Elsevier Editorial System(tm) for The Lancet Infectious Diseases

Manuscript Draft

Manuscript Number: THELANCETID-D-17-00371

Title: Critical assessment of ten-year combined molecular and epidemiological norovirus surveillance of the NoroNet network, 2005 - 2016

Article Type: Articles (Original Research)

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

Corresponding Author: Mr. Janko van Beek, MSc

Corresponding Author's Institution: RIVM

First Author: Janko van Beek, MSc

Order of Authors: Janko van Beek, MSc; Miranda de Graaf; Haider Al-hello; David Allen; Katia Ambert-Balay; Nadine Botteldoorn; Mia Brytting; Javier Buesa; Maria Cabrerizo; Martin Chan; Fiona Cloak; Ilaria Di Bartolo; Susana Guix; Joanne Hewitt; Nobuhiro Iritani; Miao Jin; Reimar Johne; Ingeborg Lederer; Janet Mans; Vito Martella; Leena Maunula; Georgina McAllister; Sandra Niendorf; Hubert G. Niesters; Alexander T Podkolzin; Mateja Poljsak-Prijatelj; Lasse Dam Rasmussen; Gábor Reuter; Gráinne Tuite; Annelies Kroneman; Harry Vennema; Marion P.G. Koopmans

Manuscript Region of Origin: NETHERLANDS

Abstract: Background
Noroviruses are the most 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=16636 norovirus sequences with associated epidemiological metadata, shared between 2005 and 2016 through NoroNet by partners from Europe, Asia, Australia, and Africa.

Findings
We show continued global dominance and evolution of specific noroviruses, particularly of genotype GII.4, but with substantial regional differences possibly reflecting differences in epidemiology, susceptibility or both. The 2-3 year periodicity of emergence of 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. Currently used sequencing protocols rarely include the main epitopes on the viral

capsid, which will become important in view of ongoing vaccine development.

Interpretation
This study highlights the need for sustained norovirus surveillance, including assessment of possible immune escape and evolution by recombination in order to provide a full overview of norovirus epidemiology for future vaccine policy decisions.

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

Critical assessment of ten-year combined molecular and epidemiological norovirus surveillance of the NoroNet network, 2005 – 2016

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

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 National Reference Centre for Gastroenteritis Viruses, University Hospital of Dijon, Dijon, France

6 Scientific service of foodborne pathogens, Institute of Public Health, Brussels, Belgium

7 Microbial Typing Unit, The Public Health Agency of Sweden, Stockholm, Sweden
8 Viral Gastroenteritis Research Group, University of Valencia, Valencia, Spain
9 Enterovirus and Viral Gastroenteritis Unit, Instituto de Salud Carlos III, Madrid, Spain
10 Department of Microbiology, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China
11 GZV (Gastroenteric/Vectorborne/Zoonotic) Unit, Health Protection Surveillance Centre, Dublin, Ireland
12 Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanita, Rome, Italy
13 Enteric Virus Laboratory, University of Barcelona, Barcelona, Spain
14 Norovirus Reference Laboratory, Institute of Environmental Science and Research, Porirua, New Zealand
15 Department of Microbiology, Osaka City Institute of Public Health and Environmental Sciences, Osaka, Japan
16 Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, China CDC, Beijng, China
17 Department of Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany
18 Reference Centres and Reference Laboratories, Austrian Agency for Health and Food Safety, Vienna, Austria
19 Department of Medical Virology, University of Pretoria, Pretoria, South Africa
20 Department of Veterinary Medicine, University of Bari, Bari, Italy
21 Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland
22 Specialist Virology Centre, Royal Infirmary Edinburgh, Edinburgh, United Kingdom
23 Consultant Laboratory for Noroviruses, Robert Koch Institute, Berlin, Germany
24 Medical Microbiology, Division of Clinical Virology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
25 Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor), Central Research Institute of Epidemiology, Moscow, Russia

26 Institute of Microbiology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
27 Department of Virus & Microbiological Special Diagnostics, Statens Serum Institut, Copenhagen, Denmark
28 Department of Medical Microbiology and Immunology, University of Pécs, Pécs, Hungary
29 National Virus Reference Laboratory, University College Dublin, Dublin, Ireland
30 http://www.noronet.nl

[BOX] Research in context

Evidence before this study

Norovirus is a major cause of acute gastroenteritis causing high disease burden and related costs globally. While numerous studies cited in PubMed report on norovirus genetic diversity in a limited setting, geographic area, or timeframe, studies presenting the long-term global norovirus diversity trends are scarce.

Added value of this study

This study reports trends in norovirus genetic diversity combined with epidemiological metadata, obtained from reports from 19 countries across four continents shared through a jointly owned database. It shows the continued dominance and evolution of specific noroviruses, but with substantial regional differences possibly reflecting differences in epidemiology, susceptibility or both. We critically assess current norovirus global 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

This study highlights the need for optimized protocols in order to have full use of the available system for international data sharing, which was developed when sequencing was less accessible. We recommend to explore the use of novel next-generation sequencing techniques combined with the development of standardised approaches to measuring immunity to novel norovirus variants, to provide data on antigenic evolution and escape to herd immunity needed for vaccine updates.

