A Study of Norovirus Molecular Epidemiology

Impact, Prevalence, Diversity and Genetic Adaptation

J. Joukje Siebenga

ISBN: 978-90-6464-415-3

The research described in this thesis was conducted at the National Institute for Public Health and the Environment, The Netherlands, and was funded by the European Commission, DG Research Quality of Life Program, under the 6th Framework (EVENT; SP22-CT-2004-502571)

Printed by Ponsen & Looijen B.V., Wageningen.
Printing of this thesis was supported by the RIVM and the Erasmus University.

The cover was created using images graciously provided by Dr. Kjell-Olof Hedlund, and the word cloud was created using www.wordle.net.

A Study of Norovirus Molecular Epidemiology

Impact, Prevalence, Diversity and Genetic Adaptation

Een studie naar de moleculaire epidemiologie van norovirus

Impact, voorkomen, diversiteit en genetische adaptatie

Proefschrift
ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 17 november 2010 om 15.30 uur door:

Jannigje Joukje Siebenga

geboren te Noordoostpolder

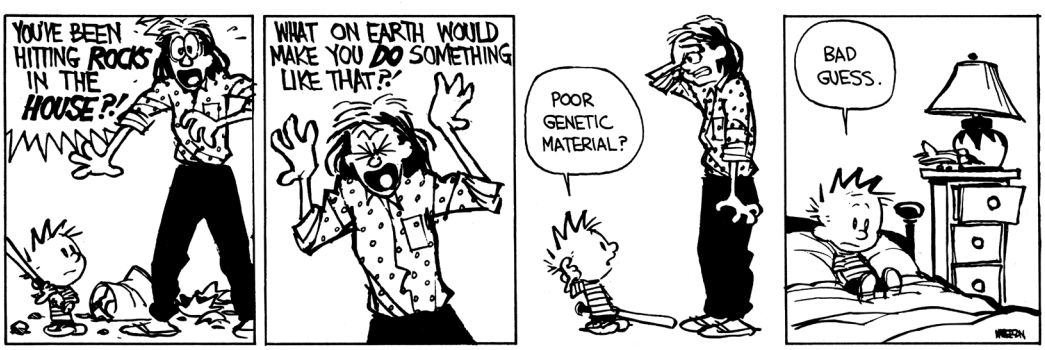
Promotie Commissie

Promotor:
Prof.dr. M.P.G. Koopmans

Overige leden:
Prof.dr. A.D.M.E. Osterhaus
Prof.dr. G.F. Rimmelzwaan
Prof.dr. D.J. Smith

Copromotor:
Dr. H. Vennema

Voor papa en mama



Contents

Chapter 1

General Introduction to Norovirus

Adapted from: Cpt. 1 Norovirus Epidemiology
In: Caliciviruses: Molecular and Cellular Virology
By: J. Joukje Siebenga, Erwin Duizer and Marion P.G. Koopmans
Edited by: Grant S. Hansman, Jason Jiang and Kim Y. Green

Halfway into the past century molecular biology emerged as a science, and it has evolved and grown very rapidly since. As a result of their very nature, infectious diseases have always had the attention of physicians and scientists, and the possibilities for research offered by molecular biology were pre-eminently suitable for studying the causative agents of these infectious diseases, and especially for studying viruses.
Noroviruses cause sudden onset gastro-intestinal illness in humans. Their ability to cause large scale outbreaks of debilitating illness, even if of quickly passing nature, has made them into a relevant topic of study. Rapid developments in molecular biology have generated diagnostic tools which enabled many labs in the world to perform norovirus diagnostics, and to perform sequence analyses of the detected strains. These sequence data form a very valuable basis for studies of the molecular epidemiology of noroviruses. By linking the sequence data to classical epidemiological data, such as time and place of illness, number of people affected, etc., a platform is made for unveiling information describing for example the spread of the virus, the impact, and the prevalence. The work presented in this thesis aims to further current knowledge of norovirus by studying its molecular epidemiology, to better enable taking public health actions aimed at decreasing the impact of disease.

Setting the scene for the main causative agent of viral gastro-intestinal illness

Norovirus is one of the five recognized genera of the family *Caliciviridae*, together with *Lagovirus* (known host species: rabbits, brown hares), *Sapovirus* (known host species: humans, swine, mink) and *Vesivirus* (known host species: sea lions, seals, other marine mammals, swine, cats, dogs, fish, cattle and primates) and the recently described *Nebovirus* genus [222] (known host species: cattle). Additionally, two other potential genera have been described, comprised of viruses detected in rhesus macaques [77], and in swine [161,327]. The animal caliciviruses have been associated with a range of clinical syndromes, including lesions, stomatitis, upper respiratory tract infection, and systemic diseases with severe hemorrhagic syndromes.
Noroviruses infect humans, but have also been detected in swine, cattle, sheep, mice, cats, lions and dogs [118,137,190,191,221,281,325]. In humans and animals, noroviruses cause acute gastroenteritis, mild gastroenteritis and asymptomatic infections. In immunocompromized mice, lacking components of the innate system, murine norovirus causes encephalitis, vasculitis, pneumonia and hepatitis, indicating a broad tissue tropism [137].
Human norovirus infection is popularly known as 'winter vomiting disease', because of the observation that disease outbreaks follow a pattern of winter-seasonality: outbreaks in the Northern hemisphere are most common between November and March. In the Southern hemisphere a similar seasonal pattern has been observed in certain countries (Australia) [303], but not in others (New Zealand) [270]. Illness caused by noroviruses is also known as

the ‘gastric flu’ or ‘stomach flu’, and recent new findings on the evolution of these viruses show that this popular name - while unintended - may be more accurate than previously thought [142,269]. Outbreaks with high media-impact among vacationers on cruise ships have also yielded the name ‘cruise ship virus’.
The unavailability of a cell culture system has been a continuing handicap in advancing knowledge of the human noroviruses [69]. The breakthrough cloning and sequencing of the prototype strain, Norwalk virus [125], in 1990, led to the development of molecular tools to study noroviruses in more detail. Since then, numerous studies have shown the importance of noroviruses as a cause of illness in people of all age groups.

History

In 1968, students and teachers of an elementary school in Norwalk, Ohio, fell ill with symptoms of acute gastroenteritis, including vomiting and diarrhea. For this illness, which had previously been described as ‘winter vomiting disease’ [4], no causative agent was found. Several attempts to identify an etiology for this infectious form of acute gastroenteritis failed, until in 1972 Albert Kapikian *et al.* described viral particles found in stool samples of volunteers experimentally infected with purified stool filtrate from a patient originally infected in the Norwalk outbreak [135], thus fulfilling Koch’s postulates. Using immune electron microscopy, small “picorna- or parvovirus-like” particles were detected, that elicited immune responses in volunteers as well as naturally infected individuals. In 1982, Kaplan *et al.* listed four criteria to identify a norovirus outbreak [136]. These criteria were i) vomiting in more than half the affected persons; ii) mean (or median) incubation period of 24-48 h; iii) mean (or median) duration of illness lasting 12-60 h; and iv) absence of a bacterial pathogen in stool cultures. Recently, these criteria were reassessed, and still found to be relatively accurate, with 68% sensitivity and 99% specificity [304].

Factors that shape the success of noroviruses in the human population

Genome properties
The noroviruses contain a positive-sense RNA genome that is approximately 7.5 kb long, covalently linked to VPg at the 5’end and polyadenylated at the 3’end (figure 1.1A) [125]. The genome encodes three open reading frames (ORFs) [127]. ORF1 encodes a polyprotein comprising all non-structural proteins, which is autocatalytically cleaved to produce the non-structural proteins. ORF2 encodes the major capsid protein, referred to in the literature as VP1. ORF3 encodes a second, minor structural protein, VP2, about which little is known. In feline calicivirus it was shown to be essential for productive replication [275], although this was not the case for RHDV [171]. Using recombinant baculovirus constructs it was

shown that Norwalk Virus VP2 expression regulates both the expression and stability of VP1 [15].

The norovirus RNA genome is replicated by the virus-encoded RNA-dependent RNA polymerase, which has no proofreading activity. The high error-rate in genomic replication causes the production of many progeny viruses with mutations (quasispecies), as other RNA viruses, making noroviruses highly flexible and increasingly diverse. Noroviruses can be divided into distinct genogroups, based on phylogenetic analyses of the capsid protein. To date, five norovirus genogroups (G) have been recognized (GI-GV). Viruses of GI, GII and GIV are known to infect humans [94,341]. GII viruses have additionally been detected in pigs [104,195,195,280,281,325,326,332] and GIV viruses have been detected in a lion cub and a dog [190,191]. GIII viruses infect cattle and sheep [332] and GV viruses infect mice [118,137]. GI, GII and GIII are further subdivided into an increasing number of genotypes; currently 8, 19 and 2 genotypes are recognized [93]. It is likely that additional viruses belonging to novel genotypes exist in humans and animals, which may be found when more research is done. The genetic variability of the norovirus genome and the regular discovery of novel genotypes have caused some inconsistencies in the nomenclature used for norovirus genotypes. A striking example is the Alphatron strain, which is classified as a GIV.1 (genogroup IV genotype 1) virus in most typing schemes [93,341], but as a GII.17 strain in another paper [131]. Similarly, GII.13 and GII.14 have been interchangeably used in different publications, as did GII.15 and GII.16 [131,341]. Recombinant norovirus strains cause additional naming difficulties, since they cannot be adequately named with the currently used nomenclature system, which is based solely on the capsid gene (ORF2). The development of a norovirus nomenclature scheme that includes genotyping designations for both ORF1 and ORF2 is being explored by a small group of laboratories involved in molecular surveillance. After completion of this discussion a typing tool applying the agreed-upon nomenclature will be available from http://www.rivm.nl/mpf/norovirus/typingtool.

Biochemical properties of the virion

Noroviruses are particularly infectious and highly stable in the environment. These properties are determined by specific characteristics of the virion. The virus particles are about 28-32 nm in diameter and have no lipid envelope (figure 1.1B) [135]. The viral capsid is made up of 180 copies of the ORF2 encoded protein, grouped as dimers, in a T=3 symmetry. The capsid protein is generally divided into the N-terminal part, the Shell domain, which forms the actual shell, and the Protruding domain, divided into P1 and P2 (figure 1.1C, and 1.1D). The Protruding domain is connected to the Shell domain by a flexible hinge region, providing the Protruding domain with flexibility, thus allowing it some freedom to change

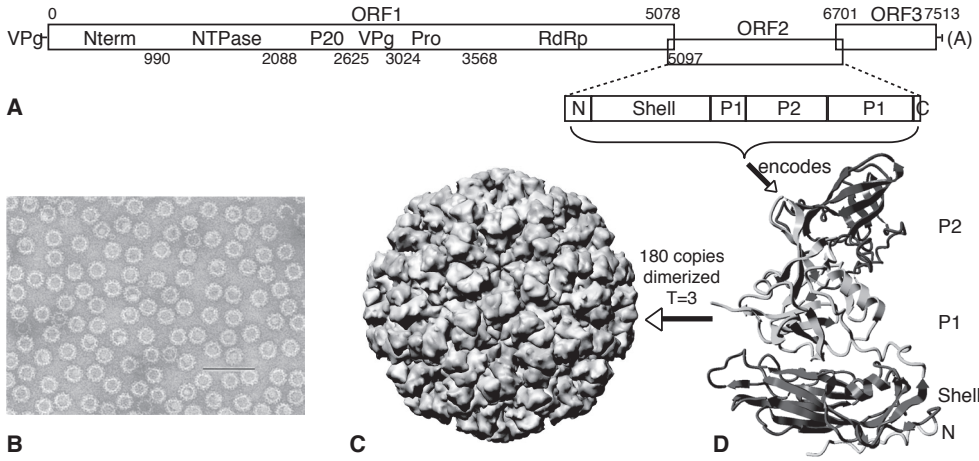


Figure 1.1. Norovirus genome and structure. A) Schematic representation of the norovirus genome. Numbering is based on the GII.4 genome. The non-structural proteins encoded in ORF1 are depicted, as are the different domains in ORF2. B) EM picture of Norwalk Virus capsid protein, self assembles into virus like particles. The bar is 100 nm. Picture from [126]. C) The norovirus particle is built from 180 copies of the capsid protein (90 dimers), and structured in a T=3 symmetry. Picture from [240]. D) Ribbon representation of the GII.4 capsid protein, modeled on the Norwalk virus model [241,269]. The N-terminal domain is shown in yellow, the Shell domain in red, the Protruding domain, split up into P1 and P2, in cyan and blue.

shape, resulting of mutations in the P domain.

These particles have been shown to be highly stable in the environment [66], although effective testing of the stability of human noroviruses has been hampered by the lack of an infectivity assay [69]. Inactivation studies have been performed using feline and canine caliciviruses as models for human norovirus. More recently the cultivable murine norovirus has been adopted as the surrogate of choice, since it is more closely related to the human noroviruses.

Studies using feline and canine caliciviruses as models for norovirus infections showed that heat inactivation of these two animal caliciviruses was highly comparable. Thermal inactivation is both temperature and time dependent with increasing survival at decreasing temperatures [66]. The noroviruses are preserved well by refrigeration and freezing. Similarly, inactivation by UV-radiation is dose and time dependent [51,214,294]. Feline calicivirus was not efficiently inactivated (defined by a > 4log10 reduction) on environmental surfaces or in suspension by 1% anionic detergents, quaternary ammonium (1:10), hypochlorite solutions with < 300 ppm free chlorine, or by different alcohol preparations (ethanol or 1- and 2-propanol) [66,89,262]. Varying efficacies of 65 to 75% alcohol are reported, however, short contact times (< 1 min) rarely result in effective inactivation. Moreover, the presence of fecal or other organic material reduces the virucidal efficacy of many chemicals tested. The efficacy of seven commercial disinfectants for the inactivation of feline calicivirus on

strawberries and lettuce was shown to be inadequate when used at the FDA permitted concentration [101]. These data were confirmed by Allwood and co-workers for sodium bicarbonate, chlorine bleach, peroxyacetic acid and hydrogen peroxide at FDA approved concentrations [7]. More recently, with murine norovirus as a surrogate in disinfectant tests [14] sensitivity was shown to 60% alcohol, alcohol hand rubs, bleach and povidone iodine-based disinfectant in suspension tests using very low level interfering substances. This is remarkable, since this would suggest a higher sensitivity for murine norovirus than for feline calicivirus for commercial disinfectants and most likely higher sensitivity to hypochlorite for murine norovirus compared to human norovirus, since the RNA of human norovirus was much better protected during hypochlorite inactivation than the RNA of feline and canine caliciviruses [66].

Shedding

Norovirus is shed in high quantities in the stool of infected persons; around 10^8 but up to 10^{10} RNA copies per gram of stool were reported for different GII viruses [165,180,302]. Similar virus concentrations were found after experimental infection with a GI virus [12]. This leads to the theoretical possibility that one gram of stool of an infected person could be enough to infect up to 5 million individuals. Projectile vomiting, which is a typical symptom for norovirus illness, was once calculated to have the theoretical potential to infect 3×10^5 to 3×10^6 people [34], and the distance of people at risk of infection to the place of vomiting was inversely related to the chance of them actually becoming infected [188].

Shedding of virus continues after clinical recovery of the patient, and may last three or four weeks in otherwise healthy people, but can be especially long in young children [12,253]. In a hospital study involving people of all ages, higher concentrations of virus in stool were found to be associated with older aged patients and also with prolonged diarrheal symptoms and increased severity of symptoms [165]. In immunocompromized patients severely prolonged illness accompanied by prolonged shedding may last up to several years, as will be described in detail later in this thesis [33,141,180,205,210,266].

Infectivity

The possibilities for infection experiments are limited by the unavailability of a cell culture model. Therefore, Teunis *et al.* used the results of volunteer studies with GI.1, Norwalk virus, to estimate infectivity. They showed norovirus to be extremely infectious, reportedly more than any other virus. The probability of infection of a single norovirus particle was estimated at 0.5, and the ID50 at 18 virus particles [293]. One recent paper estimated the reproduction number (R0) of norovirus in an outbreak situation in the absence of hygiene measures at over 14 [112]. This means that every primary case on average infects 14 secondary cases. After implementation of strict hygiene measures in this outbreak, R0 lowered to around 2.

For comparison, the R0 of the 1918 strain of pandemic influenza virus was estimated to be ~2-3 [201], and measles virus, considered the most highly contagious, has an R0 of 12 to 18 [35]. Poliovirus, also designated 'highly contagious', and also an enteric virus transmitted through the fecal-oral route, has an R0 of ~5-7.

Repeated infections

Even though there is no proof of the existence of different serotypes of norovirus by classical virus neutralization methods, the genetic diversity displayed by noroviruses likely translates into antigenic diversity, so that infection with a strain of one genotype may not confer immunity against strains of another genotype or even variants within a genotype [107,170,269]. This issue will be addressed in more detail later on in this thesis. Furthermore, volunteer studies have shown that protective immunity after infection may be absent or short-lived [128,227,336]. The combination of antigenic diversity and the apparent lack of long term protective immunity are the likely cause of the occurrence of norovirus infections in children, adults and the elderly. In effect one individual may suffer repeated infections, even with viruses belonging to the same genotype and therefore people of all ages are affected by norovirus illness, unlike with e.g., rotavirus, where symptomatic infections only seem to occur when a different serotype is encountered.

Infections and exposure

Illness and pathogenesis

The illness caused by norovirus is usually described as mild and self-limiting. Incubation time is typically 12-72 h and symptoms may last 1-3 days, although longer times up to 5 days have been reported, particularly in young children and the elderly [179,253]. Diarrhea is the most commonly reported symptom (87%), followed by vomiting (74%), abdominal pain (51%), cramps (44%) and nausea (49%). A fever was reported for 32-45% of cases and mucus in the stool was seen less commonly (19%) [245,253]. A study of 39 norovirus positive children showed viremia (virus in the serum) in 15% of the patients [285]. Illness in people with co-morbidity or in the elderly may be more severe and sometimes has very serious consequences, such as prolonged infections and excess deaths, as will be discussed in this thesis. Current information on the pathogenesis of norovirus infections has been derived from analyzing infection in volunteers and in gnotobiotic pigs. Only one of 48 pigs that developed illness had mild lesions in the proximal small intestine, and duodenal and jejunal enterocytes were shown to be infected by immunofluorescent microscopy [39]. Binding of calicivirus-like particles to duodenal and buccal tissues was also shown in pigs [38]. Volunteer studies had previously revealed villus atrophy in duodenal biopsies [5,58]. The study of duodenal biopsies of norovirus-infected people provided a basis for understanding the

cause of diarrhea, namely a combination of epithelial barrier dysfunction in the duodenum, a reduction of tight junctional proteins, increased apoptosis in duodenal epithelial cells and increased anion secretion [299]. Abdominal computed tomography (CT) scans of children with acute norovirus infections revealed wall thickening and enhancement in the different parts of the small intestine, namely the duodenum, jejunum and ileum, as well as fluid filled bowel loops [284]. Recently, a bowel perforation of the small bowel resulting of norovirus infection was reported [230].

Burden of disease estimates

The burden of hospital infections

Hospitals can be severely affected by norovirus outbreaks. The potential for person-to-person spread is great and the people at risk in this setting are often especially vulnerable to infections [196,300]. Nosocomial norovirus infections, typically acquired after a patient was admitted to the hospital, are common. A study in a Dutch academic tertiary care hospital reported that over 51% of norovirus infections in this hospital were hospital-acquired, with a conservative definition of onset of symptoms at least 5 days after admission [13]. In high risk hospitalized patients norovirus infections can have severe consequences such as prolonged illness and death. A patient who had undergone heart-transplantation was reported to shed norovirus for several years, while remaining symptomatic with diarrhea [33,210]. Such prolonged infections have been reported more often [44,164,180,205,333] and occur at alarmingly high prevalence, as we describe later in this thesis. An outbreak of necrotizing enterocolitis was reported in a neonatal unit with 2 of 8 neonates dying [305]. No cause for necrotizing enterocolitis is known, but infections of the bowels are strongly suspected to play a role and the unit was struck by a norovirus outbreak at the time of the necrotizing enterocolitis outbreak.

Attack rates among hospital patients and staff vary greatly among reports, with rates among staff (ranging from 19% to 76%) often at least as high as among patients (between 18% and 32%) [129,133,196,313].

Economic costs of hospital infections

In hospitals, norovirus outbreaks do not only form a great health burden for patients and staff, but are also associated with great financial costs. Norovirus was reported to have the highest 'closure rate' of all pathogens that may be responsible for nosocomial outbreaks; in 44.1% of norovirus outbreaks a ward had to be closed, against 38.5% for runner-up influenza and e.g. 25.9% for rotavirus and 11.8% for *Clostridium* spp [105].

One large outbreak in a 946-bed tertiary care hospital in the US was calculated to have cost over $650 000US [129]. These costs included lost revenue associated with unit closures, sick leave, replacement of supplies and cleaning expenses. Several wards were affected and 355

people fell ill with gastroenteritis. A Swiss hospital of similar size (960 beds) analyzed the cost of a smaller outbreak, affecting only 45 people at 2 different wards and calculated that this outbreak cost approximately $40 000US [342].

The burden in long term care facilities

Greig and Lee performed a review of enteric outbreak reports from long-term care facilities in several countries, and found that approximately 60% of outbreaks were caused by norovirus [96]. More importantly, over 70% of the symptomatic persons in all the studies reviewed were in these norovirus outbreaks, with 16 of 60 reported deaths caused by norovirus, exceeded in numbers only by *Salmonella* outbreaks. That norovirus has an extra great impact on nursing homes is illustrated by the finding of the high rate of illness among staff due to norovirus infections; 86% of all documented symptomatic staff were ill resulting of a norovirus infection, whereas other pathogens rarely caused illness among staff [96]. Not seldom, as in hospitals, the attack rate among staff is higher than among patients [16].
Greig and Lee reviewed recommendations for containment measures in long-term care facilities such as nursing homes. Options for containment of an outbreak are limited by the fact that the inhabitants of the facility usually live there; therefore the possibilities for ward-closure, a new-admissions stop or quarantining the affected patients are limited [96]. Limiting person-to-person spread was most effective measure in controlling an outbreak and included managing residents movements, isolation of ill residents when possible, and dedicating specific staff to care for the ill. Also, especially for norovirus outbreaks, vigorous cleaning was important in containment. A study in Dutch nursing homes with different containment protocols revealed that these different protocols were only effective when implemented rapidly after onset of the outbreak, and then positive effects often only reduced the amount of sick among the staff [79].

Economic cost of infections in long term care facilities

The overall economic burden of gastroenteritis outbreaks in long term care facilities is not known. A study of the costs of illness among elderly Americans found it difficult to calculate a reliable estimate for gastrointestinal illness due to limited availability of necessary data [311]. The annual costs for gastrointestinal illness among elderly (>65) in the US, *excluding* nursing home costs and morbidity-related indirect costs were nonetheless estimated to be around $1 billion US.

Host factors and attack rate

Differences in host-susceptibility for different genotypes have been reported, and are based on the presence or absence of specific virus receptors in the potential host. Although additional research is needed to establish more detail and to clear up some controversies, the

currently proposed receptors are encoded by the human histo blood group antigen (HBGA) genes [32,119,120,187,287,289]. Non-secretors make up approximately 20% of the Caucasian population, and similar percentages in other people [48,138]. These non-secretors are rarely infected by any genotype of norovirus, thus a 100% attack rate is equally rarely reported in an outbreak. An exception was reported by Lindesmith and co-workers, who. reported a GII.2 infection in one secretor negative individual [169] and recently a foodborne outbreak with a secretor-independent susceptibility pattern was described [213]. Some people may get infected by norovirus and consequently shed virus, but show no symptoms. Such asymptomatic shedders likely contribute to the spread of the virus, and form a reservoir for noroviruses.

Few systematic studies have reported on attack rates for outbreaks. An average attack rate of 34.7% was reported for 60 norovirus outbreaks caused by different genotypes in Catalonia, Spain, mostly ranging from 20 to 60% [298]. An earlier study, with strains belonging to different GII genotypes, reported attack rates ranging between 17-100% [322]. Interestingly, a recent Japanese study reported a lower attack rate for GII.4 food-related outbreaks compared to other genotypes [211], even though GII.4 binds to the widest variety of HBGAs of all norovirus genotypes studied so far [120]. It should be kept in mind that in outbreak situations it is often not known exactly who were exposed, and thus the given figures are estimates. Overall, reported attack rates in norovirus outbreaks vary greatly, and are probably influenced by the genotype causing the outbreak, the possible presence of immunity against this strain among the people at risk, their general health status, personal hygiene measures taken, the transmission mode and other factors.

Modes of transmission

Person-to-person transmission

The majority (88%) of reported norovirus outbreaks in a large European surveillance network (FBVE) study [157] were associated with person-to-person transmission. In a review of enteric outbreaks in long-term care facilities by Greig *et al.*, 71% of the described outbreaks in these settings were attributed to transmission by person-to-person contact [96]. This manner of transmission is important to consider, especially in healthcare facilities, where care-giving personnel are in contact with many patients or residents. Thus, when they considered all containment measures available in these settings, both Greig *et al.* and Friesema *et al.* found that control measures aimed at reducing the person-to-person spread were most effective [79,96].

Environmental transmission

Noroviruses are highly persistent in the environment. Viral particles may be transferred to new surfaces in a chain of relocation events, and during an outbreak, surfaces such

as computer keyboards or mice, door handles, and cutting-boards in a kitchen may become contaminated [43]. Contamination of surfaces in hospital wards was shown to decrease with the implementation of strict hygiene rules for staff and visitors [85,86]. If a contaminated surface is not identified and cleaned, an outbreak may become protracted as a result [72,160,315]. Projectile vomiting has been shown to play an important role in the contamination of unexpectedly large areas [34].

Food- and waterborne outbreaks

Noroviruses are a known cause of foodborne outbreaks. Studies vary in the proportion of reported outbreaks spread by foodborne transmission, likely reflecting differences in the surveillance focus between countries. Studies in Europe by the FBVE network of laboratories [157] and certain studies in the US [61] reported that approximately 10% of all reported outbreaks were associated with contaminated food. In Japan, up to 35% of norovirus outbreaks were linked to foodborne transmission [103]. In comparison, Lynch and co-workers reported that 55% of all food-borne disease with a known etiological agent in the US between 1998-2002 were caused by several bacteria, whereas 33% of the outbreaks were caused by viruses. Norovirus was the single most important known pathogen, causing the majority of diagnosed outbreaks (30%), with salmonella (27%) second in importance [183]. It has been estimated that up to 67% of all food-related illness is associated with norovirus [144,199]. Foodborne-norovirus outbreaks may be large, possibly affecting up to thousands of people, such as what happened resulting from the contamination of cake icing by a baker in the US in 1982 [158] and in Denmark, resulting from the import of a contaminated batch of frozen raspberries [73]. Such common source outbreaks may transcend borders or even continents [56,143,146,147,209,271]. The increasing global trade of food has increased the risk of international or even global foodborne outbreaks. Even though international surveillance systems have been implemented (e.g. FBVE (http://www.fbve.nl) or OzFoodNet (http://www.ozfoodnet.org.au/)), only a few large-scale outbreaks have been detected thus far. Recently, to facilitate the decision-making process for further in-depth investigation of a suspected foodborne outbreak, a selection tool was developed based on a five-year history of outbreak reporting in Europe [318]. The background data on the outbreak under investigation would minimally include information on the outbreak setting, the number of cases involved at the time of the outbreak, the country, the intensity and focus of surveillance and the suspected norovirus genotype. This tool, that is provided online, provides an estimate of the probability that an outbreak was caused by foodborne transmission. By using this approach, the estimate of outbreaks attributed to foodborne contamination was increased to 40% [318]. Similarly, a reassessment of over 4 000 food-related gastroenteritis outbreaks reported in the US between 1998 and 2002 resulted in an increase from 14% to 28% for norovirus as the likely etiological agent [304]. Still, noroviruses

are difficult to detect in food matrices; only specialized laboratories can perform this research and not many countries routinely screen for viruses in food-related outbreaks, as exemplified in a study in Japan [108]. The food industry and governments pay little attention to possible viral contamination of products [144].

