Analysis of norovirus molecular surveillance data collected through the NoroNet network, 2005 – 2016

Janko van Beek[1,2], Miranda de Graaf[1], Haider Al-Hello[3], David J Allen[4,5,6], Katia Ambert-Balay[7], Nadine Botteldoorn[8], Mia Brytting[9], prof Javier Buesa[10], Maria Cabrerizo[11,12], Martin Chan[13], Fiona Cloak[14], Ilaria Di Bartolo[15], Susana Guix[16], Joanne Hewitt[17], Nobuhiro Iritani[18], prof Miao Jin[19], prof Reimar Johne[20], Ingeborg Lederer[21], Janet Mans[22], Vito Martella[23], Leena Maunula[24], Georgina McAllister[25], Sandra Niendorf[26], prof Hubert G. Niesters[27], Alexander T Podkolzin[28], Mateja Poljsak-Prijatelj[29], Lasse Dam Rasmussen[30], prof Gábor Reuter[31], Gráinne Tuite[32], Annelies Kroneman[2], Harry Vennema[2], prof Marion P.G. Koopmans[1,2], on behalf of NoroNet[33]

Corresponding author

Marion Koopmans, m.koopmans@erasmusmc.nl. Department of Viroscience, Erasmus Medical Center, 's-Gravendijkwal 230, PO Box 3015 CE, Rotterdam, the Netherlands, +31107033289

Keywords: *Calicivirus,* norovirus, surveillance, molecular epidemiology, diversity

Author affiliations

1 Department of Viroscience, Erasmus Medical Center, Rotterdam, the Netherlands

2 Centre for Infectious Diseases Research, Diagnostics and Screening, National Institute of Public Health and the Environment, Bilthoven, the Netherlands

3 Department of Health Security, National Institute for Health and Welfare, Helsinki, Finland

4 Virus Reference Department, Public Health England, London, United Kingdom

5 Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

6 NIHR Health Protection Research Unit in Gastrointestinal Infections, UK

7 National Reference Centre for Gastroenteritis Viruses, University Hospital of Dijon, Dijon, France
8 Scientific service of foodborne pathogens, Institute of Public Health, Brussels, Belgium
9 Microbial Typing Unit, The Public Health Agency of Sweden, Stockholm, Sweden
10 Viral Gastroenteritis Research Group, Department of Microbiology, University of Valencia, Valencia, Spain
11 Enterovirus and Viral Gastroenteritis Unit, Instituto de Salud Carlos III, Madrid, Spain
12 Translational Research Network in Pediatric Infectious Diseases, Instituto de Investigación Sanitaria de la Paz (IdiPaz)
13 Department of Microbiology, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China
14 GZV (Gastroenteric/Vectorborne/Zoonotic) Unit, Health Protection Surveillance Centre, Dublin, Ireland
15 Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanita, Rome, Italy
16 Enteric Virus Laboratory, University of Barcelona, Barcelona, Spain
17 Norovirus Reference Laboratory, Institute of Environmental Science and Research, Porirua, New Zealand
18 Department of Microbiology, Osaka Institute of Public Health, Osaka, Japan
19 Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, China CDC, Beijng, China
20 Department of Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany
21 Reference Centres and Reference Laboratories, Austrian Agency for Health and Food Safety, Vienna, Austria
22 Department of Medical Virology, University of Pretoria, Pretoria, South Africa
23 Department of Veterinary Medicine, University of Bari, Bari, Italy
24 Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland
25 Specialist Virology Centre, Royal Infirmary Edinburgh, Edinburgh, United Kingdom
26 Consultant Laboratory for Noroviruses, Robert Koch Institute, Berlin, Germany

27 Department of Medical Microbiology, Division of Clinical Virology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
28 Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor), Central Research Institute of Epidemiology, Moscow, Russia
29 Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
30 Department of Virus & Microbiological Special Diagnostics, Statens Serum Institut, Copenhagen, Denmark
31 Department of Medical Microbiology and Immunology, University of Pécs, Pécs, Hungary
32 National Virus Reference Laboratory, University College Dublin, Dublin, Ireland
33 http://www.noronet.nl

Abstract

Background

Noroviruses are a common aetiology of acute gastroenteritis worldwide. Development of vaccines requires detailed understanding of global genetic diversity of noroviruses. This study describes trends in epidemiology and diversity based on global NoroNet surveillance data, and gives a future perspective on the global surveillance needs in light of these developments.

Methods

The study analysed n=16635 norovirus sequences with associated epidemiological metadata, shared between 2005 and 2016 through NoroNet by partners from Europe, Asia, Oceania, and Africa. Sequences and epidemiological data were obtained from samples collected for outbreak investigations and diagnosis of sporadic gastroenteritis cases by clinical-, public health-, and food microbiology laboratories.

Findings

During the study period, 26 different norovirus capsid genotypes circulated and 22 different recombinant genomes were found. The previously observed 2-3-year periodicity of emergence of genogroup II genotype 4 (GII.4) drift variants was not observed since 2012. Instead, the GII.4 Sydney capsid seems to persist through recombination, and we report a novel recombinant of GII.P16-GII.4 Sydney 2012 variant in Asia and Europe. The novel GII.P17-GII.17, first reported in Asia in 2014, has circulated widely in Europe. GII.4 viruses were more common in outbreaks in healthcare settings compared to other genotypes.

Interpretation

Continuous changes in the global norovirus genetic diversity highlight the need for sustained global norovirus surveillance, including assessment of possible immune escape and evolution by recombination to provide a full overview of norovirus epidemiology for future vaccine policy decisions.

Funding

115 This study was supported by the EU H2020 grant COMPARE, ZonMw TOP grant,
116 the Virgo Consortium funded by Dutch government, and by the Hungarian Scientific
117 Research Fund.
118

[BOX] Research in context

Evidence before this study

We searched Pubmed for articles published before 9th of July 2017 using keywords (worldwide OR global) AND norovirus AND genetic AND diversity in the title or abstract, and found 109 original research articles. The majority of studies reported on norovirus genetic diversity in a limited geographic area, timeframe, or focused on a single genotype. None of the studies presented long-term global norovirus diversity trends combined with epidemiological metadata, except one study focusing on the global norovirus diversity among oyster outbreaks.

Added value of this study

This study reports long-term global trends in norovirus genetic diversity combined with epidemiological metadata, obtained from reports from 19 countries across four continents/regions shared through a jointly owned database. It shows that multiple norovirus genotypes are co-circulating simultaneously with continuous and rapid changes in the norovirus genetic diversity worldwide, and with substantial regional differences, possibly reflecting differences in epidemiology, susceptibility, or both. We show differences in the preferred transmission route, preferred outbreak setting, and seasonal variation between norovirus genotypes. Finally, we discuss gaps in the norovirus surveillance and give recommendation for improvements to fulfil surveillance needs in light of vaccine development and other future interventions.

Implications of all the available evidence

Norovirus candidate vaccines are currently tested in clinical trials. This study shows that a future norovirus vaccine needs to induce broad protective immunity, or would need to be updated on a regular basis due to continuous and rapid changes in the norovirus genetic diversity. This study highlights the need for a global norovirus surveillance system using optimized sequencing protocols to monitor possible immune escape and evolution by recombination to provide data for vaccine updates. Future studies need to address the underlying factors for preferences in transmission routes, preferences in outbreak setting, and differences in seasonality among noroviruses.

Background

Acute gastroenteritis is the second greatest burden of all infectious diseases and norovirus is responsible for almost one fifth of all cases worldwide[1]. For healthy individuals, norovirus illness is typically self-limiting and of short duration, but risk groups like young children, elderly, and immunocompromised patients can suffer from prolonged symptoms[2]. In order to better understand the epidemiology and impact of norovirus and to identify (international) outbreaks, surveillance networks have been set up in some countries in the last two decades. These efforts have been challenging as norovirus surveillance is not mandatory in many countries, and if available does not always include genetic data. Despite these challenges, collaborative studies have identified international food-borne outbreaks, and substantially increased our knowledge on the norovirus diversity and antigenic evolution with the voluntary adoption of sequence-based typing[3,4]. The genus *Norovirus* is highly diverse and divided in seven genogroups (G) of which GI, GII, and GIV have been found among humans. Genogroups are further subdivided in more than 40 genotypes[5]. The epidemiology and human health impact are strongly shaped by norovirus evolution through recombination or accumulation of mutations, known as genetic drift[6]. To capture this diversity, norovirus nomenclature is based on two parameters describing the genetic lineages of the gene encoding the viral polymerase (ORF1) and the capsid protein (ORF2). Polymerase genotypes are distinguished from capsid genotypes by a P in their name (e.g. GII.P4). This dual typing approach allows for tracking of noroviruses, including recombinant forms[7]. In 2002, an informal international data sharing network was established to study noroviruses and their diversity in relation to human health impact[8]. The work from NoroNet has contributed to the understanding that noroviruses from different genetic lineages may behave differently. Genogroup II genotype 4 (GII.4) has been the predominant strain globally and responsible for approximately 70% of outbreaks since the start of NoroNet[9-11]. The antigenicity of the capsid surface alters in a stepwise manner by selection of variants under the pressure of population immunity – a process called epochal evolution[3]. In addition, frequent exchanging of genes (recombination) results in emergence of novel noroviruses. There is currently no licensed norovirus vaccine on the market, but potential candidates have been tested in phase I and II clinical trials[12,13]. Vaccine design is complicated by the large antigenic variation within the genus, and is currently targeting most commonly found genotypes. In view of the above, most likely, a future

vaccine would need to be updated on a regular basis given the flexibility of norovirus to escape natural infection-derived population immunity, hence requiring improved coverage of surveillance[14]. We analysed whether and how data obtained via the NoroNet surveillance network can be used to address the following outstanding questions regarding norovirus molecular epidemiology:

1. What are the trends in genomic diversity, recombination, and norovirus reporting?
2. Is there evidence for differences by genogroup / genotype in region, setting, and mode of transmission?
3. Where do new variants of norovirus emerge and can emerging variants be predicted from globally linked surveillance data?

Methods

NoroNet surveillance network

NoroNet links clinical-, public health-, and food microbiology laboratories willing to share norovirus molecular and epidemiological data on outbreaks and sporadic cases, and has been in existence since the mid-1990s[8,10,15]. The network started as EU funded network in 1999, continuing since 2002 as global NoroNet[8]. A jointly owned web-based database with online analysis tools was developed in which participants share and compare their data. Participation is on a give and take basis and partners have signed a code of conduct on uses of the data, after which they are granted full access to the data. Partners are expected to contribute to joint reports, and the joint database has been used for in depth studies following approval of partners.

