Predicting Directions of Changes in Genotype Proportions Between Norovirus Seasons in Japan

Yoshiyuki Suzuki [1]*, Yen Hai Doan [2], Hirokazu Kimura [3], Hiroto Shinomiya [4], Komei Shirabe [5] and Kazuhiko Katayama [6]

[1] Graduate School of Natural Sciences, Nagoya City University, Nagoya, Japan, [2] Department of Virology II, National Institute of Infectious Diseases, Musashimurayama, Japan, [3] Graduate School of Health Science, Gunma Paz University, Takasaki, Japan, [4] Department of Microbiology, Ehime Prefecctural Institute of Public Health and Environmental Science, Matsuyama, Japan, [5] Division of Virology, Yamaguchi Prefectural Institute of Public Health and Environment, Yamaguchi, Japan, [6] Kitasato Institute for Life Sciences, Kitasato University, Minato, Japan

The norovirus forecasting system (NOROCAST) has been developed for predicting directions of changes in genotype proportions between human norovirus (HuNoV) seasons in Japan through modeling herd immunity to structural protein 1 (VP1). Here 404 nearly complete genomic sequences of HuNoV were analyzed to examine whether the performance of NOROCAST could be improved by modeling herd immunity to VP2 and non-structural proteins (NS) in addition to VP1. It was found that the applicability of NOROCAST may be extended by compensating for unavailable sequence data and observed genotype proportions of 0 in each season. Incorporation of herd immunity to VP2 and NS did not appear to improve the performance of NOROCAST, suggesting that VP1 may be a suitable target of vaccines.

Keywords: genotype proportion, herd immunity, norovirus, prediction, vaccine

INTRODUCTION

Norovirus (NoV) is an etiological agent of acute gastroenteritis in humans (Kapikian et al., 1972), causing 18% of all cases (Ahmed et al., 2014) and 200,000 deaths worldwide annually (Patel et al., 2008). The human NoV (HuNoV) is more prevalent in developed countries (20%) and low-mortality developing countries (19%) than in high-mortality developing countries (14%) (Ahmed et al., 2014). Therefore, improvement in sanitation and hygiene may not be sufficient for control and prevention of HuNoV, and development of vaccines is demanded.

HuNoV belongs to the genus *Norovirus* in the family *Caliciviridae* (Clarke et al., 2012). The virion of HuNoV is non-enveloped, icosahedral, and 38 nm in diameter (Prasad et al., 1999). The genome of HuNoV is a linear, non-segmented, single-stranded, positive-sense RNA of 7.5 kb, containing ORF1-ORF3 (Jiang et al., 1993; Lambden et al., 1993). ORF1 encodes non-structural proteins (NS), whereas ORF2 and ORF3 encode structural proteins 1 (VP1) and 2 (VP2), respectively.

Based on the similarity in the amino acid sequence of VP1, NoV is divided into genogroups GI-GVII (Zheng et al., 2006; Vinje, 2015). HuNoV consists of GI, GII, and GIV strains, which are further classified into genotypes GI.1-GI.9, GII.1-GII.22, and GIV.1 and GIV.2, respectively (Kroneman et al., 2013; Vinje, 2015). Multiple genotypes of HuNoV co-circulate every season changing genotype proportions (Suzuki et al., 2016; Thongprachum et al., 2016).

Different genotypes of HuNoV appeared to be antigenically related to each other with various degrees (Hansman et al., 2006) and have been grouped into several immunotypes (Parra et al., 2017). In addition, the HuNoV infection appeared to be mediated via person-to-person, food-borne, and water-borne transmissions with some preferences depending on the genotypes (Mathijs et al., 2012). It may therefore be useful to predict changes in genotype proportions between HuNoV seasons for determining the genotypes to be included in vaccines and the human populations to be vaccinated in each season.

The norovirus forecasting system (NOROCAST) has been developed for predicting changes in genotype proportions between HuNoV seasons in Japan (Suzuki et al., 2016). The herd immunity to VP1 was taken into account based on the fitness model that was originally proposed for describing changes in proportions of influenza A virus strains (Luksza and Lassig, 2014; Suzuki, 2015). The purpose of the present study was to examine whether the performance of NOROCAST could be improved by taking into account the herd immunity to VP2 and NS in addition to VP1.

