV infect humans (11). GI, GII, and GIV are further classified into nine (GI.1–GI.9), 22 (GII.1–GII.22) and two (GIV.1 and GIV.2) genotypes, respectively (12). Here a genotype in a genogroup is defined as a cluster ($X$) of strains in the phylogenetic tree using the average ($\mu_w$) and standard deviation ($\sigma_w$) of patristic distance between the strains within cluster $X$, and the average ($\mu_b$) and standard deviation ($\sigma_b$) of patristic distance between the strains belonging to cluster $X$ and the strains belonging to the nearest cluster ($Y$) in the phylogenetic tree, to satisfy the formula $\mu_b - \mu_w > 2 \times \sigma_w + 2 \times \sigma_b$ (13, 14). Multiple genotypes of norovirus co-circulate in humans, changing the genotype composition (15). However, host immunity to norovirus is largely genotype specific (4). In addition, within a genotype, the antigenicity of norovirus evolves under positive selection (16), thus eluding host immunity (17). Furthermore, the duration of host immunity to a norovirus strain has been estimated to be 4.1–8.7 years (18). It would therefore be desirable to develop a polyvalent vaccine against norovirus that can be re-formulated periodically by predicting the genotype composition in the target season (4).

The genotype composition of norovirus in a target season is determined by the genotype composition in the season immediately before the target season (pre-target season), herd immunity, random genetic drift, environmental factors such as temperature and humidity, and other factors. For influenza A virus, the effects of herd immunity to and thermodynamic stability of the major antigenic protein HA have been modeled to define the fitness of a strain in the pre-target season (fitness model); this model has then been used to predict the proportion of that strain in the target season (19). The fitness model has been used to select candidate vaccine strains against influenza A virus (19, 20). The purpose of the present study was to incorporate the fitness model into a scheme for predicting genotype compositions in norovirus seasons in Japan.

## MATERIALS AND METHODS

### Genotype compositions in norovirus seasons in Japan

Monthly surveillance data for the genotype composition of norovirus in Japan are deposited in the IASR of the NIID, Japan. In Japan, norovirus seasons are defined as the period between September of one year and August of the following year (e.g., norovirus season 2010–2011 is from September 2010 to August 2011). Since few genotypes were identified in 2005 and earlier in IASR, we have summarized the monthly surveillance data as the genotype compositions in the seasons from 2006–2007 to 2014–2015 (supplementary Tables S1 and S2). Note that the obsolete nomenclature of norovirus genotypes in IASR has been converted to the latest nomenclature in the present study (14).

### Sequence data

On 28 October 2015, nucleotide sequences for the entire coding region of VP1 for GI and GII norovirus strains together with information on the year and month of isolation were retrieved from the INSD by conducting TBLASTN (version 2.2.26) (21) using the amino acid sequences of VP1 for GI (Hu/GI.1/8FIIa/1968/USA: INSD accession number JX023285) and GII (Hu/GII.4/MD120-12/1987/USA: JX289821) norovirus strains as the query. After eliminating sequences with ambiguous nucleotides or premature termination codons as well as those isolated before the season 2006–2007, 1262 sequences were classified into genotypes using the Norovirus Genotyping Tool (version 1.0) (22).

Additionally, 196 nucleotide sequences encoding the entire region of VP1 for GI and GII norovirus strains isolated in Japan together with information on the year and month of isolation were obtained from NIID, Japan, where the whole genomic sequences of Japanese norovirus strains have been determined as a component of surveillance of norovirus strains circulating in Japan. The INSD accession numbers for the 1458 sequences analyzed in the present study are listed in supplementary Table S3. Multiple alignments of 1458 amino acid sequences of VP1 were made using MAFFT (version 7.215) (23). It was assumed in the present study that these sequences represented those of norovirus strains circulating in Japan (supplementary Table S4).

### Fitness models

Given the genotype compositions of norovirus from the season 2006–2007 to the pre-target season, the simplest way to predict the genotype composition in the target season would be to use the genotype composition in the pre-target season as the estimate. This is equivalent to

assuming that the fitness was almost the same for all strains in the pre-target season, that is, that natural selection did not affect evolution of the genotype composition. This model of no selection was called Model 0 in the present study.