Samples and study area

Specimens were obtained for the purpose of outbreak investigations and diagnosis of sporadic gastroenteritis cases. All RT-PCR positive cases confirmed by sequencing can be shared via NoroNet. Data from partners with less than 50 submitted sequences during the study period were excluded. Based on these criteria, the study included norovirus sequences obtained from samples collected in 19 countries: Austria, Belgium, China, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, the Netherlands, New Zealand, Russia, Slovenia, South Africa, Spain, Sweden, and

the United Kingdom. Less than 50 entries had been obtained from partners in Australia, Chile and Norway.

Data analysis

All entries submitted from January 1st 2005 to November 17th 2016 were downloaded on November 18th 2016. Records from non-human origin, without sample date or with a sample date prior to 2005 were removed from the analysis. Norovirus sequences were genotyped by the online norovirus typing tool[16]. Sequences overlapping the ORF1/ORF2 for which ORF1 and ORF2 genotypes could be assigned were analysed separately. All available sequences in the NoroNet database, including those before 2005, were used for the analysis of first reports of emerging GII.4 variants. The Maximum likelihood trees were inferred with PhyML version 3.1, using the general time reversible (GTR) nucleotide substitution model with a proportion of invariant sites and a Γ distribution of among-site rate variation[17].

Role of the funding source

The funders had no role in designing the study, data collection, data analysis or interpretation of data, writing the report, or in the decision to submit the paper for publication. The corresponding author had full access to all data in the study and had full responsibility for decision to submit for publication.

Results

Surveillance coverage

Sixteen countries (Austria, Belgium, Denmark, Finland, France, Germany, Hungary, Italy, the Netherlands, Spain, China, Japan, South Africa, Sweden, United Kingdom, Russia) submitted norovirus sequences in five or more successive years of which six countries submitted sequences during the entire study period (Finland, France, Germany, Hungary, Italy, and the Netherlands). The NoroNet surveillance network is well represented in Europe and has a smaller number of collaborators in Asia, Oceania, and Africa (Table S1).

Number of reported sequences, sequence length and genome position

A median of 870 (IQR 345) ORF1 sequences and a median of 577 (IQR 594) ORF2 sequences was reported per year. Sequence reads had an average length of 351 bases and the majority of sequences were located in the RNA-dependent RNA polymerase region of ORF1 or 5' side of ORF2 (Figure 1). Only 2·7% of sequences covered the main antigenic sites located at the P2 domain of VP1. During the study period, 154 full VP1 sequences were reported including three full genome sequences (KC175323, KC631827, and KP998539). An increased number of reported ORF1 sequences was observed in years of or post introduction of new GII.4 variants (Den Haag 2006b in 2006, New Orleans 2009 in 2009, and Sydney 2012 in 2012) which could be primarily attributed to GII.P4 and GII.Pe (Figure 2A). The apparent decline in number of reported sequences in 2016 is an artefact due to the selection of sequences until November 18th 2016 and a submission delay.

Norovirus diversity at the genotype level

The number of reported sequences and GI versus GII ratio per country was analysed to get a better understanding of the genogroup coverage and diversity (Table S1). Overall, 1372 of 16635 (8·2%) sequences belonged to norovirus GI, 15256 of 16635 (91·7%) sequences belonged to GII, and 7 of 16635 (0·0%) sequences belonged to GIV.1. Austria reported the lowest GI proportion (3·2%) and Sweden the highest (22·3%) among European countries, while countries in Asia and South Africa only reported GII strains. Trends per genotype per year for GI and GII are shown in Figures 2A and 2B. The most consistently and commonly detected genotype was GII.P4 with 6125 of 11252 (54·8%) ORF1 sequences and 4184 of 6423 (65·1%) ORF2 sequences listed as GII.4 by the phylogeny based typing tool. The remaining ~40% is a diverse mixture of 31 ORF1 and 25 ORF2 genotypes with some genotypes only detected incidentally, while other genotypes were detected more often in some years.

Emergence of novel GII.17 genotype

NoroNet detected a sharp increase in the number of GII.P17 and GII.17 strains in 2015 – 2016 compared to previous years (Figure 2A and 2B). GII.P17 and / or GII.17 were widely detected among European countries (Belgium, Finland, France, Germany, Hungary, Italy, the Netherlands, Russia, and Slovenia) in 2015 – 2016, but not in all (Ireland, Spain, and United Kingdom) (Table S2A and S2B). The GII.P17

and GII.17 proportion of total number of sequences per country showed large variation among European countries (range 4·2 - 53·9% and 5·3 - 44·5%, respectively). GII.P17 and GII.17 were co-circulating with GII.P4, GII.Pe, and GII.4 strains in Europe, and were only more prevalent than GII.P4, GII.Pe, or GII.4 in France (ORF1) and Russia (ORF1 and ORF2). China and Japan submitted in total n=10 ORF1 and n=73 ORF2 sequences to NoroNet in 2015 - 2016, and China reported n=1 GII.17 strain.

Trends in GII.4 variants

The NoroNet GII.4 variant distribution time trends are shown in Figure 3. In 2006, GII.4 Hunter 2004 was replaced by GII.4 Den Haag 2006b, succeeded by GII.4 New Orleans 2009 and GII.4 Sydney 2012 in the Northern hemisphere winter seasons of 2009/2010 and 2012/2013, respectively. The GII.4 Sydney ORF2 variant circulated as recombinant with GII.Pe or GII.P4 New Orleans 2009 since it emerged in 2012, and has not (yet) developed a new ORF1 variant. The GII.4 New Orleans 2009 ORF2 variant almost disappeared as of 2013, while the corresponding GII.P4 New Orleans ORF1 variant was still widely detected due to recombination with the GII.4 Sydney 2012 ORF2 variant. The GII.4 variant group 'other' represents variants that were only detected with limited geographic distribution and at low level incidence or sequences that could not be typed to the variant level by the norovirus genotyping tool i.e. due to a short sequence length. Variants that were detected infrequently during the study period are: Camberwell 1994, Farmington Hills 2002, Asia 2003, Kaiso 2003, Yerseke 2006a, Apeldoorn 2007, and Osaka 2007. A novel GII.P16-GII.4 Sydney 2012 recombinant was detected in 2014 (n=2) (Germany and the Netherlands), not detected in 2015, and detected in Japan, China, and the Netherlands (n=13) in 2016 (see paragraph recombination for more information on the novel GII.P16-GII.4 Sydney 2012 recombinant).

Origin of novel GII.4 drift variants

To assess when and where novel drift variants originate, we assessed the sampling date and country of origin of the first reported sequence of global drift variants (Table S3). All assessed variants, except Hunter 2004, were detected 2-5 years before the global predominance of the particular strain, which may indicate that new drift variants were present at low levels in the population before their actual global

emergence. Hunter 2004 was firstly detected in the Netherlands in the year of emergence 2004.

Recombination

To assess the influence of ORF1/ORF2 recombination on the norovirus diversity, we selected all sequences (n=1047) that were overlapping the ORF1/ORF2 junction and for which both ORF1 and ORF2 sides could be genotyped by the norovirus genotyping tool. 477 of 1047 (45·6%) sequences were assigned as a recombinant strain (Table S4). No between genogroup recombination was observed. Remarkably, some polymerase types are more prone to recombine than others. Recombination within GII was most common: 457 recombinant sequences belong to GII of which GII.Pe–GII.4, GII.P21–GII.3, and GII.P7–GII.P6 are the most commonly detected recombinants. ORF2 GII.4 has been detected in combination with GII.P12, GII.P16, and GII.Pe. The GII.P12 recombinant was detected in 2005 – 2006 in combination with GII.4 Asia 2003. GII.P16 and GII.Pe are both only found in combination with GII.4 Sydney 2012 between 2014 and 2016 (data not shown). GII.P16 was found in combination with five different VP1 genotypes: GII.3, GII.4, GII.10, GII.12, and GII.13 which each form a separate clade in a maximum likelihood tree inferred from partial GII.P16 sequences (Figure S1). Three variants of GII.4 Sydney are currently co-circulating, all resulting from recombination: GII.P4 Orleans 2009-GII.4 Sydney 2012, GII.Pe-GII.4 Sydney 2012 and GII.P16-GII.4 Sydney 2012. The antigenic regions in the capsid do not contain any amino acid changes compared to previously circulating GII.4 Sydney strains, although the VP1 sequences of GII.P16-GII.4 Sydney 2012 cluster separately from other GII.Pe-GII.4 Sydney strains (Table S5 and Figure S2).

Differences by season, region, setting, and mode of transmission

The European norovirus season coincides with the Northern Hemisphere winter season (Figure 4A). GII.Pe/GII.P4-GII.4 sequences show the clearest winter seasonality patterns while GI and GII non GII.Pe/GII.P4-GII.4 strains are more continuously present throughout the year, but never exceed the number of GII.Pe/GII.P4-GII.4 sequences. The rate of norovirus submissions in Africa (all reported by South Africa) shows an elevation in the months September – November which coincides with the Southern Hemisphere spring season (Figure 4B). Asia

(reported by China and Japan) shows an elevation of the norovirus incidence in the Northern Hemisphere winter season with the peak in November, two months earlier compared to Europe (Figure 4C). Oceania (reported by New Zealand) shows highest incidence in October and November (spring) (Figure 4D).

The suspected mode of transmission was reported for n=6446 entries: 77·4% person-to-person transmission (n=4990), 19·9% foodborne transmission (n=1280), 2·1% waterborne transmission, and 0·7% other transmission mode (n=133, n=43, respectively) (Figure 5A). GII.4 is relatively more often transmitted via person-to-person compared to other genotypes.

The setting of the norovirus outbreak was reported for n=8772 entries: 29·7% hospital setting (n=2603), 36·0% residential institution (n=3154), 9·3% hotel, restaurant or caterer (n=819), 11·8% day care or school (n=1039), 13·2% other (n=1157) (Figure 5B). The majority of sequences were derived from samples obtained in health care - or residential institutions. GII.4 was relatively more often detected in healthcare settings (hospitals and residential institutions) compared to non-GII.4 genotypes.

Discussion

Despite differences in norovirus surveillance among countries and a lack of it in many others, the current NoroNet system is able to observe global trends and major shifts in the genetic composition of the virus population at the level of genotype and variant, as was shown by this study and by others[6,10,18,19].

