The Journal of Infectious Diseases

MAJOR ARTICLE

# The Evolution and Transmission of Epidemic GII.17 Noroviruses

**Jing Lu,[1,2,3] Lin Fang,[1] Huanying Zheng,[1] Jiaqian Lao,[1] Fen Yang,[1] Limei Sun,[1] Jianpeng Xiao,[1,2] Jinyan Lin,[1] Tie Song,[1] Tao Ni,[4] Jayna Raghwani,[3] Changwen Ke,[1] Nuno R. Faria,[3] Thomas A. Bowden,[4] Oliver G. Pybus,[3,a] and Hui Li[1,a]**

[1]Guangdong Provincial Center for Disease Control and Prevention, and [2]Guangdong Provincial Institution of Public Health, Guangzhou, China; [3]Department of Zoology, and [4]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom

***Background.*** In recent decades, the GII.4 norovirus genotype has predominated in epidemics worldwide and been associated with an increased rate of evolutionary change. In 2014, a novel GII.17 variant emerged and persisted, causing large outbreaks of gastroenteritis in China and sporadic infections globally. The origin, evolution, and transmission history of this new variant are largely unknown.

***Methods.*** We generated 103 full capsid and 8 whole-genome sequences of GII.17 strains collected between August 2013 and November 2015 in Guangdong, China. Phylogenetic analyses were performed by integrating our data with those for all publically available GII.17 sequences.

***Results.*** The novel emergent lineage GII.17_Kawasaki_2014 most likely originated from Africa around 2001 and evolved at a rate of $5.6 \times 10^{-3}$ substitutions/site/year. Within this lineage, a new variant containing several important amino acid changes emerged around August 2013 and caused extensive epidemics in 2014–2015. The phylodynamic and epidemic history of the GII.17_Kawasaki lineage shows similarities with the pattern observed for GII.4 norovirus evolution. Virus movements from Hong Kong to neighboring coastal cities were frequently observed.

***Conclusions.*** Our results provide new insights into GII.17 norovirus evolution and transmission and highlight the potential for a rare norovirus genotype to rapidly replace existing strains and cause local epidemics.

**Keywords.** norovirus; phylogenetic; phylogeographic; epidemic; virus transmission.

Noroviruses are divided into 7 genogroups and >30 genotypes according to the amino acid sequences of their VP1 genes. Viruses belonging to genogroups GI, GII, and GIV infect humans [1]. However, approximately 80% of norovirus outbreaks since the mid-1990s have been caused by a single norovirus genotype, GII.4 [2, 3]. New variants of GII.4 have emerged every 2–3 years and may escape population immunity through the virus' high rate of mutation and through intragenotype recombination [4, 5].

In contrast to genotype GII.4, GII.17 is an uncommon norovirus genotype that was, until recently, rarely detected in human cases [6]. However, in the winter of 2014–2015, a large acute gastroenteritis (AGE) outbreak in China was caused by GII.17 norovirus. Epidemics of GII.17 infection were reported almost simultaneously from at least 9 provinces or regions in mainland China, ranging from Beijing in the north to Guangdong Province in the south during November 2014 [7–11]. All these outbreaks were caused by a newly detected norovirus belonging to the GII.17_Kawasaki lineage [12]. Further sporadic infections with this strain were detected outside mainland China, including in Hong Kong [13], Taiwan [14], Japan [12], USA [15], Italy [16] and Romania [17]. Current surveillance data suggests that the GII.17_Kawasaki strain has been widely distributed and has the potential to cause epidemics worldwide. Previous sequence analyses of GII.17 viruses have undertaken only basic genotyping and phylogenetic analysis [12, 13] and the origin, evolution and molecular epidemiology of GII.17 noroviruses are still largely unknown. Addressing these issues is important for designing optimal surveillance and control strategies.

In this study, we sequenced 96 GII.17 noroviruses collected from 42 AGE outbreaks during the 2014–2015 and 2015–2016 epidemic seasons in Guangdong, China. In addition, sporadic norovirus cases that preceded these outbreaks were also screened, resulting in the identification of 7 GII.17 strains between 2013 and 2014. We combine our new data with all available GII.17 sequences to investigate GII.17 virus evolution. Using phylogenetic, geographic and molecular clock analyses we reveal the spread of the GII.17_Kawasaki lineage in Guangdong from its introduction across multiple epidemic seasons.

## METHODS

### Sample Collection and Sequencing

Ninety-six samples from 42 GII.17 outbreaks during the 2014–2015 and 2015–2016 epidemic seasons were collected for capsid

Received 28 March 2016; accepted 13 May 2016; published online 24 May 2016.

[a]O. G. P. and H. L. contributed equally to the study and are joint senior authors.

Correspondence: H. Li, Guangdong Provincial Center for Disease Control and Prevention, No. 160, Qunxian Road, Dashi Town, Panyu District, Guangzhou City, Guangdong Province, China (gdcdclihui@163.com).

**The Journal of Infectious Diseases® 2016;214:556–64**

© The Author 2016. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail journals.permissions@oup.com.

DOI: 10.1093/infdis/jiw208

sequencing. To trace the origin and evolutionary history of the GII.17_Kawasaki variant, we performed a retrospective study to screen previously untyped norovirus samples. In total, 103 capsid sequences and 8 whole-genome sequences of GII.17 were generated in this study and are available on GenBank (accession numbers KU557783-KU557893).

#### VP1 Data Set Preparation

We compiled 2 different norovirus GII.17 VP1 data sets, one comprising all GII.17 strains and one comprising only the newly emerged GII.17_Kawasaki lineage (see Results for details). Multiple sequence alignment was performed using ClustalW [18]. For each gene data set, the temporal accumulation of genetic divergence in the phylogeny was assessed from maximum likelihood midpoint rooted phylogenies, using the root-to-tip regression method implemented in TempEst (formerly Path-O-Gen) [19].