Food is most often contaminated by the introduction of enteric pathogens present in fecal material that is derived from either an environmental source or an infected food handler. Filter feeding bivalve mollusks (oysters and mussels) are an important source of norovirus in food-borne outbreaks. Mollusks filter water for nutrients and retain noroviruses in their gut by specific binding, often when their growth-beds have been contaminated with waste water effluents [19,81,162,163,209,279,295,296]. Also, salads and soft fruits like berries can be contaminated by irrigation water or by washing with contaminated water [185,197]. Large scale (drinking) water contaminations have also been reported [114,116,153,197,215,216]. Food-handlers who are ill or recently recovered at the time of their food-handling are also known sources of contamination. Numerous examples are known, such as the recently described norovirus illness among river rafters in the Grand Canyon [186] or the employees of a government department who fell ill after attending a reception with a lunch buffet, where the bread rolls had been prepared by an ill baker [56].

Zoonotic transmission

Direct zoonotic transmission of noroviruses to humans is thought to be absent or rare. However, the detection of noroviruses closely related to human noroviruses in animals and especially the increasing numbers of detection reports in cattle indicates a potential risk [195,325,332]. In Argentina, an unusually high prevalence of GIV noroviruses was detected in gastroenteritis outbreaks among humans [91]. GIV strains relating to illness in humans are rarely reported, however, they have been detected in several different animals [190,191]. Two genogroups (GIII and GV) comprised solely of animal noroviruses exist, of which the GIII viruses have been detected in cattle and sheep [332] and the GV viruses in mice. Cattle and swine have also been shown to be susceptible to human noroviruses [39,276] and a report of inter-genogroup recombination by Nayak and coworkers [207] opens a new possible source of norovirus diversity. On the other hand, the finding of strains related to the most highly prevalent genotype in humans (GII.4) in pig stools as well as in retail meat [195] raises the question whether this was the result of animal to human zoonoses, or rather, human to animal 'zoonoses', or possibly contamination during processing.

Vesicular Exanthema of Swine Virus (VESV), the prototype vesivirus, that infects swine, and was first detected in swine in 1932, is commonly believed to have successfully crossed the species barrier from marine mammals, in which near identical viruses have been found, to swine. Some marine vesiviruses have been described to have caused accidental laboratory infections.

Surveillance and detection

Surveillance history

In many countries, norovirus surveillance depends on outbreak reporting by peripheral institutions such as municipal health authorities to central laboratories, where diagnostic tests are performed to assign an etiological agent in samples. Criteria that have typically been used to establish if the definition of a *viral* gastroenteritis outbreak is met are the previously mentioned Kaplan criteria [136,304].

In the past, the lack of classical virological techniques for norovirus detection other than electron microscopy has greatly hampered norovirus surveillance. During the past decade, molecular detection and typing techniques for gastroenteritis viruses have become commonly used. As a result, awareness of the burden of norovirus outbreaks has grown. Sporadic outbreak investigations have clarified that outbreaks, which are generally reported in surveillance, only form the tip of the iceberg of all norovirus infections occurring in the population [53,159,229,253,329]. Surveillance systems typically cover healthcare settings, such as hospitals and nursing homes for the elderly, but not so much the general community of otherwise healthy adults and children. Thus, gastrointestinal illness among the general population often goes unreported and norovirus illness has historically been mainly associated with the elderly and the frail. Nowadays the importance of norovirus as a cause of illness in young (hospitalized) children is increasingly recognized, in developed as well as developing countries [10,13,24,44,60,73,106,116,141,180,182,193,200,202,205,206,225,234,246,254].

Detection methods

RT-PCR

Various reverse transcription-PCR (RT-PCR) assays have been designed to detect norovirus in clinical samples such as fecal samples or vomit, and also in environmental samples, such as surface swabs, or food and water. Several genomic regions may be targeted for detection. RT-PCR is relatively sensitive, offering the possibility of detecting a low quantity of virus, using degenerate primers targeting conserved genomic regions. However, due to the high degree of variation among noroviruses, some norovirus strains may not be detected. In clinical practice this may not be problematic given the dominance of a limited number of genotypes (see below) [157,268,301,303]. In other situations and in reference laboratories care should be taken to monitor test-performance against less common genotypes. Real-time PCR is increasingly used, which is more sensitive and faster than RT-PCR. Using real-time PCR with virus-specific-primer and probe combinations in multiplex assays, the detection of multiple different viruses in one test has become feasible. Additionally, real-time assays are semi-quantitative, i.e., a decrease in Ct values for the same virus indicates

that the amount of viral RNA present in samples has increased, which may be indicative for increased virulence.

Many laboratories perform sequencing of noroviruses positive samples. Determining the genotype and possible signature mutations enables linking of patients or outbreaks, or finding a common source of infection. Again, several genomic regions can be analyzed (figure 1.2) [270,323]. For surveillance purposes these partial genomic sequences are used to monitor trends, whereas the highly variable P2 domain of the capsid is sequenced for addressing questions regarding transmission routes, e.g., in assessing hospital epidemiology [321,337].

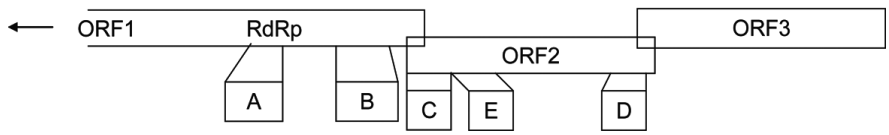


Figure 1.2. Schematic representation of the locations of the genomic regions of norovirus used for genotyping. Adapted from Vinjé et al. [321], published in Siebenga et al. [270]. RdRp: RNA dependent RNA polymerase.

Enzyme immunoassay

Enzyme immunoassay (EIA) tests have been developed for the detection of norovirus antigen in stool samples, and several of these tests are commercially available. Advantages of EIA testing over PCR based assays include simplicity (no specialized equipment or skilled personnel required) and speed (rapid bed-side tests have been developed based on an EIA that promise results within 15 min [140]). The EIAs use either monoclonal or polyclonal antibodies specific for a limited number of antigenically-distinct norovirus genotypes, which can be problematic in the detection of antigenic variants or emerging genotypes [92]. Knowledge of the local circulating norovirus genotypes is helpful in evaluating the efficiency of the EIA in a particular setting [49]. Moreover, if outbreak samples are negative by the EIA test, they should be further screened by RT-PCR. The low sensitivity of EIA tests (between 44 and 59%) [92] makes them unsuitable for diagnosing sporadic cases, unless negative results are, again, followed by RT-PCR analysis.

Outbreak algorithms

The availability of diagnostic tests for noroviruses has allowed the establishment of laboratory-based criteria for the designation of a norovirus outbreak. Duizer *et al.* proposed criteria that address the variation of sensitivity and specificity among diagnostic assays as well as the 'base-line prevalence' of norovirus in the population [68]. It has been estimated that, on average, 5.2% of the population is shedding norovirus without symptoms at a

given time, which is an important factor to consider when detecting norovirus-positive individuals in an outbreak investigation [53]. According to Duizer *et al.*, when RT-PCR is employed as the diagnostic assay, the number of norovirus-positive samples should be at least 1 when testing 2 to 4 samples, and at least 2 when testing 5 to 11 samples. When EIA tests are utilized, the norovirus-positive samples should be at least 1 when testing 2 to 6 samples, or at least 2 when testing 7 to 11 samples.

Epidemiology

Outbreaks

An outbreak is generally defined as at least two definite cases, linked in time and place, with definite cases defined as patients with a) two or more episodes of vomiting in a 12 hour period lasting 12 hours or longer, or b) two or more loose stools in a 12-hour period lasting 12 hours or longer, or c) both. Noroviruses' combined propensities of low infectious dose, high levels of shedding, high environmental persistence, long periods of asymptomatic shedding after clinical recovery, high genetic and antigenic variability and the lack of long-term immunity give the virus the ability to cause, possibly repeated, outbreaks in all possible settings of human contact. Such settings include hospitals and long-term care facilities, where outbreaks may have great impact on already weak persons. Other settings that are often associated with norovirus outbreaks are recreational facilities such as cruise ships [145,315,316,330] day care centers [182], schools [134,135], restaurants and the army [102]. Also, hundreds of hurricane evacuees fell ill in evacuees housing in the aftermath of hurricane Katrina in the south of the US in 2005 (ProMed 20050911.2693) [339].

Hedlund and co-workers report that 50% and 26% of all norovirus outbreaks reported in Sweden between 1994-1998 took place in hospitals and nursing homes, respectively [111]. Fifty-seven percent of Dutch outbreaks were reported from nursing homes and 9% from hospitals [310]. Forty-three percent of outbreaks in the US were reported from nursing homes and hospitals [75], although for a later period only 25% of outbreaks were reported for these settings [74]. However, the true prevalence of norovirus outbreaks is difficult to determine because not all outbreaks are reported whereas those in healthcare settings such as hospitals are mandatory to report in many countries.

Sporadic cases

In the Netherlands, approximately 500 000 episodes of norovirus were estimated to occur annually, in a total population of 15.7 million [52]. Even though norovirus is notorious for causing outbreaks, the major share of infections occurred in sporadic cases or in small family outbreaks that were not reported. Surveys of sporadic cases of gastroenteritis due to norovirus showed that although prevalence in young children and the elderly are the

highest, people of all ages experience norovirus episodes. A recent literature review provides a nice overview of current knowledge of the role of norovirus in sporadic gastroenteritis globally [229]. The studies included in this review reported a range of 5-36% of sporadic cases of diarrhea to be caused by norovirus infection, with a pooled proportion of 12%.

A number of studies on sporadic norovirus cases have focused on illnesses in children. A Dutch population based cohort study in 1999 indicated that 65% of all norovirus illness occurred in children below 18 years of age [53]. Patel *et al.* were able to estimate that approximately 200 000 yearly deaths result from norovirus infection among children under 5 years of age in developing countries. They also estimated that annually norovirus caused approximately 900 000 episodes of norovirus gastroenteritis that required clinic visits among children under 5 years of age in high-income countries and 64 000 children are hospitalized. Additionally, several recent studies comparing etiological agents as causes of acute gastroenteritis in young children showed that norovirus may be found as commonly as rotavirus, a virus generally regarded as a major burden in childhood disease, and some of these even report similar clinical impact [45,124,249,263].

The previously mentioned population study in the Netherlands found that 13% of norovirus illness was found among elderly over 65 years of age [53]. Death among elderly resulting from norovirus infection has been reported for individual cases [179,220,317]. A syndromic surveillance study estimates excess deaths among elderly over 65 years of age in England at least 80 per year [110].

Norovirus molecular epidemiology

Many public health laboratories nowadays include norovirus genotyping in their standard diagnostic procedures, in combination with epidemiological background data. Strains belonging to GII were found in 75%-100% of sporadic cases (for overview see Patel *et al.* [229]). From molecular surveillance data it has become clear that there is less diversity in outbreak based studies than in studies looking at sporadic cases, indicating differences in transmissibility between norovirus strains. GII.4 viruses are by far the most commonly detected strain in studies in recent years, whether describing outbreaks or sporadic cases, as will be discussed thoroughly later in this thesis. Among children and patients with prolonged shedding, GII.3 is the second most commonly reported genotype, but this data is less complete [60,233,266].

Molecular epidemiology [see Box] and the studying of strain diversity through time are important analytical tools for noroviruses, in view of the lacking of neutralization assays for noroviruses. Sites subject to selective pressures can be identified by studying sequence alignments [148].

Molecular epidemiology as a tool in linking norovirus outbreaks and tracing foodborne outbreaks

For prevention and containment purposes there is a need for tools that can be used to assess whether different outbreaks are connected to provide insight into transmission routes. Using genotyping data alone for such assessments presents the difficulty that GII.4 strains cause the great majority of all outbreaks, and within epidemic seasons genetic variation of outbreak strains may be minimal in the genetic regions assessed for diagnostic purposes. Lopman *et al.* proposed a method for assessing both molecular as well as statistical epidemiological data to verify if different outbreaks were actually linked [177]. This method takes the probability of the occurrence of the observed patterns by pure chance into account, and provides statistical evidence that linked outbreaks share greater similarity in sequences detected, discernible even for GII.4 outbreaks.

If outbreaks are proven connected, preventive measures may be more effectively executed, for example if a common source of infection is involved. A further complication in detecting diffuse food-related norovirus outbreaks and their possible source lies in the high environmental stability of the virus; it implies that the contamination detected in the food may have occurred long before the outbreak [67]. In addition, a practical problem is that the number of reported outbreaks may be too overwhelming for available infection control staff, making it impractical to spend time for proper outbreak investigation. Again, based on statistical analyses assessing the probability of occurrence of norovirus in specific settings, Verhoef *et al.* proposed to use the genotyping information to help in selecting the outbreaks for which foodborne infection is most likely [318]. So, although some obstacles and difficulties remain in detecting common-source norovirus outbreaks, promising progress is made in understanding and distinguishing them.

BOX. Phylogenetic analysis of viral sequences; molecular evolution and epidemiology

Phylogenetics is the study of evolutionary relatedness of organisms. Genomic sequences are commonly used for the reconstruction of phylogenies, and the phylogenetic tree is the graphical representation of the genetic distances between the taxa. Studying the phylogenetic relations of infectious agents such as viruses, in combination with patterns of illness ('classical epidemiology'), is generally referred to as molecular epidemiology. Molecular (genetic) characteristics of the virus causing specific patterns of illness can thus be identified, by mapping epidemiological data on the tree. This might increase insight in exact transmission chains and ultimately such knowledge may be used to control transmission of the virus [198].

Different methods and different underlying models of evolution for inferring the phylogeny and subsequent phylogenetic tree of a group of taxa exist, and choosing the most appropriate methods will follow from considerations regarding the type of sequence and the organism under study, and the question the researcher is looking to answer [198]

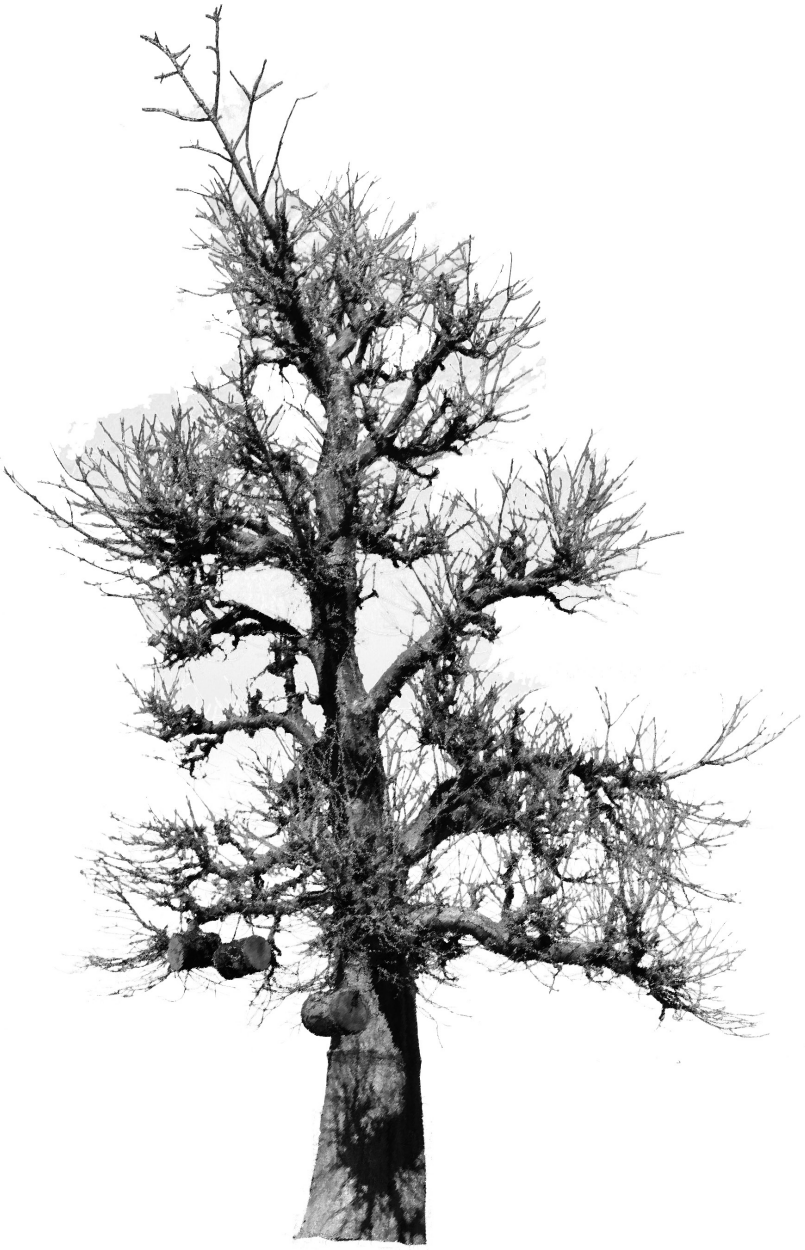
The high rate of mutation accumulation of especially RNA viruses, resulting of short generation times, large populations and lack of proofreading activity of the polymerase, makes them especially suitable for phylogenetic analyses. The evolutionary pattern observed for any virus is the resultant of the hosts' immune response (posing selective pressure and favoring antigenically 'novel' strains), and transmission bottlenecks (determining which viruses from the quasispecies cloud in one host are successfully transmitted to other hosts). Such patterns will differ per virus, and are shaped by parameters such as the type of immunity that is built up by the hosts, and the types of infection established by the virus (e.g. short infections with (partial) crossimmunity (Influenza), or persistent infections). Under the assumption of the coalescent theory this enables the reconstruction of evolutionary histories, and thus the behaviour of virus populations through time [62,64,231]. The coalescent theory assumes that all organisms in a phylogeny share a common ancestor, and through coalescing them in nodes in the tree, this (hypothetical) common ancestor is found. Including detection dates in the phylogeny can reveal substitution rates on the branches of the phylogeny, and thus population history parameters can be inferred; phylodynamics.

Scope of this thesis

The aim of this thesis was to get a better understanding of the processes underlying norovirus diversity. This knowledge of molecular evolution should provide a basis for efficient use of molecular typing data obtained in norovirus surveillance and outbreak situations, and for designing rational prioritization in intervention strategies to limit spread or prevent emergence of epidemic norovirus strains.

In **chapter 2** we present an overview of the norovirus surveillance results obtained during twelve years of structured norovirus surveillance in the Netherlands. These data showed that a large proportion of all reported norovirus outbreaks in the Netherlands had been caused by one genotype, GII.4. The genetic characteristics of strains detected throughout the surveillance period belonging to this genotype were studied in more detail, the results of which are described in **chapter 3**. A stepwise cycle of emergence of new GII.4 variants, followed by replacement of the previously circulating dominant variant was found. This so-called epochal evolution, where periods of phenotypic stasis are separated by the emergence of phenotypically distinct new variants was shown to be a global phenomenon, as we describe in **chapter 4**. Then, we used a phylogenetic approach for analyzing the vast amounts of sequence data available from our outbreak surveillance database. Based on the presumption that the evolution of norovirus is subject to a molecular clock, we showed in **chapter 5** that the number of effective GII.4 norovirus infections has recently increased. Additionally, using computational analyses for lack of reliable cell-culture systems, we studied the capsid sequences of GII.4 strains and identified amino acids with important roles in molecular adaptation of the GII.4 viruses to their ever changing environment.

An important question that arose from our findings, was from where these new genetic variants emerged. Genetic drift of noroviruses occurs at high rate, so they could very well be 'regular' drift variants. Alternative explanations are, however, also possible. Patients with chronic norovirus infections had been mentioned as potential sources for the emergence of new variants. In **chapter 6** we describe a retrospective study in a tertiary care hospital, which showed that chronic infections do not only occur at a high frequency, but that the viruses infecting these patients accumulate mutations at high rates, that are however inversely proportional to the amount of immune suppression of the patient (viruses in immunocompetent people accumulate mutations faster). So, patients with chronic infections can not be excluded as potential reservoirs for emergence of new variants. In **chapter 7** we show that infectious virus can be transmitted by these patients to other patients in the hospital, thus forming a continuous source of virus and consequently hospital outbreaks. Finally, using syndrome surveillance coupled to our norovirus surveillance data, in **chapter 8** we show that excess morbidity and mortality in the elderly population is likely attributable to norovirus infections.



Chapter 2

Molecular and Epidemiological Overview of Twelve Years of Norovirus Surveillance in the Netherlands, 1994-2005

J. Joukje Siebenga[1,2], Harry Vennema[1], Erwin Duizer[1] and Marion Koopmans[1,2]

1 National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands,
2 Erasmus Medical Center, Rotterdam, The Netherlands,

Published in condensed form in Emerging Infectious Diseases
• www.cdc.gov/eid • Vol. 13, No. 1, January 2007, 144-146 •

Abstract

Noroviruses were found to be the main causative agent of viral gastroenteritis outbreaks in the Netherlands during the passive outbreak surveillance, conducted between December 1993 and December 2005. Overall, the number of norovirus related outbreaks showed an increasing trend during the surveillance period, with four epidemic winter seasons with increased activity, in 1995-1996, 2001-2002, 2002-2003 and 2004-2005. Strains belonging to GII.4 were predominant from 1995 onward, causing 68% of all outbreaks. The epidemic rises in the afore mentioned seasons were attributable to GII.4 strains which also stand out from the other genotypes by causing significantly more healthcare setting outbreaks, and having the preferred transmission mode person to person. Although it was previously repeatedly reported that long-term immunity against noroviruses is not found, the drops in the absolute and relative numbers of GII.4 outbreaks in years following epidemic seasons suggest the development of some level of immunity in the population.

Introduction

In recent years, human noroviruses have increasingly been recognized as a common cause of gastroenteritis. Since the introduction of rapid molecular detection techniques a very high proportion of acute gastroenteritis outbreaks have been attributed to noroviruses [75,320,322]. It has been established that noroviruses are the leading cause of acute gastroenteritis and more commonly detected in outbreaks than bacterial pathogens [55,331,333].

Surveillance of viral gastroenteritis (GE) outbreaks was initiated just over a decade ago at the National Institute for Public Health and the Environment (RIVM) in the Netherlands. In order to be able to determine the exact role of norovirus and possible differences between different norovirus strains in GE outbreaks, a minimal set of epidemiological data was collected for reported outbreaks. This data was combined with molecular-biological detection and typing techniques to allow more in-depth analysis of the surveillance data.

The genus *norovirus* is classified in the family *Caliciviridae*, based on genetic and morphological properties. Five genogroups are recognized, of which the two major groups, genogroup (G) I and GII and a third smaller group, GIV, are found in humans. These genogroups are divided into a growing number of genotypes. This division is based on capsid sequences [242]. Currently in GI 8 genetic clusters and in GII 17 are recognized [341].

Here we report an overview of twelve years (December 1993 – December 2005) of passive norovirus outbreak surveillance in the Netherlands. All data available at the RIVM was

collected and studied in order to find trends in prevalence of genotypes through the years. Although many reports of norovirus outbreak surveillance are available, this is the first report of surveillance over a period longer than five years in duration, linking molecular virological data to epidemiological data.

Material and methods

Outbreak definition and norovirus inclusion criteria

An outbreak of gastroenteritis with probable viral etiology was defined as at least two definite cases, linked in time and place, with definite cases defined as patients with a) two or more episodes of vomiting in a 12 hour period lasting 12 hours or longer, or b) two or more loose stools in a 12-hour period lasting 12 hours or longer, or c) both.

An outbreak was considered to have been caused by norovirus when 25% or more of the stool specimens collected from cases were positive for norovirus [181].

Outbreak reporting

Outbreaks were typically reported by the Municipal Health Service (MHS) or the Food Inspection Services to the RIVM. Usually samples were tested for bacterial pathogens using routine methods before they were submitted to the RIVM (http://www.infectieziekten.info/index.php3). Diagnostic testing for viral pathogens at the RIVM was done free of charge.

Epidemiological data of outbreaks was provided by the MHS's or by the reporting physicians. Data requested were setting, date of onset, the number of persons affected, the (most probable) mode of transmission and the number of hospitalizations. For each field a choice of options was offered. The most probable mode of transmission was judged by the reporting physician, based on observations of the manner of spread in time and place of the outbreak. Different dates were reported. In some cases the first day of illness of the first case was reported, in other cases the first day of sampling. When these dates were not available the date of arrival of samples at the RIVM was recorded. If more dates were available per outbreak the first day of illness was preferentially used, otherwise any available date closest to the first day of illness. For the completeness analysis, if the month was known, the date was considered complete.

Sampling, detection and typing

Stool samples were sent to the RIVM by regular overnight mail at ambient temperature in prelabeled specimen kits and tested for presence of norovirus by Reverse Transcriptase Polymerase Chain Reaction (RT PCR), using broadly reactive primers. The detection methods were periodically modified according to state of the art developments in the field

[181,323]. Typing was done by sequence analysis of Region A and in selected cases of Region C. Recently, RT-PCR detection methods were complemented with real-time PCR similar to methods described elsewhere [130].

Time span

The first outbreak reported in the database occurred in 1993 (reporting date 18-12-1993). The most recent data that has been included in this overview, is from the season 2005-2006 (26-12-2005).

Because norovirus outbreaks are highly seasonal, with peaks during winter, a year was defined as running from July through June, except in the completeness analysis.

An enhanced surveillance project was done in the years 2001 and 2002 [310]. During these years, MHS's were actively stimulated to report outbreaks of viral GE and submit samples for diagnostic testing.

Data storing and processing

The epidemiologic data had initially been stored using paper reports. Diagnostic testing, typing and sequencing data were stored in Access and Excel 2002 (Microsoft) and Bionumerics (Applied Maths BVBA, Sint-Martens-Latem, Belgium) software. Data from the paper folders and lab results were linked through a unique outbreak identifier and analyzed using Excel 2002. Statistical analyses (chi square, Fisher Exact Test, Relative Risk, and odds ratio) were done using SAS9.1 software (SAS Institute Inc., North Carolina, USA).

Results

In the period from 18-12-1993 through 26-12-2005 in total 1032 GE outbreaks were reported to the RIVM. Of these, 695 (67%) outbreaks met the inclusion criteria for a norovirus outbreak used in this study (≥ 25% of samples positive). In four of these 695 outbreaks (0.6%) other pathogens (rotavirus, *Clostridium perfringens*) were found. In 50 of 1032 outbreaks (5%) norovirus was detected, but less than 25% of the samples tested positive, therefore these outbreaks were not attributed to norovirus. In 8 of these 50 outbreaks other pathogens were found as well (rotavirus and astrovirus). From these 8, 5 (62,5%) were daycare centre outbreaks (data not shown).

Data completeness

Data completeness (table 2.1) varied considerably in time and per field. Fields with the outbreak setting and the date were generally completed (on average 97% and 96%, respectively). The mode of transmission was provided in a lower percentage of outbreaks (38% on average), but increased over the years (79% since 2003). Typing data was completed

for 88% of the outbreaks, and the number of affected persons in 59%.

