Evolutionary and Molecular Analysis of Complete Genome Sequences of Norovirus From Brazil: Emerging Recombinant Strain GII.P16/GII.4

Juliana Merces Hernandez[1], Luciana Damascena Silva[2]*, Edivaldo Costa Sousa Junior[2], Jedson Ferreira Cardoso[3], Tammy Kathlyn Amaral Reymão[1], Ana Caroline Rodrigues Portela[2], Clayton Pereira Silva de Lima[3], Dielle Monteiro Teixeira[2], Maria Silvia Souza Lucena[2], Marcio Roberto Teixeira Nunes[3] and Yvone Benchimol Gabbay[2]

[1] Postgraduate Program in Biology of Infectious and Parasitic Agents, Institute of Biological Sciences, Federal University of Pará, Belém, Brazil, [2] Virology Section, Evandro Chagas Institute, Brazilian Ministry of Health, Ananindeua, Brazil, [3] Center for Technological Innovation, Evandro Chagas Institute, Brazilian Ministry of Health, Ananindeua, Brazil

Noroviruses (NoVs) are enteric viruses that cause acute gastroenteritis, and the pandemic GII.4 genotype is spreading and evolving rapidly. The recombinant GII.P16/GII.4_Sydney strain emerged in 2016, replacing GII.P31/GII.4_Sydney (GII.P31 formerly known as GII.Pe) in some countries. We analyzed the complete genome of 20 NoV strains (17 GII.P31/GII.4_ Sydney and 3 GII.P16/GII.4_Sydney) from Belém and Manaus, Brazil, collected from 2012 to 2016. Phylogenetic trees were constructed by maximum likelihood method from 191 full NoV-VP1 sequences, demonstrated segregation of the Sydney lineage in two larger clades, suggesting that GII.4 strains associated with GII.P16 already have modifications compared with GII.P31/GII.4. Additionally, the Bayesian Markov Chain Monte Carlo method was used to reconstruct a time-scaled phylogenetic tree formed by GII.P16 ORF1 sequences ($n = 117$) and three complete GII.P16 sequences from Belém. The phylogenetic tree indicated the presence of six clades classified into different capsid genotypes and locations. Evolutionary rates of the ORF1 gene of GII.P16 strains was estimated at 2.01×10^{-3} substitutions/site/year, and the most recent common ancestors were estimated in 2011 (2011–2012, 95% HPD). Comparing the amino acid (AA) sequence coding for ORF1 with the prototype strain GII.P16/GII.4, 36 AA changes were observed, mainly in the non-structural proteins p48, p22, and RdRp. GII.P16/GII.4 strains of this study presented changes in amino acids 310, 333, 373, and 393 of the antigenic sites in the P2 subdomain, and ML tree indicating the division within the Sydney lineage according to the GII.P16 and GII.P31 polymerases. Notably, as noroviruses have high recombination rates and the GII.4 genotype was prevalent for a long time in several locations, additional and continuous evolutionary analyses of this new genotype should be needed in the future.

Keywords: norovirus, GII.P16/GII.4, emergent recombinant, gastroenteritis, molecular evolution, complete genome

INTRODUCTION

Noroviruses (NoVs) are enteric viruses belonging to the *Caliciviridae* family that commonly cause pandemic acute gastroenteritis worldwide (Pires et al., 2015). The global burden of illness is estimated to be 570.3 million cases in low- and middle-income countries (Bartsch et al., 2016). In countries such as Brazil, the risk factors for noroviruses can be five times more than those for other enteropathogens (Lima et al., 2019). NoV saffect all ages, but severe disease is particularly observed in children (less than 5 years old), elderly people, and immunocompromised individuals (Cardemil et al., 2017; Woodward et al., 2017; Riera-Montes et al., 2018).

NoVs are genetically diverse and classified into at least 10 genogroups (GI-GX) and more than 45 genotypes (ICTV Virus Taxonomy, 2020). Despite the wide variability in NoVs, the GII genogroup is responsible for most cases in humans, especially the GII.4 genotype, which has been most associated with outbreaks and sporadic cases of viral gastroenteritis. The viral genome is a single-stranded positive-sense RNA genome of approximately 7.6 kb with a polyA tail. It is organized in three open reading frames (ORF1-3) (Green, 2013). ORF1 encodes a large polyprotein that after cleavage forms six non-structural proteins, including the viral RNA-dependent RNA polymerase (RdRp) involved in replication. ORF2 encodes the major capsid protein (VP1) structurally comprised of two domains, shell (S) and protruding (P). The P domain is subdivided into the P1 and P2 subdomains. Antigenic sites and carbohydrates binding-sites map into the P2 subdomain, both playing a major role in susceptibility to norovirus infection (Lindesmith et al., 2008). ORF3 encodes a minor capsid protein (VP2).

Norovirus presents multiple genotypes, but GII.4 strains are the predominant worldwide. GII.4 noroviruses have been shown the emergence of variants that present changes in antigenic sites, which facilitate escape from herd immunity elicited by previous variants (Debbink et al., 2012; Tohma et al., 2019).

In addition, Lindesmith et al. (2018), reported a strategy of camouflaging antigenic sites to carbohydrate blocking antibodies, named "particle breathing." Since 1995, many variants causing global epidemics have emerged, including US 1995/96 in 1996 (Fankhauser et al., 1998), Farmington Hills in 2002 (Lopman et al., 2004), Hunter in 2004 (Bull et al., 2006), Den Haag in 2006 (Eden et al., 2010), New Orleans in 2009 (Yen et al., 2011), and Sydney in 2012 (van Beek et al., 2013). However, in 2014, a non-GII.4 genotype, named GII.17_Kawasaki, was reported as the main cause of outbreaks in Asia (Lu et al., 2015; Matsushima et al., 2015; Thanh et al., 2016), some countries of the Americas (Parra and Green, 2015; Degiuseppe et al., 2017; Silva et al., 2017) and Europe (De Graaf et al., 2015; Giammanco et al., 2017).

The GII.4_Sydney lineage emerged in 2012, and still causes a large number of cases. In 2016, the variant GII.4_Sydney recombined in association with a novel GII.P16 ORF1, and the circulation of this variant increased in many countries such as Germany, South Korea and United States (Choi et al., 2017; Niendorf et al., 2017; Barclay et al., 2019). In Brazil, this novel variant was identified in 59.5% (22/37) of genetically characterized samples (Barreira et al., 2017). Additionally, the increase in norovirus activity at the end of 2016 to 2017 may have occurred because of the emergent GII.P16/GII.2 recombinant (Fu et al., 2017; Kwok et al., 2017; Liu et al., 2017). This GII.P16 genotype presented several amino acid changes in non-structural proteins, such as RdRp, and these changed seemed to facilitate its transmission (Cannon et al., 2017; Tohma et al., 2017; Cheung et al., 2019). In this study, we analyzed the complete genomes of GII.4 strains collected over 5 years to understand their genetic characteristics and investigate the emergence of GII.P16/GII.4.

