Infection, Genetics and Evolution 71 (2019) 67–75

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

# Molecular epidemiology of G12 rotavirus strains during eight consecutive epidemic seasons in the Basque Country (North of Spain), 2010–2018

Ainara Arana[a,*], Khuzwayo C. Jere[b,c,d], Chrispin Chaguza[b,e,f], Milagrosa Montes[a,g], Miriam Alkorta[a], Miren Iturriza-Gomara[b,h], Gustavo Cilla[a,g]

[a] Microbiology Department, Donostia University Hospital - Biodonostia Health Research Institute, San Sebastián, Spain
[b] Institute of Infection & Global Health, University of Liverpool, Ronald Ross Building, Liverpool, UK
[c] Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
[d] Department of Medical Laboratory Sciences, College of Medicine, University of Malawi, Blantyre, Malawi
[e] Parasites and Microbes, Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK
[f] Darwin College, University of Cambridge, Silver Street, Cambridge, UK
[g] Biomedical Research Centre Network for Respiratory Diseases (CIBERES), San Sebastián, Spain
[h] NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool, UK

ARTICLE INFO

*Keywords:*
G12 rotavirus
Acute gastroenteritis
Rotavirus phylodynamics
Wa-like genetic backbone
Whole genome sequencing

ABSTRACT

G12 rotaviruses were first detected in Spain (Gipuzkoa province) in December 2004. After four years with no detections, G12 strains re-emerged in the 2010–2011 epidemic season, when the first European epidemic circulation of this genotype was observed in Gipuzkoa. G12 rotaviruses were also the dominant strains in 2011–2012, 2014–2015 and 2015–2016 epidemic seasons and were sporadically detected in the remaining periods (2012–2014 and 2016–2018). The most frequently detected G-type between 2010 and 2018 was G12 (29.9%) rather than G1 rotavirus (17.8%), which historically had been the dominant genotype in our setting (1989–2009 period) and globally. Phylogenetic analysis of the VP4 and VP7 genome segments showed chronologically ordered clades, which spanned between two to four consecutive seasons. Overall, the circulating G12 rotavirus strains in Gipuzkoa between 2010 and 2018 belonged to four clades, which emerged in early 2009 potentially due to at least four importations from other regions followed by local evolution. Whole genome analysis of 16 G12 strains detected from 2010 to 2018 revealed a Wa-like genotype constellation, G12-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, and also showed that G12 strains from Gipuzkoa were similar to those identified in other countries. These findings suggest circulation of G12 rotavirus strains in different parts of the world leading to high genetic diversity.

## 1. Introduction

Group A rotaviruses (RVAs) are a leading cause of acute gastroenteritis (AGE) in children < 5 years old worldwide. In 2009, World Health Organisation (WHO) recommended the universal vaccination of infants against RVAs (WHO, 2013) and the number of rotavirus deaths nearly halved to 215,000 by 2013 (Tate et al., 2016). The RVA genome consists of 11 double-stranded RNA (dsRNA) segments that encode six viral structural proteins (VP) and six non-structural proteins (NSP). A whole-genome classification system was proposed in which a specific genotype is assigned to each of the 11 genome segments. The RVA VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5 encoding genome segments are described using the abbreviation Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx (where "x" represents the number of the genotype), respectively. The six major G- and P-type combinations of human RVAs circulating worldwide (G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8]), are in their majority assigned to two genotype constellations, Wa-like (G1/3/4/9/12-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1) or DS-1-like (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) (Matthijnssens et al., 2008).

G12 was an unusual genotype until 2008, but subsequently G12 rotaviruses have been detected in significant proportions worldwide. The first case of G12 rotavirus (P[4] P-type) infecting humans was detected in the Philippines in 1987 (Taniguchi et al., 1990). Later, in 1998, it was detected (P[9] P-type) in Thailand (Pongsuwanna et al., 2002) and the following year in the United States (USA) (Griffin et al., 2002) and Argentina (Castello et al., 2006) in combination with P[6] and P[9], respectively. During the early 2000s, high proportions of G12 strains were detected almost exclusively in countries of the Indian

* Corresponding author at: Servicio de Microbiología, Hospital Universitario Donostia, Paseo Dr. Beguiristain s/n, 20014, Donostia, San Sebastián, Spain.
*E-mail address:* ainara.aranasalaberria@osakidetza.eus (A. Arana).

https://doi.org/10.1016/j.meegid.2019.03.016
Received 21 November 2018; Received in revised form 31 January 2019; Accepted 20 March 2019
Available online 21 March 2019

subcontinent (Ahmed et al., 2010; Sharma et al., 2008; Sherchand et al., 2009). Thereafter, they emerged in the rest of the world (Rahman et al., 2007; Matthijnssens et al., 2010; Cilla et al., 2013; Kokkinos et al., 2013; Mijatovic-Rustempasic et al., 2014).

Molecular surveillance of circulating human RVAs has been conducted in Gipuzkoa (Basque Country, Spain) since 1996. In 2010, the first European epidemic circulation of G12 rotavirus was reported in Gipuzkoa (Cilla et al., 2013). The aim of this study is to describe the molecular epidemiology of G12 rotaviruses that circulated in Gipuzkoa from 2010 to 2018.

## 2. Material and methods

### 2.1. Study population

Circulation of G12 strains was observed in the context of a prospective, population-based study performed at Zumarraga and Donostia University Hospitals (Gipuzkoa, Basque Country, Spain) from July 2010 to June 2018. The Hospitals serve a population of 437,875 inhabitants of whom 22,012 are children < 5 years old (according to the 2011 census, Basque Institute of Statistics). The study included children < 5 years old, attending primary care (outpatients) or hospital (emergency department and inpatients), and from whom stool analysis was requested. An epidemic season was defined as the period time elapsed from July 1st of one year till June 30th of the following year. Repeated episodes occurring in the same semester were excluded. Rotavirus vaccines were not included in the Basque Health System Expanded Program on Immunization (EPI) schedule and were only dispensed in private practice. The estimated vaccination coverage in Gipuzkoa was < 15%.

### 2.2. RVA detection and G/P-typing

The presence of RVA was investigated by enzyme immunoassay (ELISA, ProsPecT™ Rotavirus, Oxoid Ltd., UK) from July 2010 to June 2013 and by a commercial multiplex real-time polymerase chain reaction (PCR) assay from July 2013 onwards (FTD® Viral Gastroenteritis, Fast-Track Diagnostics Ltd., Luxembourg or Allplex™ Gastrointestinal full Panel Assay, Seegene Inc., Korea). The nucleic acid of the rotavirus-positive samples was extracted using the NucliSENS EasyMAG platform (bioMérieux, France). G/P-typing was conducted using multiplex reverse transcription (RT)-PCR methods (European Rotavirus Detection and Typing Methods version 4 [http://ern.heartitdev3.co.uk/project-information/documents-and-methods/]). In each run of RT-PCR, the amplicons of at least one sample of every different genotype were sequenced for genotype confirmation (at least 30% of the PCR-genotyped samples). From July 2014, confirmatory sequencing was conducted on all the VP7 amplicons of strains initially typed as G12 since the G12F primer, intended for G12 strains detection, cross-reacts with the VP7 gene of equine-like G3 rotavirus strains (Arana et al., 2016). The amplicons were sequenced using the ABI3730xl DNA Analyzer (Applied Biosystems, USA). The NCBI BLAST search was performed for genotype assignment (Altschul et al., 1990).