METHOD

In Japan, the HuNoV season is defined such that season 2006/2007 represents the period between September of 2006 and August of 2007. Genotype frequencies in HuNoV seasons in Japan have been surveyed by the National Institute of Infectious Diseases (NIID), Japan, and deposited in the Infectious Agents Surveillance Report (IASR). The observed genotype frequencies were available for the analysis from season 2006/2007 to season 2016/2017 (**Supplementary Table S1**), which were converted into the observed genotype proportions (**Supplementary Table S2**). Multiple genotypes of GI and GII strains but no GIV strains were observed in these seasons.

Nearly complete genomic sequences of GI and GII strains containing the entire coding regions of VP1, VP2, and NS provided with the information on the isolation year and month, were retrieved from the International Nucleotide Sequence Database (INSD). The INSD accession numbers for the 404 sequences analyzed in the present study were listed in **Supplementary Table S3**. These sequences were classified according to the isolation season and the VP1 genotype, which was determined by the Norovirus Genotyping Tool (version 1.0) (Kroneman et al., 2011) (**Supplementary Table S4**). It appeared that the sequence data were biased toward more prevalent genotypes and were sometimes unavailable for less prevalent genotypes in each season. For each of VP1, VP2, and NS, multiple alignment of 404 amino acid sequences was made using the computer program MAFFT (version 7.215) (Katoh et al., 2002).

In the fitness model (Luksza and Lassig, 2014), the proportion of strain i in the target season t ($p_{i(t)}$) was predicted from that in the season immediately before the target season (pre-target season) $t - 1$ ($p_{i(t-1)}$) by

$$\hat{p}_{i(t)} = p_{i(t-1)} \times e^{f_i}, \tag{1}$$

where f_i is the fitness of strain i. In NOROCAST (Suzuki et al., 2016), f_i was defined as

$$f_i = f_0 - c_i, \tag{2}$$

where f_0 is a constant to ensure $\sum_i \hat{p}_{i(t)} = 1$. The term c_i denotes the effect of herd immunity to VP1 of strain i elicited by strain j's that have circulated from season 2006/2007 to the pre-target season, namely

$$c_i = \sum_j k_{VP1} \times \left(1 - l_{VP1}\right)^{d_{ij(VP1)}} \times 0.8^{t_i - t_j + 1} \times p_{j(t_j)}. \tag{3}$$

Here k_{VP1} is the coefficient of herd immunity to VP1, which was assumed to decline by fraction l_{VP1} for each of amino acid differences in VP1 between strains i and j designated as $d_{ij(VP1)}$ (Lindesmith et al., 2011). The terms t_i and t_j denote the isolation seasons of strains i and j, respectively. The herd immunity was assumed to decline 20% each season, so that the expected duration of human immunity was roughly 5 years (Simmons et al., 2013).

In the present study, the herd immunity to VP2 and NS was also taken into account, so that c_i was re-defined as

$$c_i = \sum_j \left[k_{VP1} \times \left(1 - l_{VP1}\right)^{d_{ij(VP1)}} + k_{VP2} \times \left(1 - l_{VP2}\right)^{d_{ij(VP2)}} \right. $$
$$\left. + k_{NS} \times \left(1 - l_{NS}\right)^{d_{ij(NS)}} \right] \times 0.8^{t_i - t_j + 1} \times p_{j(t_j)}. \tag{4}$$

Here k_{VP2}, l_{VP2}, and $d_{ij(VP2)}$ and k_{NS}, l_{NS}, and $d_{ij(NS)}$ are parameters for VP2 and NS, respectively, corresponding to k_{VP1}, l_{VP1}, and $d_{ij(VP1)}$ for VP1, respectively. For predicting genotype proportions in the target season, parameters were optimized using the observed genotype proportions from season 2006/2007 to the pre-target season. Optimization was performed with the genetic algorithm (Tomita et al., 2000) through minimizing the sum of squared deviations of the predicted genotype proportions from the observed genotype proportions over the seasons from season 2008/2009 to the pre-target season. Three random real numbers were assigned as the initial parameter values to confirm the convergence of the optimization (data not shown) (Suzuki, 2013, 2015). The optimized parameter values were adopted for predicting directions of changes in genotype proportions from the pre-target season to the target season.

The performance of NOROCAST was evaluated retrospectively by predicting directions of changes in genotype proportions setting the target seasons to be from season 2008/2009 to season 2016/2017. It should be noted that in NOROCAST (Suzuki et al., 2016), genotype proportions could not be predicted or were always predicted to be 0 when the sequence data were unavailable or the observed genotype proportions were 0 in the pre-target season, respectively. Therefore, the prediction was considered to be correct when both the observed and the predicted directions of changes in genotype proportions (increase or decrease) were the same, and incorrect when they were different. In the present study, unavailable sequence data in a particular

TABLE 1 | Numbers of observed and predicted directions of changes in genotype proportions setting the target seasons to be from season 2008/2009 to season 2016/2017.