MATERIALS AND METHODS

Clinical Samples

This study on the emergence of the GII.P16/GII.4 strain started from previous data from studies conducted with samples from the cities Belém and Manaus in northern Brazil (Hernandez et al., 2018; Reymão et al., 2018).

Stool samples analyzed in this study were collected between 2012 and 2016. From January 2010 to December 2016, 32.8% (557/1698) of the samples from the study region were positive for norovirus by the EIA test. GII.4 was the genotype most frequently detected, with the emergence of the GII.P31/GII.4-Sydney 2012, GII.17_2014 and GII.P16/GII.4 strains. Reports about GII.P31/GII.4 and GII.17_2014, as well as the temporal distribution and emergence period of these strains, have already been described (Silva et al., 2013; Hernandez et al., 2018; Reymão et al., 2018; Costa et al., 2019).

Considering the information available on norovirus epidemiological surveillance, 28 samples were selected for whole genome sequencing. The samples were previously characterized as the genotype GII.P31/GII.4 as well as three samples as GII.P16/GII.4 by direct sequencing of short genomic fragments (from 200 to 250 bp). To provide temporal representation, the samples were distributed by year and State, as shown in **Table 1**.

Ethics Approval

This study was approved by the Ethics Committee on Human Research of Evandro Chagas Institute, Brazilian Ministry of

TABLE 1 | Data about the samples selected for this study.

Year of collection	Local (City/State)	Norovirus genotype
2012	*N* = 3 (Belém/Pará)	GII.P31/GII.4
	N = 1 (Manaus/Amazonas)	GII.P31/GII.4
2013	*N* = 3 (Belém/Pará)	GII.P31/GII.4
	N = 3 (Manaus/Amazonas)	GII.P31/GII.4
2014	*N* = 3 (Belém/Pará)	GII.P31/GII.4
	N = 3 (Manaus/Amazonas)	GII.P31/GII.4
2015	*N* = 3 (Belém/Pará)	GII.P31/GII.4
	N = 2 (Manaus/Amazonas)	GII.P31/GII.4
2016	*N* = 6 (Belém/Pará)	GII.P31/GII.4 (*N* = 3); GIIP.16/GII.4 (*N* = 3)
	N = 1 (Manaus/Amazonas)	GII.P31/GII.4
Total	28 (Belém = 18, Manaus = 10)	GII.P31/GII.4 (*N* = 25); GIIP.16/GII.4 (*N* = 3)

Health (protocol No. 0039/2011 and protocol No. 0017/2014 update No. 1.318.103 of 2015) as described by Hernandez et al. (2018) and Reymão et al. (2018).

Viral Genome Extraction and Sequencing

A fecal suspension (10% w/v) prepared in Tris/HCl/Ca^{2+} buffer was used for nucleic acid extraction by a QIAamp viral RNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer's guidelines or using the silica method (Hernandez et al., 2018). The extracted RNA was quantified by a Qubit 2.0 fluorometer using the Qubit RNA BR assay kit (Thermo Fisher). Then, the samples were subjected to reverse transcription and purified using a cDNA synthesis system kit (Roche, Branford, CT, United States) to obtain pure double-stranded DNA. The DNA was quantified using the Qubit DNA BR assay kit (Thermo Fisher Scientific) and analyzed for fragmentation and quality profile using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States) as recommended by the manufacturer. Subsequently, libraries for sequencing were prepared using the Illumina Nextera XT DNA Library Prep kit and sequenced on an Illumina HiSeq 2500 instrument (Illumina, San Diego, CA, United States) with the high-output V4 2 × 100-bp sequencing kit.

Identification, Assembly, and Genome Annotation

The adapters were removed from raw sequence reads using Trim Galore pipeline v.0.4.5[1], and the reads were quality-trimmed and filtered by Prinseq-lite (Schmieder and Edwards, 2011). Reads were assembled *de novo* by IDBA-UD assembler v.1.1.3 (Peng et al., 2012). Additionally, the 3′ and 5 ends were verified using reference mapping based on norovirus prototype strains obtained from GenBank. Genome management and annotation were conducted in Geneious 9.1.2 (Kearse et al., 2012). The whole sequences of norovirus were deposited in GenBank under the accession numbers MN525275, MN525276 and MT474045/MT474046 for the GII.P16/GII.4 strains and MT474032-MT474044/MT474047- MT474052 for the GII.P31/GII.4 strains.

Phylogenetic Analyses of Capsid

The complete norovirus ORF2 sequences from this study were aligned with a database composed by 1369 genomes obtained from all norovirus GII.4 sequences of GenBank. To remove redundant sequences, CD-HIT was applied, resulting in 947 sequences. After removing identical sequences using as cut-off based on the e-value of the classification in groups with CD-HIT, 191 genomes had full VP1 sequences that were used to reconstruct the phylogeny. All sequences were aligned with Mafft (Katoh and Standley, 2013), and maximum likelihood phylogenetic analyses based on GTR nucleotide substitution model were performed using RAxML v.8.2 with 1,000 bootstraps (Stamatakis, 2014). Tree editing was performed with Evolview (Zhang et al., 2012). To confirm the clades within GII.4_Sydney,

[1]http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

we performed another ML tree, including all GII.4 variants with GII.P31, GII.P16 and GII.P4 polymerases, totalizing 342 sequences of diverse sizes. The parameters were the same previously described for RaxML.

Evolutionary History of Recombinant NoV GII.P16

To construct a time-scaled phylogenetic tree, a dataset formed of full or near-full GII.P16 ORF1 sequences (n = 117) available in GenBank combined with three complete GII.P16 sequences from Belém was generated. The dataset was subjected to Bayesian Markov Chain Monte Carlo (MCMC) analysis implemented in BEAST software v1.10.8 (Drummond et al., 2012). GTR + I + G was the nucleotide substitution model used. To choose the most appropriate molecular clock and the spatial model, several models were tested, and higher values of path sampling (ps) and stepping stone (ss) indicated the best model fit. A final clock ran in triplicate with 50 million generations.

