

Short Communication

The recombinant origin of emerging human norovirus GII.4/2008: intra-genotypic exchange of the capsid P2 domain

Tommy Tsan-Yuk Lam,^{1,2,3} Huachen Zhu,^{2,3} David K. Smith,^{2,3} Yi Guan,^{2,3} Edward C. Holmes^{4,5} and Oliver G. Pybus¹

Correspondence

Tommy Tsan-Yuk Lam
tylam.tommy@gmail.com

¹Department of Zoology, University of Oxford, Oxford, UK

²International Institute of Infection and Immunity, Shantou University Medical College, Shantou, Guangdong, PR China

³State Key Laboratory of Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR

⁴Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA, USA

⁵Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

Received 22 October 2011
Accepted 9 January 2012

GII.4 noroviruses are a major cause of acute gastroenteritis in humans. A new variant of GII.4, the 2008 variant, has recently increased its prevalence on a global scale. A previous study of this variant in Japan suggested that it might be of recombinant origin, with a breakpoint at the ORF1–ORF2 junction. Here, examination of the evolutionary origin of the 2008 variant based on a larger sample of worldwide GII.4 norovirus sequences revealed a more complex pattern of recombination between the 2006a- and 2006b-like variants of genotype GII.4, involving the P2 antigenic domain. Double (termed ‘2008i’) and triple (termed ‘2008ii’) recombinant forms of 2008 variants were identified. This study highlights the possible importance of intra-genotypic recombination over antigenic regions in driving norovirus evolution, and is suggestive of a process analogous to the antigenic shift of influenza A virus by reassortment.

Noroviruses, a major aetiological agent of gastroenteritis in humans (Atmar & Estes, 2006), are classified into five genogroups (GI–GV), which are in turn divided into a number of genotypes (Zheng *et al.*, 2006). Surveillance since the 1990s has confirmed that genotype 4 of the GII genogroup (denoted GII.4) dominates most worldwide outbreaks of norovirus-related gastroenteritis (Jin *et al.*, 2008; Kroneman *et al.*, 2008; Nayak *et al.*, 2009; Park *et al.*, 2010; Siebenga *et al.*, 2009).

Over the past 20 years GII.4 norovirus has evolved a series of genetic variants, some of which persisted and replaced the previously circulating variants (Reuter *et al.*, 2008; Siebenga *et al.*, 2010). The continuous evolution of the viral capsid VP1 protein has been proposed as a key mechanism by which new antigenic variants are generated. This is supported by observations of temporal changes in histo-blood group antigen-binding characteristics, antigenic relatedness and genome composition among GII.4 norovirus variants (Lindesmith *et al.*, 2008).

From 2006 to 2007 onwards the GII.4 ‘2006b variant’ dominated GII.4 norovirus outbreaks worldwide (Chan-It *et al.*, 2011; Chung *et al.*, 2010; Eden *et al.*, 2010; Jin *et al.*, 2008; Kittigul *et al.*, 2010). However, the ‘2008 variant’ has been reported increasingly frequently since its first isolation in 2008 (Belliot *et al.*, 2010; Motomura *et al.*, 2010) and, at the time of writing, has been identified in 12 countries on four continents (Belliot *et al.*, 2010; Eden *et al.*, 2010; Han *et al.*, 2011; Mans *et al.*, 2010; Motomura *et al.*, 2010; Pang *et al.*, 2010; Schenk *et al.*, 2010). Previous phylogenetic analysis showed that the 2008 variant emerged from the Hunter-2006a ancestral lineage, with which it shares 88% nucleotide similarity (Eden *et al.*, 2010). However, the phylogenetic position of the 2008 variant differed according to the genome region investigated (Belliot *et al.*, 2010; Eden *et al.*, 2010; Mans *et al.*, 2010; Motomura *et al.*, 2010), with a recombination breakpoint identified at the ORF1–ORF2 junction of the Japanese 2008 variants (Motomura *et al.*, 2010).

To better understand the role of recombination in the evolutionary origin of the 2008 variant, partial and complete genome sequences of GII.4 noroviruses from

Supplementary material is available with the online version of this paper.

