van Beek, Janko; de Graaf, Miranda; Al-Hello, Haider; Allen, David J; Ambert-Balay, Katia; Botteldoorn, Nadine; Brytting, Mia; Buesa, Javier; Cabrero, Maria; Chan, Martin; +23 more... Cloak, Fiona; Di Bartolo, Ilaria; Guix, Susana; Hewitt, Joanne; Iritani, Nobuhiko; Jin, Miao; Johne, Reimar; Lederer, Ingeborg; Mans, Janet; Martella, Vito; Maunula, Leena; McAllister, Georgina; Niendorf, Sandra; Niesters, Hubert G; Podkolzin, Alexander T; Poljsak-Prijatelj, Mateja; Rasmussen, Lasse Dam; Reuter, Gábor; Tuite, Gráinne; Kroneman, Annelies; Vennema, Harry; Koopmans, Marion PG; NoroNet; (2018) Molecular surveillance of norovirus, 2005-16: an epidemiological analysis of data collected from the NoroNet network. The Lancet infectious diseases, 18 (5). pp. 545-553. ISSN 1473-3099 DOI: https://doi.org/10.1016/S1473-3099(18)30059-8

Downloaded from: http://researchonline.lshtm.ac.uk/id/eprint/4647631/

DOI: https://doi.org/10.1016/S1473-3099(18)30059-8

Usage Guidelines:

Please refer to usage guidelines at https://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
Analysis of norovirus molecular surveillance data collected through the NoroNet network, 2005 – 2016


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, Beijing, 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
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.
[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. 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. 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. 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. The epidemiology and human health impact are strongly shaped by norovirus evolution through recombination or accumulation of mutations, known as genetic drift. 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. In 2002, an informal international data sharing network was established to study noroviruses and their diversity in relation to human health impact. 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. 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. 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. 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\textsuperscript{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\textsuperscript{8,10,15}. The network started as EU funded network in 1999, continuing since 2002 as global NoroNet\textsuperscript{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\textsuperscript{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 $\Gamma$ distribution of among-site rate variation\textsuperscript{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,
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 others6,10,18,19.

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 vaccines20. It was previously noted that increased notification reflect true 
increases in disease trends18,21. 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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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.
Nnorovirus 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. 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. 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. 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\textsuperscript{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\textsuperscript{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\textsuperscript{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 COMPAR (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 GIIPe-GIIL.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 GIIPe variant trends per year (n=8083, top) and ORF2 GIIL.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


34. Allen DJ, Trainor E, Callaghan A, O'Brien SJ, Cunliffe NA, Iturriza-Gomara M. Early Detection of Epidemic GI.4 Norovirus Strains in UK and Malawi: Role of
649 Surveillance of Sporadic Acute Gastroenteritis in Anticipating Global Epidemics. 
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 
Figure 1
<table>
<thead>
<tr>
<th>Year</th>
<th>GI Pe</th>
<th>GIL P17</th>
<th>GIL P4</th>
<th>GIL other</th>
<th>GI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>4.0 (9)</td>
<td>0.0</td>
<td>2.0 (2)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2006</td>
<td>3.0 (7)</td>
<td>13.0 (7)</td>
<td>5.0 (6)</td>
<td>3.0 (4)</td>
<td>0.0</td>
</tr>
<tr>
<td>2007</td>
<td>9.0 (2)</td>
<td>7.0 (9)</td>
<td>17.0 (2)</td>
<td>18.0 (2)</td>
<td>16.0 (2)</td>
</tr>
<tr>
<td>2008</td>
<td>5.0 (6)</td>
<td>5.0 (6)</td>
<td>33.0 (2)</td>
<td>32.0 (3)</td>
<td>21.0 (2)</td>
</tr>
<tr>
<td>2009</td>
<td>5.0 (7)</td>
<td>7.0 (7)</td>
<td>10.0 (9)</td>
<td>10.0 (9)</td>
<td>34.0 (2)</td>
</tr>
<tr>
<td>2010</td>
<td>2.0 (2)</td>
<td>2.0 (2)</td>
<td>17.0 (6)</td>
<td>17.0 (6)</td>
<td>17.0 (2)</td>
</tr>
<tr>
<td>2011</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2012</td>
<td>3.0 (3)</td>
<td>0.0</td>
<td>6.0 (4)</td>
<td>4.0 (4)</td>
<td>7.0 (8)</td>
</tr>
<tr>
<td>2013</td>
<td>0.0</td>
<td>0.0</td>
<td>4.0 (5)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2014</td>
<td>2.0 (2)</td>
<td>0.0</td>
<td>2.0 (2)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2015</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2016</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Total: 429 (100) | 767 (100) | 803 (100) | 802 (100) | 777 (100) | 1374 (100) | 955 (100) | 1155 (100) | 1403 (100) | 1404 (100) | 922 (100) | 817 (100)
| Year | GL.1 | GL.2 | GL.3 | GL.4 | GL.5 | GL.6 | GL.7 | GL.8 | GL.9 | GIL.1 | GIL.2 | GIL.3 | GIL.4 | GIL.5 | GIL.6 | GIL.7 | GIL.8 | GIL.9 | GIL.10 | GIL.11 | GIL.12 | GIL.13 | GIL.14 | GIL.15 | GIL.16 | GIL.17 | GIL.18 | GIL.19 | GIL.20 | GIL.21 |
|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|      | 3 (1.3) | 8 (1.3) | 4 (1.6) | 9 (1.5) | 5 (1.5) | 6 (1.8) | 7 (1.3) | 10 (1.8) | 11 (1.8) | 12 (1.8) | 13 (1.8) | 14 (1.8) | 15 (1.8) | 16 (1.8) | 17 (1.8) | 18 (1.8) | 19 (1.8) | 20 (1.8) | 21 (1.8) | 22 (1.8) | 23 (1.8) | 24 (1.8) | 25 (1.8) | 26 (1.8) | 27 (1.8) | 28 (1.8) | 29 (1.8) | 30 (1.8) | 31 (1.8) | 32 (1.8) |
| 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 4

