Gene Reports 26 (2022) 101492

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Estimating antigenic distances between GII.4 human norovirus strains

Yoshiyuki Suzuki *

Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya-shi, Aichi-ken 467-8501, Japan

ARTICLE INFO

Edited by Hee Jeong Im Sampen

Keywords:
Antigenic distance
Human norovirus
Protruding domain
Shell domain
Viral protein 1

ABSTRACT

Human norovirus (HuNoV) is an etiological agent of acute gastroenteritis. Viral protein 1 (VP1) is the major capsid protein and the determinant of antigenicity in HuNoV. GII.4 has been the dominant VP1-genotype in most of HuNoV seasons. Here an attempt was made to estimate the magnitude of antigenic difference, defined as the antigenic distance, between GII.4 HuNoV strains from the comparison of amino acid sequences for VP1. Antigenic distance was estimated more accurately assuming an exponential decline in the amount of increase in the antigenic distance along with an increment in the number of amino acid differences than assuming a constant increase. Although antigenic sites A, C, D, E, and G appeared to be the major determinant of antigenicity, estimation accuracy was sometimes improved by adding other regions. These results suggested that it may be suitable to construct a model based on the entire region of VP1 for estimating antigenic distances between GII.4 HuNoV strains.

Norovirus (NoV) constitutes the genus *Norovirus* in the family *Caliciviridae* (Clarke et al., 2012). The NoV virion is non-enveloped and icosahedral with the size of 38 nm in diameter. The NoV genome is a linear, non-segmented, single-stranded RNA of positive polarity with the length of 7.5–7.7 kb, containing ORF1-ORF3 and occasionally ORF4 (Lambden et al., 1993). ORF1-ORF4 encode non-structural proteins (NSs), viral protein 1 (VP1), viral protein 2 (VP2), and virulence factor 1 (VF1), respectively.

The NoV capsid is composed of 90 dimers of VP1 (Prasad et al., 1999). VP1 consists of the shell (S) and protruding (P) domains, which form the core and the spike of the capsid, respectively. The P domain comprises the P1 and P2 subdomains, which form the proximal and distal parts of the spike, respectively. NoV is divided into genogroups GI-GX based on the similarity in the amino acid sequence of VP1. GI, GII, GIV, GVIII, and GIX NoVs infect humans and thus constitute human norovirus (HuNoV) (Chhabra et al., 2019).

HuNoV is an etiological agent of acute gastroenteritis (AGE) (Kapikian et al., 1972). HuNoV causes 685 million cases of AGE and 212 thousand deaths worldwide annually (Lopman, 2015; Pires et al., 2015). The prevalence of HuNoV among the cases of AGE is greater in developed countries (20%) and low-mortality developing countries (19%) than in high-mortality developing countries (14%), suggesting that improvements in sanitation and hygiene may not be sufficient for control and prevention of HuNoV (Ahmed et al., 2014). It is therefore demanded to develop vaccines against HuNoV (Ford-Siltz et al., 2021).

Based on the similarity in the amino acid sequence of VP1, GI, GII, and GIV HuNoVs are further divided into VP1-genotypes GI.1–GI.9, GII.1–GII.27, and GIV.1 and GIV.2, respectively (Kroneman et al., 2013; Vinje, 2015). HuNoVs with various VP1-genotypes co-circulate every season changing their proportions (Thongprachum et al., 2016), which may be driven by the herd immunity in humans (Ruis et al., 2020) and the genomic recombination in HuNoVs (Suzuki, 2021).