Protein	Arbitrary value	Predicted	Observed	
			Increase	Decrease
VP1	0	Increase	51	59
		Decrease	3	15
	5×10^{-6}	Increase	71	59
		Decrease	3	15
	5×10^{-5}	Increase	71	59
		Decrease	3	15
	5×10^{-4}	Increase	71	59
		Decrease	3	15
VP1, VP2, and NS	0	Increase	49	59
		Decrease	5	15
	5×10^{-6}	Increase	69	58
		Decrease	5	16
	5×10^{-5}	Increase	70	58
		Decrease	4	16
	5×10^{-4}	Increase	71	58
		Decrease	3	16

season were compensated by the available sequence data in the closest season before or after that season or both within the same genotype (**Supplementary Table S5**). The observed genotype proportions of 0 were also replaced with an arbitrary value (5×10^{-6}, 5×10^{-5}, or 5×10^{-4}) that was smaller than 5.099×10^{-4} ($= \frac{1}{1961}$), which was the lowest positive value obtainable in **Supplementary Table S2**.

RESULTS

The accuracy of NOROCAST in predicting directions of changes in genotype proportions was examined setting the target seasons to be from season 2008/2009 to season 2016/2017 (**Supplementary Tables S6–S14**). When only the herd immunity to VP1 was taken into account without replacing the observed genotype proportions of 0, the numbers of correct and incorrect predictions were 66 and 62, respectively, yielding the accuracy of 0.516 ($P = 0.724$ by the χ^2 test) (**Table 1**). However, by replacing the observed genotype proportions of 0 with 5×10^{-6}, 5×10^{-5}, or 5×10^{-4}, directions were predicted correctly and incorrectly in 86 and 62 cases, respectively, so that the accuracy was 0.581 ($P = 0.0485$ by the χ^2 test). The accuracy was elevated mainly because changes in genotype proportions became predictable even when the observed genotype proportions were 0 in the pre-target season (**Table 1**).

However, the accuracy of NOROCAST did not appear to be improved by taking into account the herd immunity to VP2 and NS in addition to VP1 (**Supplementary Tables S6–S14**). That is, without replacing the observed genotype

proportions of 0, both the numbers of correct and incorrect predictions were 64, yielding the accuracy of 0.5 ($P = 1$ by the χ^2 test) (**Table 1**). By replacing the observed genotype proportions of 0 with 5×10^{-6}, 5×10^{-5}, or 5×10^{-4}, directions were predicted correctly and incorrectly in 85–87 and 61–63 cases, respectively, so that the accuracy was 0.574–0.588 ($P = 0.0326$–0.0705 by the χ^2 test) (**Table 1**).

It should be noted that the direction of change in genotype proportion was always (9 out of 9 cases with VP1; $P = 0.00391$ by the binomial test) or almost always (8 out of 9 cases with VP1, VP2, and NS; $P = 0.0352$ by the binomial test) predicted correctly for GII.4, which was the most prevalent genotype in HuNoV (**Supplementary Tables S6–S14**). Interestingly, GII.4 has been classified as the evolving genotype, whereas other genotypes have been classified as the static genotypes (Parra et al., 2017). It is possible that GII.4 strains may be influenced by the herd immunity more strongly than the strains of other genotypes, which may have resulted in faster evolution and higher predictability of the direction of change in genotype proportion for GII.4 than other genotypes.

DISCUSSION

In the previous study (Suzuki et al., 2016), NOROCAST was developed for predicting directions of changes in genotype proportions between HuNoV seasons in Japan modeling the herd immunity to VP1. From the analysis of 1458 VP1 sequences setting the target seasons to be from season 2008/2009 to season 2014/2015, directions were predicted correctly and incorrectly in 45 and 30 cases, respectively, yielding the accuracy of 0.6 ($P = 0.0833$ by the χ^2 test). In the present study, the effect of modeling herd immunity to VP2 and NS in addition to VP1 was examined from the analysis of 404 nearly complete genomic sequences of HuNoV.