In a study on influenza A virus, the proportion of strain $i$ in the target season $t+1$ ($p_{i(t+1)}$) was predicted from that in the pre-target season $t$ ($p_{i(t)}$) using its fitness ($f_i$), calculated as

$$\hat{p}_{i(t+1)} = p_{i(t)} \times e^{f_i}. \quad [1]$$

(fitness model) (19). In the present study, this fitness model was used to predict the genotype composition of norovirus defining $f_i$ as

$$f_i = f_0 - c_i \quad [2],$$

where $f_0$ is a constant to ensure that $\Sigma_i \hat{p}_{i(t+1)} = 1$. The term $c_i$ denotes the effect of herd immunity on strain $i$, which was quantified as the sum of cross immunity to strain $i$ induced by strain $j$'s that have circulated in the 2006–2007 season to the pre-target season. The magnitude of cross immunity was modeled in several ways in the present study. In Model 1, the magnitude of cross immunity was assumed to be the same ($k$) for all strain $j$'s belonging to the same genotype as strain $i$, that is

$$c_i = \Sigma_j k \qquad t_j \leq t_i, \quad g_j = g_i \quad [3],$$

where $t_i$ and $t_j$ denote the seasons of isolation of strains $i$ and $j$, respectively, and $g_i$ and $g_j$ the genotypes of strains $i$ and $j$, respectively. The duration of host immunity to a norovirus strain is reportedly 4.1–8.7 years (18). Therefore, in Model 2, the magnitude of cross immunity was assumed to decline by 20% per season; thus, the expected duration of cross immunity is roughly 5 years. That is

$$c_i = \Sigma_j k \times 0.8^{t_i - t_j + 1} \qquad t_j \leq t_i, \quad g_j = g_i. \quad [4].$$

The magnitude of cross immunity may also decline as the amino acid sequence of VP1 for strain $i$ diverges from those for strain $j$'s (17). Therefore, in Model 3, the magnitude of cross immunity was assumed to decline at rate $l$ per difference in the amino acid sequence of VP1, that is

$$c_i = \Sigma_j k \times (1-l)^{d_{ij}} \qquad t_j \leq t_i, \quad g_j = g_i \quad [5]$$

where $d_{ij}$ denotes the number of amino acid sites in VP1 that differ between strains $i$ and $j$. In Model 4, which is a mixture of Models 2 and 3, cross immunity was assumed to decline according to both differences in the season of isolation and in the amino acid sequence of VP1 between strains $i$ and $j$, that is,

$$c_i = \Sigma_j k \times 0.8^{t_i - t_j + 1} \times (1-l)^{d_{ij}} \qquad t_j \leq t_i, \quad g_j = g_i. \quad [6].$$

Finally, in Model 5, cross immunity was assumed to be induced by all strains of norovirus irrespective of genotype, its magnitude being specified by the same formula as for Model 4, that is

$$c_i = \Sigma_j k \times 0.8^{t_i - t_j + 1} \times (1-l)^{d_{ij}} \qquad t_j \leq t_i. \quad [7].$$

### Prediction of genotype compositions

The performance of Models 0–5 in predicting genotype compositions in norovirus seasons in Japan was evaluated retrospectively by applying each model to predict the genotype compositions in the seasons from 2008–2009 to 2014–2015. It should be noted that in Models 1–5 the proportion of a strain in the target season was predicted based on a comparison of the amino acid sequence of VP1 for the strain with those for other strains. Therefore, the predicted proportion of a genotype, which is the sum of the predicted proportions of the strains belonging to it, could not be calculated when no sequence data were available for that genotype in the pre-target season. In such situations, the observed proportion of the genotype in the pre-target season was used as the predicted proportion in the target season, as is the case for Model 0.

When predicting the genotype composition in each of the target seasons from 2008–2009 to 2014–2015 using each of models 1–5, $k$ and $l$ were optimized by minimizing the sum of squared deviations of the predicted proportions of genotypes from the observed proportions over the season 2007–2008 to the pre-target season using a genetic algorithm (24). Three sets of random real numbers were assigned as the initial parameter values to assess convergence of the results (20, 25). After confirmation of convergence (data not shown), the estimates of $k$ and $l$ that were associated with the smallest value of the sum of squared deviations were adopted to predict the genotype composition in the target season.