GII.4 has been the dominant VP1-genotype in most of HuNoV seasons (Farahmand et al., 2022). Since the mid-1990s, GII.4 HuNoVs caused six pandemics named Grimsby 1995, Farmington Hills 2002, Hunter 2004, Den Haag 2006b, New Orleans 2009, and Sydney 2012, and five epidemics named Lanzhou 2002, Sakai 2003, Yerseke 2006a, Osaka 2007, and Apeldoorn 2007. Pandemic and epidemic GII.4

Abbreviations: α, amount of increase in the antigenic distance; α_1, amount of increase in the antigenic distance by the first amino acid difference; AGE, acute gastroenteritis; d_{ij}, antigenic distance between strains *i* and *j*; HBGA, histo-blood group antigen; HuNoV, human norovirus; INSD, International Nucleotide Sequence Database; *m*, number of antigenic distances; *n*, number of amino acid differences; NoV, norovirus; NS, non-structural protein; P, protruding; P2ACDEG, antigenic sites A, C, D, E, and G; P2ΔACDEG, P2 subdomain excluding antigenic sites A, C, D, E, and G; *r*, decreasing rate of the amount of increase in the antigenic distance; RdRp, RNA-dependent RNA polymerase; RMSE, root mean square error; S, shell; $t_{B(ij)}$, HBGA blockade titer for strains *i* and *j*; $\bar{t}_{B(ij)}$, geometric mean of $t_{B(ij)}$; $t_{N(ij)}$, neutralization titer for strains *i* and *j*; VF1, virulence factor 1; VLP, virus-like particle; VP1, viral protein 1; VP2, viral protein 2.

* Corresponding author at: Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya, Aichi 467-8501, Japan.
E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp.

https://doi.org/10.1016/j.genrep.2021.101492
Received 1 October 2021; Received in revised form 17 November 2021; Accepted 1 December 2021
Available online 10 January 2022

HuNoVs can be antigenically distinguished, which may be attributed to differences in amino acids at antigenic sites A, C, D, E, and G (Ford-Siltz et al., 2021). In addition, positive selection has been detected at the antigenic sites, indicating that GII.4 HuNoV may have been escaping from the herd immunity through antigenic evolution (Motoya et al., 2017). These observations suggested that the vaccine against GII.4 HuNoV may have to be re-formulated periodically for matching the antigenicity of the seed strain to that of the prevalent strain in the target season.

In the re-formulation of vaccines, selection of seed strains may be facilitated by predicting antigenic evolution of GII.4 HuNoV (Luksza and Lassig, 2014; Suzuki, 2015). Antigenic evolution of viruses may be predicted through evaluation of probabilities for occurrence of possible mutations as well as their effects on viral productivity and antigenicity. Possible mutations with a high probability of occurrence raising productivity and changing antigenicity may be predicted to drive antigenic evolution (Suzuki, 2013). Here it is required to quantify effects of possible mutations on antigenicity. The purpose of the present study was to make an attempt to estimate the magnitude of antigenic difference between GII.4 HuNoV strains from the comparison of amino acid sequences for VP1.

The magnitude of antigenic difference between viral strains i and j is usually quantified using the titer of anti-serum raised against strain i to neutralize the infectivity of strain j (neutralization titer: $t_{N(ij)}$) (Lau et al., 2021). The antigenic distance between strains i and j (d_{ij}) can be defined as

$$d_{ij} = log_2 \frac{t_{N(ii)}}{t_{N(ij)}}$$

(Lees et al., 2010).

In HuNoV, infection is initiated with the binding of the P2 subdomain to the cellular receptor, which is as yet unidentified, as well as the histo-blood group antigens (HBGAs), which serve as the attachment factor (Lindesmith et al., 2003). Although the cell culture system for measuring the neutralization titer is of limited use in HuNoV, it has been demonstrated that the titer of anti-serum raised against the virus-like particle (VLP) of strain i to block the HBGA binding activity of the VLP of strain j (HBGA blockade titer: $t_{B(ij)}$) can be used as a surrogate for $t_{N(ij)}$ (Atmar et al., 2020; Ford-Siltz et al., 2020).