GenBank were examined. After alignment using MUSCLE v3.8 (Edgar, 2004), a maximum-likelihood (ML) phylogeny of GII.4 ORF2 sequences (1620 nt; $n=600$) was estimated using PhyML v3.0 (Guindon & Gascuel, 2003). Topological robustness was assessed by performing ML

bootstrap analysis (1000 pseudoreplicates; using PhyML) and Bayesian phylogeny sampling (5000 samplings, each at every 2000 states; using MrBayes 3.1.2; Ronquist & Huelsenbeck, 2003). This revealed a distinct monophyletic lineage of variant strains ($n=35$; Fig. 1) that was previously

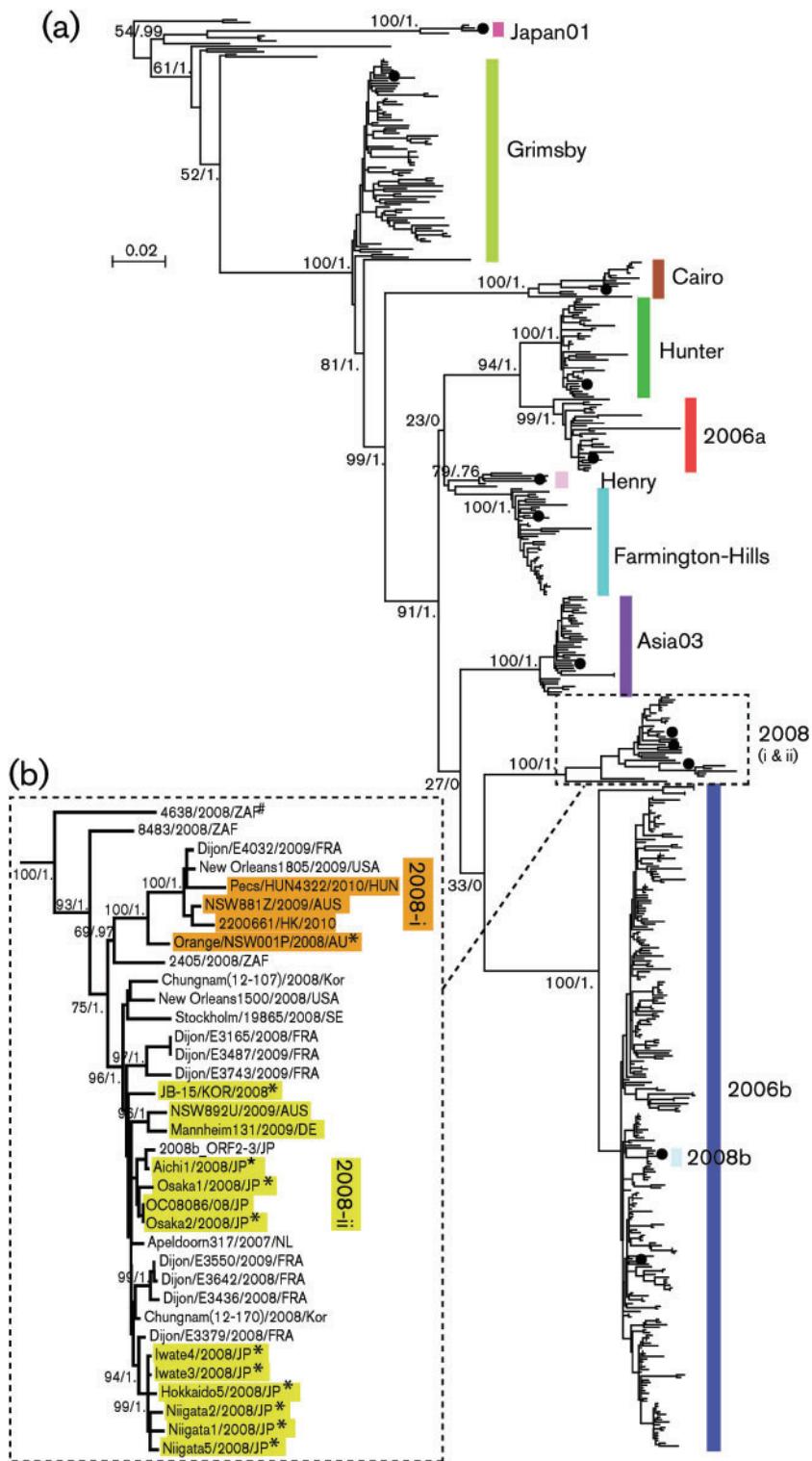


Fig. 1. ML phylogeny of ORF2 of GII.4 norovirus ($n=600$). (a) ML tree of ORF2 of GII.4. Variant lineages are compressed for clarity. They are Japan01 (Kaiso/030556/2003/JP), Grimsby (Grimsby), Cairo (Cairo2/2006/EGY), Hunter (2004/NL), 2006a (Yerseke38/2006/NL), Henry (Houston/TCH186/2002/US), Farmington-Hills (Langen1061/2002/DE), Asia03 (Sakai/04-179/2005/JP), 2008 (JB-15/KOR/2008, Osaka1/2008/JP, Orange/NSW-001P/2008/AU), 2006b (DenHaag89/2006/NL) and 2008b (Hokkaido4/2008/JP). Representative strains (in parentheses), mostly from Siebenga *et al.* (2009) are indicated by black circles. Bootstrap support and clade posterior probabilities are shown on the left and right side of the slashes. (b) The lineage of 2008 variants extracted from the larger tree. Putative 2008i and 2008ii variants are indicated in orange and yellow, respectively. Virus strains for which complete genome sequences are available for analysis are indicated with asterisks. A 2008-derived recombinant is indicated by a hash. Branch lengths are in unit of substitutions per site.