#### Dated Phylogenetic Analysis

Molecular clock phylogenetic analysis was performed using the Bayesian Markov chain Monte Carlo (MCMC) framework implemented in BEAST. We used a SRD06 nucleotide substitution model [20] and an uncorrelated lognormal relaxed clock model to accommodate variation in substitution rates among branches [21]. To reconstruct changes in the genetic diversity of GII.17 we used the Bayesian skyline plot coalescent model, which estimates effective population size, using a piecewise constant model of effective population size [22]. The output of the skyline plot model represents the product of effective population size ($N_e$) and generation time ($\tau$) through time. Three independent MCMC runs of $1 \times 10^8$ steps were computed. Convergence and behavior of MCMC chains was inspected using Tracer v1.6 (available at: http://tree.bio.ed.ac.uk) [23]. A subset of 500 trees was randomly drawn from the combined posterior distribution of trees and used as an empirical distribution for subsequent analysis [23].

#### Phylogeography

We used a Bayesian discrete phylogeographic approach to investigate viral spatial transmission among 4 geographic regions (Africa, America, Asia, and Europe; Figure 2) for the full data set and 13 local Chinese regions for the local data set (Figure 3). To provide a more realistic model of the direction of virus transmission, we used an asymmetric continuous-time Markov chain model [24] to estimate ancestral locations and to estimate location posterior probabilities for each node in the time-scaled phylogenies. A Bayesian stochastic search variable selection method was used to identify well-supported viral lineages movements and to summarize epidemiological connectivity between locations [25]. We also undertook posterior inference of the complete Markov jump history through time, which allowed us quantify changes in location and the time spent in a particular location along each phylogenetic branch.

## RESULTS

#### GII.17 Outbreaks

A sharp increase in the number of norovirus outbreaks was observed in Guangdong during November 2014 (Figure 1*A*). In total, 50 outbreaks were recorded between November 2014 and April 2015, representing a >4-fold increase in outbreaks as compared to the 2012–2013 season (10 outbreaks) and the 2013–2014 season (12 outbreaks). Laboratory testing identified that a novel variant of GII.17 norovirus replaced GII.4_Sydney_2012 norovirus as the primary cause of AGE outbreaks [10].

GII.4_Sydney_2012, first detected in August 2012 in a sporadic infection case in Guangdong [26], caused 11 outbreaks in 4 prefecture-level cities between 2013 and 2015 (Figure 1*B*). By contrast, the GII.17 strain that emerged in Guangdong in August 2013 has already caused 48 outbreaks in 11 cities. Moreover, GII.4_Sydney_2012 outbreaks mainly occurred in schools in Guangdong (10 of 11 [91%]), whereas the AGE outbreaks caused by GII.17 were reported in a wider range of locations, including schools (31 of 44 [70%]), factories (9 of 44 [20%]), hospitals (2 of 44 [5%]), and kindergartens (2 of 44 [5%]). Partial data for the winter season of 2015–2016 suggests that GII.17 has reemerged, although the epidemic size and geographic distribution of GII.17 outbreaks appear more limited than in 2014–2015 (Figure 1*A* and 1*B*). These data suggest that the infectivity of the novel GII.17 variant is at least comparable with that of GII.4. The high epidemic risk of the GII.17 strain prompted us to further investigate its evolution and transmission in this study.

#### Molecular Clock Phylogeny and Phylogeography

Before phylogenetic analysis, recombination in GII.17 was examined by combining 8 GII.17 noroviruses collected in 2013–2015 in Guangdong with all 16 genomes available from GenBank (as of November 2015). Recombination in the RNA-dependent RNA polymerase (RdRp) gene of GII.17 has been proposed previously [12, 27], and more recombination events can be seen when updated GII.17 sequences are included (Supplementary Figure 1). Owing to frequent recombination in RdRp sequences, the phylogenetic analyses below were based on VP1 (capsid gene) sequences, and no recombination could be detected in those individual data sets.

In this study, we generated 103 full capsid sequences of GII.17 norovirus collected from 2013 to 2015 in Guangdong. We combined these sequences with 109 full or partial sequences from GenBank to obtain the full data set of 212 sequences. Poor phylogenetic temporal signal is obtained if the phylogeny includes the oldest GII.17 strain (GII.17_GUF_1978; Supplementary Figure 2*A*). In contrast, rooting the tree by using the oldest strain within the GII.17_Kawasaki lineage (JX629786_KR_2009) reveals a good correlation between sampling date and root-to-tip genetic distance (Supplementary Figure 2*B*), suggesting that different GII.17 lineages may evolve at different rates.