Reported outbreaks per year

During all years of surveillance norovirus was found as a causative agent in outbreaks. Overall, a gradual increasing trend in the number of reported outbreaks per year was observed (figures 2.1 and 2.2, table 2.2). The years 2001 and 2002 were enhanced surveillance years, and had relatively high numbers of outbreak reports. In the years 1995-1996, 2001-2002, 2002-2003 and 2004-2005 increased numbers of norovirus related outbreak reports (66, 90, 154 and 161, respectively) were observed.

The number of outbreaks per month was highest during the winter months, as also reported in [277], and reached its maximum in different months, ranging from October to March, but most often in December and January (figure 2.1). The peak could not always be defined as clearly in the earlier years of surveillance. In three seasons (1999-2000, 2000-2001 and 2001-2002) two peaks were observed. This was most pronounced in 2001-2002, where the peaks were three months apart (January and April).

Table 2.1. Completeness of data per year. OBs: outbreaks. Avg: average. Tr Mode: transmission mode, Typed: polymerase typed, Aff: number of affected patients. For description of fields please refer to methods section.

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	*Total*Avg	
OBs	**8**	**29**	**58**	**20**	**12**	**32**	**39**	**65**	**180**	**47**	**123**	**81**	*694*	
	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)	#	%
Setting	8 (100)	28 (97)	52 (90)	19 (95)	12 (100)	32 (100)	36 (92)	64 (98)	175 (97)	46 (98)	118 (96)	78 (96)	*668*	97
Mode	4 (50)	5 (17)	1 (2)	4 (20)	1 (8)	4 (13)	22 (56)	34 (52)	7 (4)	36 (77)	100 (81)	64 (79)	*282*	38
Date	8 (100)	29 (100)	54 (93)	20 (100)	12 (100)	32 (100)	39 (100)	49 (75)	174 (97)	43 (91)	123 (100)	81 (100)	*664*	96
Typed	6 (75)	18 (62)	58 (100)	19 (95)	11 (92)	23 (72)	35 (90)	63 (97)	170 (94)	42 (89)	117 (95)	68 (84)	*630*	87
Aff	2 (25)	15 (52)	29 (50)	0 (0)	1 (8)	3 (9)	8 (21)	27 (42)	117 (65)	39 (83)	104 (85)	64 (79)	*409*	59
Avg %	70	66	67	62	62	59	72	73	73	88	91	88		

Yearly genotype prevalence

Viruses from GII were predominant throughout all years (table 2.2). A total of 577 (91%) out of 631 outbreaks for which the genotype was determined were caused by GII noroviruses, 36 (6%) by GI or GIV viruses, and 18 (3%) were mixed infections of different genotypes, mainly existing of GII strains. The period 1993-1994 was the only season during which no outbreaks caused by GII.4 were found (figure 2.2 and table 2.2). After that GII.4 has consistently been the most commonly detected genotype. The fraction of GII.4 outbreaks varied from 23% in 2000-2001 to 89% in 2004-2005 and had an undulating distribution (figure 2.2). During the four seasons in which rises in the number of outbreaks were observed, the fraction of GII.4 strains also peaked. In 1995-1996, 2001-2002, 2002-2003 and 2004-2005 the percentages were 82%, 71%, 83% and 89%, respectively. In the years following these epidemic years, the fraction of GII.4 dropped to below its average level (68%), to 39%, 55% and 32% in

the seasons 1996-1997, 2003-2004 and the first half of the 2005-2006 season, but not during the season following 2001-2002. The diversity of genotypes per season was highest during the three consecutive years between 1999 and 2002, when 9, 12 and 8 different norovirus genotypes were detected, against numbers ranging from 2 to 6 different types per season in the other seasons (table 2.2). The majority of all winter outbreaks were caused by strains of GII.4, whereas strains of other genotypes did not exhibit such clear seasonality or rises in numbers (figure 2.2; the line of GII.4 outbreaks dictates the line of all outbreaks).

Strains from the genotypes GIIb and GIIc were first found in the Netherlands in the season 2000-2001, and GIIb has continuously been found since then. Strains with GIIc polymerase sequences were only found in 2001. The GIIb and GIIc strains are recombinant viruses, having a unique polymerase sequence coupled to differing capsids [19,215,250,252] and have not yet been assigned to one of the numerical genotypes, since they have no own capsid protein [94]. The total number of outbreaks in which these strains were detected in the Netherlands amounted to 38. For 17 of these recombinant strains the capsid genotype was determined, showing that 10 had GII.3 capsid proteins, 6 of GII.1 and 1 of GII.2 (data not shown). Strains from GIIc were only found in one season.

Table 2.2. Numbers of outbreaks per norovirus genotype found per season. Totals per genogroup are given in the rows in bold print. Percentages for each genotype are given in the last column. In the bottom row the total numbers of genotypes found per season are given. All genotypes found in mixed outbreaks were counted individually (data not shown). # GTs: number of genotypes detected in one season.

Genotype	1993-1994	1994-1995	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	*Total*
I.2							1	2	1				1	*5 (0.72)*
I.3					1		1		6	3				*11 (1.58)*
I.4			1				1	3						*5 (0.72)*
I.5									1				1	*2 (0.29)*
I.6				1	2			1	1	4	2			*11 (1.58)*
GI total			1	1	3		3	6	9	7	2		2	*34 (4.9)*
II.1		1	1			1	6	8						*17 (2.45)*
II.2								2	5		1	1		*9 (1.29)*
II.3	3		1	3	4	3	2							*16 (2.30)*
II.4		10	54	7	8	5	16	13	64	128	18	143	6	*472 (67.9)*
II.5	1					2		2						*5 (0.7)*
II.7				5		1	2		1	1	3	3	1	*17 (2.5)*
II.8						1								*1 (0.14)*
IIb								7	7	4	7	1	2	*28 (4.0)*
IIc								10						*10 (1.4)*
IINA				1			1							*2 (0.3)*
GII total	4	11	56	16	12	13	27	42	77	133	29	148	9	*577 (83)*
IV					1							1		*2 (0.3)*
Mixed	2			1		2	1	7	3		1	1		*18 (2.6)*
Untyped	1	3	9		1	5	9	1	1	14	1	11	8	*64 (9.2)*
Total	*7*	*14*	*66*	*18*	*17*	*20*	*40*	*56*	*90*	*154*	*33*	*161*	*19*	*695 (100)*
# GTs	5	2	4	6	5	6	9	12	8	5	6	6	5	*17*

In eighteen outbreaks more than one norovirus genotype was found, up to four different types in one outbreak. The transmission mode for 8 of these was reported and was foodborne in 3 outbreaks, waterborne in 1 and person to person in 4 outbreaks. In the three foodborne outbreaks two (GI.4 and GIIa), four (GI.1, GI.4, GIIb and GIV) and two (GII.3 and GII.7) genotypes were detected. In the waterborne outbreak two genotypes were detected (GI.3 and GI.6). The person to person mixed outbreaks occurred in daycare centers (2 outbreaks) and in healthcare settings (2 outbreaks, hospital and residential institution).

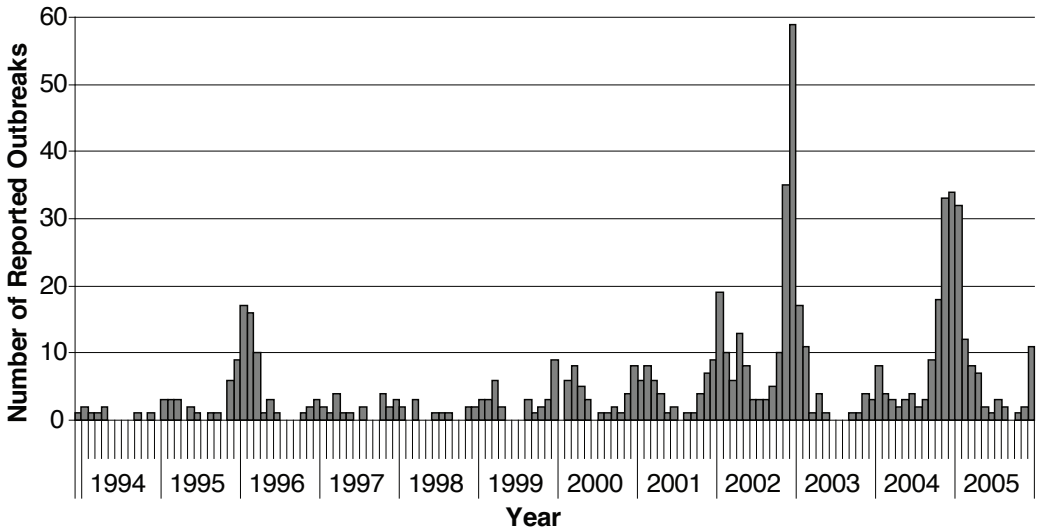


Figure 2.1. Number of norovirus outbreaks reported per month in the Netherlands.

Setting and genotype-distribution per setting

In total, 548 (79%) outbreaks were reported in healthcare settings (table 2.3, figure 2.3a) versus 104 (15%) in non-healthcare settings and for 43 (6%) outbreaks no data was available about the setting.

In the healthcare settings (hospital and residential institution) 407 of 502 outbreaks (81%, not counting the outbreaks where the genotype was not determined) were caused by norovirus strains belonging to GII.4. In the non-healthcare settings the share of GII.4 was 43%, 39 OBs out of a total of 91 (blanks for genotype not counted) (figure 2.3a). This difference is significant ($p<0.0001$, no matter whether blanks for genotype and / or setting were included as nonGII.4 or non-healthcare), and the relative risk of obtaining a GII.4 infection in a healthcare setting is at least 2.17 times bigger than in other settings (when blanks are counted as nonGII.4 and non-healthcare setting).

Transmission route per genotype

Only for 273 outbreaks (39%) a mode of transmission was entered (table 2.3). In 73% of outbreaks where the main transmission route was reported to be person to person GII.4 strains were found (figure 2.3b), versus 44% GII.4 in food related transmission modes. This difference is significant ($p<0.0001$) and when the blanks for genotype are counted as non-GII.4 the relative risk of finding GII.4 in a person to person transmitted outbreak is 2.3.

In 11% of the outbreaks where the mode of transmission was reported to be possibly foodborne, multiple norovirus strains were found and for the transmission mode mainly foodborne no mixed infections were reported. For other modes of transmission the share of mixed norovirus infections was much lower (ranging from 0-3%). The numbers are too low for reliable statistical analysis.

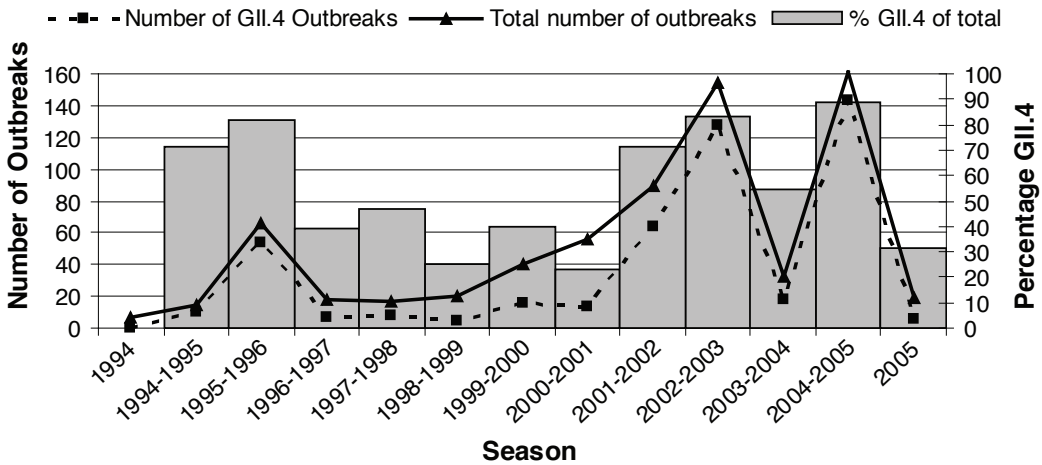


Figure 2.2. Total number of outbreaks per season and fraction of GII.4 outbreaks reported in the Netherlands. Total number is indicated by the solid line, number of GII.4 outbreaks by the dotted line (values on left Y-axis). Bars indicate the percentage of GII.4 outbreaks of the total number (values on right Y-axis). Arrows indicate the epidemic seasons. Seasons run from July through June.

Discussion

The completeness and uniformity of epidemiological data provided in outbreak reports is crucial for enabling the linking of epidemiological data to virological data and proper interpretation. A good cooperation with the reporting MHS's has been crucial for this surveillance system, and it is vital for the dataset that both epidemiological and molecular data are known. Better and more uniform completion of data would improve the value of the dataset. Also, adding extra data with information about the affected patients, such as age, blood group and secretor status would allow more detailed analyses of the epidemiology of individual norovirus genotypes. This would, however, also mean a greater invasion of patient's privacy.

Table 2.3. Modes of Transmission and Settings per genotype. GT: genotype. Left side of table: Numbers (#) and percentages (%) per genotype, per indicated Mode of Transmission. B: Blank, MW: Mainly Waterborne, MF: Mainly Foodborne, PF: Possibly Foodborne, PTP: Person to Person, B / U / O: Blank, Unknown or Other. Right side of table: Numbers (#) and percentages (%) per genotype, per Setting. T: Transport, D: Daycare Centre, C: Caterer, H: Hospital, H/GH: Hotel/ Guest House, PH: Private House, RI: Residential Institution, R/C: Restaurant / Canteen, S: School, S/R: Shop / Retailer, B / O / U: Blank, Unknown or Other.

	Mode of Transmission											**Setting**																					
	MW		**MF**		**PF**		**PTP**		**B/ U / O**		**Total**	**T**		**D**		**C**		**H**		**H / GH**		**PH**		**RI**		**R / C**		**S**		**S / R**		**B / U / O**		***Total***
GT	#	%	#	%	#	%	#	%	#	%		#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	
I.2			1	*20*			1	*20*	3	*60*	*5*					1	20					1	20	1	20						2	40	*5*
I.3	1	*9*							10	*91*	*11*					1	9	1	9			1	9	2	18	3	27	2	18			1	9	*11*
I.4					1	*20*	2	*40*	2	*40*	*5*	1	20											3	60	1	20							*5*
I.5									2	*100*	*2*							1	50					1	50									*2*
I.6					2	*18*	3	*27*	6	*55*	*11*	1	9					2	18	1	9			2	18	2	18	1	9			2	18	*11*
II.1							7	*41*	10	*59*	*17*			1	6			3	18					12	71			1	6					*17*
II.2			1	*11*	1	*11*	1	*11*	6	*67*	*9*			1	11	1	11	2	22					2	22	1	11	1	11			1	11	*9*
II.3			2	*13*	2	*13*	2	*13*	10	*63*	*16*			2	13	2	13	1	6					8	50	1	6	1	6	1	6			*16*
II.4			7	*1*	11	*2*	163	*35*	291	*62*	*472*	3	1	10	2	7	1	91	19	4	1	1	0	316	67	13	3	1	0			26	6	*472*
II.5							1	*20*	4	*80*	*5*	1	20					2	40					2	40									*5*
II.7			1	*6*	1	*6*	4	*24*	11	*65*	*17*							3	18					11	65	2	12	1	6					*17*
II.8									1	*100*	*1*													1	100									*1*
IIb					4	*14*	11	*39*	13	*46*	*28*			1	4	1	4	2	7	1	4			15	54	5	18					3	11	*28*
IIc					2	*20*	7	*70*	1	*10*	*10*			1	10	2	20	4	40					2	20	1	10							*10*
IINA									2	*100*	*2*													2	100									*2*
IV			1	*50*			1	*50*			*2*													1	50	1	50							*2*
mix	1	*6*			3	*17*	4	*22*	10	*56*	*18*			2	11	2	11	3	17					6	33			2	11			3	17	*18*
B			3	*5*	3	*5*	17	*27*	41	*64*	*64*			2	3	1	2	8	13					38	59	7	11	2	3			6	9	*64*
Total	*2*	*0*	*16*	*2*	*30*	*4*	*224*	*32*	*423*	*61*	*695*	*6*	*1*	*20*	*3*	*18*	*3*	*123*	*18*	*6*	*1*	*3*	*0*	*425*	*61*	*37*	*5*	*12*	*2*	*1*	*0*	*44*	*6*	*695*

The number of reported outbreaks per year in the Netherlands has generally increased over the past decade. The most likely explanation for this rise is improved reporting, as physicians and healthcare consultants became aware of norovirus as a problem and benefited from the free viral GE outbreak diagnosis performed at the RIVM. Alternatively, the number of outbreaks may have increased. This can not be proven based on this passive surveillance data alone, but is suggested by the increased prevalence of GII.4 in recent years. We reported a shift in the predominant GII.4 strain in 2002, which was observed throughout Europe [31]. Similarly, in 2004 we found another emerging variant of GII.4, slightly different from the 2002 variant, which again was associated with high numbers of outbreak reports [156]. This, in combination with the data presented in this study, shows that the emergence of new variants of GII.4 has preceded epidemic rises in reported norovirus outbreaks during the seasons 1995-1996, 2001-2002, 2002-2003 and 2004-2005.

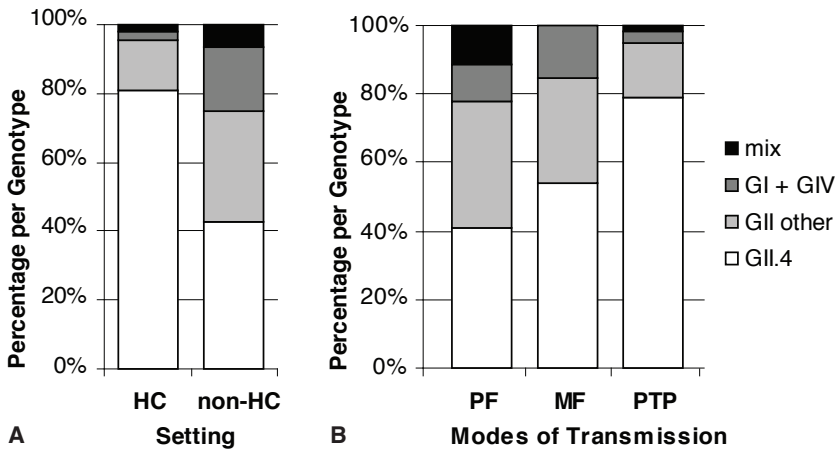


Figure 2.3. Genotype distributions for settings and transmission mode. A) Genotype distribution for healthcare (HC) (residential institutions and hospitals) and non-healthcare (non-HC) settings. Outbreaks for which no data for the setting (n=38) was available or for which no genotype was determined (n=64) are not depicted. B) Genotype distribution per mode of transmission. PF: Possibly Foodborne, MF: Mainly Foodborne, PtP: Person to Person. Waterborne outbreaks (n=2), outbreaks for which no data for the transmission mode was available (n=400) or outbreaks for which no genotype was determined (n=46) are not depicted.

Strains from GII.4 were predominant in the Netherlands throughout all years of surveillance, as was reported for other countries [17,75,334], although we did not find them in outbreaks in 1993 and 1994. It is not known whether GII.4 noroviruses were first introduced in the Netherlands at that time, or whether they were present and circulating in the population before. During the seasons with increased numbers of outbreaks the share of GII.4 simultaneously peaked (figure 2.2). This shows that the GII.4 strains were responsible for these norovirus epidemics, and had somehow obtained one or more biological advantages compared to previous variants of GII.4 as well as over other norovirus genotypes. Three of

these four epidemic seasons were followed by years in which not only the total number of outbreaks was lower, but the percentage of GII.4 strains also dropped markedly, suggesting that the population built up immunity against these predominant strains. The fact that for the year 2001-2002 a high number of outbreaks was reported, is attributable to the emergence of a new variant in the spring of 2002, causing uncharacteristic numbers of outbreaks in April, May and June of 2002, and the subsequent epidemic in the winter of 2002-2003.

As in other countries, strains belonging to the recombinant genotypes GIIb were first discovered in the year 2000 and GIIc in 2001 (table 2.2). In France GIIb strains were first found in September 2000 [19], in Sweden in May 2001 [215] and in Hungary in October 2000 [250,252]. In the Netherlands the first GIIb strain was found in November 2000. If and how the epidemiology of norovirus is influenced by recombination events remains to be determined. For this, standardized double typing for both the polymerase and the capsid gene is needed.

As in previous surveillance overviews from the Netherlands [320,322], person to person spread was reported to be the major transmission mode for norovirus (figure 2.3b, table 2.3). It is likely that the number of outbreaks reported for the healthcare settings is overrepresented in this survey, as well as in studies done by others, because reporting of outbreaks of illness in healthcare settings is mandatory. Outbreaks in other settings, such as private homes, are likely to be underreported. Thus it is not inconceivable that the contribution to GE outbreaks of a number of genotypes, other than GII.4, is underestimated.

For the outbreaks in which multiple norovirus strains were detected, where the mode of transmission was known, not more than 50% of the routes that were reported were food- or waterborne, the other 50% were attributed to person to person transmission. Outbreaks in which multiple norovirus genotypes are detected relatively often have contaminated foodstuffs as a starting point. It should be noted that the level of evidence for transmission mode entries is based on expert judgment, which may not always be correct. If the initial cases in a foodborne outbreak are overlooked, the high rate of secondary, person to person, transmissions may mask this foodborne start within days. Therefore, if foodborne outbreaks are not immediately detected, by the time the investigation is done they will mimic person to person outbreaks.

We found that 81% of the outbreaks in healthcare settings were caused by GII.4, as has been observed in other studies covering a shorter reporting period [17,19,75,82]. This preferential occurrence of GII.4 in healthcare settings suggests that GII.4 norovirus are more transmissible than other norovirus genotypes in healthcare settings, where close contact of

many people favors rapid person to person transmission. Increased transmissibility could result from increased levels or prolonged shedding of GII.4, or increased stability of the virus particles outside the host compared with other genotypes. Other factors may be lesser hygiene in these environments and reduced resistance to viral infections of the people at risk involved [27].

Over twelve years of viral gastroenteritis surveillance in the Netherlands has generated a valuable dataset for more fundamental virological research of the causative agents of these outbreaks. Continuing this surveillance and gathering more complete data in the future in order to enable better comparison will further increase the value of this dataset. The European surveillance network that has been set up in recent years to further expand our understanding of viral gastroenteritis will also contribute to this [146,178]. Faster gathering of more complete data will enable us to detect possible common source outbreaks more easily and more quickly. Nonetheless, the dataset that is currently available should be interpreted with care, as reporting biases are likely to influence it in a number of areas.

Based on this surveillance overview we conclude that noroviruses are the most common cause for acute gastroenteritis outbreaks in the Netherlands. The GII.4 strains stand out because of the high number of outbreaks they cause. Additionally, properties like their preferred outbreak-setting and transmission route, and the fact that they seem to be the main contributors to the seasonality observed, at least during the past decade, set them apart from the other norovirus strains. Perhaps even more importantly their ability to cause large and worldwide epidemics on top of their ever-presence in the population are features that make it a worthwhile challenge to research what virological and molecular properties of this particular genotype make them more successful in infection of humans throughout the years. Future research is addressed at understanding the molecular basis for the observed changes in epidemiology, with the ultimate goal to work towards improved control.

Acknowledgements

We thank E. de Bruin and B. van de Veer technical assistance. For assistance with statistical analysis we thank L. Verhoef. Finally we wish to thank A. Kroneman for assistance with the databases.
This work was supported by the European Commission, DG Research Quality of Life Program, 6th Framework (SP22-CT-2004-502571).

Chapter 3

Epochal Evolution of GII.4 Norovirus Capsid Proteins from 1995 to 2006

J. Joukje Siebenga[1,2], Harry Vennema[1], Bernadet Renckens[3], Erwin de Bruin[1],Bas van der Veer[1], Roland J. Siezen[3] and Marion Koopmans[1,2]

1 National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands,
2 Erasmus Medical Center, Rotterdam, The Netherlands,
3 Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands

Published in JOURNAL OF VIROLOGY
• http://jvi.asm.org • Vol. 81, No. 18, Sept. 2007, p. 9932–9941 • DOI:10.1128/JVI.00674-07

Abstract

Noroviruses are the causative agents of the majority of viral gastroenteritis outbreaks in humans. During the past 15 years, noroviruses of genotype GII.4 have caused four epidemic seasons of viral gastroenteritis, during which four novel variants (termed epidemic variants) emerged and displaced the resident viruses. In order to understand the mechanisms and biological advantages of these epidemic variants, we studied the genetic changes in the capsid proteins of GII.4 strains over this period. A representative sample was drawn from 574 GII.4 outbreak strains collected over 15 years of systematic surveillance in The Netherlands, and capsid genes were sequenced for a total of 26 strains. The three-dimensional structure was predicted by homology modeling, using the Norwalk virus (Hu/NoV/GGI.1/Norwalk/1968/US) capsid as a reference. The highly significant preferential accumulation and fixation of mutations (nucleotide and amino acid) in the protruding part of the capsid protein provided strong evidence for the occurrence of genetic drift and selection. Although subsequent new epidemic variants differed by up to 25 amino acid mutations, consistent changes were observed in only five positions. Phylogenetic analyses showed that each variant descended from its chronologic predecessor, with the exception of the 2006b variant, which is more closely related to the 2002 variant than to the 2004 variant. The consistent association between the observed genetic findings and changes in epidemiology leads to the conclusion that population immunity plays a role in the epochal evolution of GII.4 norovirus strains.

Introduction

Since the beginning of viral gastroenteritis outbreak surveillance in the early 1990s, noroviruses have become recognized as the major cause of reported outbreaks of acute viral gastroenteritis worldwide. Noroviruses form a genus within the family Caliciviridae and are genetically and antigenically highly variable. Currently, five distinct genogroups are recognized. Strains belonging to GI, GII, and GIV are known to cause infections in humans. The genogroups have been subdivided further into genotypes, defined by a minimum amino acid sequence identity over the complete capsid sequence of 80% [9].

The strains most commonly identified as the cause of outbreaks belong to genotype GII.4. In The Netherlands, this was the case for 68% of all norovirus outbreaks that were characterized during 12 years of surveillance and for up to 81% of all health care-related outbreaks. Since their first detection in The Netherlands in January 1995, the GII.4 strains have consistently been present in the Dutch population [268]. These observations are in agreement with those of other surveillance studies worldwide [17,27,74,82,156,218,330].

During the past 15 years, four epidemic norovirus seasons have occurred, in the winters of 1995-1996, 2002-2003, 2004-2005, and 2006-2007. These worldwide epidemics were invariantly caused by the predominant genotype, GII.4, and were attributed to the emergence of new variant lineages of this genotype [27,175,212,320,322]. These genetic variants, which have been identified previously by partial sequencing of either the RNA-dependent RNA polymerase (RdRp) or the capsid gene, have been given several names across the world. Here they are referred to by using the first year of their detection, supplemented where necessary with an extra suffix. The following variants have been identified: <1996, 1996, 2002, 2004, 2006a, and 2006b.

The pattern of emergence of new lineages followed by large-scale epidemics suggests that new variants obtained one or more decisive advantages over the previously circulating predominant variant. It is unknown what the nature of this advantage is, but its basis is likely to be found in VP1, since this protein is needed for essential properties and functions in the viral life cycle, such as antigenicity, host specificity, host cell binding and virus entry properties, and assembly of new particles.

