

RESEARCH ARTICLE

Open Access

Molecular and evolutionary characterization of norovirus GII.17 in the northern region of Brazil



Larissa Cristina Prado das Neves Costa^{1,2*}, Dielle Monteiro Teixeira³, Ana Caroline Rodrigues Portela⁴, Ian Carlos Gomes de Lima⁴, Renato da Silva Bandeira³, Edivaldo Costa Sousa Júnior³, Jones Anderson Monteiro Siqueira⁴, Hugo Reis Resque⁴, Luciana Damascena da Silva⁴ and Yvone Benchimol Gabbay⁴

Abstract

Background: Currently, norovirus (NoV) is associated with one-fifth of all acute gastroenteritis (AGE) cases worldwide. The NoV GII.17_2014 variant has been associated with gastroenteritis outbreaks in several Asian countries, replacing the previously dominant Sydney 2012 variant. There is limited data about circulation of this new strain in Brazil. This study aimed to describe the phylogenetic and evolutionary characteristics of the GII.17_2014 strains in the Northern region of Brazil.

Methods: NoV was detected by enzyme immunoassay (EIA) in 645 stool samples of AGE cases that were reported in Pará and Amazonas states during 2015–2016. All positive samples were tested for NoV GI and GII by reverse transcription polymerase chain reaction (RT-PCR) and the amplicons were subjected to genome sequencing. The GII.17-positive samples were retested by PCR using different sets of designed primers, which target a highly conserved capsid gene region. Next, the amplicons were sequenced and phylogenetically analyzed using Bayesian inferences.

Results: Of the 645 samples tested, 208 (32.2%) tested were positive for NoV by EIA, among which 95 (45.7%) were genotyped. Among the genotyped samples, 12 (12.6%) were characterized as GII.17_2014 with the first case detected in November 2015 (1/30, 3.3%) and the others in 2016 (11/65, 16.9%). All strains found in our study were clustered in clade D (epidemic strain). The uncorrelated log-normal model estimations calculated the rate of evolution for GII-17 strains as 1.95×10^{-3} (1.28×10^{-3} – 2.63×10^{-3}). In total, 36 nucleotide changes were observed after analyzing the VP1 sequence, among which 28 occurred in the P2 region.

Conclusions: These data demonstrate the evolutionary dynamics in NoV GII.17_2014 strains, which indicated high mutation rates with nucleotide substitutions and indels that are related to the elevated levels of antigenic diversity. This partly explains the increase in viral prevalence.

Keywords: Norovirus, Emergent, Gastroenteritis

* Correspondence: prado.cris.lnc@hotmail.com

¹Postgraduate Program in Parasitic Biology in the Amazon, Universidade do Estado do Pará, Instituto Evandro Chagas, Belém, PA, Brazil

²Virology Section, Evandro Chagas Institute, Brazilian Ministry of Health, Rodovia BR-316, Km 7 s/n, Levilândia, Ananindeua, Pará 67030-000, Brazil

Full list of author information is available at the end of the article



Background

Nowadays, norovirus (NoV) is associated with one-fifth of all acute gastroenteritis (AGE) cases worldwide and accounts for approximately 200,000 deaths annually in developing countries. NoV is also implicated in outbreaks of non-bacterial gastroenteritis in closed communities, such as schools, hospitals, camps, cruise ships, and nursing homes, affecting all age groups [1].

NoV is currently classified into ten genogroups (GI to GX) based on the amino acid diversity of VP1 protein. These viruses are further subdivided into 49 confirmed capsid genotypes based on amino acids of the complete VP1 and 60 confirmed types based on partial nucleotide sequences of RNA-dependent RNA polymerase regions [2].

Morphologically, NoV is a small, non-enveloped virus, exhibiting icosahedral symmetry with a diameter ranging from 27 to 40 nm. NoV has a single-stranded, positive-sense RNA genome, which is organized in three open reading frames (ORF). The ORF1 encodes six non-structural proteins, which are involved in viral replication. The ORF2 encodes the largest structural protein (VP1), which is subdivided into shell (S) and protruding (P) domains. Within the P domain there is a highly variable region called P2 that is essential for virion binding with the host cell histo-blood group antigens (HBGA). The ORF3 encodes the smallest structural protein (VP2) [3–5]. The mutations causing amino acid substitutions, insertions, or deletions in VP1 alter the properties of the virus and affect the binding and recognition by the immune system and subsequently result in the emergence of new NoV antigenic variants [6, 7].

Globally, GII.4 genotype has been responsible for most gastroenteritis outbreaks and sporadic cases related to NoV since the 1990s. Pandemic variants of GII.4 emerge every two or three years and cause multiple epidemics of gastroenteritis [8]. During the 2014–2015 season, the GII.17_2014 (Kawasaki) pandemic variant emerged and caused gastroenteritis outbreaks in schools, colleges, factories, and kindergartens in the Guangdong province, China [9]. In China, 24 (82.8%) of the 29 outbreaks (2340 cases) reported during 2014–2015 were caused by GII.17_2014. Between 2013 and 2014, only nine outbreaks with 949 cases were recorded. These outbreaks were predominantly caused by the GII.4 Sydney 2012 genotype [9].

Since 2013–2014, a GII.17_2014 variant has been associated with outbreaks of gastroenteritis in several countries and in some cases replacing the dominant GII.4 Sydney 2012 genotype [9, 10]. In Latin America, the GII.17_2014 variant was reported in Brazil and Argentina in children and adults [11–13].

This study aimed to describe the phylogenetic and evolutionary characteristics of circulating GII.17_2014

strains in the northern region of Brazil, which will contribute to a greater understanding of its distribution in the country. It is important to understand the antigenic and evolutionary characteristics of this variant as this strain has epidemiological importance in some regions of the country [10, 14].

Methods

Samples collections

The NoV-positive cases were identified from the database of the National Viral Gastroenteritis Surveillance Program, Brazilian Ministry of Health. This program includes epidemiological surveillance of diarrhea cases, which was obtained from inpatients who attended the public health facilities of the Northern region of Brazil (states of Amazonas, Pará, Roraima, Amapá, and Tocantins). Fecal specimens from children under 5 years of age were also included in this study. Previously, these fecal samples tested negative for rotavirus. In this study, 32.2% (208/645) samples tested positive for NoV by immunochromatography (RIDA[®]QUICK Norovirus) or enzyme immunoassay (EIA) (RIDASCREEN[®] Norovirus 3rd Generation). These NoV-positive samples were subjected to one-step RT-PCR to amplify the ORF1/ORF2 junction region using the MON431/MON 432 and G2SKR (557 bp) primers, which can detect all NoV genotypes [15, 16].

Preparation of fecal suspensions and nucleic acid extraction

The fecal suspensions (10%) were prepared in 0.01 M Tris/HCl/Ca⁺⁺ salt solution (pH 7.2), resulting in a final volume of 140 µL. The nucleic acid was extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions.