P2 Region Analysis

The P2 subdomain of four GII.P31/GII.4 and two GII.P16/GII.4 strains were analyzed to explore amino acid changes in antigenic sites of the VP1 protein using the Geneious 9.1.2 (Kearse et al., 2012).
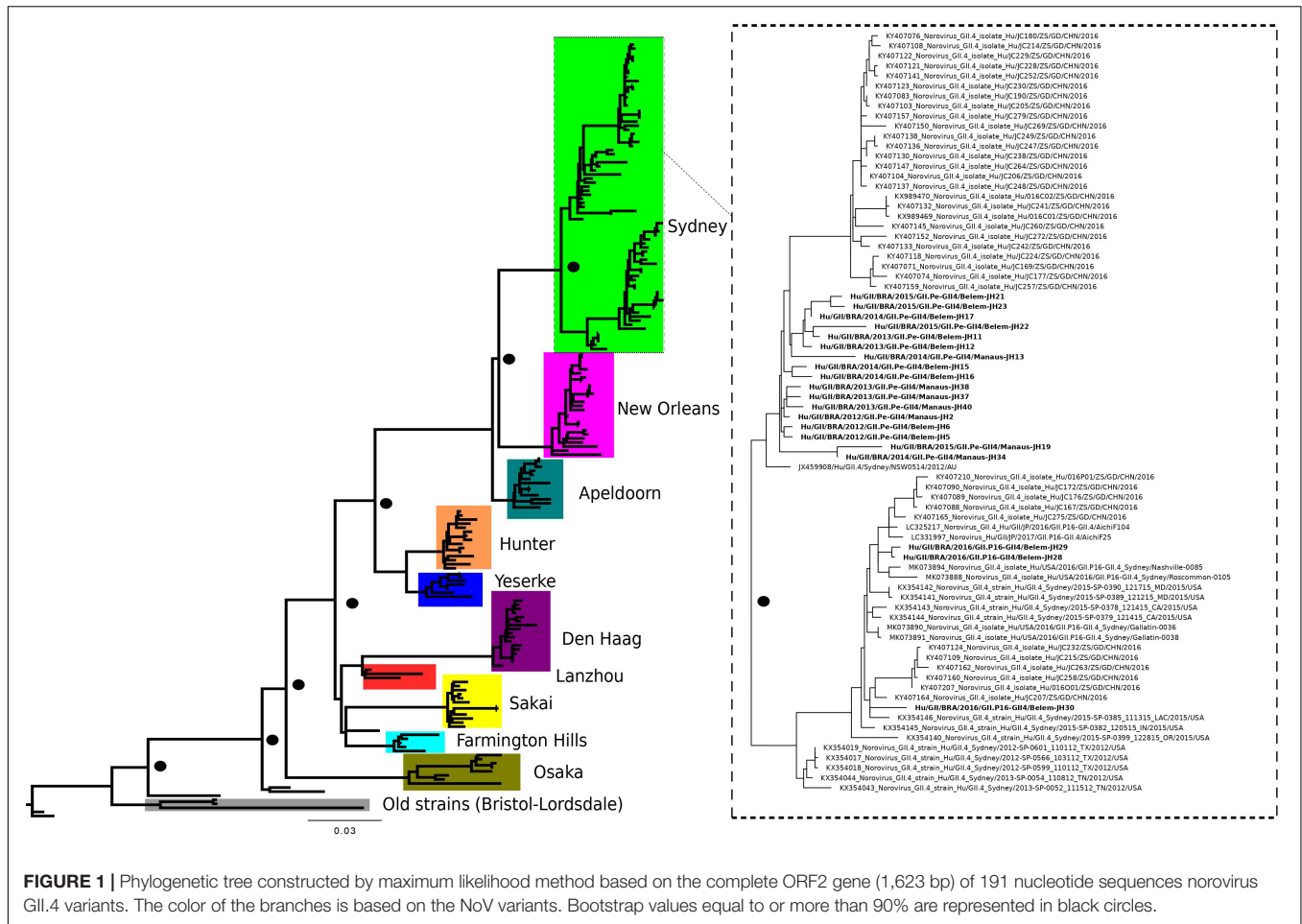
RESULTS

Twenty complete or near complete genomes of norovirus strains were obtained from the samples sequenced using Illumina HiSeq. An average of 11,510,081 reads were generated per sample. The number of reads and the average coverage of each base for NoV is shown in **Supplementary Table S1**. The sequences generated presented 7,503 nt in length on average [excluding the poly(A) tail]. The open reading frames (ORFs) were 5,094 (ORF1), 1,623 (ORF2), and 807 nt (ORF3) in length.

The maximum likelihood phylogenetic tree of the complete ORF2 gene (1,623 nt) demonstrated that the epidemic strain segregated in two larger clades within the Sydney lineage, collected from 2012 to 2016. The GII.P16/GII.4 sequences of this study clustered with other emerging recombinants from China and the United States (**Figure 1**). Sequence analysis in BLASTn demonstrated that these sequences had 98.8 and 98.6% nucleotide identity with other GII.P16/GII.4 recombinant strains from the United States (MK764013) and United Kingdom (KY887601), respectively. High nucleotide identity (>98%) was also observed with the first report of a complete genome sequence of norovirus GII.P16/GII.4, isolated in Japan (Matsushima et al., 2016). The ML tree with large number of sequences confirms the division within the Sydney lineage, according to the GII.P16 and GII.Pe polymerases (**Figure 2**). The GII.4_Sydney strains that have the GII.P16 polymerase form an independent clade of the other Sydney variants.

The capsid gene of this new recombinant (GII.P16/GII.4) shared 96.8% nucleotide identity with the capsid gene of the

FIGURE 1 | Phylogenetic tree constructed by maximum likelihood method based on the complete ORF2 gene (1,623 bp) of 191 nucleotide sequences norovirus GII.4 variants. The color of the branches is based on the NoV variants. Bootstrap values equal to or more than 90% are represented in black circles.

GII.4_Sydney prototype isolated in 2012 (JX459908) and 96.4% with GII.4[P31] from the same city of Brazil (MT474032), indicating that the nucleotide substitutions accumulated over the years, as expected for norovirus strains. GII.P16/GII.4 strains of this study presented changes in amino acids 310, 333, 373, and 393 and of the antigenic sites in the P2 subdomain (**Figure 3**).