## RESULTS

### Fluctuation of genotype compositions in norovirus seasons in Japan

Numbers of norovirus strains belonging to different genotypes sampled in the seasons from 2006–2007 to

2014–2015 in Japan were summarized from the monthly surveillance data deposited in IASR of NIID, Japan (supplementary Table S1) and were converted to the genotype compositions in these seasons (supplementary Table S2). Although no GIV norovirus strain was observed during these seasons, it has been confirmed that multiple genotypes of GI and GII norovirus strains have co-circulated in Japan (15). GII.4 accounted for more than 30% of norovirus strains sampled in all nine seasons examined and was the most prevalent of all genotypes in eight of those seasons. However, the genotype composition appears to have fluctuated season by season. In particular, GI.3, GI.6, and GII.17 exceeded 5% in one season, GI.4, GII.6, and GII.14 in two seasons, GII.2 in three seasons and GII.3 in six seasons. These observations suggest that a vaccine against norovirus should be polyvalent and that the genotypes of vaccine strains in the polyvalent vaccine should be matched to those of major strains in circulation by predicting the genotype composition in the target season (4).

## Ability of Models 0–5 to predict the proportions of genotypes

The genotype composition of norovirus in each of the seasons 2008–2009 to 2014–2015 was predicted with Models 0–5 with the aim of retrospectively evaluating the performance of each model. The genotype compositions predicted with Models 0–5 are presented in supplementary Tables S5–S10, respectively. The predicted proportions in Models 1–5 are presented in green when they are closer to and blue when they are further from the observed proportions than the predicted proportions in Model 0 (supplementary Tables S5–S10). The predicted proportions in Models 1–5 are presented in red when they are the same as both the observed proportions in the pre-target season and the predicted proportions in Model 0 (supplementary Tables S5–S10) because of a lack of sequence data in the pre-target season (supplementary Table S4). In addition, the sums of squared deviations of predicted proportions from observed proportions, which are considered as indicators of the overall accuracy in prediction, are presented in green and blue in Models 1–5 when they are smaller and greater, respectively, than the corresponding values in Model 0 (supplementary Tables S5–S10). The excess in the sum of squared deviations in Models 1–5 over that in Model 0 is presented in Figure 1.

In model 0, the observed genotype composition in the pre-target season was used as the predicted genotype composition in the target season on the assumption that selection had not contributed to causing the cross immunity between norovirus strains. This model was employed as the reference for evaluating the performance of Models 1–5 in predicting genotype compositions, in which the magnitude of cross immunity was modeled in different ways. In Model 1, it was assumed that cross immunity was genotype specific and of constant magnitude regardless of differences in the season of isolation and amino acid sequence of VP1 between norovirus strains. However, in this model non-zero parameter values were obtained only for the target seasons of 2008–2009 to 2011–2012 and the sum of squared deviations was mostly greater than that for Model 0 (Figure 1 and supplementary Tables S5 and S6), suggesting that Model 1 is not appropriate for predicting the genotype composition. Model 2, in which a 20% decrease in the magnitude of cross immunity per season is assumed, is an extension of Model 1 (18). However, the results obtained from Models 1 and 2 were similar (Figure 1 and supplementary Tables S6 and S7), suggesting that the decrease in magnitude of cross immunity assumed in Model 2 was not large enough to influence prediction of genotype composition.