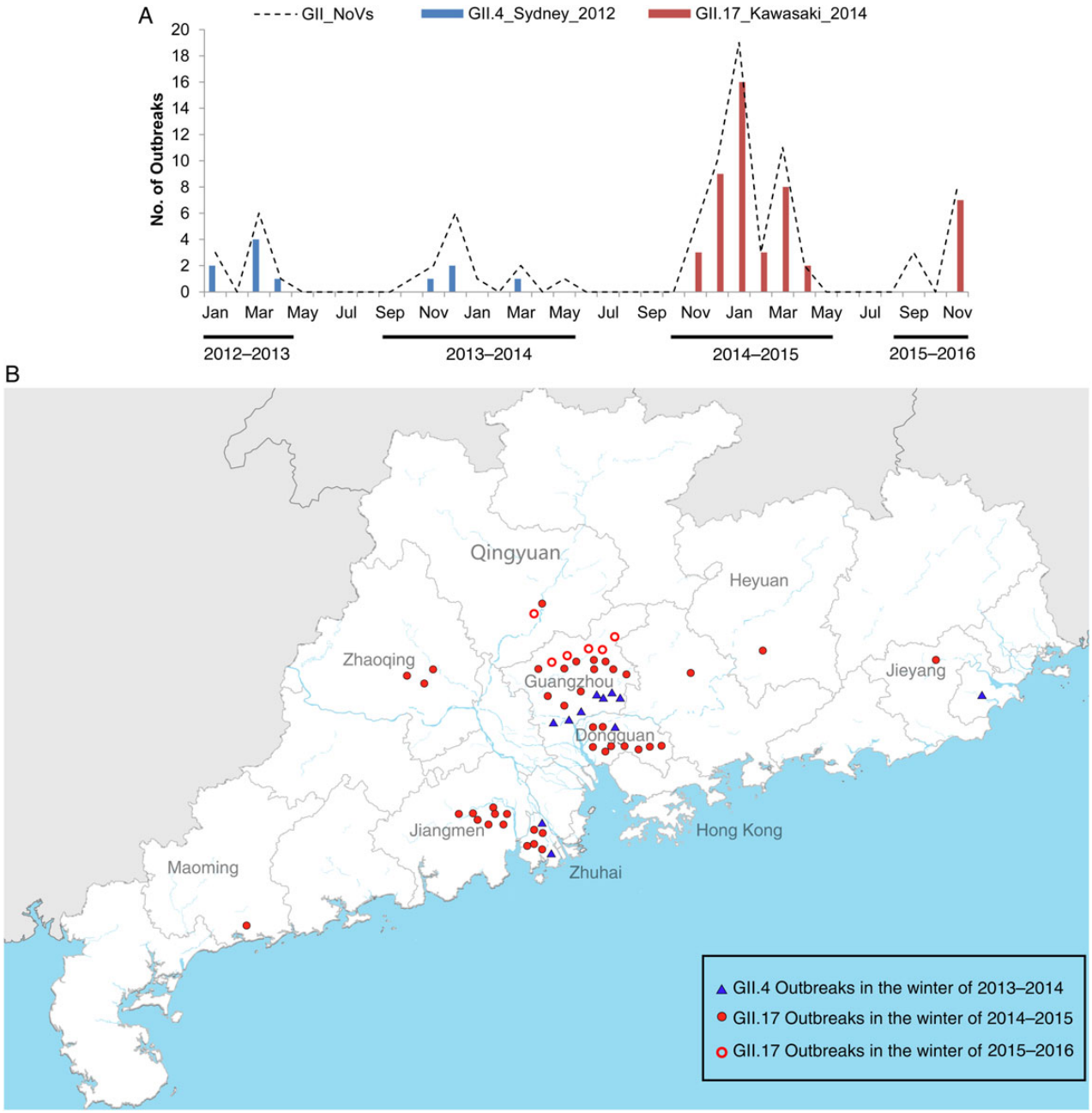


**Figure 1.** Outbreaks of GII.4 and GII.17 infections in Guangdong China, 2013–2015. *A*, Time series of the number of GII norovirus (NoV) outbreaks during the study period, from January 2013 to November 2015. GII.4 (blue) and GII.17 (red) were detected as major causes of outbreaks in different epidemic seasons and are presented with bar charts. The dotted line represented the total number of NoV outbreaks. *B*, Geographic distribution of GII.4 and GII.17 NoV outbreaks in Guangdong from January 2013 to November 2015.

Consequently, parameters of the molecular clock model were estimated separately for the full GII.17 data set and the GII.17_Kawasaki lineage data set (Table 1).

A molecular clock and phylogeographic analysis of the full GII.17 data set was undertaken using a Bayesian inference approach. As Figure 2 shows, the GII.17 VP1 gene phylogeny contains 3 distinct lineages. Lineage A, which includes the oldest strain GII.17_GUF_1978, shares a common ancestor around 1966 (1974–1994, 95% highest posterior density [HPD]). The GII.17 strains within this lineage have been rarely detected recently, except for strain GII.17_NS682_2015, which was isolated in Hong Kong in 2015. This demonstrates that the lineage A GII.17 norovirus is still circulating. Lineage B shares a common ancestor around 1987 (1973–1994, 95% HPD) and has been recently detected in different continents, including Asia, Europe, and Africa. Lineage B contains viral strains detected in wastewater