### 2.3. Next generation sequencing

#### 2.3.1. Sample selection and sequencing

Next Generation Sequencing was performed on 16 G12 strains collected during the rotavirus epidemics of 2010–2018. The strains were selected on the basis of having sufficient material leftover and the presence of high viral load inferred from the presence of high intensity bands of VP7 and VP4 amplicons when visualized under ultraviolet light in agarose electrophoresis gel or a low cycle threshold value ($C_t$ < 21) in the real-time PCR performed for RVA detection. The dsRNA extraction, library preparation and sequencing were performed following previously described methods (Arana et al., 2016). Briefly, ScriptSeq v2 RNA-Seq Library Preparation or ScriptSeq Complete Gold (Epidemiology) kits (Epicentre, Chicago, IL, USA) were used to generate libraries. Libraries were sequenced using HiSeq 2500 or MiSeq Illumina platforms at the Centre for Genomic Research of the University of Liverpool (United Kingdom) or at Life Sequencing (Valencia, Spain), respectively. Sequences of whole genomes of 16 G12P[8] strains and two G1P[8] strains circulating in Gipuzkoa in 2010–2018 were deposited in GenBank [accession numbers: MH171306 – MH171481 and MK417774 – MK417795]. The whole genomes of two G1P[8] strains, collected in the same epidemic season when the first epidemic of G12 rotavirus occurred in Gipuzkoa, were included in the phylogenetic analysis in order to assess if a reassortment event could have occurred locally among these Wa-like strains and the G12P[8] strains circulating in 2010–2011. The partial sequences of 120 VP4 (610 bp) and 92 VP7 (795 bp) genome segments of G12P[8] strains collected in Gipuzkoa in 2010–2018 were also deposited in GenBank [accession numbers: MK417562 – MK417773].

#### 2.3.2. Statistical and genomic data analyses

The paired-end reads were assembled using Geneious R8 (Kearse et al., 2012). An automated genotyping tool for RVAs, RotaC$^{2.0}$, was used to assign genotypes to the remaining nine genome segments (Maes et al., 2009). The pairwise identities of G12P[8] strains identified in Gipuzkoa in 2010–2018 with most similar strains in GenBank as well as some reference strains were calculated using Geneious R8 (Kearse et al., 2012). The sequences of the partial VP4 (P genotype) and VP7 (G genotype) coding genes were aligned together with full reference sequences for each gene using MUSCLE v3.8.31 (Edgar, 2004), and MEGA v6 (Tamura et al., 2013) was used to convert the sequence alignments from nucleotide to amino acid. Variable nucleotide and amino acid alignment positions were detected using Snp-Sites v2.3.1 (Page et al., 2016). Phylogenetic analyses were conducted using RAxML v8.2.8 (Stamatakis, 2006) and MEGA, using the Maximum Likelihood (ML) method with 1000 bootstrap replications for branch support. The best phylogenetic substitution model was obtained by ModelTest from MEGA 6 (Tamura et al., 2013) while the generalised time reversible (GTR) model (Tavaré, 1986) was used for RAxML. Before doing temporal genomic evolution analysis in BEAST v1.7.5, the molecular clock signal was assessed in phylogenetic trees of VP4 and VP7 encoding genes using TempEST v1.5.1 (Rambaut et al., 2016). The presence of the molecular clock-like signal was inferred when there was a significant association for the linear regression coefficient of the root-to-tip distance and time of sampling. The input data for BEAST was generated using BEAUTi v1.7.5. We used different molecular clock models (strict, lognormal relaxed, exponential, random and fixed) and a random starting tree as a prior. The population dynamics were inferred using the Gaussian Markov Random Field (GMRF) Bayesian Skyride model (Minin et al., 2008). The evolution parameters were summarised in Tracer v1.6 (Rambaut et al., 2018) while maximum clade credibility (MCC) time-dated phylogenies were generated with TreeAnnotator v1.7.5 (http://tree.bio.ed.ac.uk/software/). Other heuristic methods for estimating nucleotide substitution (mutation) rates were also used namely LSD v0.3beta (To et al., 2016), TreeDater v0.2.0 (Volz and Frost, 2017), TreeTime (Sagulenko et al., 2018) and TempEST v1.5.1 (Rambaut et al., 2016). The strains were clustered into clades using K-Pax2 unsupervised clustering approach (Pessia et al., 2015), which has been used to identify clusters in viral genomic data. All statistical analyses were performed by GraphPad Instat software, v3.05 (GraphPad Software, California, USA) and GGplot package (Wickham, 2009) in R v3.4.1 (R Core Team, 2014).

## 3. Results and discussion

### 3.1. Circulation of G12 rotavirus (2010–2018)

From July 2010 to June 2018, the presence of RVA was investigated

**Table 1**
Characteristics of the population included in the study (2508 RVA-positive episodes).

| Population characteristics | | N (%) |
|---|---|---|
| Sex | Male | 1408 (56.1) |
| | Female | 1100 (43.9) |
| Age (months) | < 12 | 1061 (42.3) |
| | 12 – < 24 | 1024 (40.8) |
| | 24 – < 36 | 279 (11.1) |
| | 36 – < 48 | 94 (3.8) |
| | 48 – < 60 | 50 (2.0) |

**Table 2**
Comparison of epidemiological variables of G1 and G12 rotaviruses circulating in Gipuzkoa (Basque Country, Spain), July 2010 – June 2018.

| | G1 | G12 | *p-value*[1] |
|---|---|---|---|
| Analysed episodes | 404 | 678 | – |
| Sex-male (n, %) | 241 (59.6%) | 364 (53.7%) | 0.065[2] |
| Age in months (mean ± SD) | 15.2 ± 10.0 | 14.4 ± 9.1 | 0.258[3] |
| Hospitalization rate (n, %) | 51 (12.6%) | 93 (13.7%) | 0.675[2] |

[1] A *p-value* of < 0.05 was considered statistically significant.
[2] Chi-square test.
[3] Mann-Whitney test.

in 21,135 stool samples collected from children < 5 years old with AGE, and rotavirus was detected in 2577 (12.2%) samples, corresponding to 2508 episodes. Population characteristics are shown in Table 1. A total of 2303 (91.8%) and 2270 (90.5%) rotavirus-positive episodes were P- typed or G-typed, respectively. In the study period (2010–2018), the most frequently detected P-types were P[8] (91.4%) and P[4] (8.5%), while the most frequently detected G-type was G12 (29.9%) followed by G9 (23.3%), G1 (17.8%), G3 (16.8%), G2 (7.5%) and G4 (2.8%). G12 rotavirus was first detected in Spain in December 2004 (three strains) in Gipuzkoa (Cilla et al., 2013). In 2010–2011 season, following four years in which G12 strains were not identified, epidemic circulation of G12 rotavirus was observed for the first time in this region. G12 rotavirus was the dominant genotype in 2010–2011 and also in 2011–2012 (Fig. 1). In the following two seasons, 2012–2013 and 2013–2014, G12 rotavirus was detected sporadically. However, from 2014 to 2015 and 2015–2016, G12 rotaviruses re-emerged and became a co-dominant genotype along with G1 and G9 strains, respectively. Subsequently, from 2016 to 2017 and 2017–2018, G12 strains were found in 2.6% and 6.2% of the characterised rotavirus-positive samples, respectively. Excluding mixed rotavirus infections, a total of 13.7% (93/678) of patients infected with G12 rotavirus were hospitalised due to community-acquired infections. Epidemiological variables of G12 rotavirus strains were compared to those of G1 and no significant differences were observed regarding gender, age or hospitalization rates (Table 2).