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, incidentally resulting in death[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 capsid proteins (the external surface of the virus particles), and of genes encoding the viral polymerase. 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]. It escapes population immunity by altering the antigenicity of the capsid surface in a stepwise manner – a process called epochal evolution[3]. In addition, frequent mixing 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 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 herd immunity, hence requiring improved coverage of surveillance[14]. In addition, the rapid development of novel methods for genomic epidemiology may change the surveillance landscape, opening up opportunities for higher coverage of surveillance, for instance when next-generation sequencing is implemented in clinical laboratories, or environmental laboratories that may monitor which viruses circulate in communities[15]. In view of these developments, we analysed how data obtained via the NoroNet surveillance network can be used to address the following outstanding questions regarding the impact of virus evolution on norovirus disease:

1. What are the trends in norovirus reporting and genomic diversity?
2. Is there evidence for differences by genotype in time, region, setting and mode of transmission?
3. Where do new variants of norovirus emerge?
4. Can emerging variants be predicted from globally linked surveillance data?
5. What is the impact of recombination on the reported trends?

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[8,10,16]. 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

All sequences were derived from human faecal samples. 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 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 without sample date or with a sample date prior to 2005 were removed from the analysis. A subset of the data was selected based on the availability of data needed for each individual analysis. Norovirus sequences were genotyped by the online norovirus typing tool[17]. Sequences overlapping the ORF1/ORF2 for which ORF1 and ORF2 genotypes could be assigned were analysed separately.

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 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, Australia, and Africa (Table S1).

Number of reported sequences, sequence length and genome position

The total number of reported sequences fluctuated from 429 to 1403 per year for ORF1 and from 188 to 893 for ORF2 and has increased in recent years, especially for ORF2 sequences (Figure 1A and 1B). 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 2). 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. 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, indicated by separate lines in Figure 1A.

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). Countries in Asia and South Africa only reported GII strains while other countries showed a GI proportion up to 22·3%. Overall, 1372 of 16636 (8·2%) sequences belong to norovirus GI, 15256 of 16636 (91·7%) sequences belong to GII. Additionally, eight sequences were submitted belonging to GIII or GIV (n=1, n=7, respectively). Trends per genotype per year for GI and GII are shown in Figures 1A and 1B. The most consistently and commonly detected genotypes 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 or with increased prevalence in some years.

Emergence of novel GII.17 genotype

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[18,19]. NoroNet detected a sharp increase in the number of GII.P17 and GII.17 strains in 2015 - 2016 (Figure 1A and 1B). Asian countries (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 (data not shown). GII.P17 and GII.17 were widely detected among several 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). The GII.P17 / GII.17 fraction is smaller than GII.Pe / GII.P4 / GII.4 in the majority of European countries except for France (ORF1) and Russia (ORF1 and ORF2). Nevertheless, the viruses

were co-circulating with GII.4 strains that remained the most commonly reported genotype, (data not shown).

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 a 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).

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 1). 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 are at low levels present 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 sides could be genotyped by the norovirus genotyping tool. 477 of 1047 (45·6%) sequences were assigned as a recombinant strain (Table 2). 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 for a separate clade in a maximum likelihood tree inferred from partial GII.P16 sequences (Figure 4). 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 do not contain any amino acid changes that have not been observed in 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 (Supplementary Figure 2 and 3).

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

The European norovirus season coincides with the Northern Hemisphere winter season (Figure 5a). 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 5b). 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 5c). Australia (reported by New Zealand) shows highest incidence in October and November (Figure 5d).

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 6A). 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 6B). 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 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[6,10,18,20]. GII.4 Sydney 2012 is the predominantly detected variant 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. While there was no evidence for antigenic evolution of the GII.4 Sydney capsids, it is remarkable that these capsids seem to be circulating as recombinant forms, suggesting that recombination somehow favours virus maintenance in the population. The driving forces for this phenomenon are currently unknown. It is possible that some recombinant viruses have increased fitness due to differences in replication and/or transmission efficiency[21]. For GII.4, recombination has only been with closely related genotypes GII.Pe and GII.P12, which are both suggested to be derived from an ancestor of GII.P4[22]. The novel recombinant GII.P16-GII.4 Sydney 2012 may allow further exploration of antigenic diversity to escape host immunity in the future.

Norovirus surveillance is done on a voluntary basis since funding for the network is unavailable. This is reflected by an unstable submission behaviour of many countries. Unstandardized sampling and 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 potential use is the identification of international

outbreaks, which have been observed during periods of sustained funding[4,23]. 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 protocols. To overcome this problem, we are exploring the use of next generation sequencing to allow whole genome sequencing as a new standard. 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 protein. A protocol for sequencing this particular region has been described[24].

Norovirus vaccine candidates are currently tested 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 most likely need to be updated on a regular basis due to the ability of the virus to escape herd immunity by altering the epitopes on the surface of the capsid protein, especially by the predominant GII.4[25]. 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. Enhanced capsid surveillance or enhanced full genome sequencing via next generation sequencing techniques could provide a better insight in the evolution of the virus and could provide data to monitor the antigenic distance between the future vaccine and circulating strains.

One of the major questions within the norovirus research field is whether we are capable of predicting emerging variants in the near future. All the recent major drift variants were already circulating years before they became dominant, suggesting early warning surveillance for variant emergence would be possible. If we assume that new variants develop in the human population and could emerge anywhere in the word, as shown by this study, this would require a surveillance system with global coverage including large-scale genomics to capture both capsid diversity and recombination. 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

None of the authors declared any conflict 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).

Figure descriptions

Figure 1 Number of reported ORF1 (A) and ORF2 (B) sequences (n=11252 and n=6423, respectively) stratified per genotype group, genotype, and year. Note that n=1047 sequences overlapping ORF1/ORF2 are counted for both ORF1 and ORF2.

Figure 2 Position of n=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).

Figure 3 GII.4 variant trends per year for ORF1 (top) and ORF2 (bottom).

Figure 4 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). 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[26]. GII.P16-GII.4 Sydney 2012 sequences are indicated in red.

Figure 5 Norovirus seasonality patterns in Europe (A) (n=13935), Africa (B) (n=195), Asia (C) (n=262), Australia (D) (n=806) stratified per genotype group.