Design and validation of NoV primers

Specific primer sets were designed to sequence the GII.17_2014 strains. These primers amplified the complete ORF2 region of NoV GII.17 into larger fragments of the NoV genome that can be used for more accurate phylogenetic analyses.

The primers were designed based on multiple alignments of GII.17 norovirus genome sequences available in GenBank (including Brazilian strain) using the Geneious 10.0.6 and Primer3 softwares. The parameters used for primer designing were as follows: minimum size of 1000 bp for the first pair (LNOV 5KF/5KR) and 550 bp for the second pair (LNOV 6KF/6KR), melting temperature (tm) of approximately 60 °C and medium GC content of 55%. The polymerase region was amplified using the IL17-3F, 3R and IL17-4F, 4R primers as described by Xue et al. [17].

The RNA extraction was performed using the QIAamp Viral RNA MiniKit (Qiagen, Hilden, Alemanha). Initially, a one-step RT-PCR was performed with the temperature gradient ranging from 44.9–58 °C to identify the ideal annealing temperature. The validation tests were performed in triplicates on three different days by two analysts using the same samples (control sample GII.17_2014 previously sequenced) under the same conditions for a period of up to 48 h [18].

Positive and negative controls (CP = 5; CN = 5) were tested in three rounds in the same thermocycler by the same two analysts. The detection limit (LD) was determined using the serially diluted (10^1 – 10^9) total RNA samples (RNAse-free water was used for dilution). The fluorometric quantitation of total RNA was performed for the previously sequenced GII.17_2014 sample in a Qubit™ Fluorometer (Invitrogen) apparatus using the Qubit® RNA BR Assay Kit (Invitrogen). The initial concentration of extracted undiluted viral RNA was 6.0×10^3 ng/μL. The RT-PCR was performed in a single step using the One-Step SuperScript III RT-PCR System kit (Invitrogen, Carlsbad, CA, USA). The amplification of the NoV partial genome was performed using five sets of primers to amplify the viral polymerase (ORF1) and capsid (ORF 2) regions. The amplicons were visualized in 1.5% agarose gels.

The performance of the method was confirmed based on the conformity hypothesis test, considering the definition of pair of answers (presence of concordant and discordant results using CP and CN). This test verified whether the method detected the viral nucleic acid (the results are “concordant”) or does not detect the viral nucleic acid (the results are “discordant”). Based on these data, we verified the probability of accepting or rejecting the null hypothesis (agreement method), considering a significance level of 0.05%.

Based on the results obtained, it was possible to verify some parameters that evaluate the performance of the method such as, repeatability, reproducibility, accuracy, sensitivity, and specificity. These parameters were compared using Cohen Kappa Index [18].

Detection GII.17_2014 strains by RT-PCR

Four genomic fragments containing the NoV polymerase and VP1 regions were amplified by RT-PCR. The RT-PCR was performed in a 25-μL reaction volume comprising 8 μL of viral RNA, 16 μL of reagent mixture (0.5 μL H₂O-free DNase/RNase, 12.5 μL of 2X Reaction Mix, 1 μL of each primer, 1 μL SuperScript® III RT/Platinum™ Taq Mix), and 1 μL of DMSO. The PCR conditions were as follows: For primers IL17-3F,3R and IL17-4F,4R [45 °C for 30 min, 94 °C for 3 mins, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, and 68 °C for 90 s, and finally 68 °C for 5 min]; For primers LNOV

5KF/5KR [45 °C for 30 min, 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 58 °C for 1 min, and 72 °C for 1 min and finally 72 °C for 10 min]; primers LNOV 6KF/6KR [45 °C for 30 min, 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 1 min, and finally 72 °C for 10 min]. The amplicons were visualized on 1.5% agarose gels stained with SYBR Safe DNA Gel Stain (Invitrogen, USA).

Sequencing

The amplicons were purified using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany). Direct sequencing was performed using the same primer set used in the PCR and Big Dye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) in the 3130xl Genetic Analyzer automatic sequencer (Applied Biosystems, Foster City, CA, USA).

NoV genotyping and evolutionary analysis

The genotype was identified using the Norovirus Genotyping Tool 1.0, Basic Local Alignment Search Tool (BLAST) and Human calicivirus Typing tool (<https://norovirus.phiresearchlab.org/>). All genomes or complete ORF2 GII.17 sequences registered in the GenBank were used for evolutionary analysis. The final database comprised 85 sequences after data formatting using the CD-Hit program (Cluster Database at High Identity with Tolerance) to split the sequences with 100% identity (Additional file 1) [19]. The sequences obtained in this study were deposited in the GenBank database under accession numbers: MN045201, MN045199, MN045191, MN045193, MN045195, MN045196, MN045197, MN045200, MN045203, MN045192, MN045194, and MN045198.

The sequences were aligned using the Mafft 7 algorithm in Aliview 1.17.1 alignment viewer and editor software [20, 21]. Phylogenetic analyses were performed using the MEGA 6.06 software and the phylogenetic trees were generated by the maximum likelihood method with GTR+I+G4 nucleotide substitution model in jModelTest v. 2.1 [22] with 1000 bootstraps. Evolutionary distances were calculated using the Kimura 2-parameter model [23].

The evolutionary analyses were performed using Bayesian inference by Markov Chain Monte Carlo (MCMC) method in the BEAST software 1.10.4 [24]. Molecular clock models calculated the time at which the divergence between clades in the phylogeny occurred and determined the time of the most recent common ancestor (TMRCA).

Two clock models (strict clock and lognormal relaxed) and three coalescence models (Constant Size, Bayesian Skyline, and GMRF Bayesian Skyride) were used. The estimation of the data and the best model were evaluated

using the test marginal likelihood estimation (MLE) and path sampling (PS)/stepping-stone sampling (SS) [25, 26].

The runs were carried out with three Markovian chains with 50 million generations, using the best model of the molecular clock and evolutionary coalescence together with the phylogeographic data. The results were analyzed in the Tracer 1.66 software [27]. The files were combined in LogCombiner and the trees were annotated in the Treeannotator (both are part of the BEAST package). The graphic result was visualized in the FigTree 1.4.2 program [28]. Amino acid sequence of VP1 was analyzed using the Geneious 8.1.3 software.

Protein modeling

Three-dimensional structure of proteins was constructed employing a protein homology modeling. The initial search and selection were performed using Protein Data Bank (PDB), with the NoV capsule GII.17_2014 as the initial parameter. Modeller 9.15 software was used to construct the three-dimensional models. After protein modeling, these models were validated using Procheck and Verify 3D to confirm the quality of the biochemical parameters [29, 30]. Pymol 1.8 software was used for molecular visualization. Molecular docking was performed using the three dimensional structures of P2 protein constructed by homology modelling and HBGA using the Autodock Vina software, which uses Genetic

Algorithm to define the best binding site and superimposes with the structures [31].