G1 rotavirus has been the predominant RVA genotype globally and previously in Gipuzkoa (Cilla et al., 2010) but was less frequently detected in Gipuzkoa from 2010 onwards in conjunction with an increased circulation of emerging genotypes, especially G9 and G12. The replacement of G1 by G12 strains has been demonstrated in Ghana following the introduction of rotavirus vaccination in 2012 (Lartey et al., 2018), which is consistent with findings obtained in the Australian states and territories using RotaTeq® vaccine (Roczo-Farkas et al., 2018). The 10th Annual report of EuroRotaNet (European Rotavirus surveillance Network) showed that G1 rotavirus was the most prevalent genotype each epidemic season between 2007 and 2014 (Hungerford and Iturriza-Gómara, 2016), but its prevalence decreased to 13% overall among the 14 EuroRotaNet countries in 2014–2015, and this decrease was seen both in countries with or without rotavirus vaccination. Furthermore, G1 rotavirus circulated widely again in Gipuzkoa in 2013–2014 and 2014–2015 (Fig. 1). Therefore, the declining trend of G1 should be interpreted with caution, as it could be the result of natural fluctuation of genotypes.

### 3.2. Phylogenetic and phylodynamic analyses of G12P[8] rotavirus

#### 3.2.1. VP4 and VP7 genome segments

The phylogenetic analyses of the partial nucleotide sequences for the VP4 and VP7 genome segments, 610 bp and 795 bp, respectively revealed chronological clustering of the G12P[8] strains from Gipuzkoa (Fig. 2). The clusters were inferred using an unsupervised genomic sequence clustering algorithm, which detected seven clades (clusters) for VP4 and four clades for VP7 genome segments. However, for the VP4 segment most strains clustered in five different clades which included strains collected in 2010–2011 (clade A), 2011–2013 (clade B), 2013–2014 (clade D), 2014–2015 (clade F) and 2014–2018 (clade G) periods, circulating strains of the latter clade during four consecutive epidemic seasons (Figs. 2a and 4a). Interestingly, the VP4 segment of strains collected in 2014–2015 season split into two clades, F and G. In Europe, as in our setting, G12 strains have been more frequently associated with P[8] than with P[6] and even less commonly with P[4] and P[9] (Iturriza-Gomara et al., 2011). Among G12 strains from Gipuzkoa, two were found to be associated with P[4] (Cilla et al., 2013) and six did not cluster with other P[8] collected in the same season (clade C and E). The six divergent P[8] sequences were similar or

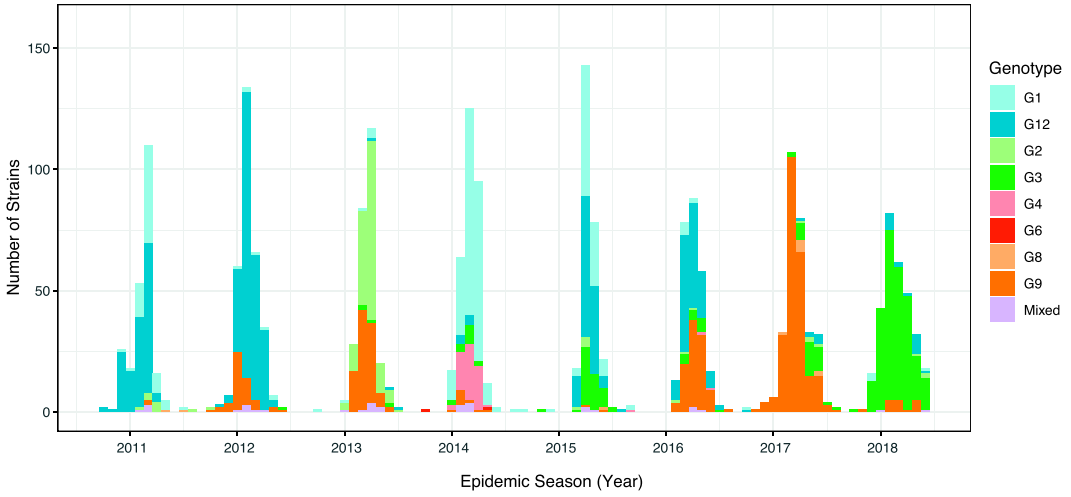


**Fig. 1.** Circulation and distribution of G12 and other RVA G-types in children < 5 years-old in Gipuzkoa (Basque Country, Spain), July 2010 – June 2018.

identical to local P[8] strains associated with different G-types (Supplementary Fig. 1), similar to previous observations in Central Italy in 2012–2013 season (Delogu et al., 2015). Thus, these strains may have acquired the VP4 genome segment from other co-circulating strains through genetic reassortment. All VP4 and VP7 genome segments of G12P[8] strains from Gipuzkoa belonged to lineage III, which is the most common lineage worldwide (da Silva et al., 2017; Delogu et al., 2015; Komoto et al., 2014; Langa et al., 2016; Matthijnssens et al., 2010; Mijatovic-Rustempasic et al., 2014; Ndze et al., 2014; Wangchuk et al., 2014).

### 3.2.2. Emergence of G12 strains in Gipuzkoa

G12P[8] RVA circulated in Gipuzkoa in all the epidemic seasons from 2010 to 2018. However, it was not possible to know to what extent the sporadic introductions of additional strains of other regions occurred and how the local persistent strains evolved to contribute to the circulation and diversity of G12 strains during the eight consecutive epidemic seasons. Evolutionary analysis of VP7 gene confirmed the emergence of four different clades of G12 rotavirus strains which diverged in 2010–2011 (clade A), 2011–2012 (clade B) and 2012–2013 (clades C and D) epidemic seasons in Gipuzkoa (Fig. 3). These results are supported by the relative low nucleotide sequence identities observed among strains within the different clades when the 11 genes were compared (Supplementary Tables 1–4). Strains that caused the epidemic of 2011–2012 (clade B) and strains detected in 2014–2015 (clade D) spanned over two and four consecutive epidemic seasons, respectively (Fig. 4b). This implied that the long persistence of this genotype in Gipuzkoa was due to the successive turnover of different clades of G12 strains. Typically, the clades would be replaced with new ones, except for clade C for VP4, first observed in 2013 that re-emerged in 2017, and clade B for VP7 that was observed between 2011 and 2013 and re-emerged in 2016–2017 season. These strains may have persisted since then undetectable in our setting. A study that analysed the complete genomes of 58 Wa-like strains (36 G1P[8], 18 G3P[8], and 4 G12P[8]) (McDonald et al., 2012), showed that some allele constellations persisted over 1 to 3 years. Although in our study detection of RVA was carried out uninterruptedly throughout the study period, G12 rotavirus was not detected in the long interepidemic periods (> 5 months, > 900 faecal samples analysed per interepidemic period) (Fig. 1). However, this study, like most rotavirus surveillance studies, captured only medically attended symptomatic disease and it is possible that strains are maintained in a population through subclinical or mild infections not requiring or seeking diagnostic testing. Studies of RVA detection in sewage would provide a useful approach in order to fully understand the true circulation of any rotavirus strain across the entire population and throughout the year, and to understand in-depth rotavirus population dynamics.

### 3.2.3. Evolution of VP4 and VP7 genome segments

In order to investigate the evolutionary dynamics of these G12 strains, a Bayesian maximum clade credibility time-dated tree based on the partial VP7 nucleotide sequences was constructed (Fig. 3). There was insufficient molecular clock signal in VP4 gene possibly reflecting a history of reassortments which may have distorted the temporal signal therefore it was excluded from coalescent analysis in BEAST. The time-dated phylogenies inferred using different clock models were similar with effective sample sizes (ESS) > 1000 for inferred parameters including mutation rates and time to the most recent common ancestor (tMRCA). Using the time-dated phylogeny inferred by the exponential clock model, the tMRCA for the G12 strains from Gipuzkoa was February 2009 (95% highest posterior density [HPD] interval December 2007 – December 2009) (Figs. 3 and 4c). This tMRCA is plausible