Figure 6 Norovirus transmission route (n=8772) (A) and suspected outbreak setting (n=6446) (B) stratified per genotype group.

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. Nordahl Petersen T, Rasmussen S, Hasman H, et al. Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Sci Rep* 2015; **5**: 11444.
16. 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.
17. Kroneman A, Vennema H, Deforche K, et al. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 2011; **51**(2): 121-5.

18. 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).
19. 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.
20. 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.
21. Arias A, Thorne L, Ghurburrun E, Bailey D, Goodfellow I. Norovirus Polymerase Fidelity Contributes to Viral Transmission In Vivo. *mSphere* 2016; **1**(5).
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. 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.
24. 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.
25. Parra GI, Squires RB, Karangwa CK, et al. Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity. *PLoS Pathog* 2017; **13**(1): e1006136.
26. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003; **52**(5): 696-704.

Tables

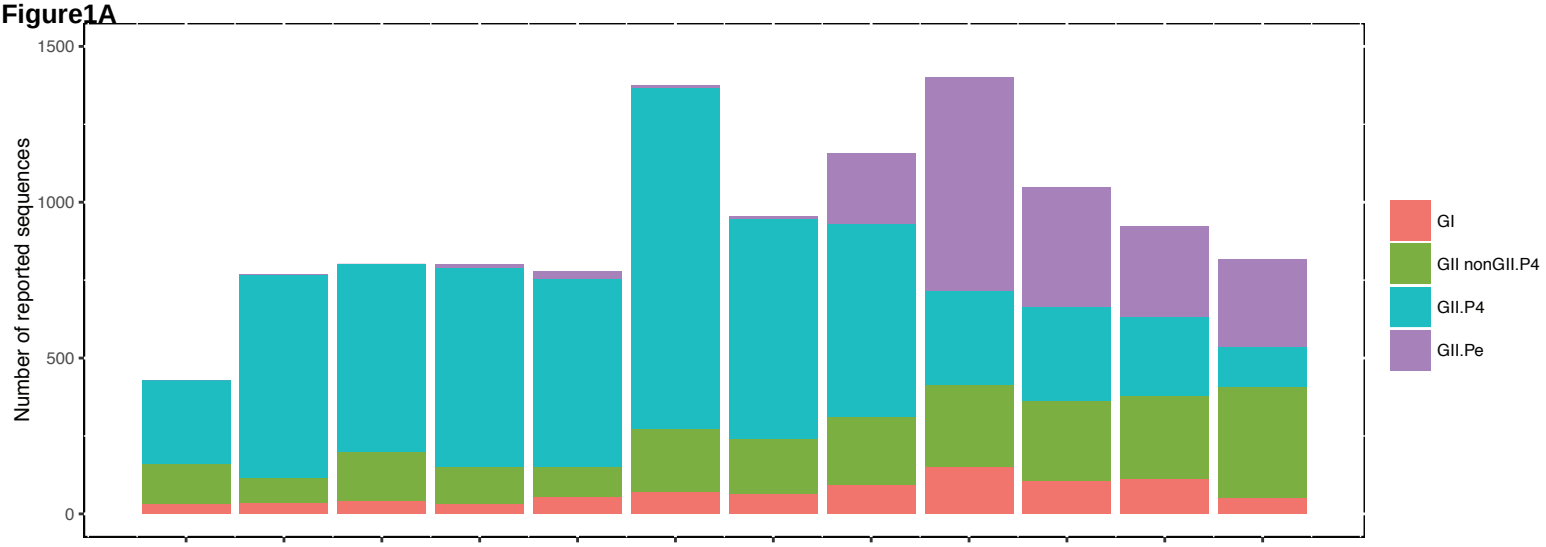
Table 1 First detections of global GII.4 drift variants

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

Table 2 ORF1 / ORF2 combinations (n=1047) detected by NoroNet 2005 - 2016

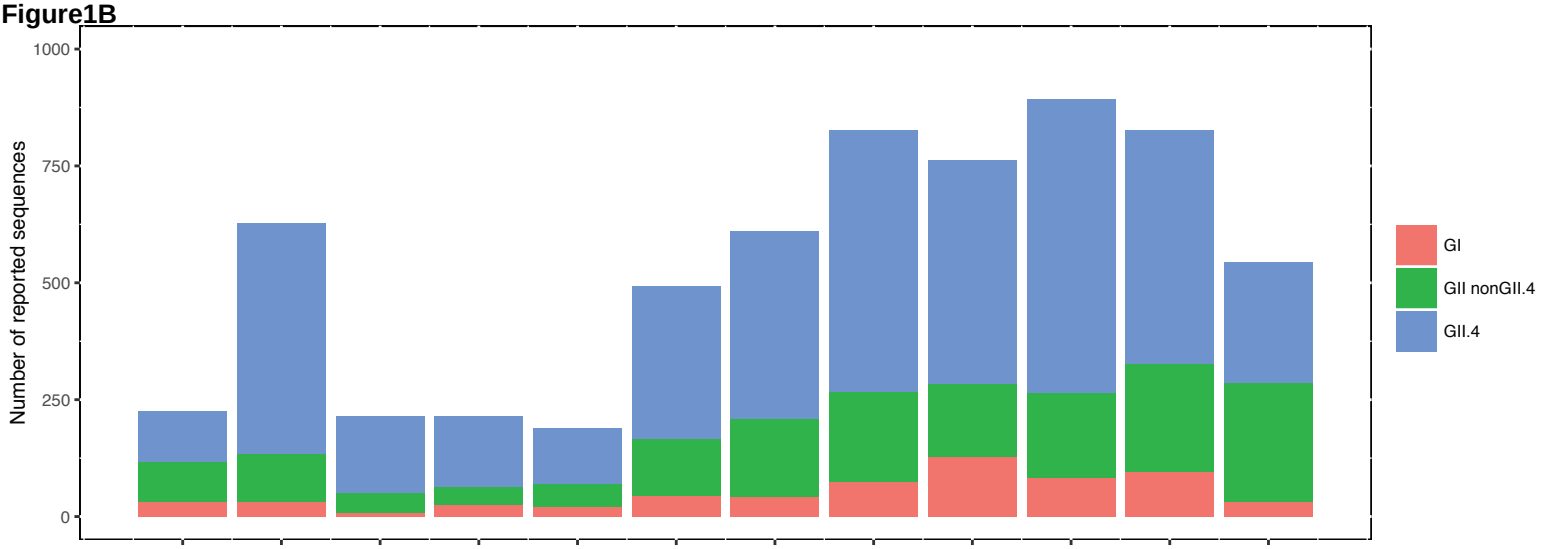
	GI.1	GI.2	GI.3	GI.4	GI.5	GI.6	GII.1	GII.2	GII.3	GII.4	GII.5	GII.6	GII.7	GII.10	GII.12	GII.13	GII.14	GII.17	Total
GI.P1	**9**																		9
GI.P2		**10**																	10
GI.P3			**26**																26
GI.P4				**15**															15
GI.P5					**9**														9
GI.P7			1																1
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

