

Preadaptation of pandemic GII.4 noroviruses in unsampled virus reservoirs years before emergence

Christopher Ruis,^{1,*}† Lisa C. Lindesmith,² Michael L. Mallory,²
Paul D. Brewer-Jensen,² Josephine M. Bryant,¹ Veronica Costantini,^{3,‡}
Christopher Monit,¹ Jan Vinjé,³ Ralph S. Baric,² Richard A. Goldstein,^{1,*}§,**
and Judith Breuer^{1,4, **}

¹Division of Infection and Immunity, University College London, London WC1E 6BT, UK, ²Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA, ³Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA and ⁴Department of Microbiology, Virology and Infection Control, Great Ormond Street Hospital for Children, London, UK

*Corresponding author: E-mail: r.goldstein@ucl.ac.uk; E-mail: christopher.ruis.10@ucl.ac.uk

†<https://orcid.org/0000-0003-0977-5534>

‡<https://orcid.org/0000-0002-1532-4345>

§<https://orcid.org/0000-0001-5148-4672>

**These authors contributed equally to this work.

Abstract

The control of re-occurring pandemic pathogens requires understanding the origins of new pandemic variants and the factors that drive their global spread. This is especially important for GII.4 norovirus, where vaccines under development offer promise to prevent hundreds of millions of annual gastroenteritis cases. Previous studies have hypothesized that new GII.4 pandemic viruses arise when previously circulating pandemic or pre-pandemic variants undergo substitutions in antigenic regions that enable evasion of host population immunity, as described by conventional models of antigenic drift. In contrast, we show here that the acquisition of new genetic and antigenic characteristics cannot be the proximal driver of new pandemics. Pandemic GII.4 viruses diversify and spread over wide geographical areas over several years prior to simultaneous pandemic emergence of multiple lineages, indicating that the necessary sequence changes must have occurred before diversification, years prior to pandemic emergence. We confirm this result through serological assays of reconstructed ancestral virus capsids, demonstrating that by 2003, the ancestral 2012 pandemic strain had already acquired the antigenic characteristics that allowed it to evade prevailing population immunity against the previous 2009 pandemic variant. These results provide strong evidence that viral genetic changes are necessary but not sufficient for GII.4 pandemic spread. Instead, we suggest that it is changes in host population immunity that enable pandemic spread of an antigenically pre-adapted GII.4 variant. These results indicate that predicting future GII.4 pandemic variants will require surveillance of currently unsampled reservoir populations. Furthermore, a broadly acting GII.4 vaccine will be critical to prevent future pandemics.

Key words: phylogenetics ; phylodynamics ; antigenic change ; serology.

1. Introduction

Noroviruses are the leading cause of acute gastroenteritis in humans worldwide, causing an estimated 684 million gastroenteritis episodes, 200,000 deaths and \$65 billion of health and societal costs annually (Kirk et al. 2015; Pires et al. 2015; Bartsch et al. 2016). While more than thirty norovirus genotypes have been described based on sequence variation in the VP1 capsid protein, the GII.4 genotype is responsible for the majority of human cases and outbreaks and has caused six major pandemics since the mid-1990s, each associated with a distinct pandemic variant: US95/96, Farmington Hills 2002, Hunter 2004, Den Haag 2006, New Orleans 2009, and Sydney 2012 (where the year denotes the year of onset of the respective pandemic) (Kroneman et al. 2008; Siebenga et al. 2009; White 2014; Vinje 2015; Motoya et al. 2017; Parra et al. 2017), as well as other geographically limited outbreaks caused by epidemic variants (Siebenga et al. 2009; Eden et al. 2014). Pandemic and epidemic variants are defined as well-supported clusters in the GII.4 phylogenetic tree (Chhabra 2019). The onset of a new GII.4 pandemic is well defined from epidemiological data. The new pandemic variant rapidly and simultaneously increases in prevalence to dominate outbreaks across many geographical regions. For example, Sydney 2012 rapidly emerged between November 2012 and January 2013 to become the dominant GII.4 variant on at least four continents (White 2014).

Vaccines currently under development offer promise to mitigate the global economic and health impact of new GII.4 pandemics (Lindesmith et al. 2015). However, effective vaccine design and distribution depend on understanding the sources from which new pandemic variants emerge and the factors that drive their global circulation especially if, like influenza, vaccine updates are necessary. It has been proposed that new pandemic GII.4 viruses generally evolve from one of the preceding pandemic variants (Siebenga et al. 2007; White 2014) through acquisition of substitutions in the capsid VP1 protein that alter antigenicity and enable evasion of host population immunity (Lindesmith et al. 2008, 2012; Debbink et al. 2013; Lindesmith et al. 2013; Eden et al. 2014). However, despite their rapid expansion upon pandemic emergence, the four most recent pandemic variants caused cases up to 5 years prior to pandemic emergence (Sdiri-Loulizi et al. 2009; Siebenga et al. 2009; Eden et al. 2014; van Beek et al. 2018), leading to suggestions that GII.4 variants circulate at low levels until they acquire the VP1 substitutions necessary to drive rapid pandemic emergence (Eden et al. 2014; White 2014; Tohma et al. 2019). Since several new GII.4 variants harbored a novel ORF1 gene, recombination has also been proposed to play a role in pandemic emergence (Eden et al. 2013), although precisely how is unknown.

Here, we combine phylogenetic and serological analyses to formulate and test a new hypothesis for GII.4 pandemic emergence. We demonstrate that pandemic GII.4 variants diversify over wide geographical areas over several years prior to pandemic onset. In depth analyses of sequence data show that genetic substitutions and recombination events that may be important for pandemic emergence are acquired years before such emergence occurs. Serological assays incorporating reconstructed ancestral strains of the Sydney 2012 pandemic variant demonstrate that key antigenic characteristics required for emergence had already been acquired by 2003, 9 years prior to pandemic spread. Together, our results show that viral genetic changes (substitutions and/or recombination events) are

necessary but not sufficient for pandemic spread and suggest a role for changes in host immunity in the emergence of new variants.

2. Results and discussion

To understand the origins and spread of norovirus pandemics, we reconstructed the temporal history of each genomic region of GII.4. This supports a conclusion that GII.4 was present for at least 50 years prior to the first documented pandemic (Fig. 1A, Supplementary Fig. S1, and Table S1), approximately 20 years earlier than previous estimates (Bok et al. 2009) due to inclusion of an additional early sequence that diverges from a more ancient node within the GII.4 phylogeny (Supplementary Fig. S2).

The phylogenetic relationships among GII.4 RdRp, VP1, and VP2 sequences are incompatible with the hypothesis that new pandemic/epidemic viruses evolve from previous pandemic/epidemic variants (Fig. 1A, Supplementary Fig. S1, and Table S2). Instead, the deep phylogenetic nodes suggest that GII.4 variants diverge from one another long before emerging to spread pandemically. While pandemic emergence is rapid (Fig. 1B), the long tree branch lengths indicate ongoing undetected circulation of pre-pandemic variants at low level during the period between divergence and pandemic emergence. For example, Den Haag 2006 and New Orleans 2009 were circulating as unique-independent lineages by 1997 (VP1, 95% highest probability density (HPD) 1995–2000) and 2004 (VP1, 95% HPD 2002–5), respectively. This undetected persistence over long time periods means that multiple future pandemic/epidemic variants co-circulated simultaneously. For example, at least six unsampled lineages co-circulated in the year 2000, four of which gave rise to five subsequent pandemics (Fig. 1A and Supplementary Table S2). These results are supported by previous reports of pre-pandemic circulation of Hunter 2004 (Sdiri-Loulizi et al. 2009), Den Haag 2006 (Siebenga et al. 2009; van Beek et al. 2018), New Orleans 2009 and Sydney 2012 (Eden et al. 2014; van Beek et al. 2018). We further identified previously sequenced pre-pandemic Farmington Hills 2002, Hunter 2004, New Orleans 2009 and Sydney 2012, and pre-epidemic Osaka 2007 sequences (Supplementary Fig. S3 and Tables S3 and S4). These pre-pandemic sequences sit closer to the root of their respective clade than sequences collected during the pandemic/epidemic (Supplementary Fig. S3) with placements in the tree in agreement with their sampling dates, supporting our inferred ancestor and divergence dates. The presence of multiple long future pandemic lineages that are only rarely sampled suggests that highly diverse viral populations are circulating within reservoirs that are not included in current surveillance.