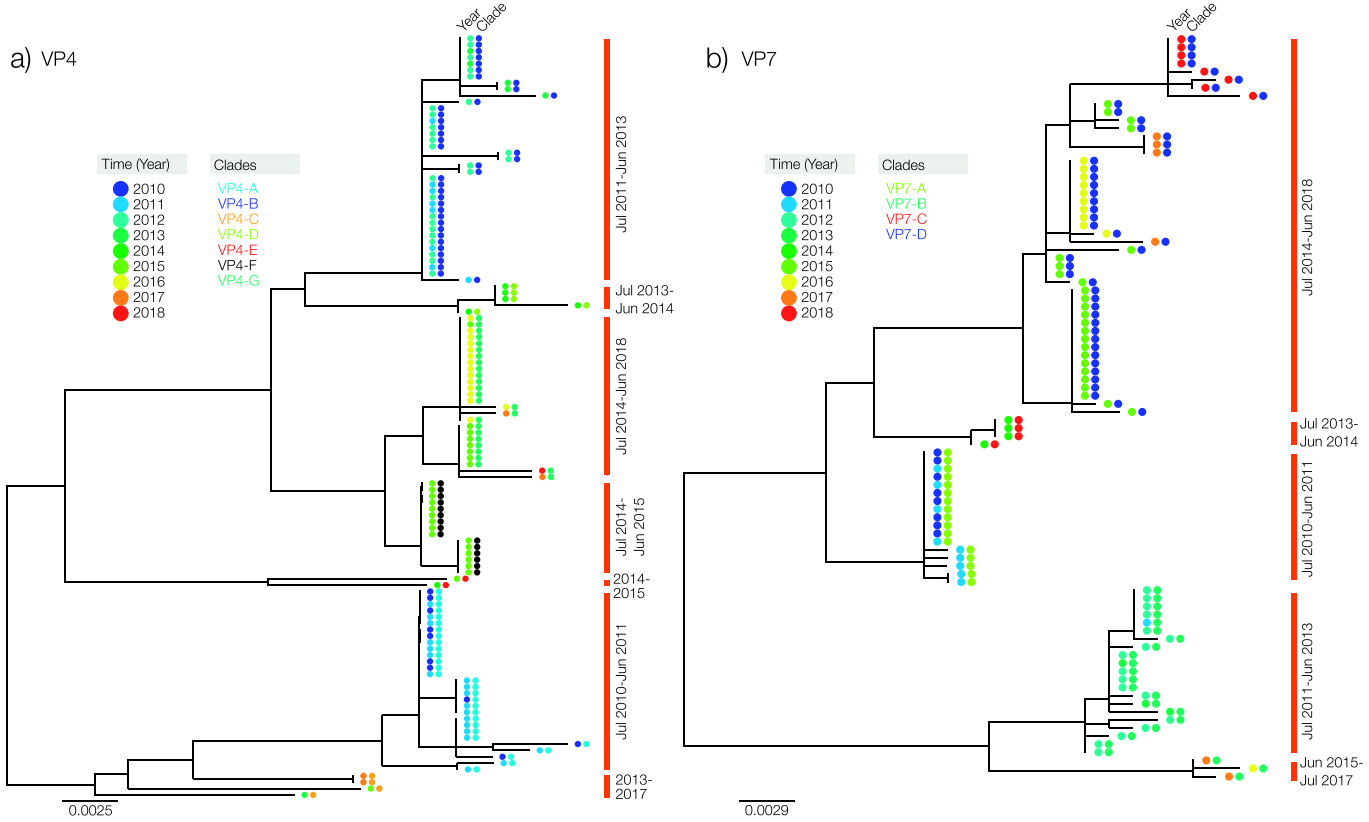


**Fig. 2.** Maximum likelihood phylogeny of rotavirus genomic segments for G12P[8] strains from Gipuzkoa (Spain). a) Phylogenetic tree constructed using partial VP4 genome segment (610 bp) and b) partial VP7 genome segment (795 bp). The taxon tips are coloured by year of collection and clades defined by K-Pax2 (Pessia et al., 2015). The trees were rooted at mid-point of the branch separating the most divergent strains in the phylogeny.

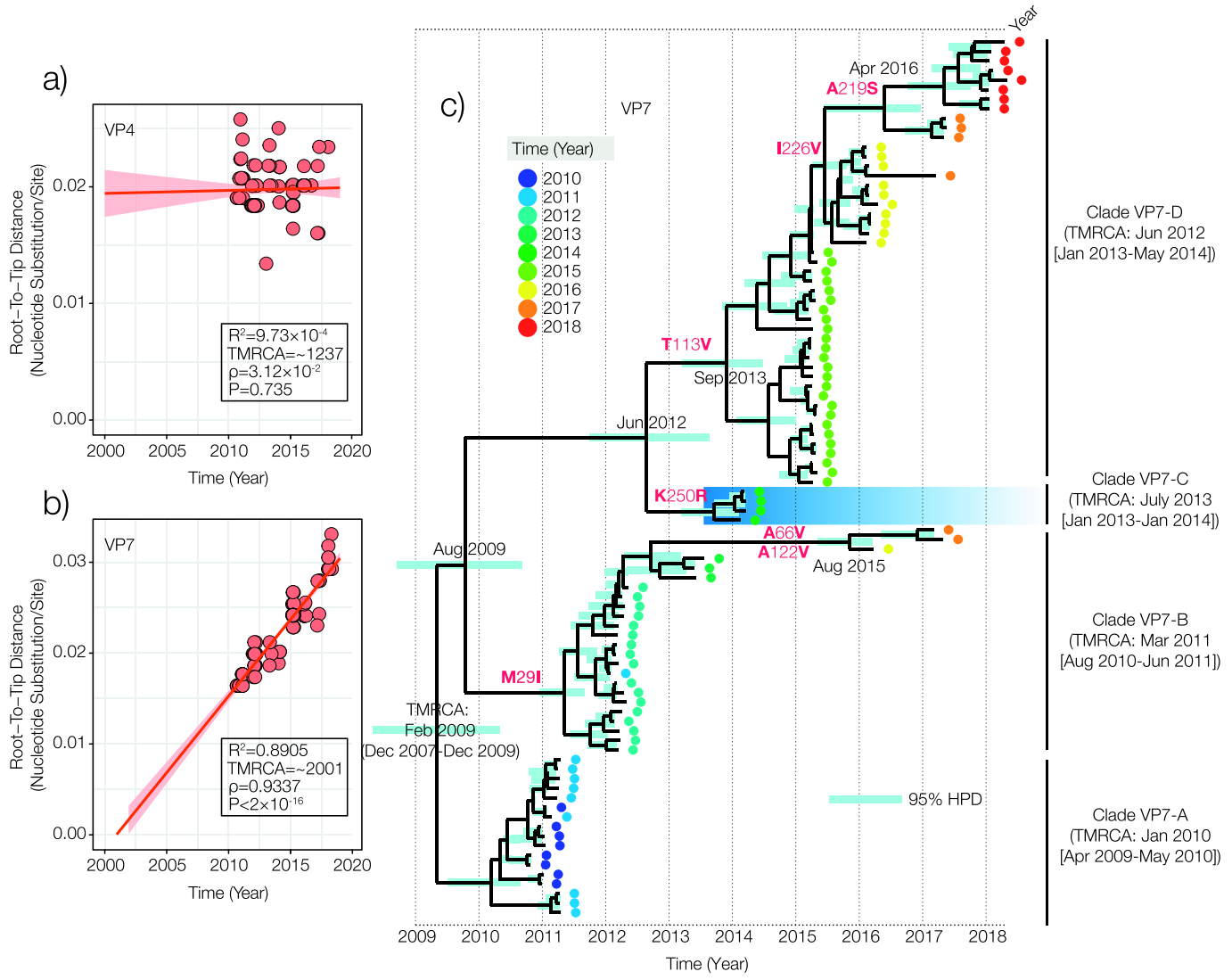


**Fig. 3.** Bayesian time-dated maximum clade credibility phylogeny of G12 VP7 rotavirus segment from Gipuzkoa (Spain). Calibration of the molecular clock signal by plotting phylogenetic root-to-tip distance versus time of collection using tempEST for a) VP4 and b) VP7. Maximum likelihood trees in Fig. 2 were used but were re-rooted to identify a branch which maximised the temporal association. There was insufficient molecular clock signal for VP4 therefore only VP7 was used for the BEAST analysis. c) Time-dated phylogeny inferred by BEAST showing divergence times of the strains and their most recent common ancestors (MRCA). The clades defined by K-Pax2 in Fig. 2 are labelled on the phylogeny including divergence times of their MRCAs. Divergence times of different clades are also labelled as well as other branches. Each internal node of the phylogeny contains a horizontal rectangular strip showing 95% highest posterior density [HPD] (Bayesian equivalent for confidence interval) for the estimates of its ages. Each clade in the tree is coloured differently and adjacent to the taxon tips are circles showing dates of collection. Major amino acid changes associated with formation of different clades are labelled on the phylogeny in red text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