Noroviruses have a positive-strand RNA genome of ~7.6 kb, which is subdivided into three open reading frames (ORFs). ORF1 encodes a polyprotein which is posttranslationally processed into the nonstructural proteins, including the RdRp. Conserved regions within the RdRp are commonly used as targets for diagnostic PCR assays. At the National Institute for Public Health and the Environment in The Netherlands (RIVM), region A (nucleotides 4279 to 4604; Lordsdale genome numbering [GenBank accession no. X86557]) is commonly

used for genotyping outbreak strains. The second ORF (ORF2) encodes the major structural protein VP1. Ninety dimers of this capsid protein form a T=3 icosahedral shell [241]. In the virion, a small number of copies of the protein encoded by ORF3 are present. The precise role of this protein is not clear, although it has been suggested that it functions both in upregulation of VP1 expression and as a histone-like protein in stabilizing the capsid-RNA complex [15,90,113].

The understanding of immunity against noroviruses remains limited. Between the different genogroups and genotypes, antigenic differences as well as cross-reactivities have been demonstrated using virus-like particles and polyclonal antisera [107]. Short-term immunity was reported, but preexisting antibodies were not protective against reinfection with the same genotype [128,227,336]. Studies looking at neutralizing antibodies have not been possible due to the lack of cell culture or small-animal model systems [69]. The high level of genetic diversity between different genogroups and even between genotypes within the same genogroup resulting from the high mutation rate and from recombination events contributes to a large degree of antigenic diversity.

Host genetic factors determining the presence or absence of virus receptors also play an important role in susceptibility [109,119]. These receptors, the histo-blood group antigens, show virus strain-specific binding patterns, determining the ability of virus to infect potential host cells. Because noroviruses belonging to GII.4 have the broadest range of binding to the histo-blood group antigens of all genotypes assayed to date, this may explain part of the relative success of these viruses [120]. Other success factors may include a higher stability of the viral particles outside the host, a higher replication rate, or other factors that need to be investigated more thoroughly.

To obtain more insight into the genetic and structural bases of the selective advantage of new GII.4 variants over the old GII.4 variants, we determined the complete capsid sequences of a systematic sample of GII.4 norovirus outbreak strains found in The Netherlands during 13 years of surveillance of viral gastroenteritis and studied their genetic diversity and predicted structure [268]. Because a high-resolution three-dimensional (3D) model of GII noroviruses was lacking at the time this study was initiated, a homology model of the capsid protein was made *in silico* based on the known 3D structure of the Norwalk virus (NV; GI.1) capsid protein.

Materials and methods

Strain selection

Norovirus outbreak strains for which the capsid sequence was determined were selected from the norovirus surveillance database used at RIVM. In this Bionumerics database (Applied Maths BVBA, Sint-Martens-Latem, Belgium), epidemiological and virological

data for all norovirus strains found in The Netherlands since January 1994 are collected as previously described [268]. The norovirus surveillance system is a laboratory-based passive reporting system to which municipal health services can submit patient samples from suspected viral gastroenteritis outbreaks. Because there is mandatory reporting for outbreaks of illness in institutions and RIVM has been the only laboratory providing diagnostic services for noroviruses in The Netherlands, the collection represents a national sample of reported outbreaks.

As the first step, all GII.4 strains detected between January 1994 and December 2004 were selected from the database. A phylogenetic tree was made based on partial polymerase sequences of 145 nucleotides for the older sequences to 250 nucleotides for the strains isolated after 2001 (amplified with primer pair JV12 and JV13 or modifications thereof [region A]) [314,322]. The branching of the tree was used to guide the selection of outbreak strains for this study, with at least two strains per branch selected when sufficient material was available. Following reports of unusual outbreaks in the spring of 2006 [145,155], six strains from this period were included in the study. A minimum spanning tree (MST) was made on the basis of 145 nucleotides of the polymerase gene, using the default settings in Bionumerics, to give an overview of the distribution of strains available in the database. An MST is a tree that connects all samples from a database in such a manner that the summed distance between all samples or branches is minimized. An MST is particularly useful for representing large (genomic) data sets with relatively high similarity levels and, as such, has been shown to enable representation of microevolution or population modeling [258,259]. Another condition is that the data set should represent the biodiversity for the organism under study and therefore should have been gathered over a time period that is short relative to the expected rate of change for the organism. During tree formation, the sample with the largest number of related samples is chosen as the root node, and subsequent branches are added in order of relatedness.

Viral RNA isolation, cDNA preparation, and sequencing

Stool specimens taken in selected outbreaks were collected from the biobank. Specimens were stored as undiluted stools at 4°C, as 10% fecal suspensions at 4°C, and as RNA extracts frozen at –80°C. Where available, extracted RNAs were used as the template. When this failed to yield a PCR product, a new RNA extraction was done from diluted stool or fecal samples. Sequencing of these samples was done as described previously [49]. Briefly, RNA was reverse transcribed in overlapping fragments, using avian myeloblastosis virus reverse transcriptase (Invitrogen), and subsequently, the obtained cDNA was amplified and sequenced using an ABI Prism BigDye Terminator v3.0 ready reaction cycle sequencing kit. The primers that were used were the following: TCTCAGATCTGAGCACGTGG (GR19A), AACAGTTAAGATTGGGACG (GR19B), GTCTCTTGTCGAGTTCTCACG (GR20), GGTGAATTGAACACTACCCAGC (GR21), CTCGACCC-

GTGCCCACAAAGC (GR22), CATTATAATGCACGCCTGCGCC (GR23), GGGTCAACCAGTTCTACACAC (GR24), CCAGCTGAAGAACCTAGTCTCG (GR25), ACGTGCCCAGGCAAGAGCCAAT (GRJS1), TAACATCTACTATTATATGGG (GR-JS2), TCATATTTGCAGCAGTCCCA (GR20A), CTCTGAAGGTGCAGATGTTG (GR21A), TGTGAATCCAGACACAGGTAG (GR24A), and ACGGGCCGCATCTGCTGTGGAA (GR25A).

Data processing

DNA sequences were processed using SeqMan and EditSeq (DNAStar Inc., Konstanz, Germany) and aligned and analyzed using the BioEdit sequence alignment editor (Isis Pharmaceuticals Inc.). Alignments were done manually or using ClustalW alignment algorithms in BioEdit. Informative sites were determined by ProSeq 2.91. Sites were considered informative when at least two strains had an identical amino acid mutation in the alignment. Informative sites discriminating subsequent epidemic variants were also determined. Epidemic variants were defined as GII.4 strains that were dominant for at least one outbreak season following initial detection. Silent mutations (nucleotide mutations which caused no amino acid mutation) or replacement mutations (nucleotide mutations which caused amino acid mutations) were determined using ProSeq 2.91, with an insertion considered a single mutation.

Phylogenetic analyses were done in Bionumerics, using neighbor joining and the unweighted-pair group method using average linkages, with 1,000 bootstrap resamplings and no correction and with the gap cost set at 5%. Trees were plotted using the program Treeview (version 1.6.6) [224] or Treecon (version 1.3b) [308], with the exception of the MST, which was calculated as well as plotted in Bionumerics. For phylogenetic analysis of the partial polymerase and capsid sequences, the following sequences from GenBank were included: Grimsby strain (Hu/NoV/GGII.4/Grimsby/1995/UK; GenBank accession no. AJ004864), Farmington Hills strain (Hu/NoV/GGII.4/Farmington Hills/2002/USA; accession no. AY502023), Hunter strain (Hu/NoV/GII.4/Hunter 284E/2004/AU [accession no. DQ078794] for the capsid and Hu/NoV/GII.4/Hunter 532D/2004/AU [accession no. DQ078801] for the partial polymerase sequence), and Camberwell strain (Hu/NoV/GGII.4/1994/AU [accession no. U46500] and others for the capsid analysis).

Sequences were checked for possible recombination events by using Simplot (version 3.2), where the window size was varied from 80 to 150, with steps of 20 nucleotides, and a distance model with Jukes-Cantor correction was used. The capsid and RdRp sequences were analyzed independently, as well as after concatemerization of region A and the capsid sequences, to look for possible crossover in the joining region.

Homology modeling

The three-dimensional structure of the NV capsid protein (PDB code 1IHM) [240] was used as a template for homology modeling of the GII.4 capsid protein. Sequence alignments

were made using the program MUSCLE [70]. Compared to the NV capsid protein, the GII.4 capsid protein has four insertions of three to seven amino acids which cannot be modeled. Generally, such insertions are located in surface-exposed loops of proteins. Based on the alignment of the two sequences and on the 3D structure of the NV capsid protein, the most likely places for insertion were predicted. The GII.4 capsid protein also has one deletion of two amino acids compared to the NV capsid protein. The place of this deletion can be modeled and was chosen in the same way as that for the insertions. Homology modeling was performed with WhatIf/Yasara Twinset software (Yasara) [324].

Nucleotide sequence accession numbers

The complete capsid nucleotide sequences determined in this study are accessible in the DNA DataBank of Japan under accession numbers AB303922 through AB303941 and EF126961 through EF126966.

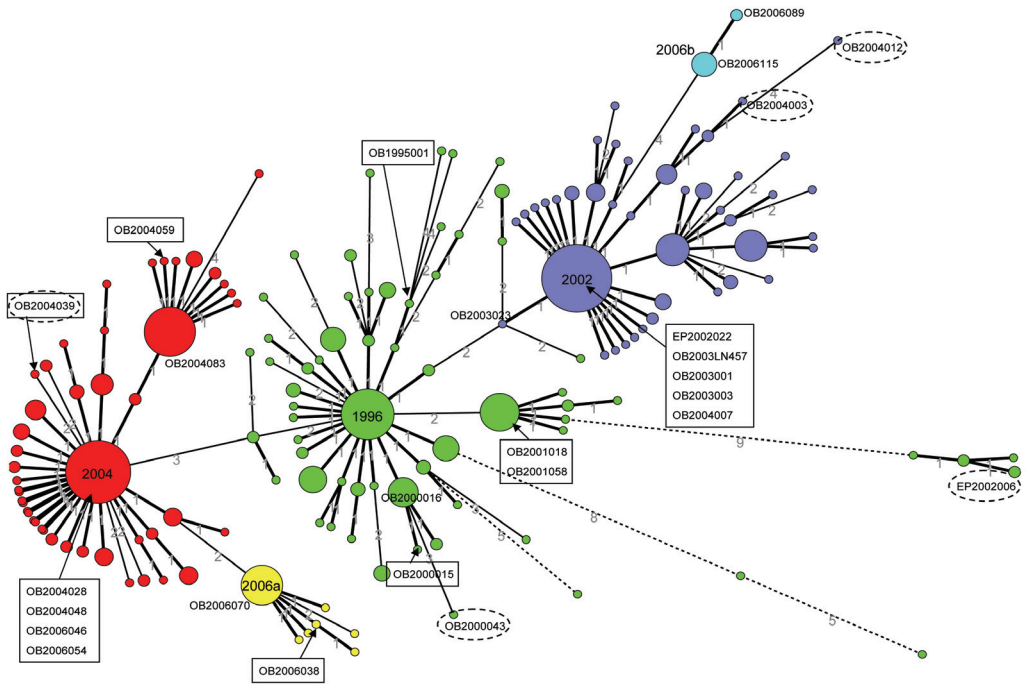


Figure 3.1. Minimum Spanning Tree. Based on alignment of 145 nucleotides of the polymerase gene sequences (region A) of all GII.4 strains found in The Netherlands between January 1995 and August 2006 (n = 574). Colors represent different variants, as indicated in the figure. The sizes of the circles are drawn to scale with their member counts. The smallest circles represent 1 strain, and the largest circle (the center of the 2002 cluster) represents 70 strains. Genetic distances between the circles, in numbers of nucleotides, are given on connecting lines. The total distance is 230 nucleotides. Strains included in this study are indicated. The strains shown as circles with dotted lines are considered outliers.

Results

Comparative phylogenies of the polymerase and capsid genes and sequence analysis

A phylogenetic tree (unweighted-pair group method using average linkages) was made to enable us to make a representative selection of strains for capsid sequencing. All GII.4 strains found in The Netherlands from the start of viral gastroenteritis surveillance in January 1994 up to August 2006 (n = 574) for which partial polymerase sequences (region A) were available were included (tree not shown). The minimum identity was 89.44%. The strains segregated into three major branches, with multiple outbreak strains per branch (n = 166, 161, and 180); some smaller clusters; and outlying strains that did not fit into any of the major groups. The major groups were each subdivided into smaller clusters. Strains were selected using this tree, aiming to obtain two sequences from the largest clusters from each major group and from five outlying strains. The total number of strains analyzed was 26.

The MST for the partial polymerase sequences is shown in figure 3.1. The total distance of the tree is 230 nucleotides. This tree was not used to make the selection of the strains. However, it illustrates the grouping of the different variants and the positioning of the selected strains because it takes into account the localization of nucleotide changes. Strains OB2000043, EP2002006, OB2004003, OB2004012, and OB2004039 are considered outliers based on their positions in the neighbor-joining tree for all polymerase sequences (not shown).

RNA sequences of the complete capsid genes were determined and aligned. All capsid sequences found belonged to GII.4. The neighbor-joining tree for the capsid sequences is shown in figure 3.2. Strains from GenBank were included for reference. Polymerase-based and capsid-based groupings were congruent for all strains, and thus no intragenotypic recombination was observed. SimPlot analysis of complete capsids as well as of complete capsids combined with the sequences of region A revealed no potential intergenotypic recombination sites (data not shown). The relatively high level of homology between all strains may obscure possible recombination events, and therefore recombination between strains belonging to the same variant of GII.4 cannot be ruled out.

It should be noted that three strains showed different groupings upon analysis of region A in the RdRp from those obtained in the analysis of the capsids. Three of the five strains that were outliers when comparing partial polymerase nucleotide sequences (OB2004039, OB2000043, and OB2004003) (figure 3.1) did fit into the capsid amino acid tree and fell into their respective variant groups (figure 3.2). Although OB2004012 clustered with the 2002 variant, it was still an outlying strain. EP2002006 was phylogenetically more similar to the eldest strains from GenBank and was therefore used as an ancestral strain in further analyses.

Analysis of the capsid gene and changes over time

Informative sites in the capsid sequences were then determined. An alignment of all informative sites in the capsid is represented in figure 3.3. Sites were considered informative when at least two strains had an identical amino acid or nucleotide mutation in the alignment.

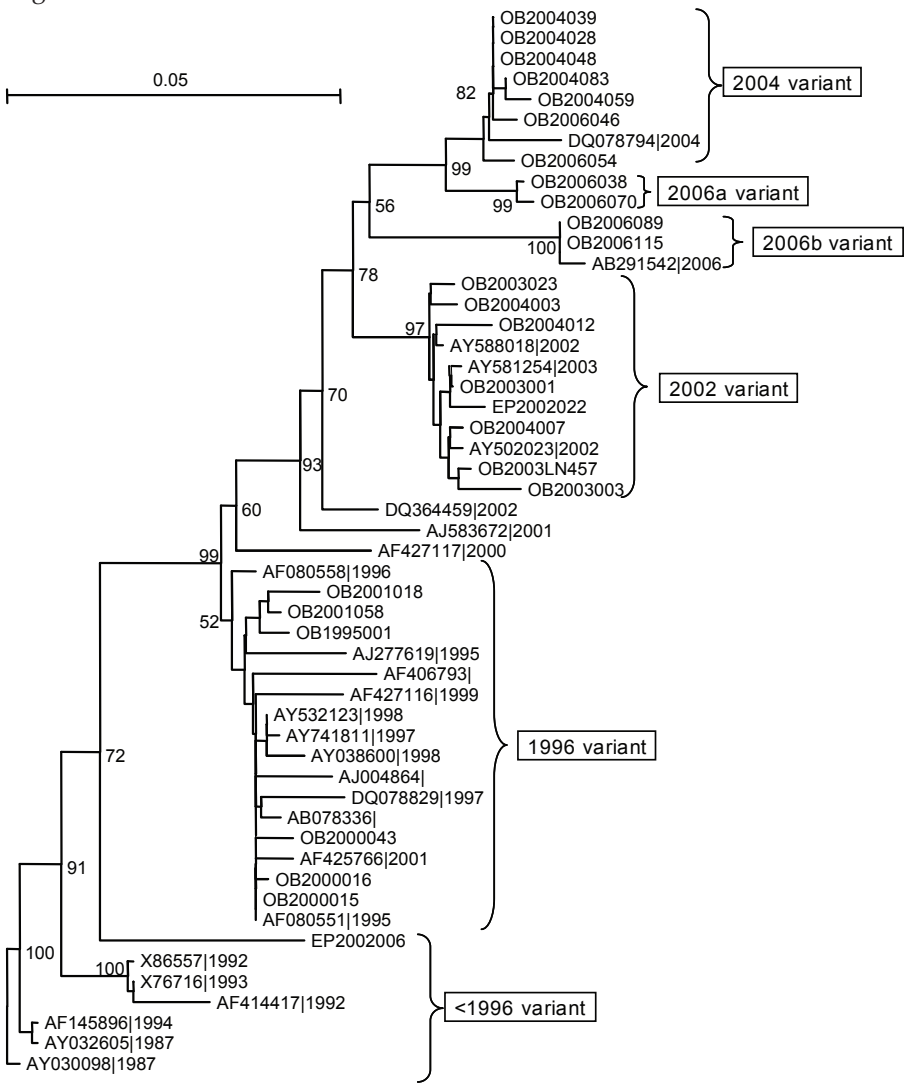


Figure 3.2. Neighbor-joining tree for complete capsid amino acid sequences. Type strains from GenBank were used in order to emphasize and confirm the groupings. Branch lengths are drawn to scale. Bootstrap values are percentages of 1,000 iterations.

At the amino acid level, 48 sites (9% of 541 amino acids) were informative (table 3.1). Thirty of these sites were located in the P2 domain (24% of the amino acids in this domain). In all

other domains, the numbers of informative sites were significantly lower (chi-square test; $P < 0.001$), as follows: 3 of 40 amino acids (8%) in the N-terminal domain, 4 of 182 amino acids (2%) in the shell domain, 10 of 184 amino acids (5%) in the P1 domain, and 1 of 10 amino acids (10%) in the C-terminal domain. Hypervariable sites (sites at which three or more different amino acids were found over the 12-year period) were found only in both protruding domains, with 13 in P2 and 2 in P1 (figure 3.3). One amino acid insertion was observed, at position 395, and was first detected in strains from The Netherlands in January 2002. This insertion was located in a highly variable loop region on top of the P2 domain.

At the nucleotide level, 267 sites were found to be informative (table 3.1). The P2 domain had a higher percentage of informative mutations than did the other domains: 24% of the nucleotide mutations in P2 were informative, 8% of those in the N-terminal domain were informative, 13% of those in the shell domain were informative, 16% of those in the P1 domain were informative, and 20% of those in the C-terminal region were informative. The differences with P2 were significant for the N-terminal and shell domains (chi-square test; $P < 0.01$), not for the P1 and C-terminal domains. Relatively high percentages of first- and second-position nucleotide mutations (20% and 22%, respectively) were seen for the P2 region compared to both the percentages found for the N-terminal and shell domains taken together (13% and 5%, respectively) and those found for the P1 domain (19% and 5%, respectively). Of these mutations, 44% were replacement mutations, versus 30% (3 of 10 mutations) in the N-terminal domain, 6% (4 of 70 mutations) in the shell domain, 18% (16 of 91 mutations) in the P1 domain, and 17% (1 of 6 mutations) in the C-terminal region. These differences with P2 are significant for comparisons of the P2 domain with the shell and P1 domains or the whole capsid sequence (chi-square test; $P < 0.001$).

As shown in figure 3.3, changes in informative sites occurred stepwise rather than gradually, with the steps coinciding in time with the emergence of each respective new epidemic variant (2002, 2004, 2006a, and 2006b). When consecutive variants were compared and EP2002006 was considered the precursor (<1996) of the 1996 variant, the numbers of stable amino acid mutations per emerging new variant were 14 (<1996 variant versus 1996 variant), 25 (1996 variant versus 2002 variant), 21 (2002 variant versus 2004 variant), 8 (2004 variant versus 2006a variant), 25 (2004 variant versus 2006b variant), and 23 (2002 variant versus 2006b variant). Both 2006a and 2006b were compared to the 2004 variant, since this was their temporal precursor. The 2006b variant was also compared to the 2002 variant, since these variants are genetically more closely related based on phylogenetic clustering (neighbor joining) (figure 3.2).

Prevalence of GII.4 variants in The Netherlands from January 1995 to February 2007

Since the capsid changes showed clustering in time of GII.4 variants and the capsid-based variant assignment was consistent with that based on the partial RdRp sequences used

for routine surveillance, we plotted the presence of the different GII.4 variant types in The Netherlands over time (figure 3.4). This figure shows that new variants invariantly replaced their predecessors within 5 months of cocirculation.

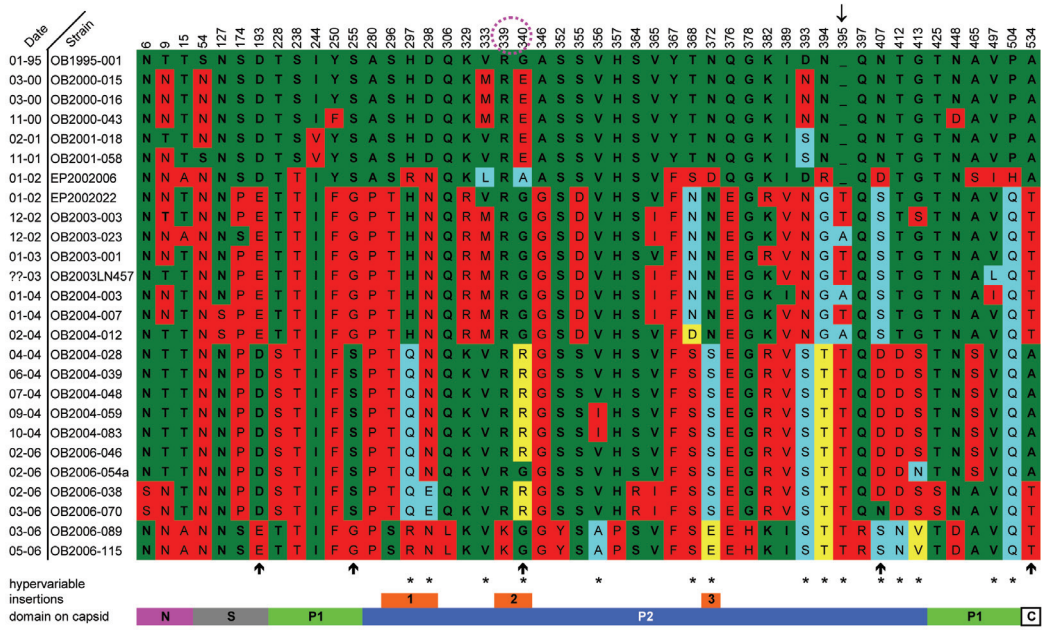


Figure 3.3. Fixed amino acid changes (informative sites) in capsid sequences of GII.4 outbreak strains collected between 1995 and 2006. The informative sites throughout the protein are listed from left to right. Amino acid numbering is indicated at the top, and outbreak dates (month-year of isolation, e.g., 01-95 is January 1995) and names are given on the left. From top to bottom, the same color indicates identical amino acids, and different colors are distinct amino acids. Colors were assigned by frequency; amino acids that occurred most are shown in green, followed by red, blue, and yellow (diminishing frequencies). The amino acids circled in magenta are part of the additional RGD motif present in the 2002 variant and the earliest strain. The arrow at the top indicates where an amino acid insertion occurred. The orange bars at the bottom indicate the locations of insertions in GII.4 compared to NV and correspond to insertions 1 to 3 in Fig. 5A. Asterisks indicate hypervariable sites (with more than one mutation), and the arrows below the sequences indicate the sites where an amino acid mutation occurs at each variant change (not including 2006a). Domains are indicated in the bar below the figure.

Analysis of structural polymorphism of the capsid protein

The sequence of OB2004039, a 2004 variant, was used as a reference for modeling of the basic 3D structure of capsid proteins used in this study. Amino acid differences in all other strains were plotted in this 3D model. Compared to the NV capsid protein, the GII.4 capsid protein has four insertions, of six, three, seven, and three amino acids, with the first three occurring in the P2 region and the fourth occurring in the second coding region of P1. These insertions were not modeled because of the poor reliability of such predictions. However, we predict that these insertions are located close together three-dimensionally, both intra-

and interdimerically. Three of the four insertions, all located in the P2 domain, had one or more informative sites, in contrast to the fourth insertion, located in the P1 domain, which had none. The locations of the insertions are shown as orange bars in figure 3.3 (note that the fourth insertion is not shown in this figure) and as orange arrows in figure 3.5A. In figure 3.5B, the inter- and intradimerical interactions of one dimer pair and one-half of the neighboring dimer are shown. In this figure, an extra RGD motif that is present in the earliest strain as well as in the 2002 variant strains is indicated by blue arrows. This motif is located at amino acids 339 to 341.

Most of the informative sites mapped to the surface of the P2 domain

Subsequent variants were compared pairwise to identify possible informative sites that consistently changed with every new epidemic variant. When informative sites of ensuing variant pairs (figure 3.6A, panels i through v) were listed, five amino acids changed between every variant pair when 2006a was not included in the analysis (figures 3.3 and 3.6B). These were amino acids 193 (D_{1996}->E_{2002}->D_{2004}->E_{2006b}), 255 (S_{1996}->G_{2002}->S_{2004}->G_{2006b}), 340 (E_{1996}->G_{2002}->R_{2004}->G_{2006b}), 407 (N_{1996}->S_{2002}->D_{2004}->S_{2006b}), and 534 (A_{1996}->T_{2002}->A_{2004}->T_{2006b}).

Discussion

During the past 15 years, four worldwide epidemics of acute gastroenteritis caused by emerging variants of GII.4 noroviruses have been described. Emerging new variant lineages replaced the previously circulating dominant types rapidly and completely (figure 3.4) [155,175]. The mechanisms underlying the emergence of these new lineages as well as the biological advantages they possessed over other circulating strains are not yet well understood. Most biological properties that are relevant for variables such as stability, assembly, antigenicity, host cell binding, and host specificity are incorporated into the major capsid protein of the virus. We studied the genetic variation in capsids of successive variant lineages to find possible clues about the improved fitness of the successive emerging variants.

In the analysis of the informative sites, for both the nucleotide sequences and the amino acid sequences, mutations were fixed at a number of sites. Every successive variant had a number of distinct, lineage-defining mutations, which were found throughout the capsid sequence. The highest densities of informative sites were located on the surfaces of the protruding regions of the capsid (figures 3.3 and 3.5). The P2 domain had significantly more mutations than the rest of the capsid protein and, more specifically, many more replacement mutations (0.11 per nucleotide, versus 0.01 to 0.03 per nucleotide for the other domains of the capsid) (table 3.1). This is clear evidence of selective force providing new variant viruses having certain mutations with an advantage over previously circulating variants.

Table 3.1. Informative sites in GII.4 capsid sequences[a]

	Amino acid			Nucleotide			No (%) of mutations at codon position				
	aa's	Length (aa)	No. (%) Inf	nt's	Length (nt)	No. (%) Inf	First	Second	Third	No. (%) Silent	No. (%) Non-sil
N	1-40	40	3 (8)	1-120	120	10 (8)	2 (20)	2 (20)	6 (60)	7 (70)	3 (30)
S	41-222	182	4 (2)	121-666	546	70 (13)	8 (11)	2 (3)	60 (86)	66 (94)	4 (6)
P1	223-276, 402-531	184	10 (5)	667-828, 1204-1593	552	91 (16)	17 (19)	5 (5)	69 (76)	75 (82)	16 (18)
P2	227-401	125	30 (24)	829-1203	375	90 (24)	18 (20)	20 (22)	52 (58)	50 (56)	40 (44)
C	531-541	10	1 (10)	1593-1623	30	6 (20)	1 (17)	0 (0)	5 (83)	5 (83)	1 (17)
total		*541*	*48 (9)*		*1623*	*267 (16)*	*46 (3)*	*29 (2)*	*192 (12)*	*203*	*64*

[a] Percentages for informative sites are given as fractions of the total domain length, in numbers of amino acids or nucleotides.
Position numbers are given for GII.4 strains. aa, amino acid(s); Inf, informative; nt, nucleotide(s).