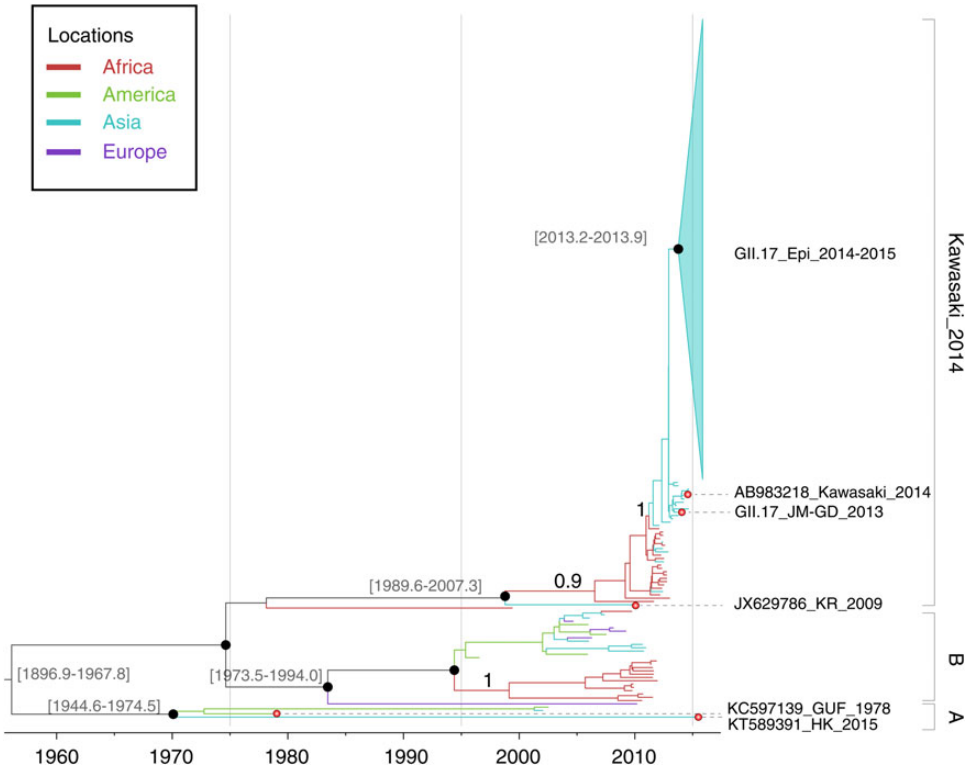


**Figure 2.** Molecular clock phylogeny of GII.17 VP1 gene sequences. The tree shown is a maximum clade credibility phylogeny, Branch colors represent the most probable ancestral locations of each branch, inferred using a phylogeographic model (Supplementary Materials and Methods). Long internal branches with uncertain locations are in gray. Three major lineages of GII.17 noroviruses are denoted on the right. Black circles indicate posterior probabilities of >0.80 at selected nodes. Location posterior probabilities are noted above 3 specific branches.

in South Africa in 2010–2011 [28] and was detected in samples from healthy children in Cameroon in 2009 [29].

The third lineage of GII.17 was named GII.17_Kawasaki in a previous study [12], and AB983218 (Kawasaki323) is regarded as the first identified strain of the new GII.17 variant. However, our surveillance data suggest the existence of earlier GII.17 cases belonging to this lineage (eg, an isolate from a sporadic infection case in Guangdong in August 2013; GII.17_JM-GD_2013; Figure 2). The inclusion of a greater number of partial capsid sequences during the analysis provides a better description of the evolution of this lineage, and strong posterior probability support (1.0) was observed for the basal node of this lineage (Figure 2). Our analysis estimates that the MRCA of this lineage existed around 2001 (1989–2007, 95% HPD interval), consistent with our analysis that used only GII.17_Kawasaki lineage sequences (Table 1 and Supplementary Figure 3). As illustrated in Figure 2, the strains near the root of GII.17_Kawasaki lineage were detected mainly in Africa between 2007 and 2012, with occasional detection in Asian countries. After 2013, GII.17_Kawasaki strains were detected increasingly frequently in sporadic infections from Korea, Japan, and China, suggesting that the lineage has become established in Asia. In the following winter, all sequenced isolates from the GII.17 epidemic outbreaks were placed in a single clade of low genetic diversity (denoted GII.17_Epi_2014-2015). Phylogenetic analysis (Figure 2) indicates that the MRCA of the GII.17 epidemic strains existed around August 2013 (April 2013–December 2013, 95% HPD).

#### GII.17 Evolution, Dynamic, and Transmission in Guangdong

The continuous surveillance system in Guangdong allows us to study how the novel GII.17 variant evolved and spread within the region. GII.17 epidemics occurred at the same time in neighboring Hong Kong [13], so viral sequences from Hong Kong were included in the analysis. A Bayesian phylogeographic analysis was performed to reconstruct the spatial transmission of GII.17 in Guangdong. The dashed vertical lines in Figure 3 separate the epidemic from the preepidemic period and the 2014–2015 and 2015–2016 epidemic seasons. A few strains circulated in Guangdong and Hong Kong during the preepidemic period, but these were divergent and shared a common ancestor with the outbreak lineage in 2008 (Figure 3*A*). During the 2014–2015 epidemic season, GII.17 noroviruses in Guangzhou and Hong Kong were split into 2 distinct clades. The smaller clade (labeled Epi_GZ in Figure 3*A*) appears to have moved from Guangzhou (also shown in Figure 3*C*). The larger clade (labeled Epi_HK in Figure 3*A*) is inferred to have originated in Hong Kong and led to major