because it came before the first detection of G12 strains in Gipuzkoa in September 2010 and its subsequent rapid expansion in 2010–2011 epidemic season. As illustrated in the ML tree (Fig. 2b), the VP7 gene segregated into four aforementioned clades and their tMRCAs are shown in Fig. 3. Clades A, B, C and D included G12 strains from Gipuzkoa collected in 2010–2011, 2011–2013, 2013–2014 and 2014–2018, respectively.

In total, 20 and 24 amino acid substitutions were identified among the VP4 and VP7 genes of the G12P[8] strains from Gipuzkoa, respectively. These amino acid substitutions exhibited different phylogenetic and temporal patterns for example M29I, K250R and T113 V substitutions in VP7 attained fixation over time and defined different clades namely B, C and D, respectively. The increased frequency and dominance of the substitutions suggests that these mutations offered an advantage to the G12 strains (Figs. 3 and 5). Regarding VP8* antigenic epitope (Zeller et al., 2012), amino acid differences at positions D116E, S146G or G195D were observed in some strains (Supplementary Fig. 2). In contrast, only one amino acid substitution, S147 L in VP7 gene, was located at mapped antigenic regions (Supplementary Fig. 3). This mutation was identified in 12.5% of the strains circulating in 2011 but it was not detected in the following seasons potentially reflecting that it was not highly favoured by selection. Other mutations, located outside the mapped antigenic regions, showed different patterns including random patterns such as P71L and S95R in VP4 and I17V and Y67H in VP7 while others showed a sharp increase to become the predominant mutation in a single epidemic season e.g. A66V and K250R in VP7 but are immediately purged from the population within a single epidemic season, which may suggest they are not favoured by selection. The mean evolutionary rate for the VP7 gene was $\sim 2.5 \times 10^{-3}$ nucleotide substitutions per site per year (Fig. 4d). The Bayesian Skyride Plot analysis showed that the effective population size of G12P[8] strains from Gipuzkoa remained stable during the study period but started to decrease in 2017 (Fig. 4e), which was consistent with the

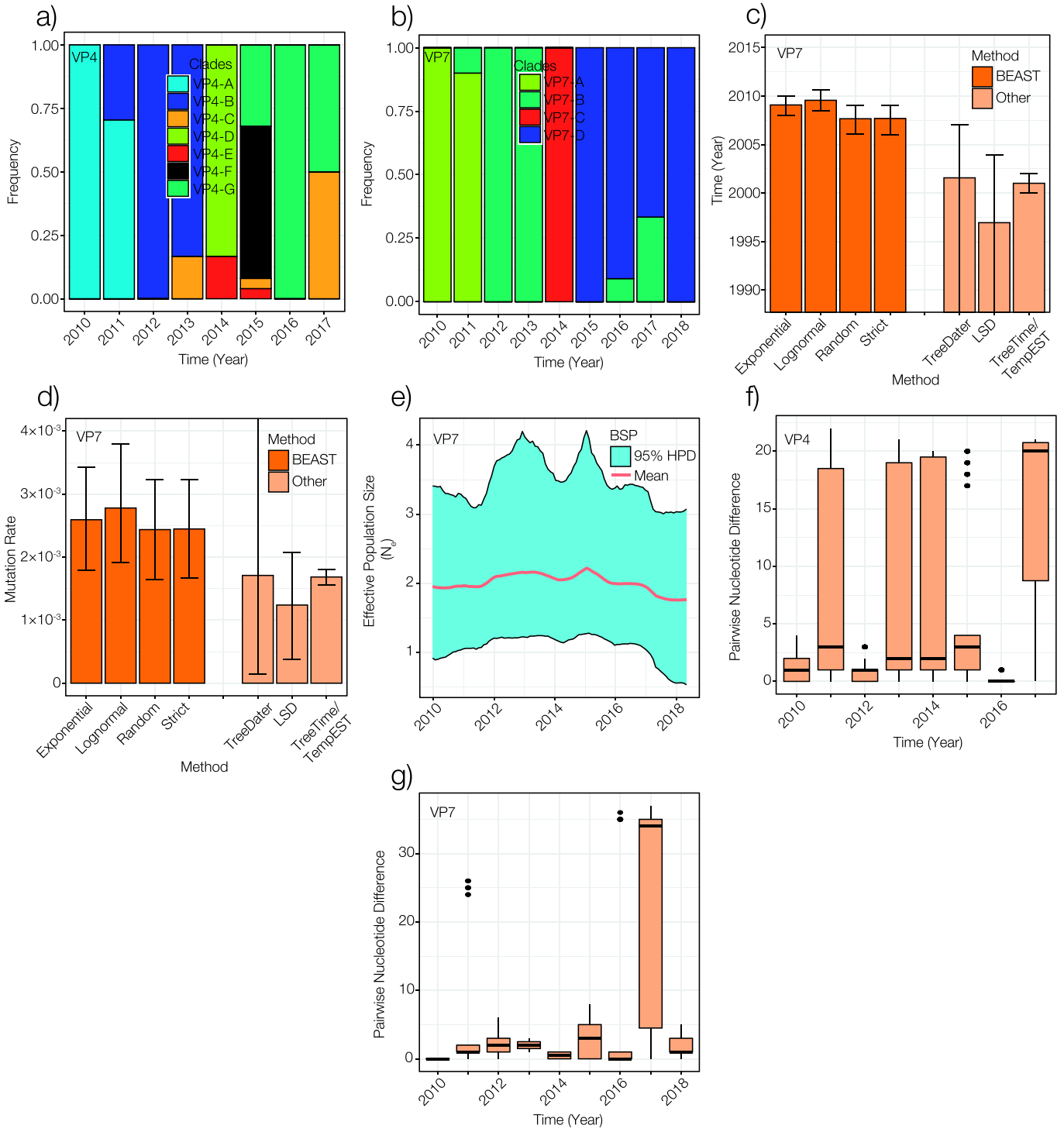


**Fig. 4.** Evolutionary dynamics of G12P[8] rotavirus segments from Gipuzkoa (Spain). a) Temporal changes in frequency of G12P[8] clades for VP4 segment. 2018 was not included in the figure since only one VP4 sequence was available in this period. b) Temporal changes in frequency of G12P[8] clades for VP7 segment. c) Times to the most recent common ancestor (tMRCA) inferred by different clock models in BEAST and heuristic methods for VP7 segment. The effective sample sizes are labelled on each bar in the plots. d) Nucleotide substitution (mutation) rate for VP7 segment estimated using different clock models in BEAST in comparison with estimates obtained using quicker but heuristic methods. e) Gaussian Markov Random Field (GMRF) Bayesian Skyride Plot showing changes in effective population size ($N_e$) for VP7 segment. Distribution of the number of single nucleotide polymorphisms (SNP) identified by pairwise comparisons of strains sampled in the same year for f) VP4 and g) VP7 segments.

reduction in frequency of G12 from the 2017 epidemic season shown in Fig. 1. Unlike for VP4, which showed higher sequence diversity than VP7, the stability of the population size of the latter is further supported by the less variability in the sequence diversity of the genome segment (Fig. 4f and g).