A. Europe

- GI
- GII.non GII.Pe/GII.P4-GII.4
- GII.Pe/GII.P4-GII.4

B. Africa

- GII.non GII.Pe/GII.P4-GII.4
- GII.Pe/GII.P4-GII.4

C. Asia

- GII.non GII.Pe/GII.P4-GII.4
- GII.Pe/GII.P4-GII.4

D. Oceania

- GI
- GII.non GII.Pe/GII.P4-GII.4
- GII.Pe/GII.P4-GII.4
Figure 5

A. Outbreak setting

- **Group**: GI, GI Pe/GH P4, GI Pe/GH P4-GH.4
- **Outbreak setting**:
  - **Other**: 
  - **Day care / school**: 
  - **Hotel/ restaurant/ caterer**: 
  - **Residential institution**: 
  - **Hospital**: 

B. Transmission route

- **Group**: GI, GI Pe/GH P4, GI Pe/GH P4-GH.4
- **Transmission route**:
  - **Other**: 
  - **Waterborne**: 
  - **Foodborne**: 
  - **Person to person**: 

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

<table>
<thead>
<tr>
<th>Continent</th>
<th>Country</th>
<th>GI (%)</th>
<th>GII (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>Austria</td>
<td>6 (3.2)</td>
<td>180 (96.8)</td>
<td>186</td>
</tr>
<tr>
<td>Europe</td>
<td>Belgium</td>
<td>41 (11.4)</td>
<td>319 (88.6)</td>
<td>360</td>
</tr>
<tr>
<td>Asia</td>
<td>China</td>
<td>0 (0)</td>
<td>142 (100)</td>
<td>142</td>
</tr>
<tr>
<td>Europe</td>
<td>Denmark</td>
<td>67 (10.4)</td>
<td>580 (89.6)</td>
<td>647</td>
</tr>
<tr>
<td>Europe</td>
<td>Finland</td>
<td>96 (8.5)</td>
<td>1037 (91.5)</td>
<td>1133</td>
</tr>
<tr>
<td>Europe</td>
<td>France</td>
<td>267 (8.2)</td>
<td>3004 (91.8)</td>
<td>3271</td>
</tr>
<tr>
<td>Europe</td>
<td>Germany</td>
<td>183 (16.4)</td>
<td>932 (83.6)</td>
<td>1115</td>
</tr>
<tr>
<td>Europe</td>
<td>Hungary</td>
<td>43 (5.2)</td>
<td>791 (94.8)</td>
<td>834</td>
</tr>
<tr>
<td>Europe</td>
<td>Ireland</td>
<td>11 (7)</td>
<td>147 (93)</td>
<td>158</td>
</tr>
<tr>
<td>Europe</td>
<td>Italy</td>
<td>23 (7.7)</td>
<td>276 (92.3)</td>
<td>299</td>
</tr>
<tr>
<td>Asia</td>
<td>Japan</td>
<td>0 (0)</td>
<td>293 (100)</td>
<td>293</td>
</tr>
<tr>
<td>Europe</td>
<td>Netherlands</td>
<td>327 (6)</td>
<td>5100 (94)</td>
<td>5427</td>
</tr>
<tr>
<td>Australia</td>
<td>New Zealand</td>
<td>148 (18.4)</td>
<td>658 (81.6)</td>
<td>806</td>
</tr>
<tr>
<td>Europe</td>
<td>Russia</td>
<td>23 (7.5)</td>
<td>283 (92.5)</td>
<td>306</td>
</tr>
<tr>
<td>Europe</td>
<td>Slovenia</td>
<td>15 (6.7)</td>
<td>209 (93.3)</td>
<td>224</td>
</tr>
<tr>
<td>Africa</td>
<td>South Africa</td>
<td>0 (0)</td>
<td>195 (100)</td>
<td>195</td>
</tr>
<tr>
<td>Europe</td>
<td>Spain</td>
<td>16 (5.5)</td>
<td>274 (94.5)</td>
<td>290</td>
</tr>
<tr>
<td>Europe</td>
<td>Sweden</td>
<td>69 (22.3)</td>
<td>241 (77.7)</td>
<td>310</td>
</tr>
<tr>
<td>Europe</td>
<td>United Kingdom</td>
<td>37 (5.9)</td>
<td>595 (94.1)</td>
<td>632</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GI (%)</td>
<td>GL other (%)</td>
<td>GILF (%)</td>
<td>GILP (%)</td>
</tr>
<tr>
<td>Austria</td>
<td>6 (3.3)</td>
<td>14 (8.2)</td>
<td>184 (5.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Belgium</td>
<td>18 (2.8)</td>
<td>14 (5.2)</td>
<td>162 (3.1)</td>
<td>3 (0.0)</td>
</tr>
<tr>
<td>China</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>2 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Denmark</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>4 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Finland</td>
<td>45 (8.3)</td>
<td>16 (2.4)</td>
<td>116 (2.1)</td>
<td>6 (0.0)</td>
</tr>
<tr>
<td>France</td>
<td>45 (8.3)</td>
<td>16 (2.4)</td>
<td>116 (2.1)</td>
<td>6 (0.0)</td>
</tr>