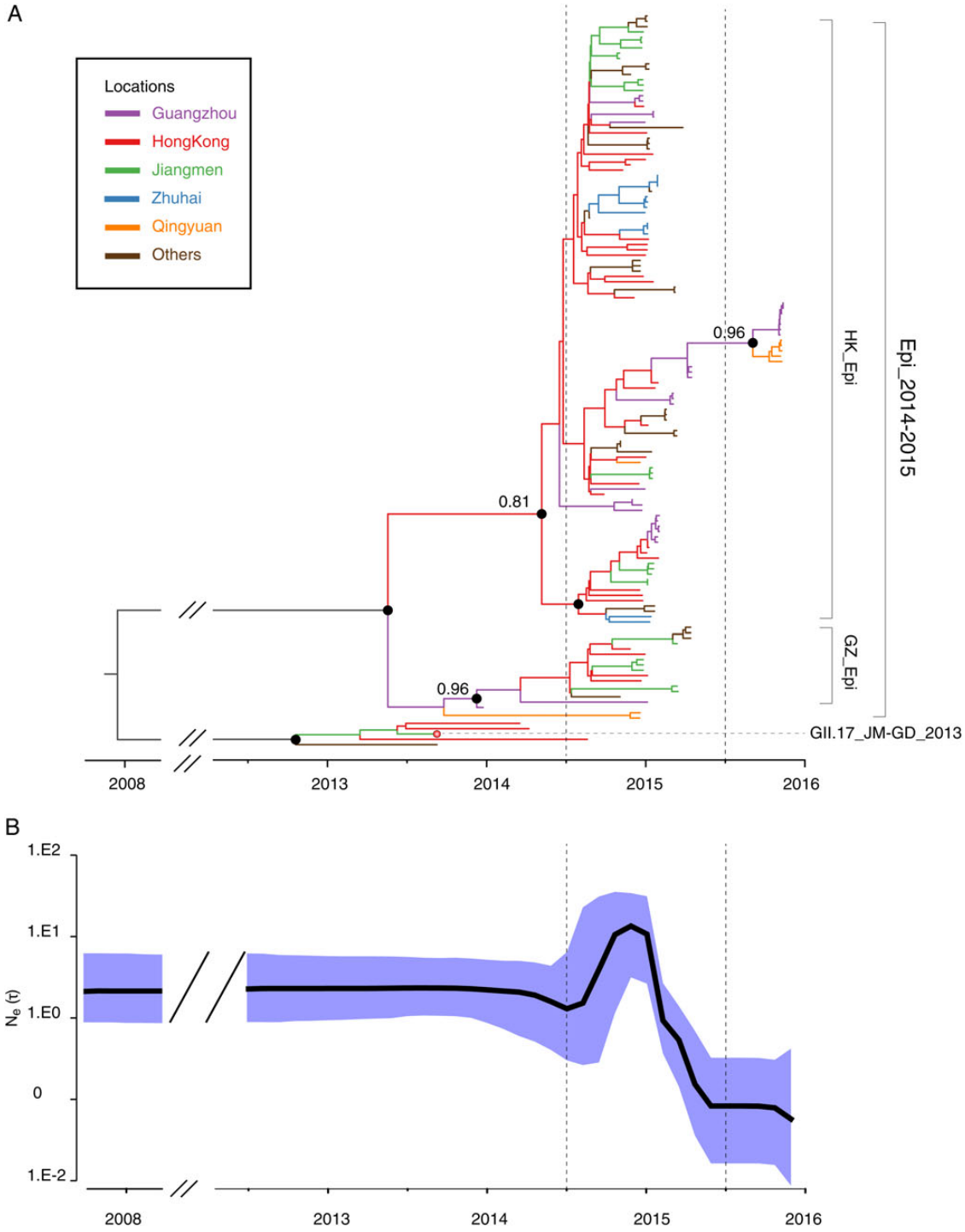


**Figure 3.** Inferred phylogeny, demographic history, and spatial dissemination of GII.17 norovirus in Guangdong. *A*, maximum clade credibility phylogeny with branches colored according to the most probable posterior ancestral location (see legend). Black circles indicate posterior probabilities of >0.80 at selected nodes. Long internal branches with uncertain locations are in gray. *B*, Demographic history of GII.17 norovirus in Guangdong, inferred using the Bayesian Skyline plot approach. *C*, Well-supported virus translocation events (Bayes factor >3) are represented on a map of Guangdong. Arrow thickness represents the relative number of well-supported movement events for each pair of locations. The inset on the bottom right shows the water flows among the cities in Guangdong among which virus movement was inferred. $N_e(\tau)$, effective population size.

outbreaks in Guangdong and Hong Kong. To avoid possible biases in phylogeographic inference that may arise from an over-representation of strains from Hong Kong in the full data set, we undertook random downsampling so that the numbers of sequences from the 3 most densely sampled locations were equal (Supplementary Table 1). Across these randomly subsampled