Model 3, in which a decrease in the magnitude of cross immunity per difference in the amino acid sequence of VP1 between norovirus strains is assumed, is also an extension of Model 1 (17). The sum of squared deviations for Model 3 was greater than that for Model 0 in the first three seasons (2008–2009 to 2010–2011), but was smaller in the following three seasons (2011–2012 to 2013–2014) (Figure 1 and supplementary Tables S5 and S8). Although the relationship reverted in the last season (2014–2015), these observations indicate that Model 3's ability to predict the genotype composition may improve in later seasons because of an increase in the data available for optimizing parameters. Indeed, prediction of the proportions of genotypes was improved in later seasons, as shown in Figure 2, in which the rates of better prediction in Models 1–5 than in Model 0 are plotted. However, the performance of Model 3 is still comparable to that of Model 0, as evidenced by the fact that the numbers of predicted proportions of genotypes presented in green and blue are the same (23 cases each) in the later four seasons (2011–2012 to 2014–2015) (supplementary Table S8). These results indicate that it is difficult to predict the proportions of genotypes accurately, partly because of the effect of random genetic drift (26, 27). Model 4, in which a 20% decrease in the magnitude of cross immunity per season is assumed, is an extension of Model 3; however, the results obtained with Model 4 are similar to those obtained with Model 3 (Figure 1 and supplementary Tables S8 and S9), as is true for Models 1 and 2, as described above. Finally, Model 5, which takes into account the cross immunity induced by norovirus

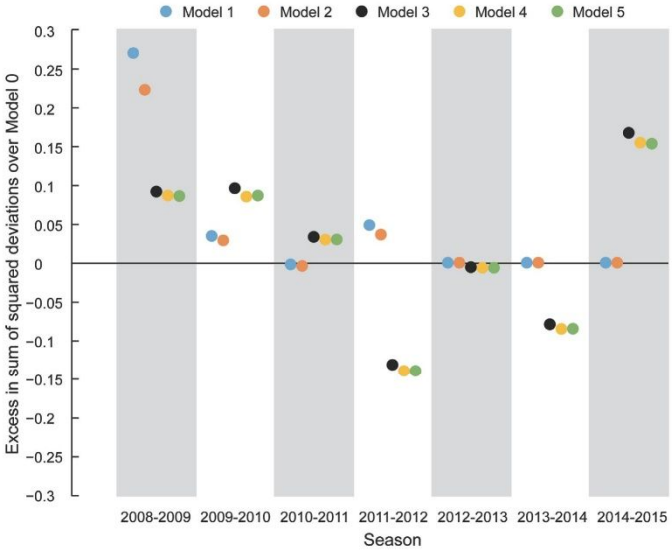


**Fig. 1. Excess in the sum of squared deviations for predicted proportions of genotypes from observed proportions in Models 1–5 over that in Model 0 in the seasons from 2008–2009 to 2014–2015.** The dots are spread horizontally in each column to enable them to be distinguished from each other.

strains of all genotypes, is an extension of Model 4. However, the results obtained with Model 5 are again similar to those obtained with Model 4 (Figure 1 and supplementary Tables S9 and S10), probably because norovirus strains belonging to different genotypes were too diverged to induce cross immunity (4). Our results indicate that Model 3 is the most appropriate of the five models for predicting genotype compositions in norovirus seasons in Japan.

### Performance of Models 1–5 in predicting the direction of change in the proportions of genotypes

A major difference between Models 1–5 and Model 0 is that the former models did predict the direction of change in the proportions of genotypes from the pre-target to the target season, whereas the latter model did not. Therefore, the predicted direction of change in the proportions of genotypes is indicated by the symbols ↑, ↓, and → for increase, decrease, and no change, respectively, and the former two symbols are colored green and blue when the direction has been predicted correctly and incorrectly, respectively (supplementary Tables S5–S10). In addition, the rates of correct prediction in Models 1–5 are plotted in Figure 3. Here the focus is particularly on the performance of Model 3 because this model was considered to be the most appropriate of the five models for predicting the genotype compositions. It was found that Model 3 often predicted the direction correctly even when the predicted proportion of genotype was further from the observed proportion than the predicted proportion in Model 0 (Figure 3 and supplementary Table S8). In total, the direction was predicted correctly in 49 cases, which is significantly more than the number of incorrect predictions of direction (26 cases) ($P = 0.0106$ by the two-tailed binomial test) over the seasons from 2008–2009 to 2014–2015 (supplementary Table S8). These results further support the effectiveness of Model 3 in predicting genotype compositions in norovirus seasons in Japan.