Results

During 2015–2016, 32.2% (208/645) of the samples tested positive for NoV by both immunochromatography and EIA tests. Of these 208 samples, 95 (45.7%) were successfully genotyped. The GII.17_2014 genotype emerged in the Amazonas state in November 2015 with a total frequency of 12.6% (12/95) with the first positive sample detected in 2015 (1/30, 3.3%) and the others in 2016 (11/65, 16.9%). During this period, GII.Pe/GII.4 was the most frequent genotype (50.5%, 48/95), followed by GII.16/GII.4 (16.9%, 16/95) (Fig. 1). Other genotypes, such as GII.P7/GII.6, GII.P16/GII.30, GII.P7/GII.7, GII.Pg/GII.1, and GII.P21/GII.13 (20%, 19/95) were also detected.

The RT-PCR assays were satisfactorily validated. The optimum annealing temperature for the LNOV 5KR/5KF and LNOV 6KR/6KF primers was 58 °C and 55 °C, respectively. The qualitative RT-PCR analysis using both the undiluted and diluted viral RNA revealed that the genetic material could be amplified at 10⁻⁵ dilution. Thus, considering an initial undiluted viral RNA concentration of 6.0 × 10³ ng/μL, the viral genome could be detected till the concentration of 6.0 × 10⁻² ng/μL in the assays using the two pairs of primers (LNOV 5KF/5KR and LNOV 6KF/6KR).

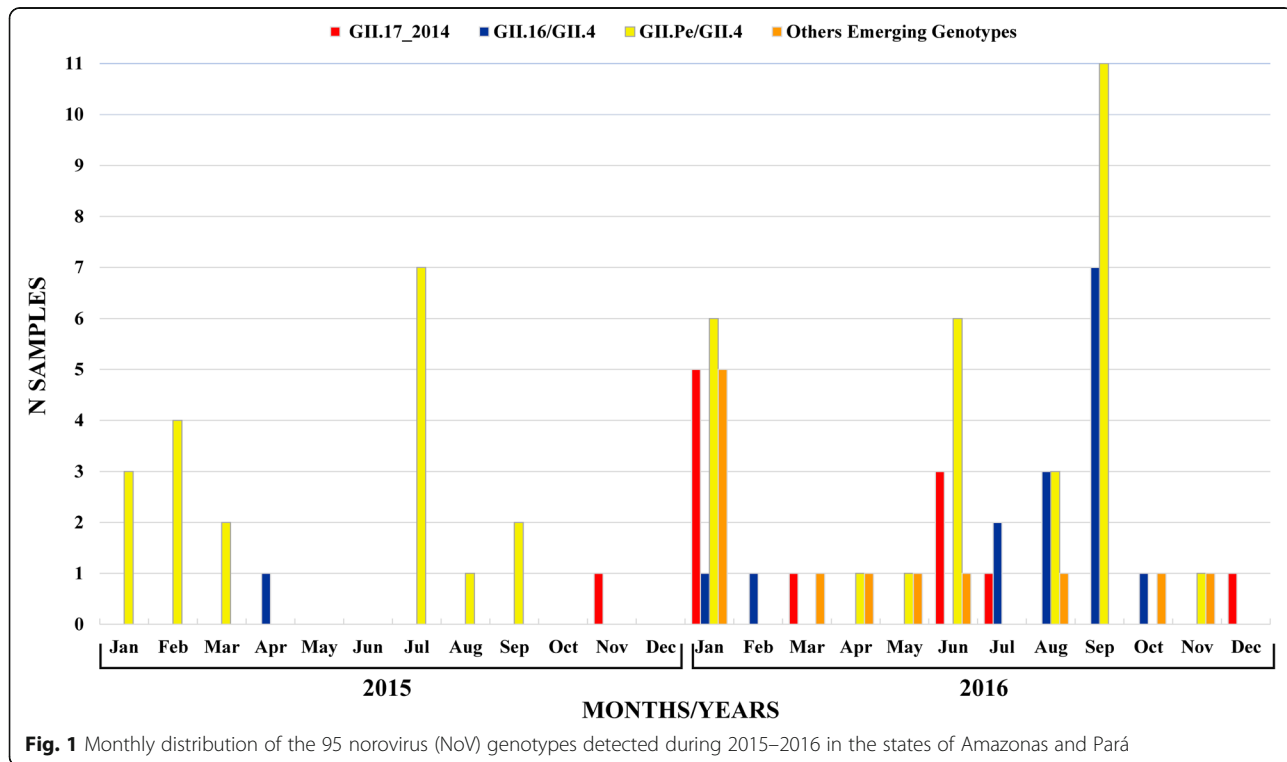


Fig. 1 Monthly distribution of the 95 norovirus (NoV) genotypes detected during 2015–2016 in the states of Amazonas and Pará

The method performance was verified based on the parameters, such as sensitivity, specificity, accuracy, and Kappa index (Table 1). Based on the data obtained, good correlation of results was verified. It is noteworthy that an equal number of negative controls (DNase- and RNase-free water) were also tested and all showed negative results.

The phylogenetic analysis revealed that the reported NoV GII.17 sequences are grouped into four different clades: clade A, which comprises strains circulating in the year 1978, clade B comprising strains detected from 2002 to 2007, and clades C and D, comprising strains detected after 2013. The strains found in our study were all clustered in clade D (Figs. 2 and 3).

The distribution of the p-distances was evaluated to establish the distance matrix and to facilitate the description of the amino acid variation among the clades. The results of this study revealed that the amino acid variation between clades C and D was 4.72% (Table 2).

To investigate the temporal evolutionary dynamics of the GII.17 norovirus in the Amazon, we applied Bayesian analysis with different population dynamic models. The strict molecular clock and GMRF Bayesian Skyride coalescence model were the most suitable models. The selection of the evolutionary model was made by comparing a posteriori probabilities of the models under investigation. Following Bayes factor calculation, the TMRCA was estimated in 1881 (95% HPD: 1832–1921). The evolution rate for GII-17 strains was 1.95×10^{-3} (1.28×10^{-3} – 2.63×10^{-3}) (Additional file 2).

The Bayesian inference revealed that the Brazilian GII.17_2014 strains detected in this study have high similarities and belonged to clade D (epidemic strain) with TMRCA estimated in 2010 (2009–2012). Temporal and spatial analysis estimated that the strains detected in this study most likely originated from China. The strains were spread in Brazil from Hong Kong/China through Amazonas state. Lineages of GII.17_2014 were observed in Amazonas and Pará State. The TMRCA for these

lineages was estimated in the year 2013 (2013–2015) with mutation rates of 1.97×10^{-3} (1.28×10^{-3} – 2.66×10^{-3}) and 2.05×10^{-3} (1.38×10^{-3} – 2.77×10^{-3}) substitutions/site/year, respectively (Fig. 4).