The subsequent variants of GII.4 accumulated mutations in chronological order, and each descended from its predecessor in time, with the exception of the 2006b variant. At the amino acid level, this variant seemed more related to the 2002 variant (figure 3.2). This newly emerging variant is likely a descendant of a virus strain older than the 2004 variant that has accumulated quite a few mutations while not causing many outbreaks in the population.
The situation that is currently unfolding is highly intriguing. In the spring of 2006, two distinct new variants emerged, named 2006a and 2006b. These two new variants have been detected and reported worldwide, often in cruise ship-related outbreaks [145,155]. It has not been reported before that two norovirus variants can cause epidemic-scale outbreaks simultaneously. The 2006a variant shows 8 amino acid mutations compared to its predecessor, the 2004 variant, whereas the 2006b variant shows 25 amino acid mutations compared to the 2004 variant, its temporal and therefore immunologic predecessor. Both 2006 variants emerged almost simultaneously [243]. It will be interesting to see if both variants continue to cause outbreaks simultaneously in the population or if one proves to be more successful than the other, perhaps with differing patterns across the world.

The viral strains used for this analysis all originated from our outbreak surveillance database. Strains that are intermediate between the epidemic variants are likely to have reduced viral fitness and are therefore less likely to be detected on the basis of sampling from outbreaks. Although we did look for intermediate strains bridging the different variants by choosing to sequence a number of outliers from the polymerase alignment, no capsid sequences that could be considered intermediates between the different variants were found. EP2006006 does not fit with any variant of the strains included in this study. It does, however, show resemblance to the older strains from GenBank that were included in the neighbor-joining tree.

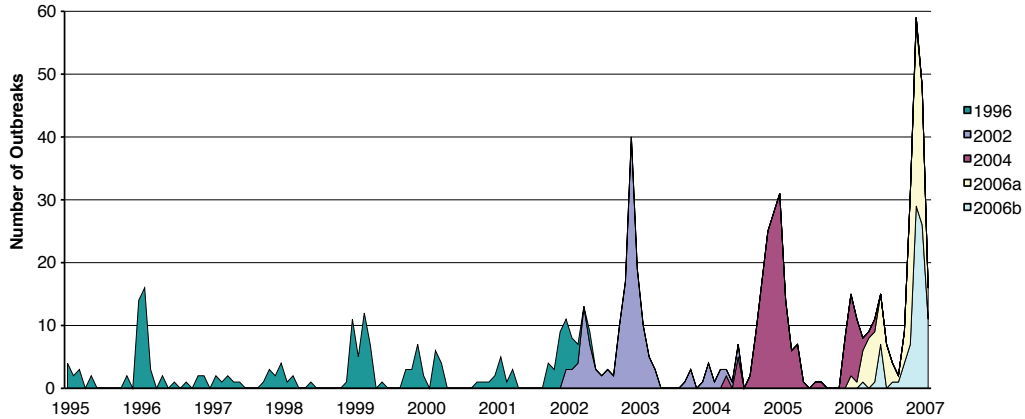


Figure 3.4. Prevalence of GII.4 variant types in The Netherlands between January 1995 and February 2007. Genotype and variant type assignments were done based on partial polymerase sequence data (region A) available from the Dutch norovirus surveillance database.

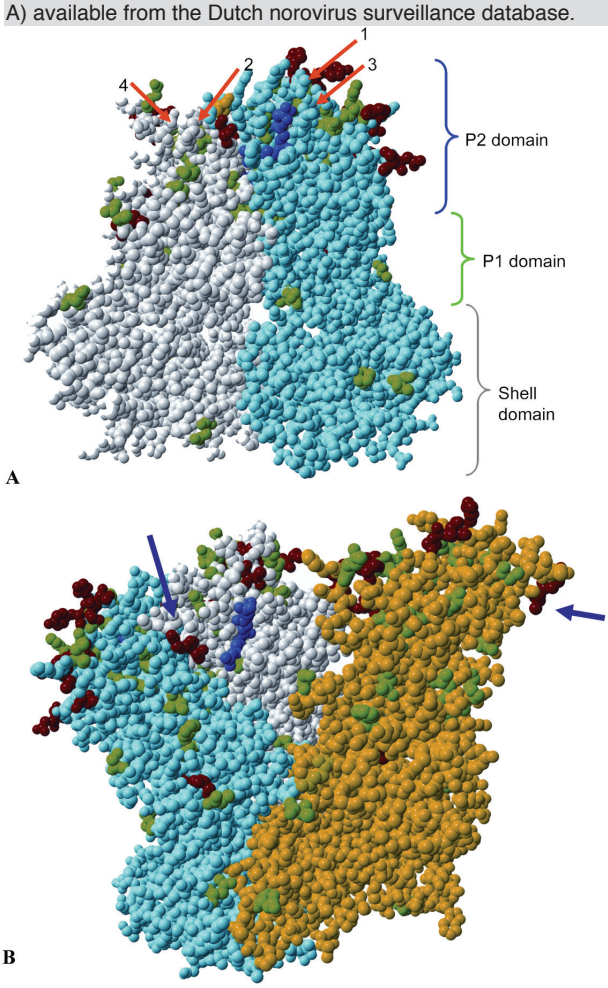


Figure 3.5. Informative sites mapped on 3D model of GII.4 capsid proteins. Sites with two distinct amino acid changes over the 12-year period are depicted in green, and sites with three or more amino acid changes are shown in red. The conserved RGD motif is shown in blue. A) Dimeric subunit of two capsids, with one in gray and one in light blue. The extra RGD motif is indicated in yellow. The locations of the insertions compared to the NV capsid, which have not been modeled, are indicated by orange arrows 1 to 4. The brackets on the right indicate the different domains. The shell domain is indicated in grey, the P1 domain is shown in green, and the P2 domain is shown in blue. B) Three capsid proteins, including a dimer with one-half of a neighboring dimer. The gray and light blue areas form one dimer, and the yellow capsid belongs to another dimer. The inserted RGD motif is indicated by the blue arrows.

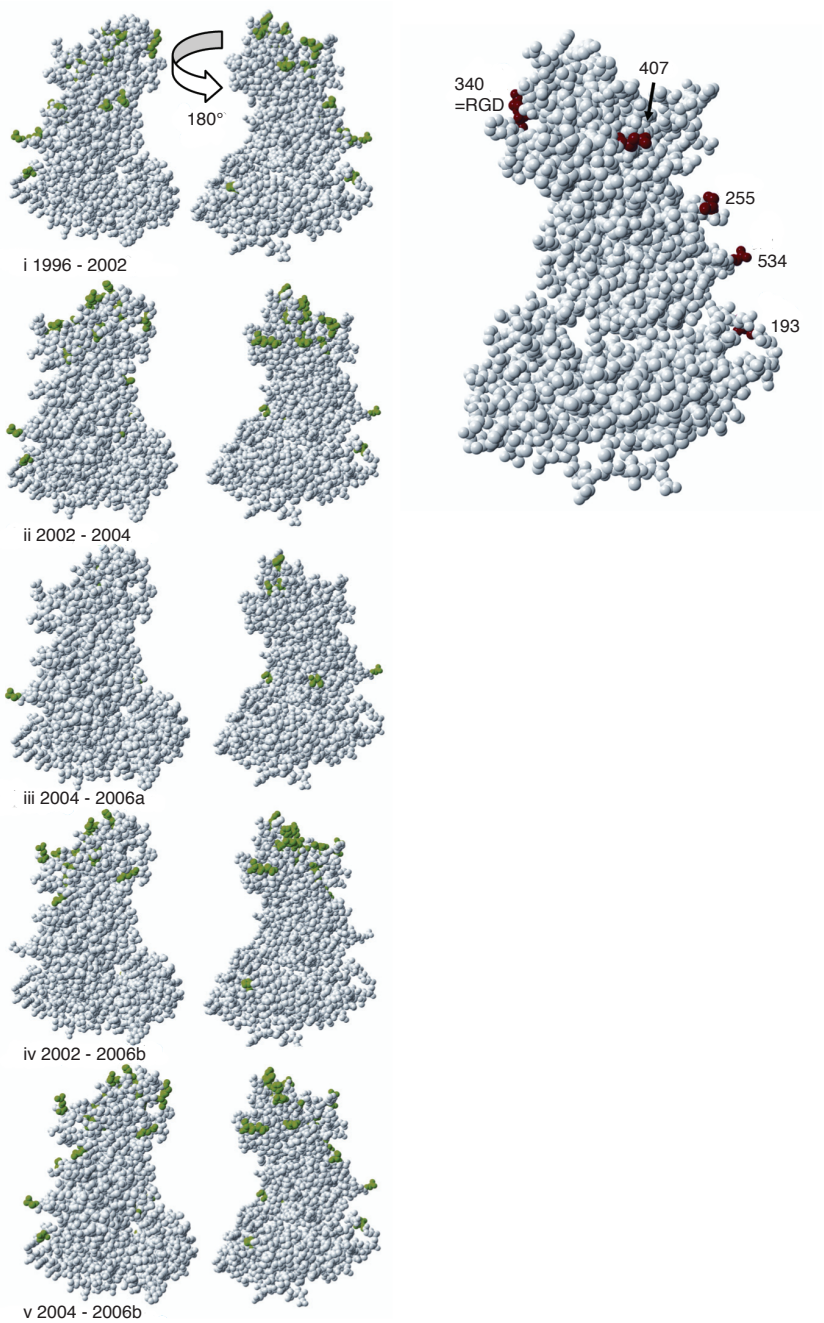


Figure 3.6. Changes in informative sites (green) derived from amino acid comparisons between subsequent epidemic variants. A) (i to v) For each comparison, two views of the capsid protein are given, with one frontal view and one from the rear. For the 2006b variant, two comparisons were made, with the phylogenetic precursor (2002) and the chronologic precursor (2004). B) Amino acids that change between every subsequent variant group, with 2006a not included.

Since no real intermediate strains were found, the origin of emerging variants or the reservoir in which they accumulate their defining mutations thus remains a subject for speculation. The most logical place is the general population. While not causing (many) outbreaks, strains may circulate in the population and not come to the attention of surveillance, slowly accumulating mutations until the built-up genetic variety results in enough antigenic variety to be able to successfully cause (more) outbreaks and become a dominant variant. Alternatively, animal reservoirs, a limited number of which have been recognized [38,39,76], or chronically infected patients [84,210] may be places where the virus can accumulate mutations.

Neutralizing epitopes were previously reported for the surface-exposed P2 domain, and a role in antigenicity was indicated for this domain in several studies with human as well as animal caliciviruses [36,40,173,194,208,226,247,291,297].

Tan and coworkers reported the conserved RGD motif to be involved in host cell binding [47,286]. Highly variable regions were found in close spatial proximity to the conserved RGD motif (figure 3.5). Tan and coworkers also reported three amino acids, neighboring the RGD motif, which were suggested to have a role in ligand (histo-blood group antigen) binding specificity [286]. One of these surrounding amino acids, designated IV in their paper, is an informative site in our study. Before the 2002 variant, this amino acid was Q376, and it mutated into E376 from the 2002 variant onward. Studies are needed to determine if these mutations lead to changes in host binding specificities.

A second RGD motif (amino acids 339 to 341) was present in the earliest strain sequenced as well as in the 2002 variant. Since it was absent from variants after 2002, it does not seem to confer a great binding advantage. The location of this motif, in spatial proximity to the reported conserved RGD motif on the surface of the molecule and as an insertion compared to the NV genome, does suggest a possible role in ligand binding.

The five amino acids that were informative when comparing all chronologic sets of variants were spread over the capsid. One that stands out is amino acid 340 (E_{1996}->G_{2002}->R_{2004}->G_{2006b}), which in the 2002 variant was also part of an additional RGD motif. The functional implications of these mutations remain to be determined. For our structural analyses of the polymorphisms in the GII.4 variants, we used a computer-derived model of the VP1 protein. After submission of the present study, Cao *et al*. published a paper on the cocrystallization of the P protein of a GII.4 strain norovirus with its receptor [32]. No differences between our computer model and this high-resolution structure were found to be of influence to the data presented here.

One could speculate that the location and positioning of the P2 domain of the capsid might explain part of the great prevalence of mutations in this area. The protruding region is connected to the shell domain with a hinge region, and an additional point of flexibility between P1 and P2 was reported [41,240]. This provides flexibility to slightly adjust to the position of the protruding region on top of the shell domain if needed, thus allowing for more conformational changes and thus for more mutations in this region than in the rest of the protein [41]. This does not explain the epidemiological observations, however, and therefore we do not think it is the complete story.

Similarly, a possible advantage that new lineages of GII.4 might have obtained by the accumulation of mutations is increased stability. However, even though increased stability of the viral particles outside the host would increase the number of infectious particles of the more stable variant available for infection, it does not explain the rapid and complete replacement of previous variants that circulated in the population [175]. Improved binding or a broadened host range also does not provide a tight explanation for the replacement of previously circulating strains.

The most likely advantage for new variants over older ones is that of immune evasion. Noroviruses, particularly strains of GII.4, are highly prevalent in the population. During epidemic seasons, up to 86% of norovirus outbreaks were caused by the predominant genotype, GII.4, followed by a sharp drop in the prevalence of this genotype in the subsequent season [268]. Then, only after the emergence of a genetically distinct new lineage of this genotype, the prevalence of GII.4 strains rose again to cause a new epidemic.

A similar pattern of so-called epochal evolution has been described very elegantly for influenza A virus (H3N2) [142], where periods of phenotypic stasis are separated by the stepwise emergence of phenotypically distinct new variants, as we also see here for noroviruses. During the periods of phenotypic (and antigenetic) stasis, neutral or almost neutral mutations do occur and accumulate if they are beneficial or at least not disadvantageous. For influenza virus, this pattern of evolution and emergence of genetically novel variants is attributed to host population immunity and subsequent antigenic escape by the virus. The striking parallel observed here for norovirus suggests that this pattern of epidemics is driven by (population) immunity as well.

No long-term immunity to norovirus infection has been reported so far. Short-term protective antibodies have been reported, however, and repeated exposure, which is likely to occur with the high prevalence of norovirus, will lengthen the duration of specific protection. Studies with NV in volunteers suggested that immune protection wanes after 6 months

without reexposure [128,227].

In agreement with the hypothesis that immunity to the predominant GII.4 variant built up in the population, Nilsson and coworkers reported on the in vivo evolution of a GII.3 strain infecting a chronically ill immunocompromized patient [210]. They observed the accumulation of amino acid mutations in the capsid protein and suggested that these changes gave rise to a new phenotype, through immune response-driven evolution. Similar to our findings, they found most amino acid mutations in the P2 domain of the capsid. This observation supports the idea that new variants may possibly emerge from chronically infected patients.

The data presented in this paper underpin observations that the elevated numbers of norovirus outbreaks in the winter seasons of 1995-1996, 2002-2003, 2004-2005, and 2006-2007 [27,175,268,322] were mainly, if not solely, due to the emergence of new variants of the GII.4 genotype. The gradual increase in nucleotide mutations in the sequences of norovirus GII.4 strains confirms that genetic drift occurs in the virus. Additionally, the stepwise fixation of numbers of amino acid mutations in the capsid of this predominant genotype, mainly in the surface-exposed P2 domain, is likely to be caused by selective pressure due to population immunity, which resulted in emerging variants which have caused worldwide epidemic rises in outbreak numbers.

Further immunological studies of this variation in the capsid protein are urgently needed to shed light on the mechanisms of immune evasion utilized by the most prevalent genotype of norovirus.

Acknowledgments

This work was supported by the European Commission, DG Research Quality of Life Program, under the 6th Framework (EVENT; SP22-CT-2004-502571).

Chapter 4

Norovirus Illness Is a Global Problem: Emergence and Spread of Norovirus GII.4 Variants, 2001–2007

J. Joukje Siebenga[1,2], Harry Vennema[1], Du-Ping Zheng[3], Jan Vinjé[3], Bonita E. Lee[4], Xiao-Li Pang[4], Eric C. M. Ho[5], Wilina Lim[5], Avinash Choudekar[7], Shobha Broor[7], Tamar Halperin[8], Nassar B. G. Rasool[9], Joanne Hewitt[10], Gail E. Greening[10], Miao Jin[6], Zhao-Jun Duan[6], Yalda Lucero[11], Miguel O'Ryan[11], Marina Hoehne[12], Eckart Schreier[12], Rodney M. Ratcliff[13], Peter A. White[14], Nobuhiro Iritani[15], Gábor Reuter[16] and Marion Koopmans[1,2]

1 National Institute for Public Health and the Environment, Bilthoven, and
2 Erasmus Medical Centre, Rotterdam, the Netherlands
3 Centers for Disease Control and Prevention, Atlanta, Georgia, United States
4 Provincial Public Health Laboratory (Microbiology), Edmonton, Alberta, Canada
5 Centre for Health Protection, Hong Kong
6 Department of Viral Diarrhea, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
7 All India Institute of Medical Sciences, New Delhi, India
8 Central Laboratory, Israel Defense Force Medical Corps, Israel
9 University of Malaya, Kuala Lumpur, Malaysia
10 Institute of Environmental Science and Research, Porirua, New Zealand
11 Institute of Biomedical Sciences, University of Chile, Santiago, Chile
12 Robert Koch Institute, Berlin, Germany
13 Institute of Medical and Veterinary Science and School of Molecular and Biomedical Science, University of Adelaide, Adelaide Australia
14 University of New South Wales, Sydney, Australia
15 Osaka City Institute of Public Health and Environmental Sciences, Osaka, Japan
16 ÁNTSZ Regional Institute of State Public Health Service, Pécs, Hungary

Published in The Journal of Infectious Diseases
• http://www.journals.uchicago.edu/toc/jid/current • 2009; 200 (5):802–812 •
DOI: 10.1086/605127

Abstract

Noroviruses are the most common cause of viral gastroenteritis. Their high incidence and importance in health care facilities result in a great impact on public health. Studies from around the world describing increasing prevalence have been difficult to compare because of differing nomenclatures for variants of the dominant genotype, GII.4. We studied the global patterns of GII.4 epidemiology in relation to its genetic diversity.

Data from norovirus outbreaks with dates of onset from January 2001 through March 2007 were collected from 15 institutions on 5 continents. Partial genome sequences (*n*=775) were collected, allowing phylogenetic comparison of data from different countries.

The 15 institutions reported 3098 GII.4 outbreaks, 62% of all reported norovirus outbreaks. Eight GII.4 variants were identified. Four had a global distribution - the 1996, 2002, 2004, and 2006b variants. The 2003Asia and 2006a variants caused epidemics, but they were geographically limited. Finally, the 2001Japan and 2001Henry variants were found across the world but at low frequencies.

Norovirus epidemics resulted from the global spread of GII.4 strains that evolved under the influence of population immunity. Lineages show notable (and currently unexplained) differences in geographic prevalence. Establishing a global norovirus network by which data on strains with the potential to cause pandemics can be rapidly exchanged may lead to improved prevention and intervention strategies.

Introduction

Noroviruses are the leading cause of acute viral gastroenteritis worldwide. People of all ages are affected, but outbreaks are most often reported in health care settings (such as nursing homes and hospitals), where infections in high-risk groups (such as elderly and immunocompromized people) can have a serious impact by causing prolonged morbidity and mortality [220,266]. Outbreaks are difficult to control and may lead to considerable economic costs resulting from closure of wards, increased length of hospitalization, hiring of extra personnel, and use of extra supplies [105,129].

The incidence of norovirus illness is high. A population study of gastroenteritis in the Netherlands estimated that 500,000 norovirus cases occurred among a population of 15.7 million in 1999 (a nonepidemic year) [53]. A study from England (1993–1996) reported that noroviruses were the most commonly identified pathogen in cases of intestinal infectious disease [329]. On the basis of these and historical data, the prevalence in the United States was estimated to be 23 million cases annually [199]. In developing countries, where diarrhea is a leading cause of death in young children [152], relatively little is known about the etiological role played by noroviruses, but estimates indicating that >1.1 million hospitalizations and almost 220,000 deaths occur among children <5 years old each year in such countries have been published recently [229].

Noroviruses are extremely contagious, owing to a very low infectious dose (estimated median infectious dose, ~18 viral particles) [293] and to levels of shedding $>1*10^{10}$ RNA copies per gram of stool [37,302]. Additionally, although patients usually recover from symptoms within 2 or 3 days, shedding may last several weeks [302]. Transmission occurs through the fecal-oral route, either by direct contact with infected individuals or via contaminated surfaces, food, or water. No or limited long-term immunity results from infection [128,227], so one person may be repeatedly infected. Nevertheless, short-term immunity combined with the high prevalence of noroviruses amounts to herd immunity [170,269].

Noroviruses belong to a highly genetically and antigenically diverse genus of the family Caliciviridae. They segregate into 5 genogroups. In genogroup I, 8 genotypes are currently recognized; in genogroup II, 19 are recognized [93]. Viruses of the GII.4 genotype have been predominant during the past decade in the United States, Europe, and Oceania, causing 70%–80% of all norovirus outbreaks (particularly in health care settings)[157]. In 2002, the number of reported norovirus outbreaks increased sharply in many countries, followed by 4 epidemic winters in 2002–2003, 2004–2005, 2006–2007, and 2007–2008 in the Northern Hemisphere. During these seasons, incidence rates were likely much higher than reported in

the previously mentioned population studies, which were all conducted before 2002. Genetic analyses showed that these epidemics coincided with the emergence of novel GII.4 variants except for the 2007–2008 epidemic, which was a continuation of the 2006–2007 epidemic [2 7,212,219,250,251,268,269,301,303]. The global molecular epidemiology of GII.4 lineages has not been systematically evaluated; reports describing the molecular epidemiology of the GII.4 variants have focused on local or regional studies. Furthermore, the lack of a unified nomenclature for GII.4 variants has hampered comparison of the available data. Thus, the extent to which reported epidemics were truly global remains unclear.

The growing awareness and knowledge of both the scale and impact of norovirus epidemiology raised the question as to how patterns suggested by molecular data analyses matched observations from public health surveillance across the world. We also wanted to create a common data set for norovirus strains to function as a starting point for the newly established global norovirus collaboration network, NoroNet. This global network aims at limiting the impact and scale of future norovirus epidemics, for which monitoring of circulating strains is essential. Here, we present a global overview of the molecular epidemiology of GII.4 noroviruses for the period from January 2001 through March 2007. During this period, 4 of the described GII.4 variants had a truly global distribution, whereas 4 had a different distribution or prevalence. Our results demonstrate that emergent GII.4 norovirus strains spread throughout the world very rapidly and, therefore, that epidemiology needs to be assessed on a global scale in order to further the development of the basic knowledge that will underpin future research and prevention and control measures.

Methods

Epidemiologic data

Data obtained from January 2001 through March 2007 were included in the present study. Institutions on 5 continents were contacted for participation. Selected institutions were included because they had structured norovirus surveillance programs that included the molecular characterization of outbreak strains (n=1) or because they provided data on geographic areas that could not otherwise be evaluated (n=6).

The aggregated data set contained monthly reports describing all reported norovirus outbreaks attributable to each GII.4 variant. Norovirus outbreaks were identified according to criteria used by each institution. Epidemiologic data describing the population covered, the number of outbreaks reported annually, the number of confirmed norovirus outbreaks, and the percentage of GII.4 outbreaks were collected.

Sequence data

Each participating institution provided sequences from representative strains detected during the beginning, middle, and end of the circulation periods of each variant for genotyping; sequences obtained were from the preferred genomic regions of the strains (regions A, C, D, and E) (figure 4.1) [321].

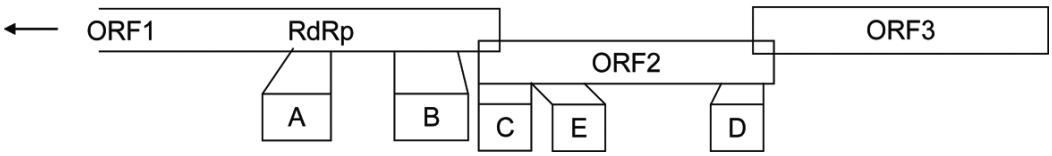


Figure 4.1. Schematic representation of the locations of the genomic regions of norovirus used for genotyping. Adapted from Vinjé et al. [27]. Region B was not used for the analyses in the present study because it does not enable discrimination between GII.4 variants. ORF, Open Reading Frame; RdRp, RNA-Dependent RNA Polymerase.

These data enabled comparison of the nomenclature used by the participating institutions and in the literature (table 4.1), standardization of assignment of variant names, a check of the quality of the submitted data, and phylogenetic comparison of sequences detected in different geographic regions of the world. As the basis for variant assignment, we used the recently published amino acid sequences of the complete VP1 [269]. This method defines variants on the basis of phylogenetic clustering combined with epidemiologic patterns. Variant names include the first reported year of detection supplemented by a geographic region or suffix when necessary. Table 4.1 summarizes these names in addition to names used in other publications.

Data analyses

Variant assignment was done by the participating institutions, using either their own referencing systems, a tool offered by the Netherlands National Institute for Public Health and the Environment (RIVM) for regions A and C (Norovirus Quicktyping; available at: http://www.rivm.nl/bnwww), or a set of reference sequences compiled for this purpose that spanned the continuous genomic region covering regions A, C, E, and D.

Epidemiologic and sequence data were stored and analyzed in Microsoft Excel and BioNumerics software (version 4.61; Applied Maths). Additional reference sequences were obtained from GenBank. At RIVM, phylogenetic analysis of all sequences was performed using PhyML software (version 2.4.4) [100].

Table 4.1. Nomenclature for GII.4 variants.

GenBank ref. strain	Epidemic season	Name used in present article	First cit./ full description	Other names
X86557	Before 1995	…	…	Lordsdale, GII/4 b [232]
AF145896	Before 1995	…	…	Camberwell, GII.4-1987 [170], GII/4-a [219], GII/4 c [232]
X76716	Before 1995	…	…	Bristol, GII/4-a [219], GII/4 b [232]
AJ004864	1995–1996	1996	[184]	Grimsby, Burwash Landing, GII.4-1997 [170], GII/4-b[219], GII/4-g [232]
AB294779	NA	2001Japan	[219]	GII/4-c [219], GII/4 a [232]
EU310927	NA	2001Henry	[226]	Houston
AY485642	2002–2003	2002	[175]	Farmington Hills, GII/4-d [219], GII/4 e [232]
AB220922	NA	2003Asia	[220]	Sakai, GII.4-2005 [170]
AY883096	2004–2005	2004	[27]	Hunter, GII/4 f [232]
EF126963, EF126964	2006–2007	2006a	[269]	Laurens, V4 [83][a]
EF126965, EF126966	2006–2007	2006b	[269]	Minerva, Den Haag [170], GII/4-e [219], GII/4-f e [219], Kobe034, V6 [83] [a]

NOTE. Reference strains were the first complete capsid sequences of each variant submitted to the public databases. NA, not applicable. [a] Not all divisions made by Gallimore *et al.*, which were based on motifs, fit the phylogenetic grouping.