Figure1A



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

Figure1B



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

Figure2

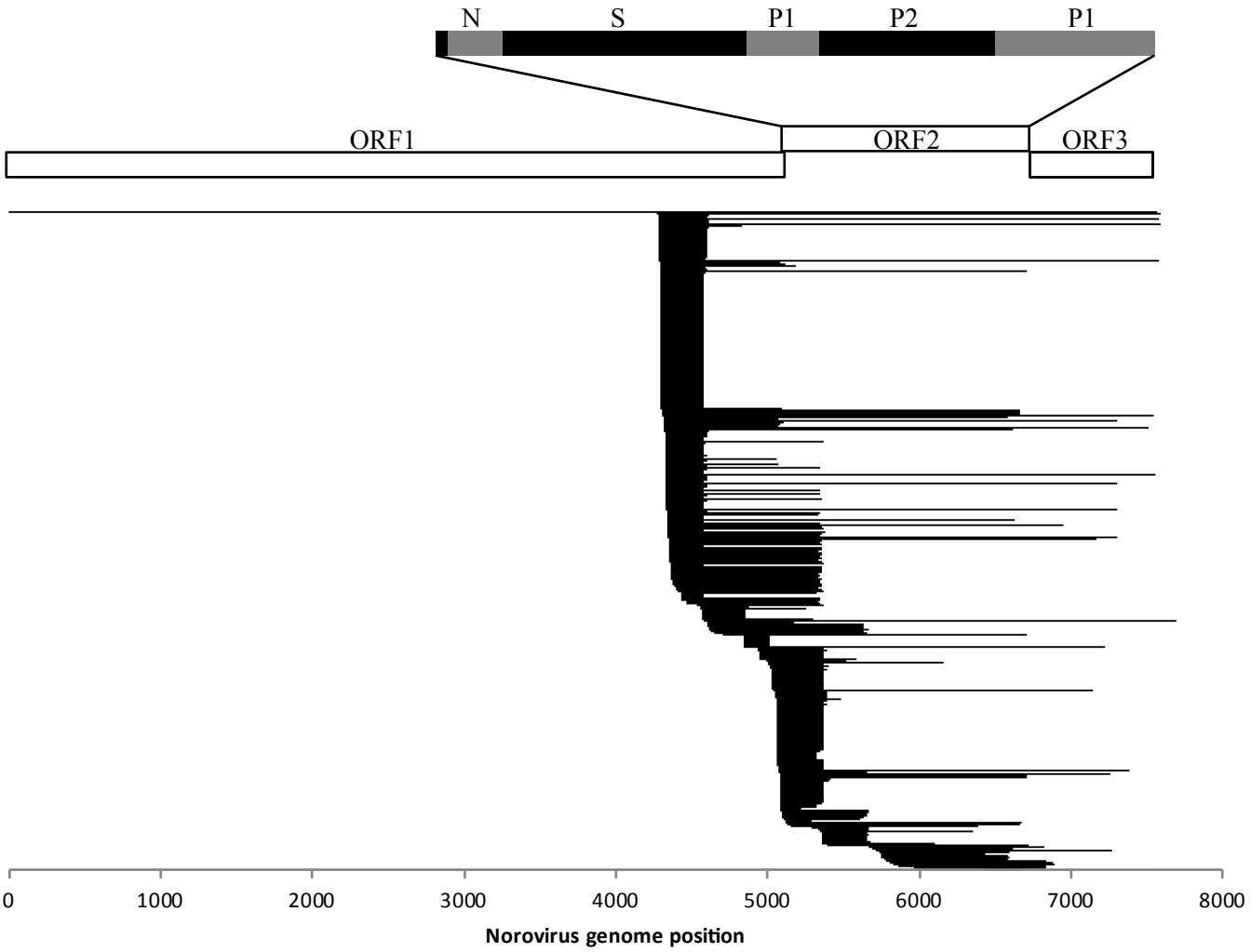

N
S
P1
P2
P1
ORF1
ORF2
ORF3
0
1000
2000
3000
4000
5000
6000
7000
8000
Norovirus genome position