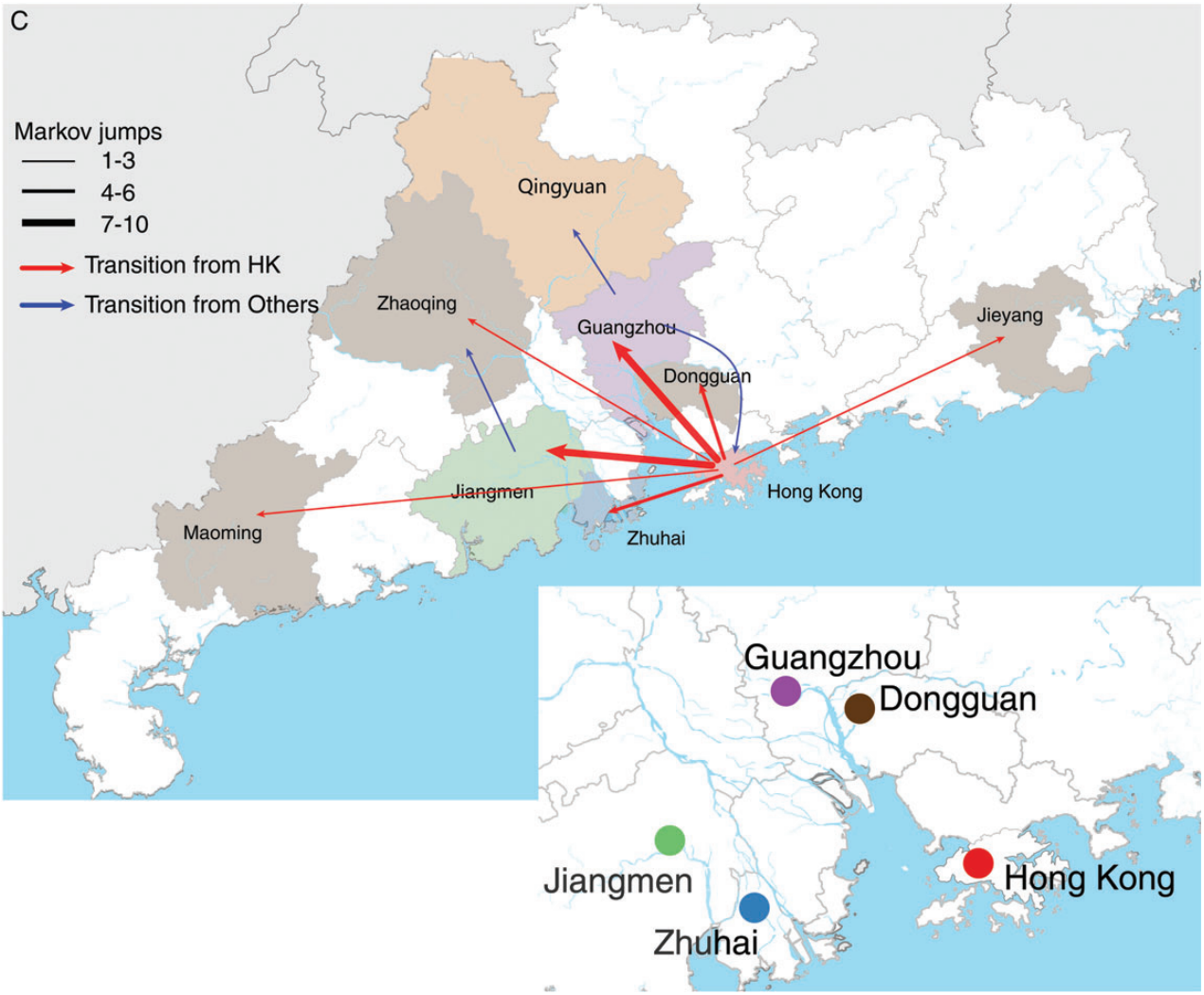


*Figure 3. Continud*

data sets, we consistently found that the major clade of GII.17 epidemic strains was most likely to have a common ancestor in Hong Kong (Supplementary Figure 4). In November 2015, this clade of GII.17 outbreaks reemerged in Guangzhou and the neighboring city of Qingyuan. Our analysis reveals that the GII.17 strain that reemerged in the 2015–2016 epidemic season likely descended from strains previously circulating in Guangzhou in early 2015, thereby suggesting that the virus might have persisted locally between epidemic seasons.

A Bayesian skyline plot coalescent model was used to infer the demographic history of the GII.17 lineage in Guangdong. The skyline plot (Figure 3*B*) shows a rise and fall in effective population size ($N_e[\tau]$) that coincides with the epidemic peak during the 2014–2015 winter season observed by the norovirus surveillance systems in Guangdong (Figure 1*A*). The skyline plot is flat before this owing to an absence of coalescence events in the phylogeny. Specifically, preepidemic GII.17 norovirus was detected in Guangdong on August 2013 (GD-JM_2013; Figure 3*A*), and replacement of the GD-JM_2013 variant by the GII.17_Epi_2014-2015 lineage coincided with a sharp increase in the effective number of infections. However, only 1 lineage from the 2014–2015 season has been observed in the winter of 2015–2016, leading to a sudden loss of viral genetic diversity (possible population bottleneck), resulting a steep decline in the skyline plot near the present.

The Markov jump history through time generated by the discrete phylogeographic analysis was used to quantify among-location changes along each phylogenetic branch. Figure 3*C* shows all virus lineage movements that were statistically well supported (Bayes factor >3). Hong Kong appears to have acted as an epicenter during local GII.17 outbreaks. In total, 82% of virus lineage movements during the evolutionary history of GII.17 were away from Hong Kong and toward other cities in Guangdong. Specifically, the estimated number of GII.17 lineage movements was high for Hong Kong to Jiangmen (n = 10.7), Hong Kong to Guangzhou (n = 9.0),

**Table 1. Evolutionary Parameters for the GII.17 VP1 Gene Inferred by Bayesian Analysis**

| Parameter | Data Set | |
|---|---|---|
| | GII.17_Kawasaki_2014 (n = 178) | GII.17 Full (n = 212) |
| RdRp genotype(s) | GII.P17 | GII.P4, GII.P16, GII.Pe |
| Evolutionary rate, $\times 10^{-3}$ substitutions/site/year[a] | 5.68 (3.76–7.87) | 4.83 (3.36–6.36) |
| TMRCA[a] | 2001 (1989–2008) | 1944 (1897–1968) |
| Coefficient of variation[a] | 0.98 (0.48–1.38) | 1.07 (0.72–1.46) |

Abbreviations: RdRp, RNA-dependent RNA polymerase; TMRCA, the most recent common ancestor.

[a] Data are median of the posterior distribution (95% highest posterior density credible interval).

Hong Kong to Dongguan (n = 6.2), and Hong Kong to Zhuhai (n = 5.4).

## DISCUSSION

GII.17 was, until 2014, a rarely detected norovirus genotype, with only limited cases of infection documented since 1978 [6, 30]. However, a novel lineage, termed GII.17_Kawasaki, emerged in August 2013 in Guangdong, where it replaced GII.4 and caused a significant increase in AGE outbreaks during the winter of 2014–2015 [10]. Epidemics caused by this lineage were detected almost simultaneously in several provinces of China, and sporadic infection cases were also reported worldwide [6]. In November 2015, outbreaks of GII.17 reemerged in Guangdong, providing evidence that the virus had persisted in local populations. Here we undertook continuous surveillance, virus genome sequencing and phylogenetic analysis to characterize the origin, evolution, and transmission of GII.17 norovirus.

Owing to its previously low epidemic activity, the evolutionary dynamics of GII.17 have not been previously been studied in detail. Chen et al [13] recently reported a relatively high rate of evolutionary change for GII.17. In present study, we characterized the evolution of GII.17 using 2 data sets, one comprising all GII.17 strains and one comprising only the newly emerged GII.17_Kawasaki lineage.

GII.17 appears to exhibit significant heterogeneity in rate of evolutionary change among branches. The Bayesian phylogenetic analysis of both data sets gave high values for the coefficient of variation parameter (Table 1). In addition, rooting the full data set phylogenetic tree on the oldest strain GII.17_GUF_1978 is problematic (Supplementary Figure 2). These results suggest that GII.17 evolves at a variable rate; this variance may be caused by different factors. First, the variation may reflect the fact the phylogeny includes both well-sampled epidemic lineages as well as long internal branches that might represent the persistence of noroviruses in their natural reservoir, which is as yet unknown (Figure 2). A similar model has been proposed for the Ebola virus in West Africa; the Ebola virus rate of evolutionary change during the 2014 epidemic was higher than the rate between outbreaks, which represents transmission in the zoonotic reservoir [31]. Second, it is also possible that variation in the rate of evolutionary change in the VP1 gene is affected by recombination, because recombination results in the same VP1 lineage being associated with different RdRp proteins. As shown in Supplementary Figure 1 and Table 1, 3 GII.17_A strains adopt 3 different RdRp proteins (GII.P4, GII.P16, and GII.Pe). The RdRp of GII.17_Kawasaki is divergent from those previously identified and named GII.P17 [12]. Recombination between RdRp and VP1 may have allowed GII.17 to acquire a RdRp with a lower fidelity and/or increased replicative ability, resulting in a higher rate of evolutionary change for the GII.17 VP1 gene. The molecular clock results suggest that the GII.17_A and GII.17_Kawasaki lineages shared a common ancestor some time ago (point estimate = 1944; Table 1 and Figure 2).