Results

Participants

Table 4.2 lists the participants of the present study and outlines their surveillance methods. Overall, most reported outbreaks occurred in health care settings. In the countries of most participating institutions, it was mandatory to report outbreaks caused by any pathogen in health care settings. Ten institutions had an ongoing passive norovirus surveillance program (table 4.2, top section). The Provincial Public Health Laboratory, Canada, had limited surveillance data for 2001; therefore, this year was not included. Five additional institutions, plus a targeted population study from the University of Chile, that did not have structured norovirus surveillance program provided target-study data and sequences with detection dates (table 4.2, bottom section). The University of Malaya, Malaysia, had different target studies mainly involving hospitalized children. The Israel Defense Force Medical Corps Central Laboratory provided data on norovirus outbreaks in military units in Israel [102]. The Institute of Biomedical Sciences, University of Chile, submitted data from intermittent population surveillance and targeted population studies but usually without genotyping data [319]. The All India Institute of Medical Sciences provided data from a study comprising 4 different study groups [246].

Table 4.2. Participating institutions and brief description of their surveillance setup.

Category, Institution	Geographic Area	Coverage[a]	Typing Region[b]	S/O[c]	Publication(s)	Note
Participants with population level surveillance						
National Institute for Public Health and the Environment	Netherlands	Population, 100% Input mainly HC**	A	O	[268,268,269]	
Centers for Disease Control and Prevention	United States	Population, 100% Input mainly HC	C, D, E	O	[17,17,17,74]	
Provincial Public Health Laboratory	Alberta, Canada	Population, 100% Input mainly HC	E	O		No data for 2001
Osaka City Institute of Public Health and Environmental Sciences	Osaka Prefecture, Japan	Population, 100%	C	O	[264]	
Center for Health Protection, Hong Kong	Hong Kong, China	Population, 100%	A, C since 2005	O+S	[115]	
Environmental Science & Research	New Zealand	Population, 100% Input mainly HC	B until 2004, D	O		
Universidad de Chile	Santiago, Chile	100%	A, C	O	[319]	No data for 2004-2005, no genotyping for 2006
Regional Institute of State Public Health Service	Hungary	Population, 100% Input mainly HC	A	O	[250,251,251]	
Robert Koch-Institute	Germany	Population, 100% Input mainly HC	A	O	[117,260,260]	
Institute of Medical and Veterinary Science	Adelaide, Australia	Population, 100% Input mainly HC	C, D	O		
Participants with targeted population studies		Target population				
University of New South Wales	New South Wales, Australia	All	C	O	[27,27,301,301,303]	
Department of Viral Diarrhea National Institute	Various provinces and cities in China	Children <5 years with acute gastroenteritis	A, C since 2006	S		
All India Institute of Medical Sciences	India	Various*	A	S	[246]	
Universidad de Chile	Chile	Islands (2003) and cruise ship (2005)	A	S		
Central Laboratory	Israel	Military	D	O	[102]	
University of Malaya	Malaysia	Hospitalized children (0-6 years old), 1 adult	E, D	S	Rasool et al. unpublished	

* Including children's home, hospital, health centre. [a] HC indicates health care settings, including hospitals, nursing homes, and long-term-care facilities.
[b] Genomic region used for typing detected norovirus strains (see figure 1). [c] Indicates whether the institution reports sporadic cases (S), outbreaks (O), or both.

Epidemiology of norovirus outbreaks, by geographic area

In total, 4988 norovirus outbreaks were reported, of which 3089 (62%) were confirmed to be GII.4 outbreaks (figure 4.2). Winter seasonality was seen in almost all geographic areas analyzed and was largely associated with elevated numbers of GII.4 outbreaks.

Data were comparable among the European countries, although the 2002–2003 and 2004–2005 seasons in Hungary were not as pronounced as those in Germany and the Netherlands. Germany and the Netherlands reported an off-seasonal peak in April and May 2002 that preceded the epidemic winter of 2002–2003 [175].

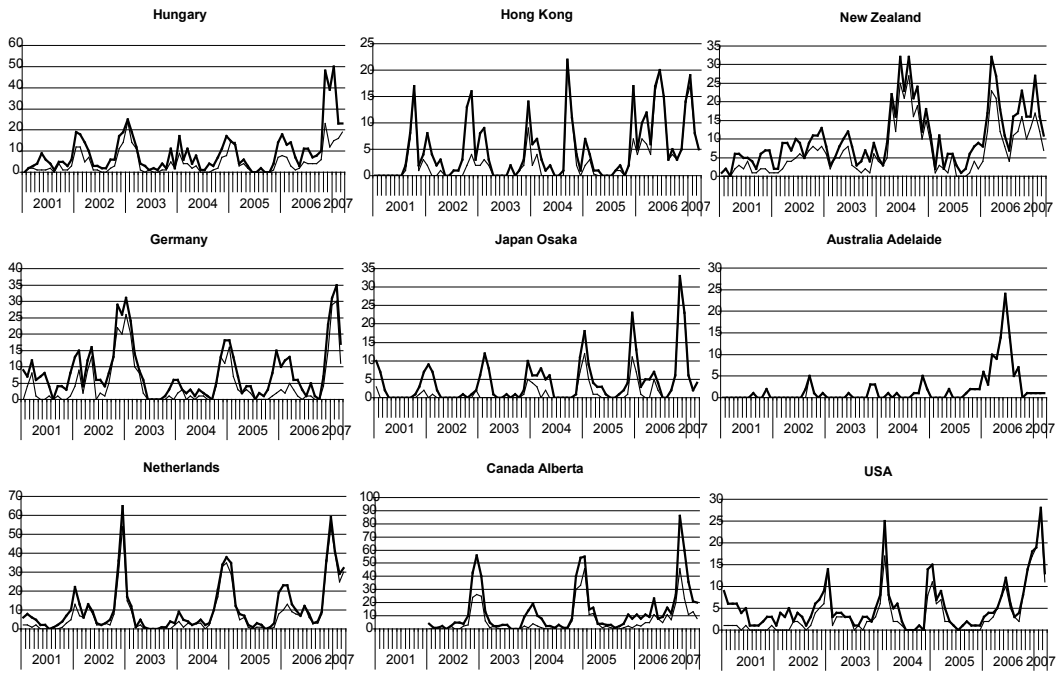


Figure 4.2. Monthly numbers of reported norovirus outbreaks. The thick line shows all norovirus outbreaks and the thin line the GII.4 outbreaks, by geographic area. Note the different scales for the Y-axes. The total numbers of reported norovirus outbreaks (left of virgule) and the total numbers of GII.4 outbreaks (right of virgule) for each area are given above the graphs.

Similar summer peaks were seen more widely in 2006, caused by 2 emerging variants: 2006a and 2006b. These peaks were observed in the Netherlands; Germany; Hungary [251]; Australia; Alberta, Canada; Japan; and Hong Kong. In Japan and Hong Kong, the peaks were caused by the 2006b variant only. In Hong Kong, winter peaks started earlier (September–October) than in the other analyzed regions of the Northern Hemisphere (November–December), and the year 2006 in Hong Kong was marked by many outbreaks, with those occurring during the 2005–2006 winter being only partly related to GII.4 outbreaks. In

Japan, the number of reported norovirus outbreaks increased gradually, and, especially during the 2006–2007 winter, the proportion of GII.4 viruses was very high (almost 100%). Unlike most other countries, the United States reported many GII.4 outbreaks during the 2003–2004 winter.

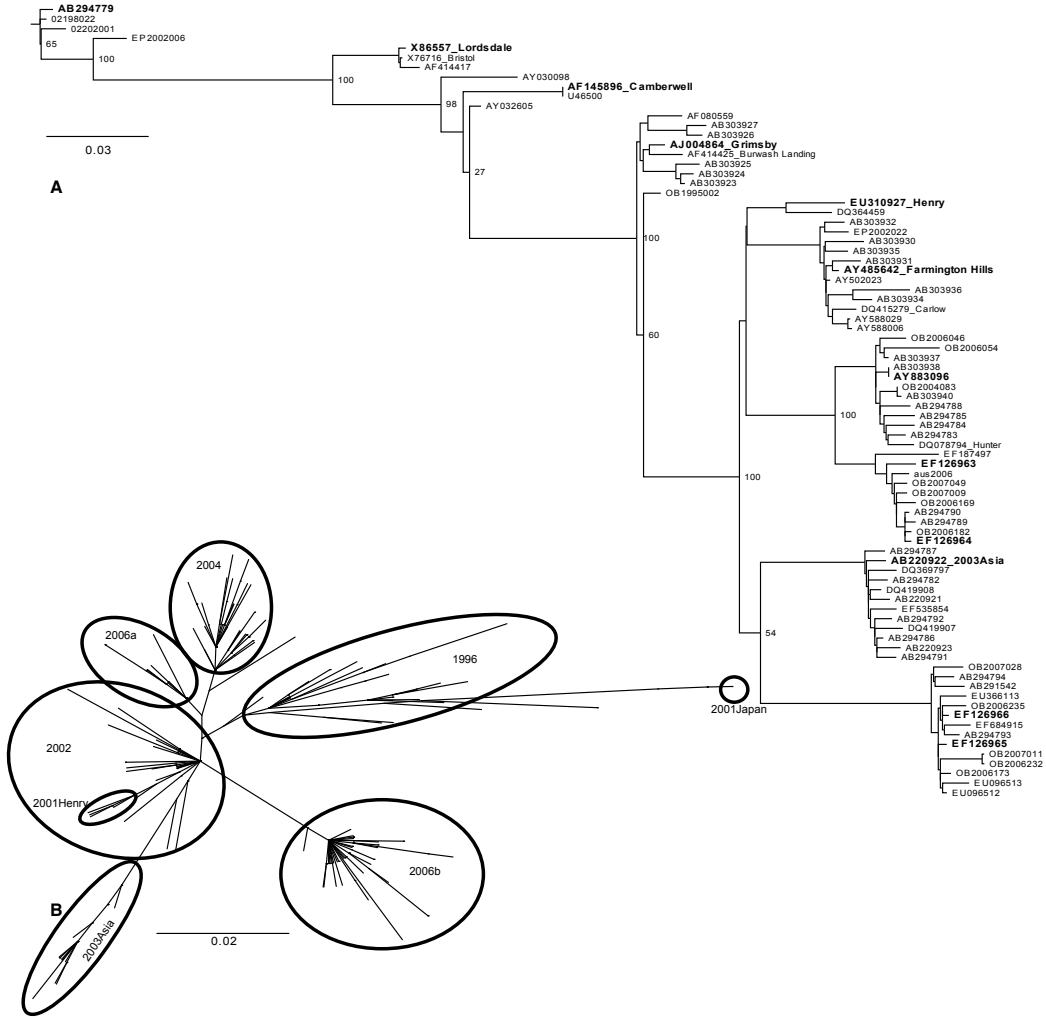


Figure 4.3. Phylogenetic analyses of the sequences in the database. A) Phylogenetic tree of complete capsid sequences, constructed by means of PhyML software (49 sequences, 1623 nt, SYM+I+Γ4 [symmetric, with allowance for invariant sites and gamma substitution rate variation], 1000 bootstraps). Reference strains (first complete capsid sequences submitted to the public databases) are shown in boldface type. The model was selected using MrAIC software (J. A. A. Nylander, 2004. MrAIC.pl. Program distributed by author. Evolutionary Biology Centre, Uppsala University). B) Phylogenetic tree of all available region C sequences, constructed by means of PhyML software (301 sequences, 266 nt, SYM+I+Γ4). Strains detected in Australia, China, Germany, Ghana, Hong Kong, Hungary, Japan, the Netherlands, New Zealand, and the United States were included in this tree.

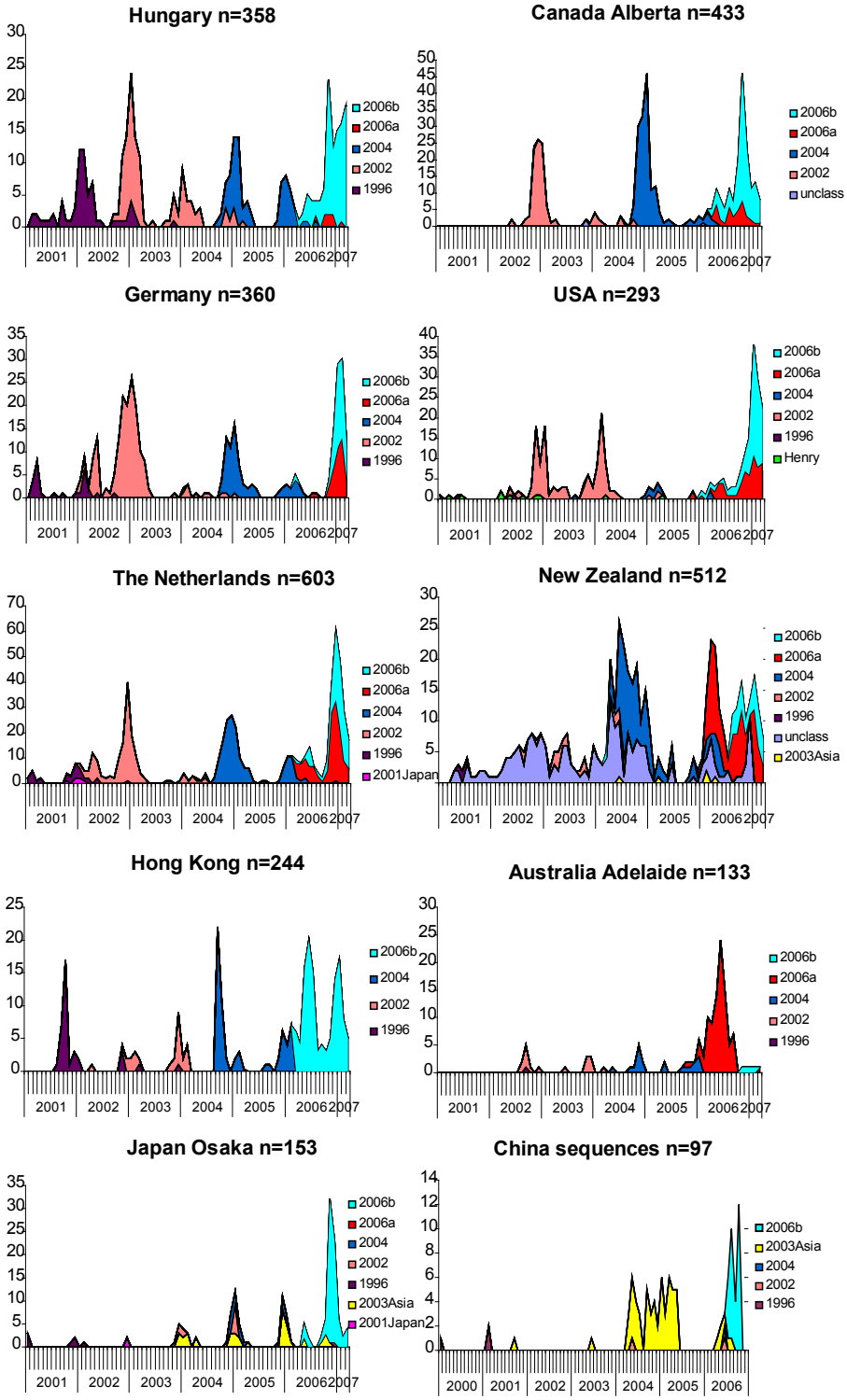

Hungary n=358
2006b
2006a
2004
2002
1996
Canada Alberta n=433
2006b
2006a
2004
2002
unclass
Germany n=360
2006b
2006a
2004
2002
1996
USA n=293
2006b
2006a
2004
2002
1996
Henry
The Netherlands n=603
2006b
2006a
2004
2002
1996
2001Japan
New Zealand n=512
2006b
2006a
2004
2002
1996
unclass
2003Asia
Hong Kong n=244
2006b
2004
2002
1996
Australia Adelaide n=133
2006b
2006a
2004
2002
1996
Japan Osaka n=153
2006b
2006a
2004
2002
1996
2003Asia
2001Japan
China sequences n=97
2006b
2003Asia
2004
2002
1996
2001 2002 2003 2004 2005 2006 2007
2000 2001 2002 2003 2004 2005 2006

Figure 4.4. Prevalence of GII.4 variants, by geographic area. All data are from areas where ongoing population surveillance is being conducted; data from China are from sporadic diarrhea cases occurring in children <5 years old (derived from sequence data submitted for this study). The number of reported outbreaks is indicated for each area. Note the different scales for the Y-axes. The large number of unclassified (UC) strains reported by New Zealand were sequenced in region B only. In Hong Kong, 2003Asia strains were detected only in sporadic cases (especially between January 2004 and January 2005) but not in outbreaks.

The Southern Hemisphere countries Australia and New Zealand had different pictures. Interestingly, winter seasonality was absent from New Zealand. In the Adelaide area of Australia, the frequency of norovirus outbreaks increased 5–10-fold during the 2006 epidemic, compared with that during the previous outbreak years of 2002 and 2004. This increase did not coincide with alterations in the surveillance scheme.

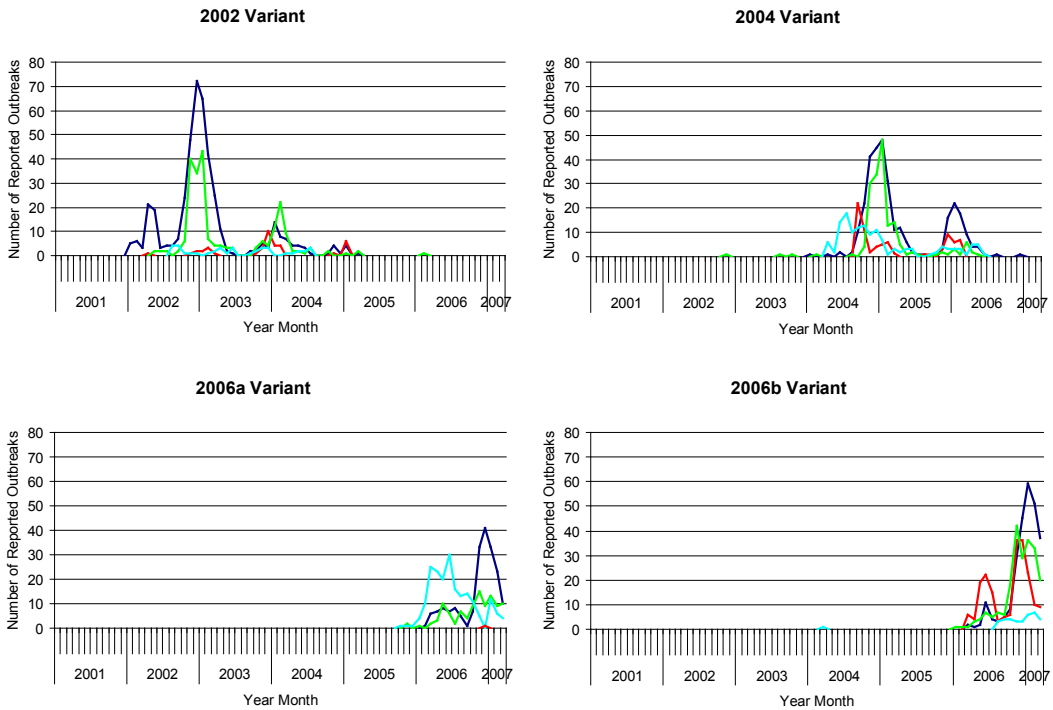


Figure 4.5. Timeline of reported outbreaks accumulated on each continent, by GII.4 variant. Dark blue indicates Europe (the Netherlands, Hungary, and Germany); red, Asia (Hong Kong and Japan); green, North America (United States and Canada); and light blue, Oceania (New Zealand and Australia [the Adelaide area]). Note that absolute numbers are depicted and that they vary by continent because of different sizes of reporting areas and populations.

Sequence analysis and phylogeny

A total of 781 sequences were analyzed, including sequences of 19 strains from the public databases of Brazil (n=12), Malawi (n=1), and Ghana (n=6) (to represent geographic areas

that were not covered by surveillance) and 11 reference sequences. Twenty-four of the submitted sequences were too short for reliable analysis and were excluded from further analysis.

Full capsid sequences and partial sequences from regions A, C, E, and D (figure 4.1) were phylogenetically analyzed (figures 4.3A and 4.3B). Discriminatory power was not sufficient to distinguish different variants in region B; therefore, sequences from region B were not collected. Partial genome sequences were assigned a variant type on the basis of similarity with complete capsid sequences and associated partial polymerase sequences. Clustering into variants was consistent, as illustrated for region C sequences in figure 4.3B, except for the 2001Henry variant. Although the capsid sequence for this variant was clearly distinct, partial sequences clustered with other variants. As a result, discrimination of this strain may not have been accurate, and assessing its prevalence was difficult. Geographic trends within variant clusters were not identified. Analyses of regions A, E, and D resulted in identical clustering (data not shown).

Eight GII.4 variants were identified (table 4.1 and figure 4.3A); the 5 main variants were 1996, 2002, 2004, 2006a, and 2006b, and the 3 minor variants were 2003Asia, 2001Japan, and 2001Henry. The minor variants were detected at multiple geographic locations but at low frequencies. The variant 2003Asia is a recombinant, with an open reading frame 1 sequence belonging to GII.12 (Wortley-like) and open reading frame 2 and 3 sequences belonging to GII.4. This strain caused many outbreaks in Asia; however, in Hong Kong it was detected only among sporadic norovirus cases, not in outbreaks. In Hong Kong, 993 sporadic norovirus cases were analyzed, of which 58 (6%) were 2003Asia strains, detected from late 2002 through 2005. The peak prevalence of 2003Asia in sporadic cases in Hong Kong was observed between January 2004 and January 2005 (15% [33/223] of cases; data not shown). The 2001Japan sublineage showed the most genetic resemblance to the Bristol and Lordsdale reference strains. The 2001Henry variant was identified in the United States and China (DQ364459), but prevalence was low.

All variants from the 2002 variant onward (i.e., the 2001Henry, 2003Asia, 2002, 2004, 2006a, and 2006b variants) had a 1-aa insertion (at position 393) in the P2 domain of the capsid.

Quality assurance

The accuracy of variant assignment by participants was independently verified by phylogenetic analysis of submitted sequences. Of the 775 submitted sequences, 767 (99%) had been assigned to the correct variant. Furthermore, sequences of 2 separate genomic regions had been submitted for a subset of 81 outbreaks, and, with the exception of 5 (6%), they consistently belonged to the same variant in both regions. The 5 remaining strains may have been recombinants, or 2 different viruses may have been present in 1 outbreak.

Epidemiology of GII.4 variants

Epidemiologic analysis of the GII.4 variants was done using the 3089 submissions for which collection dates had been provided (figure 4.4). In 2001–2002, the 1996 variant caused significant numbers of outbreaks in Europe and Hong Kong but was rare in other geographic areas. The 2002, 2004, and 2006b variants caused epidemics in all areas with population surveillance, but there was a marked difference in the magnitude of the seasonal peaks. The 2002 variant heavily affected Europe, the United States, and Canada [175] but was observed less frequently in Asia and Oceania, despite significant numbers of reported outbreaks. During the same period, high numbers of GII.4 strains were reported in New Zealand, but they could not be assigned to a GII.4 variant because they were sequenced in region B only. Variants 2004 and 2006b were truly global and were prevalent in all analyzed geographic areas; 2006a is closely related phylogenetically to the 2004 variant, but both the complete capsid sequence and its distinct epidemiology (with clearly discernible epidemic peaks in 2004–2005 and 2006–2007) justify its classification as a separate variant [269]. Interestingly, 2006a was rarely reported by participants in Asia - 1 outbreak was reported in Japan, and no sequences belonging to the 2006a variant were submitted by other participants in Asia (China, Japan, Hong Kong, India, and Malaysia [not all data shown]).

The minor variants were not confined to a single geographic area. For example, the 2001Japan variant has been detected in Japan, the Netherlands, Malawi (GenBank), and Chile, and the 2003Asia variant has been detected in China, Japan [203,220], Hong Kong, New Zealand, and the United Kingdom [83] - although rarely in the latter two and, interestingly, only in sporadic cases but not in outbreaks in Hong Kong.

Pilot studies in geographic areas with no structured outbreak surveillance data

Data from participants with targeted population studies - including participants in Australia, China, India, Chile, Israel, and Malaysia (table 4.2, bottom section), as well as GenBank sequences detected in Brazil, Malawi, and Ghana - were included to establish a more comprehensive overview of the global GII.4 variant distribution. The data thus obtained fit the overall picture of common GII.4 variants except for the data from Chile and China. Seven of 10 Chilean strains belonged to the 2001Japan variant. The sequences from China (n=97) (figure 4.4) represented studies of children <5 years old and showed no evidence of epidemics caused by the 2002, 2004, or 2006a variants. However, a high number of 2003Asia strains were identified, especially in 2004 and 2005, and a high number of 2006b strains were identified in 2006.

Timeline of GII.4 variant emergence and spread over different continents

Figure 4.5 shows accumulated outbreaks per variant per continent. For the 2002 variant, the off-seasonal peak in Europe was seen first, followed by the epidemic winter in Oceania

(although the latter is difficult to observe in the graphs because of the low numbers relative to those for the other continents). Three variants emerged first in Oceania or Asia; the 2004 and 2006a variants clearly first emerged in Oceania, and the 2006b lineage caused high numbers of outbreaks first in Asia (Hong Kong), followed by an off-seasonal peak in Europe and then Oceania.

Discussion

This article represents the first effort to describe the molecular epidemiology of GII.4 noroviruses on a global scale. We have demonstrated that GII.4 strains have been dominant all over the world during the past 7 years and were responsible for an increased number of reported norovirus outbreaks globally. The impact of epidemic seasons was reflected in the outbreak-based surveillance, which focused on health care settings, although outbreaks also occurred in other places where close human contact occurs (such as military establishments, cruise ships, a hurricane Katrina refugee complex, and schools) [102,316,339].

Although the emergence of the 1996 variant falls outside the time range of this study, it was reported on all continents and likely had global coverage [212]. From the 1996 variant onward 8 distinct variants were identified, of which the 2002, 2004, and 2006b variants caused global epidemics. They displaced their predecessors rapidly and completely, although the 2006b variant initially cocirculated with the 2006a variant before continuing to cause the majority of outbreaks during the winter of 2007–2008, whereas the 2006a variant decreased in incidence but remained present (data not shown; personal communication with participating institutions). The 2006a and 2003Asia variants did cause epidemics but not in all analyzed geographic areas, whereas outbreaks of 2001Japan and 2001Henry occurred sporadically although at multiple geographic locations. Additional minor variants may have circulated but not been identified because of low frequencies. We have not proposed a nomenclature system for future variant naming here because a proposal for norovirus nomenclature is currently being prepared.

Successive GII.4 variants differ antigenically and in their patterns of host cell binding, as shown by detailed molecular characterization of systematically collected outbreak data, the elucidated structure of the capsid protein, and binding experiments [170,269]. The data presented here suggest that norovirus GII.4 strains evolved and spread in a manner similar to that of influenza A virus, with a rapid global spread of emerging variants. The limited information available from Africa and South America fits this general pattern, although the available data are very patchy. Clearly, more (and more-thorough) surveillance in these areas would improve our understanding of norovirus epidemiology. Unlike what could

be concluded from regional surveillance data alone, the present study of epidemiologic and virologic data from norovirus outbreaks on 5 continents has shown that the dominant norovirus genotype, GII.4, has a global impact. Additional studies, particularly in developing countries, are needed to further elucidate the burden of disease due to norovirus in various communities. Although it remains possible that the increased number of reported norovirus outbreaks resulted from improved diagnostics and reporting, this does not diminish our finding that the spread of the GII.4 variants was rapid and global and was unseen before the 1996 variant. Given that the incidence of norovirus infection is high among young children and that diarrhea is one of the leading causes of death in children <5 years old, it is critical that studies be conducted to assess the role played by noroviruses in childhood diarrhea [152,229].