### 3.2.4. *Whole genome analysis of G12 strains*

Whole genome sequencing revealed that the 16 G12 strains sampled between 2010 and 2018 from Gipuzkoa had a Wa-like genetic backbone: G12-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, which is in line with findings from other studies (Freeman et al., 2009; Matthijnssens and

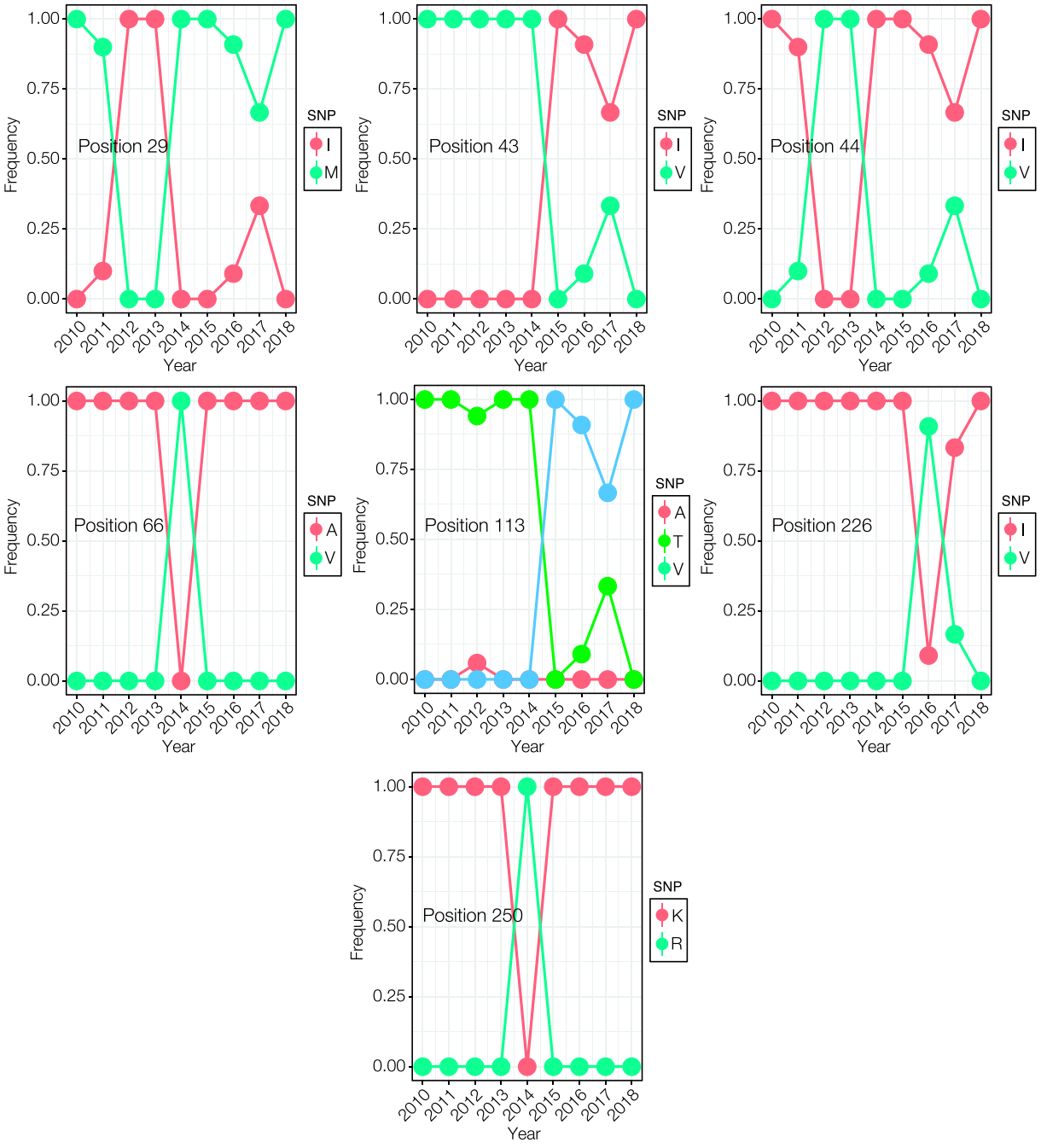


**Fig. 5.** Temporal dynamics of amino acid changes in rotavirus G12P[8] VP7 segment in Gipuzkoa (Spain). Frequencies of amino acids which changed significantly during the sampling period are shown. Annual frequency of each mutation is shown by different colours in the key. Specific positions in VP7 where the mutational changes occurred are shown inside each plot. All mutational changes for VP4 and VP7 are shown in Supplementary Figs. 2 and 3, respectively.

Van Ranst, 2012; Rahman et al., 2007). Previous studies suggested that G12P[8] strains may have originated by the introduction of a G12 VP7 genome segment within the genetic backbone of a globally common Wa-like strain (Freeman et al., 2009). The temporal clustering in the VP7 and VP4 phylogenetic trees was maintained when the whole genome was analysed (Supplementary Fig. 4). Based on the concatenated coding regions of the 11 genome segments, the G12P[8] rotavirus strains of this study showed a nucleotide sequence identity ≥96.3%, which was ≥98.9% among strains belonging to the same clade (data not shown). De Grazia and colleagues (De Grazia et al., 2016) conducted a phylogenetic analysis of the complete genome of 22 G12P[8] strains collected from 2012 to 2014 in Sicily and found that all strains, except two, clustered into three major subgroups which indicate repeated introductions of G12 RVA strains in Sicily in a short period of time. Other studies (Stupka et al., 2012; Ide et al., 2015) found that only one clone of G12P[8] strains circulated in their settings (Argentina from 2008 to 2009 and Myanmar in 2011), although limited number of strains were analysed. These data suggested that in this early-stage of the expansion of G12, the circulation is mainly clonal in defined areas even though different genetic clones may co-exist during the same epidemic season.

The phylogenetic analysis of the complete genome of G12P[8] strains showed that the strains from Gipuzkoa exhibited a high nucleotide identity (> 98%) with Wa-like G12 strains collected worldwide. In particular, the first strains detected in Gipuzkoa (2010–2011 epidemic) clustered with CU616-TK strain collected in Thailand in 2009

(Theamboonlers et al., 2014). Only the strains detected in 2013–2014 epidemic season showed less identity with other international strains, sharing the highest nucleotide identity with a South African strain (MRC-DPRU2140) collected in 2005 (98.0%). Despite belonging to the same cluster, the VP6 genome segment of strains collected in 2014–2016 showed low nucleotide identities (95.6–95.7%) with those of strains detected in 2017–2018, which indicates that a reassortment event may have occurred (Supplementary Tables 1–4). The origin of the four main clades responsible for the G12 RVA circulation in Gipuzkoa remains unclear, since VP7 and VP4 genome segments as well as the remaining nine genes showed a close relationship with those of other strains collected in different parts of the world in different years. Wa-like is the most prevalent and stable genotypic constellation among human RVA strains, which indicate it is well adapted to the human host. Similarly to what happened with G9P[8] rotavirus from the mid-1990s onwards, this genetic backbone may have conferred an evolutionary advantage upon G12 strains, leading to a remarkably fast global distribution (Ghosh and Kobayashi, 2011; Matthijnssens et al., 2008; Matthijnssens and Van Ranst, 2012). This together with the high genetic diversity of G12 strains makes it difficult to define an origin for the different clones of G12P[8] rotavirus circulating in Gipuzkoa.

