Emergence of novel GII.4 strains with an amino acid insertion at the start of antigenic site A in the major capsid protein

Preeti Chhabra[1#*], Damien C. Tully[2#], Janet Mans[3], Sandra Niendorf[4], Leslie Barclay[1], Jennifer L. Cannon[1], Anna M. Montmayeur[1], Chao-Yang Pan[5], Nicola Page[3,6], Rachell Williams[7], Helena Tutil[7], Sunando Roy[7], Cristina Celma[8], Stuart Beard[8], Gédéon Prince Manouana[9,10], Thirumalaisamy P. Velavan[9,11,12], Peter G. Kremsner[9,10]; Ayola Akim Adegnik[9,13,14], Stéphane Hué[2], Judith Breuer[13], Jan Vinjé[1]

[1]Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[2]London School of Hygiene and Tropical Medicine, London, UK

[3]University of Pretoria, South Africa

[4]Robert Koch-Institut, Berlin, Germany

[5]California Department of Public Health, Richmond, California, USA

[6]National Institute for Communicable Diseases, Johannesburg, South Africa

[7]UCL Great Ormond Street Institute of Child Health, Guilford Street, London, UK

[8]UK Health Security Agency, London, UK

[9]Institute of Tropical Medicine, Universitätsklinikum Tübingen, Germany

[10]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon

[11]Vietnamese-German Center for Medical Research (VG-CARE), Hanoi, Vietnam

[12]Faculty of Medicine, Duy Tan University, Da Nang, Vietnam

[13]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon

[14]German Center for Infection Research (DZIF), Tübingen, Germany

[15]University College London, London, UK

[#] These authors contributed equally to this article

*Correspondence: Preeti Chhabra, Centers for Disease Control and Prevention, Atlanta, Georgia, USA,

PChhabr@cdc.gov

Disclaimer:

Abstract:

We detected a novel GII.4 norovirus strain with an amino acid insertion at the start of epitope A in the

VP1 gene in a childcare center outbreak in San Francisco, USA in 2017. Similar strains were detected in

Gabon, South Africa, and the UK between 2018 and 2023.

Background

Norovirus is the most common cause of acute gastroenteritis in all age groups worldwide. The virus has

a positive-strand RNA genome of approximately 7.5 kb in length that is organized in three open reading

frames (ORFs). ORF1 encodes a polyprotein that is post-translational cleaved into six non-structural

proteins including NS7, the viral RNA-dependent RNA polymerase (RdRp). ORF2 and ORF3 encode

the major (VP1) and minor (VP2) capsid proteins. Noroviruses are genetically diverse viruses which are

classified into at least 10 different genogroups (G), with genogroup II, genotype 4 (GII.4) viruses

causing the majority of illness worldwide (2). Since 1995, six major norovirus GII.4 epidemic variants

have been identified (US 95–96, Farmington Hills 2002, Hunter 2004, Den Haag 2006, New Orleans

2009, and Sydney 2012) (3). The emergence of new GII.4 variants has been associated with changes in

antibody binding epitopes A to G which have been mapped to the surface exposed P2 subdomain of VP1

(4). Antibodies mapping to these epitopes can block the interaction between VP1 and histo-blood group

antigens (HBGA) which serve as binding ligands for the virus to cells (5). Since 2012, GII.4 Sydney variant has been the most prevalent genotype globally (2).

The Study

Several surveillance networks are tracking trends in norovirus strain diversity including CaliciNet in the US (6) and NOROPATROL in the UK. In July 2017, sequences of a norovirus outbreak in a childcare center in San Francisco, CA that had > 2% sequence difference in the 5-end of ORF2 (300nt-region C) from existing GII.4 Sydney viruses were uploaded to CaliciNet. Genetically similar strains were identified in stool specimens from 3 UK outbreaks in Newcastle (2019), London (2021) and Brighton (2021- oyster samples consumption) in the UK, in sporadic samples from children in 2018 and 2019 in four towns (Libreville, Franceville, Oyem and Koulamoutou) in Gabon (7) and in children and adolescents from three provinces (Western Cape, Gauteng, Mpumalanga) in South Africa in 2018-2022. We sequenced the complete genome and/or complete VP1 sequences of these strains and compared the complete VP1 and RdRp sequences (if available) with existing GII.4 norovirus sequences (https://calicivirustypingtool.cdc.gov/).