The EC50 HBGA blockade titers have been measured for 24 representative strains of 6 pandemic and 4 epidemic GII.4 HuNoVs in Kendra et al. (2021) (Supplementary Table S1). Each of anti-sera raised against the VLPs of 10 strains was used for measuring the HBGA blockade titers against the VLPs of the same ($t_{B(ii)}$) and different ($t_{B(ij)}$) strains in four replicates. In the present study, the antigenic distance was measured based on the geometric means of $t_{B(ii)}$ ($\bar{t}_{B(ii)}$) and $t_{B(ij)}$ ($\bar{t}_{B(ij)}$) over the four replicates. Here the titer below the detection limit (<50) was regarded to be 50. The antigenic distance between strains i and j was obtained as (Supplementary Table S2).

$$d_{ij} = log_2 \frac{\bar{t}_{B(ii)}}{\bar{t}_{B(ij)}}$$

VP1 of GII.4 HuNoV consists of 540 amino acids, in which positions 1–215 and 216–540 correspond to the S and P domains, respectively. Within the P domain, positions 216–280 and 416–540 constitute the P1 subdomain, whereas positions 281–415 constitute the P2 subdomain. Antigenic sites A, C, D, E, and G are located in the P2 subdomain; antigenic site A: positions 294–298, 368, 372, and 373; antigenic site C: positions 339–341 and 375–378; antigenic site D: positions 393–397; antigenic site E: positions 407 and 411–414; and antigenic site G: positions 352, 355–357, 359, and 364 (Tohma et al., 2019).

Nucleotide sequences encoding the entire region of VP1 (1620 nucleotides) for the 24 strains of GII.4 HuNoVs, which were used for obtaining antigenic distances above, were retrieved from the International Nucleotide Sequence Database (INSD) or kindly provided by Drs. Gabriel I. Parra and Kentaro Tohma at Center for Biologics Evaluation and Research, Food and Drug Administration (Kendra et al., 2021). The INSD accession numbers for the 24 nucleotide sequences are listed in Supplementary Table S1.

In the present study, two models (models 1 and 2) were considered for estimating antigenic distances between GII.4 HuNoV strains from the comparison of amino acid sequences for VP1. In model 1, the antigenic distance was assumed to increase by a constant amount along with an increment in the number of amino acid differences in the region under consideration. Thus, the antigenic distance between strains i and j was estimated by

$$\widehat{d}_{ij} = an,$$

where a and n denote the amount of increase in the antigenic distance along with an increment in the number of amino acid differences and the number of amino acid differences in the region under consideration, respectively. In model 2, the antigenic distance was assumed to increase by an exponentially decreasing amount along with an increment in the number of amino acid differences in the region under consideration. Thus, the antigenic distance between strains i and j was estimated by

$$\widehat{d}_{ij} = \frac{a_1[1-(1-r)^n]}{r},$$

where a_1, r, and n denote the amount of increase in the antigenic distance by the first amino acid difference, the decreasing rate of the amount of increase in the antigenic distance along with an increment in the number of amino acid differences, and the number of amino acid differences in the region under consideration, respectively. In each model, the antigenic distance was considered to be determined by amino acid differences in the entire region of VP1 (540 amino acids), S domain (215 amino acids), P domain (325 amino acids), P1 subdomain (190 amino acids), P2 subdomain (135 amino acids), P2 subdomain excluding antigenic sites A, C, D, E, and G (P2ΔACDEG) (104 amino acids), or antigenic sites A, C, D, E, and G (P2ACDEG) (31 amino acids).

To examine the performance of models 1 and 2 in estimating antigenic distances, 24 strains of GII.4 HuNoVs were split into two groups based on their phylogenetic relationships. One group was adopted as the training data set for optimizing parameter values in the model, and the other group was adopted as the test data set for evaluating estimates of antigenic distances obtained with the optimized parameter values in the model.

Multiple alignments for nucleotide sequences (1620 nucleotides) and amino acid sequences (540 amino acids) of VP1 were made for 24 strains of GII.4 HuNoVs using the computer program MAFFT (version 7.475) (Katoh et al., 2002). Both alignments did not contain any gaps (data not shown). Phylogenetic trees were constructed for nucleotide sequences and amino acid sequences by the neighbor-joining method (Saitou and Nei, 1987) with the p distance, which is known to produce reliable topologies when closely related sequences are analyzed (Nei and Kumar, 2000; Yoshida and Nei, 2016), using MEGA (version 10.1.8) (Kumar et al., 2018).