The GII.P31/GII.4_Sydney sequence isolates in this study showed a high degree of pairwise identity (more than 98%) with Japanese (AB972473), Brazilian (MN307994, MN308017), and American (KY486271) strains. BLAST analysis showed that the capsid gene sequence (1,623 bp) shared 97.2–99% identity with the Sydney prototype isolate obtained in 2012 (JX459908).
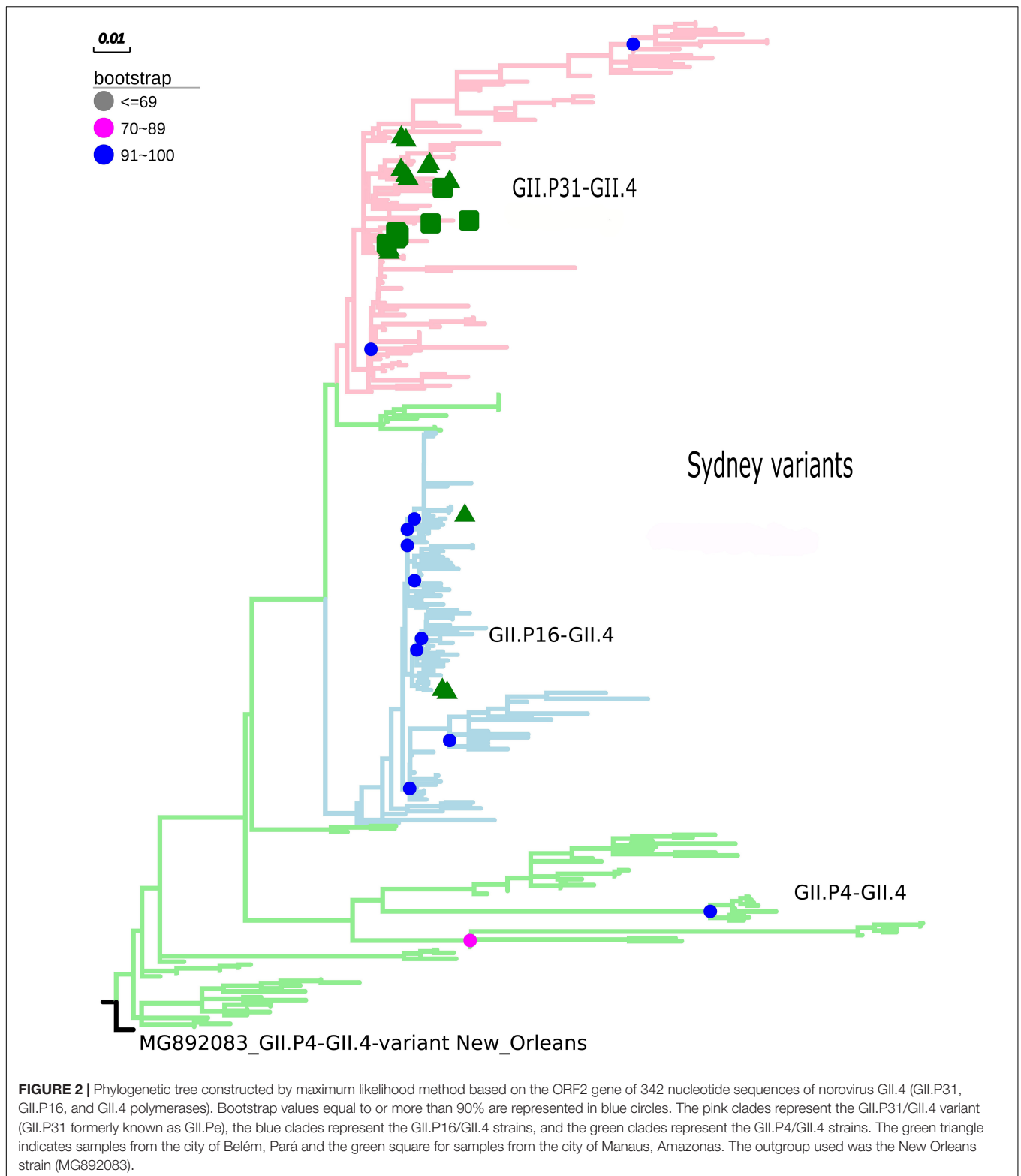
To reconstruct the time-scaled phylogenetic tree of the ORF1 gene (5,100 nt) for the RdRp of norovirus GII.P16, the Bayesian MCMC method was used with the best model lognormal GMRF Bayesian Skyrideon a strict clock. The path sampling (PS) and stepping-stone sampling (SS) parameters used in the evolutionary analysis are shown in **Supplementary Table S3**. The phylogenetic tree formed six clades classified into different capsid genotypes and locations. The cluster GII.P16/GII.4 was closer to the GII.P16/GII.2 Chinese strains. Notably, GII.P16-GII.2 re-emerged in Japan (2013–2014) with nucleotide modifications. Furthermore, the cluster

GII.P16-GII.4_Sydney_2012 contained samples from the United Kingdom, United States and Japan described in 2016, including the LC175467 prototype (Matsushima et al., 2016; **Figure 4**). The evolutionary rates of the ORF1 gene of norovirus GII.P16 strains was estimated at 2.01×10^{-3} substitutions/site/year. Moreover, the most recent common ancestor (TMRCA) of GII.P16 strains was estimated in 2011 (2011–2012, 95% HPD) (**Figure 4**).

In addition, comparison of the amino acid (AA) sequence coding for ORF1 with the sequence of the prototype strain GII.P16/GII.4 (LC175468) identified 36 AA changes in the ORF1 gene, 11 of which had not yet been described in the literature (**Supplementary Table S2**). We found mutations in non-structural proteins such as p48, p22, and RdRp. Additionally, eleven of these non-synonymous changes modified the chemical nature and polarity of the AAs (**Supplementary Table S2**).

DISCUSSION

Here, we demonstrated the evolutionary and molecular analysis of GII.4 strains detected in the states of Pará and Manaus, Brazil. In Brazil, the epidemiology of the norovirus GII.4 strain has been

FIGURE 2 | Phylogenetic tree constructed by maximum likelihood method based on the ORF2 gene of 342 nucleotide sequences of norovirus GII.4 (GII.P31, GII.P16, and GII.4 polymerases). Bootstrap values equal to or more than 90% are represented in blue circles. The pink clades represent the GII.P31/GII.4 variant (GII.P31 formerly known as GII.Pe), the blue clades represent the GII.P16/GII.4 strains, and the green clades represent the GII.P4/GII.4 strains. The green triangle indicates samples from the city of Belém, Pará and the green square for samples from the city of Manaus, Amazonas. The outgroup used was the New Orleans strain (MG892083).

well established in recent years. Previous studies conducted by Hernandez et al. (2018) and Reymão et al. (2018) demonstrated the emergence of the GII.P31/GII.4_Sydney lineage in 2012 and its high frequency from 2012 to 2016. Also, Fioretti et al. (2014)

observed the rapid evolutionary dynamics in GII.4 variants in several regions of Brazil.

GII.4_Sydney variant predominated for several years with the GII.P31 (GII.Pe) polymerase. However, in 2015–2016 emerged as