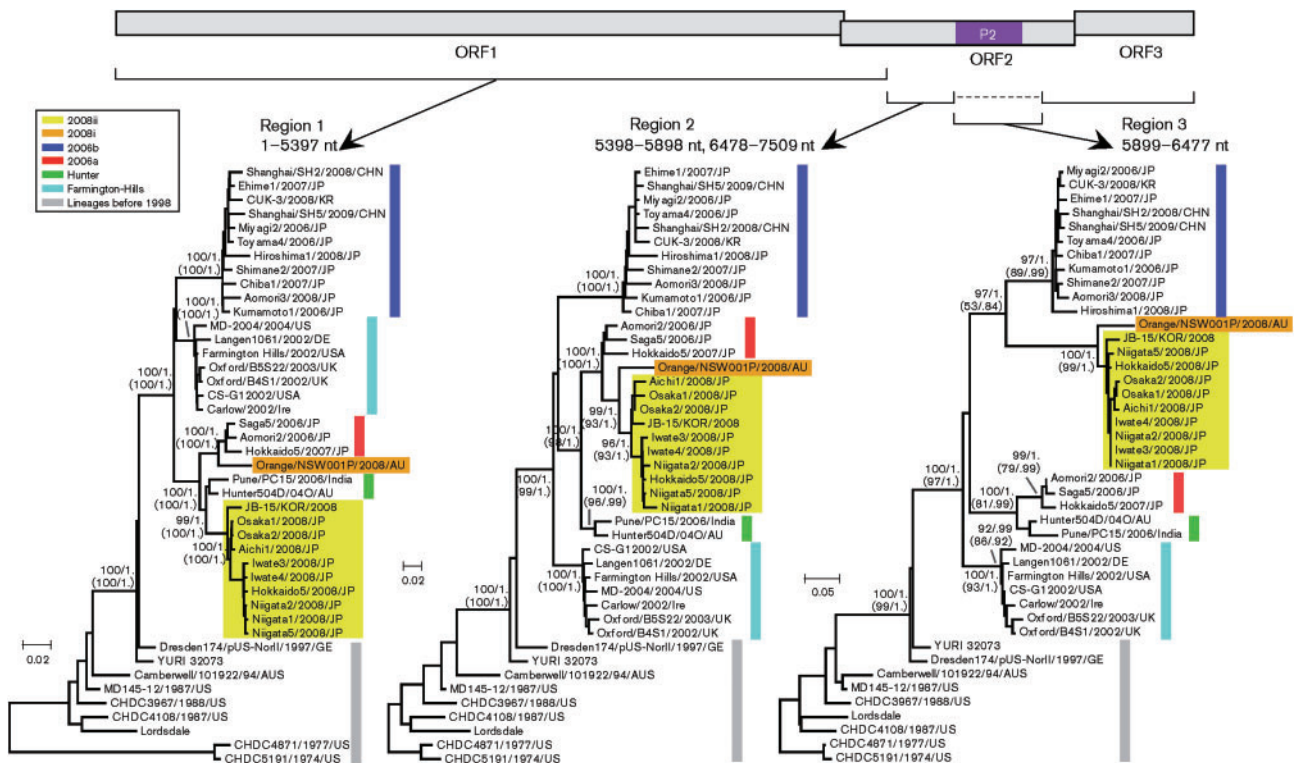


Fig. 2. Recombination analysis of the full genome sequences of GII.4 2008 variants. ML phylogenies were inferred from the three non-recombinant regions 1, 2 and 3. Bootstrap support (left) and posterior probabilities (right) for complete sequences are separated by slashes and support values for trees based on synonymous site sequences are shown below these in parentheses. Branch lengths are in unit of substitutions per site.

defined as ‘2008 variant’ (Belliot *et al.*, 2010; Mans *et al.*, 2010), and was also known as ‘2008a variant’ according to Motomura *et al.* (2010). Other known GII.4 variants are also identified in Fig. 1 (colour bars).

Systematic screening for recombination was undertaken using the suite of statistical tests implemented in RDP v3.41 (Martin & Rybicki, 2000). Sequences of 2008 variant viruses (1620 nt; $n=35$; Supplementary Table S1, available in JGV Online) consistently exhibited a mosaic fragment about 300–500 nt long in the ORF2 P2 domain (statistically significant in >2 RDP tests, after Bonferroni correction). Results suggest that the 2006b and 2006a variants were the ancestors of the small and large ORF2 fragments, respectively. Next, we studied the complete genomes of 2008 variants (7509 bp; $n=11$; Supplementary Table S1) that revealed similar breakpoints in the P2 region. However, all 2008 variants (except Orange/NSW001P/2008/AU, denoted ‘NSW001P’) also exhibited an additional breakpoint near the ORF1–ORF2 junction (statistically significant in >4 RDP tests, after Bonferroni correction). Further analysis using similarity plots in Simplot v3.5.1 (Lole *et al.*, 1999) gave concordant results (Supplementary Fig. S1a, b, available in JGV Online). Using the best recombination model selected by the phylogenetic GARD method (Kosakovsky Pond *et al.*,

2006), the recombination breakpoints were estimated to be at nucleotide positions 5398, 5866 (5932 for NSW001P) and 6477 (Supplementary Fig. S1a, b; right axis).

The estimated breakpoints thus define three non-recombinant genomic regions of the 2008 variant: region 1 (1–5397 nt), region 2 (5398–5898 nt plus 6478–7509 nt) and region 3 (5899–6477 nt) (Fig. 2). Separate phylogenies of these three regions were estimated using both ML (in PhyML) and Bayesian (in MrBayes) methods (Fig. 2). The GTR+I+ Γ_4 substitution model was used and 1000 bootstrap replicates were computed. The MrBayes analysis was performed using 10^7 steps, sampled every 2000. In the phylogeny of region 1 (mainly ORF1; Fig. 2 left), NSW001P clustered with the 2006a variants, while the other 2008 variant viruses ($n=10$; Supplementary Table S1) clustered with the Hunter variant. For region 2, all 2008 variants clustered with 2006a variants (Fig. 2 centre), whereas in region 3 they clustered with the 2006b variants (Fig. 2 right).