Figure3

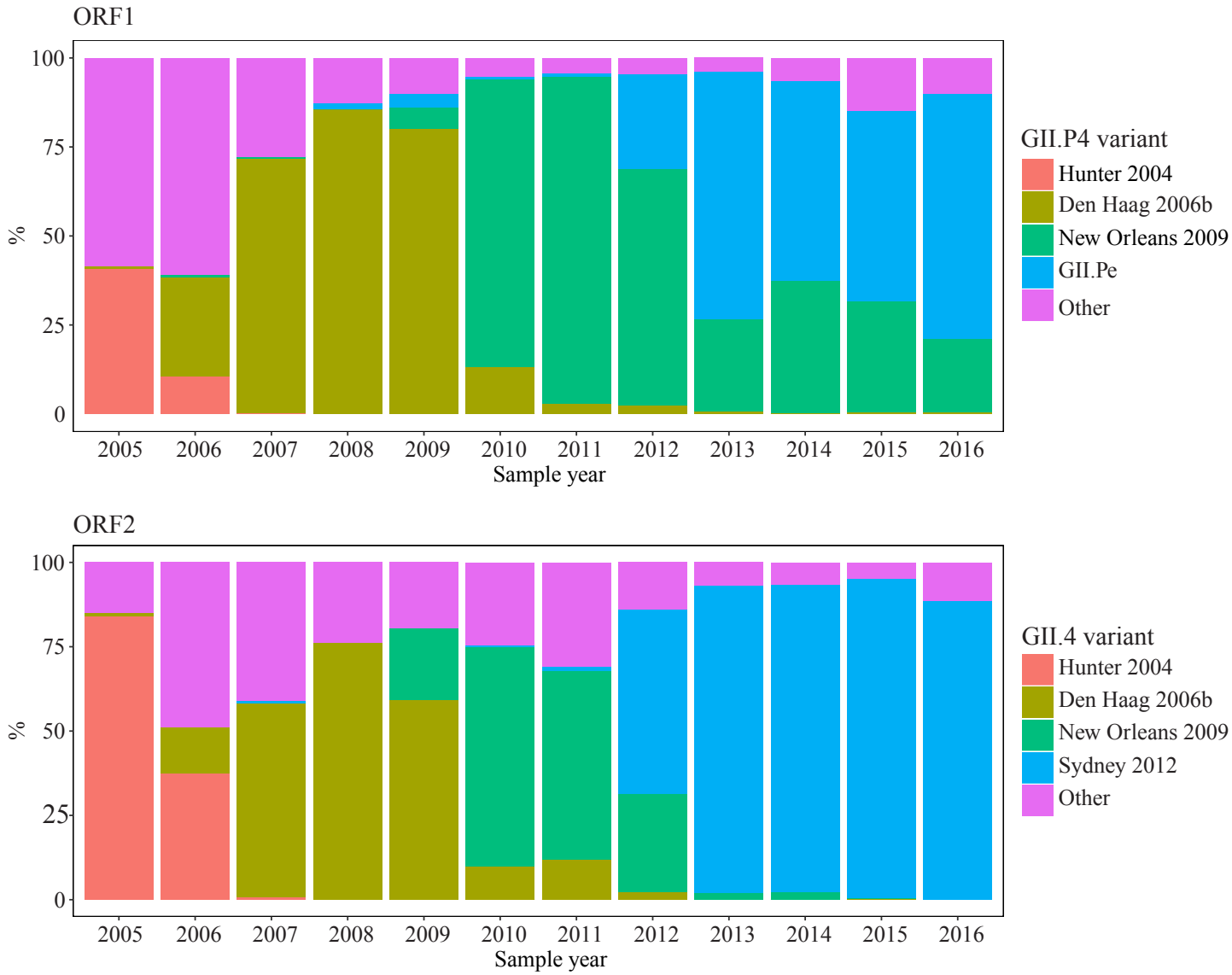

ORF1
100
75
50
25
0
%
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Sample year
GII.P4 variant
Hunter 2004
Den Haag 2006b
New Orleans 2009
GII.Pe
Other
ORF2
100
75
50
25
0
%
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Sample year
GII.4 variant
Hunter 2004
Den Haag 2006b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