When only the herd immunity to VP1 was taken into account without replacing the observed genotype proportions of 0, directions were predicted correctly and incorrectly in 52 and 48 cases, respectively, for the same target seasons as above, so that the accuracy was 0.52 ($P = 0.689$ by the χ^2 test). Compared with the previous results, the number of predictions was raised because unavailable sequence data were compensated but the accuracy was lowered possibly because a smaller number of sequences were analyzed in the present study. However, the accuracy was recovered by replacing the observed genotype proportions of 0 with an arbitrary value that was smaller than the lowest positive value obtainable as the observed genotype proportion over all HuNoV seasons.

In contrast, the performance of NOROCAST did not appear to be improved by incorporation of herd immunity to VP2 and NS, suggesting that VP1-derived virus-like particle (VLP) and VP1-P domain-derived P-particle may be suitable as vaccines (Hansman et al., 2006; Kocher and Yuan, 2015). Nevertheless, VP2 and NS may still be incorporated into the fitness model for improving

the performance of NOROCAST. It may be interesting to model the fitness effects of environmental conditions such as temperature, humidity, and pH, which may affect the functions of VP2 and NS as well as VP1 (de la Noue et al., 2014; Samandoulgou et al., 2015). In addition, the fitness effect of herd immunity to VP1 may depend not only on the number of amino acid differences between HuNoV strains but also on their impacts on the binding affinity to the HuNoV receptor, which has not been identified yet (Murakami et al., 2013; Haga et al., 2016; Orchard et al., 2016). It may be critical to identify the amino acid sites in VP1, VP2, and NS involved in these functions.

When the target season was set to be season 2017/2018, it was consistently predicted that the proportions of GII.2-GII.4 and GII.17 may decrease, whereas those of other genotypes may increase, regardless of whether the observed genotype proportions of 0 were replaced with 5×10^{-6}, 5×10^{-5}, or 5×10^{-4} and whether the herd immunity to VP2 and NS was taken into account in addition to VP1 (**Supplementary Table S15**). The predicted genotype proportions can be seen on the NOROCAST website (URL: http://www.nsc.nagoya-cu.ac.jp/~yossuzuk/norocast.html). To validate these results and predict genotype proportions in the future HuNoV seasons, it may be critical to update the observed genotype frequencies and the sequence data of HuNoV.

AUTHOR CONTRIBUTIONS

YS, YHD, HK, HS, KS, and KK designed this study and wrote the manuscript. YS and YHD analyzed the data. HK, HS, KS, and KK supervised this study. All authors read and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00116/full#supplementary-material

REFERENCES

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., et al. (2014). Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect. Dis.* 14, 725–730. doi: 10.1016/S1473-3099(14)70767-4

Clarke, I. N., Estes, M. K., Green, K. Y., Hansman, G. S., Knowles, N. J., Koopmans, M. K., et al. (2012). "Caliciviridae," in *Virus Taxonomy: The Classification and Naming of Viruses. The Ninth Report of The International Committee on Taxonomy of Viruses*, eds A. M. Q. King, M. J. Adams, E. B. Carstens, E. J. Lefkowitz (San Diego: Elsevier Academic Press), 977–986.

de la Noue, A. C., Estienney, M., Aho, S., Perrier-Cornet, J.-M., de Rougemont, A., Pothier, P., et al. (2014). Absolute humidity influences the seasonal persistence and infectivity of human norovirus. *Appl. Environ. Microbiol.* 80, 7196–7205. doi: 10.1128/AEM.01871-14

Haga, K., Fujimoto, A., Takai-Todaka, R., Miki, M., Doan, Y. H., Murakami, K., et al. (2016). Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6248–E6255. doi: 10.1073/pnas.1605575113

Hansman, G. S., Natori, K., Shirato-Horikoshi, H., Ogawa, S., Oka, T., Katayama, K., et al. (2006). Genetic and antigenic diversity among noroviruses. *J. Gen. Virol.* 87, 909–919. doi: 10.1099/vir.0.81532-0

Jiang, X., Wang, M., Wang, K., and Estes, M. K. (1993). Sequence and genomic organization of Norwalk virus. *Virology* 195, 51–61. doi: 10.1006/viro.1993.1345

Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R., and 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., and and, Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kocher, J., and Yuan, L. (2015). Norovirus vaccines and potential antinorovirus drugs: recent advances and future perspectives. *Fut. Virol.* 10, 899–913. doi: 10.2217/fvl.15.57

Kroneman, A., Vega, E., Vennema, H., Vinje, J., White, P. A., Hansman, G., et al. (2013). Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* 158, 2059–2068. doi: 10.1007/s00705-013-1708-5