We next used datasets of VP1 sequences to reconstruct a more detailed temporal history of the two most recent pandemic variants, New Orleans 2009 and Sydney 2012 (Fig. 2 and Supplementary Fig. S4). Similar to other pandemic GII.4 variants (Siebenga et al. 2010), New Orleans 2009 and Sydney 2012 VP1 regions underwent a large increase in relative genetic diversity coinciding with their pandemic emergence in 2009 and 2012, respectively (Supplementary Fig. S4). However, each variant had already diversified into many lineages prior to pandemic emergence, at least 67 (95% HPD 41–100) and 88 (95% HPD 59–113) lineages for New Orleans 2009 and Sydney 2012, respectively (Fig. 2A and B, Supplementary Fig. S4). Each of the other pandemic GII.4 variants also exhibits this pre-pandemic divergence (Fig. 1). Therefore, pandemic variants not only arise long before the pandemic is observed, but also undergo extensive

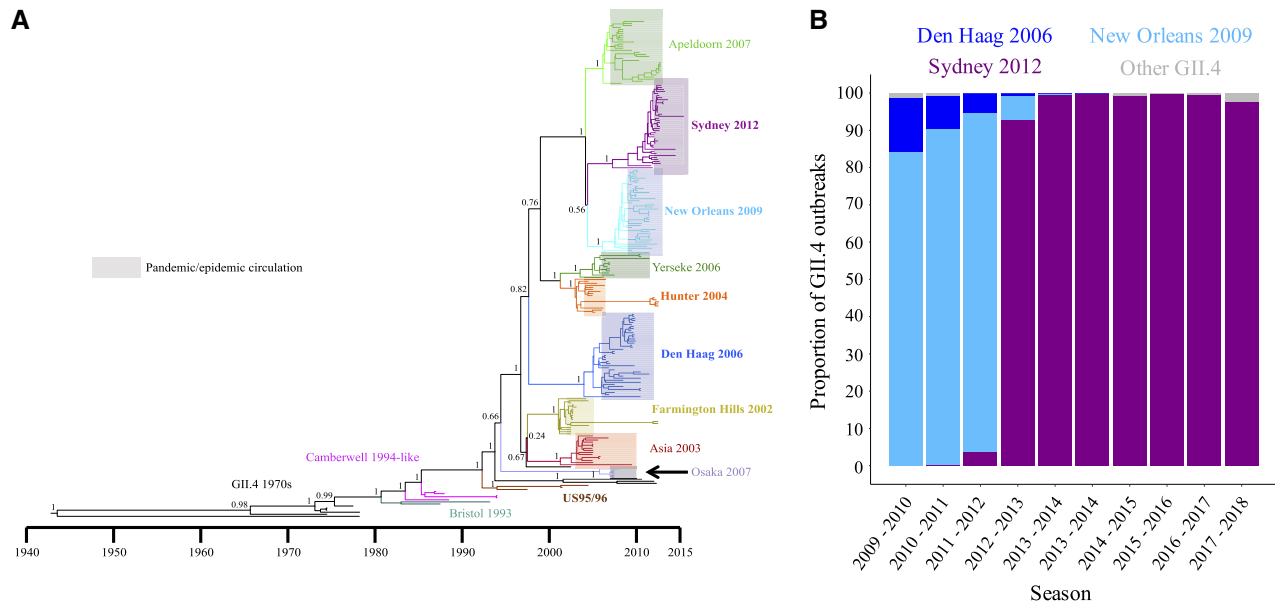


Figure 1. Pandemic GII.4 variants arise years before pandemic emergence and do not arise from preceding pandemic variants. (A) Temporal MCC tree of GII.4 VP1 sequences from major pandemic and epidemic variants. Variants diverge from all other sampled variants years before their emergence as a pandemic or epidemic (represented by the shaded area). Long branches throughout the tree indicate a high level of unsampled diversity through time. Posterior supports are shown on trunk nodes. (B) Proportion of GII.4 outbreaks in the USA caused by each pandemic GII.4 variant between 2009 and 2018, using outbreak data from CaliciNet (Vega et al. 2011). Sydney 2012 was first detected at low-level in 2011 then rapidly emerged to replace New Orleans 2009 in 2012.

diversification into multiple-related lineages that circulate at low levels for years preceding pandemic emergence.

To determine the extent of circulation prior to pandemic emergence, we reconstructed the spatiotemporal history of New Orleans 2009 and Sydney 2012, which demonstrated entry into each continent prior to pandemic emergence (Fig. 2). Specifically, New Orleans 2009 likely entered Africa, Asia and Oceania more than 2 years prior to pandemic emergence while Sydney 2012 was likely present in Africa, Europe, and Oceania more than 4 years before pandemic emergence (Fig. 2C and D). Following their introduction, these data suggest sustained intra- and inter-continental circulation of New Orleans 2009 and Sydney 2012 both before (at lower levels, Supplementary Fig. S4) and after (at higher levels) pandemic emergence (Fig. 2E and F).

The pandemic variant common ancestor date occurring years prior to pandemic emergence indicates either that the important characteristics for pandemic spread were acquired years before such spread occurred or that such changes were acquired convergently following diversification into multiple lineages. The extent of diversification prior to pandemic emergence argues strongly against the latter scenario. Not only would important changes have to occur in a large number of individual lineages located on multiple continents, but these changes would have to occur approximately simultaneously after a delay of multiple years. We therefore hypothesized that the key characteristics for pandemic spread are acquired by the variant common ancestor years prior to pandemic emergence. To test this hypothesis, we used a surrogate neutralization assay (antibody ‘blockade’ of ligand binding) to investigate the antigenic properties of two reconstructed Sydney 2012 ancestors: Sydney^{Anc}_{All}, the common ancestor of all sequences genotyped as Sydney 2012 (estimated date: late 2003, 95% HPD early 2000–early 2007) and Sydney^{Anc}_{Pand}, the common ancestor of all pandemic Sydney 2012 viruses (estimated date: late 2008, 95% HPD

late 2006–early 2010) (Fig. 3A). The blockade assay measures the ability of polyclonal sera or monoclonal antibodies to block interaction with HBGA attachment factors (Harrington et al. 2002, Lindesmith et al. 2019). While this assay therefore only measures a subset of the total neutralizing antibody repertoire, blockade in this assay correlates with protection from infection (Reeck et al. 2010, Bok et al. 2011) and virus neutralization in cell culture systems (Alvarado et al. 2018; Costantini et al. 2018). Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} exhibit similar or greater resistance to anti-New Orleans 2009 human polyclonal sera (Fig. 3B), mouse polyclonal sera (Fig. 3C), and mouse monoclonal antibodies (mAbs, Supplementary Fig. S5) compared with a reference Sydney 2012 virus (Sydney^{Ref}) collected during the pandemic. These data indicate that substitutions in VP1 acquired prior to Sydney^{Anc}_{All} provided resistance to the anti-New Orleans 2009 antibody response at least 9 years prior to the onset of the Sydney 2012 pandemic and 6 years prior to the pandemic emergence of New Orleans 2009.