<tr>
<td>Germany</td>
<td>83 (7.7)</td>
<td>108 (10.0)</td>
<td>116 (1.1)</td>
<td>1076 (8.3)</td>
</tr>
<tr>
<td>Hungary</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Italy</td>
<td>45 (4.5)</td>
<td>31 (3.9)</td>
<td>76 (3.5)</td>
<td>24 (2.6)</td>
</tr>
<tr>
<td>Japan</td>
<td>9 (7.9)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>221 (5.3)</td>
<td>829 (19.7)</td>
<td>1250 (5.9)</td>
<td>4209 (10.3)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>71 (18.3)</td>
<td>102 (26.3)</td>
<td>167 (2.1)</td>
<td>106 (0.3)</td>
</tr>
<tr>
<td>Norway</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Russia</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>South Africa</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Spain</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Sweden</td>
<td>41 (4.7)</td>
<td>12 (1.5)</td>
<td>24 (0.0)</td>
<td>172 (0.0)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>21 (2.2)</td>
<td>11 (1.4)</td>
<td>140 (1.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Country</td>
<td>GI (%)</td>
<td>GII other (%)</td>
<td>GII.4 (%)</td>
<td>GII.17 (%)</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>---------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>Austria</td>
<td>0 (0,0)</td>
<td>1 (50,0)</td>
<td>1 (50,0)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Belgium</td>
<td>14 (12,2)</td>
<td>24 (20,9)</td>
<td>77 (67,0)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>China</td>
<td>0 (0,0)</td>
<td>36 (37,1)</td>
<td>60 (61,9)</td>
<td>1 (1,0)</td>
</tr>
<tr>
<td>Denmark</td>
<td>28 (18,8)</td>
<td>73 (49,0)</td>
<td>47 (31,5)</td>
<td>1 (0,7)</td>
</tr>
<tr>
<td>Finland</td>
<td>5 (17,2)</td>
<td>3 (10,3)</td>
<td>21 (72,4)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>France</td>
<td>130 (9,5)</td>
<td>230 (16,8)</td>
<td>1007 (73,5)</td>
<td>3 (0,2)</td>
</tr>
<tr>
<td>Germany</td>
<td>63 (17,9)</td>
<td>118 (33,6)</td>
<td>170 (48,4)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Hungary</td>
<td>2 (3,8)</td>
<td>26 (50,0)</td>
<td>24 (46,2)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Ireland</td>
<td>4 (3,4)</td>
<td>14 (11,8)</td>
<td>101 (84,9)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Italy</td>
<td>15 (10,4)</td>
<td>23 (16,0)</td>
<td>106 (73,6)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Japan</td>
<td>0 (0,0)</td>
<td>0 (0,0)</td>
<td>265 (100)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>26 (4,8)</td>
<td>84 (15,6)</td>
<td>428 (79,6)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>77 (18,4)</td>
<td>115 (27,5)</td>
<td>226 (54,1)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Russia</td>
<td>6 (6,4)</td>
<td>58 (61,7)</td>
<td>30 (31,9)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Slovenia</td>
<td>7 (5,7)</td>
<td>20 (16,3)</td>
<td>96 (78,0)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>South Africa</td>
<td>0 (0,0)</td>
<td>65 (33,3)</td>
<td>128 (65,6)</td>
<td>2 (1,0)</td>
</tr>
<tr>
<td>Spain</td>
<td>10 (7,2)</td>
<td>47 (34,1)</td>
<td>81 (58,7)</td>
<td>0 (0,0)</td>
</tr>
<tr>
<td>Sweden</td>
<td>65 (22,7)</td>
<td>116 (40,6)</td>
<td>103 (36,0)</td>
<td>2 (0,7)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>37 (6,5)</td>
<td>76 (13,4)</td>
<td>456 (80,1)</td>
<td>0 (0,0)</td>
</tr>
</tbody>
</table>
### Supplementary Table 3 First detections of global GIL4 drift variants