The first question addressed in this study is about the trends in norovirus genomic diversity, recombination, and norovirus reporting. During the study period, we observed circulation of at least 26 ORF2 genotypes when looking at diversity of the capsid gene. The viral capsid contains epitopes that are targeted by protective antibody responses, and understanding this diversity is important for evaluation of candidate vaccines[20]. It was previously noted that increased notification reflect true increases in disease trends[18,21]. Therefore, the observed increase in reported sequences post emergence of new GII.4 variants is probably related to an increase in norovirus activity. GII.4 Sydney 2012 is the predominantly detected variant worldwide since 2012 and, given the replacement cycle of two to three years shown for previous

variants, a new antigenic variant has been anticipated for some years. This trend in antigenic evolution, however, was not observed in the period described here. Instead, viruses with GII.4 Sydney capsids, have evolved by recombination, suggesting that recombination somehow favours virus maintenance in the population. For GII.4, recombination has previously only been with the closely related sequence types GII.Pe and GII.P12, which are both suggested to be derived from an ancestor of GII.P4[22]. The drivers for emergence of recombinant genomes in a population previously exposed to the same capsid sequences remains to be understood. The novel recombinant GII.P16-GII.4 Sydney 2012 may have increased fitness due to changes in the RNA dependent RNA polymerase (RdRp) that alter the polymerase fidelity and interaction with VP1, leading to differences in replication and/or transmission efficiency[23-26].

In addition to the globally prevalent GII.4 viruses, recent studies from Asia reported a major shift in genotype composition from the predominant GII.4 to the novel GII.P17-GII.17 norovirus strain (GII.17 Kawasaki 2014) late 2014 and onwards[19,27]. The number of detected GII.P17-GII.17 strains among Asian countries within our network was limited and likely caused by a filtered submission of the respective countries. The GII.P17-GII.17 strain was widely detected among most European countries in 2015 and 2016 and showed substantial differences in prevalence among countries. This strain has not (yet) fully replaced GII.4 strains.

The great genetic diversity of noroviruses is typically not considered in epidemiological or clinical studies, but may translate to differences in the epidemiology. Therefore, we compared distribution of reported modes of transmission and settings for the reported outbreaks by genotype (question 2). The most commonly reported transmission mode for the GII.4 outbreaks reported to NoroNet was person-to-person transmission and the most commonly reported setting was residential institution [10]. Underlying driving factors for these differences compared with other genotypes are unknown. We observed substantial regional variation in the norovirus genotype distribution possibly reflecting differences in epidemiology, susceptibility of the population, or both.

Norovirus surveillance is done on a voluntary basis since funding for the network is unavailable. This is reflected by unstable reporting behaviour of many countries and a potential bias in this study. A limitation of the NoroNet network is that unstandardized convenience sampling and irregular submission affects the ability of the network to robustly identify the effect of introduction of new genotypes and variants on the norovirus impact and severity. Another limitation of the study are the gaps on the surveillance map with missing or limited data from most countries in Africa, Middle East, North – and South America, Oceania, and Asia. The USA and Australia do have norovirus surveillance, but use separate databases to store and analyse their data. Future integration of surveillance databases could help to improve our understanding of the norovirus (molecular) epidemiology.

A potential use of the NoroNet network is the identification of international outbreaks, which have been observed during periods of sustained funding[4,28]. The currently provided sequence data can be used to genotype a virus to the level of genotype and variant, but is less suitable for phylogenetic analysis for the purpose of international outbreak investigations due to the lack of standardisation of sequencing protocols. The use of next generation sequencing is explored to allow whole genome sequencing as a new standard to overcome this problem[29-31]. Most countries currently upload data to the NoroNet database batch wise, which leads to a submission delay and identification of international outbreaks potentially months after their occurrence. Countries would need to upload data on a weekly basis to be able to set effective public health measures (i.e. withdraw of a contaminated food product from the market).

Norovirus vaccine candidates are currently in phase I and II trials and although vaccine cross-protection, efficacy, and effectiveness need to be evaluated, especially in vulnerable patient populations, it seems likely that a norovirus vaccine will be available in the near future. Such a vaccine will likely need to be updated on a regular basis due to escape of the virus from population immunity, especially by the predominant GII.4[32]. Essential data about the antigenic changes, especially those located in the P2 domain of the major capsid of the virus, can be obtained via a global surveillance system. As a minimum, a shared protocol for sequencing is needed, preferably including the ORF1 / ORF2 overlap to genotype both the viral RNA-

dependent RNA polymerase and VP1, and to detect recombinant strains. A protocol for sequencing this particular region has been described[33]. In addition to this protocol, a subset of specimens could be monitored for changes in the antigenic regions using a protocol spanning the P domain of VP1. Whole genome sequencing via next generation sequencing techniques could replace both protocols and potentially provide a better insight in the evolution of the virus, including the not well studied VP2.

One of the major questions within the norovirus research field is whether we are capable of predicting emerging variants in the near future, the third and last question addressed in our study. All recent major drift variants were already circulating years before they became dominant as shown by this study and by others, suggesting early warning surveillance for variant emergence would be possible[34]. If we assume that new variants develop in the human population and could emerge anywhere in the world, as shown by this study and by others, this would require a surveillance system with global coverage including large-scale genomics to capture both capsid diversity and recombination[35,36]. A next step would be to predict antigenic properties from the genomic diversity, although this is likely to be challenging and requires development of phenotypic assays to assess antigenicity and immunity, similar to the model of the global influenza virus surveillance network. More research and new funding sources are needed to address these issues.

Contributors

MK, MG, and JB designed the study. MK, MG and JB analysed and interpreted the data, and MG and JB prepared the tables and figures. MK, MG and JB wrote the manuscript. AK, MC, HV and NI collected data and critically read the manuscript. All other authors contributed by submitting data during the study period.

Declaration of interests

We declare that we have no conflicts of interest.

Acknowledgement

This study uses data from nineteen countries shared via the NoroNet network and we gratefully thank all people (including Kate Templeton and Zhaojun Duan) who

contributed to the study by collecting and sharing of data. This study was supported by the EU H2020 grant COMPARE (grant agreement number 643476), ZonMw TOP grant under grant number 91213058, the Virgo Consortium funded by Dutch government (FES0908), and by a grant from the Hungarian Scientific Research Fund (OTKA/NKFIH K111615).

Author declaration

DJA is affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. DJA is based at The London School of Hygiene and Tropical Medicine, and Public Health England. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Figure legends

Figure 1 Position of 16628 sequence reads on the norovirus genome. Each sequence represents a line in the figure. Boxes above the graph represent the norovirus open reading frames (ORFs) of reference GII.Pe-GII.4 Sydney 2012 (Genbank accession: JX459908). ORF1 encodes for a polyprotein that is post-translationally cleaved by the virus-encoded protease (Pro) into six non-structural proteins (p48, NTPase, p22, VPg, Pro, and RNA-dependent RNA polymerase (RdRp)). ORF2 encodes for the major capsid protein (VP1) which consists of a shell (S) and protruding domains P1 and P2 with antigenic epitopes A, D, and E. ORF3 encodes for the minor capsid protein VP2.

Figure 2 Number of reported ORF1 sequences (n=11252) stratified per genotype group, genotype, and year (**A**) and number of reported ORF2 sequences (n=6423) stratified per genotype group, genotype, and year (**B**). Note that n=1047 sequences overlapping ORF1/ORF2 are counted for both ORF1 and ORF2.

Figure 3 ORF1 GII.P4 variant trends per year (n=8083, top) and ORF2 GII.4 variant trends per year (n=4184, bottom). The relatively high proportion of viruses/sequences typed as "other" in the oldest category of submissions is an artefact due to the typing tool that was used. This tool performs a phylogeny based assignment of norovirus sequences to genera, genotypes, and variants. For correct assignment of variants, the reference sequences need to be periodically updated, when a new variants arise. By focusing on correct assignment of recent sequences, older strains may then be labelled as "unknown" with the current version of the typing tool.

Figure 4 Norovirus seasonality patterns in Europe (n=13935) (**A**), Africa (n=195) (**B**), Asia (n=262) (**C**), and Oceania (n=806) (**D**), stratified per genotype group. Records without sample month were removed for this analysis.

Figure 5 Norovirus transmission route (n=8772) (**A**) and suspected outbreak setting (n=6446) (**B**), stratified per genotype group. Records without known transmission route or suspected outbreak setting were removed. Outbreaks with suspected foodborne origin and subsequent person-to-person transmission were recoded as foodborne.

Figure S1 Maximum likelihood tree for region B of ORF1 sequences displaying the genetic diversity of GII.P16 sequences that are found in combination with different VP1 sequences (used sequence length 289 nucleotides, n=34). GII.P16-GII.4 Sydney 2012 sequences are indicated in red.

Figure S2 Maximum likelihood tree inferred from all complete GII.4 VP1 sequences displaying the genetic diversity of GII.4 sequences that are detected in combination with different polymerase genotypes. GII.P16-GII.4 Sydney 2012 sequences are indicated in red.

References

1. Ahmed SM, Hall AJ, Robinson AE, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect Dis* 2014; **14**(8): 725-30.
2. Lindsay L, Wolter J, De Coster I, Van Damme P, Verstraeten T. A decade of norovirus disease risk among older adults in upper-middle and high income countries: a systematic review. *BMC infectious diseases* 2015; **15**: 425.
3. Siebenga JJ, Vennema H, Renckens B, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J Virol* 2007; **81**(18): 9932-41.
4. Verhoef L, Kouyos RD, Vennema H, et al. An integrated approach to identifying international foodborne norovirus outbreaks. *Emerg Infect Dis* 2011; **17**(3): 412-8.
5. de Graaf M, van Beek J, Koopmans MP. Human norovirus transmission and evolution in a changing world. *Nat Rev Microbiol* 2016; **14**(7): 421-33.
6. Siebenga JJ, Vennema H, Zheng DP, et al. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. *J Infect Dis* 2009; **200**(5): 802-12.
7. Kroneman A, Vega E, Vennema H, et al. Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol* 2013; **158**(10): 2059-68.
8. Koopmans M, Vennema H, Heersma H, et al. Early identification of common-source foodborne virus outbreaks in Europe. *Emerg Infect Dis* 2003; **9**(9): 1136-42.
9. Siebenga JJ, Vennema H, Duizer E, Koopmans MP. Gastroenteritis caused by norovirus GGII.4, The Netherlands, 1994-2005. *Emerg Infect Dis* 2007; **13**(1): 144-6.
10. Kroneman A, Verhoef L, Harris J, et al. 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* 2008; **46**(9): 2959-65.
11. Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinje J. Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. *J Clin Microbiol* 2014; **52**(1): 147-55.
12. Bernstein DI, Atmar RL, Lyon GM, et al. Norovirus vaccine against experimental human GII.4 virus illness: a challenge study in healthy adults. *J Infect Dis* 2015; **211**(6): 870-8.
13. Treanor JJ, Atmar RL, Frey SE, et al. A novel intramuscular bivalent norovirus virus-like particle vaccine candidate--reactogenicity, safety, and immunogenicity in a phase 1 trial in healthy adults. *J Infect Dis* 2014; **210**(11): 1763-71.
14. Debbink K, Lindesmith LC, Baric RS. The state of norovirus vaccines. *Clin Infect Dis* 2014; **58**(12): 1746-52.
15. Kroneman A, Harris J, Vennema H, et al. Data quality of 5 years of central norovirus outbreak reporting in the European Network for food-borne viruses. *J Public Health (Oxf)* 2008; **30**(1): 82-90.
16. Kroneman A, Vennema H, Deforche K, et al. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 2011; **51**(2): 121-5.
17. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003; **52**(5): 696-704.