We identified six sites in the antigenic P2 domain that are shared by the three Sydney 2012 VLPs but differ from the New Orleans 2009 VLP (Fig. 3D). These sites all reside within known epitopes: A (site 294), C (sites 341, 377), D (site 396), E (site 413), and G (site 359) (Lindesmith et al. 2012; Tohma et al. 2019). It is therefore likely that the substitution at one or more of these sites is responsible for the observed antigenic differences (Fig. 3B and C, Supplementary Fig. S5). An additional eight amino acid substitutions occurred in VP1 between Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} (Fig. 3A), of which sites 310 and 368 remain highly conserved within Sydney 2012 (Supplementary Fig. S6), indicating that their acquisition by Sydney^{Anc}_{Pand} may have been important for its subsequent emergence as a new pandemic. Site 368 is located within epitope A and was previously demonstrated to alter recognition of mAbs raised against New Orleans 2009 (Debbink et al. 2013) while site 310 is located within the NERK motif that regulates particle breathing and antibody

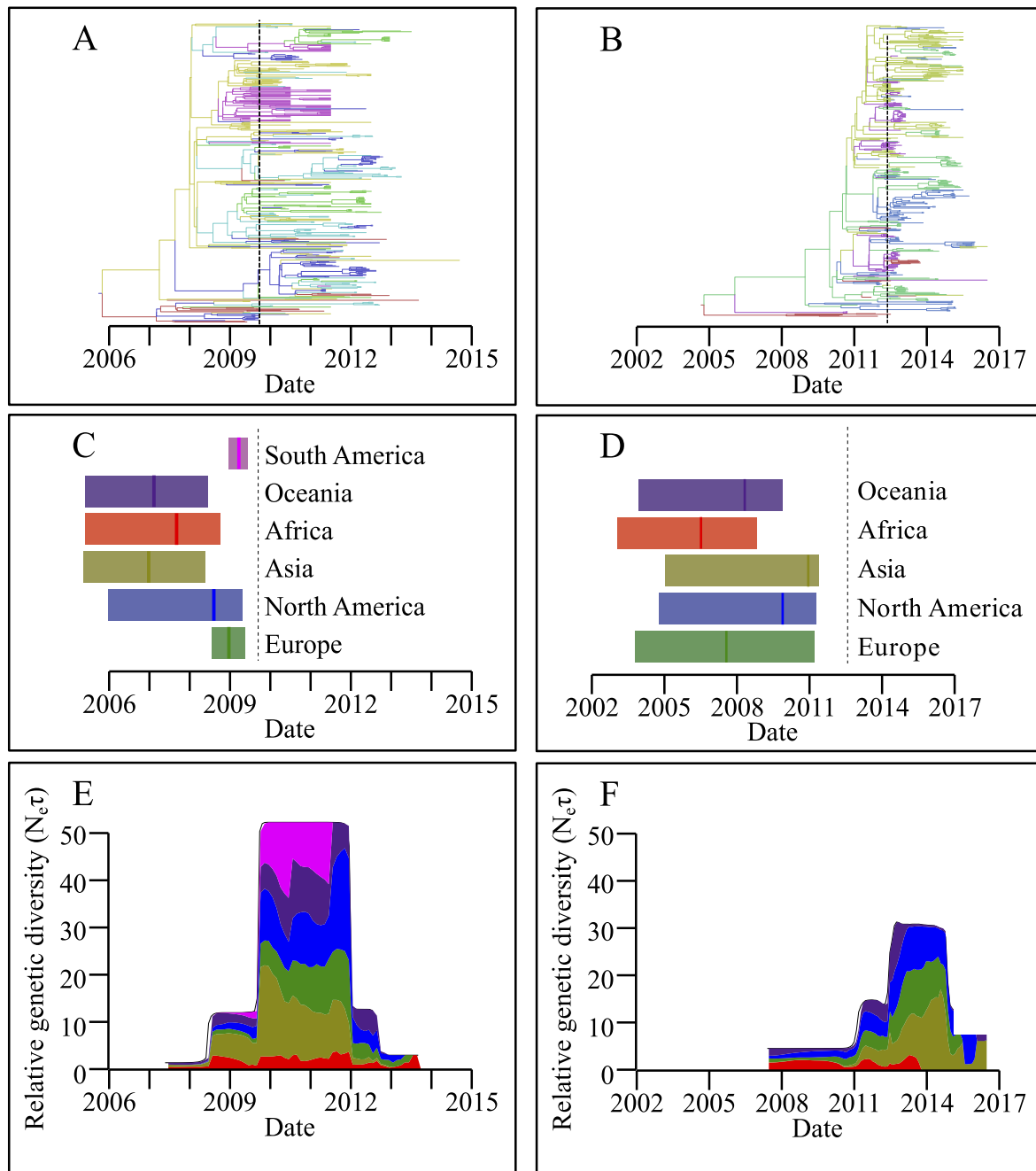


Figure 2. GII.4 variants New Orleans 2009 and Sydney 2012 diversified and spread widely prior to pandemic emergence. (A and B) Spatiotemporally resolved MCC trees for New Orleans 2009 (A) and Sydney 2012 (B) with each branch colored by inferred location, as in panels C and D. (C and D) Summary of continent import dates for New Orleans 2009 (C) and Sydney 2012 (D); the vertical line is the median import date and the shaded area the 95 per cent HPD. The dashed vertical black line is the inferred date of pandemic emergence. (E and F) Summary of the spatiotemporal distribution of lineages from New Orleans 2009 (E) and Sydney 2012 (F). The proportion of lineages on each continent through time is plotted as a stacked area plot, scaled to the estimated relative genetic diversity.

access to epitopes (Lindesmith et al. 2014, 2018). The acquisition of these substitutions between Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} may have been important to enable pandemic emergence by further altering antigenicity or by increasing transmissibility, receptor binding, particle stability or other properties. However, these changes were acquired by 2008 and were therefore not proximal

to the rapid pandemic emergence in 2012. While these substitutions may have influenced viral fitness, the serological assays (Fig. 3B and C, Supplementary Fig. S5) indicate that not only were the critical antigenic properties for pandemic emergence acquired by 2003 but these properties were also maintained in Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand}.