Figure4

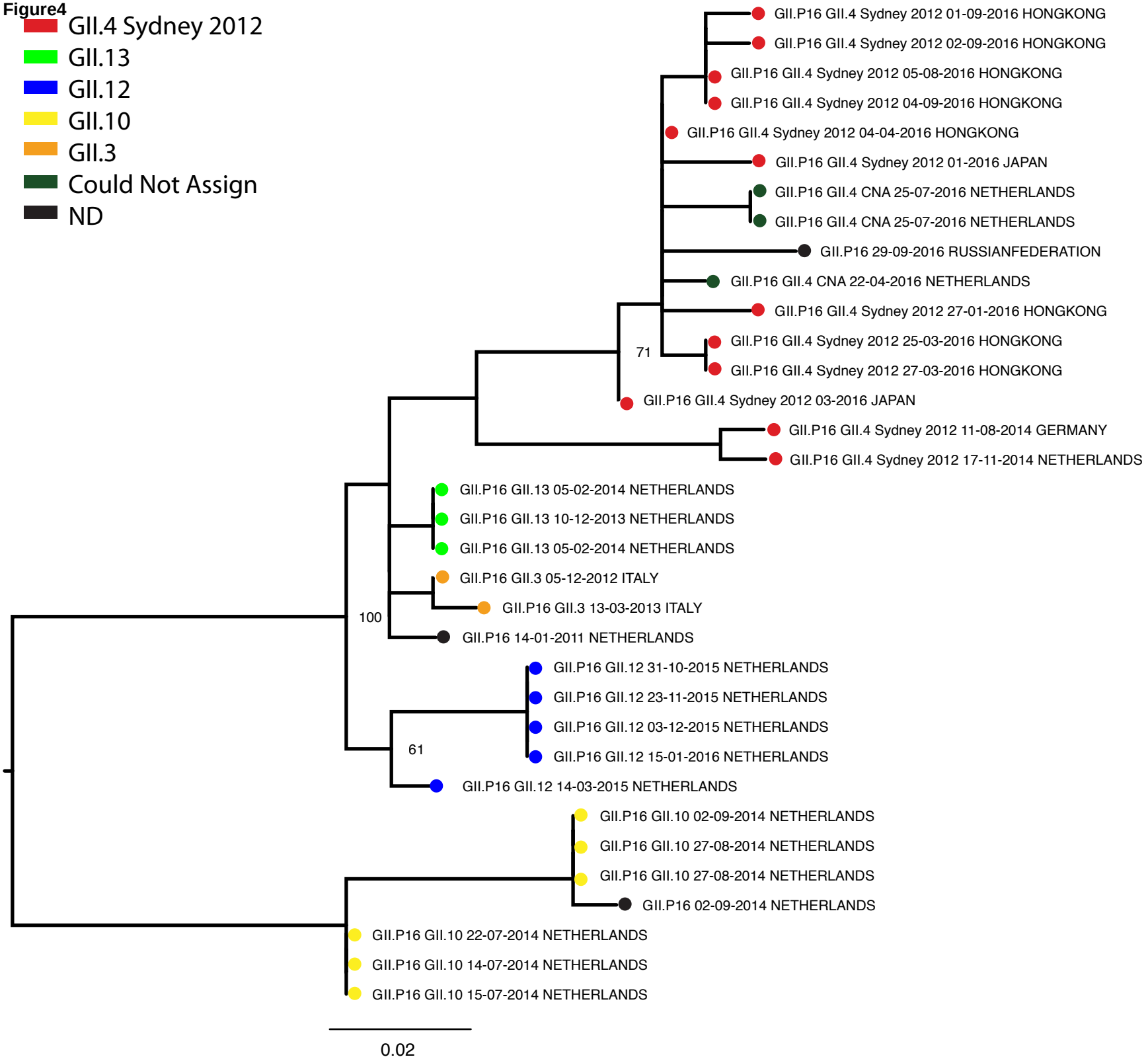

GII.4 Sydney 2012
GII.13
GII.12
GII.10
GII.3
Could Not Assign
ND
GII.P16 GII.4 Sydney 2012 01-09-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 02-09-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 05-08-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 04-09-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 04-04-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 01-2016 JAPAN
GII.P16 GII.4 CNA 25-07-2016 NETHERLANDS
GII.P16 GII.4 CNA 25-07-2016 NETHERLANDS
GII.P16 29-09-2016 RUSSIANFEDERATION
GII.P16 GII.4 CNA 22-04-2016 NETHERLANDS
GII.P16 GII.4 Sydney 2012 27-01-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 25-03-2016 HONGKONG
GII.P16 GII.4 Sydney 2012 27-03-2016 HONGKONG
71
GII.P16 GII.4 Sydney 2012 03-2016 JAPAN
GII.P16 GII.4 Sydney 2012 11-08-2014 GERMANY
GII.P16 GII.4 Sydney 2012 17-11-2014 NETHERLANDS
GII.P16 GII.13 05-02-2014 NETHERLANDS
GII.P16 GII.13 10-12-2013 NETHERLANDS
GII.P16 GII.13 05-02-2014 NETHERLANDS
GII.P16 GII.3 05-12-2012 ITALY
GII.P16 GII.3 13-03-2013 ITALY
100
GII.P16 14-01-2011 NETHERLANDS
GII.P16 GII.12 31-10-2015 NETHERLANDS
GII.P16 GII.12 23-11-2015 NETHERLANDS
GII.P16 GII.12 03-12-2015 NETHERLANDS
GII.P16 GII.12 15-01-2016 NETHERLANDS
61
GII.P16 GII.12 14-03-2015 NETHERLANDS
GII.P16 GII.10 02-09-2014 NETHERLANDS
GII.P16 GII.10 27-08-2014 NETHERLANDS
GII.P16 GII.10 27-08-2014 NETHERLANDS
GII.P16 02-09-2014 NETHERLANDS
GII.P16 GII.10 22-07-2014 NETHERLANDS
GII.P16 GII.10 14-07-2014 NETHERLANDS
GII.P16 GII.10 15-07-2014 NETHERLANDS
0.02