Most groupings are supported by high bootstrap scores and Bayesian posterior probabilities (Fig. 2). It is noteworthy that the 2008 variant group diverged from a node [indicated by clade support values 97/1 (53/.84) in the rightmost tree of Fig. 2] that is genetically distant (0.129

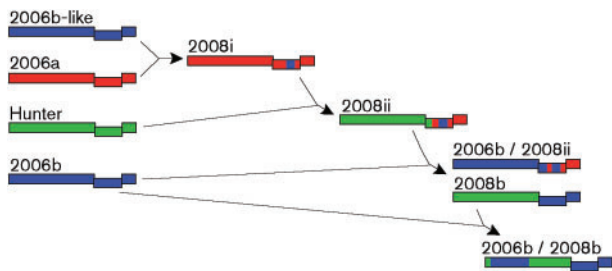


Fig. 3. Schematic diagram showing a possible recombination history for the genesis of 2008 variants and their descendant recombinants. Only those with complete genome sequences are shown. The 2006b/2008ii and 2008b recombinants were generated independently from different recombination events of 2008ii and 2006b. Toyama5/2008/JP and Iwate5/2007/JP are the sporadic 2006b/2008ii and 2006b/2008b recombinant strains, respectively. The approximate positions (starting from the first nucleotide of ORF1) of the breakpoints in the recombinant strains are: 5899 and 6478 for 2008i; 5398, 5899 and 6478 for 2008ii; 5599, 5899 and 6478 for 2006b/2008ii; 5097 for 2008b; 357, 2583 and 5097 for 2006b/2008b.

substitutions per site) to the most recent common ancestor of 2006b in region 3. This suggests a possible much older origin in region 3 of the 2008 variant, compared with regions 1 and 2 of the 2008 variant originated from Hunter and 2006a variants. Hereafter, we use the term '2006b-like' to describe the more distant relationship of 2006b lineage with the 2008 variant. Osaka sequences of the Cairo variant are clustered with, but ancestral to, the 2008 and 2006b variants in region 3 (Supplementary Fig. S3, available in JGV Online).

Phylogenetic-based analyses of recombination may be affected by convergent evolution leading to similar sets of amino acid substitutions in independent lineages. To examine such effects, we aligned the consensus sequences of different GII.4 variants, and removed the codon positions that are non-synonymous across variant lineages from the alignment. ML and Bayesian phylogenies were then estimated from these synonymous-site sequences (4710, 1250 and 412 bp in length for the regions 1, 2 and 3, respectively). The resultant tree topologies are largely congruent with regions 1, 2 and 3 phylogenies obtained from the full data (Fig. 2). Clustering of 2008 and 2006b variants in the region 3 synonymous-site phylogeny has a lower ML bootstrap support (53%), but quite high Bayesian posterior probability (0.84).

There are 35 viral isolates falling inside the 2008 variant lineage in ORF2 phylogeny (Fig. 1). Only 11 of them have complete genomes that were confidently classified into two recombinant forms (Figs 2 and 3): (i) a 2006a/2006b-like double recombinant and (ii) a 2006a/2006b-like/Hunter triple recombinant, which we term '2008i' and '2008ii', respectively. The remaining 24 of them showed a recombinant pattern in the P2 domain similar to that

observed in the 11 complete genomes, hence these 24 are, at least, double recombinants like 2008i. Phylogenetic analyses of six of them (those for which sufficient sequence before nucleotide position 5397 is available to distinguish between 2008i and 2008ii; see Supplementary Fig. S2a, available in JGV Online) suggests that three have the 2008i recombination structure (orange without asterisk in Fig. 1b) and three have the 2008ii structure (yellow without asterisk in Fig. 1b). Furthermore, Han *et al.* (2011) reported five 2008 variants in Korea based on their partial sequences at ORF1–ORF2 junction (GenBank accession numbers: HM635099–HM635103; the available ORF2 region is too short to be included in the ORF2 phylogeny of Fig. 1). Our re-analysis suggests that three of them might have the 2008ii genome structure, one might have the 2008i structure and one (Seoul/0654/2009/KOR) might be a Hunter variant, because it branched from a node closer to the Hunter lineage than to the 2008ii lineage, but could also be derived from the early ancestor of 2008ii (Supplementary Fig. S2a). Complete genomes of these isolates would help to confirm these speculations.