Kroneman, A., Vennema, H., Deforche, K., Avoort, H. V., Penaranda, S., Oberste, M. S., et al. (2011). An automated genotyping tool for enteroviruses and noroviruses. *J. Clin. Virol.* 51, 121–125. doi: 10.1016/j.jcv.2011.03.006

Lambden, P. R., Caul, E. O., Ashley, C. R., and Clarke, I. N. (1993). Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* 259, 516–519. doi: 10.1126/science.8380940

Lindesmith, L. C., Donaldson, E. F., and Baric, R. S. (2011). Norovirus GII.4 strain antigenic variation. *J. Virol.* 85, 231–242. doi: 10.1128/JVI.01364-10

Lukssa, M., and Lassig, M. (2014). A predictive fitness model for influenza. *Nature* 507, 57–61. doi: 10.1038/nature13087

Mathijs, E., Stals, A., Baert, L., Botteldoorn, N., Denayer, S., Mauroy, A., et al. (2012). A review of known and hypothetical transmission routes for noroviruses. *Food Environ. Virol.* 4, 131–152. doi: 10.1007/s12560-012-9091-z

Murakami, K., Kurihara, C., Oka, T., Shimoike, T., Fujii, Y., Takai-Todaka, R., et al. (2013). Norovirus binding to intestinal epithelial cells is independent of histo-blood group antigens. *PLoS ONE* 8:e66534. doi: 10.1371/journal.pone.0066534

Orchard, R. C., Wilen, C. B., Doench, J. G., Baldridge, M. T., McCune, B. T., Lee, Y. C., et al. (2016). Discovery of a proteinaceous cellular receptor for a norovirus. *Science* 353, 933–936. doi: 10.1126/science.aaf1220

Parra, G. I., Squires, R. B., Karangwa, C. K., Johnson, J. A., Lepore, C. J., Sosnovtsev, S. V., et al. (2017). Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLoS Pathog.* 13:e1006136. doi: 10.1371/journal.ppat.1006136

Patel, M. M., Widdowson, M.-A., Glass, R. I., Akazawa, K., Vinje, J., and Parashar, U. D. (2008). Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg. Infect. Dis.* 14, 1224–1231. doi: 10.3201/eid1408.071114

Prasad, B. V. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., and Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* 286, 287–290. doi: 10.1126/science.286.5438.287

Samandoulgou, I., Hammami, R., Morales-Rayas, R., Fliss, I., and 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, 7680–7686. doi: 10.1128/AEM.01278-15

Simmons, K., Gambhir, M., Leon, J., and Lopman, B. (2013). Duration of immunity to norovirus gastroenteritis. *Emerg. Infect. Dis.* 19, 1260–1267. doi: 10.3201/eid1908.130472

Suzuki, Y. (2013). Predictability of antigenic evolution for H3N2 human influenza A virus. *Genes Genet. Syst.* 88, 225–232. doi: 10.1266/ggs.88.225

Suzuki, Y. (2015). Selecting vaccine strains for H3N2 human influenza A virus. *Meta Gene* 4, 64–72. doi: 10.1016/j.mgene.2015.03.003

Suzuki, Y., Doan, Y. H., Kimura, H., Shinomiya, H., Shirabe, K., and Katayama, K. (2016). Predicting genotype compositions in norovirus seasons in Japan. *Microbiol. Immunol.* 60, 418–426. doi: 10.1111/1348-0421.12384

Thongprachum, A., Khamrin, P., Maneekarn, N., Hayakawa, S., and Ushijima, H. (2016). Epidemiology of gastroenteritis viruses in Japan: prevalence, seasonality, and outbreak. *J. Med. Virol.* 88, 551–570. doi: 10.1002/jmv.24387

Tomita, M., Hashimoto, K., Takahashi, K., Matsuzaki, Y., Matsushima, R., Saito, K., et al. (2000). The E-CELL project: towards integrative simulation of cellular processes. *New Gener. Comput.* 18, 1–12. doi: 10.1007/BF03037563

Vinje, J. (2015). Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* 53, 373–381. doi: 10.1128/JCM.01535-14

Zheng, D.-P., Ando, T., Fankhauser, R. L., Beard, R. S., Glass, R. I., and Monroe, S. S. (2006). Norovirus classification and proposed strain nomenclature. *Virology* 346, 312–323. doi: 10.1016/j.virol.2005.11.015