Viral RNA was extracted and complete genome or VP1 sequences for US, UK and South African strains were obtained according to published methods (6, 7, 8; 9, 10. Complete VP1 from Gabon strains were amplified by nested RT-PCR using oligo dT and oligonucleotide primers NV6e (s) 5'-ACCAYTWTGATGCAGACTA; NV6f (s) 5'-ACCAYTATGATGCTGATTA; NV6g (s) 5'-ATCAYTATGATGCWGAYTA and NV357a (as) 5'-CGCCAGTCCAGGAGTCCAAAATY to amplify VP1 and VP2 of the genome. Briefly, cDNA was synthesized using Oligo dT and Lunascript Master Mix Kit (New England Biolabs®, https://www.newenglandlab.com/) for 30 min at 55°C; 1 min at 95°C. Both PCRs were carried out using OneTaq 2X Master Mix (New England Biolabs®) for 30 cycles at 94°C for 10 sec, 45°C for 30 sec and 72°C for 3 min followed by final extension of 72°C for 2 min. All amplicons were sequenced. Complete VP1 amino acid and partial RNA-dependent RNA

polymerase sequences (762 nt) were aligned with GII.4 reference strains representing all known emerging and epidemic GII.4 viruses using with Clustal W in MEGA X (11). Phylogenetic trees were computed with Maximum-likelihood phylogeny using the Jones-Taylor-Thornton (amino acid) and Tamura-Nei (nucleotide) models with gamma distribution of evolutionary rates among sites and 100 bootstrap replications. 3-D structures of GII.4 San Francisco viruses were predicted using the ChimeraX v1.4 (12) and alphafold prediction tool (13) with the GII.4 Sydney (PDB 4OP7) P-domain as the backbone. Nucleotide sequences of GII.4 San Francisco strains have been deposited in GenBank under accession numbers aaaa – bbbb.

The widely used norovirus genotyping region (region C) of all GII.4 San Francisco sequences was closest to the GII.4 Sydney and GII.4 Den Haag reference strains (Figure 1A) with maximum nucleotide sequence identities ranging from 91% - 95% and formed a separate cluster from all other known GII.4 variant viruses. In addition, the complete VP1 amino acid sequences formed a distinct cluster from GII.4 New Orleans and GII.4 Sydney viruses with 5-10% amino acid difference (Figure 1B). RdRp sequences of all strains were typed as GII.P31 (data not shown).

Interestingly, VP1 sequences of all GII.4 San Francisco strains had an alanine insertion at position 293/294 at the start of Epitope A, coinciding with a unique SVTQTAT/A motif at positions 289-295 (Figure 2A). Compared with GII.4 Sydney 2012 and GII.4 New Orleans viruses, major mutations at amino acid residues 256 and 438 in the P1-region and 294, 310, 340, 341, 356, 372, 373, 377, 393 and 395 in the P2-region (hypervariable region). (Figure 2A).

Homology modeling of the GII.4 San Francisco P-domain using GII.4 Sydney 2012 as a backbone (PDB 4OP7; GenBank accession number JX459908) (Figure 2B) showed structural changes near and within Epitope A. Several charged amino acids present in GII.4 Sydney 2012 were replaced by neutral amino acids (Figure 2B) with the introduction of the insertion and the SVTQTAT/A motif of GII.4 San Francisco. Changes in amino acid charge or hydrophobicity was also observed in monoclonal antibody-

binding epitope G (A256N) and within and around the HBGA binding region D391N, S393D and

T395A (except South African strains) and surrounding sites (data not shown in 3-D model).

Conclusion

We report a novel norovirus GII.4 strain, provisionally named "GII.4 San Francisco", detected on three

continents in human stool specimens. These novel strains have a unique amino acid insertion in the VP1

major capsid protein at the start of epitope A. A similar unique insertion on antigenic site D was

observed in GII.4 variant Farmington Hills, which emerged in 2002 replacing the GII.4 US 95/96

viruses which had been circulating globally since 1995 (14). Whether the GII.4 San Francisco strains

will be able to replace the currently dominant GII.4 Sydney strains is not clear yet.

GII.4 San Francisco strains showed mutations at residues S393D and T395A in HBGA binding site 2. In

the GII.4 Camberwell variant, aspartic acid (D) at position 393 was reported to promote B-type binding

(15). Also, the insertion of threonine at position 395 in a GII.4 Farmington Hills 2004 variant resulted in

a markedly lower level of binding to A and B HBGA types as compared to the previous US95/96 GII.4

variant suggesting biological relevance (5). In addition, threonine at position 395 has been consistent in

all GII.4 variants that emerged after GII.4 Farmington Hills. However, in GII.4 San Francisco strains

from the US, Gabon and the UK, threonine has been replaced by alanine. Whether these changes in

binding site 2 can alter the binding affinity/specificity with host HBGAs requires additional studies.