The topologies were slightly different between the phylogenetic trees for nucleotide sequences (Supplementary Fig. S1) and amino acid sequences (Supplementary Fig. S2). Based on the topology of the phylogenetic tree for nucleotide sequences (Supplementary Fig. S1), 24 strains were split into groups IA and IB; group IA included 12 strains with 44 antigenic distances and group IB included 12 strains with 66 antigenic distances. Similarly, based on the topology of the phylogenetic tree for amino acid sequences (Supplementary Fig. S2), 24 strains were split into groups IIA and IIB; group IIA included 11 strains with 50 antigenic distances and group IIB included 13 strains with 60 antigenic distances. Group IA and IB strains were used as the training and test data sets and *vice versa*, and group IIA and IIB strains were used as the training and test data sets and *vice versa*.

The parameter values in models 1 (a) and 2 (a_1 and r) were optimized

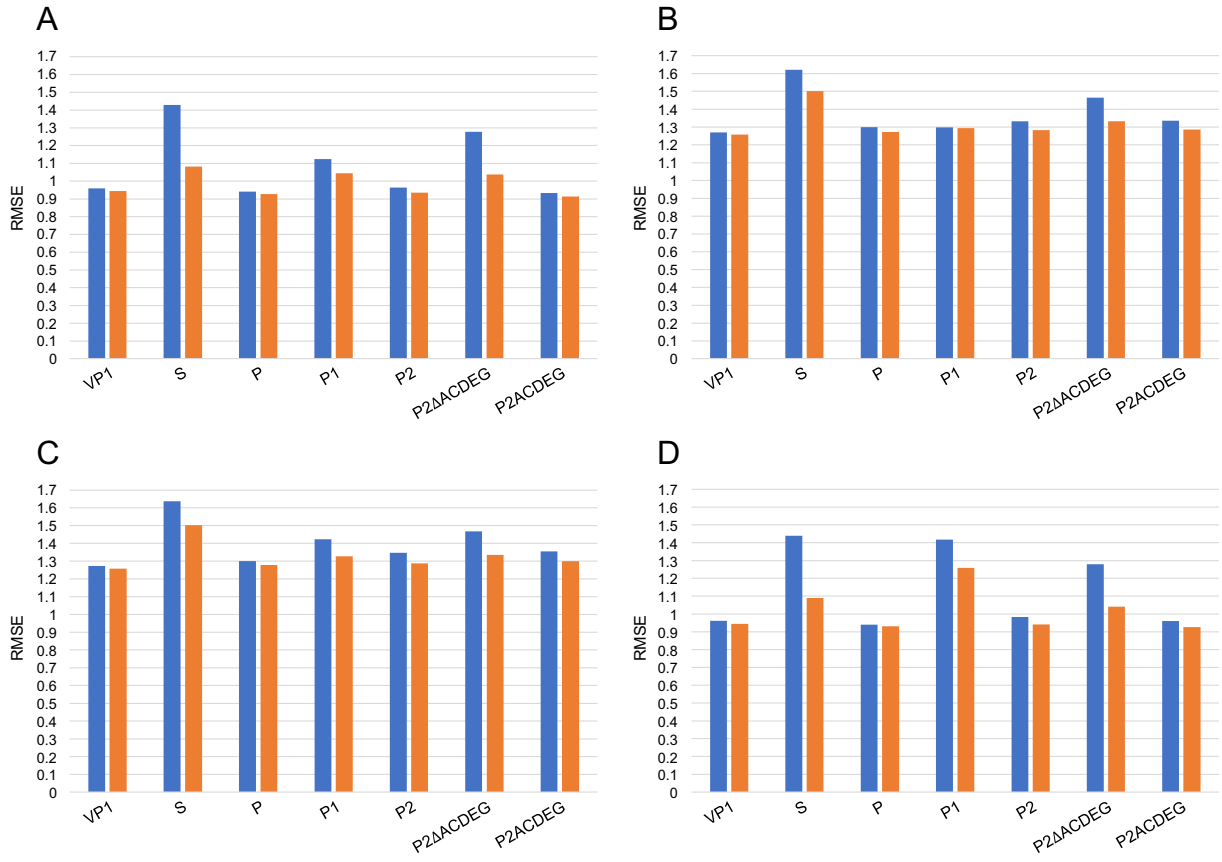

Fig. 1. The RMSE in the optimization of parameter values when group (A) IA and (B) IB strains were adopted as the training data set and in the evaluation of estimates for antigenic distances when group (C) IB and (D) IA strains were adopted as the test data set. The results obtained from models 1 and 2 are presented with blue and orange columns, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on the training data set by minimizing the root mean square error (RMSE) of estimates for antigenic distances between all pairs of strains *i* and *j* for which antigenic distances were available, that is