That some variants became epidemic only within a limited geographic region leads to interesting speculation as to what determines the success of different lineages (or the apparent lack thereof). The 2003Asia variant was observed in Asia but was rarely seen on other continents; conversely, the 2006a variant was rare in Asia. This was seen in our aggregated data and is substantiated in the literature, including in a recent article in which the Norovirus Surveillance Group of Japan reported a large cluster of 2003Asia and very low numbers of 2006a strains [203]. One possible explanation is that the 2006a variant shares neutralizing epitopes with the 2003Asia variant and, therefore, encountered an immune population in Asia but not in other regions. However, although the capsid of the 2003Asia variant does share 12 informative sites with that of the 2006a variant (data not shown), the same amino acid motifs are also present in the 2004 and 2006b variants, which were geographically unrestricted. Thus, prior population immunity appears to be an unlikely explanation.

A second possibility is that variants may differ in their affinity for host ligands involved in virus attachment before entry into host cells. Noroviruses are known to exhibit different strain-specific host-ligand-binding patterns [288]. Recently, shifts in the binding patterns of ensuing GII.4 variants were reported [170]. The lack of a receptor for the 2006a variant in Asian populations might explain why epidemics of this variant were not noted in Asia. In a similar manner, the 2003Asia variant was rarely detected outside Asia. Therefore, we compared the predicted receptor-binding interface published by Cao *et al.* [32] in sequences of different variants, especially the 2003Asia and 2006a variants. No variation between the 2003Asia and 2006a variants could be identified at the loci involved in ligand binding other than the previously reported aa 393, which is highly polymorphic among norovirus strains ([32] and X. Jiang, personal communication). Similar results indicating that more-detailed molecular characterization is needed to verify the hypotheses on binding have

been reported by Lindesmith *et al.* [170].

A third possible explanation for the differing success of GII.4 variants is the occurrence of seeding events (i.e., efficient initial large-scale introduction and subsequent transmission of viruses). Diffuse, large-scale, food-related outbreaks have been described, such as the seeding of a specific strain across Europe during the winter of 2000–2001 through contaminated shellfish [146]. Such outbreaks are extremely difficult to detect given the current state of surveillance for noroviruses but are estimated to be common [318].

We looked for a pattern of emergence that could be used in a global early-detection system for new epidemic strains. The date of the first reported detection is too strongly dependent on the intensity of surveillance and chance to be a reliable indicator. However, when the first epidemic peak observed in each participant's region was considered, 3 of 4 high-impact variants were first reported in Asia and Oceania. This may indicate that these antigenic variants arose in Asia and then spread throughout the world, as has been suggested for influenza A virus [256]. Additionally, the data imply that international communication of surveillance results could help prepare health care settings for "hot seasons." Oceania appeared to be ahead of countries in the Northern Hemisphere for the epidemic peaks of at least 2 variants, possibly resulting from its inverted seasons compared with the Northern Hemisphere. Epidemic peaks caused by emerging variants were preceded by off-seasonal peaks during spring and summer in the Northern Hemisphere [175,251,316]. As the number of outbreaks in Northern Hemisphere countries decreased during summer before becoming fully epidemic the following winter, winter started in Oceania, creating opportunities for the emerging variant there.

Our study demonstrates that norovirus epidemics result from the rapid and global spread of successful GII.4 strains. These strains have been shown to evolve under the pressure of population immunity. Notable differences in the prevalence of certain lineages have been observed that cannot be explained by our current knowledge of noroviruses. This highlights that joint surveillance will increase our understanding of norovirus infection and epidemiology. The participants of this retrospective study will continue to collaborate and have initiated a global norovirus surveillance network, NoroNet, that has already expanded to include more institutions. Through this global network, more-timely identification of unusual activity and new variants will improve efforts to limit the scale of epidemics.

Acknowledgments

We thank Dr. Jutta Preiksaitits (Provincial Public Health Laboratory, Canada) for her support and Sallene Wong (Provincial Public Health Laboratory, Canada), J. K. R. Thomas, M. A. C. Hamat, H. Kamyon, and M. R. Idid (University of Malaya, Malaysia) for practical assistance. J.J.S. kindly thanks all coauthors for generously sharing their data and answering queries.

Chapter 5

Phylodynamic Reconstruction Reveals Norovirus GII.4 Epidemic Expansions and their Molecular Determinants

J. Joukje Siebenga[1, 2*], Philippe Lemey[3*], Sergei L. Kosakovsky Pond[4], Andrew Rambaut[5], Harry Vennema[1], Marion Koopmans[1,2]

* Equal contributions

Affiliations

1. National Institute for Public Health and the Environment, RIVM, Bilthoven, The Netherlands

2. Erasmus Medical Centre, Rotterdam, The Netherlands

3. Rega Institute, Katholieke Universiteit Leuven, Leuven, Belgium

4. Antiviral Research Center, University of California San Diego, San Diego, California, USA

5. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, and Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.

Published in PLoS Pathogens • www.plospathogens.org • 2010;6(5):e1000884 • DOI:10.1371/journal.ppat.1000884

Abstract

Noroviruses are the most common cause of viral gastroenteritis. An increase in the number of globally reported norovirus outbreaks was seen the past decade, especially for outbreaks caused by successive genogroup II genotype 4 (GII.4) variants. Whether this observed increase was due to an upswing in the number of infections, or to a surveillance artifact caused by heightened awareness and concomitant improved reporting, remained unclear. Therefore, we set out to study the population structure and changes thereof of GII.4 strains detected through systematic outbreak surveillance since the early 1990s.

We collected 1383 partial polymerase and 194 full capsid GII.4 sequences. A Bayesian MCMC coalescent analysis revealed an increase in the number of GII.4 infections during the last decade. The GII.4 strains included in our analyses evolved at a rate of $4.3 – 9.0 \times 10^{-3}$ mutations per site per year, and share a most recent common ancestor in the early 1980s. Determinants of adaptation in the capsid protein were studied using different maximum likelihood approaches, to identify sites subject to diversifying or directional selection and sites that co-evolved. While a number of the computationally determined adaptively evolving sites were on the surface of the capsid and possible subject to immune selection, we also detected sites that were subject to constrained or compensatory evolution due to secondary RNA structures, relevant in virus-replication. We highlight codons that may prove useful in identifying emerging novel variants, and, using these, indicate that the novel 2008 variant is more likely to cause a future epidemic than the 2007 variant.

While norovirus infections are generally mild and self-limiting, more severe outcomes of infection frequently occur in elderly and immunocompromized people, and no treatment is available. The observed pattern of continually emerging novel variants of GII.4, causing elevated numbers of infections, is therefore a cause for concern.

Introduction

Noroviruses are the most common cause of acute viral gastroenteritis [11,229], with the numbers of reported outbreaks peaking characteristically between November and March in the northern hemisphere [228]. Illness is usually self-limiting and symptoms, comprising acute onset vomiting and watery diarrhea, subside within one to three days [253]. The relevance of studying norovirus lies in their high prevalence in the population [53], and in the more severe and prolonged illness that is seen among elderly and immunosuppressed patients [110,220,266]. Noroviruses are highly infectious, due to the combination of an extremely low infectious dose (an estimated ID50 of less than 20 viral particles [293]), very high levels of shedding (around 10^8 but up to $>10^{10}$ RNA copies per gram of stool) and prolonged shedding after clinical recovery [165,302]. Norovirus outbreaks, which may affect hundreds of people and are notoriously difficult to control, are primarily associated with places where people are in close contact, for example hospitals and long-term care facilities.

Noroviruses are a genetically diverse group of positive sense single-stranded RNA viruses from the *Caliciviridae* family. Their 7.5 kb genome includes three open reading frames (ORFs). The first ORF encodes a polyprotein that is post-translationally processed to form the non-structural proteins, the second and third ORFs encode the major and minor structural proteins; VP1 or the capsid protein and VP2. The viral capsid is formed by 180 copies of the major capsid protein, and governs antigenicity, host-specificity and environmental stability.

Noroviruses are classified into five distinct genogroups, which are further subdivided into genotypes, based on their amino acid capsid sequence. Molecular epidemiological studies have shown that in recent years approximately 70% of norovirus outbreaks among humans have been caused by one dominant genotype, GII.4 [17,82,157,268,301,303].

With continuous surveillance systems in place in some countries since the mid-1990s, it has become apparent that the number of reported norovirus outbreaks, and especially those caused by GII.4 strains has risen since the appearance of the 2002 variant of GII.4 [17,27,175,268,269,330]. Since then, genetically distinct GII.4 variants have emerged, and spread rapidly across the world causing epidemic waves of norovirus illness [269,270]. To date, three variants named after the year when they were first detected have been identified in populations across the world (the 2002, 2004 and 2006b variants [270]). The emergence of each of these three variants was followed by 'hot' norovirus winters with sharply increased numbers of reported outbreaks. Older strains belonging to the lineage designated 1996 were also detected around the world, although surveillance was limited at that time.

The pattern of continuous lineage turnover, referred to as epochal evolution [170], with emerging new variants replacing previously predominant circulating ones, is strongly reminiscent of what is observed in the molecular epidemiology of Influenza A virus (IAV). The evolutionary interaction between IAV and the human immune response results in antigenic drift, illustrated by the characteristic ladder-like tree shapes for hemagglutinin and neuraminidase surface proteins [98]. Long-term partial immunity to the virus induces sharp fitness differences among strains and drives rapid amino acid replacement at key antigenic sites, pinpointed by *in vitro* and *in silico* analyses [28,217]. Whereas antigenic data can be readily generated for IAV, allowing the comparative mapping of antigenic and genetic evolution [272], research of norovirus antigenic properties has been hampered by the lack of a simple cell culture model [69]. However, recent publications indicate that the genetic differences between norovirus genotypes, and also between variants of the GII.4 genotype, translate into distinct antigenic types, although molecular determinants remain largely unclear [107,170]. Thus, individuals may be repeatedly infected by strains belonging to different genotypes, and also, because immunity against norovirus infection is short-lived at best [128,227,336], possibly repeatedly by strains of the same genotype. As a result, the impact of immune responses on norovirus epidemiology remains poorly understood and phylodynamic and molecular adaptation studies may provide some key insights.

In this study, we aimed to provide a rigorous measurement of norovirus GII.4 diversity through time, and we investigated viral population expansions in relationship to the increased numbers of infections reported in recent years. Evolutionary and population dynamics of GII.4 noroviruses were estimated by a Bayesian coalescent approach, using two different datasets of sequences from strains with known detection dates, between 1987 and 2008. One set of sequences contained full capsid sequences, the other short partial polymerase sequences, which had been obtained for standard-procedure genotyping in norovirus surveillance in Europe [157] (http://www.noronet.nl/fbve/) and from the global norovirus surveillance network NoroNet [270] (http://www.noronet.nl/).

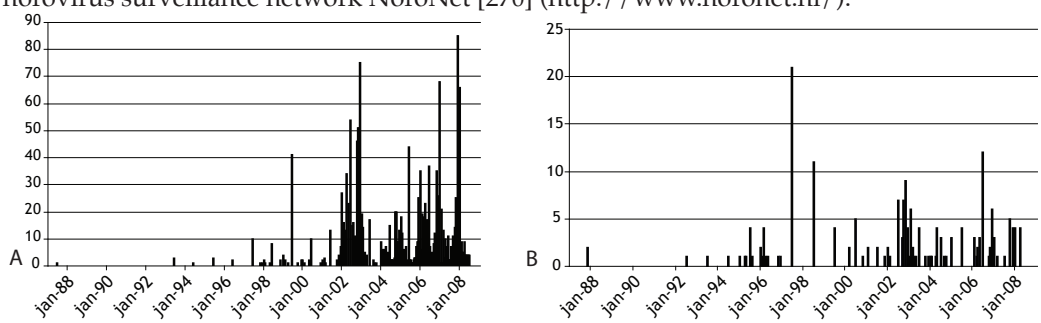


Figure 5.1. Detection dates of norovirus GII.4 strains included in the study.
A) The polymerase sequence detection dates. B) The capsid sequence detection dates.

We also tested whether these dynamics differed from neutral expectations, so whether and how they were shaped by selective pressure, and we attempted to further elucidate the molecular determinants of norovirus evolutionary and epidemiological dynamics using *in silico* techniques. To identify the molecular characteristics of norovirus GII.4 strain replacement, we investigated both directional and diversifying selection and elucidated capsid protein positions showing evidence for co-evolutionary dynamics acting between sites.

Materials and methods

Datasets and data preparation

We compiled two different norovirus GII.4 datasets. First, partial polymerase gene sequences with known detection month and year were collected. These sequences, encoding a short genomic region commonly referred to as Region A, have been collected for genotyping purposes, as an essential part of the ongoing surveillance practice in institutions around the world [270,321]. This dataset includes sequences from participants of the Foodborne Viruses in Europe (www.fbve.nl) and of the NoroNet (www.noronet.nl) networks, the contributing institutions of which are listed in the acknowledgements. Sequences of sufficient length (i.e. covering at least the final region) were included, generating a dataset of 1383 taxa, 247 nt in length. Strains originated from systematic surveillance collections, and form the best representative reflection of the circulating strains currently available.

Second, complete capsid sequences of GII.4 norovirus strains with known sampling date were collected. This resulted in a dataset of 194 taxa, 1623 nt long. To allow comparison of results from capsid based versus polymerase based analysis, a set of 172 partial polymerase gene sequences matching the sequences in the capsids dataset (identical variant typing and similar detection dates) were selected from the total polymerase dataset. The 2003Asia variant was excluded from this mirror-dataset, since it was identified as recombinant and ORF1 does not belong to the GII.4 genotype. Details on the nature of the strains comprised by the two generated datasets are provided in table 5.1. The distribution of the sampling dates of the included sequences is depicted in figure 5.1.

For recombination analyses, a set comprising 20 sequences, two of each GII.4 variant, spanning the genome region between Region A, in ORF1, and the complete capsid sequence was collected. Sequences were aligned using the Clustal W algorithm implemented in BioEdit (version 7.0.9.0) and edited where necessary. Sequence alignments can be obtained from the authors on request.

Recombination analysis

Recombination within the genomic area under study invalidates the use of phylogenetic

approaches. Therefore, we checked for possible recombination signal by analyzing 20 sequences (two for each variant) spanning Region A through the complete capsid protein (2404 nt). Different evolutionary histories across this genome region were inferred using the genetic algorithm for recombination detection (GARD) [150] and specific recombinants were identified using a modified VisRD algorithm [166] and using RDP3 [192]. In addition, we used the Phi test, shown to perform well under strong population growth and to be able to distinguish recurrent mutations from recombination events, to identify recombination signal in the norovirus alignments [22].

Table 5.1. Background information on strains comprised by the two datasets. Geog.: geographical, NA: Not Assigned, C'well: Camberwell

Capsid dataset

Geog.origin	*Oceania*	*North America*	*South America*	*Asia*	*Europe*	*Africa*	*Total*					
Strains (%)	7 (3.6)	22 (11.3)	1 (0.5)	29 (15)	135 (69.6)	0 (0)	194 (100)					
Variant	*1996*	*2002*	*2004*	*2006a*	*2006b*	*2001Japan*	*2002/CN / Henry*	*2003Asia*	*Bristol*	*C'well*	*NA*	*Total*
Strains (%)	67 (34.5)	45 (23.2)	12 (6.2)	17 (8.8)	24 (12.4)	4 (2.1)	2 (1.0)	12 (6.2)	2 (1.0)	3 (1.6)	6 (3.1)	194 (100)

Polymerase dataset

Geog. origin	*Oceania*	*North America*	*South America*	*Asia*	*Europe*	*Africa*	*Total*					
Strains (%)	3 (0.2)	6 (0.4)	1 (0.1)	60 (4.3)	1312 (94.9)	1 (0.1)	1383 (100)					
Variant	*1996*	*2002*	*2004*	*2006a*	*2006b*	*2001Japan*	*2002/CN / Henry*	*2003Asia*	*Bristol*	*C'well*	*NA*	*Total*
Strains (%)	175 (12.7)	433 (31.3)	214 (15.5)	159 (11.5)	377 (27.3)	16 (1.2)	1 (0.1)	NA	1 (0.1)	5 (0.4)	2 (0.1)	1383 (100)

Time-measured phylodynamic analyses and associated neutrality test

Evolutionary dynamics were estimated using a Bayesian Markov chain Monte Carlo (MCMC) approach implemented in BEAST (BEAST version 1.4.7 [63]). BEAST MCMC analysis estimates marginal posterior distributions for every parameter in a full probabilistic model comprising the timed evolutionary history, based on the incorporation of sampling times in a molecular clock model, the substitution process and demographic history. We used the GTR + I + Γ_4 model of substitution and the uncorrelated lognormal relaxed clock model to accommodate variation in substitution rates among different branches [62].

To test selective neutrality of GII.4 molecular evolution, we adopted the genealogical framework presented by Drummond and Suchard [65]. This involves the full model-based Bayesian analysis to obtain a posterior distribution of trees, genealogical summary statistics, and posterior predictive simulation to detect departures from the neutral expectations for these statistics. We employed the genealogical Fu and Li statistic (D_F), which compares the length of terminal branches to the total length of the coalescent genealogy. Strongly

negative values for this statistic indicate terminal branch lengths being larger than expected, which reflects an excess of slightly deleterious mutations on these branches. This statistic has proven to be most sensitive in uncovering non-neutral evolution, and has for example rejected neutrality for human IAV hemagglutinin genes [65]. Posterior predictive simulation is performed according to the same demographic model as used in to obtain the posterior tree distribution. To evaluate the impact of large demographic trends, we compared analyses using both constant and exponential growth population size priors. To further investigate the impact of demographic detail on the neutrality test we extended the simulation procedure to highly parametric demographic models, including piecewise constant demographic functions that define a Bayesian skyline plot model. We validated the simulation procedure by comparing the reconstructed Bayesian skyline plot from the trees generated by posterior predictive simulation with the Bayesian skyline plot inferred from the sequence data, which yielded consistent results.

To reconstruct the norovirus GII.4 demographic history in more detail, we employed the Bayesian skyline plot (BSP) model, which generates piecewise constant population size trajectories [64]. In this coalescent setting, demography is measured as the product of effective population size (N_e) and generation time (τ), $N_e\tau$, through time. To obtain a detailed measurement of norovirus GII.4 diversity through time, given the large dataset, we specified 40 groups in the piecewise constant population size function. All chains were run sufficiently long to achieve stationarity after burn-in, as checked using TRACER (http://tree.bio.ed.ac.uk/software/tracer/).

Additionally, the polymerase dataset was split up into separate subsets, each comprising all available sequences from a major GII.4 variant, and these were analyzed individually using the same model and settings as was used for the whole polymerase dataset.

To examine how different sampling densities through time can impact our demographic estimates, we performed an additional analysis of a subset of the polymerase data, containing sequences selected to best mirror the sequences in the capsid dataset, using the same specifications as described above.

Analyses of Molecular Adaptation

Codon substitution analyses

Selective pressure in the capsid genes was analyzed using the nonsynonymous/synonymous rate ratio (*dN*/*dS*) in a codon model framework (reviewed in [57]). Probabilistic models of codon substitution can be used to identify various types of selection pressures in the evolutionary history of a gene. Diversifying positive selection is characterized by $dN/dS > 1$, which indicates that (adaptive) nonsynonymous substitutions have accumulated faster than synonymous substitutions. We considered a generalization of the random effects *Dual* model [235], where synonymous and non-synonymous rates are drawn from independent

general discrete distributions. In this study, we fitted a general bivariate discrete distribution (GBDD) of substitution rates to the data using maximum likelihood in HyPhy [151,237]. A GBDD on D rates estimates $3D$-2 parameters: a pair of rates (synonymous and non-synonymous) and a weight for each class. Two constraints that must be satisfied by the parameters are: the weights must sum to one, and the mean for synonymous rates must also be one to avoid confounding the mean of substitution rates and the length of the tree (see [235] for details). This distribution allows dN and dS to co-vary, by directly estimating the values for dS and dN for D rate classes, instead of assuming that they are selected independently. The value of D was determined by a step-up procedure on D (starting with D=1), where the fit of models with D and D+1 rate classes was compared using small sample AIC. Posterior distributions of substitution rate parameters were approximated using sampling importance resampling. We identified sites under diversifying or positive selection using an empirical Bayes procedure. We chose an empirical Bayes Factor cut-off of 20 to classify positively selected sites [235]. Additionally we carried out a selection analysis using a random effects model (REL), and a fixed effects likelihood (FEL) approach [148] to confirm the predictions of the REL method and performed dataset-specific simulations [274] on internal braches, as considering internal branches is more relevant than the analysis of terminal branches in population level selection [236], to estimate the operating characteristics of the site-specific likelihood ratio test.

Directional evolution analyses

To detect directional selection we used the DEPS test that identifies statistically significant shifts in amino acid residue frequencies over the tree and/or an unusually large number of substitutions towards a particular residue in a maximum likelihood context [149]. DEPS tests whether the amino acid substitution rate towards a particular residue as estimated using a directionally biased model is significantly different from baseline, reversible substitution rates. For this procedure, we used the amino acid substitution rates estimated under the general reversible protein model (REV) for the baseline and directionally biased models and employed the rooted maximum clade credibility from the BEAST analysis.

Epistatic effects, or co-evolutionary analyses of amino acids

Co-evolving sites, showing evidence for concurrent amino acid substitutions, were identified using Bayesian graphical models (BGM) in a phylogenetic framework [239]. Briefly, this method infers the history of substitution events using a codon-based maximum likelihood phylogenetic approach. Correlated patterns of nonsynonymous substitutions are subsequently identified using BGMs and an order-MCMC algorithm. Two modes of analyses were performed allowing either a single or two co-dependencies per site. We reported sites identified with $P > 0.9$ by either mode.

RNA secondary structure predictions

Secondary structure predictions of the 5'end of the ORF2 encoding RNA were generated using the webbased RNA Fold Webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) [99].

Results

Recombination analysis

To examine the extent to which recombination has shaped norovirus evolution, we analyzed an alignment of 20 GII.4 sequences, two for each variant, spanning the genome from region A (a gene region in ORF1 that is commonly used for genotyping purposes, see the Materials and methods section, and [270] and [321]) up to the 3'end of the capsid. Using GARD (Genetic Algorithm for Recombination Detection), VisRD (Visual Recombination Detection) and RDP3 (Recombination detection program), significant phylogenetic variability was identified in this genome region, which could be attributed to a recombination event for the 2003Asia variant, a GII.4 variant previously identified as a recombinant lineage, mainly detected in Asia, and rarely in Europe, Oceania or the Americas [270]. The crossover point lay in the ORF1/2 overlap, a position previously identified as a recombination hotspot in norovirus [25,26]. Further analyses below were based on polymerase and capsid gene sequences that do not include this breakpoint, and no recombination could be detected in those individual data sets using the Phi test [22].

Time-measured phylodynamic analyses

To test whether GII.4 evolution deviated from selective neutrality, we applied a genealogical neutrality test that involved Bayesian coalescent inference, a tree-based summary statistic (D_F), and posterior predictive simulation [65]. Using a constant population size demographic prior, the capsid data set (194 sequences, 1623 nt) showed significantly more negative D_F than expected ($P < 0.01$), suggesting a selective process that generated a significant excess of mutations on terminal branches. The same was true using an exponential growth prior, but a Bayes factor test did not support an exponential growth scenario (ln BF constant versus exponential growth = -2.47). A Bayes factor comparison also favored a constant population size model over an exponential growth model for the matched polymerase data set (172 sequences, 247 nt) (lnBF = -2.23). In this case, we did not observe a significant difference in D_F ($P = 0.15$). Because the polymerase sequences were considerably shorter, they provided less information to evaluate branch length properties. To counteract the loss of power we increased the number of sequences, and indeed, an analysis of the complete polymerase sequences dataset (1383 sequences, 247 nt) rejected the model of neutral evolution ($P = 0.001$).

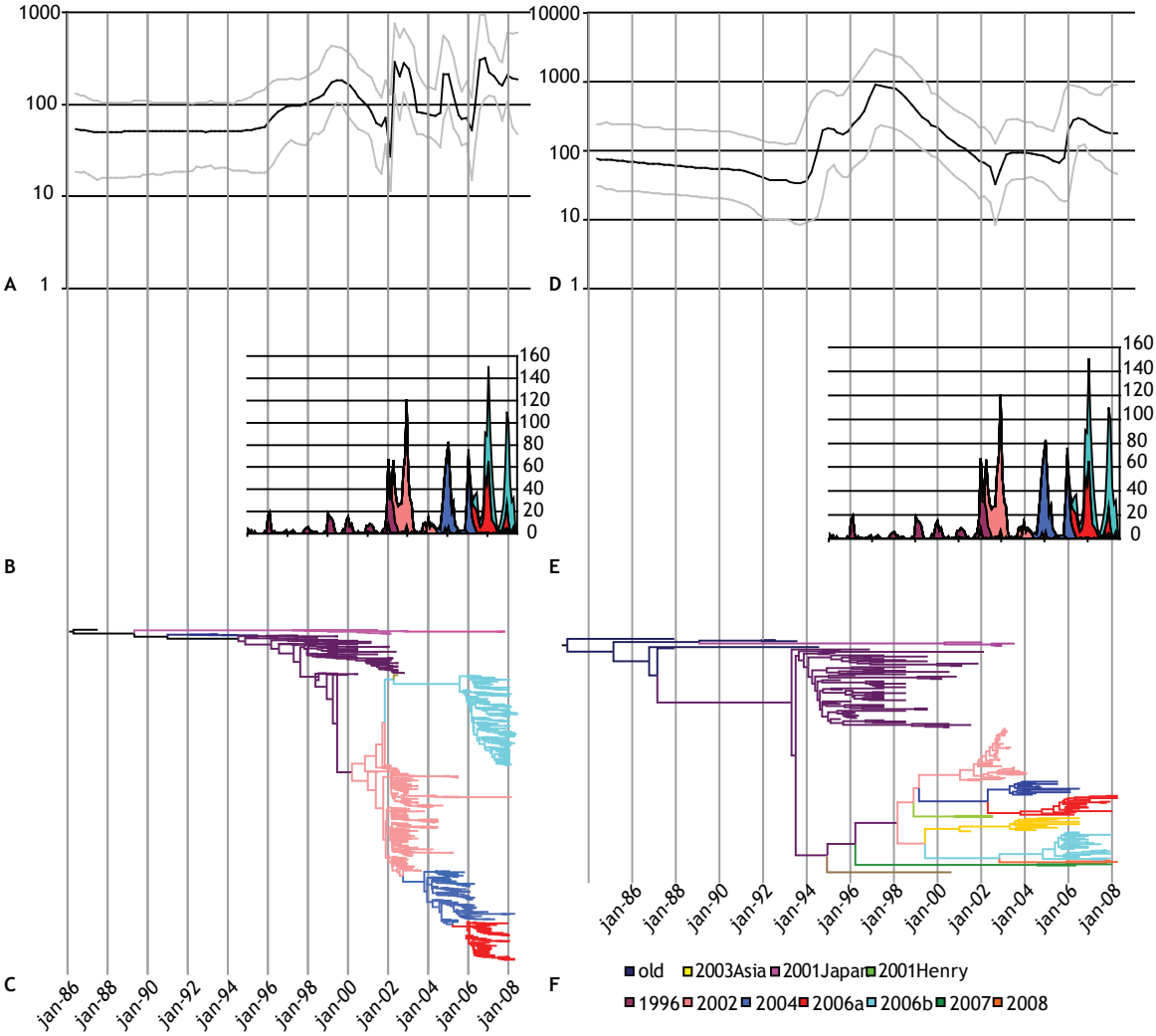


Figure 5.2. Phylodynamics of the GII.4 noroviruses. The left panel (A, B, C) describes analysis of the polymerase dataset, the right panel (D, E, F) the analysis of the capsid dataset. A) Bayesian Skyline Plot (BSP) of GII.4 norovirus partial polymerase sequences, representing the relative genetic diversity, a measure for the number of effective infections, of circulating GII.4 noroviruses through time. The black line represents the median posterior value, the grey lines the 95% Highest Probability Density (HPD) intervals. The Y-axis depicts the value of Neτ on a logarithmic scale. B and E) Surveillance data of norovirus GII.4 strains, detected in outbreaks, included for comparison. Only European surveillance data was obtained from the FBVE database and from Dutch surveillance [12,13]. The different GII.4 variants detected in the outbreaks are represented by different colors, showing the displacement of subsequent variants through time. The same colors were used in panels D and F, the color legend is shown under panel F. C) Maximum Clade Credibility (MCC) Tree of norovirus GII.4 partial polymerase dataset. The MCC tree represents an estimate of the posterior distribution of tree topologies and branch lengths. Different variants are represented in different colors, the tree is scaled in units of time with tips constrained at to strain detection dates. D) BSP of the GII.4 norovirus capsid dataset. F) MCC tree of the GII.4 norovirus capsid dataset.