Figure5

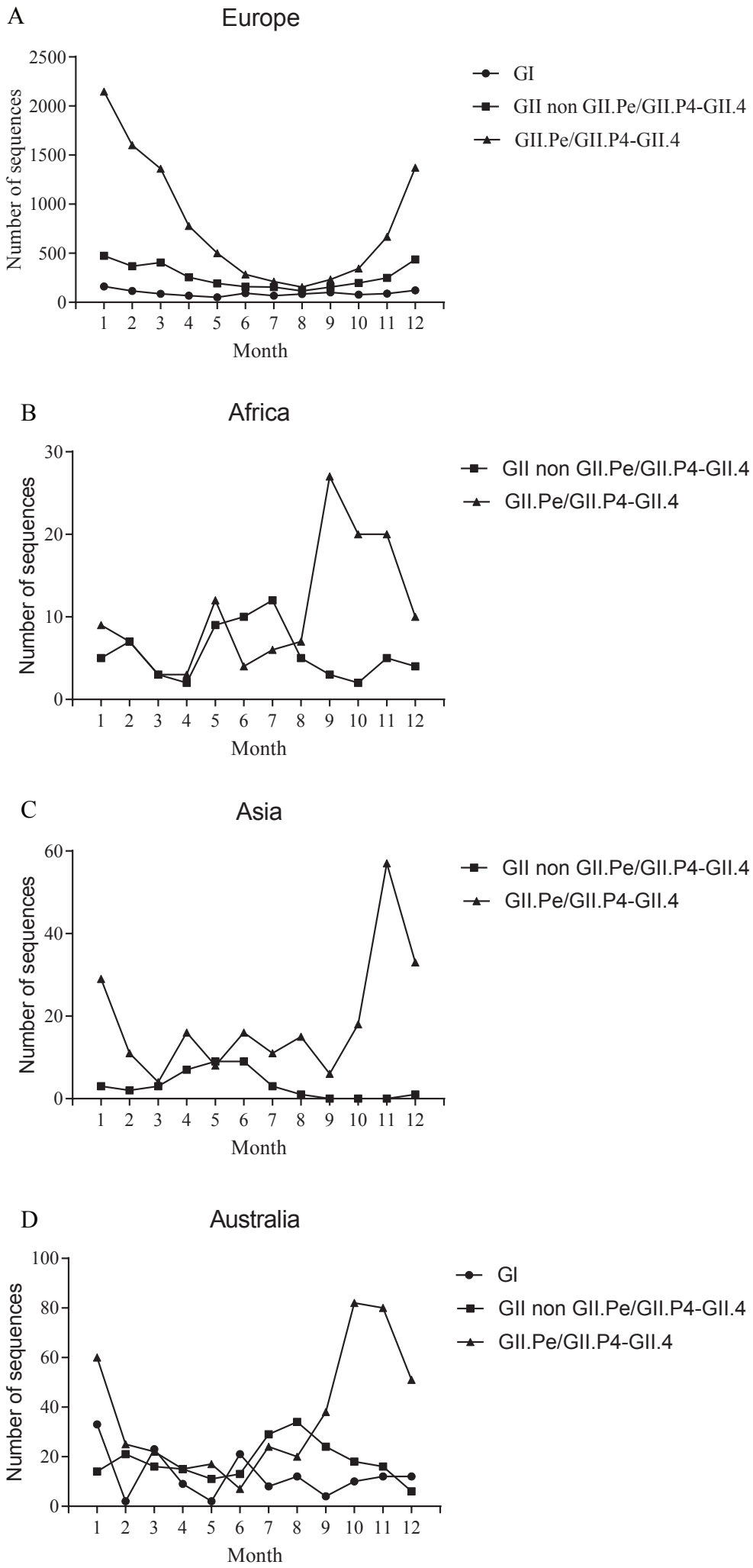
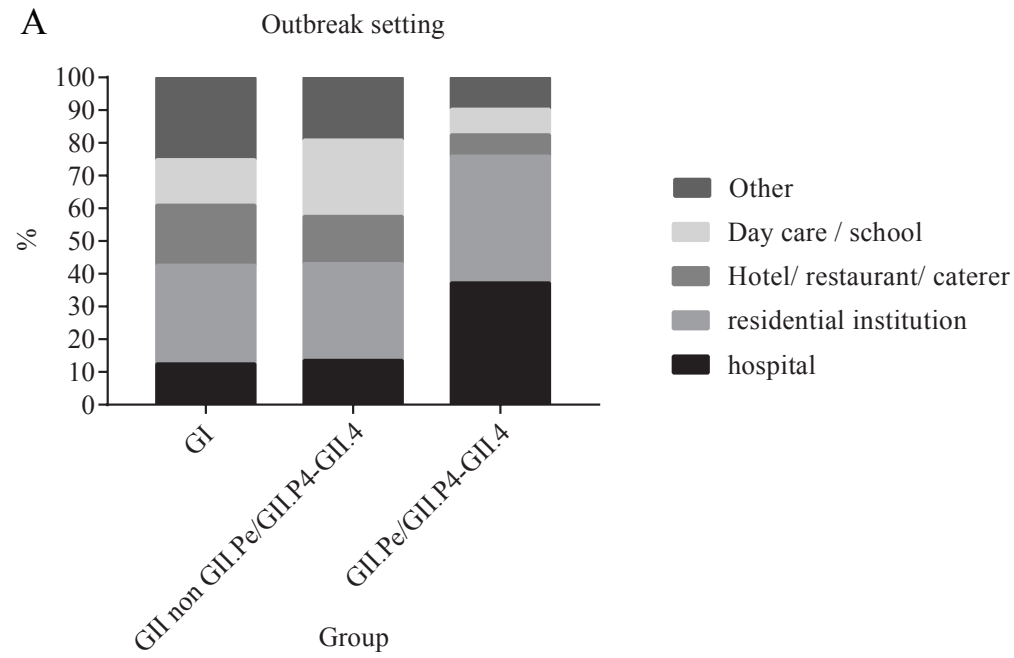
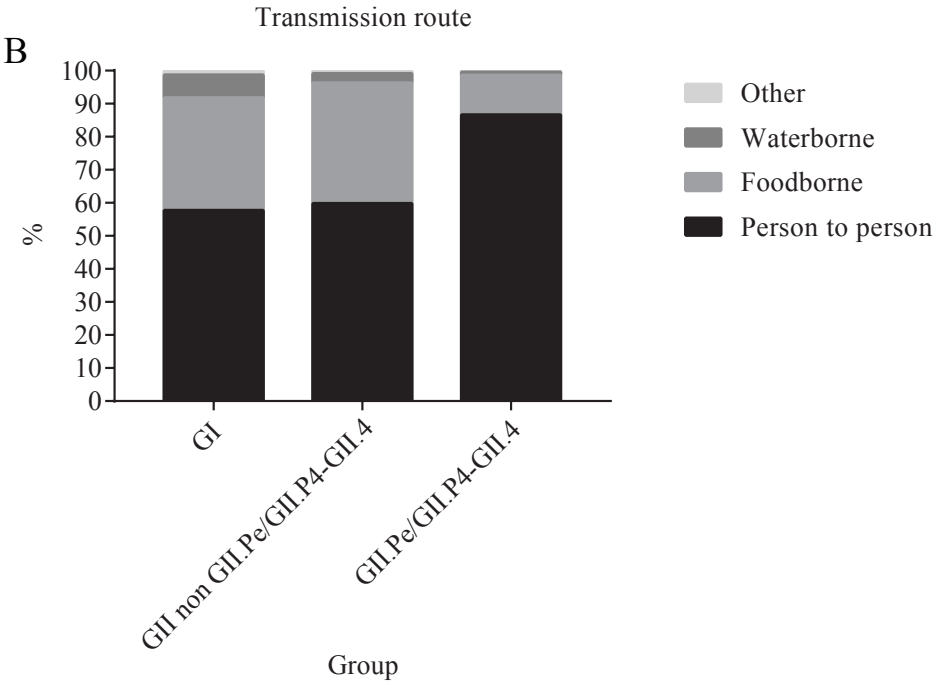