## DISCUSSION

A fitness model for predicting the proportion of a strain in the target season from that in the pre-target season based on its fitness has been developed for influenza A virus. This model was developed by modeling the effects

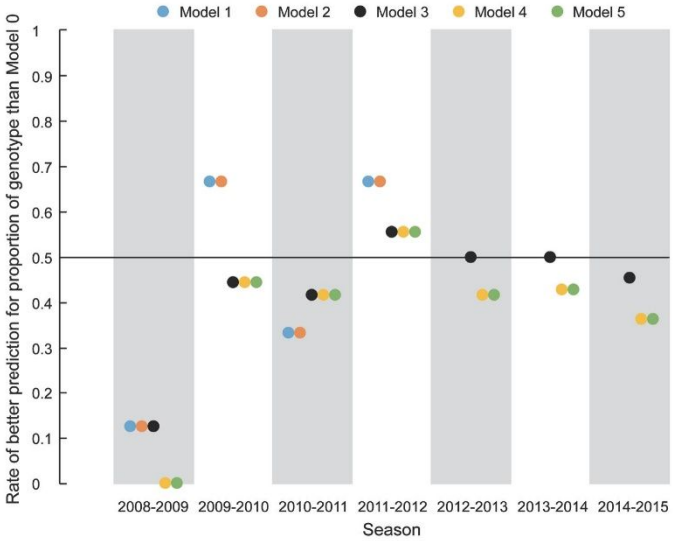


**Fig. 2.** Rate of better prediction of the proportions of genotypes in Models 1–5 than in Model 0 in the seasons from 2008–2009 to 2014–2015. There are no data points for Models 1 and 2 in the seasons from 2012–2013 to 2014–2015. The dots are spread horizontally in each column to enable them to be distinguished from each other.

of herd immunity to and thermodynamic stability of HA using information on the amino acid sites involved in neutralization epitopes, receptor binding sites and so on (19, 20). Since some information such as receptor binding sites is not available for norovirus (28), the fitness model for predicting genotype compositions in norovirus seasons in Japan in the present study takes into account only the fitness effect of herd immunity.

Of the five models with varying assumptions concerning the magnitude of cross immunity between norovirus strains, Model 3 was found to be the most appropriate for predicting genotype composition. Of note, Model 3 is the closest of the five models to the original fitness model developed for influenza A virus, in which the magnitude of cross immunity is assumed to decline with differences in the amino acid sequence of HA but not with differences in the season of isolation between strains (19, 20). Although it was difficult to predict the proportions of genotypes accurately, Model 3 did predict the direction of change in the proportions of genotypes relatively accurately. This model can be used to predict the genotype composition of norovirus in future seasons (4); it has predicted that GII.3 and GII.4 may contract in the season 2015–2016 (supplementary Table S8). Although the predicted proportions of genotypes should be viewed with caution, GII.17 is predicted to expand and predominate in this season (supplementary table S8) (29, 30, 31). Models 4 and 5 also yielded similar results (supplementary Tables S9 and S10).

The procedure for predicting genotype compositions in norovirus seasons described in the present study has been implemented in the norovirus forecasting system (NOROCAST) (32). The system should be updated as concerning genotype compositions and the VP1 sequences of norovirus accumulate. In addition, the system should be upgraded by improving the model to predict the genotype compositions more accurately. It may be interesting to incorporate into the model the fitness effects of the interactions of genetic factors of humans such as histo-blood group antigen binding specificity (33) and environmental factors such as temperature (34) and humidity (35) with specific amino acid residues in VP1 as well as other proteins of norovirus such as RNA-dependent RNA polymerase. Predicting genotype compositions in norovirus seasons as described in the present study may be useful for formulating the content of genotypes in polyvalent vaccines against norovirus.