In this study, 36 indels were obtained by comparing the VP1 sequences of clades C and D mainly in the sub-domain P2, which contains the antigenic epitopes and host receptor binding domains. These indels occurred in the antigenic sites of the outer capsid region, the amino acid substitutions occurred at the positions: 294, 295, 296, 395 and 407 (Additional file 3 and Fig. 5a). Additionally, two insertions at the positions 377 and 397 and one deletion at the position 393 were observed (Additional file 3). It is worth mentioning that the 377/378 residue commonly binds to the HBGA fucose fraction forming a hydrogen bond, as shown in Fig. 5b.

The modeling used the published crystal structure of Protruding domain of GII.17 norovirus Kawasaki 308 (PDB: 5lkc) as a template [21]. The quality of the model was evaluated and validated using Procheck, based on the Ramachandran graph to verify the structure. The results of Procheck demonstrated that this model showed 94.2% of residues in the most favored regions, 5.8% of residues in the additional allowed regions, 0% in generously allowed regions, and 0% in regions generously allowed and disapproved (Additional file 4).

Therefore, the predicted model is considered of high quality due to the percentage distribution of the amino acid residues. The results obtained by Verify 3D, showed that 84.56% of the residuals of the models presented a score of ≥ 0.2 in the 3D-1D profile, which indicated the good quality of the general structure (Additional file 5).

The molecular docking shows the interaction between the amino acids in the following positions: 347, 348, 349, 375, 378, 385, 442, 443, and 444. The minor energy for this system was -4.7 kcal/mol.

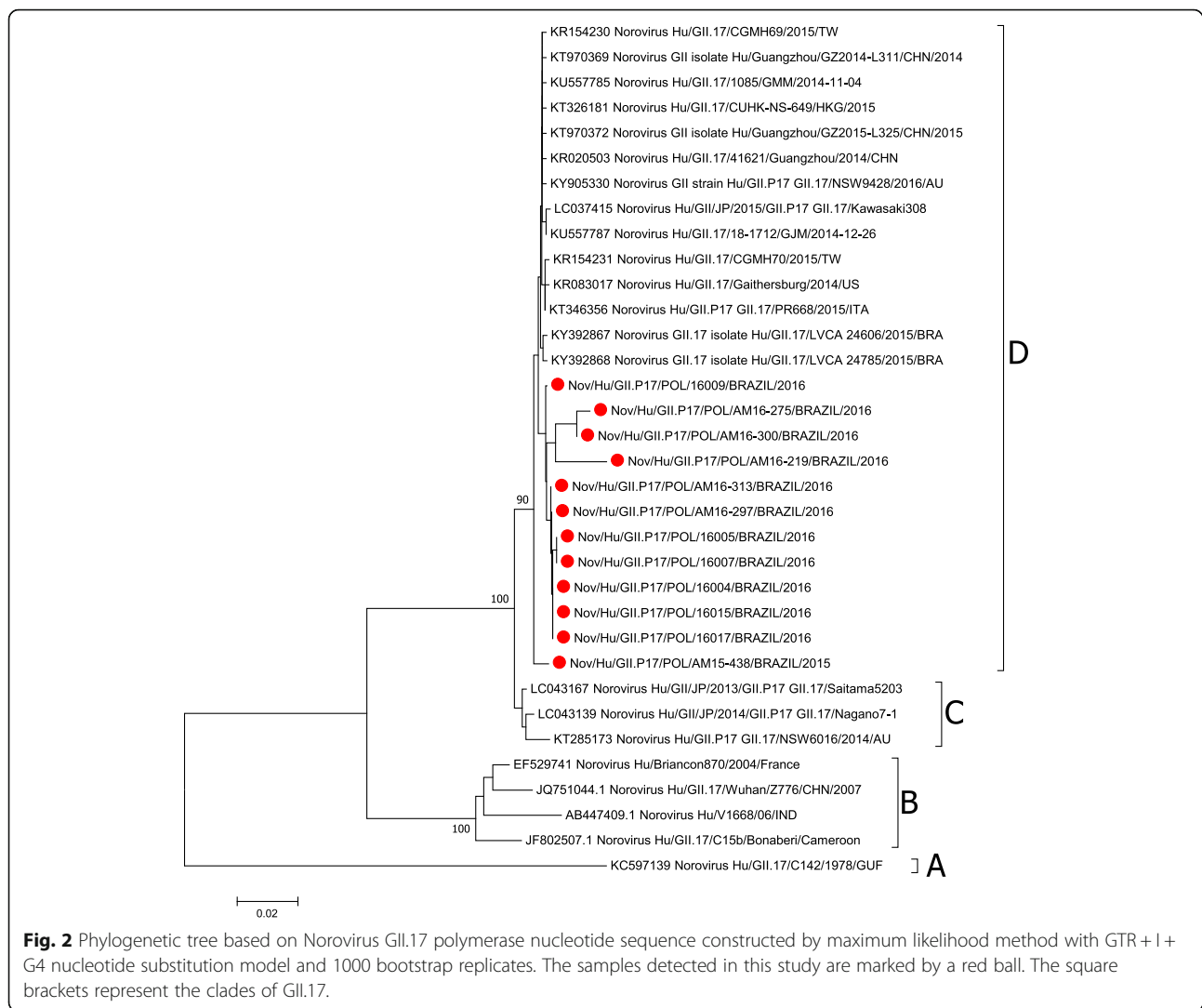
Discussion

NoV causes non-bacterial gastroenteritis, especially among children [32]. The GII.4 genotype and its variants were mainly responsible for the numerous gastroenteritis outbreaks and sporadic cases for the last 20 years [33]. However, new strains have emerged, and GII.17_2014 has assumed a significant epidemiological role for causing large outbreaks in Asian countries (replacing the GII.4 Sydney 2012 strain) with rapid spread to other parts of the world, which has alarmed the surveillance system of various regions [10, 14, 34].

In the Amazon, there was an increase in the circulation of emerging NoV strains during 2015–2016. In 2016, Silva et al. [11] reported the presence of GII.17_2014 genotype among children hospitalized in northern Brazil for the first time, with a frequency of 38.5% (5/13). In a study with Argentinian patients with AGE, this

Table 1 Evaluation of the analytical sensitivity of reverse transcription polymerase chain reaction for detection of NoV genotype GII.17 strains

Parameters	Values			
	Assay 1		Assay 2	
	Analyst 1	Analyst 2	Analyst 1	Analyst 2
Reproducibility	93.3%	80.0%	100.0%	100.0%
Repeatability	86.6%	80.0%	100.0%	100.0%
Sensitivity	93.3%	80.0%	100.0%	100.0%
Specificity	100.0%	100.0%	100.0%	100.0%
Accuracy	96.6%	90.0%	100.0%	100.0%
Kappa index	0.93	0.80	1.0	1.0



variant was detected in 25% (1/4) of the samples. Additionally, phylogenetic analysis revealed that this genotype presented characteristics of the emergent strain that circulated in Asian countries [13].