$$RMSE = \sqrt{\frac{\sum_{i,j}\left(d_{ij} - \widehat{d}_{ij}\right)^2}{m}},$$

where m denotes the number of antigenic distances available in the training data set. The optimization was accomplished with the genetic algorithm (Tomita et al., 2000), in which three random real numbers were set as the initial parameter values to confirm the convergence of the optimization (data not shown). Models 1 and 2 with optimized parameter values were applied to the test data set to estimate antigenic distances between all pairs of strains *i* and *j* for which antigenic distances were available. The RMSE of estimates for antigenic distances was computed for examining the performance of the models.

In the optimization of parameter values on the training data set, it was observed that the RMSE was always smaller in model 2 than in model 1 (Fig. 1; Supplementary Fig. S3). This result reflected the fact that the number of parameters was greater in model 2 than in model 1, and model 1 was nested in model 2 corresponding to the condition of $r = 0$. Among the regions analyzed, the S domain was mostly associated with the largest RMSE in both models 1 and 2. The only exception was the case when group IIA strains were adopted as the training data set in model 1, in which the P2ΔACDEG was associated with the largest RMSE (Fig. 1; Supplementary Fig. S3). The regions associated with the smallest RMSE were the same in models 1 and 2; the entire region of VP1, P domain, and P2ACDEG when group IB strains, group IIB strains, and group IA and IIA strains were adopted as the training data set, respectively (Fig. 1; Supplementary Fig. S3).

In the evaluation of estimates for antigenic distances on the test data set, it was again observed that the RMSE was always smaller in model 2 than in model 1 (Fig. 1; Supplementary Fig. S3). Also, among the regions analyzed, the S domain was again mostly associated with the largest RMSE in both models 1 and 2. However, the P1 subdomain and P2ΔACDEG were associated with the largest RMSE in the analyses of group IA strains in model 2 and group IIA strains in model 1, respectively (Fig. 1; Supplementary Fig. S3). Various regions were associated with the smallest RMSE; the entire region of VP1 in the analysis of group IB strains in models 1 and 2, the P domain in the analyses of group IA strains in model 1 and group IIB strains in models 1 and 2, the P2 subdomain in the analysis of group IIA strains in model 2, and the P2ACDEG in the analyses of group IA strains in model 2 and group IIA strains in model 1 (Fig. 1; Supplementary Fig. S3).

In the present study, an attempt was made to estimate antigenic distances between GII.4 HuNoV strains from the comparison of amino acid sequences for VP1. Model 2 appeared to be superior to model 1 in estimating antigenic distances, indicating that model 2 was more suitable than model 1 for describing antigenic evolution of GII.4 HuNoV. This result may be derived from the fact that the antigenic distance had an upper limit because the HBGA blockade titer below the detection limit (< 50) was regarded to be 50 in the present study. In addition, co-evolution among amino acids may be involved in antigenic evolution, which may generate a non-linear relationship between the number of amino acid differences and the antigenic distance (Kendra et al., 2021).

In the optimization of parameter values and the evaluation of estimates for antigenic distances, the regions associated with the smallest RMSE were the entire region of VP1, P domain, P2 subdomain, and P2ACDEG, which included the P2ACDEG. In contrast, the regions associated with the largest RMSE were the S domain, P1 subdomain, and