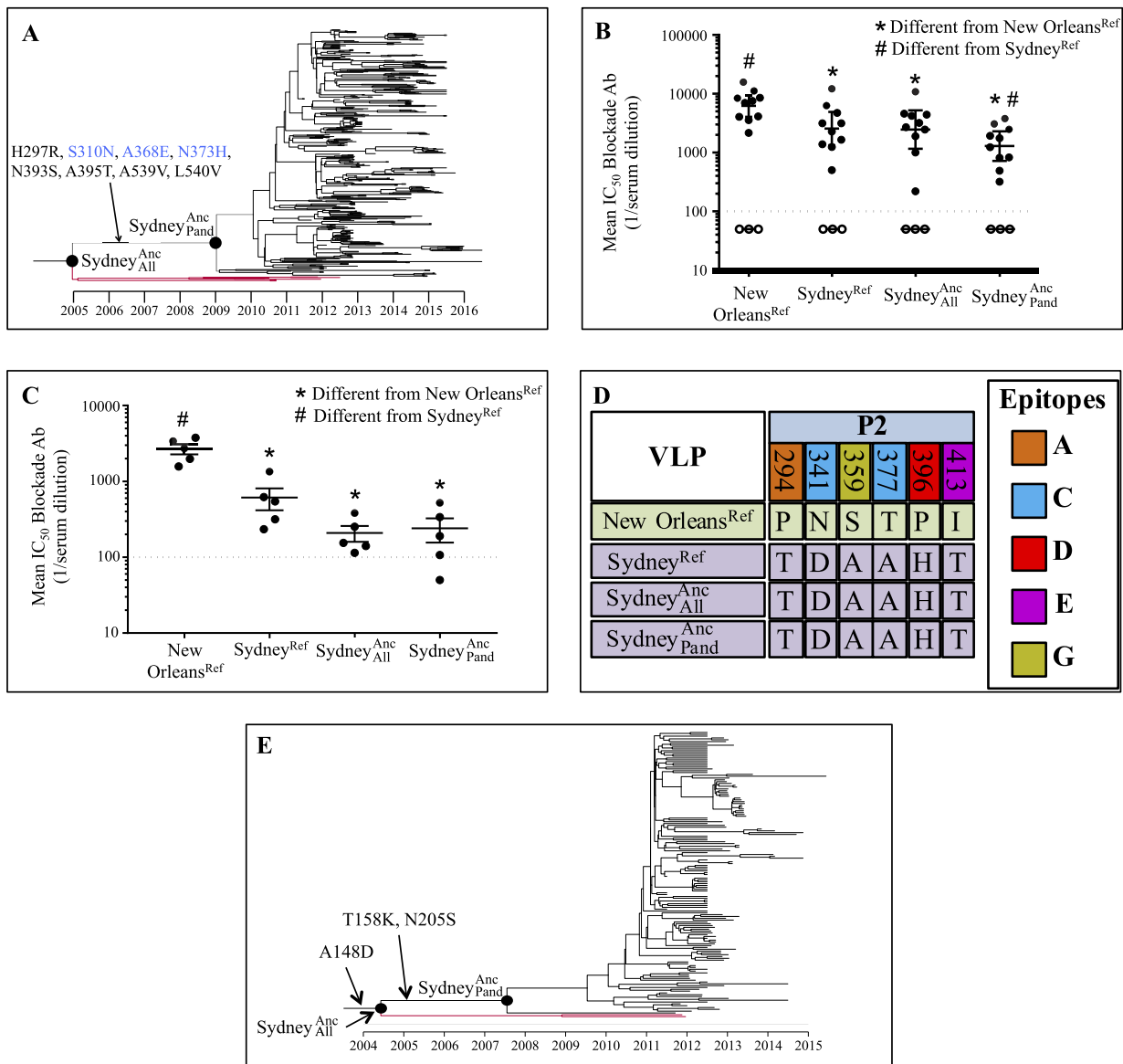


Figure 3. Sydney 2012 could resist anti-New Orleans 2009 immunity by 2003. (A) Temporally resolved Sydney 2012 tree with Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} labeled. The lineages that diverged between Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} (shown in red) did not persist in the population. Nonsynonymous substitutions that occurred leading to Sydney^{Anc}_{Pand} are labeled. Substitutions labeled in blue remained highly conserved in Sydney 2012 (Supplementary Fig. S6). (B) Blockade of Sydney^{Anc}_{All}, Sydney^{Anc}_{Pand}, New Orleans^{Ref}, and Sydney^{Ref} interaction with PGM by polyclonal sera from patients infected with New Orleans 2009 (closed circles) or healthy blood donors (open circles, did not block VLPs at the assay limit of detection). Bars are geometric mean values with 95 per cent confidence intervals. Dashed line is assay limit of detection. Statistical significance calculated using the Wilcoxon test. (C) As in B but using polyclonal sera collected from mice exposed to New Orleans 2009. (D) Six amino acid sites in the antigenic VP1 P2 subdomain exhibit a different residue in all three Sydney 2012 VLPs compared with New Orleans 2009, all of which are located within putative epitopes. (E) Temporally resolved Sydney 2012 VP2 phylogeny with Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} labeled. Nonsynonymous substitutions leading to Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} are labeled.

We next examined the potential influence of other genomic regions on the pandemic emergence of Sydney 2012. It is unlikely that the nonstructural polyprotein drove Sydney 2012 pandemic spread, as this variant co-circulated with the nonstructural polyprotein from the unrelated GII.P31, GII.P4, and (more recently) GII.P16 genotypes (Supplementary Fig. S7) (Wong et al. 2013; Ruis et al. 2017). Substitutions occurred within VP2 leading to Sydney^{Anc}_{All} (site 148) and Sydney^{Anc}_{Pand} (sites 158, 205) and remained highly conserved within the Sydney 2012

clade (Fig. 3E and Supplementary Fig. S6). Without better structural or functional characterization of the VP2 protein, interpreting the contribution of these changes is difficult. However, the Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} VP2 proteins occurred in mid-2004 (95% HPD mid-1999–early 2009) and early 2008 (95% HPD mid-2005–early 2010), respectively (Fig. 3E). Therefore, as with VP1, if these substitutions were important for pandemic emergence, they were acquired years earlier and were not therefore the proximate driver of pandemic emergence.

A similar process occurred for the other four pandemic GII.4 variants to have emerged since 2002, where in each case, key nonsynonymous substitutions in the nonstructural polyprotein, VP1 and VP2 that characterize the new pandemic variant occurred along the branch leading to the respective common ancestor long before these variants spread pandemically (Supplementary Tables S5 and S6). The fact that the VP1 substitutions included sites within known blockade epitopes (Lindesmith et al. 2012, Tohma et al. 2019) provides further support for the hypothesis that the antigenic changes required for pandemic emergence were acquired years before pandemic emergence actually occurred.

While recombination has previously been suggested to drive GII.4 pandemics (Eden et al. 2013), we find that each recombination event in which a variant acquired a new nonstructural polyprotein or VP2 occurred years prior to pandemic emergence (Supplementary Table S7 and text). These results indicate that recombination events are not the proximate drivers of new norovirus pandemics.

Together, our results indicate that pandemic GII.4 variants arise, diversify and spread widely years before they emerge to cause a pandemic. The onset of a new GII.4 pandemic is well-defined epidemiologically, with the new pandemic variant rapidly increasing in prevalence to dominate outbreaks globally (Fig. 1B, White 2014), as reflected in pandemic variant Bayesian skyline plots (Supplementary Fig. S4). Our results therefore raise the question of what drives a variant that has been circulating widely and cryptically for years to suddenly increase in frequency, dominate outbreaks worldwide and rapidly replace the preceding pandemic variant. If a new pandemic was triggered by one or more viral genetic changes altering antigenicity or another viral characteristic, a single lineage would rapidly increase in prevalence, as is observed for influenza A H3N2 (Bedford et al. 2015). Our data instead strongly support a scenario where the key antigenic and other changes are acquired through substitutions and/or recombination events years before pandemic emergence. Therefore, a pandemic event is not proximally driven by genetic changes in any of the genomic regions altering antigenicity, receptor binding or another property. It is, however, highly likely that virus genetic factors are important for pandemic expansion as multiple GII.4 variants were present at the onset of each GII.4 pandemic (Fig. 1A), but only one variant underwent pandemic spread. There must therefore be a mechanism that interacts with one or more virus genetic factors to determine which of the present variants is selected to spread pandemically. This factor needs to change through time to enable the expansion of GII.4 variants that have been present for years. It is difficult to conceive of any factor that fulfills these criteria except a host factor and the host factor that most clearly changes rapidly through time, can interact with virus genetic factors and can explain both the disappearance of preceding pandemic variants and the rapid expansion of a new variant is host immunity. We therefore suggest that GII.4 pandemics are driven by an interplay between pre-circulating antigenic diversity in GII.4 variants and temporal changes in host immunity.

There is substantial evidence of antigenic differences between pandemic GII.4 variants (Lindesmith, Donaldson, and Baric 2011; Lindesmith et al. 2012; Debbink et al. 2013) and each newly emerging pandemic GII.4 variant is antigenically distinct from all preceding variants. While there have only been limited studies of re-infection with GII.4 noroviruses, re-infection with consecutive pandemic variants has been demonstrated (Saito et al., 2014) and re-infection with the same pandemic variant occurs less frequently than expected (Sakon et al., 2015). This

supports antigenic differences between GII.4 variants being important for re-infection and therefore for substantial onward transmission and pandemic expansion.