The GII.17 circulation caused several gastroenteritis outbreaks replacing even the GII.4 variant. Several studies have described the origin and dispersion of this emerging strain, mainly in Asian countries. Lu et al. [9] demonstrated that between January 2014 and January 2015, GII.17_2014 genotype was detected in 10 provinces of Guangdong, China, and accounted for 83% of all outbreaks in this region. In 2014, thirty residents of a hospital in Shanghai had AGE symptoms and 43.3% were positive for GII.17_2014 and no other pathogens were identified [35].

With the emergence of the new GII.17_2014 strain, there was a need to design new primers that efficiently amplify the VP1 region of the viral genome. Therefore, the primers designed in this study aimed to amplify the whole VP1 region into two large fragments. The primers

described in the literature were included in this study to amplify the virus genome analysis region (ORF1 and ORF3).

The RT-PCR performed using the primers designed in this study presented high sensitivity and considerably increased the detection rate of NoV GII.17_2014 in the analyzed samples. Therefore, the evaluation of the performance of the method showed very satisfactory results regarding sensitivity and other evaluation parameters. However, small variations were observed when analyzing different samples due to factors such as, amounts of virus, mode of storage, preparation, and handling.

The results obtained in this study was 100 times more sensitive than those achieved by Xue et al. [17], who evaluated primers developed to detect the complete genome of NoV GII.17.

A study conducted in Japan to evaluate the limit of detection obtained values ranging from 10^{-2} to 10^{-4} in Multiplex-PCR for the following viruses: Aichi virus, Human Parechovirus, Enterovirus, and Human

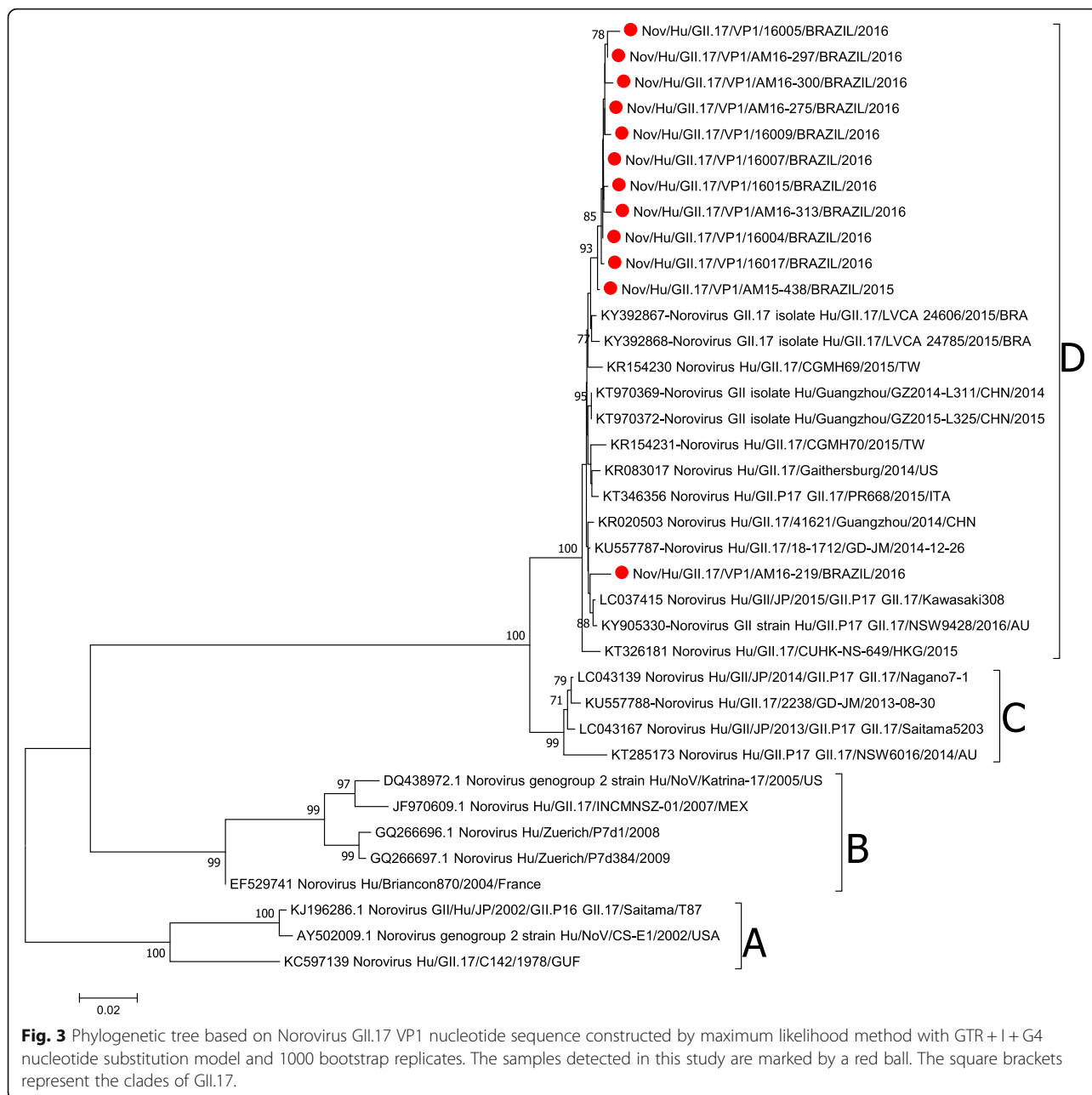
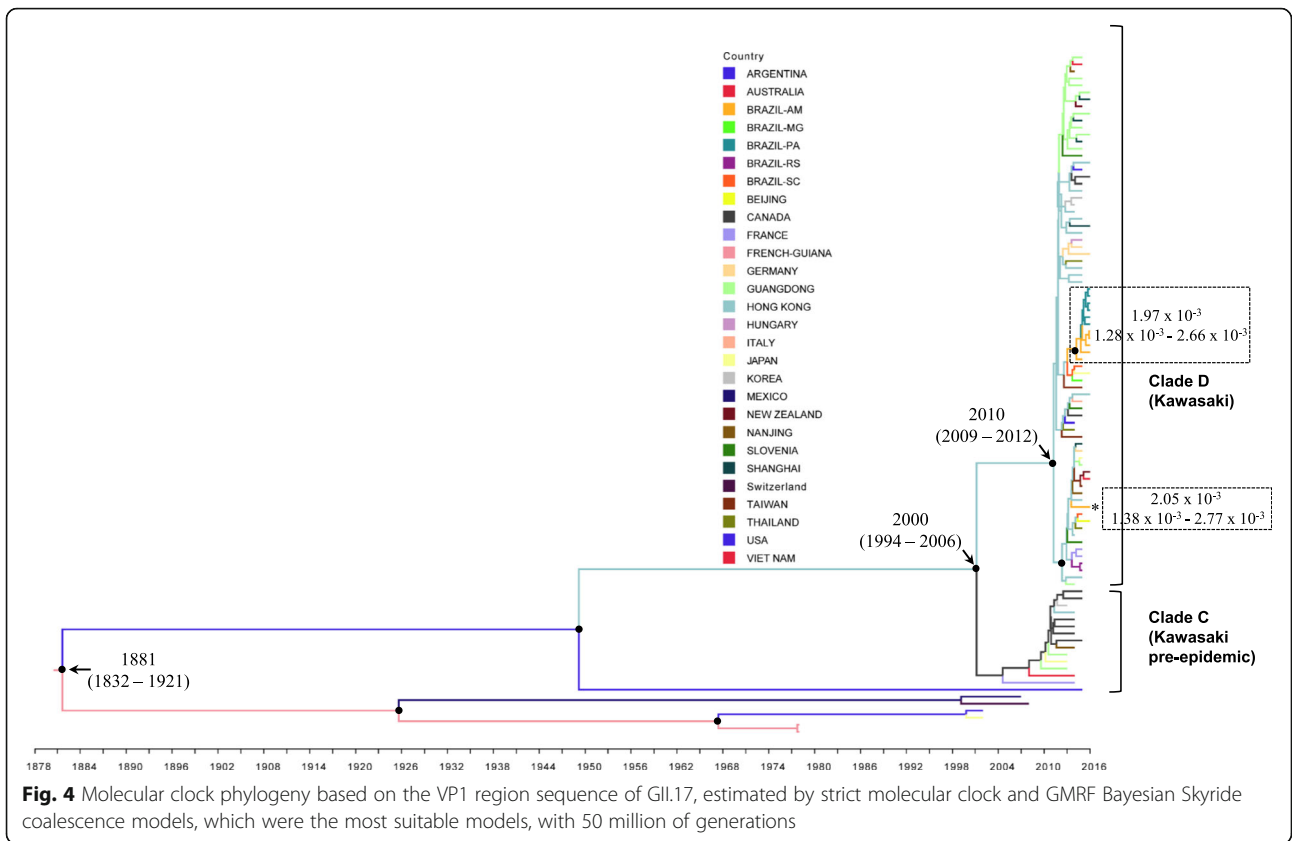