A
Europe
Number of sequences
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GI
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
B
Africa
Number of sequences
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12
Month
GII non GII.Pe/GII.P4-GII.4
GII.Pe/GII.P4-GII.4
C
Asia
Number of sequences
60
40
20
0
1 2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

Figure6


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 Table 1
Click here to download Necessary additional data: Supplementary table 1.docx

Supplementary Table 1 Number of reported GI and GII sequences per continent 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	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
Europe	Russia	23 (7,5)	283 (92,5)	306

Supplementary Figure 2

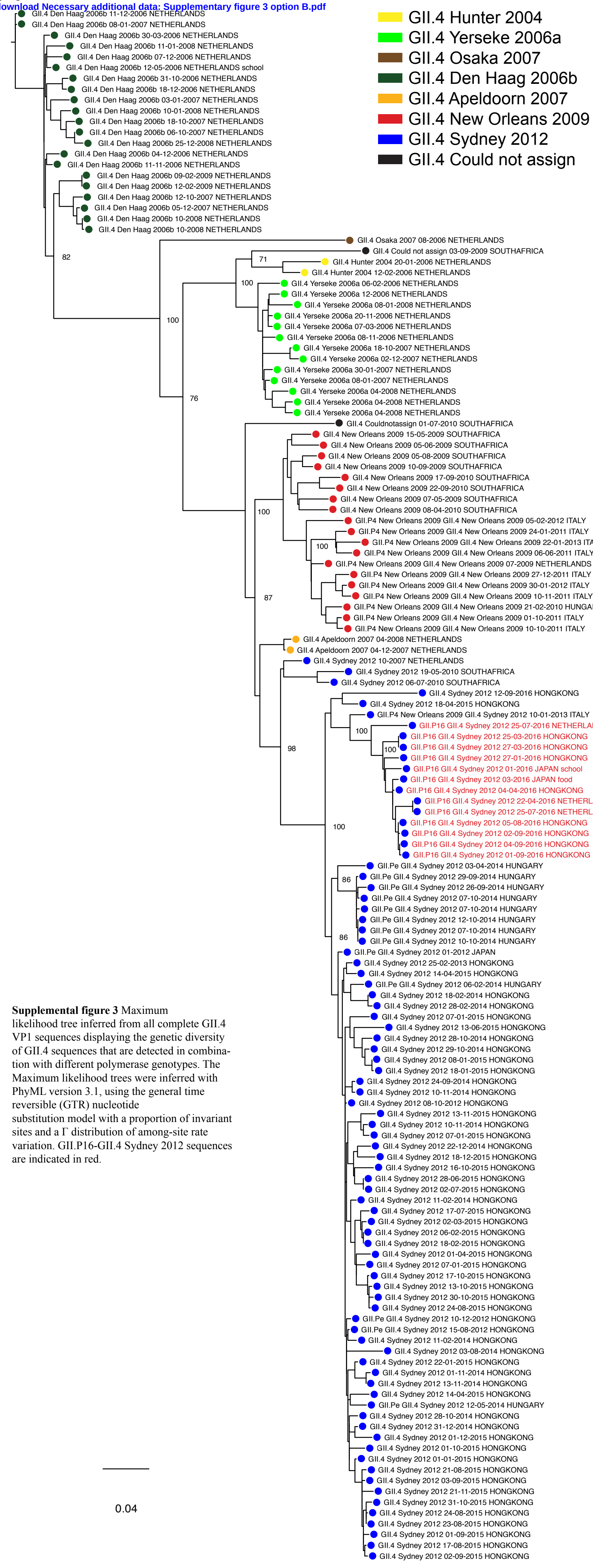
Click here to download Necessary additional data: Supplementary figure 2.pdf

				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

Supplementary figure 2 Amino acid (aa) comparison of the blockade epitopes A, D, and E between GII.Pe-GII.4 Sydney 2012 reference strains (Genbank JX459908.1 and JX459907.1) and novel GII.P16-GII.4 Sydney 2012 recombinant. Epitope A consists of aa 294, 296, 297, 298, 368, and 372, epitope D of aa 393-395, and epitope E of 407, 412, and 413.

Supplementary Figure 3

Click here to download Necessary additional data: Supplementary figure 3 option B.pdf



Supplemental figure 3 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. 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. GII.P16-GII.4 Sydney 2012 sequences are indicated in red.