Recent studies examining the host serological repertoire have demonstrated generation of both homotypic and heterotypic immunity upon GII.4 vaccination/infection in adults (Lindesmith et al. 2015, 2019). This homotypic immunity against variant-specific epitopes likely lasts at least several years (Lindesmith et al. 2012; Bernstein et al., 2015), while heterotypic immunity against conserved GII.4 epitopes wanes more rapidly (Lindesmith et al. 2015). Therefore, infection of a large number of individuals during a pandemic will result in significant population homotypic immunity against the pandemic variant, causing this variant to decline. We hypothesize that heterotypic immunity in adults prevents low-level GII.4 variants from expanding during a pandemic. However, as the dominant variant declines, heterotypic immunity will also begin to decline (Lindesmith et al. 2015), allowing a new antigenically distinct GII.4 variant to expand in the adult population. The generation of long-lived variant-specific antibody clonotypes (Lindesmith et al. 2015, 2019) will prevent previous dominant variants from re-emerging in the future. Therefore, while virus genetic changes altering antigenicity are essential to enable pandemic spread, pandemic onset is proximally driven by changes in host population immunity. Additional factors, such as competition between variants for a replication niche and stochastic transmission, may also influence variant emergence.

These results raise the question of where pre-pandemic variants circulate over the years prior to emergence. Both immunocompromised patients and animals have been mooted as a potential source of pandemic GII.4 variants (Debbink et al. 2014a; Karst and Baric 2015, Kocher et al. 2018). While it is possible that either of these hosts could be the source of the ancestral variant, both seem unlikely to be the source of the diversifying pre-pandemic lineage. It is unlikely that multiple pandemic lineages could emerge simultaneously from a single immunocompromised host at pandemic onset. It is also difficult to explain how multiple immunocompromised patients could form the inter-connected intercontinental transmission network required for this pre-diversification, supporting recent suggestions that immunocompromised patients are an unlikely reservoir (Eden et al. 2017). In addition, while GII.4 viruses have occasionally been detected in stool samples from cows, pigs, and dogs (Mattison et al. 2007; Summa, von Bonsdorff, and Maunula 2012), concurrent emergence of multiple lineages would require multiple zoonotic transmissions and no such transmissions have been observed (Wilhelm et al. 2015). A more parsimonious explanation for our findings is that pandemic GII.4 variants circulate within the community and are not detected by current surveillance efforts that largely target outbreaks, predominantly in hospital and institutional settings comprised adults (Inns et al. 2017). More extensive co-circulation of viral lineages has been noted in influenza A H1N1 and influenza B compared with influenza A H3N2 and has been correlated with slower rates of antigenic drift and a lower average age of infection (Bedford et al. 2015; Vijaykrishna et al. 2015), suggesting that pre-pandemic GII.4 lineages might circulate within children. The broad heterotypic immunity generated upon infection of adults does not appear to be generated upon infection in children (Malm et al. 2014), potentially because multiple infections are required to generate a broadly neutralizing serological repertoire (Lindesmith et al. 2019). We hypothesize that this lack of heterotypic immunity allows a wide diversity of GII.4 variants to circulate and persist in children. In addition,

while noroviruses are prevalent in individuals of all age groups, the infection rate is highest in young children (Lopman et al. 2016; O'Brien et al. 2016). In support of pre-pandemic variants initially circulating in children, fifteen of the sixteen identified pre-pandemic samples for which age information is available are either from children or were sampled in a nursery or primary school (Supplementary Fig. S3 and Table S3). The ability of the ancestral Sydney 2012 VLPs to evade anti-New Orleans 2009 polyclonal sera raised in mice that have not previously been exposed to any other human norovirus (Fig. 3C) suggests that these viruses would have been able to evade homotypic immunity raised against other variants in young children. Thus, while continued strain monitoring, as captured by norovirus outbreak surveillance networks such as CaliciNet (Vega et al. 2011) and NoroNet (van Beek et al. 2018), has significant value for vaccine development, additional efforts should focus on identifying potential reservoirs from which future pandemic norovirus variants could emerge, including symptomatic and asymptomatic children (Rouhani et al. 2016) in healthcare and community settings. A natural corollary of our proposed model is that future pandemic GII.4 variants are continuing to circulate and diversify undetected within the reservoir until changes in host immunity favor the emergence of a new pandemic variant. Therefore, as our results indicate that key antigenic changes are acquired years before pandemic expansion, combining an efficient surveillance system with antigenic screening will enable detection of low-level antigenically novel variants that may cause future pandemics, as is currently carried out with influenza virus (Li et al. 2016).

Our results have also important implications for current efforts to develop norovirus vaccines. Should the new pandemic variant emerge from the preceding variant, a vaccine targeting the current variant may prevent the next pandemic. However, under our hypothesis, a vaccine that boosts immunity to the current variant may actually hasten emergence of the next pandemic. It is therefore essential that norovirus vaccines provide broad immunity against GII.4 viruses (Lindesmith et al. 2015, 2019). Combining comprehensive surveillance with antigenic testing will enable identification of low-level antigenically distinct variants that might require vaccine reformulation. Future studies characterizing the antigenic and serological diversity through time will be essential to understand why specific GII.4 variants caused pandemics at specific points in time. This will enable robust vaccine reformulation and decisions on which low-level variants to include in such reformulations, should multiple antigenically distinct variants be present.

3. Materials and methods

3.1 Reconstruction of the temporal history of GII.4 norovirus

We collected all norovirus sequences available on GenBank as of 30 October 2015 containing the complete RNA-dependent RNA polymerase (RdRp), VP1, and VP2 genome regions. Each sequence was genotyped using the norovirus genotyping tool (Kroneman et al. 2011) and the 871 sequences containing the GII.4 VP1 were retained. Due to inter-genotype recombination events close to the ORF1–ORF2 boundary (Eden et al. 2013), the dataset contains sequences with the GII.P1, GII.P4, GII.P12, and GII.P31 RdRps.

Due to the presence of recombination close to ORF boundaries, we ran all analyses on the RdRp, VP1, and VP2 separately. Each genome region was aligned at the amino acid level using

MUSCLE (Edgar 2004). We screened for the presence of intra-genetic recombination in the RdRp, VP1, and VP2 separately using the Single Breakpoint (SBP) method implemented in HyPhy (Pond et al. 2006), identifying nineteen sequences as potential recombinants in VP1 (Supplementary Table S8). These sequences clustered differently with strong support on either side of the putative breakpoint in phylogenetic trees reconstructed with RAxML v8.1 (Stamatakis 2014) with the GTR model and gamma rate heterogeneity with four gamma classes. SBP was then run on the alignment again following removal of these samples to ensure the recombination signal had been removed. A summary of the number of remaining sequences from each GII.4 variant and each RdRp genotype is shown in Supplementary Table S9. The variant names used here are those returned by the norovirus genotyping tool (Kroneman et al. 2011).

Methods employing sequence data and sampling dates to infer divergence times are valid only if there is a temporal evolutionary signal in the dataset (Rambaut et al. 2016). To assess whether each of our datasets exhibits a temporal evolutionary signal, we reconstructed a maximum likelihood tree using RAxML, as above. We identified the best-fitting root position using TempEst v1.5 (Rambaut et al. 2016) and calculated the R^2 correlation between root-to-tip distance and sampling date (Supplementary Fig. S8). There was a significant temporal evolutionary signal within each dataset ($P < 0.001$).