Table 2 The mean amino acid distance between clades of the NoV GII.17_2014

	A	B	C	D
A	1.51%			
B	7.07%	2.02%		
C	11.76%	11.80%	0.28%	
D	13.44%	13.24%	4.72%	0.44%

Bocavirus [36]. These results demonstrated that the maximum detection limit that can be achieved is 10^{-4} dilution. Our study demonstrated a sensitivity of 10^{-5} for the designed primer pairs.

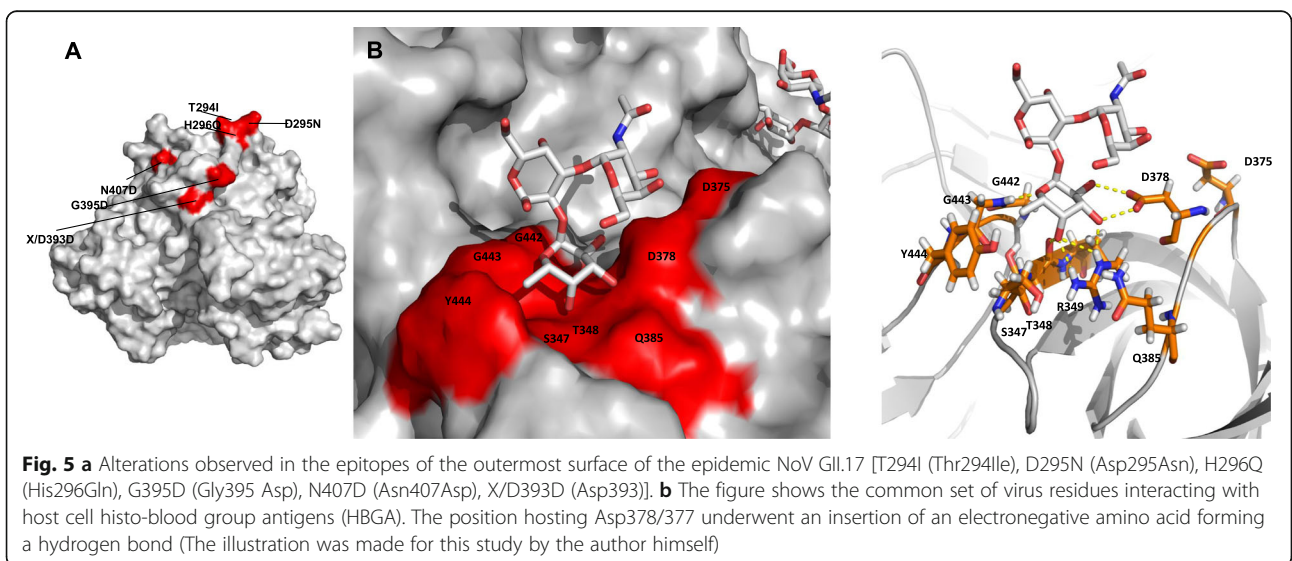
The phylogenetic trees obtained in this study demonstrated that all the Brazilian strains were grouped in clade D, which were the strains responsible for the epidemic that occurred in Asia since 2014. Sang et al. [37] suggested that strains from clades C and D, which were responsible for outbreaks in different countries, are two distinct variants as they are genetically distinct and that the minimum amino acid distance between them is



4.2%. The clade D variant replaced clade C in the 2014–2015 epidemics. Our study obtained similar amino acids variation (4.72%), corroborating with the earlier findings.

The mutation rate for strain GII.17 determined in this study was 1.95×10^{-3} substitutions/site/year with the TMRCA estimated in 1881. Sang et al. [37] identified an

evolution rate of 1.68×10^{-3} substitutions/site/year for GII.17, which is similar to the rate obtained in this study. Our investigation also corroborated with the results obtained by Parra et al. [38], which indicated that GII.17 were a static genotype accumulating only a few amino acid mutations over years.



The phylogeographic tree suggests that one introduction of the GII.17_2014 strain occurred in the Amazon Region. In previous studies, Andrade et al. [12] described four independent introduction events of this lineage in Brazil during the football World Cup 2014.

Amino acid analysis of the complete VP1 of the strains detected in the Amazon region demonstrated higher accumulation of mutations in the antigenic sites when compared to other pre-epidemic strains. A molecular characterization study conducted with several GII.17 strains circulating in Jiangsu Province, China between 2014 and 2015 identified differences in amino acids that occurred mainly in the P2 domain, particularly at positions 295–297, 376, 398–400, and 414 [39, 40].