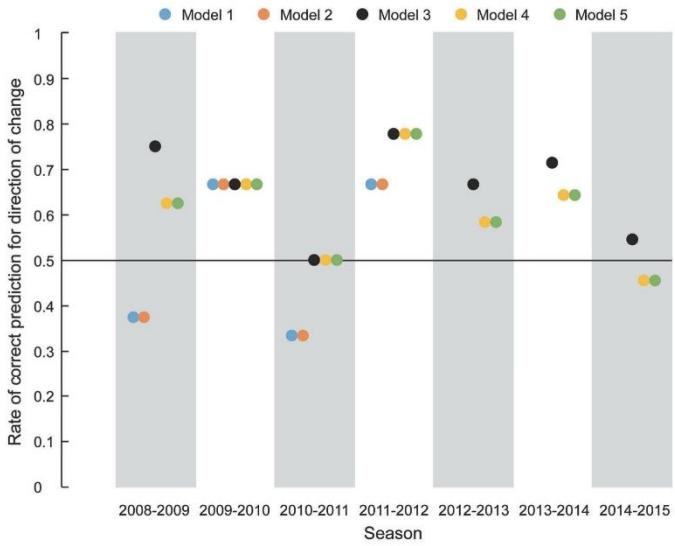


**Fig. 3.** Rate of correct prediction for the direction of change in the proportions of genotypes in Models 1–5 in the seasons from 2008–2009 to 2014–2015. There are no data points for Models 1 and 2 in the seasons from 2012–2013 to 2014–2015. The dots are spread horizontally in each column to enable them to be distinguished from each other.

## ACKNOWLEDGMENTS

The authors thank the following researchers in Japan for their cooperation in the surveillance of norovirus strains circulating in Japan: Shima Yoshizumi at Hokkaido Institute of Public Health; Sayaka Kogawa at Aomori Prefectural Public Health and Environment Center; Yo Ueki at Miyagi Prefectural Institute of Public Health and Environment; Takumi Motoya at Ibaraki Prefectural Institute of Public Health; Fuminori Mizukoshi at Tochigi Prefectural Institute of Public Health and Environmental Science; Hiroyuki Tsukagoshi, Miho Kobayashi, Yoshiko Sasaki and Kunihisa Kozawa at Gunma Prefectural Institute of Public Health and Environmental Sciences; Michiyo Shinohara at Saitama Institute of Public Health; Rieko Suzuki at Kanagawa Prefectural Institute of Public Health; Hideaki Shimizu, Mariko Ishikawa and Nobuhiko Okabe at Kawasaki City Institute for Public Health; Shinichiro Shibata and Akari Kodaira at Nagoya City Public Health Research Institute; Tomoko Takahashi at Iwate Prefectural Meat Inspection Center; Akira Iwakiri at Miyazaki Prefecture Kobayashi Meat Inspection Center; Makoto Kuroda, Tsuyoshi Sekizuka and Akifumi Yamashita at National Institute of Infectious Diseases; Keiji Sahara at Shizuoka Institute of Environment and Hygiene; Naomi Tanaka-Sakon at Osaka Prefectural Institute of Public Health; Naoki Shigemoto at Hiroshima Prefectural Technology Research Institute; Reiko Okamoto at Yamaguchi Prefectural Institute of Public Health and Environment; Seiya Harada and Kenta Yoshioka at Kumamoto Prefectural Institute of Public-Health and Environmental Science; and Yasutaka Yamashita at Ehime Prefectural Institute of Public Health and Environmental Science. The authors are also indebted to two anonymous reviewers for valuable comments. This work was supported by the Japan Agency for Medical Research and Development, AMED.

## DISCLOSURE

The authors declare no conflicts of interest.

## REFERENCES

1. Kapikian A.Z., Wyatt R.G., Dolin R., Thornhill T.S., Kalica A.R., Chanock R.M. (1972) Visualization by immune electron microscopy of a particle associated with acute infectious nonbacterial gastroenteritis. *J Virol* **10**: 1075–81.
2. Ahmed S.M., Hall A.J., Robinson A.E., Verhoef L., Premkumar P., Parashar U.D., Koopmans M., Lopman B.A. (2014) Global prevalence of norovirus in cases of gastroenteritis: A systematic review and meta-analysis. *Lancet Infect Dis* **14**: 725–30.