18. 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**(1): 8-9.
19. 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**(26).
20. Lindesmith LC, Costantini V, Swanstrom J, et al. Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. *J Virol* 2013; **87**(5): 2803-13.
21. Verhoef L, Depoortere E, Boxman I, et al. Emergence of new norovirus variants on spring cruise ships and prediction of winter epidemics. *Emerg Infect Dis* 2008; **14**(2): 238-43.
22. Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA. Recombination within the pandemic norovirus GII.4 lineage. *J Virol* 2013; **87**(11): 6270-82.
23. Bull RA, Eden JS, Rawlinson WD, White PA. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog* 2010; **6**(3): e1000831.
24. Arias A, Thorne L, Ghurburrun E, Bailey D, Goodfellow I. Norovirus Polymerase Fidelity Contributes to Viral Transmission In Vivo. *mSphere* 2016; **1**(5).
25. Ruis C, Roy S, Brown JR, Allen DJ, Goldstein RA, Breuer J. The emerging GII.P16-GII.4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contains polymerase changes that may increase virus transmission. *PloS one* 2017; **12**(6): e0179572.
26. Subba-Reddy CV, Yunus MA, Goodfellow IG, Kao CC. Norovirus RNA synthesis is modulated by an interaction between the viral RNA-dependent RNA polymerase and the major capsid protein, VP1. *J Virol* 2012; **86**(18): 10138-49.
27. 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.
28. Verhoef L, Hewitt J, Barclay L, et al. Norovirus genotype profiles associated with foodborne transmission, 1999-2012. *Emerg Infect Dis* 2015; **21**(4): 592-9.
29. Brown JR, Roy S, Ruis C, et al. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. *J Clin Microbiol* 2016; **54**(10): 2530-7.
30. Bavelaar HH, Rahamat-Langendoen J, Niesters HG, Zoll J, Melchers WJ. Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. *J Clin Virol* 2015; **72**: 122-5.
31. van Beek J, de Graaf M, Smits SL, et al. Whole genome next generation sequencing to study within-host evolution of chronic norovirus infection among immunocompromised patients. *J Infect Dis* 2017.
32. Parra GI, Squires RB, Karangwa CK, et al. Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity. *PLoS Pathog* 2017; **13**(1): e1006136.
33. van Beek J, van der Eijk AA, Fraaij PL, et al. Chronic norovirus infection among solid organ recipients in a tertiary care hospital, the Netherlands, 2006-2014. *Clin Microbiol Infect* 2016.
34. Allen DJ, Trainor E, Callaghan A, O'Brien SJ, Cunliffe NA, Iturriza-Gomara M. Early Detection of Epidemic GII-4 Norovirus Strains in UK and Malawi: Role of

649 Surveillance of Sporadic Acute Gastroenteritis in Anticipating Global Epidemics.
650 *PloS one* 2016; **11**(4): e0146972.
651 35. Bull RA, Tu ET, McIver CJ, Rawlinson WD, White PA. Emergence of a new
652 norovirus genotype II.4 variant associated with global outbreaks of
653 gastroenteritis. *J Clin Microbiol* 2006; **44**(2): 327-33.
654 36. Eden JS, Hewitt J, Lim KL, et al. The emergence and evolution of the novel
655 epidemic norovirus GII.4 variant Sydney 2012. *Virology* 2014; **450-451**: 106-13.
656

Figure 1

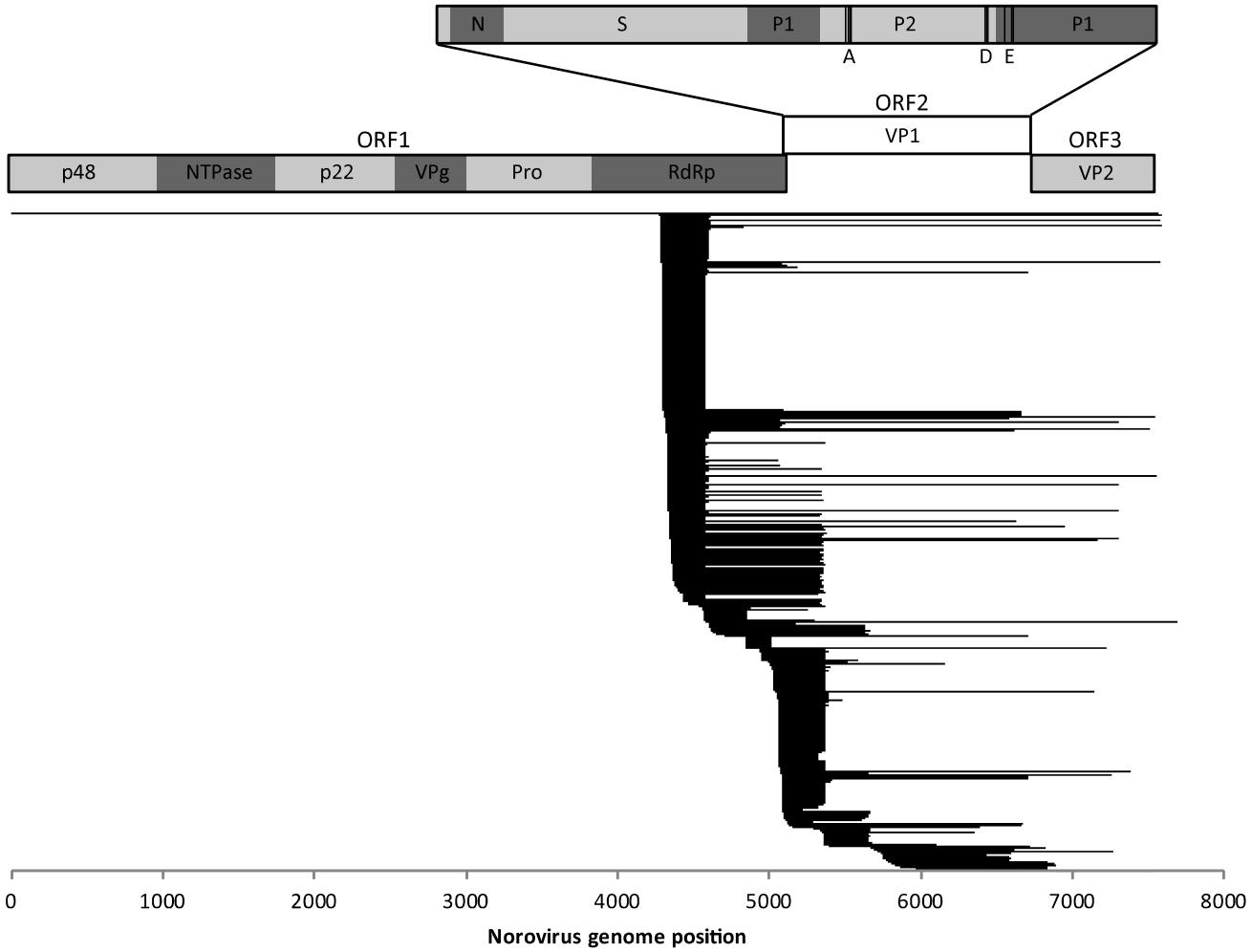
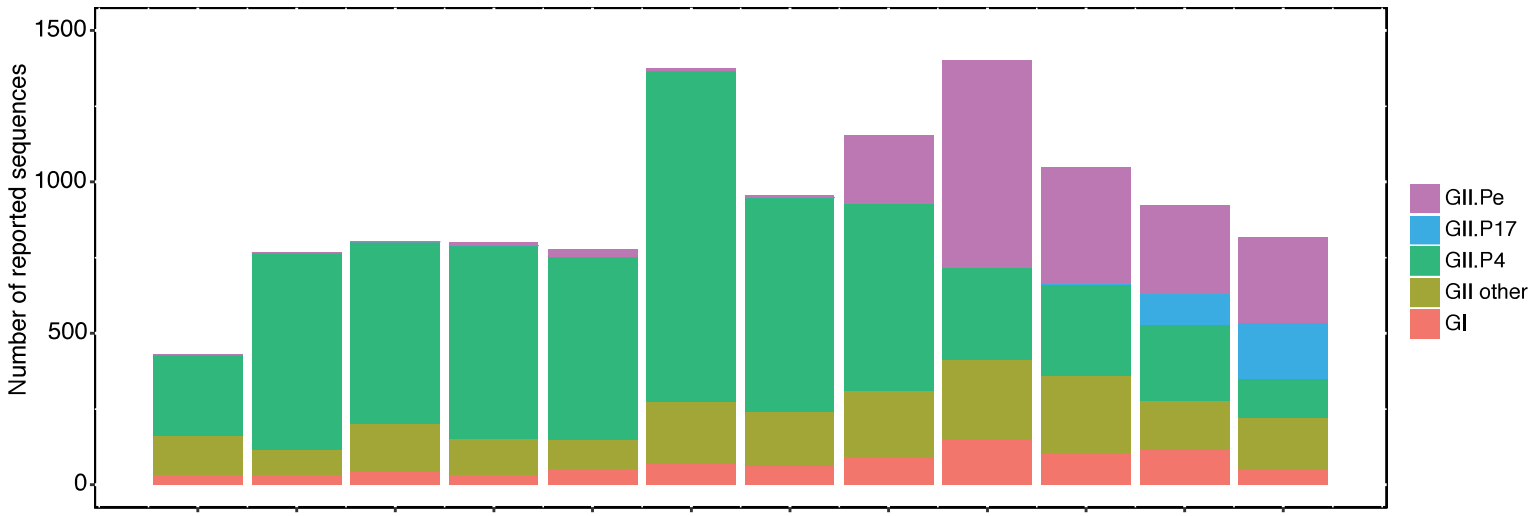
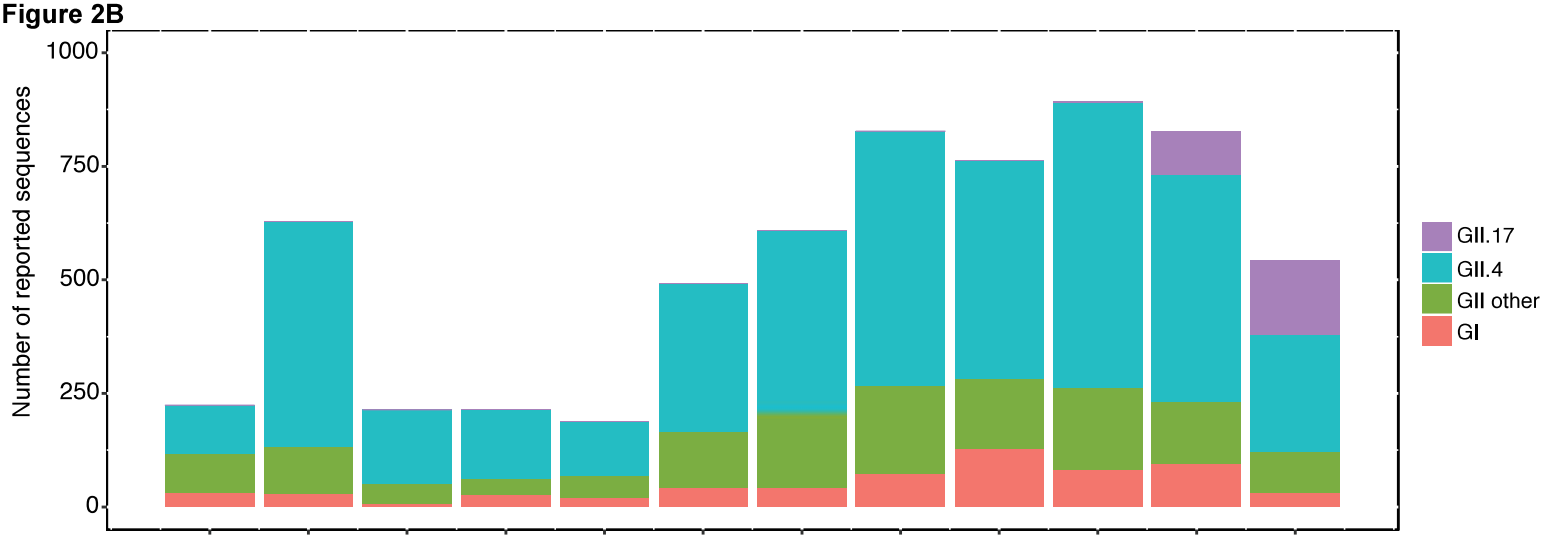