P2ΔACDEG, which did not include the P2ACDEG. These observations supported the idea that the P2ACDEG was the major determinant of antigenicity in GII.4 HuNoV (Ford-Siltz et al., 2021). Note that the number of variable amino acids was smaller in the S domain, P1 subdomain, and P2ΔACDEG than in the entire region of VP1, P domain, P2 subdomain, and P2ACDEG, which may also lead to a greater RMSE in the former regions than in the latter regions (Supplementary Table S3). The S domain was mostly associated with the largest RMSE, suggesting that the S domain may make the least contribution to antigenic evolution of GII.4 HuNoV (Alvarado et al., 2021). The P1 subdomain and P2ΔACDEG appeared to make a greater contribution compared to the S domain (Parra et al., 2016). Reportedly, however, even the S domain contains B-cell and T-cell epitopes (van Loben Seltz and Green, 2019). Indeed, the entire region of VP1, P domain, and P2 subdomain, which were associated with the smallest RMSE, comprise the S domain, P1 subdomain, and P2ΔACDEG in addition to the P2ACDEG. These results suggested that it may be suitable to construct a model based on the entire region of VP1 for estimating antigenic distances between GII.4 HuNoV strains. To attain a high estimation accuracy, the model may be designed to take into account not only the number of amino acid differences but also the physicochemical properties and the locations in the three-dimensional structure of different amino acids in the entire region of VP1 (Suzuki, 2013). The antigenic distances may be integrated with the information on the occurrence of genomic recombinations creating novel combinations of the VP1-genotype and the genotype for RNA-dependent RNA polymerase (RdRp-genotype) for predicting antigenic evolution of GII.4 HuNoV.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genrep.2021.101492.

CRediT authorship contribution statement

Yoshiyuki Suzuki: Conceptualization, Methodology, Investigation, Visualization, Writing.

Declaration of competing interest

The author declares no conflict of interest.

Acknowledgements

The author thanks two anonymous reviewers for valuable comments. This work was supported by JSPS KAKENHI Grant Number JP19K12221 and AMED Grant Number JP21fk0108120h0702 to Y.S.

References

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, 725–730.

Alvarado, G., Salmen, W., Ettayebi, K., Hu, L., Sankaran, B., Estes, M.K., Prasad, B.V.V., Drowe Jr., J.E., 2021. Broadly cross-reactive human antibodies that inhibit genogroup I and II noroviruses. Nat. Commun. 12, 4320.

Atmar, R.L., Ettayebi, K., Ayyar, B.V., Neill, F.H., Braun, R.P., Ramani, S., Estes, M.K., 2020. Comparison of microneutralization and histo-blood group antigen blocking assays for functional norovirus antibody detection. J. Infect. Dis. 221, 739–743.

Chhabra, P., de Graaf, M., Parra, G.I., Chan, M.C.-W., Green, K., Martella, V., Wang, Q., White, P.A., Katayama, K., Vennema, H., Koopmans, M.P.G., Vinje, J., 2019. Updated classification of norovirus genogroups and genotypes. J. Gen. Virol. 100, 1393–1406.

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.). Elsevier Academic Press, San Diego, pp. 977–986.

Farahmand, M., Moghoofei, M., Dorost, A., Shoja, Z., Ghorbani, S., Kiani, S.J., Khales, P., Esteghamati, A., Sayyahfar, S., Jafarzadeh, M., Minaeian, S., Khanaliha, K., Naghdalipour, M., Tavakoli, A., 2022. Global prevalence and genotype distribution of norovirus infection in children with gastroenteritis: a meta-analysis on 6 years of research from 2015 to 2020. Rev. Med. Virol. 32, e2237.

Ford-Siltz, L.A., Wales, S., Tohma, K., Gao, Y., Parra, G.I., 2020. Genotype-specific neutralization of norovirus is mediated by antibodies against the protruding domain of the major capsid protein. J. Infect. Dis. jiaa116.

Ford-Siltz, L.A., Tohma, K., Parra, G.I., 2021. Understanding the relationship between norovirus diversity and immunity. Gut Microbes 13, 1–13.

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, 1075–1081.

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, 3059–3066.

Kendra, J.A., Tohma, K., Ford-Siltz, L.A., Lepore, C.J., Parra, G.I., 2021. Antigenic cartography reveals complexities of genetic determinants that lead to antigenic differences among pandemic GII.4 noroviruses. Proc. Natl. Acad. Sci. U. S. A. 118, e2015874118.