In recent decades, the worldwide predominance of the GII.4 genotype in norovirus epidemics suggests that GII.4 noroviruses may possess greater viral fitness in human populations as compared to other norovirus genotypes. GII.4 norovirus has a high rate of evolutionary change, which may increase its ability to evade immune responses [2]. We estimated the rate of evolutionary change of GII.17 VP1 to be $5.68 \times 10^{-3}$ substitutions/site/year (95% HPD, $3.76\times10^{-3}$–$7.87\times10^{-3}$) for the GII.17_Kawasaki lineage, and $4.83 \times 10^{-3}$ substitutions/site/year (95% HPD, $3.36\times10^{-3}$–$6.36\times10^{-3}$) for the GII.17 full data set (Table 1). For both data sets, the rate of evolutionary change of GII.17 is comparable to that estimated for GII.4 VP1 sequences ($5.3$–$6.3 \times 10^{-3}$ substitutions/site/year) [32, 33] and is higher than that estimated for the GII.3 ($1.961 \times 10^{-3}$ substitutions/site/year) and GII.7 ($2.36 \times 10^{-3}$ substitutions/site/year) genotypes [33]. Therefore, the capacity of GII.17 to evolve, adapt, and undergo antigenic drift may equal to that of the GII.4 genotype.

Elucidating the circulation pattern of viruses from genomic data provides valuable information for epidemic prediction and management. The most significant feature of the GII.4 epidemic during the last decade is its epochal nature, with periods of epidemic stasis followed by the emergence of a novel GII.4 variant coinciding with a new epidemic peak [32, 34–36]. A similar behavior is exhibited in our study by the GII.17_Kawasaki lineage. To begin, the preepidemic GII.17 strains (GD-JM_2013–like viruses) were detected around August 2013 and caused only a few sporadic infections. Next, the GD-JM_2013–like viruses were replaced with GII.17_Epi_2014-2015 variants and coincided with the occurrence of a series of amino acid changes (Figure 3*A* and Supplementary Figure 5*A*); the GII.17_Epi_2014-2015 variants were associated with the epidemic spread of GII.17 in Guangdong. The Bayesian skyline plot analysis also shows a sharp rise in viral effective population

size after this replacement event. Following the dominance GII.17 in outbreaks during the 2014–2015 epidemic season, we observed more-limited GII.17 outbreaks during winter 2015–2016, although data for that season are incomplete. This coincides with a notable decrease in viral genetic diversity in the skyline plot; this viral population bottleneck could result from herd immunity against GII.17 or from seasonal variation in transmission. A similar pattern has been observed for GII.4_2004 and GII.4_2006a variants; specifically, the emergence of a genetically and antigenically novel GII.4 variant was followed by an epidemic peak and later diminishment [32]. In addition, spontaneous mutations in the capsid gene contribute to a long branch between the preepidemic GII.17 strains and the GII.17_Epi_2014-2015 strains (Figure 3*A* and Supplementary Figure 3). This phylogenetic feature is also frequently observed during GII.4 evolution and is associated with antigenic shift and possible immune evasion [32, 34]. We assessed the potential function of amino acid substitutions during GII.17 evolution by comparing preepidemic, epidemic 2014–2015, and re-emerged 2015–2016 viral strains (Supplementary Figure 5*A*) and mapped the locations of these changes onto the homodimeric crystal structure of the GII.4 P domain (strain VA387; Protein Data Bank accession number 2OBT [37]). We found that 8 of these substitutions were located in previously identified epitopes of potential functional importance (Supplementary Results and Discussion) and may possibly influence the antigenicity, pathogenicity, and infectivity of the viruses.

Our phylogeographic analysis of GII.17 in Guangdong suggests that Hong Kong was an epicenter during the GII.17 outbreaks in 2014–2015 (Figure 3). Virus translocations from Hong Kong to 4 other cities in Guangdong are frequently observed throughout the GII.17 phylogeny (Figure 3*A* and 3*C*), even though these locations are not geographically proximate. Crucially, random downsampling of overrepresented locations gives consistent estimates of the inferred locations of major nodes in the phylogeny (Supplementary Figure 4) and of viral lineage movements (Supplementary Table 2), although the 95% confidence intervals for the number of estimated movement events were increased in the downsampled data sets owing to smaller sample sizes. This indicates that our results are robust and that the ancestral locations and movement events that we infer for the real data are not simply reflecting the relative frequency of sampling locations.

The major reservoir(s) of GII.17 that contribute to transmission among Hong Kong and Guangdong coastal cities are still unknown. Seawater intrusion into freshwater aquifers, which always occurs during the winter season in the Pearl River Delta, may partially contribute to virus spread, as it causes pollution of inshore shellfish cultivation. Indeed, the new variant of GII.17_Epi_2014-2015 was identified in cultured oysters as early as March 2014 (X. W. Zhong, unpublished data). Recent norovirus sequences from Korea (KT383964 and KT438801) also suggest the presence of GII.17 in seawater and short-necked clams during March 2014–March 2015. Taken together, these results suggest that GII.17 noroviruses in the environmental water and farmed seafood of coastal cities should be closely monitored for future disease control and prevention.