N
S
P1
P2
P1
A
D E
ORF2
VP1
ORF1
ORF3
p48
NTPase
p22
VPg
Pro
RdRp
VP2
0
1000
2000
3000
4000
5000
6000
7000
8000
Norovirus genome position

Figure 2A



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
GI.P1	4 (0,9)	0 (0)	2 (0,2)	0 (0)	0 (0)	1 (0,1)	2 (0,2)	0 (0)	4 (0,3)	3 (0,3)	6 (0,7)	1 (0,1)
GI.P2	3 (0,7)	13 (1,7)	5 (0,6)	3 (0,4)	0 (0)	3 (0,2)	3 (0,3)	5 (0,4)	2 (0,1)	17 (1,6)	10 (1,1)	17 (2,1)
GI.P3	9 (2,1)	7 (0,9)	17 (2,1)	18 (2,2)	16 (2,1)	4 (0,3)	7 (0,7)	10 (0,9)	34 (2,4)	39 (3,7)	51 (5,5)	7 (0,9)
GI.P4	9 (2,1)	7 (0,9)	5 (0,6)	5 (0,6)	33 (4,2)	32 (2,3)	21 (2,2)	14 (1,2)	40 (2,9)	4 (0,4)	10 (1,1)	12 (1,5)
GI.P5	2 (0,5)	1 (0,1)	0 (0)	0 (0)	1 (0,1)	0 (0)	0 (0)	0 (0)	6 (0,4)	4 (0,4)	7 (0,8)	3 (0,4)
GI.P6	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0,3)	0 (0)	0 (0)	0 (0)
GI.P7	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	20 (1,5)	3 (0,3)	6 (0,5)	7 (0,5)	2 (0,2)	3 (0,3)	0 (0)
GI.P8	2 (0,5)	0 (0)	0 (0)	1 (0,1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	0 (0)	0 (0)	0 (0)
GI.P9	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	0 (0)	0 (0)	0 (0)	4 (0,3)	4 (0,4)	0 (0)	1 (0,1)
GI.Pa	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	2 (0,2)	2 (0,1)	0 (0)	0 (0)	0 (0)
GI.Pb	2 (0,5)	4 (0,5)	13 (1,6)	5 (0,6)	2 (0,3)	10 (0,7)	25 (2,6)	50 (4,3)	43 (3,1)	21 (2)	22 (2,4)	4 (0,5)
GI.Pd	3 (0,7)	2 (0,3)	1 (0,1)	0 (0)	1 (0,1)	0 (0)	1 (0,1)	1 (0,1)	3 (0,2)	9 (0,9)	0 (0)	4 (0,5)
GI.Pf	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	6 (0,5)	1 (0,1)	2 (0,2)	5 (0,5)	1 (0,1)
GII.P2	15 (3,5)	10 (1,3)	9 (1,1)	8 (1)	11 (1,4)	23 (1,7)	11 (1,2)	17 (1,5)	39 (2,8)	35 (3,3)	34 (3,7)	43 (5,3)
GII.P3	1 (0,2)	2 (0,3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
GII.P4	269 (62,7)	649 (84,6)	603 (75,1)	639 (79,7)	603 (77,6)	1094 (79,6)	709 (74,2)	617 (53,4)	302 (21,5)	301 (28,7)	252 (27,3)	127 (15,5)
GII.P6	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	1 (0,1)	0 (0)	0 (0)	0 (0)	0 (0)
GII.P7	59 (13,8)	18 (2,3)	39 (4,9)	28 (3,5)	28 (3,6)	31 (2,3)	67 (7)	95 (8,2)	93 (6,6)	81 (7,7)	62 (6,7)	33 (4)
GII.P8	1 (0,2)	0 (0)	0 (0)	1 (0,1)	1 (0,1)	3 (0,2)	0 (0)	1 (0,1)	2 (0,1)	2 (0,2)	0 (0)	0 (0)
GII.P11	1 (0,2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
GII.P12	3 (0,7)	6 (0,8)	2 (0,2)	0 (0)	0 (0)	0 (0)	0 (0)	7 (0,6)	1 (0,1)	0 (0)	0 (0)	3 (0,4)
GII.P13	0 (0)	7 (0,9)	5 (0,6)	2 (0,2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
GII.P15	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	1 (0,1)	0 (0)	1 (0,1)	0 (0)	1 (0,1)	0 (0)	0 (0)
GII.P16	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	0 (0)	5 (0,5)	17 (1,5)	19 (1,4)	17 (1,6)	5 (0,5)	31 (3,8)
GII.P17	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	4 (0,4)	102 (11,1)	185 (22,6)
GII.P20	3 (0,7)	1 (0,1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	0 (0)
GII.P21	42 (9,8)	39 (5,1)	101 (12,6)	70 (8,7)	46 (5,9)	52 (3,8)	31 (3,2)	30 (2,6)	75 (5,3)	92 (8,8)	41 (4,4)	49 (6)
GII.P22	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	0 (0)	0 (0)	3 (0,3)	16 (1,1)	2 (0,2)	0 (0)	3 (0,4)
GII.Pc	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0,3)	2 (0,2)	0 (0)	0 (0)
GII.Pe	0 (0)	1 (0,1)	0 (0)	12 (1,5)	24 (3,1)	7 (0,5)	7 (0,7)	225 (19,5)	686 (48,9)	384 (36,6)	291 (31,6)	281 (34,4)
GII.Pg	1 (0,2)	0 (0)	1 (0,1)	7 (0,9)	7 (0,9)	93 (6,8)	61 (6,4)	47 (4,1)	14 (1)	22 (2,1)	20 (2,2)	12 (1,5)
GII.Pm	0 (0)	0 (0)	0 (0)	3 (0,4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Total	**429 (100)**	**767 (100)**	**803 (100)**	**802 (100)**	**777 (100)**	**1374 (100)**	**955 (100)**	**1155 (100)**	**1403 (100)**	**1048 (100)**	**922 (100)**	**817 (100)**