In this study, the identified changes modified the chemical nature of amino acid at positions 294 to 296, 393/395, and 407, which may have an impact on the properties and structure of the P2 domain of the virus. Additionally, all these positions are located at the outer points in the viral capsid structure and are part of exposed loops. The study by Lindesmith et al. [41] demonstrated changes in residues 393–396 that may have contributed to the emergence of new variants of GII.17. These findings reinforce the notion that a new antigenic variant with the pandemic potential may occur by the accumulation of mutations in several sites in the subdomain P2 [8, 38].

Furthermore, two insertions were observed at positions Asp377 and Gly397. These residues interact with the fucose of the HBGA in the GII.4 strains. According to Singh et al. [38], this insertion at the loop harboring Asp378 does not change the orientation of the side chain of the protein and consequently does not alter the binding mechanism at this position. These inserts were also described by Chan et al. [42], but the role of these changes in virus binding with HBGA remains unclear.

One of the limitations of this study was the low percentage of genotyped samples (45.7%) that may have influenced the number of GII.17_2014 detection. Therefore, further studies are needed to understand the molecular epidemiological characteristics of emerging NoV strains that cause AGE.

Conclusion

In 2016, there was an increase in the circulation of the emerging GII.17_2014 strain in the Amazon Region. All the positive samples for NoV GII.17_2014 variant detected in this study were grouped in clade D, which was responsible for numerous outbreaks of AGE worldwide. One introduction of the GII.17_2014 strains originating from Hong Kong/China was observed in the Amazon region in 2013. Several accumulations of mutations were observed at several points in the NoV VP1 region,

mainly in the subdomain P2, which is a highly variable region.

These data demonstrate that the NoV GII.17 strains undergo a dynamic process of evolution, and the number of mutations exposed on their surface may explain amino acid changes in antigenic sites, which may be related to the increase in their prevalence. This is an extremely relevant inference for the studies involving immunogenicity and vaccines, which are developed considering the epidemiological importance of NoV infections worldwide.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12879-019-4628-5>.

Additional file 1. The final database comprised 85 sequences of GII.17 and 12 described sequences evaluated in this study.

Additional file 2. Best suited molecular clock and coalescence models.

Additional file 3. Changes in the amino acids found in the VP1 region of the NoV GII.17 strains that circulated in the Amazon region during 2015–2016. ↓ Red arrow down symbol nucleotide substitutions: Thr294Ile, Asp295Asn, His296Gln, Gly395Asp, and Asp407Asn. * insertions: Asp377 and Gly397. * Red Asterisk symbol Deletion: Asp393.

Additional file 4. PROCHECK results showed residues in the most favored (A, B, L), additional admissible (a, b, l, p), and generously allowed regions (~ a, ~ b ~ l, ~ p). Parameters, such as residues in the favored, allowed, and generously allowed regions are the determinants of a good model.

Additional file 5. Graph showing the average 3D-1D score for each residue. Plot and scores are generated using VERIFY 3D.

Abbreviations

AGE: Acute gastroenteritis; BLAST: Basic Local Alignment Search Tool; CD-Hit program: Cluster Database at High Identity with Tolerance; EIA: Enzyme immunoassay; HBGA: Histo-blood group antigens; LD: Detection limit; MCMC: Markov Chain Monte Carlo; MLE: Marginal likelihood estimation; NoV: Norovirus; ORF: Open reading frames; P: Protruding; PDB: Protein Data Bank; PS: Path sampling; RT-PCR: Reverse transcription polymerase chain reaction; S: Shell; SS: Stepping-stone sampling; Tm: Melting temperature; TMCA: Time of the most recent common ancestor

Acknowledgements

The authors would like to acknowledge the technical support provided to the entire staff of the Norovirus and other Enteric Viruses Laboratory (LNVE). A special thanks to the Central Laboratory (LACEN) of Amazonas and Para States involved in the Brazilian Network Surveillance Program of Viral Gastroenteritis, who conducted the samples collection.

Authors' contributions

LCPNC: prepared the manuscript, analyzed and performed the laboratory tests; LDS: Reviewed the manuscript and analyzed the results; DMT, JAMS, ICGL, and HRR: performed the laboratory tests, reviewed the manuscript and analyzed the results; RSB and ECSJ: provided support for phylogenetic analysis of the detected samples; ACRP: performed the laboratory tests; YBG: Reviewed the manuscript, analyzed the applied methodology and results and was involved in general discussions with the first author. All authors read and approved the final manuscript.

Funding

This study was supported by the Evandro Chagas Institute, Brazilian Ministry of Health. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. We are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the scholarship provided to the first author

(LCPNC) during her Master's degree in the Postgraduate Program in Parasitic Biology of the Amazon.

Availability of data and materials

All data generated or analyzed during this study are included in this article. Also, all the sequences obtained are available in GenBank database <https://www.ncbi.nlm.nih.gov/genbank/> (accession numbers: MN045201, MN045199, MN045191, MN045193, MN045195, MN045196, MN045197, MN045200, MN045203, MN045192, MN045194 and MN045198).

Ethics approval and consent to participate

All the fecal specimens used in this research were collected from the National Program for Surveillance of Rotavirus/Norovirus Gastroenteritis, in coordination with Brazilian Ministry of Health (governmental sector responsible for the administration and maintenance of Public Health in the country). This study involved only molecular characterization of viral strains, the consent to participate from the under 16 years old was waived and was submitted to and approved by the Ethics Committee on Human Research of Evandro Chagas Institute in Brazil (ethical report number 1.318.103) according to Resolution 466/12 published in the DOU n° 12, June 13, 2013 (Section 1, Page 59) and its complementary, of the Council National Health, Ministry of Health, Brazil.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Postgraduate Program in Parasitic Biology in the Amazon, Universidade do Estado do Pará, Instituto Evandro Chagas, Belém, PA, Brazil. ²Virology Section, Evandro Chagas Institute, Brazilian Ministry of Health, Rodovia BR-316, Km 7 s/n, Levilândia, Ananindeua, Pará 67030-000, Brazil. ³Postgraduate Program in Virology, Instituto Evandro Chagas, Secretaria de Vigilância em Saúde, Ministério da Saúde, Ananindeua, PA, Brazil. ⁴Virology Section, Instituto Evandro Chagas, Secretaria de Vigilância em Saúde, Ministério da Saúde, Ananindeua, PA, Brazil.