3. Patel M.M., Widdowson M-A., Glass R.I., Akazawa K., Vinje J., Parashar U.D. (2008) Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg Infect Dis* **14**: 1224–31.
4. Lopman B. (2015) Global burden of norovirus and prospects for vaccine development. *Centers for Disease Control and Prevention*.
5. Clarke I.N., Estes M.K., Green K.Y., Hansman G.S., Knowles N.J., Koopmans M.K., Matson D.O., Meyers G., Neill J.D., Radford A., Smith A.W., Studdert M.J., Thiel H-J., Vinje J. (2012) Caliciviridae. In King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J., eds. *Virus Taxonomy: the Classification and Nomenclature of Viruses. The Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego: Elsevier Academic Press, pp. 977–86.
6. Prasad B.V.V., Hardy M.E., Dokland T., Bella J., Rossmann M.G., Estes M.K. (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**: 287–90.
7. Jiang X., Wang M., Wang K., Estes M.K. (1993) Sequence and genomic organization of Norwalk virus. *Virology* **195**: 51–61.
8. Lambden P.R., Caul E.O., Ashley C.R., Clarke I.N. (1993) Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259**: 516–9.
9. Hansman G.S., Natori K., Shirato-Horikoshi H., Ogawa S., Oka T., Katayama K., Tanaka T., Miyoshi T., Sakae K., Kobayashi S., Shinohara M., Uchida K., Sakurai N., Shinozaki K., Okada M., Seto Y., Kamata K., Nagata N., Tanaka K., Miyamura T., Takeda N. (2006) Genetic and antigenic diversity among noroviruses. *J Gen Virol* **87**: 909–19.
10. Kocher J., Yuan L. (2015) Norovirus vaccines and potential antinorovirus drugs: recent advances and future perspectives. *Future Virol* **10**: 899–913.
11. Zheng D.-P., Ando T., Fankhauser R.L., Beard R.S., Glass R.I., Monroe S.S. (2006) Norovirus classification and proposed strain nomenclature. *Virology* **346**: 312–23.
12. Vinje J. (2015) Advances in laboratory methods for detection and typing of norovirus. *J Clin Microbiol* **53**: 373–81.
13. Katayama K., Shirato-Horikoshi H., Kojima S., Kageyama T., Oka T., Hoshino F.B., Fukushi S., Shinohara M., Uchida K., Suzuki Y., Gojobori T., Takeda N. (2002) Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**: 225–39.
14. Kroneman A., Vega E., Vennema H., Vinje J., White P.A., Hansman G., Green K., Martella V., Katayama K., Koopmans M. (2013) Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol* **158**: 2059–68.
15. Thongprachum A., Khamrin P., Maneekarn N., Hayakawa S., Ushijima H. (2016) Epidemiology of gastroenteritis viruses in Japan: prevalence, seasonality, and outbreak. *J Med Virol* **88**: 551–70.
16. Kobayashi M., Yoshizumi S., Kogawa S., Takahashi T., Ueki Y., Shinohara M., Mizukoshi F., Tsukagoshi H., Sasaki Y., Suzuki R., Shimizu H., Iwakiri A., Okabe N., Shirabe K., Shinomiya H., Kozawa K., Kusunoki H., Ryo A., Kuroda M., Katayama K., Kimura H. (2015) Molecular evolution of the capsid gene in norovirus genogroup I. *Sci Rep* **5**: 13806.
17. Lindesmith L.C., Donaldson E.F., Baric R.S. (2011) Norovirus GII.4 strain antigenic variation. *J Virol* **85**: 231–42.
18. Simmons K., Gambhir M., Leon J., Lopman B. (2013) Duration of immunity to norovirus gastroenteritis. *Emerg Infect Dis* **19**: 1260–7.
19. Luksza M., Lassig M. (2014) A predictive fitness model for influenza. *Nature* **507**: 57–61.
20. Suzuki Y. (2015) Selecting vaccine strains for H3N2 human influenza A virus. *Meta Gene* **4**: 64–72.
21. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–10.
22. Kroneman A., Vennema H., Deforche K., Avoort H.V., Penaranda S., Oberste M.S., Koopmans V.J. (2011) An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* **51**: 121–5.
23. Katoh K., Misawa K., Kuma K.-I., Miyata T. (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–66.