Figure 2B



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
GI.1	3 (1,3)	6 (1)	0 (0)	0 (0)	1 (0,5)	4 (0,8)	3 (0,5)	0 (0)	5 (0,7)	2 (0,2)	6 (0,7)	7 (1,3)
GI.2	0 (0)	8 (1,3)	2 (0,9)	4 (1,9)	0 (0)	2 (0,4)	1 (0,2)	7 (0,8)	4 (0,5)	15 (1,7)	18 (2,2)	2 (0,4)
GI.3	11 (4,9)	4 (0,6)	4 (1,9)	11 (5,1)	5 (2,7)	6 (1,2)	8 (1,3)	22 (2,7)	42 (5,5)	36 (4)	40 (4,8)	11 (2)
GI.4	12 (5,4)	8 (1,3)	1 (0,5)	3 (1,4)	14 (7,4)	16 (3,2)	18 (3)	13 (1,6)	34 (4,5)	5 (0,6)	10 (1,2)	4 (0,7)
GI.5	2 (0,9)	2 (0,3)	0 (0)	1 (0,5)	1 (0,5)	1 (0,2)	0 (0)	0 (0)	7 (0,9)	3 (0,3)	5 (0,6)	5 (0,9)
GI.6	3 (1,3)	2 (0,3)	1 (0,5)	5 (2,3)	0 (0)	7 (1,4)	11 (1,8)	22 (2,7)	28 (3,7)	17 (1,9)	13 (1,6)	3 (0,6)
GI.7	0 (0)	1 (0,2)	0 (0)	1 (0,5)	0 (0)	8 (1,6)	2 (0,3)	9 (1,1)	4 (0,5)	1 (0,1)	4 (0,5)	0 (0)
GI.8	0 (0)	0 (0)	0 (0)	1 (0,5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
GI.9	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	5 (0,7)	3 (0,3)	0 (0)	0 (0)
GII.1	2 (0,9)	0 (0)	0 (0)	4 (1,9)	0 (0)	31 (6,3)	29 (4,8)	35 (4,2)	12 (1,6)	14 (1,6)	15 (1,8)	3 (0,6)
GII.2	6 (2,7)	7 (1,1)	7 (3,3)	7 (3,3)	9 (4,8)	6 (1,2)	15 (2,5)	11 (1,3)	30 (3,9)	29 (3,2)	34 (4,1)	22 (4,1)
GII.3	21 (9,4)	15 (2,4)	22 (10,3)	10 (4,7)	16 (8,5)	15 (3)	45 (7,4)	29 (3,5)	17 (2,2)	53 (5,9)	23 (2,8)	34 (6,3)
GII.4	107 (47,8)	493 (78,5)	163 (76,2)	151 (70,6)	118 (62,8)	327 (66,3)	402 (65,9)	559 (67,6)	479 (62,8)	628 (70,3)	500 (60,5)	257 (47,3)
GII.5	0 (0)	2 (0,3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0,5)	13 (1,7)	3 (0,3)	0 (0)	0 (0)
GII.6	23 (10,3)	25 (4)	9 (4,2)	6 (2,8)	8 (4,3)	10 (2)	42 (6,9)	70 (8,5)	34 (4,5)	63 (7,1)	35 (4,2)	11 (2)
GII.7	28 (12,5)	43 (6,8)	0 (0)	1 (0,5)	2 (1,1)	4 (0,8)	18 (3)	26 (3,1)	28 (3,7)	3 (0,3)	7 (0,8)	8 (1,5)
GII.8	1 (0,4)	3 (0,5)	0 (0)	1 (0,5)	0 (0)	1 (0,2)	0 (0)	1 (0,1)	1 (0,1)	0 (0)	2 (0,2)	0 (0)
GII.10	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0,2)	0 (0)	6 (0,7)	1 (0,1)	0 (0)
GII.12	0 (0)	1 (0,2)	3 (1,4)	1 (0,5)	9 (4,8)	37 (7,5)	4 (0,7)	6 (0,7)	1 (0,1)	2 (0,2)	8 (1)	2 (0,4)
GII.13	1 (0,4)	2 (0,3)	0 (0)	1 (0,5)	0 (0)	13 (2,6)	6 (1)	4 (0,5)	10 (1,3)	4 (0,4)	5 (0,6)	4 (0,7)
GII.14	2 (0,9)	0 (0)	1 (0,5)	5 (2,3)	3 (1,6)	1 (0,2)	2 (0,3)	2 (0,2)	7 (0,9)	1 (0,1)	6 (0,7)	6 (1,1)
GII.15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0,1)	1 (0,1)	0 (0)	0 (0)
GII.16	1 (0,4)	2 (0,3)	0 (0)	0 (0)	0 (0)	1 (0,2)	1 (0,2)	1 (0,1)	0 (0)	0 (0)	0 (0)	0 (0)
GII.17	0 (0)	1 (0,2)	0 (0)	1 (0,5)	1 (0,5)	1 (0,2)	1 (0,2)	0 (0)	1 (0,1)	3 (0,3)	94 (11,4)	164 (30,2)
GII.20	1 (0,4)	3 (0,5)	0 (0)	0 (0)	1 (0,5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
GII.21	0 (0)	0 (0)	1 (0,5)	0 (0)	0 (0)	2 (0,4)	2 (0,3)	3 (0,4)	0 (0)	1 (0,1)	0 (0)	0 (0)
Total	**224 (100)**	**628 (100)**	**214 (100)**	**214 (100)**	**188 (100)**	**493 (100)**	**610 (100)**	**827 (100)**	**763 (100)**	**893 (100)**	**826 (100)**	**543 (100)**

Figure 3

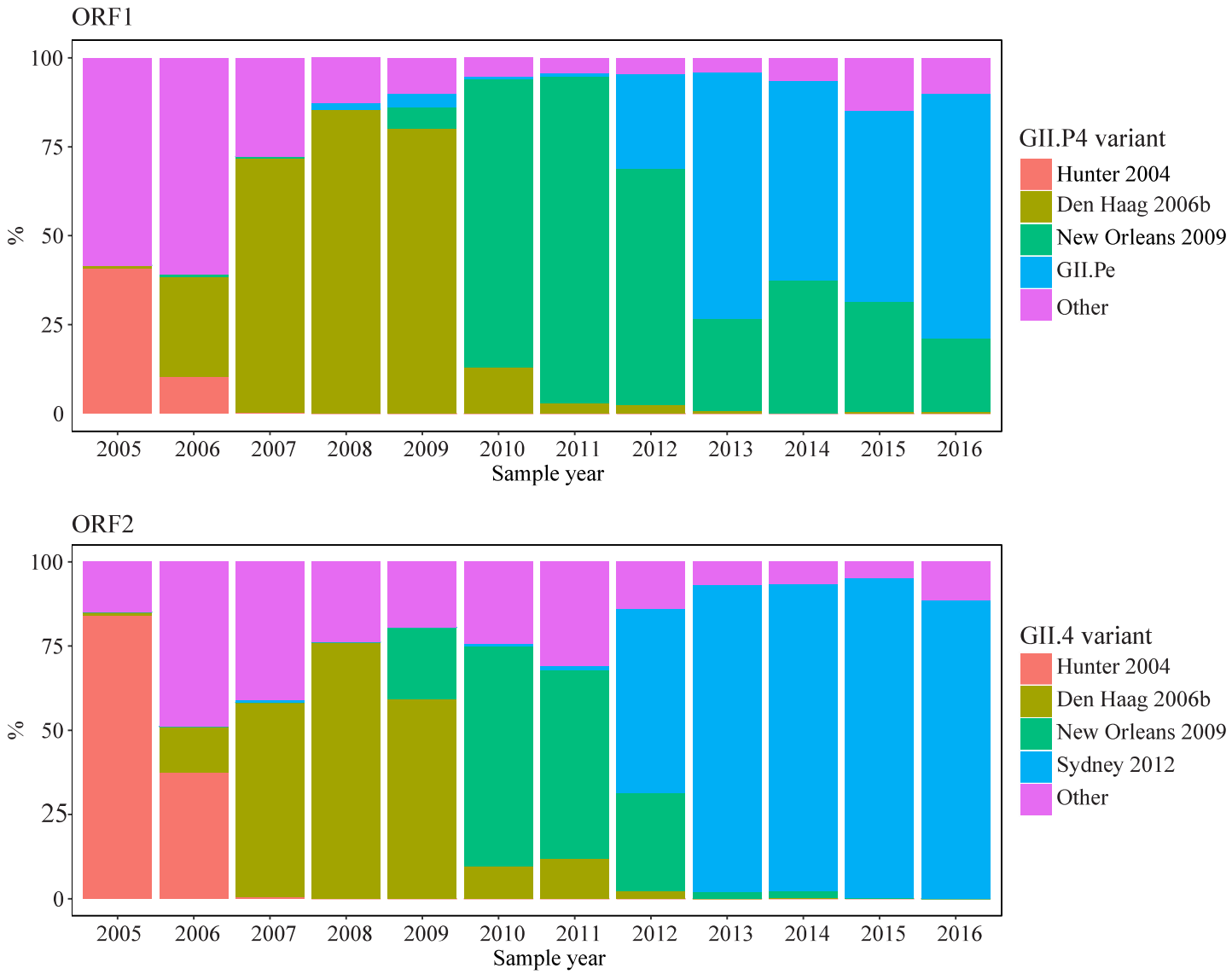

ORF1
100
75
50
25
0
%
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Sample year
GII.P4 variant
Hunter 2004
Den Haag 2006b
New Orleans 2009
GII.Pe
Other
ORF2
100
75
50
25
0
%
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Sample year
GII.4 variant
Hunter 2004
Den Haag 2006b
New Orleans 2009
Sydney 2012
Other