GII.4 Variant	Genotype	Accession number	Year	Local	Capsid Protein residues (A–H epitopes)
Hu/GII.4/Sydney/NSW0514/2012/AU (**Reference strain**)	GII.P31/GII.4	JX459908	2012	Australia	T G S R N E D R V I T E G T T H R S R N T H Y S A D A R N D
Hu/USA2012/GII.Pe_GII.4_Sydney/WI2138/GII	GII.P31/GII.4	KY496327	2012	USA	T G S R N E D H V I A E G T T H R S R N T H Y S A D A R N N
Hu/NP/2013/GII.Pe-GII.4_Sydney_2012/ETR-NV-080/GII.Pe-GII.4_Sydney_2012	GII.P31/GII.4	MH702258	2013	Nepal	T G S R N E D H V I T E G T T H R S R N T H Y S A D A R N N
Hu/GII.Pe_GII.4/JH5/2012/PA/BRA/GII.PE/GII.4 (This study)	GII.P31/GII.4	MT474051	2012	Belem-Brazil	T G S R N E D H V I T E S T T H R S R N N H Y S A D A R N N
Hu/GII.Pe_GII.4/JH2/2012/AM/BRA/GII.PE/GII.4 (This study)	GII.P31/GII.4	MT474044	2012	Amazonas-Brazil	T G S R N E D H V I T E S T T H R S R N N H Y S A D A R N N
Hu/GII.Pe_GII.4/JH11/2013/PA/BRA/GII.PE/GII.4 (This study)	GII.P31/GII.4	MT474032	2013	Belem-Brazil	T G S H N E N H V I T E S T T H R S R N N H Y S A D A R N N
Hu/GII.4/JH17/2014/PA/BRA/GII.PE/GII.4 (This study)	GII.P31/GII.4	MT474039	2014	Belem-Brazil	T G S H N E N H V I T E S T T H R S R N N H Y S A D A R N N
Hu/GII/JP/2016/GII.P16_GII.4_Sydney2012	GII.P16/GII.4	LC175468	2016	Kawasaki-Japan	T G S R N E D H M I T E S T T H R S R N N H Y S A D A R N N
Hu/USA2015/GII.P16-GII.4Sydney/Pasadena3477/GII.P16.GII.4	GII.P16/GII.4	KY947550	2015	USA	T G S R N E D H M I T E S T T H R S R N N H Y S A D A R N N
Norovirus_GII.4_isolate_Hu/USA/2016/GII.P16-GII.4_Sydney/Roscommon-0105	GII.P16/GII.4	MK073888	2016	Roscommon-USA	T G S R N E D H V I T E S T T H R S R N T H Y S A D A R N N
Norovirus_GII.4_isolate_Hu/USA/2016/GII.P16-GII.4_Sydney/Nashville-0085	GII.P16/GII.4	MK073894	2016	Nashville-USA	T G S R N E D H M I T E S T T H R S R N T H Y S A D A R N N
Hu/BRA/2016/GII.P16_GII.4/ES_LVGI99/GII.P16_GII.4	GII.P16/GII.4	MF681696	2016	Brazil	T G S R N E D H M I T E S T T H R S R N N H Y S A D A R N N
Hu/GII/BRA/2016/GII.P16-GII4/Para-JH28 (This study)	GII.P16/GII.4	MN525275	2016	Belem-Brazil	T G S R N E D H M I T E S T T H R S R N T H Y S A D A R N N
Hu/GII/BRA/2016/GII.P16-GII4/Para-JH29 (This study)	GII.P16/GII.4	MN525276	2016	Belem-Brazil	T G S R N E D H M I T E S T T H R S R N T H Y S A D A R N N
Norovirus_Hu/GII/JP/2017/GII.P16-GII.4/AichiF25	GII.P16/GII.4	LC331997	2017	Aichi-Japan	T G S R N E D H M I T E S T T H R S R N N H Y S A D A R N N
Grant0016/GII.P16-GII.4Sydney	GII.P16/GII.4	MK752933	2018	USA	T G S H N E N H M I T E G T T H R S R N N H Y S A D A R N N

■ Residues that differ from Sydney variant isolated in 2012

FIGURE 3 | Alignment of antigenic residues of major capsid protein (P2 region) of GII.P31/GII.4 (GII.P31 formerly known as GII.Pe) and GII.P16/GII.4 strains. Residues mapping on previously characterized A-H epitopes. The prototype strain representative of GII.P31/GII.4 variant that emerged 2012 is highlighted in gray. Amino acid residues that differ from those of the prototype is highlighted in blue. Sequences of this study are in bold.
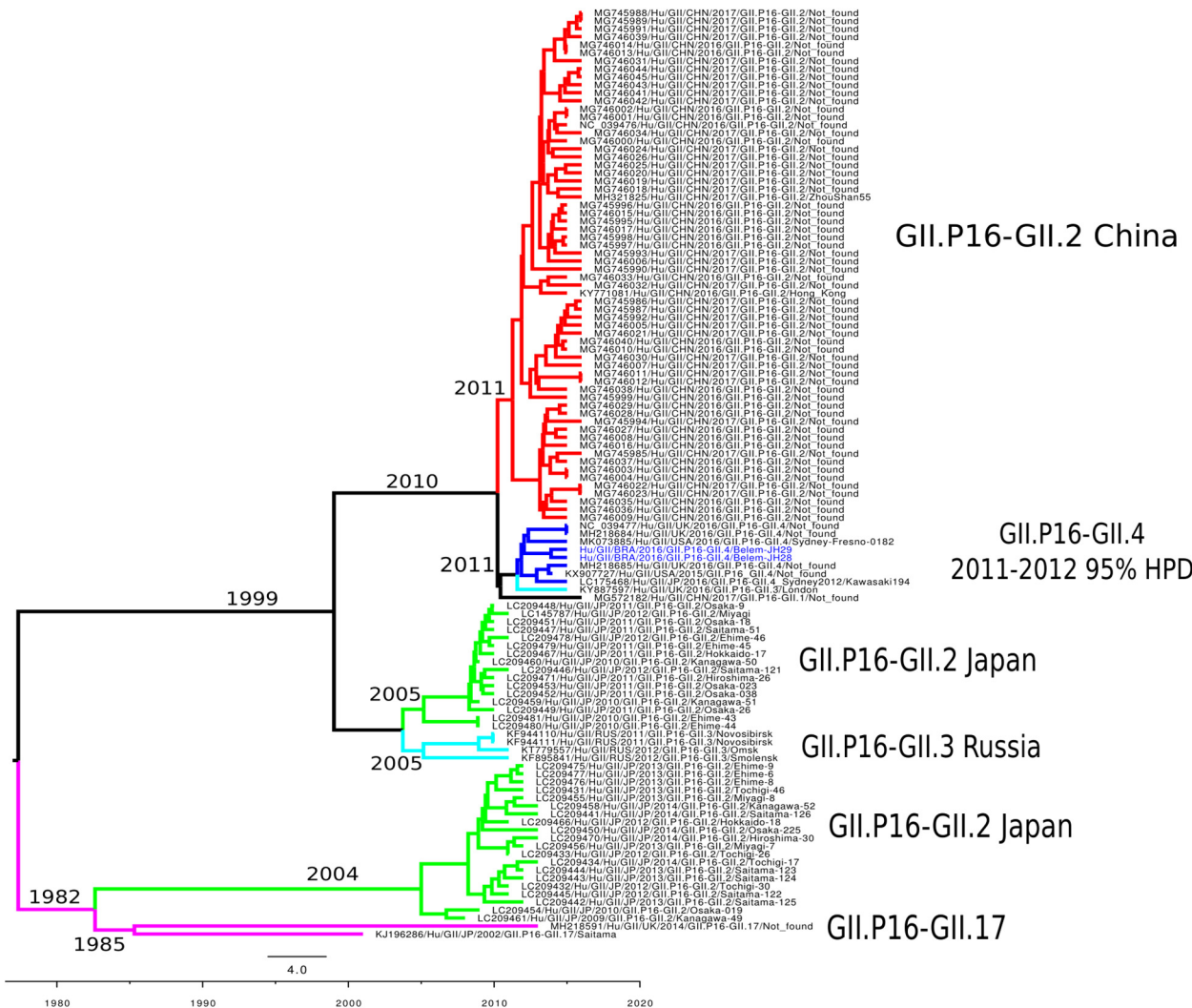
FIGURE 4 | Time-scaled phylogenetic tree of the 117 sequences based on the whole nucleotide sequence of the ORF1 gene of human norovirus GII.P16 genotype strains. The tree was constructed using the Bayesian Markov Chain Monte Carlo method (MCMC) and the GTR + I + G nucleotide substitution model. Branches are colored according to the distinct norovirus genotype. Norovirus strains isolated in this study are colored blue. The scale bar represents the unit of time (years).