A simple recombination history may explain the emergence of the norovirus GII.4 2008 variants (Fig. 3). An ancestral 2006b-like virus might have recombined (by contributing a fragment of its P2 domain) with an ancestral 2006a virus creating the 2008i recombinant. Some of these continue to circulate (e.g. NSW001P). Others may have subsequently acquired a 5' genomic region (5397 nt in length) from an ancestral Hunter variant to form the triple recombinant 2008ii. This scenario involves the least number of recombination steps, but conflicts with the observation that the 2008i lineage is slightly more distant than the 2008ii lineage to the most recent common ancestor of the 2008 lineage (Fig. 1b). To resolve the order of recombination with more certainty, full-length genome sequences of early 2008 variants (such as 8483/2008/ZAF and 2405/2008/ZAF) and molecular dating analyses (Lam *et al.*, 2008; Tee *et al.*, 2009) are needed.

There are subsequent recombinants derived from 2008 and other GII.4 variants. Recently, Motomura *et al.* (2010) reported a recombinant group, called 2008b, that is 2008-like in ORF1 and 2006b-like in other genomic regions. Our phylogenetic analysis (Supplementary Fig. S2b) suggests that the origin of the ORF1 of 2008b is most likely 2008ii: the 2008b group diverged before (but quite close to) the 2008ii lineage in a phylogeny of positions 1–5096 nt, and is placed inside the 2008ii in a phylogeny of positions 4264–5096 nt (Supplementary Fig. S2b). There are two other Japanese sporadic recombinant strains, Iwate5/2007/JP (termed 2007b in Motomura *et al.*, 2010), which is a recombinant of 2008b and 2006b (Supplementary Fig. S2c), and Toyama5/2008/JP, which is a recombinant of 2006b and 2008ii (Supplementary Fig. S2d). A South African isolate (4638/2008/ZAF) for which only an ORF2 sequence is available, may be a mosaic of Hunter and 2008i variants, but with a different breakpoint location (at position 5966 nt) to that of the 2008ii strain (Supplementary Fig. S2e).

Both mutation and recombination are important mechanisms in norovirus evolution (Bull & White, 2011; Rohayem *et al.*, 2005). As the entire clade of 2008 variant viruses appear to have a recombinant origin, intra-genotypic recombination may be more important in the emergence of new norovirus GII.4 variants than previously suggested (Lindesmith *et al.*, 2008; Siebenga *et al.*, 2007). By exchanging the P2 domain, which contains numerous receptor-binding and antigenic sites (Donaldson *et al.*, 2010; Lindesmith *et al.*, 2008; Prasad *et al.*, 1999; Siebenga *et al.*, 2007; Tan *et al.*, 2003) between two GII.4 lineages, the virus may be able to move more rapidly across its fitness landscape (Burke, 1997) in the face of high levels of immunity in the host population. Incorporation by recombination of a P2 domain from a minority strain that exhibits epitopes unfamiliar to host immunity may result in a fitter recombinant progeny. This scenario may explain the emergence of the 2008 variant, whose 2006b-like parent may represent a minority circulating lineage that has not been discovered in the past. The putative antigenic change caused by recombination within the P2 domain of norovirus may resemble the antigenic shift caused by reassortment in human influenza virus, whereby gene segments encoding viral surface proteins are replaced by those of other (usually zoonotic) origins and representing novel antigenicity to human. However, such an analogy must remain speculative until a robust way of classifying noroviruses by antigenicity becomes available.

Although the pathogenicity and transmissibility of the 2008 variant remain to be studied experimentally, its novel intra-genotypic recombinant nature and widespread global distribution suggest it should be a surveillance target. Although the prevalence of the 2008 variant among GII.4 infections varies geographically [$\sim 1\%$ in Canada (Pang *et al.*, 2010), $\sim 8\%$ in France (Belliot *et al.*, 2010), $\sim 22\%$ in Korea (Han *et al.*, 2011), and up to 80% in South Africa (Mans *et al.*, 2010)], its increasing prevalence (Belliot *et al.*, 2010; Motomura *et al.*, 2010) and involvement in other recombinants raise a global public health concern. Despite this, we note that emergence of new GII.4 variants do not always cause dramatic outbreaks, e.g. during the 2009–2010 winter in the USA, when the 2008 variant first appeared there (Yen *et al.*, 2011). In general, recombination of the norovirus antigenic region could have major implications for the design and effectiveness of the norovirus vaccines.