Figure 4

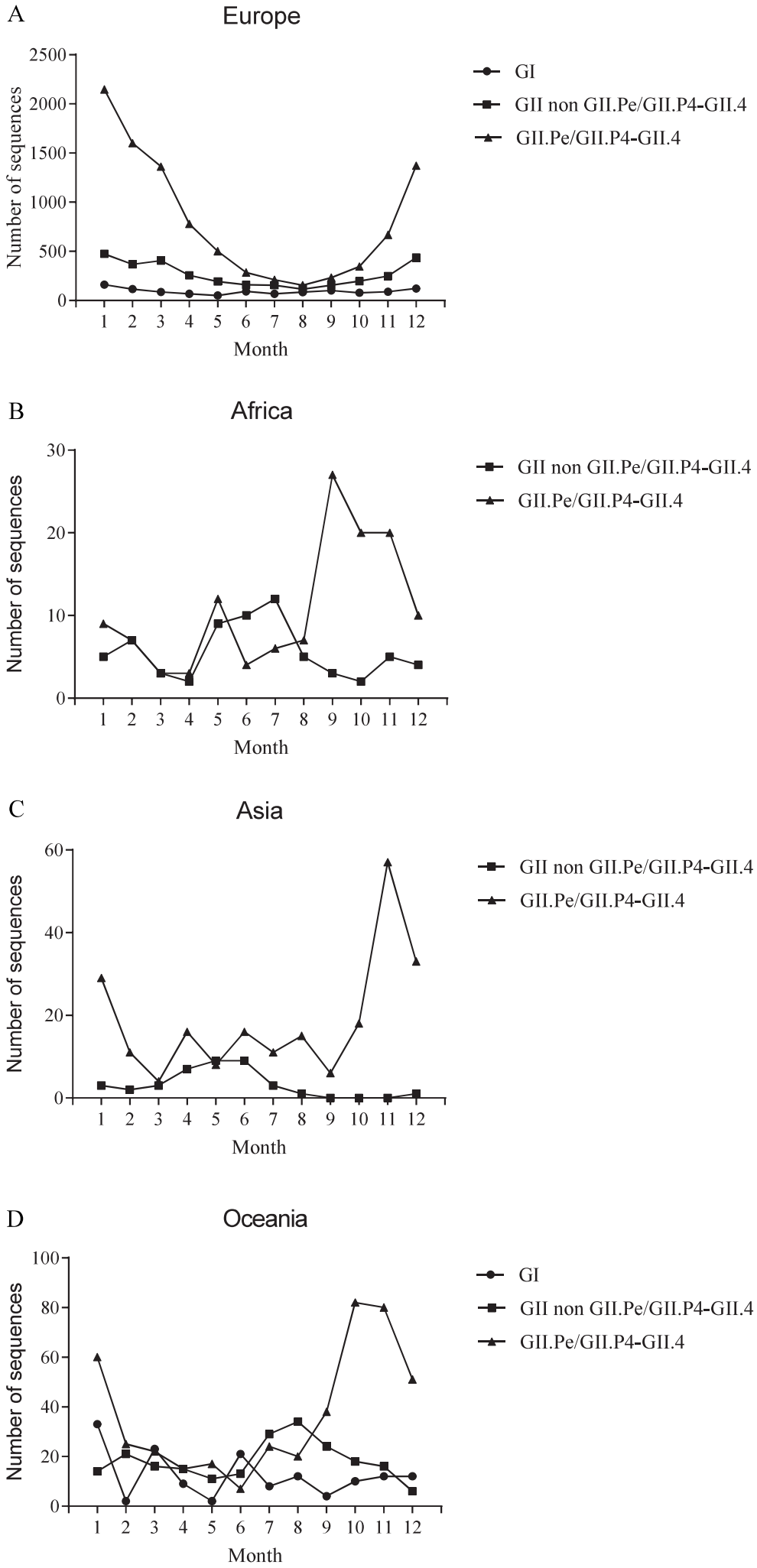
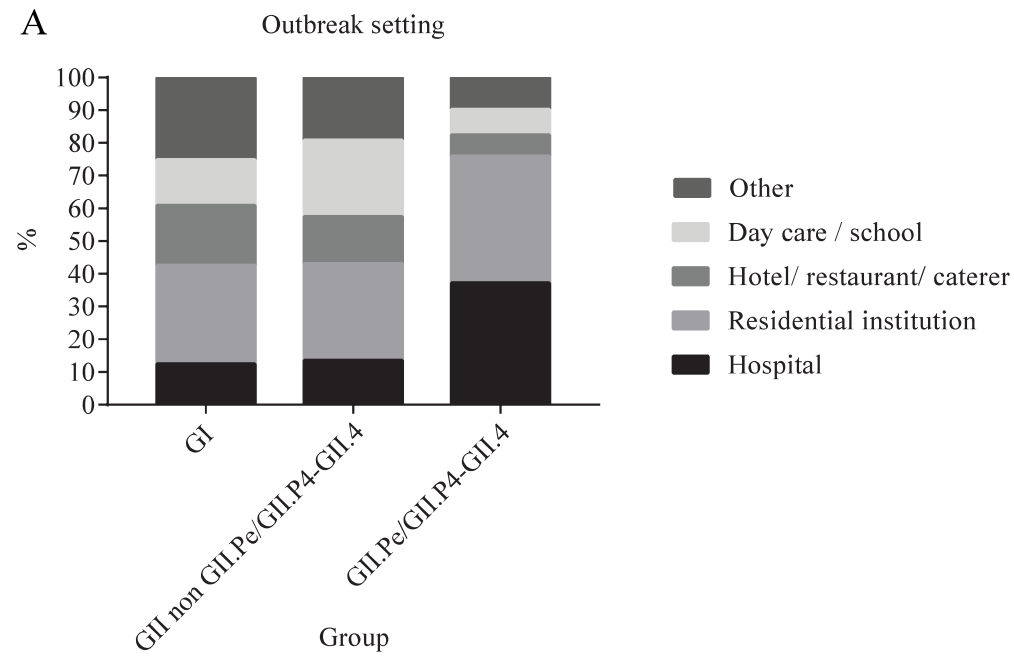
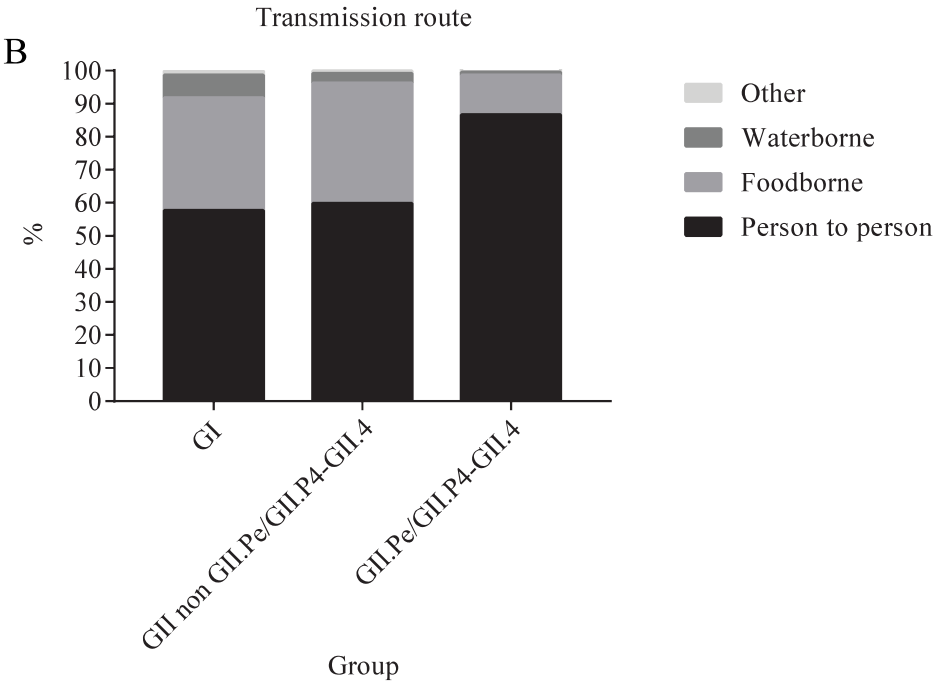

A
Europe
Number of sequences
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GI
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
B
Africa
Number of sequences
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
C
Asia
Number of sequences
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
D
Oceania
Number of sequences
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GI
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4

Figure 5


A
Outbreak setting
100
90
80
70
60
50
40
30
20
10
0
%
GI
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
Group
Other
Day care / school
Hotel/ restaurant/ caterer
Residential institution
Hospital


B
Transmission route
100
90
80
70
60
50
40
30
20
10
0
%
GI
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
Group
Other
Waterborne
Foodborne
Person to person

Supplementary tables

Supplementary Table 1 Number of reported GI and GII sequences per continent/region and country

Continent	Country	GI (%)	GII (%)	Total
Europe	Austria	6 (3,2)	180 (96,8)	186
Europe	Belgium	41 (11,4)	319 (88,6)	360
Asia	China	0 (0)	142 (100)	142
Europe	Denmark	67 (10,4)	580 (89,6)	647
Europe	Finland	96 (8,5)	1037 (91,5)	1133
Europe	France	267 (8,2)	3004 (91,8)	3271
Europe	Germany	183 (16,4)	932 (83,6)	1115
Europe	Hungary	43 (5,2)	791 (94,8)	834
Europe	Ireland	11 (7)	147 (93)	158
Europe	Italy	23 (7,7)	276 (92,3)	299
Asia	Japan	0 (0)	293 (100)	293
Europe	Netherlands	327 (6)	5100 (94)	5427
Australia	New Zealand	148 (18,4)	658 (81,6)	806
Europe	Russia	23 (7,5)	283 (92,5)	306
Europe	Slovenia	15 (6,7)	209 (93,3)	224
Africa	South Africa	0 (0)	195 (100)	195
Europe	Spain	16 (5,5)	274 (94,5)	290
Europe	Sweden	69 (22,3)	241 (77,7)	310
Europe	United Kingdom	37 (5,9)	595 (94,1)	632

Supplementary Table 2A Number of reported norovirus ORF1 sequences stratified per genogroup/genotype, country, and time

	2005-2014						2015-2016					
Country	**GI (%)**	**GII other (%)**	**GII.P4 (%)**	**GII.P17 (%)**	**GII.Pe (%)**	**Total**	**GI (%)**	**GII other (%)**	**GII.P4 (%)**	**GII.P17 (%)**	**GII.Pe (%)**	**Total**
Austria	6 (3,3)	28 (15,2)	145 (78,8)	0 (0,0)	5 (2,7)	184	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Belgium	14 (8,6)	45 (27,8)	62 (38,3)	0 (0,0)	41 (25,3)	162	3 (6,7)	12 (26,7)	14 (31,1)	5 (11,1)	11 (24,4)	45
China	0 (0,0)	0 (0,0)	0 (0,0)	1 (33,3)	2 (66,7)	3	0 (0,0)	8 (100)	0 (0,0)	0 (0,0)	0 (0,0)	8
Denmark	43 (8,3)	116 (22,4)	351 (67,6)	0 (0,0)	9 (1,7)	519	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Finland	83 (7,7)	108 (10,0)	862 (80,1)	0 (0,0)	23 (2,1)	1076	8 (30,8)	7 (26,9)	0 (0,0)	3 (11,5)	8 (30,8)	26
France	80 (7,0)	104 (9,1)	682 (59,4)	3 (0,3)	279 (24,3)	1148	17 (5,0)	34 (10,1)	71 (21,1)	143 (42,4)	72 (21,4)	337
Germany	81 (14,7)	167 (30,4)	249 (45,3)	0 (0,0)	53 (9,6)	550	20 (18,7)	31 (29,0)	11 (10,3)	11 (10,3)	34 (31,8)	107
Hungary	33 (4,5)	120 (16,3)	535 (72,9)	0 (0,0)	46 (6,3)	734	9 (11,8)	10 (13,2)	3 (3,9)	25 (32,9)	29 (38,2)	76
Italy	9 (7,0)	9 (7,0)	38 (29,7)	0 (0,0)	72 (56,3)	128	1 (4,2)	7 (29,2)	4 (16,7)	1 (4,2)	11 (45,8)	24
Japan	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	1 (100)	1	0 (0,0)	2 (100)	0 (0,0)	0 (0,0)	0 (0,0)	2
Netherlands	221 (5,3)	829 (19,7)	2516 (59,8)	1 (0,0)	642 (15,3)	4209	106 (10,3)	207 (20,2)	272 (26,5)	51 (5,0)	389 (38,0)	1025
New Zealand	71 (18,3)	102 (26,3)	47 (12,1)	0 (0,0)	168 (43,3)	388	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Russia	0 (0,0)	12 (92,3)	0 (0,0)	0 (0,0)	1 (7,7)	13	0 (0,0)	19 (21,3)	4 (4,5)	48 (53,9)	18 (20,2)	89
Slovenia	8 (13,3)	11 (18,3)	41 (68,3)	0 (0,0)	0 (0,0)	60	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
South Africa	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Spain	6 (4,2)	22 (15,5)	114 (80,3)	0 (0,0)	0 (0,0)	142	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Sweden	4 (16,7)	12 (50)	4 (16,7)	0 (0,0)	4 (16,7)	24	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
United Kingdom	21 (12,2)	11 (6,4)	140 (81,4)	0 (0,0)	0 (0,0)	172	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0