## 4. Conclusion

G12 rotavirus strains emerged in Gipuzkoa (Basque Country, Spain) in 2010 and became the dominant genotype in some years, partially replacing G1 rotavirus, which historically had been the dominant genotype in this territory (1989–2009) and globally. Circulation of G12 rotavirus in Gipuzkoa was driven by the introduction of strains from other regions and their limited local persistence and evolution. Moreover, our findings suggest circulation of G12 rotavirus strains in different parts of the world leading to high genetic diversity favoured by the increasingly frequent international movements of population that facilitate the contact among distinct rotavirus strains and thus the generation of new RVA strains by reassortment.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.03.016.

## Funding

This work was partly funded by a grant from the Basque Country Government to the University Research Groups [grant number: IT656–13] and from the EuroRotaNet project. A. Arana was supported by a grant for the training of predoctoral researchers, Basque Government, Vitoria-Gasteiz, Spain [grant number: PRE_2014_2_83]. M. Iturriza-Gómara received support from the National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections at the University of Liverpool [grant number: NIHR HPRU 2012–10038] and the Wellcome Trust. K.C. Jere received funding from the Wellcome Trust [grant number: 201945/Z/16/Z].

## Author contributions

A. Arana, C. Chaguza, M. Iturriza-Gómara and G. Cilla designed the study and drafted the manuscript. M. Alkorta carried out rotavirus detection, analysis and sample collection. M. Montes and A. Arana performed the G-/P-genotyping. K.C. Jere and A. Arana generated the libraries for Next Generation Sequencing. C. Chaguza, A. Arana and K.C. Jere conducted sequence data analysis. All authors contributed to the interpretation of the data, writing of the manuscript, read and approved the final manuscript.

## Potential conflicts of interest

M. Iturriza-Gómara has received research grant support from GlaxoSmithKline Biologicals and Sanofi Pasteur MSD and Merck for work on rotavirus. K.C. Jere has received grants support from GlaxoSmithkline Biologicals for work on rotavirus. The other authors declare that they have no conflict of interest.

## Disclaimer

The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

## References

Ahmed, K., Ahmed, S., Mitui, M.T., Rahman, A., Kabir, L., Hannan, A., Nishizono, A., Nakagomi, O., 2010. Molecular characterization of VP7 gene of human rotaviruses from Bangladesh. Virus Genes 40, 347–356.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arana, A., Montes, M., Jere, K.C., Alkorta, M., Iturriza-Gomara, M., Cilla, G., 2016. Emergence and spread of G3P[8] rotaviruses possessing an equine-like VP7 and a DS-1-like genetic backbone in the Basque Country (north of Spain), 2015. Infect. Genet. Evol. 44, 137–144.

Castello, A.A., Arguelles, M.H., Rota, R.P., Olthoff, A., Jiang, B., Glass, R.I., Gentsch, J.R., Glikmann, G., 2006. Molecular epidemiology of group a rotavirus diarrhea among children in Buenos Aires, Argentina, from 1999 to 2003 and emergence of the infrequent genotype G12. J. Clin. Microbiol. 44, 2046–2050.

Cilla, G., Montes, M., Gomariz, M., Pineiro, L., Perez-Trallero, E., 2010. Rotavirus genotypes in children in the Basque Country (northern Spain) over a 13-year period (July 1996-June 2009). Eur. J. Clin. Microbiol. Infect. Dis. 29, 955–960.

Cilla, G., Montes, M., Gomariz, M., Alkorta, M., Iturzaeta, A., Perez-Yarza, E.G., Perez-Trallero, E., 2013. Rotavirus genotypes in children in the Basque Country (north of Spain): rapid and intense emergence of the G12[P8] genotype. Epidemiol. Infect. 141, 868–874.

De Grazia, S., Doro, R., Bonura, F., Marton, S., Cascio, A., Martella, V., Banyai, K., Giammanco, M.G., 2016. Complete genome analysis of contemporary G12P[8] rotaviruses reveals heterogeneity within Wa-like genomic constellation. Infect. Genet. Evol. 44, 85–93.

Delogu, R., Ianiro, G., Camilloni, B., Fiore, L., Ruggeri, F.M., 2015. Unexpected spreading of G12P[8] rotavirus strains among young children in a small area of Central Italy. J. Med. Virol. 87, 1292–1302.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Freeman, M.M., Kerin, T., Hull, J., Teel, E., Esona, M., Parashar, U., Glass, R.I., Gentsch, J.R., 2009. Phylogenetic analysis of novel G12 rotaviruses in the United States: a molecular search for the origin of a new strain. J. Med. Virol. 81, 736–746.

Ghosh, S., Kobayashi, N., 2011. Whole-genomic analysis of rotavirus strains: current status and future prospects. Future Microbiol 6, 1049–1065.

Griffin, D.D., Nakagomi, T., Hoshino, Y., Nakagomi, O., Kirkwood, C.D., Parashar, U.D., Glass, R.I., Gentsch, J.R., National Rotavirus Surveillance System, 2002. Characterization of nontypeable rotavirus strains from the United States: identification of a new rotavirus reassortant (P2A[6],G12) and rare P3[9] strains related to bovine rotaviruses. Virology 294, 256–269.

Hungerford, D., Iturriza-Gómara, M., 2016. 2017. EuroRotaNet, Annual report. http://ern.heartitdev3.co.uk/.

Ide, T., Komoto, S., Higo-Moriguchi, K., Htun, K.W., Myint, Y.Y., Myat, T.W., Thant, K.Z., Thu, H.M., Win, M.M., Oo, H.N., Htut, T., Wakuda, M., Dennis, F.E., Haga, K., Fujii, Y., Katayama, K., Rahman, S., Nguyen, S.V., Umeda, K., Oguma, K., Tsuji, T., Taniguchi, K., 2015. Whole genomic analysis of human G12P[6] and G12P[8] rotavirus strains that have emerged in Myanmar. PLoS One 10, e0124965.

Iturriza-Gomara, M., Dallman, T., Banyai, K., Bottiger, B., Buesa, J., Diedrich, S., Fiore, L., Johansen, K., Koopmans, M., Korsun, N., Koukou, D., Kroneman, A., Laszlo, B., Lappalainen, M., Maunula, L., Marques, A.M., Matthijnssens, J., Midgley, S., Mladenova, Z., Nawaz, S., Poljsak-Prijatelj, M., Pothier, P., Ruggeri, F.M., Sanchez-Fauquier, A., Steyer, A., Sidaraviciute-Ivaskeviciene, I., Syriopoulou, V., Tran, A.N., Usonis, V., VAN Ranst, M., DE Rougemont, A., Gray, J., 2011. Rotavirus genotypes co-circulating in Europe between 2006 and 2009 as determined by EuroRotaNet, a pan-European collaborative strain surveillance network. Epidemiol. Infect. 139, 895–909.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kokkinos, P.A., Ziros, P.G., Monini, M., Lampropoulou, P., Karampini, A., Papachatzi, E., Mantagos, S., Ruggeri, F.M., Vantarakis, A., 2013. Rare types of rotaviruses isolated from children with acute gastroenteritis in Patras, Greece. Intervirology 56, 237–241.

Komoto, S., Wandera Apondi, E., Shah, M., Odoyo, E., Nyangao, J., Tomita, M., Wakuda, M., Maeno, Y., Shirato, H., Tsuji, T., Ichinose, Y., Taniguchi, K., 2014. Whole genomic analysis of human G12P[6] and G12P[8] rotavirus strains that have emerged in Kenya: identification of porcine-like NSP4 genes. Infect. Genet. Evol. 27, 277–293.

Langa, J.S., Thompson, R., Arnaldo, P., Resque, H.R., Rose, T., Enosse, S.M., Fialho, A., de Assis, R.M., da Silva, M.F., Leite, J.P., 2016. Epidemiology of rotavirus a diarrhea in Chokwe, southern Mozambique, from February to September, 2011. J. Med. Virol. 88, 1751–1758.