GII.P16/GII.4_Sydney recombinant. In the same period, it was observed the co-circulation of all circulating pandemic strains related worldwide in samples from Northern region of Brazil, including the GII.P16/GII.4 (Costa et al., 2019). In addition, Barreira et al. (2017) detected GII.P16/GII.4 circulating in the Southeastern Brazilian coast state among children up 11 years old. Therefore, this study was conducted to investigate the molecular features of NoV-GII.4 strains in two states of the Northern region of Brazil, to better understand the circulation of these pandemic viruses in this country.

The GII.P16/GII.4 strain emerged in the Amazon region in 2015 and was detected in 16.9% (16/95) of the cases during 2015–2016, being the second most prevalent virus since GII.P31/GII.4, which corresponded to 50.5% (48/95) of the NoV-positive cases (Costa et al., 2019). In this study, nucleotide analysis by BLASTn indicated that the strains circulating in Brazil had high identity with sequences of GII.P16/GII.4 from the United States (Barclay et al., 2019) and United Kingdom (Ruis et al., 2017). These data suggest a possible importation of this strain from other countries, considering previous descriptions about the circulation of this genotype in other places.

Several studies have investigated the evolutionary pattern of norovirus, especially using partial regions of the genome and the VP1 gene (Qiao et al., 2016; Parra et al., 2017; Tohma et al., 2018, 2019). However, a very limited number of studies have evaluated the evolutionary rate of the ORF1 gene. We found a time-scaled evolutionary tree of the complete ORF1 similar to that observed using GII.P16/GII.2 strains (2.03×10^{-3} s/s/y) (Nagasawa et al., 2018). Ozaki et al. (2018) estimated the evolutionary rate at 2.82×10^{-3} s/s/y for strains of the GII genogroup considering RdRp; however, each genotype has an evolutionary mechanism. Evolutionary analyses may vary depending on the genotype and the database used (Ozaki et al., 2018). The evolutionary rate found in our study is consistent for a conserved region of the genome such as ORF1. Viral genes involved in replication tend to undergo fewer changes in their proteins compared to the viral capsid. Thus, any change in these genes can have a major impact on viral fitness as proposed for the GII.P16/GII.4 strains.

While in this study we removing identical sequences to decrease the computational time-consuming, creating a selection-bias, we possibly ignored important temporal information. This limitation may have influenced to fully recapitulate the evolutionary history of GII.P16 sequences, including the tMRCA values. Nagasawa et al. (2018) found a common ancestor for the GII.P16/GII.2 strain detected between 2016 and 2017 in 1999, 17 years earlier that when they predominated. The only study that analyzed the tMRCA of GII.P16/GII.4 strains found values 2 years later (2013) (Ruis et al., 2017) of those obtained in this study (2011), which reinforces the importance of conducting more studies like this one. This difference may be linked to the non-random selection of sequences, however, more studies should be conducted to clarify the ancestry of this recombinant.

The GII.P16 genotype has already been associated with several different capsid genotypes (GII.P16-GII.2, GII.P16-GII.3,

GII.P16-GII.4, GII.P16-GII.1, GII.P16-GII.12, GII.P16-GII.10) (Ruis et al., 2017; Nagasawa et al., 2018; Barclay et al., 2019). Previous studies suggested that the GII.4_Sydney strain acquired GII.P16 from the GII.P16-GII.2 strains (Tohma et al., 2017; Lun et al., 2018). This hypothesis was supported in this study, since strains clustered closer in the temporal reconstruction tree, and the TMRCA of both strains emerged in 2010. In conclusion, the results obtained strongly suggest that the recombinant GII.P16/GII.4 originated from the strains GII.P16/GII.2 and GII.P31/GII.4_Sydney.

In recent years, most studies on norovirus immunogenicity have focused on the antigenic sites involved in the blocking of HBGA carbohydrates, namely A-G (Debbink et al., 2012; Tohma et al., 2019). Several studies have shown that amino acid changes in these epitopes result in the ability to escape from the action of the immune system, leading to global epidemics, as demonstrated in the GII.4 Sydney_2012 lineage. In this study, we analyzed the epitopes (A-G), extensions of previously predicted and new uncharacterized antigenic sites proposed by Tohma et al. (2019). The GII.P16/GII.4 strains presented the same changes in amino acids 310, 333, 373, and 393 of the antigenic sites in the P2 subdomain already described previously among the GII.P16_GII.4 strains (**Figure 3**) (Cannon et al., 2017; Ruis et al., 2017), which were not presented in the reference GII.P31/GII.4_Sydney_2012 strain. Some of these changes seem to be very frequent in GII.P31/GII.4 strains. These corroborates with previous studies that observed the new GII.P16 polymerase recombinant with GII.4 capsid as a component of viral fitness, which might have influenced the transmissibility and spread of this virus (Ruis et al., 2017; Tohma et al., 2019).

Therefore, the cause of the emergence of this strain in several countries could be changes occurring in the ORF1 region. We found AA changes in non-structural proteins such as p48, p22, and RdRp that could directly affect viral replication. Considering specifically p22 and p48, diverse pathways can be altered, as these proteins have a role in blocking the host secretory pathway, which avoids cytokine secretion and the complement system cascade (Roth and Karst, 2016). Considering that noroviruses have high recombination rates and that the GII.4 genotype was prevalent in several places for a long time, additional and continuous evolutionary analyses of this genotype should be needed in the future.

While in this study we sequenced only samples from two states, making it difficult to fully recapitulate the evolutionary history of these pandemic viruses in Brazil, we present the first complete (or nearly complete) genome sequences of GII.P16/GII.4 viruses in Brazil. This information could facilitate the understanding of geographical distribution of this genotype at global level.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee on Human Research of Evandro 117 Chagas Institute, Brazilian Ministry of Health (protocol No. 0039/2011 and protocol No. 118 0017/2014 update No. 1.318.103 of 2015). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JH contributed to all sections, including laboratory tests, molecular analyses, and writing the manuscript. LS revised the manuscript and performed the study design. TR wrote and revised the manuscript. ES and JC performed phylogenetic analysis of the sequences and all the bioinformatics approaches. AP and CL participated in laboratory analyses. ML and DT processed the clinical samples. YG and MN contributed to data analysis and the editing of the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01870/full#supplementary-material

REFERENCES

Barclay, L., Cannon, J. L., Wikswo, M. E., Phillips, A. R., Browne, H., Montmayeur, A. M., et al. (2019). Emerging novel GII.*P16* noroviruses associated with multiple capsid genotypes. *Viruses* 11:E535. doi: 10.3390/v11060535

Barreira, D. M. P. G., Fumian, T. M., Tonini, M. A. L., Volpini, L. P. B., Santos, R. P., Ribeiro, A. L. C., et al. (2017). Detection and molecular characterization of the novel recombinant norovirus GII.*P16*-GII.4 Sydney in southeastern Brazil in 2016. *PLoS One* 12:e0189504. doi: 10.1371/journal.pone.0189504

Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J., and Lee, B. Y. (2016). Global economic burden of norovirus gastroenteritis. *PLoS One* 11:e0151219. doi: 10.1371/journal.pone.0151219

Bull, R. A., Tu, E. T., McIver, C. J., Rawlinson, W. D., and White, P. A. (2006). Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis. *J. Clin. Microbiol.* 44, 327–333. doi: 10.1128/jcm.44.2.327-333.2006

Cannon, J. L., Barclay, L., Collins, N. R., Wikswo, M. E., Castro, C. J., Magaña, L. C., et al. (2017). 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.* 55, 2208–2221. doi: 10.1128/jcm.00455-17

Cardemil, C. V., Parashar, U. D., and Hall, A. J. (2017). Norovirus infection in older adults: epidemiology, risk factors, and opportunities for prevention and Control. *Infect. Dis. Clin. North Am.* 31, 839–870. doi: 10.1016/j.idc.2017.07.012

Cheung, S., Kwok, K., Zhang, L. Y., Mohammad, K. N., Lui, G., Lee, N., et al. (2019). Higher Viral load of emerging norovirus GII.*P16*-GII.2 than Pandemic GII.4 and Epidemic GII.17, Hong Kong, China. *Emerg. Infect. Dis.* 25, 119–122. doi: 10.3201/eid2501.180395

Choi, Y. S., Koo, E. S., Kim, M. S., Choi, J. D., Shin, Y., and Jeong, Y. S. (2017). Re-emergence of a GII. 4 Norovirus Sydney 2012 variant equipped with GII. P16 RdRp and its predominance over novel variants of GII.17 in South Korea in 2016. *Food Environ. Virol.* 9, 168–178. doi: 10.1007/s12560-017-9278-4

Costa, L. C. P. N., Teixeira, D. M., Portela, A. C. R., de Lima, I. C. G., da Silva Bandeira, R., Sousa Júnior, E. C., et al. (2019). Molecular and evolutionary characterization of norovirus GII.17 in the northern region of Brazil. *BMC Infect. Dis.* 19:1021. doi: 10.1186/s12879-019-4628-5

De Graaf, M., Van Beek, J., Vennema, H., Podkolzin, A., Hewiit, J., Bucardo, F., et al. (2015). Emergence of a novel GII.*17* norovirus—end of the GII.4 era? *Euro. Surveill.* 20:21178. doi: 10.2807/1560-7917.es2015.20.26.21178

Debbink, K., Donaldson, E. F., Lindesmith, L. C., and Baric, R. S. (2012). Genetic mappingof a highly variable norovirus GII.4 blockade epitope: potential role in escape from humanherd immunity. *J. Virol.* 86, 1214–1226. doi: 10.1128/JVI.06189-11

Degiuseppe, J. I., Gomes, K. A., Hadad, M. F., Parra, G. I., and Stupka, J. A. (2017). Detection of novel GII.*17* norovirus in Argentina, 2015. *Infect. Genet. Evol.* 47, 121–124. doi: 10.1016/j.meegid.2016.11.026

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Eden, J., Bull, R. A., Tu, E., McIver, C. J., Lyon, M. J., Marshall, J. A., et al. (2010). Norovirus GII.4 *variant* 2006b caused epidemics of acute gastroenteritis in Australia during 2007 and 2008. *J. Clin. Virol.* 49, 265–271. doi: 10.1016/j.jcv.2010.09.001

Fankhauser, R. L., Noel, J. S., Monroe, S. S., Ando, T., and Glass, R. I. (1998). Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United States. *J. Infect. Dis.* 178, 1571–1578. doi: 10.1086/314525

Fioretti, J. M., Bello, G., Rocha, M. S., Victoria, M., Leite, J. P. G., and Miagostovich, M. P. (2014). Temporal dynamics of norovirus GII.4 variants in Brazil between 2004 and 2012. *PLoS One* 9:e92988. doi: 10.1371/journal.pone.0092988

Fu, J. G., Shi, C., Xu, C., Lin, Q., Zhang, J., Yi, Q. H., et al. (2017). Outbreaks of acute gastroenteritis associated with a re-emerging GII.*P16*-GII.2 norovirus in the spring of 2017 in Jiangsu, China. *PLoS One* 12:e0186090. doi: 10.1371/journal.pone.0186090

Giammanco, G. M., De Grazia, S., Bonura, F., Cappa, V., Muli, S. L., Pepe, A., et al. (2017). Norovirus GII.*17* as major epidemic strain in Italy, Winter 2015-16. *Emerg. Infect. Dis.* 23, 1206–1208. doi: 10.3201/eid2307.161255

Green, K. (2013). "Caliciviridae: the noroviruses," in *Fields' Virology*, eds D. M. Knipe and P. M. Howley (Philadelphia, PA: Lippincott Williams & Wilkins).

Hernandez, J. M., Silva, L. D., Junior, E. C. S., Bandeira, R. S., Rodrigues, E. A. M., Lucena, M. S. S., et al. (2018). Molecular epidemiology and temporal evolution of norovirus associated with acute gastroenteritis in Amazonas state, Brazil. *BMC Infect. Dis.* 18:147. doi: 10.1186/s12879-018-3068-y

ICTV Virus Taxonomy (2020). *Genus: Norovirus.* Available online at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/caliciviridae/1167/genus-norovirus (accessed April 02, 2020).

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-havas, S., Cheung, M., and Sturrock, S. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kwok, K., Niendorf, S., Lee, N., Hung, T. N., Chan, L. Y., Jacobsen, S., et al. (2017). Increased detection of emergent recombinant norovirus GII.*P16*-GII.2 Strains in Young Adults, Hong Kong, China, 2016-2017. *Emerg. Infect. Dis.* 23, 1852–1855. doi: 10.3201/eid2311.170561

Lima, A. A. M., Oliveira, D. B., Quetz, J. S., Havt, A., Prata, M. M. G., Lima, I. F. N., et al. (2019). Etiology and severity of diarrheal diseases in infants at the semiarid region of Brazil: a case-control study. *PLoS Negl. Trop. Dis.* 13:e0007154. doi: 10.1371/journal.pntd.0007154

Lindesmith, L. C., Donaldson, E. F., Lobue, A. D., Cannon, J. L., Zheng, D. P., Vinje, J., et al. (2008). Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5:e31. doi: 10.1371/journal.pmed.0050031

Lindesmith, L. C., Mallory, M. L., Debbink, K., Donaldson, E. F., Brewer-Jensen, P. D., Swann, E. W., et al. (2018). Conformational occlusion of blockade antibody epitopes, a novel mechanism of GII.4 human norovirus immune evasion. *mSphere* 3:e00518-17. doi: 10.1128/mSphere.00518-17

Liu, L. T., Kuo, T. Y., Wu, C. Y., Liao, W. T., Hall, A. J., and Wu, F. T. (2017). Recombinant GII.*P16*-GII.2 Norovirus, Taiwan, 2016. *Emerg. Infect. Dis.* 23, 1180–1183. doi: 10.3201/eid2307.170212