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, 2059–2068.

Kumar, S., Stecher, G., Li, M., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

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, 516–519.

Lau, E.H.Y., Tsang, O.T.Y., Hui, D.S.C., Kwan, M.Y.W., Chan, W.-H., Chiu, S.S., Ko, R.L. W., Chan, K.H., Cheng, S.M.S., Perera, R.A.P.M., Cowling, B.J., Poon, L.L.M., Peiris, M., 2021. Neutralizing antibody titres in SARS-CoV-2 infections. Nat. Commun. 12, 63.

Lees, W.D., Moss, D.S., Shepherd, A.J., 2010. A computational analysis of the antigenic properties of haemagglutinin in influenza a H3N2. Bioinformatics 26, 1403–1408.

Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X.I., Lindblad, L., Stewart, P., LePendu, J., Baric, R., 2003. Human susceptibility and resistance to Norwalk virus infection. Nat. Med. 9, 548–553.

Lopman, B., 2015. Global Burden of Norovirus And Prospects for Vaccine Development. Centers for Disease Control and Prevention.

Luksza, M., Lassig, M., 2014. A predictive fitness model for influenza. Nature 507, 57–61.

Motoya, T., Nagasawa, K., Matsushima, Y., Nagata, N., Ryo, A., Sekizuka, T., Yamashita, A., Kuroda, M., Morita, Y., Suzuki, Y., Sasaki, N., Katayama, K., Kimura, H., 2017. Molecular evolution of the VP1 gene in human norovirus GII.4 variants in 1974–2015. Front. Microbiol. 8, 2399.

Nei, M., Kumar, S., 2000. Molecular Evolution And Phylogenetics. Oxford University Press, Oxford, New York.

Parra, G.I., Sosnovtsev, S.V., Abente, E.J., Sandoval-Jaime, C., Bok, K., Dolan, M.A., Green, K.Y., 2016. Mapping and modeling of a strain-specific epitope in the Norwalk virus capsid inner shell. Virology 492, 232–241.

Pires, S.M., Fischer-Walker, C.L., Lanata, C.F., Devleesschauwer, B., Hall, A.J., Kirk, M. D., Duarte, A.S.R., Black, R.E., Angulo, F.J., 2015. Aetiology-specific estimates of the global and regional incidence and mortality of diarrhoeal diseases commonly transmitted through food. PLoS ONE 10, e0142927.

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, 287–290.

Ruis, C., Lindesmith, L.C., Mallory, M.L., Brewer-Jensen, P.D., Bryant, J.M., Costantini, V., Monit, C., Vinje, J., Baric, R.S., Goldstein, R.A., Breuer, J., 2020. Preadaptation of pandemic GII.4 noroviruses in unsampled virus reservoirs years before emergence. Virus Evol. 6, veaa067.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Suzuki, Y., 2013. Predictability of antigenic evolution for H3N2human influenza A virus. Genes Genet. Syst. 88, 225–232.

Suzuki, Y., 2015. Selecting vaccine strains for H3N2 human influenza A virus. Meta Gene 4, 64–72.

Suzuki, Y., 2021. Effect of recombinations on changes in genotype proportions between norovirus seasons in Japan. Meta Gene 29, 100943.

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, 551–570.

Tohma, K., Lepore, C.J., Gao, Y., Ford-Siltz, L.A., Parra, G.I., 2019. Population genomics of GII.4 noroviruses reveal complex diversification and new antigenic sites involved in the emergence of pandemic strains. mBio 10, e02202–e02219.

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.

van Loben Seltz, J.M., Green, K.Y., 2019. The antigenic topology of norovirus as defined by B and T cell epitope mapping: implications for universal vaccines and therapeutics. Viruses 11, 432.

Vinje, J., 2015. Advances in laboratory methods for detection and typing of norovirus. J. Clin. Microbiol. 53, 373–381.

Yoshida, R., Nei, M., 2016. Efficiencies of the NJp, maximum likelihood, and Bayesian methods of phylogenetic construction for compositional and noncompositional genes. Mol. Biol. Evol. 33, 1618–1624.