We reconstructed the temporal evolutionary history of each genomic region using the Bayesian Markov chain Monte Carlo approach implemented in BEAST version 2.2.1 (Bouckaert et al. 2014). Analyses were run independently on the RdRp, VP1, and VP2. As the Den Haag 2006 and New Orleans 2009 variants have a greater number of sequences compared with other variants (Supplementary Table S9), we took three random subsamples of forty-one sequences from each of these variants (with forty-one chosen to match the number of sequences in the third most numerous variant); results were insensitive to the subsampled dataset (Supplementary Table S10). Each sequence was labeled with the most accurate collection date possible: the day of collection if available, the month of collection if the day of collection was not available or the year of collection if the month of collection was not available. All sample collection dates were checked in the relevant literature and in cases where a more precise collection date was present in the literature, we used the more precise date. Each dataset was analyzed using the GTR substitution model with gamma rate heterogeneity and partitioned so codon positions 1 and 2 shared a substitution model and codon position 3 had a different substitution model. We employed both the strict and relaxed lognormal clock models to examine variation in the substitution rate within each dataset. We used a lognormal prior distribution for each dataset with mean 4.3×10^{-3} substitutions/site/year and standard deviation (SD) 0.1 for the VP1 (Bok et al. 2009) and with mean 4.32×10^{-3} substitutions/site/year and SD 0.1 for the RdRp (Siebenga et al. 2010). At the time of our analysis, there was no previously published substitution rate for VP2 across the GII.4 clade. We therefore employed the same prior as for the VP1 dataset. We applied a coalescent Bayesian skyline tree prior. Three replicate runs with different starting values were performed for each dataset and clock model and run until convergence, as assessed using Tracer v1.6 (Rambaut et al. 2018). The replicate runs were combined with removal of suitable burnin using LogCombiner v2.2.1 and maximum clade credibility (MCC) trees were obtained using TreeAnnotator v2.2.1. In each case there was strong support to reject the strict clock model in favor of the relaxed lognormal

clock model. Therefore, the results employing the relaxed lognormal clock model were used for all further analyses.

We estimated variant divergence dates and recombination dates by combining the posterior distribution of trees from each subsampled dataset into a single posterior distribution. We inferred variant divergence dates by calculating the date of the most recent common ancestor between each pair of variants in each tree in this posterior distribution. To calculate the date of each recombination event, we identified the mean and 95 per cent HPD of the distribution of branch start and end times of the corresponding branch in each tree in the posterior distribution.

3.2 Identification of pre-pandemic and pre-epidemic GII.4 sequences

We defined pre-pandemic/pre-epidemic sequences as those that cluster with a GII.4 variant but were collected prior to the year in which that variant emerged pandemically or epidemically. We genotyped all GII.4 VP1 sequences present on GenBank as of 9 February 2017 containing more than 400 nucleotides (Kroneman et al. 2011). We identified fifty sequences with a reported collection date earlier than the year of pandemic/epidemic emergence of the respective variant. We estimated the collection date of each of these sequences using BEAST v2.4.2 (Bouckaert et al. 2014), assuming a uniform prior distribution with minimum 1974.5 and maximum 2015.446575 (the dates of the earliest and latest collected sequences in the main dataset). The 95 per cent HPD of the estimated collection date overlapping with the reported collection date was taken as evidence to support, but not confirm, the reported collection date.

3.3 Reconstruction of the temporal history of New Orleans 2009 and Sydney 2012

We compiled expanded datasets from GenBank on 3 February 2016 containing 460 and 533 P2 domain sequences for New Orleans 2009 and Sydney 2012, respectively. Each dataset was aligned and inferred to exhibit a temporal evolutionary signal ($P < 0.001$) as above. We reconstructed the evolutionary dynamics of New Orleans 2009 and Sydney 2012 independently using BEAST v2.2.1 (Bouckaert et al. 2014), using the HKY nucleotide substitution model with four gamma classes. We employed a lognormal prior on the substitution rate with mean 6.83×10^{-3} and SD 0.1 to accommodate the mean and 95 per cent HPD of our estimate of the substitution rate across the complete GII.4 VP1 clade. We applied the strict and relaxed lognormal clock models and found strong support (log10 Bayes factor > 100) to reject the strict clock model in favor of the relaxed lognormal clock model in each case. We therefore used the results from the relaxed lognormal clock runs in all further analyses. Bayesian skyline plots were reconstructed using Tracer v1.5. The date at which the Bayesian skyline plot exhibits a large increase in relative genetic diversity was used as the time of pandemic onset. The number of lineages present at the onset of the pandemic was calculated as the number of lineages present at this point in time in each tree in the posterior distribution.

3.4 Reconstruction of the spatiotemporal history of New Orleans 2009 and Sydney 2012

We collected datasets containing all available VP1 sequences on GenBank as of 9 February 2017 from New Orleans 2009 ($n = 565$) and Sydney 2012 ($n = 708$). Each dataset exhibited temporal

signal as above. Examination of the sampling locations showed that there was typically only a small number of sequences from each country (Supplementary Table S11). We therefore used the continent of collection as the location label. The New Orleans 2009 dataset contained a large number of sequences from Asia and Oceania relative to the other continents, while the Sydney 2012 dataset contained a large number of sequences from Asia relative to the other continents (Supplementary Table S11). Should sequences from the same continent cluster together within the tree, an excess of sequences from one continent is unlikely to alter estimates of ancestral locations. However, if sequences from the different continents are typically interspersed within the tree, an excess of sequences from one continent could result in artifactual support for this continent being the location of ancestral nodes. The sequences from each continent are interspersed throughout the tree in both New Orleans 2009 and Sydney 2012 (Supplementary Fig. S9). We therefore randomly down-sampled New Orleans 2009 sequences from Asia and Oceania and Sydney 2012 sequences from Asia to match the number of sequences from the next most commonly represented continent. We carried out three random subsamples and performed all analyses on each subsampled dataset. All results were insensitive to the subsampled dataset.

We used discrete phylogeography (Lemey et al. 2009) implemented in BEAST v2.4.2 (Bouckaert et al. 2014) to reconstruct the spatiotemporal history of New Orleans 2009 and Sydney 2012. Sequences were labeled with the most accurate collection date possible, as described above. We modeled the nucleotide substitution process using the HKY substitution model and gamma rate heterogeneity with four gamma classes. We applied both the strict and relaxed lognormal clock models. In each case there was strong support to reject the strict clock model in favor of the relaxed lognormal clock model (log10 Bayes factor 66–83) and we therefore used the relaxed lognormal clock model for our inferences. However, the results with the strict clock model are qualitatively very similar, indicating that potential over-parameterization due to the large number of branch-specific rates with the relaxed lognormal clock model has not influenced our results. We applied a lognormal prior on the substitution rate with mean 7×10^{-3} for New Orleans 2009 and 6.4×10^{-3} for Sydney 2012 and SD 0.1 in each case, based on the posterior estimates of the variant substitution rates inferred previously. We employed a Bayesian coalescent skyline tree prior for each dataset. We applied a discrete phylogeographic model to describe lineage migrations within each dataset (Lemey et al. 2009), consisting of a symmetric transition matrix for the migration rates and a set of Bayesian stochastic search variable selection (BSSVS) indicator variables. We assumed a Poisson prior for the number of 'on' BSSVS variables with mean 5 for the New Orleans 2009 datasets and mean 4 for the Sydney 2012 datasets to incorporate a broad range on the number of 'on' transition rates. We used an exponential prior with mean 1.0 migration rate per lineage per year for the overall rate of geographical transition to incorporate a broad range of reasonable transition rates while placing a high prior probability on a high rate of inter-continental transmission, as suggested by the high level of interspersed sequences from different continents. We employed a gamma prior with shape 1.0 and scale 1.0 for each of the relative geographical transition rates. Three replicate runs with different starting parameters were carried out with each dataset and run until convergence, as assessed using Tracer v1.6 (Rambaut et al. 2018). Runs from each subsampled dataset were combined into a single posterior distribution for downstream analyses.

We identified the date of first import into each continent by calculating either the root date if the continent was inferred to be the root location or the earliest branch midpoint where the downstream node was inferred to be within the continent. We calculated this date within each tree in the posterior distribution. We therefore assume that migration events occurred at the midpoint of the branch. We obtain similar results using the earliest non-root node inferred to be within the continent, which assumes that migration occurs at the end of the branch. We used the program posterior analysis of coalescent trees (PACT) v0.9.4 to compute the proportion of lineages present on each continent through time.