<table>
<thead>
<tr>
<th>GIL4 variant</th>
<th>Year of emergence</th>
<th>First record ORF1</th>
<th>First ORF1 country</th>
<th>first record ORF2</th>
<th>First ORF2 country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunter 2004</td>
<td>2004</td>
<td>6-Apr-2004</td>
<td>The Netherlands</td>
<td>6-Apr-2004</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>Den Haag 2006b</td>
<td>2006</td>
<td>14-Feb-2002</td>
<td>Germany</td>
<td>30-Sep-2003</td>
<td>Japan</td>
</tr>
<tr>
<td>New Orleans 2009</td>
<td>2009</td>
<td>12-Dec-2006</td>
<td>France</td>
<td>24-Apr-2009</td>
<td>South Africa</td>
</tr>
<tr>
<td>Sydney 2012</td>
<td>2012</td>
<td>-</td>
<td>-</td>
<td>Oct-2007</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>GI1</td>
<td>GI2</td>
<td>GI3</td>
<td>GI4</td>
<td>GI5</td>
<td>GI6</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>GLP1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP2</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP3</td>
<td></td>
<td></td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP4</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>GLP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>GLP7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLPb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLPd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.P22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.Pc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.Pe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIH.Pg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>10</td>
<td>37</td>
<td>15</td>
<td>9</td>
</tr>
</tbody>
</table>
### Supplementary Table 5

<table>
<thead>
<tr>
<th>Accession nr</th>
<th>Sample location</th>
<th>Sample date</th>
<th>GII.4 ORF2 variant</th>
<th>Recombinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>JX459908.1</td>
<td>Australia</td>
<td>Mar-12</td>
<td>GII.Pe - GII.4 Sydney 2012</td>
<td>294 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>JX459907.1</td>
<td>Australia</td>
<td>May-12</td>
<td>GII.Pe - GII.4 Sydney 2012</td>
<td>296 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>OH16002</td>
<td>Japan</td>
<td>Jan-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>297 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-886</td>
<td>Hong Kong</td>
<td>Jan-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>298 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>OC16023</td>
<td>Japan</td>
<td>Mar-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>299 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-937</td>
<td>Hong Kong</td>
<td>Mar-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>300 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-938</td>
<td>Hong Kong</td>
<td>Jul-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>301 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>5061600252</td>
<td>Netherlands</td>
<td>Jul-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>302 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>5061600253</td>
<td>Netherlands</td>
<td>Jul-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>303 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>5061600205</td>
<td>Netherlands</td>
<td>Apr-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>304 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-943</td>
<td>Hong Kong</td>
<td>Apr-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>305 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-1002</td>
<td>Hong Kong</td>
<td>Aug-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>306 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-1037</td>
<td>Hong Kong</td>
<td>Sep-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>307 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-1038</td>
<td>Hong Kong</td>
<td>Sep-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>308 T S R N E D G T T S N T</td>
</tr>
<tr>
<td>CUHK-NS-1044</td>
<td>Hong Kong</td>
<td>Sep-16</td>
<td>GII.P16 - GII.4 Sydney 2012</td>
<td>309 T S R N E D G T T S N T</td>
</tr>
</tbody>
</table>