Acknowledgements

T.T.-Y.L. is supported by Newton International Fellowship from Royal Society. H.Z., D.K.S. and Y.G. are supported by Li Ka Shing Foundation and Research Grant Council of Hong Kong SAR (HKU 765809M). O.G.P. is supported by Royal Society, UK. Additional bioinformatic and computational resources were provided by the Computer Centre at The University of Hong Kong (HKU). We thank the two anonymous reviewers for useful suggestions. We also thank W. K. Kwan, Frankie Cheung and Lilian Chan (HKU) for technical support.

References

- Atmar, R. L. & Estes, M. K. (2006). The epidemiologic and clinical importance of norovirus infection. *Gastroenterol Clin North Am* **35**, 275–290, viii.
- Belliot, G., Kamel, A. H., Estienney, M., Ambert-Balay, K. & Pothier, P. (2010). Evidence of emergence of new GII.4 norovirus variants from gastroenteritis outbreak survey in France during the 2007-to-2008 and 2008-to-2009 winter seasons. *J Clin Microbiol* **48**, 994–998.
- Bull, R. A. & White, P. A. (2011). Mechanisms of GII.4 norovirus evolution. *Trends Microbiol* **19**, 233–240.
- Burke, D. S. (1997). Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis* **3**, 253–259.
- Chan-It, W., Thongprachum, A., Okitsu, S., Nishimura, S., Kikuta, H., Baba, T., Yamamoto, A., Sugita, K., Hashira, S. & other authors (2011). Detection and genetic characterization of norovirus infections in children with acute gastroenteritis in Japan, 2007–2009. *Clin Lab* **57**, 213–220.
- Chung, J. Y., Han, T. H., Park, S. H., Kim, S. W. & Hwang, E. S. (2010). Detection of GII-4/2006b variant and recombinant noroviruses in children with acute gastroenteritis, South Korea. *J Med Virol* **82**, 146–152.
- Donaldson, E. F., Lindesmith, L. C., Lobue, A. D. & Baric, R. S. (2010). Viral shape-shifting: norovirus evasion of the human immune system. *Nat Rev Microbiol* **8**, 231–241.
- Eden, J. S., Bull, R. A., Tu, E., McIver, C. J., Lyon, M. J., Marshall, J. A., Smith, D. W., Musto, J., Rawlinson, W. D. & White, P. A. (2010). Norovirus GII.4 variant 2006b caused epidemics of acute gastroenteritis in Australia during 2007 and 2008. *J Clin Virol* **49**, 265–271.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Han, T. H., Kim, C. H., Chung, J. Y., Park, S. H. & Hwang, E. S. (2011). Emergence of norovirus GII-4/2008 variant and recombinant strains in Seoul, Korea. *Arch Virol* **156**, 323–329.
- Jin, M., Xie, H. P., Duan, Z. J., Liu, N., Zhang, Q., Wu, B. S., Li, H. Y., Cheng, W. X., Yang, S. H. & other authors (2008). Emergence of the GII.4/2006b variant and recombinant noroviruses in China. *J Med Virol* **80**, 1997–2004.
- Kittigul, L., Pombubpa, K., Taweekate, Y., Diraphat, P., Sujirarat, D., Khamrin, P. & Ushijima, H. (2010). Norovirus GII-4 2006b variant circulating in patients with acute gastroenteritis in Thailand during a 2006–2007 study. *J Med Virol* **82**, 854–860.
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098.
- Kroneman, A., Verhoef, L., Harris, J., Vennema, H., Duizer, E., van Duynhoven, Y., Gray, J., Iturriza, M., Böttiger, B. & other authors (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. *J Clin Microbiol* **46**, 2959–2965.
- Lam, T. Y., Hon, C. C., Wang, Z., Hui, R. K., Zeng, F. & Leung, F. C. (2008). Evolutionary analyses of European H1N2 swine influenza A virus by placing timestamps on the multiple reassortment events. *Virus Res* **131**, 271–278.
- Lindesmith, L. C., Donaldson, E. F., Lobue, A. D., Cannon, J. L., Zheng, D. P., Vinje, J. & Baric, R. S. (2008). Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med* **5**, e31.

- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.
- Mans, J., de Villiers, J. C., du Plessis, N. M., Avenant, T. & Taylor, M. B. (2010). Emerging norovirus GII.4 2008 variant detected in hospitalised paediatric patients in South Africa. *J Clin Virol* **49**, 258–264.
- Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563.
- Motomura, K., Yokoyama, M., Ode, H., Nakamura, H., Mori, H., Kanda, T., Oka, T., Katayama, K., Noda, M. & other authors (2010). Divergent evolution of norovirus GII.4 by genome recombination from May 2006 to February 2009 in Japan. *J Virol* **84**, 8085–8097.
- Nayak, M. K., Chatterjee, D., Nataraju, S. M., Pativada, M., Mitra, U., Chatterjee, M. K., Saha, T. K., Sarkar, U. & Krishnan, T. (2009). A new variant of norovirus GII.4/2007 and inter-genotype recombinant strains of NVGII causing acute watery diarrhoea among children in Kolkata, India. *J Clin Virol* **45**, 223–229.
- Pang, X. L., Preiksaitis, J. K., Wong, S., Li, V. & Lee, B. E. (2010). Influence of novel norovirus GII.4 variants on gastroenteritis outbreak dynamics in Alberta and the Northern Territories, Canada between 2000 and 2008. *PLoS ONE* **5**, e11599.
- Park, K. S., Jeong, H. S., Baek, K. A., Lee, C. G., Park, S. M., Park, J. S., Choi, Y. J., Choi, H. J. & Cheon, D. S. (2010). Genetic analysis of norovirus GII.4 variants circulating in Korea in 2008. *Arch Virol* **155**, 635–641.
- Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G. & Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287–290.
- Reuter, G., Pankovics, P. & Szucs, G. (2008). Genetic drift of norovirus genotype GII-4 in seven consecutive epidemic seasons in Hungary. *J Clin Virol* **42**, 135–140.
- Rohayem, J., Münch, J. & Rethwilm, A. (2005). Evidence of recombination in the norovirus capsid gene. *J Virol* **79**, 4977–4990.
- Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- Schenk, S., Petzold, A., Hoehne, M., Adam, R., Schrotten, H. & Tenenbaum, T. (2010). Severe gastroenteritis with secondary fever in a 10-month-old boy. *J Clin Virol* **47**, 107–109.
- Siebenga, J. J., Vennema, H., Renckens, B., de Bruin, E., van der Veer, B., Siezen, R. J. & Koopmans, M. (2007). Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J Virol* **81**, 9932–9941.
- Siebenga, J. J., Vennema, H., Zheng, D. P., Vinjé, J., Lee, B. E., Pang, X. L., Ho, E. C., Lim, W., Choudekar, A. & other authors (2009). Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J Infect Dis* **200**, 802–812.
- Siebenga, J. J., Lemey, P., Kosakovsky Pond, S. L., Rambaut, A., Vennema, H. & Koopmans, M. (2010). Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. *PLoS Pathog* **6**, e1000884.
- Tan, M., Huang, P., Meller, J., Zhong, W., Farkas, T. & Jiang, X. (2003). Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. *J Virol* **77**, 12562–12571.
- Tee, K. K., Pybus, O. G., Parker, J., Ng, K. P., Kamarulzaman, A. & Takebe, Y. (2009). Estimating the date of origin of an HIV-1 circulating recombinant form. *Virology* **387**, 229–234.
- Yen, C., Wikswo, M. E., Lopman, B. A., Vinje, J., Parashar, U. D. & Hall, A. J. (2011). Impact of an emergent norovirus variant in 2009 on norovirus outbreak activity in the United States. *Clin Infect Dis* **53**, 568–571.
- Zheng, D. P., Ando, T., Fankhauser, R. L., Beard, R. S., Glass, R. I. & Monroe, S. S. (2006). Norovirus classification and proposed strain nomenclature. *Virology* **346**, 312–323.