3.5 Reconstruction of ancestral Sydney 2012 viruses and identification of substitutions leading to each GII.4 variant

We collected a dataset containing all 2198 available GII.4 VP1 sequences as of 9 February 2017, including sequences from all of the major GII.4 variants (Supplementary Table S12). Alignment and phylogenetic reconstruction was carried out as above. We used RAxML to optimize branch lengths within the maximum likelihood phylogenetic tree and ten bootstrap tree topologies using the amino acid alignment and the WAG substitution model with optimized base frequencies. We used multiple tree topologies to assess the influence of tree topology on our inferences. We carried out ancestral reconstruction at the amino acid level with PAML v4.9 (Yang 2007) using the WAG substitution model and optimized base frequencies. The ancestral sequence at Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} was identical with each tree topology and the residue inferred at each site was supported by posterior probability >0.95. We identified VP1 substitutions by comparing the sequence of the variant root ancestor with the sequence of the immediately upstream node in each tree.

To identify substitutions with the nonstructural polyprotein and VP2 we collected datasets containing all available GII and GIV nonstructural polyprotein sequences (Supplementary Table S13) and all available GII.4 VP2 sequences (Supplementary Table S12), respectively. We identified substitutions leading to each GII.4 variant using the same process described for VP1 above.

3.6 Surrogate neutralization assay (antibody blockade assay)

Sydney^{Anc}_{All} and Sydney^{Anc}_{Pand} VP1 genes were codon optimized for mammalian expression and synthesized by Bio Basic, Inc. (Amherst, NY) and VLPs were expressed in baby hamster kidney cells from Venezuelan equine encephalitis virus replicons (11); 0.25 µg/ml VLP was pretreated with decreasing concentrations of antibody/serum for 1 h before addition to pig gastric mucin (PGM) type III (PGM, Sigma Aldrich, St. Louis, MO) coated plates for 1 h. Bound VLP was detected with anti-GII.4 2012 ref rabbit hyperimmune sera. Percent control binding is defined as the binding with antibody/serum pretreatment compared to the binding without multiplied by 100. All incubations were done at room temperature. The blockade data were fit using sigmoidal dose-response curve analysis of nonlinear data in GraphPad Prism 702 and IC50 titers with 95 per cent confidence intervals calculated. Antibodies that did not block 50 per cent of binding at the highest dilution tested were assigned an IC50 of two times the assay limit of detection for statistical comparison (11). Anti-New Orleans 2009 polyclonal sera were collected from ten patients naturally infected during a New Orleans 2009 outbreak

in a long-term care facility in March 2010. Infection was confirmed by symptoms and norovirus detection in acute stool. The human serum samples were collected from a study that was approved by institutional review board of the Centers for Disease Control and Prevention (CDC Protocol no. 5051). Mouse anti-GII.4 New Orleans 2009^{Ref} sera were generated in immunized mice as described in Debbink et al. (2014b). Mouse monoclonal antibodies to GII.4 New Orleans 2009^{Ref} (Lindesmith et al. 2013) were generated by VLP immunization.

3.7 Data availability

All alignments and phylogenetic trees used in this manuscript are available at https://github.com/chrisruis/Norovirus_data. A summary of sequences used is provided in Supplementary Table S14.

Acknowledgments

The authors would like to thank Victoria Madden of Microscopy Services Laboratory, Department of Pathology and Laboratory Medicine, University of North Carolina-Chapel Hill for expert technical support.

Data availability

All alignments, phylogenetic trees and BEAST XML files are available at https://github.com/chrisruis/Norovirus_data. VLPs are available from the authors.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

Funding

This work was supported by the Wellcome Trust (grant number 203268/Z/16/Z to J.B.), a studentship from UCL CoMPLEX (to C.R.), the National Institute for Health Research UCL-UCLH Biomedical Resource Centre, the Medical Research Council (grant number U117573805 to R.A.G.), the National Institutes of Health, Allergy and Infectious Diseases (grant number U19 AI109761 to R.S.B.), the EU FP7 PATHSEEK grant and a Medical Research Council studentship. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

- Alvarado, G. et al. (2018) 'Human Monoclonal Antibodies That Neutralize Pandemic GII.4 Noroviruses', *Gastroenterology*, 155: 1898–907.
- Bartsch, S. M. et al. (2016) 'Global Economic Burden of Norovirus Gastroenteritis', *PLoS One*, 11: e0151219.
- Bedford, T. et al. (2015) 'Global Circulation Patterns of Seasonal Influenza Viruses Vary with Antigenic Drift', *Nature*, 523: 217–20.