Supplementary Table 2B Number of reported norovirus ORF2 sequences stratified per genogroup/genotype, country, and time

	2005-2014					2015-2016				
Country	**GI (%)**	**GII other (%)**	**GII.4 (%)**	**GII.17 (%)**	**Total**	**GI (%)**	**GII other (%)**	**GII.17 (%)**	**GII.4 (%)**	**Total**
Austria	0 (0,0)	1 (50,0)	1 (50,0)	0 (0,0)	2	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Belgium	14 (12,2)	24 (20,9)	77 (67,0)	0 (0,0)	115	10 (26,3)	10 (26,3)	2 (5,3)	16 (42,1)	38
China	0 (0,0)	36 (37,1)	60 (61,9)	1 (1,0)	97	0 (0,0)	0 (0,0)	1 (2,2)	44 (97,8)	45
Denmark	28 (18,8)	73 (49,0)	47 (31,5)	1 (0,7)	149	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Finland	5 (17,2)	3 (10,3)	21 (72,4)	0 (0,0)	29	0 (0,0)	2 (100)	0 (0,0)	0 (0,0)	2
France	130 (9,5)	230 (16,8)	1007 (73,5)	3 (0,2)	1370	40 (9,6)	67 (16,1)	151 (36,2)	159 (38,1)	417
Germany	63 (17,9)	118 (33,6)	170 (48,4)	0 (0,0)	351	19 (17,6)	30 (27,8)	11 (10,2)	48 (44,4)	108
Hungary	2 (3,8)	26 (50,0)	24 (46,2)	0 (0,0)	52	0 (0,0)	0 (0,0)	1 (4,5)	21 (95,5)	22
Ireland	4 (3,4)	14 (11,8)	101 (84,9)	0 (0,0)	119	7 (17,9)	8 (20,5)	0 (0)	24 (61,5)	39
Italy	15 (10,4)	23 (16,0)	106 (73,6)	0 (0,0)	144	0 (0,0)	7 (25,9)	3 (11,1)	17 (63,0)	27
Japan	0 (0,0)	0 (0,0)	265 (100)	0 (0,0)	265	0 (0,0)	0 (0,0)	0 (0,0)	28 (100)	28
Netherlands	26 (4,8)	84 (15,6)	428 (79,6)	0 (0,0)	538	35 (8,3)	73 (17,4)	37 (8,8)	275 (65,5)	420
New Zealand	77 (18,4)	115 (27,5)	226 (54,1)	0 (0,0)	418	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Russia	6 (6,4)	58 (61,7)	30 (31,9)	0 (0,0)	94	17 (15,5)	18 (16,4)	49 (44,5)	26 (23,6)	110
Slovenia	7 (5,7)	20 (16,3)	96 (78,0)	0 (0,0)	123	0 (0,0)	0 (0,0)	3 (7,3)	38 (92,7)	41
South Africa	0 (0,0)	65 (33,3)	128 (65,6)	2 (1,0)	195	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
Spain	10 (7,2)	47 (34,1)	81 (58,7)	0 (0,0)	138	0 (0,0)	0 (0,0)	0 (0,0)	10 (100)	10
Sweden	65 (22,7)	116 (40,6)	103 (36,0)	2 (0,7)	286	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0
United Kingdom	37 (6,5)	76 (13,4)	456 (80,1)	0 (0,0)	569	0 (0,0)	11 (17,7)	0 (0,0)	51 (82,3)	62

Supplementary Table 3 First detections of global GII.4 drift variants

GII.4 variant	Year of emergence	First record ORF1	First ORF1 country	first record ORF2	First ORF2 country
Hunter 2004	2004	6-Apr-2004	The Netherlands	6-Apr-2004	The Netherlands
Den Haag 2006b	2006	14-Feb-2002	Germany	30-Sep-2003	Japan
New Orleans 2009	2009	12-Dec-2006	France	24-Apr-2009	South Africa
Sydney 2012	2012	-	-	Oct-2007	The Netherlands

Supplementary Table 4 ORF1 / ORF2 combinations (n=1047) detected by NoroNet 2005 - 2016

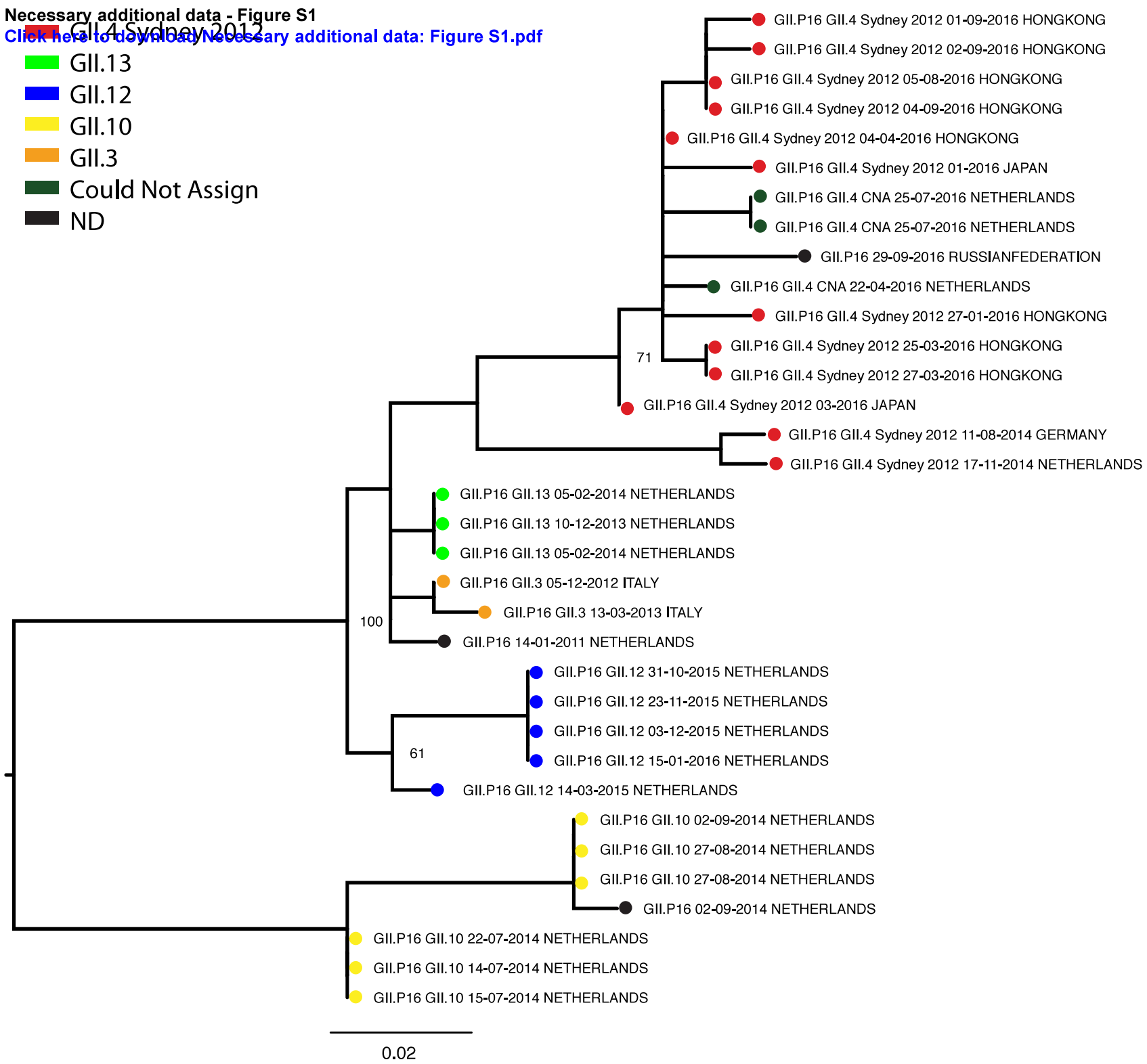
	GI.1	GI.2	GI.3	GI.4	GI.5	GI.6	GII.1	GII.2	GII.3	GII.4	GII.5	GII.6	GII.7	GII.10	GII.12	GII.13	GII.14	GII.17	Total
GI.P1	**9**																		9
GI.P2		**10**																	10
GI.P3			**26**																26
GI.P4				**15**															15
GI.P5					**9**														9
GI.P7			1																1
GI.Pb						9													9
GI.Pd			10																10
GII.P2								**12**			1								13
GII.P4										**441**									441
GII.P7												27	**9**				6		42
GII.P12									1	3									4
GII.P16									2	15				6	5	3			31
GII.P17																		**39**	39
GII.P21									63							2			65
GII.P22											2							1	3
GII.Pc							3												3
GII.Pe								2		301									303
GII.Pg							8								6				14
Total	9	10	37	15	9	9	11	14	66	760	3	27	9	6	11	5	6	40	1047

Supplementary Table 5 Amino acid (aa) comparison of the blockade epitopes A, D, and E between reference GII.Pe-GII.4 Sydney 2012 and novel GII.P16-GII.4 Sydney 2012 recombinant strains.

				A	A	A	A	A	A	D	D	D	E	E	E
Accession nr	**Sample location**	**Sample date**	**GII.4 ORF2 variant**	**294**	**296**	**297**	**298**	**368**	**372**	**393**	**394**	**395**	**407**	**412**	**413**
JX459908.1	Australia	Mar-12	GII.Pe - GII.4 Sydney 2012	T	S	R	N	E	D	G	T	T	S	N	T
JX459907.1	Australia	May-12	GII.Pe - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
Outbreak number	**Sample location**	**Sample date**	**Recombinant**	**294**	**296**	**297**	**298**	**368**	**372**	**393**	**394**	**395**	**407**	**412**	**413**
OH16002	Japan	Jan-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-886	Hong Kong	Jan-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
OC16023	Japan	Mar-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-937	Hong Kong	Mar-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-938	Hong Kong	Jul-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
5061600252	Netherlands	Jul-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
5061600253	Netherlands	Jul-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
5061600205	Netherlands	Apr-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	G	T	T	S	N	T
CUHK-NS-943	Hong Kong	Apr-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-1002	Hong Kong	Aug-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-1037	Hong Kong	Sep-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-1038	Hong Kong	Sep-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T
CUHK-NS-1044	Hong Kong	Sep-16	GII.P16 - GII.4 Sydney 2012	T	S	R	N	E	D	S	T	T	S	N	T

Necessary additional data - Figure S1

Click here to download Necessary additional data: Figure S1.pdf



Necessary additional data - Figure S2

Click here to download Necessary additional data: Figure S2.pdf