## Supplementary Data

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

## Notes

***Financial support***. This work was supported by the National Natural Science Foundation of China (grant 81501754), the Natural Science Foundation of Guangdong Province (grant 2015A030313709), the Science and Technology Planning Project of Guangdong Province (grant 2014A020212243), the China Scholarship Council (grant 201508440009), the Medical Research Council (grant MR/L009528/1), the Wellcome Trust (grant 090532/Z/09/Z), and the European Research Council, under the European Union's Seventh Framework Programme (FP7/2007-2013/ERC grant 614725-PATHPHYLODYN to O. G. P.)

***Potential conflicts of interests***. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. Kroneman A, Vega E, Vennema H, et al. Proposal for a unified norovirus nomenclature and genotyping. Arch Virol **2013**; 158:2059–68.
2. Centers for Disease Control and Prevention. Emergence of new norovirus strain GII.4 Sydney–United States, 2012. MMWR Morb Mortal Wkly Rep **2013**; 62:55.
3. van Beek J, Ambert-Balay K, Botteldoorn N, et al. Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. Euro Surveill **2013**; 18:8–9.
4. Debbink K, Lindesmith LC, Donaldson EF, et al. Emergence of new pandemic GII.4 Sydney norovirus strain correlates with escape from herd immunity. J Infect Dis **2013**; 208:1877–87.
5. Lindesmith LC, Donaldson EF, Lobue AD, et al. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med **2008**; 5:e31.
6. de Graaf M, van Beek J, Vennema H, et al. Emergence of a novel GII.17 norovirus - End of the GII.4 era? Euro Surveill **2015**; 20:pii:21178.
7. Fu J, Ai J, Jin M, 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:pii:21157.
8. Gao Z, Liu B, Huo D, et al. Increased norovirus activity was associated with a novel norovirus GII.17 variant in Beijing, China during winter 2014–2015. BMC Infect Dis **2015**; 15:574.
9. Han J, Ji L, Shen Y, Wu X, Xu D, Chen L. Emergence and predominance of norovirus GII.17 in Huzhou, China, 2014–2015. Virol J **2015**; 12:139.
10. Lu J, Sun L, Fang L, et al. Gastroenteritis Outbreaks Caused by Norovirus GII.17, Guangdong Province, China, 2014–2015. Emerg Infect Dis **2015**; 21:1240–2.
11. Wang X, Yong W, Shi L, et al. An outbreak of multiple norovirus strains on a cruise ship in China, 2014. J Appl Microbiol **2015**; 120:226–33.
12. Matsushima Y, Ishikawa M, Shimizu T, 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:pii:21173.
13. Chan MC, Lee N, Hung TN, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. Nat Commun **2015**; 6:10061.
14. Lee CC, Feng Y, Chen SY, Tsai CN, Lai MW, Chiu CH. Emerging Norovirus GII.17 in Taiwan. Clin Infect Dis **2015**; 61:1762–4.
15. Parra GI, Green KY. Genome of Emerging Norovirus GII.17, United States, 2014. Emerg Infect Dis **2015**; 21:1477–9.
16. Medici MC, Tummolo F, Calderaro A, et al. Identification of the novel Kawasaki 2014 GII.17 human norovirus strain in Italy, 2015. Euro Surveill **2015**; 20:pii:30010.

17. Dinu S, Nagy M, Negru DG, Popovici ED, Zota L, Oprisan G. Molecular identification of emergent GII.P17-GII.17 norovirus genotype, Romania, 2015. Euro Surveill **2015**; 20:pii:30010.
18. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics **2007**; 23:2947–8.
19. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol **2016**; 2:vew007.
20. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol **2006**; 23:7–9.
21. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol **2006**; 4:e88.
22. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol **2005**; 22:1185–92.
23. Lemey P, Rambaut A, Bedford T, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog **2014**; 10:e1003932.
24. Edwards CJ, Suchard MA, Lemey P, et al. Ancient hybridization and an Irish origin for the modern polar bear matriline. Curr Biol **2011**; 21:1251–8.
25. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol **2009**; 5:e1000520.
26. Sun L, Li H, Tan X, et al. [Epidemiological characteristics of norovirus variant of GII.4 Sydney and the outbreaks caused by norovirus variant of GII.4 Sydney in Guangdong province, 2012–2014]. Zhonghua Yu Fang Yi Xue Za Zhi **2015**; 49:615–20.
27. Mahar JE, Bok K, Green KY, Kirkwood CD. The importance of intergenic recombination in norovirus GII.3 evolution. J Virol **2013**; 87:3687–98.
28. Murray TY, Mans J, Taylor MB. Human calicivirus diversity in wastewater in South Africa. J Appl Microbiol **2013**; 114:1843–53.
29. Ayukekbong J, Lindh M, Nenonen N, Tah F, Nkuo-Akenji T, Bergstrom T. Enteric viruses in healthy children in Cameroon: viral load and genotyping of norovirus strains. J Med Virol **2011**; 83:2135–42.
30. Rackoff LA, Bok K, Green KY, Kapikian AZ. Epidemiology and evolution of rotaviruses and noroviruses from an archival WHO Global Study in Children (1976–79) with implications for vaccine design. PLoS One **2013**; 8:e59394.
31. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science **2014**; 345:1369–72.
32. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. PLoS Pathog **2010**; 6:e1000884.
33. Bull RA, Eden JS, Rawlinson WD, White PA. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog **2010**; 6:e1000831.
34. Bok K, Abente EJ, Realpe-Quintero M, et al. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. J Virol **2009**; 83:11890–901.
35. Donaldson EF, Lindesmith LC, Lobue AD, Baric RS. Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. Immunol Rev **2008**; 225:190–211.
36. Siebenga JJ, Vennema H, Renckens B, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. J Virol **2007**; 81:9932–41.
37. Bu W, Mamedova A, Tan M, Xia M, Jiang X, Hegde RS. Structural basis for the receptor binding specificity of Norwalk virus. J Virol **2008**; 82:5340–7.