Lartey, B.L., Damanka, S., Dennis, F.E., Enweronu-Laryea, C.C., Addo-Yobo, E., Ansong, D., Kwarteng-Owusu, S., Sagoe, K.W., Mwenda, J.M., Diamenu, S.K., Narh, C., Binka, F., Parashar, U., Lopman, B., Armah, G.E., 2018. Rotavirus strain distribution in Ghana pre- and post- rotavirus vaccine introduction. Vaccine 36, 7238–7242.

Maes, P., Matthijnssens, J., Rahman, M., Van Ranst, M., 2009. RotaC: a web-based tool for the complete genome classification of group a rotaviruses. BMC Microbiol. 9, 238-2180-9-238.

Matthijnssens, J., Van Ranst, M., 2012. Genotype constellation and evolution of group a rotaviruses infecting humans. Curr. Opin. Virol. 2, 426–433.

Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S.M., Palombo, E.A., Iturriza-Gomara, M., Maes, P., Patton, J.T., Rahman, M., Van Ranst, M., 2008. Full genome-based classification of rotaviruses reveals a common origin between human Wa-like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. J. Virol. 82, 3204–3219.

Matthijnssens, J., Heylen, E., Zeller, M., Rahman, M., Lemey, P., Van Ranst, M., 2010. Phylodynamic analyses of rotavirus genotypes G9 and G12 underscore their potential for swift global spread. Mol. Biol. Evol. 27, 2431–2436.

McDonald, S.M., McKell, A.O., Rippinger, C.M., McAllen, J.K., Akopov, A., Kirkness, E.F., Payne, D.C., Edwards, K.M., Chappell, J.D., Patton, J.T., 2012. Diversity and relationships of cocirculating modern human rotaviruses revealed using large-scale comparative genomics. J. Virol. 86, 9148–9162.

Mijatovic-Rustempasic, S., Teel, E.N., Kerin, T.K., Hull, J.J., Roy, S., Weinberg, G.A., Payne, D.C., Parashar, U.D., Gentsch, J.R., Bowen, M.D., 2014. Genetic analysis of G12P[8] rotaviruses detected in the largest U.S. G12 genotype outbreak on record. Infect. Genet. Evol. 21, 214–219.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471.

Ndze, V.N., Esona, M.D., Achidi, E.A., Gonsu, K.H., Doro, R., Marton, S., Farkas, S., Ngeng, M.B., Ngu, A.F., Obama-Abena, M.T., Banyai, K., 2014. Full genome characterization of human rotavirus a strains isolated in Cameroon, 2010-2011: diverse combinations of the G and P genes and lack of reassortment of the backbone genes. Infect. Genet. Evol. 28, 537–560.

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom. 2, e000056.

Pessia, A., Grad, Y., Cobey, S., Puranen, J.S., Corander, J., 2015. K-Pax2: Bayesian identification of cluster-defining amino acid positions in large sequence datasets. Microb. Genom. 1, e000025.

Pongsuwanna, Y., Guntapong, R., Chiwakul, M., Tacharoenmuang, R., Onvimala, N., Wakuda, M., Kobayashi, N., Taniguchi, K., 2002. Detection of a human rotavirus with G12 and P[9] specificity in Thailand. J. Clin. Microbiol. 40, 1390–1394.

Rahman, M., Matthijnssens, J., Yang, X., Delbeke, T., Arijs, I., Taniguchi, K., Iturriza-Gomara, M., Iftekharuddin, N., Azim, T., Van Ranst, M., 2007. Evolutionary history and global spread of the emerging g12 human rotaviruses. J. Virol. 81, 2382–2390.

Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). Virus Evol. 2, vew007.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian Phylogenetics using tracer 1.7. Syst. Biol. 67, 901–904.

Roczo-Farkas, S., Kirkwood, C.D., Cowley, D., Barnes, G.L., Bishop, R.F., Bogdanovic-Sakran, N., Boniface, K., Donato, C.M., Bines, J.E., 2018. The impact of rotavirus vaccines on genotype diversity: a comprehensive analysis of 2 decades of Australian surveillance data. J. Infect. Dis. 218, 546–554.

Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 4, vex042.

Sharma, S., Ray, P., Gentsch, J.R., Glass, R.I., Kalra, V., Bhan, M.K., 2008. Emergence of G12 rotavirus strains in Delhi, India, in 2000 to 2007. J. Clin. Microbiol. 46, 1343–1348.

Sherchand, J.B., Nakagomi, O., Dove, W., Nakagomi, T., Yokoo, M., Pandey, B.D., Cuevas, L.E., Hart, C.A., Cunliffe, N.A., 2009. Molecular Epidemiology of Rotavirus Diarrhea among Children Aged.

da Silva, M.F., Fumian, T.M., de Assis, R.M., Fialho, A.M., Carvalho-Costa, F.A., da Silva Ribeiro de Andrade,J., Leite, J.P., 2017. VP7 and VP8* genetic characterization of group a rotavirus genotype G12P[8]: emergence and spreading in the eastern Brazilian coast in 2014. J. Med. Virol. 89, 64–70.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stupka, J.A., Degiuseppe, J.I., Parra, G.I., Argentinean National Rotavirus Surveillance Network, 2012. Increased frequency of rotavirus G3P[8] and G12P[8] in Argentina during 2008-2009: whole-genome characterization of emerging G12P[8] strains. J. Clin. Virol. 54, 162–167.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Taniguchi, K., Urasawa, T., Kobayashi, N., Gorziglia, M., Urasawa, S., 1990. Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup I specificity and long RNA pattern: implication for new G serotype specificity. J. Virol. 64, 5640–5644.

Tate, J.E., Burton, A.H., Boschi-Pinto, C., Parashar, U.D., World Health Organization-Coordinated Global Rotavirus Surveillance Network, 2016. Global, regional, and National Estimates of rotavirus mortality in children < 5 years of age, 2000-2013. Clin. Infect. Dis. 62 (Suppl. 2), S96–S105.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on mathematics in the life sciences. American Mathematical Society 17, 57–86.

Theamboonlers, A., Maiklang, O., Thongmee, T., Chieochansin, T., Vuthitanachot, V., Poovorawan, Y., 2014. Complete genotype constellation of human rotavirus group a circulating in Thailand, 2008-2011. Infect. Genet. Evol. 21, 295–302.

To, T.H., Jung, M., Lycett, S., Gascuel, O., 2016. Fast dating using least-squares criteria and algorithms. Syst. Biol. 65, 82–97.

Volz, E.M., Frost, S.D.W., 2017. Scalable relaxed clock phylogenetic dating. Virus evolution 3, vex025-vex025.

Wangchuk, S., Mitui, M.T., Tshering, K., Yahiro, T., Bandhari, P., Zangmo, S., Dorji, T., Tshering, K., Matsumoto, T., Nishizono, A., Ahmed, K., 2014. Dominance of emerging G9 and G12 genotypes and polymorphism of VP7 and VP4 of rotaviruses from Bhutanese children with severe diarrhea prior to the introduction of vaccine. PLoS One 9, e110795.

WHO, 2013. Rotavirus vaccines. WHO position paper - January 2013. Wkly Epidemiol. Rec. 88, 49–64.

Wickham, H., 2009. Ggplot2.Elegant graphics for data analysis, first Ed. springer-Verlag New York.

Zeller, M., Patton, J.T., Heylen, E., De Coster, S., Ciarlet, M., Van Ranst, M., Matthijnssens, J., 2012. Genetic analyses reveal differences in the VP7 and VP4 antigenic epitopes between human rotaviruses circulating in Belgium and rotaviruses in Rotarix and RotaTeq. J. Clin. Microbiol. 50, 966–976.