Received: 30 May 2019 Accepted: 11 November 2019

Published online: 02 December 2019

References

- Lopman BA, Steele D, Kirkwood CD, Parashar UD. The vast and varied global burden of Norovirus: prospects for prevention and control. *PLoS Med.* 2016; 13(4):e1001999.
- 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):1393–406.
- Vinjé J. Advances in laboratory methods for detection and typing of norovirus. *J Clin Microbiol.* 2015;53(2):373–81.
- Green KY. Caliciviridae: the Noroviruses. In: Knipe DM, Howley PM, editors. *Fields Virology*. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 582–604p.
- Thorne LG, Goodfellow IG. Norovirus gene expression and replication. *J en Virol.* 2014;95(2):278–91.
- Debbink K, Donaldson EF, Lindsmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *J. Virol.* 2012a;86(2):1214–26.
- Debbink K, Lindsmith LC, Donaldson EF, Baric RS. Norovirus immunity and the great escape. *PLoS Pathogens.* 2012b;8(10):e1002921.
- Bok K, Abente EJ, Realpe-Quintero M, Mitra T, Sosnovtsev SV, Kapikian AZ, Green KY. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. *J Virol.* 2009;83(22):11890–901.
- Lu J, Sun L, Fang L, Yang F, Mo Y, Lao J, et al. Gastroenteritis outbreaks caused by Norovirus GII.17, Guangdong Province, China, 2014–2015. *Emerg Infect Dis.* 2015;21(7):1240–2.
- De Graaf M, Van Beek J, Vennema H, Podkolzin AT, Hewitt J, Bucardo F, Templeton K, Mans J, Nordgren J, Reuter G, Lynch M, Rasmussen LD, Iritani N, Chan MC, Martella V, Ambert-Balay K, Vinjé J, White PA, Koopmans MP. Emergence of a novel GII.17 norovirus – End of the GII.4 era? *Euro Surveill.* 2015;20(26).
- Silva LD, Bandeira RD, Junior EC, Lima IC, Da Penha Júnior ET, Teixeira DM, Siqueira JA, Resque HR, De Abreu Campos EM, Justino MC, Linhares AC, Gabbay YB. Detection and genetic characterization of the emergent GII.17_2014 norovirus genotype among children with gastroenteritis from northern Brazil. *Infect Genet Evol.* 2017;48:1–3.
- Andrade JSR, Fumian TM, Leite JPG, Assis MR, Bello G, Mir D, Miagostovich MP. Detection and molecular characterization of emergent GII.17/Norovirus in Brazil, 2015. *Infect Genet Evol.* 2017;51:28–32.
- Degiuseppe JI, Gomes KA, Hadad MF, Parra GI, Stupka JA. Detection of novel GII.17 norovirus in Argentina, 2015. *Infect Genet Evol.* 2017;47:121–4.
- Matsushima Y, Ishikawa M, Shimizu T, Komane A, Kasuo S, Shinohara M, et al. 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.* 2015;20(2):21173.
- Kojima S, Kageyama T, Fukushi S, Hoshino FB, et al. Genogroup-specific PCR primers for detection of Norwalk-like viruses. *J Virol Methods.* 2002;100:107–14.
- Beuret C, Kohler D, Baumgartner A, Lüthi TM. Norwalk-like virus sequences in mineral waters: one year monitoring of three brands. *Appl Environ Microbiol.* 2002;68:1925–31.
- XUE L, Wu Q, Cai W, Zhang J, Guo W. Molecular characterization of new emerging GII.17 norovirus strains from South China. *Infect Genet Evol.* 2016; 40:1–7.
- Brasil. Ministério Da Agricultura Pecuária E Abastecimento. Manual de verificação de desempenho de métodos para diagnóstico molecular de doenças infecciosas na rede nacional de laboratórios agropecuários. Brasília: Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária; 2015. p. 57. MAPA/ACS.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658–9.
- Katoh S. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
- Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* 2014;30:3276–8.
- Posada D. jModelTest: phylogenetic model averaging. *Molec Biol Evol.* 2008; 25:1253–6.
- TAMURA K, STECHER G, PETERSON D, FILIPSKI A, KUMAR S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013; 30(12):2725–9.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1. 10. *Virus Evol.* 2018;4(1):vey016.
- Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. *Syst Biol.* 2006;55:195–207.
- Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol.* 2011;60: 150–60.
- Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6, 2014. <http://beast.bio.ed.ac.uk/Tracer>. Accessed 19 Oct 2016.
- Rambaut, A. FigTree – Molecular Evolution, phylogenetics and Epidemiology, 2007. <http://tree.bio.ed.ac.uk/software/figtree/>>. Accessed 14 Feb 2018.
- Laskowski RA, MacArthur MW, Moss JM, Thornton DS. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26(2):283–91.
- EISENBERG D, LÜTHY R, BOWIE JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 1997;277: 396–404.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput Chem.* 2009;2009(16):2785–91.
- ATMAR RL, ESTES MK. The epidemiologic and clinical importance of norovirus infection. *Gastroenterol Clin N Am.* 2006;35:275–90.
- SIEBENGA JJ, VENNEMA H, ZHENG DP, VINJÉ J, LEE B. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J Infect Dis.* 2009;200:802–12.
- Lee CC, Feng Y, Chen SY, Tsai CN, Lai MW, Chiu CH. Emerging norovirus GII.17 in Taiwan. *Clin Infect Dis.* 2015;61(11):1762–4.

35. Cui C, Pan L, Wang Y, Xue C, Zhu W, Zhu L, Ye C, Lu X, Song H, Fu Y, Sun Q. An outbreak of acute GII.17 norovirus gastroenteritis in a long-termcare facility in China: the role of nursing assistants. *J Infect Public Health*. 2017; 10(6):725–9.
36. Pham NT, TRINH QD, Chan-It W, Khamrin P, Shimizu H, Okitsu S, Mizuguchi M, Ushijima H. A novel RT-multiplex PCR for detection of Aichi virus, human parechovirus, enteroviruses, and human bocavirus among infants and children with acute gastroenteritis. *J Virol Methods*. 2010;169(1):193–7.
37. SANG S, YANG X. Evolutionary dynamics of GII.17 norovirus. *PeerJ*. 2018;6: e4333.
38. 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 Pathogens*. 2017;13:e1006136. <https://doi.org/10.1371/journal.ppat.1006136>.
39. Singh BK, Koromyslova A, Hefele L, Gürth C, Hansman GS. Structural Evolution of the Emerging 2014–2015 GII.17 Noroviruses. *J Virol*. 2015;90(5): 2710–5.
40. Fu J, Ai J, Jin M, Jiang C, Zhang J, Shi C, et al. Emergence of a new GII.17 norovirus variant in patients with acute gastroenteritis in Jiangsu, China, September 2014 to march 2015. *Euro Surveill*. 2015;20(24).
41. Lindesmith LC, Kocher JF, Donaldson EF, Debbink K, Mallory ML, Swann EW, Brewer-Jensen PD, Baric RS. Emergence of novel human Norovirus GII.17 strains correlates with changes in blockade antibody epitopes. *J Infect Dis*. 2017;216(10):1227–34.
42. Chan MC, Lee N, Hung TN, Kwok K, Cheung K, Tin EK, Lai RW, Nelson EA, Leung TF, Chan PK. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat Commun*. 2015;6:10061.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