Lopman, B., Vennema, H., Kohli, E., Pothier, P., Sanchez, A., Negredo, A., et al. (2004). Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* 363, 682–688. doi: 10.1016/s0140-6736(04)15641-9

Lu, J., Sun, L., Fang, L., Yang, F., Mo, Y., Lao, J., et al. (2015). Gastroenteritis outbreaks caused by norovirus GII.*17*, Guangdong Province, China, 2014-2015. *Emerg. Infect. Dis.* 21, 1240–1242. doi: 10.3201/eid2107.150226

Lun, J. H., Hewitt, J., Yan, G. J. H., Enosi Tuipulotu, D., Rawlinson, W. D., and White, P. A. (2018). Recombinant GII.*P16*/GII.4 Sydney 2012 was the dominant norovirus identified in Australia and New Zealand in 2017. *Viruses* 10:E548. doi: 10.3390/v10100548

Matsushima, Y., Ishikawa, M., Shimizu, T., Komane, A., Kasuo, S., Shinohara, M., et al. (2015). Genetic analyses of GII.*17*norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro. Surveill.* 20:21173. doi: 10.2807/1560-7917.es2015.20.26.21173

Matsushima, Y., Shimizu, T., Ishikawa, M., Komane, A., Okabe, N., Ryo, A., et al. (2016). Complete Genome sequence of a recombinant GII.*P16*-GII.4 norovirus detected in Kawasaki City, Japan, in 2016. *Genome Announc.* 4:e001099-16. doi: 10.1128/genomeA.01099-16

Nagasawa, K., Matsushima, Y., Motoya, T., Mizukoshi, F., Ueki, Y., Sakon, N., et al. (2018). Genetic Analysis of human norovirus strains in Japan in 2016-2017. *Front. Microbiol.* 9:1. doi: 10.3389/fmicb.2018.00001

Niendorf, S., Jacobsen, S., Faber, M., Eis-Hübinger, A., Hofmann, J., Zimmermann, O., et al. (2017). Steep rise in norovirus cases and emergence of a new recombinant strain GII.*P16*-GII.2, Germany, winter 2016. *Euro. Surveill.* 22:30447. doi: 10.2807/1560-7917.ES.2017.22.4.30447

Ozaki, K., Matsushima, Y., Nagasawa, K., Motoya, T., Ryo, A., Kuroda, M., et al. (2018). Molecular evolutionary analyses of the RNA-dependent RNA polymerase region in norovirus genogroup II. *Front. Microbiol.* 9:3070. doi: 10.3389/fmicb.2018.03070

Parra, G. I., and Green, K. Y. (2015). Genome of emerging norovirus GII.*17*, United States, 2014. *Emerg. Infect. Dis.* 21, 1477–1479. doi: 10.3201/eid2108.150652

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

Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler forsingle-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., et al. (2015). Aetiology-specific estimates of the global and regional incidence and mortality of diarrheal diseases commonly transmitted through food. *PLoS One* 10:e0142927. doi: 10.1371/journal.pone.0142927

Qiao, N., Wang, X. Y., and Liu, L. (2016). Temporal evolutionary dynamics of norovirus GII.4 variants in China between 2004 and 2015. *PLoS One* 11:e0163166. doi: 10.1371/journal.pone.0163166

Reymão, T. K. A., Fumian, T. M., Justino, M. C. A., Hernandez, J. M., Bandeira, R. S., Lucena, M. S. S., et al. (2018). Norovirus RNA in serum associated with increased fecal viral load in children: detection, quantification and molecular analysis. *PLoS One* 13:e0199763. doi: 10.1371/journal.pone.0199763

Riera-Montes, M., O'Ryan, M., and Verstraeten, T. (2018). Norovirus and rotavirus disease severity in children: systematic review and meta-analysis. *Pediatr. Infect. Dis. J.* 37, 501–505. doi: 10.1097/INF.0000000000001824

Roth, A. N., and Karst, S. M. (2016). Norovirus mechanisms of immune antagonism. *Curr. Opin. Virol.* 16, 24–30. doi: 10.1016/j.coviro.2015.11.005

Ruis, C., Roy, S., Brown, J. R., Allen, D. J., Goldstein, R. A., and Breuer, J. (2017). The emerging GII.*P16*-GII.4 Sydney 2012 norovirus lineage is circulating worldwide, arose bylate-2014 and contains polymerase changes that may increase virus transmission. *PLoS One* 12:e0179572. doi: 10.1371/journal.pone.0179572

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Silva, L. D., Bandeira, R. D., Junior, E. C., Lima, I. C., Penha Júnior, E. T., Teixeira, D. M., et al. (2017). Detection and genetic characterization of the emergent GII.*17*_2014 norovirus genotype among children with gastroenteritis from Northern Brazil. *Infect. Genet. Evol.* 48, 1–3. doi: 10.1016/j.meegid.2016.11.027

Silva, L. D., Rodrigues, E. L., Lucena, M. S., Lima, I. C., Oliveira, D. S., Soares, L. S., et al. (2013). Detection of the pandemic norovirus variant GII.4 Sydney 2012 in Rio Branco, state of Acre, northern Brazil. *Mem. Inst. Oswaldo. Cruz.* 108, 1068–1070. doi: 10.1590/0074-0276130293

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Thanh, H. D., Nguyen, T. H., Lim, I., and Kim, W. (2016). Emergence of norovirus GII.*17* variants among children with acute gastroenteritis in South Korea. *PLoS One* 11:e0154284. doi: 10.1371/journal.pone.0154284

Tohma, K., Lepore, C. J., Ford-Siltz, L. A., and Parra, G. I. (2017). Phylogenetic analyses suggest that factors other than the capsid protein play a role in the epidemic potential of GII.2 *Norovirus. mSphere* 2:e00187-17. doi: 10.1128/mSphereDirect.00187-17

Tohma, K., Lepore, C. J., Ford-Siltz, L. A., and Parra, G. I. (2018). Evolutionary dynamics of non-GII genotype 4 (GII.4) noroviruses reveal limited and independent diversification of variants. *J. Gen. Virol.* 99, 1027–1035. doi: 10.1099/jgv.0.001088

Tohma, K., Lepore, C. J., Gao, Y., Ford-Siltz, L. A., and 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:e002202-19. doi: 10.1128/mBio.02202-19

van Beek, J., Ambert-Balay, K., Botteldoorn, N., Eden, J. S., Fonager, J., Hewitt, J., et al. (2013). Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, *late* 2012. *Euro. Surveill.* 18, 8–9.

Woodward, J., Gkrania-Klotsas, E., and Kumararatne, D. (2017). Chronic norovirus infection and common variable immunodeficiency. *Clin. Exp. Immunol.* 188, 363–370. doi: 10.1111/cei.12884

Yen, C., Wikswo, M. E., Lopman, B. A., Vinje, J., Parashar, U. D., and Hall, A. J. (2011). Impact of an emergent norovirus variant in 2009 on norovirus outbreak activity in the United States. *Clin. Infect. Dis.* 53, 568–571. doi: 10.1093/cid/cir478

Zhang, H., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* 40, W569–W572. doi: 10.1093/nar/gks576