Although routine typing targeting the 5'-end of ORF2 suggested a possible new GII.4 variant, the

unique amino acid insertion in epitope A of VP1 confirmed that GII.4 San Francisco is a new GII.4

variant circulating on at least 4 continents (North America, Europe, Africa) and Asia

(https://www.norosurv.org/; data not shown). Our analyses confirm a recommendation to expand the

standard region for genotyping of noroviruses to include at least complete VP1 sequences which allows

early identification of new emerging norovirus strains and share the information rapidly to assess their

pandemic potential. This will provide important information for future norovirus vaccines that are currently in clinical trials.

References:

1. Chhabra P, de Graaf M, Parra GI, Chan MC, Green K, Martella V, Wang Q, White PA, Katayama K, Vennema H, Koopmans MPG, Vinjé J. Updated classification of norovirus genogroups and genotypes. J Gen Virol. 2019;100(10):1393-1406.

2. Farahmand M, Moghoofei M, Dorost A, Shoja Z, Ghorbani S, Kiani SJ, Khales P, Esteghamati A, Sayyahfar S, Jafarzadeh M, Minaeian S, Khanaliha K, Naghdalipour M, Tavakoli A. 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. 2022;32(1):e2237.

3. Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, Green K, Martella V, Katayama K, Koopmans M. Proposal for a unified norovirus nomenclature and genotyping. Arch Virol. 2013;158(10):2059-68.

4. Tohma K, Lepore CJ, Gao Y, Ford-Siltz LA, Parra GI. Population Genomics of GII.4 Noroviruses Reveal Complex Diversification and New Antigenic Sites Involved in the Emergence of Pandemic Strains. mBio. 2019;10(5):e02202-19.

5. de Rougemont A, Ruvoen-Clouet N, Simon B, Estienney M, Elie-Caille C, Aho S, Pothier P, Le Pendu J, Boireau W, Belliot G. Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. J Virol. 2011;85(9):4057-70.

6. Cannon JL, Barclay L, Collins NR, Wikswo ME, Castro CJ, Magaña LC, Gregoricus N, Marine RL, Chhabra P, Vinjé J. Genetic and Epidemiologic Trends of Norovirus Outbreaks in the United States from 2013 to 2016 Demonstrated Emergence of Novel GII.4 Recombinant Viruses. J Clin Microbiol. 2017;55(7):2208-2221.

7. Manouana GP, Nguema-Moure PA, Mbong Ngwese M, Bock CT, Kremsner PG, Borrmann S, Eibach D, Mordmüller B, Velavan TP, Niendorf S, Adegnika AA. Genetic Diversity of Enteric Viruses in Children under Five Years Old in Gabon. Viruses. 2021;13(4):545

8. Brown JR, Roy S, Ruis C, Yara Romero E, Shah D, Williams R, Breuer J. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. J Clin Microbiol. 2016;54(10):2530-7.

9. Mans J, Murray TY, Taylor MB. Novel norovirus recombinants detected in South Africa. Virol J. 2014;11:168.

10. Parra GI, Squires RB, Karangwa CK, Johnson JA, Lepore CJ, Sosnovtsev SV, Green KY. Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity. PLoS Pathog. 2017;13(1):e1006136.

11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol.;35(6):1547-1549.

12. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70-82.

13. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

14. Widdowson MA, Cramer EH, Hadley L, Bresee JS, Beard RS, Bulens SN, Charles M, Chege W, Isakbaeva E, Wright JG, Mintz E, Forney D, Massey J, Glass RI, Monroe SS. Outbreaks of acute

gastroenteritis on cruise ships and on land: identification of a predominant circulating strain of

norovirus--United States, 2002. J Infect Dis. 2004;190(1):27-36.

15. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, Baric RS.

Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med. 2008;5(2):e31.

Figure Legends:

Figure 1: Phylogenetic trees of the **A)** common genotyping region C and **B)** complete VP1 amino acid

sequences of novel GII.4 San Francisco strains and GII.4 variants including recently identified clusters

(e.g., GII.4 Hong Kong). Evolutionary analyses were conducted in MEGA X using the Maximum

Likelihood method based on the Tamura-Nei (region-c) and JTT matrix-based (VP1) models. A discrete

Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G,

parameter = 0.2174)). The trees were drawn to scale, with branch lengths measured in the number of

substitutions per site. Scale bar represents nucleotide substitutions per site. Bootstrap (100) values are

indicated at the nodes.

Figure 2: A) Amino acid differences in the P1, P2 and HBGA site B of the GII.4 San Francisco strains

compared with reference strains of other epidemic GII.4 variants. B) The structural changes near and

within Epitope A antigenic region are shown on GII.4 San Francisco P-domain 3-D structure modeled

and overlayed on a GII.4 Sydney 2012 backbone (PDB 4OP7). Negatively (red) and positively (green)

charged amino acids of GII.4 Sydney (panel A) were replaced with neutral amino acids (blue) in the

GII.4 San Francisco strain and a hydrophobic (yellow) amino acid, alanine, was inserted between T293

and T294.