24. Tomita M., Hashimoto K., Takahashi K., Matsuzaki Y., Matsushima R., Saito K., Yugi K., Miyoshi F., Nakano H., Tanida S., Saito Y., Kawase A., Watanabe N., Shimizu T., Nakayama Y. (2000) The E-CELL project: towards integrative simulation of cellular processes. *N Gener Comput* **18**: 1–12.
25. Suzuki Y. (2013) Predictability of antigenic evolution for H3N2 human influenza A virus. *Genes Genet Syst* **88**: 225–32.
26. Kimura M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, New York, Melbourne: Cambridge University Press.
27. Nei M. (2013) *Mutation-Driven Evolution*. Oxford: Oxford University Press.
28. Murakami K., Kurihara C., Oka T., Shimoike T., Fujii Y., Takai-Todaka R., Park Y., Wakita T., Matsuda T., Hokari R., Miura S., Katayama K. (2013) Norovirus binding to intestinal epithelial cells is independent of histo-blood group antigens. *PLoS ONE* **8**:e66534.
29. Fu J., Ai J., Jin M., Jiang C., Zhang J., Shi C., Lin Q., Yuan Z., Qi X., Bao C., Tang F., Zhu Y. (2015) Emergence of a new GII.17 norovirus variant in patients with acute gastroenteritis in Jiangsu, China, September 2014 to March 2015. *Euro Surveill* **20:** 21157.
30. Lu J., Sun L., Fang L., Yang F., Mo Y., Lao J., Zheng H., Tan X., Lin H., Rutherford S., Guo L., Ke C., Hui L. (2015) Gastroenteritis outbreaks caused by norovirus GII.17, Guangdong Province, China, 2014–2015. *Emerg Infect Dis* **21:** 1240–42.
31. Matsushima Y., Ishikawa M., Shimizu T., Komane A., Kasuo S., Shinohara M., Nagasawa K., Kimura H., Ryo A., Okabe N., Haga K., Doan Y.H., Katayama K., Shimizu H. (2015) Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro Surveill* **20**: 21173.
32. NOROCAST: the norovirus forecasting system. [Cited 2 Feb 2016] Available from URL: http://www.nsc.nagoya-cu.ac.jp/~yossuzuk/norocast.html
33. Shanker S., Choi J.-M., Sankaran B., Atmar R.L., Estes M.K., Prasad B.V.V. (2011) Structural analysis of histo-blood group antigen binding specificity in a norovirus GII.4 epidemic variant: implications for epochal evolution. *J Virol* **85**: 8635–45.
34. Samandoulgou I., Hammami R., Morales-Rayas R., Fliss I., Jean J. (2015) Stability of secondary and tertiary structures of virus-like particles representing noroviruses: effects of pH, ionic strength, and temperature and implications for adhesion to surfaces. *Appl Environ Microbiol* **81**: 7680–6.
35. de la Noue A.C., Estienney M., Aho S., Perrier-Cornet J.-M., de Rougemont A., Pothier P., Gervais P., Belliot G. (2014) Absolute humidity influences the seasonal persistence and infectivity of human norovirus. *Appl Environ Microbiol* **80**: 7196–205.


## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site

**Table S1**. Numbers of norovirus strains isolated in Japan in the seasons from 2006–2007 to 2014–2015 summarized from IASR of NIID, Japan.

**Table S2**. Observed genotype compositions of norovirus in Japan in the seasons from 2006–2007 to 2014–2015.
**Table S3**. The INSD accession numbers of norovirus sequences analyzed in the present study.
**Table S4**. Numbers of norovirus strains whose VP1 sequences were analyzed in the present study.
**Table S5**. Predicted genotype compositions of norovirus in Japan using Model 0 in the seasons from 2008–2009 to 2015–2016.
**Table S6**. Predicted genotype compositions of norovirus in Japan using Model 1 in the seasons from 2008–2009 to 2015–2016.
**Table S7**. Predicted genotype compositions of norovirus in Japan using Model 2 in the seasons from 2008–2009 to 2015–2016.
**Table S8**. Predicted genotype compositions of norovirus in Japan using model 3 in the seasons from 2008–2009 to 2015–2016.
**Table S9**. Predicted genotype compositions of norovirus in Japan using Model 4 in the seasons from 2008–2009 to 2015–2016.
**Table S10**. Predicted genotype compositions of norovirus in Japan using Model 5 in the seasons from 2008–2009 to 2015–2016.