- Bernstein, D. I et al. (2015) 'Norovirus Vaccine against Experimental Human GII.4 Virus Illness: A Challenge Study in Healthy Adults', *The Journal of Infectious Diseases*, 211: 870--8.10.1093/infdis/jiu497 25210140
- Bok, K. et al. (2009) 'Evolutionary Dynamics of GII.4 Noroviruses over a 34-year Period', *Journal of Virology*, 83: 11890--901.
- et al. (2011) 'Chimpanzees as an Animal Model for Human Norovirus Infection and Vaccine Development', *Proceedings of the National Academy of Sciences*, 108: 325--30.
- Bouckaert, R et al. (2014) 'BEAST 2: A Software Platform for Bayesian Evolutionary Analysis', *PLoS Computational Biology*, 10: e1003537.
- Chhabra, P. et al. (2019) 'Updated Classification of Norovirus Genogroups and Genotypes', *Journal of General Virology*, 100: 1393--406.
- Costantini, V. et al. (2018) 'Human Norovirus Replication in Human Intestinal Enteroids as Model to Evaluate Virus Inactivation', *Emerging Infectious Diseases Journal*, 24: 8.
- Debbink, K. et al. (2013) 'Emergence of New Pandemic GII.4 Sydney Norovirus Strain Correlates with Escape from Herd Immunity', *The Journal of Infectious Diseases*, 208: 1877--87.
- Debbink, K et al. (2014) 'Chimeric GII.4 Norovirus Virus-LIKE-Particle-Based Vaccines Induce Broadly Blocking Immune Responses', *Journal of Virology*, 88: 7256--66.
- Debbink, K. et al. (2014b) 'Within-host Evolution Results in Antigenically Distinct GII.4 Noroviruses', *Journal of Virology*, 88: 7244--55.
- Eden, J.-S. et al. (2013) 'Recombination within the Pandemic Norovirus GII.4 Lineage', *Journal of Virology*, 87: 6270--82.
- et al. (2014) 'The Emergence and Evolution of the Novel Epidemic Norovirus GII.4 Variant Sydney 2012', *Virology*, 450--451: 106--13.
- et al. (2017) 'Persistent Infections in Immunocompromised Hosts are Rarely Sources of New Pathogen Variants', *Virus Evolution*, 3: 219--22.
- Edgar, R. C. (2004) 'MUSCLE: multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792--7.
- Harrington, P. R. et al. (2002) 'Binding of Norwalk Virus-like Particles to ABH Histo-blood Group Antigens is Blocked by Antisera from Infected Human Volunteers or Experimentally Vaccinated Mice', *Journal of Virology*, 76: 12335--43.
- Inns, T. et al. (2017) 'Community-based Surveillance of Norovirus Disease: A Systematic Review', *BMC Infectious Diseases*, 17: 657.
- Karst, S. M., and Baric, R. S. (2015) 'What is the Reservoir of Emergent Human Norovirus Strains?', *Journal of Virology*, 89: 5756--9.
- Kirk, M. D. et al. (2015) 'World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis', *PLoS Medicine*, 12: e1001940.
- Kocher, J. F. et al. (2018) 'Bat Caliciviruses and Human Noroviruses are Antigenically Similar and have Overlapping Histo-blood Group Antigen Binding Profiles', *mBio*, 9:e00869--18
- Kroneman, A. et al. (2008) 'Analysis of Integrated Virological and Epidemiological Reports of Norovirus Outbreaks Collected within the Foodborne Viruses in Europe Network from 1 July 2001 to 30 June 2006', *Journal of Clinical Microbiology*, 46: 2959--65.
- Kroneman, A et al. (2011) 'An Automated Genotyping Tool for Enteroviruses and Noroviruses', *Journal of Clinical Virology*, 51: 121--5.
- Lemey, P et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.
- Li, C. et al. (2016) 'Selection of Antigenically Advanced Variants of Seasonal Influenza Viruses', *Nature Microbiology*, 1: 1--10.
- Lindesmith, L. C. et al. (2008) 'Mechanisms of GII.4 Norovirus Persistence in Human Populations', *PLoS Medicine*, 5: e31.
- , Donaldson, E. F., and Baric, R. S. (2011) 'Norovirus GII.4 Strain Antigenic Variation', *Journal of Virology*, 85: 231--42.
- et al. (2012) 'Immunogenetic Mechanisms Driving Norovirus GII.4 Antigenic Variation', *PLoS Pathogens*, 8: e1002705.
- et al. (2013) 'Emergence of a Norovirus GII.4 Strain Correlates with Changes in Evolving Blockade Epitopes', *Journal of Virology*, 87: 2803--13.
- et al. (2014) 'Particle Conformation Regulates Antibody Access to a Conserved GII.4 Norovirus Blockade Epitope', *Journal of Virology*, 88: 8826--42.
- et al. (2015) 'Broad Blockade Antibody Responses in Human Volunteers after Immunization with a Multivalent Norovirus VLP Candidate Vaccine: Immunological Analyses from a Phase I Clinical Trial', *PLoS Medicine*, 12: e1001807.
- et al. (2018) 'Conformational Occlusion of Blockade Antibody Epitopes, a Novel Mechanism of GII.4 Human Norovirus Immune Evasion', *mSphere*, 3:e00518--17
- et al. (2019) 'Sera Antibody Repertoire Analyses Reveal Mechanisms of Broad and Pandemic Strain Neutralizing Responses after Human Norovirus Vaccination', *Immunity*, 50: 1530--41.e8.
- Lopman, B. A. et al. (2016) 'The Vast and Varied Global Burden of Norovirus: Prospects for Prevention and Control', *PLoS Medicine*, 13: e1001999.
- Malm, M. et al. (2014) 'High Serum Levels of Norovirus Genotype-specific Blocking Antibodies Correlate with Protection from Infection in Children', *Journal of Infectious Diseases*, 210: 1755--62.
- Mattison, K. et al. (2007) 'Human Noroviruses in Swine and Cattle', *Emerging Infectious Diseases*, 13: 1184--8.
- Motoya, T. et al. (2017) 'Molecular Evolution of the VP1 Gene in Human Norovirus GII.4 Variants in 1974--2015', *Frontiers in Microbiology*, 8:2399.
- O'Brien, S. J. et al. (2016) 'Age-specific Incidence Rates for Norovirus in the Community and Presenting to Primary Healthcare Facilities in the United Kingdom', *Journal of Infectious Diseases*, 213: S15--8.
- Parra, G. I. et al. (2017) 'Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity', *PLoS Pathogens*, 13: e1006136.
- Pires, S. M. et al. (2015) 'Aetiology-specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food', *PLoS One*, 10: e0142927.
- Pond, S. L. K et al. (2006) 'Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm', *Molecular Biology and Evolution*, 23: 1891.
- Rambaut, A. et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.
- et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901--4.
- Reeck, A. et al. (2010) 'Serological Correlate of Protection against Norovirus-induced Gastroenteritis', *The Journal of Infectious Diseases*, 202: 1212--8.
- Rouhani, S. et al. (2016) 'Norovirus Infection and Acquired Immunity in 8 Countries: Results from the MAL-ED Study', *Clinical Infectious Diseases*, 62: 1210--7.
- Ruis, C. et al. (2017) 'The Emerging GII.16-GII.4 Sydney 2012 Norovirus Lineage is Circulating Worldwide, Arose by

- Late-2014 and Contains Polymerase Changes That May Increase Virus Transmission', *PLoS One*, 12: e0179572.
- Saito, M et al. (2014) 'Multiple Norovirus Infections in a Birth Cohort in a Peruvian Periurban Community', *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 58: 483–91.
- Sakon, N et al. (2015) 'Impact of Genotype-Specific Herd Immunity on the Circulatory Dynamism of Norovirus: A 10-Year Longitudinal Study of Viral Acute Gastroenteritis', *The Journal of Infectious Diseases*, 211: 879–88.
- Sdiri-Loulizi, K. et al. (2009) 'Molecular Epidemiology of Norovirus Gastroenteritis Investigated Using Samples Collected from Children in Tunisia during a Four-year Period: Detection of the Norovirus Variant GGII.4 Hunter as Early as January 2003', *Journal of Clinical Microbiology*, 47: 421–9.
- Siebenga, J. J. et al. (2007) 'Epochal Evolution of GGII.4 Norovirus Capsid Proteins from 1995 to 2006', *Journal of Virology*, 81: 9932–41.
- Siebenga, J. J. et al. (2009) 'Norovirus Illness is a Global Problem: Emergence and Spread of Norovirus GII.4 Variants, 2001–2007', *The Journal of Infectious Diseases*, 200: 802–12.
- et al. (2010) 'Phylogenetic Reconstruction Reveals Norovirus GII.4 Epidemic Expansions and Their Molecular Determinants', *PLoS Pathogens*, 6: e1000884.
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies', *Bioinformatics*, 30: 1312–3.
- Summa, M., von Bonsdorff, C. H., and Maunula, L. (2012) 'Pet Dogs—A Transmission Route for Human Noroviruses?', *Journal of Clinical Virology*, 53: 244–7.
- Tohma, K. et al. (2019) 'Population Genomics of GII.4 Noroviruses Reveal Complex Diversification and New Antigenic Sites Involved in the Emergence of Pandemic Strains', *mBio*, 10: e02202–19
- van Beek, J. et al. (2018) 'Molecular Surveillance of Norovirus, 2005–16: An Epidemiological Analysis of Data Collected from the NoroNet Network', *Lancet Infectious Diseases*.18545.
- Vega, E. et al. (2011) 'Novel Surveillance Network for Norovirus Gastroenteritis Outbreaks, United States', *Emerging Infectious Diseases*, 17: 1389–95.
- Vijaykrishna, D. et al. (2015) 'The Contrasting Phylodynamics of Human Influenza B Viruses', *eLife*, 4: e05055.
- Vinje, J. (2015) 'Advances in Laboratory Methods for Detection and Typing of Norovirus', *Journal of Clinical Microbiology*, 53: 373–81.
- White, P. A. (2014) 'Evolution of Norovirus', *Clinical Microbiology and Infection*, 20: 741–5.
- Wilhelm, B. et al. (2015) 'A Scoping Review of the Evidence for Public Health Risks of Three Emerging Potentially Zoonotic Viruses: Hepatitis E Virus, Norovirus, and Rotavirus', *Preventive Veterinary Medicine*, 119: 61–79.
- Wong, T. H. N. et al. (2013) 'Whole Genome Sequencing and De Novo Assembly Identifies Sydney-like Variant Noroviruses and Recombinants during the Winter 2012/2013 Outbreak in England', *Virology Journal*, 10: 335.
- Yang, Z. (2007) 'PAML 4: Phylogenetic Analysis by Maximum Likelihood', *Molecular Biology and Evolution*, 24: 1586–91.