As previously introduced [37], this neutrality test relies on relative restricted demographic models governed by a limited number of parameters to capture large-scale demographic trends. To investigate the sensitivity to demographic detail, we extended the posterior predictive simulation procedure to accommodate highly parametric demographic models, which result in a more complex picture of norovirus dynamics. Using a Bayesian skyline plot (BSP) model as demographic function [38], similar conclusions could be drawn from the neutrality test: significantly more negative D_F values than expected under neutrality were observed for the capsid data set (P = 0.019) and the complete polymerase data set (P = 0.018), whereas the matched polymerase lacked power in rejecting neutrality (P = 0.029).

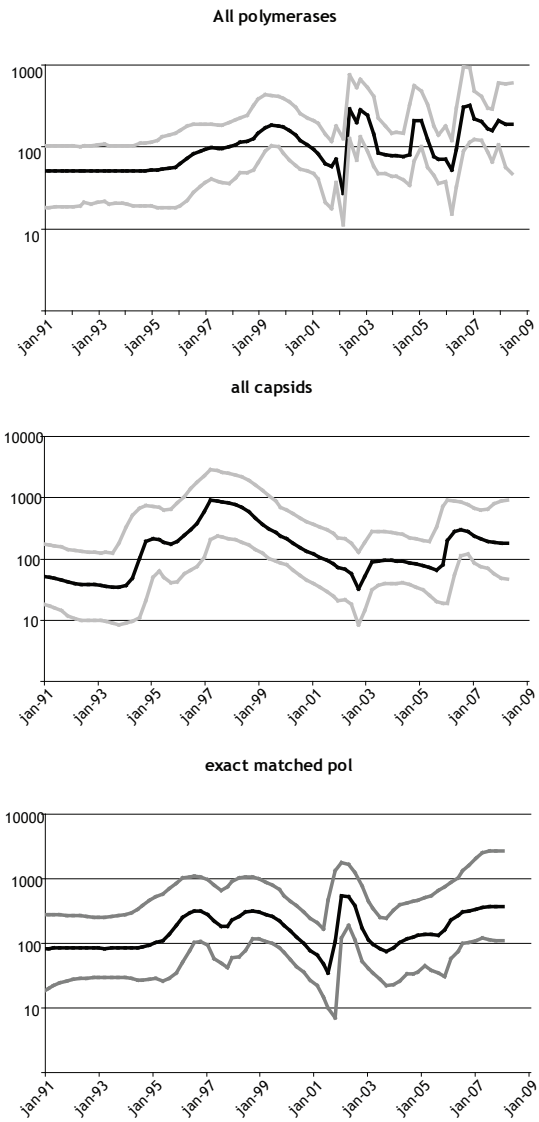


Figure 5.3. Comparison of Bayesian Skyline Plots. Graphs shown of the polymerase dataset, the capsid dataset, and the dataset comprised of partial polymerase sequences matching the sequences in the capsid dataset.

The demographic inference using the BSP model is summarized in figures 5.2a and 5.2b, which essentially plot $N_e\tau$ as a function of time. $N_e\tau$ can be considered a measure of relative genetic diversity that, in turn, reflects the number of effective infections established by the virus (see also the Materials and methods section). Uncertainty in the estimated parameters was evaluated using 95% Highest Probability Density (HPD) intervals. The Maximum Clade Credibility (MCC) trees from the same Bayesian analyses (figures 5.2c, d) summarize the norovirus evolutionary histories, and the stepwise emergence of the subsequent variants on a time scale. For comparison, surveillance data of reported norovirus outbreaks with confirmed GII.4 variant type were imposed on the BSPs.

The changing patterns of norovirus genetic diversity generally revealed seasonal dynamics, albeit with markedly varying resolution among the two datasets. The BSP for the polymerase dataset (figure 5.2a) showed peaks for $N_e\tau$ that coincided with the epidemic peaks observed in norovirus surveillance systems in the northern hemisphere winters 2002-03, 2004-05 and 2006-07. The BSP obtained from the capsid dataset (figure 5.2b) showed a pattern that was more difficult to reconcile with epidemiological observations. Values for $N_e\tau$ were highest in the years 1997-1999, and the emergence of the 2002 variant, which had a strong impact in the population according to surveillance data, did not coincide with a pronounced upsurge in the BSP. Comparison of the BSPs obtained for both genes, illustrated that unraveling seasonal population dynamics with associated population bottlenecks for viruses like norovirus, may require a sufficiently high sampling density. In fact, reducing the partial polymerase dataset to a similar number of sequences drastically diminished the resolution of the BSP analysis (figure 5.3). In particular the 2004-05 and 2006-07 epidemics were not well reflected in the BSP derived from this sub-set of polymerase sequences that matched the capsid dataset in size, both genetically and temporally. The 2002-03 epidemic, following the replacement of the 1996 variant by the 2002 variant, wás however clearly noticeable in the matched polymerase set, whereas it was not in the capsid data set. Considering the associated MCC trees (figures 5.2c, 5.2f), it is conceivable that following the relatively long build-up of genetic variation during the circulation of the 1996-variant, its replacement by the 2002 variant signified a massive and sudden loss of diversity; a population bottleneck. The 2002 variant split into two distinct subclusters for the polymerase dataset. These lineages arose almost immediately after the emergence of the 2002 variant, and individually coalesced to a Most Recent Common Ancestor (MRCA) shortly before their diversification. The capsid 2002 variant cluster also grouped in two sublineages, but they coalesced more gradually to their MRCA.

Comparison of the variant dynamics in the MCC trees to their respective BSPs suggested that variant replacement was not always absolute across subsequent epidemic seasons. To

investigate this in more detail, we performed the coalescent analyses on partial polymerase datasets for the individual major GII.4 variants separately (figure 5.4). Whereas the pattern of rapid emergence, followed by an (epidemic) peak and later peaks of diminishing size observed for the 2002, 2004 and 2006a variants were very similar, the patterns obtained for 1996 and 2006b were quite different. The 1996 variant, that was detected in the population during a relatively long period, but at low reporting frequencies after the initial epidemic (winter of 1995-1996, in the northern hemisphere) (figure 5.2a/b), showed an increasing trend in the $N_e\tau$ values persisting long after this first peak. The 2006b strain showed a less defined pattern, with multiple, smaller peaks.

The demographic component is part of a full Bayesian model that enables the inference of time-scaled evolutionary histories and rates of molecular evolution from temporally-spaced sequence data. Rates of nucleotide substitution and the MRCA's of the included GII.4 sequences were listed in table 5.2. The substitution rates found for the less densely sampled datasets, namely the complete capsid sequences (5.33×10^{-3} substitutions per site per year) and the matched polymerases set (4.32×10^{-3} substitutions per site per year) are lower than the rate found for the large partial polymerase dataset (8.98×10^{-3} substitutions per site per year). The estimated MRCA for these GII.4 strains lies in the first half of the 1980s (table 5.2).

Table 5.2. Evolutionary parameters inferred by Bayesian analysis.

Dataset (number seqs, length of seq)	Median MRCA (HPD)	Median Rate (x 10^{-3}) Subst / site / year (HPD)
GII.4 Capsids N=194, length = 1623nt	July '82 (Oct '77 – May '86)	5.33 (4.62 – 6.02)
GII.4 Polymerases N=1383, length = 247 nt	Oct '85 (Sept '82 – Jun '87)	8.98 (7.73 – 10.4)
GII.4 Polymerases matched to capsids N=172, length = 241 nt	Jun '85 (Jun '79 – Mar '88)	4.32 (3.50 – 5.20)

Molecular adaptation

Because the genealogical test rejected selective neutrality for the capsid gene, we attempted to identify the molecular determinants of this selective process through two different approaches that were not previously applied on norovirus data, namely DEPS [149] and co-evolutionary analysis of amino acids [238], and complemented this with novel extensions of previously performed codon substitution model analyses [18,170]. The partial polymerase sequences under analysis in this study are very short and have therefore not been included in these analyses.

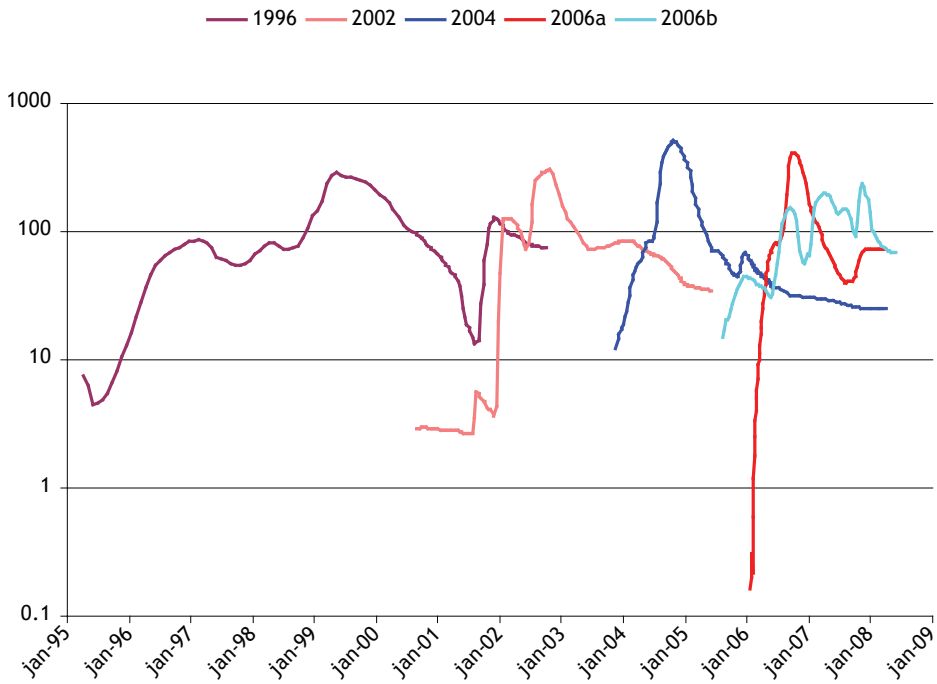


Figure 5.4. Bayesian Skyline Plots of different GII.4 variants analyzed separately. Only the mean population sizes through time were plotted for clarity.

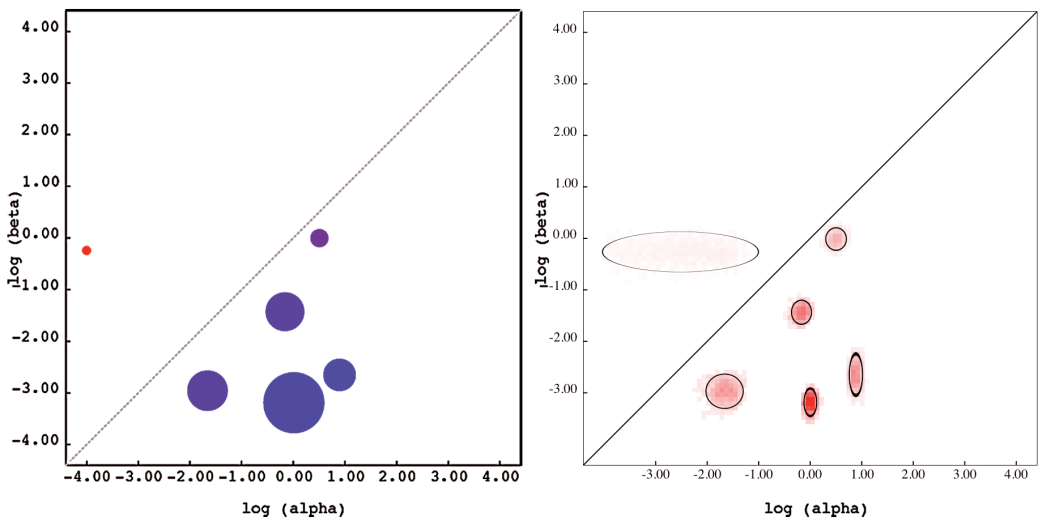


Figure 5.5. Codon substitution rate classes of GII.4 capsids. A) Inferred weights for each evolutionary rate class, showing the relative proportion of sites belonging to the given classes. The diagonal line divides the regimes of positive selection (dN/dS > 1, above) and negative selection (dN/dS < 1, below). B) Approximate inferred posterior distribution of synonymous (alpha) and non-synonymous (beta) substitution rates, showing the variance of each rate estimate. Color intensity is proportional to the square root of the density, and solid oval plots delineate approximate 95% confidence sets.

Codon substitution model

In order to apply a codon model based on a general bivariate discrete distribution (GBDD) of dN and dS [151], we employed a small sample AIC, which suggested that six rate classes (D=6) provided the best fit to the capsid data. The proportion of sites within these classes and corresponding dN and dS estimates are represented by figures 5.5A and 5.5B. The model included one rate class describing positive selection (dN (=0.77) > dS (=0.00)), with an estimated 0.93 % of sites occupying this class. An empirical Bayes approach identified sites 6, 9, 15, 47 and 534 (0.93%) to be under diversifying positive selection (table 5.3). Three of the sites (6, 9 and 534) were confirmed by a site-by-site Fixed Effects Likelihood (FEL) analysis at $p<0.05$, while the remaining two (15 and 47) were borderline significant ($p=0.06$ and $p=0.10$). The rates of false positives for FEL analyses at $p=0.05$ was approximately 0.04 and 0.08 at $p=0.1$, based on dataset-matched neutral simulations, suggesting that the putatively selected sites were not due to elevated rates of false positives at given nominal significance values.

Table 5.3. Codons under selection, identified by site-by-site FEL analysis.

Codon	Empirical Bayes Factor	FEL p-value	Substitution Pattern
6	125	0.01	0 syn 6 non-syn (N:S)
9	3125	0.007	0 syn 9 non-syn (N:S/T)
15	177	0.06	0 syn 7 non-syn (A:T)
47	37	0.10	0 syn 6 non-syn (I:V)
534	761	0.04	0 syn 9 non-syn (A:T/V)

To uncover population level selection processes, FEL analysis may be more appropriately applied to internal branches (iFEL) [236]. That this approach suited our data was also suggested by our genealogical tests, which identified an excess of slightly deleterious mutations on terminal branches, indicative of within host evolution. Using iFEL enabled us to avoid this effect and revealed 8 codons under positive selection at the population level including 6, 9, 47, 352, 372, 395, 407, 534, with $p \leq 0.05$.

Directional selection model

Codon models are powerful tools to detect an unusually high rate of nonsynonymous replacement, which generally occurs under a scenario of diversifying selection. However selection of episodic nature, e.g. directional selection or frequency-dependent selection is more difficult to detect and involves the question of which residues are being selected for or against [149]. A directional evolution in protein sequences analysis (DEPS) of norovirus

capsid sequences revealed elevated substitution rates towards 4 residues: V, S, A, T. Four sites were identified to be involved in this directional evolution; amino acids 9 (with inferred amino acid substitution pattern: N→T/S→N), 294 ((V→)A→S/P→A→T), 333 (L→M/V/L→M→V), and 395 (- →T→A) (figure 5.6).

Epistatic effects, or co-evolutionary analyses of amino acids

In folded proteins amino acids are not arranged linearly; many functionally interact, making their evolution dependant on that of others. Various types of interactions exist, and interacting sites are not necessarily direct neighbors in either the protein sequence or in the 3D protein structure. We used Bayesian graphical models (BGM) to detect co-evolving sites. The sites identified, are shown as a network in figure 5.7, and sites for which co-evolution was detected but seemed less supported are shown in figures 5.8A and 5.8B. Two values for the posterior probabilities are given, obtained from the analyses allowing for either one or two co-dependencies. Sites 231 and 209, and 238 and 504, which co-evolved as two coupled sets (figures 5.8A and 5.8B), were not involved in recent variant transitions. Therefore we conclude that they were not under selective pressure that governed variant replacement dynamics. All sites for which molecular adaptation was detected are listed in table 5.4. We marked relevant sites located on top of the capsid dimer in figure 5.9.

RNA secondary structure predictions

Our study identified codon 6 to be under positive, diversifying selection and codon 9 to be under diversifying as well as directional selection; others performing dN/dS analyses detected positive selection at these sites as well [18,170]. The variation in codon 6 is due to A<u>A</u>T (Asn) to A<u>G</u>T (Ser) changes. The signal for codon 9 is solely attributable to substitutions at the second codon position of this site; A<u>A</u>C (Asn), A<u>C</u>C (Thr) and A<u>G</u>C (Ser). The N-terminal region of ORF2 was otherwise highly conserved. Because contrary to the other amino acids that were identified to be under selective pressure, the part of the protein encoded by these two amino acids is located on the inside of the virus capsid structure, and not surface exposed, we investigated the potential RNA secondary structure encoded by this region. *In silico* replacement of nucleotides at position 17 (codon 6) did not lead to secondary structure changes (not shown). Secondary structure predictions of the RNA encoding the 5'end of ORF2 were performed with all four possible nucleotides modeled at position 26 (figure 5.10).

The 4 nucleotides upstream of the ATG, generally thought to form the boundary of the subgenomic RNA [42] were included in the predictions. The presence of A, C, or G generated similar structures, when however a T (U) was modeled, extra pairing possibilities arose, lengthening the stem of the first stem-loop structure, and thus shortening the stretch of free nucleotides, available for ribosome binding, from 11 to 9.

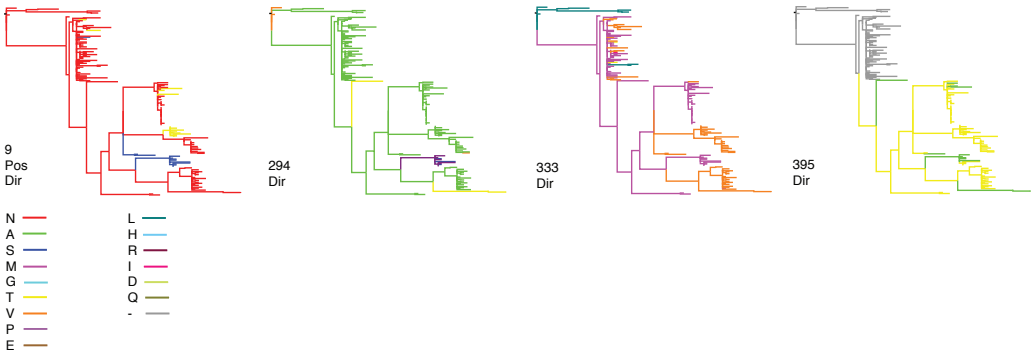


Figure 5.6. Sites identified to be under directional selection by DEPS analysis. Sites are depicted by MCC trees colored for which amino acid was present on each branch.

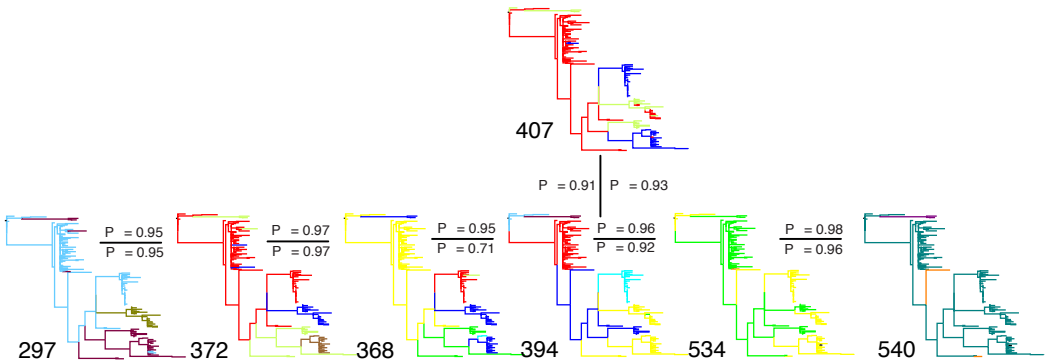


Figure 5.7. Sites identified by Bayesian graphical models to co-evolve. Amino acid sites 297, 372, 368, 394, 407, 534 and 540 of the GII.4 capsid protein were depicted as trees showing different amino acids in different colors. Each connection is associated with posterior probabilities (P) for single (above or left) or two (below or right) dependencies.

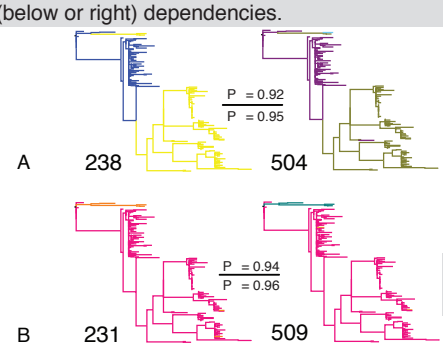


Figure 5.8 Other co-evolving sites.
A) Sites 238 and 504 B) Sites 231 and 509.

Discussion

Following the emergence of the genetically and antigenically novel GII.4 2002 variant in 2002, a sharp rise in the number of reported norovirus outbreaks was recorded in several surveillance systems throughout the world [17,175,330]. Earlier, in 1995-96, a similar rise

had been reported resulting from the emergence of another cluster of GII.4 - the 1996 variant [320]. The 1996 and 2002 variants have by now been succeeded by successive, highly prevalent GII.4 variants. Some uncertainty remained, however, as to whether the increase in the number of reported outbreaks was the result of a true rise in the numbers of infections, or of improved detection techniques combined with better reporting due to heightened awareness. In order to resolve this ambiguity in norovirus epidemic history we applied Bayesian statistical phylodynamic techniques (Bayesian Skyline Plots, or BSPs) to alignments of a large quantity of GII.4 sequences. We showed that these techniques confirm the epidemic behavior of the GII.4 variants during the past decade, and that the increase in the number of norovirus infections suggested from global surveillance data coincides with the rise of the GII.4 variant. Additionally we applied *in silico* techniques to identify sites under various types of selective pressure and to unravel epistatic interactions in the capsid protein.

BSP reveals most recent epidemics, but no additional ones

The first rise in the BSP estimates of $N_e\tau$ (a measure of effective population size) in the polymerase dataset was seen just before January 1996, around the time the first global GII.4 epidemic was noted [212] (figure 5.2). Interestingly, rather than appearing as a defined epidemic peak lasting one winter season, $N_e\tau$ increased continuously through January 2000. Surveillance data from around the world identified relatively few GII.4 outbreaks among norovirus positive outbreaks in the period between 1997 and spring 2002 [268,270]; instead, a relatively high diversity of other, non-GII.4 genotypes was detected in this period. The high $N_e\tau$ observed for this period is congruent with the long branch-lengths in the MCC tree during this period. Given the prolonged circulation time of the 1996 variant compared to the later variants, it seems likely that the BSP in this instance is a better representative of the relatively high genetic diversity built up in an extended period of co-circulation of the 1996 variant with other genotypes, rather than of the number of infections with this particular variant. The three most recent GII.4 epidemics caused by emerging GII.4 variants were clearly visible in the polymerase BSP. The 2002 epidemic was preceded by a sharp decline in $N_e\tau$, indicative of a purifying selection event against the previously dominating variant, in which the diversity that had been gradually built up by the long circulation of the 1996 variant collapsed. Next, coinciding with the off-seasonal peak observed in public health surveillance systems in the northern hemisphere, $N_e\tau$ in the BSP rose sharply. After a brief and insignificant decline another peak occurred in the winter season of 2002-03. During the epidemic caused by the 2004 variant $N_e\tau$ peaked slightly less high and less long than during the 2002 epidemic peak. During the 2006-07 epidemic, with two distinct GII.4 variants circulating, $N_e\tau$ was the highest (figure 5.2), and after this winter $N_e\tau$ values stayed high, corresponding with continued circulation of the 2006b variant, as also observed in

surveillance.
Altogether, from the first half of the 1990s to present time, two major changes were observed. First, epidemic waves have arisen that could not be detected in earlier times, and second, the number of infections has gone up. The baseline before 1996 corresponds to the trough levels in between contemporary epidemics. Hence, the increased amount of GII.4 outbreaks observed in outbreak surveillance seems to reflect an actual increase of GII.4 infections in the population.
These conclusions are undoubtedly impacted by the comparatively few available sequences from earlier years, which are, unfortunately, not widely available, and as shown by the recent publication by Bok *et al.*, not easily attainable [18]. A search in archival stool samples revealed that during the years 1974-1981 and 1987-1991 GII.4 was not the most prevalent norovirus genotype in hospitalized children with gastroenteritis, but GII.3 was. Our sampling, albeit at lower frequency during this period, would probably have detected possible unidentified epidemic surges had they occurred. Altogether, although surveillance data and demographic estimates are very different types of information, their dynamics match remarkably well. The surveillance data is presented on a linear scale and reflects the *reported* outbreaks of norovirus-gastroenteritis, which is probably an incomplete description of norovirus circulation, as many cases of norovirus illness remain unreported. The BSPs are presented on a log-scale, which makes for less sharp peaks than the peaks observed in the surveillance data graphs.

Different GII.4 variants reached similar levels of genetic diversity
The analysis of each major GII.4 variant separately (figure 5.4) indicated that $N_e\tau$ values, a measure for relative genetic diversity, and a proxy for the number of effective infections, of each variant reached approximately similar proportions. The 2002, 2004 and 2006a variants each had one epidemic season, but the turnover was relatively slow compared to e.g. influenza [248], resulting in repeated seasons (of decreasing magnitude) of illness caused by the same variant. For example, one main peak for $N_e\tau$ was observed for the 2004 variant during the 2004-05 winter, but a smaller peak followed during the 2005-06 winter. This slow turnover may very well have been caused by the fact that only incomplete immunity is mounted against norovirus after infection or alternatively that the pool of susceptibles is not depleted within one season. Different BSPs were obtained for GII.4 variants 1996 and 2006b, that persisted longer than one season. Interestingly $N_e\tau$ for the 1996 variant increased just before the emergence of the 2002 variant. The 2006b variant showed no defined epidemic peak, but a series of smaller peaks that did not coincide with annual winter-peaks. Two sublineages of the 2006b variant, distinguishable by up to 5 amino acid differences in the full capsid sequence [265] circulated in the population. Of these only S368G has been recognized as a polymorphism relevant for antigenic properties. Interestingly, while only two 2006b strains

with this mutation are present in our capsid dataset, later full capsid sequencing revealed more strains of this sublineage, and the polymerase dataset included more sequences from this cluster. Thus, the 2006b variant may have persisted in the population by changing its antigenic properties.

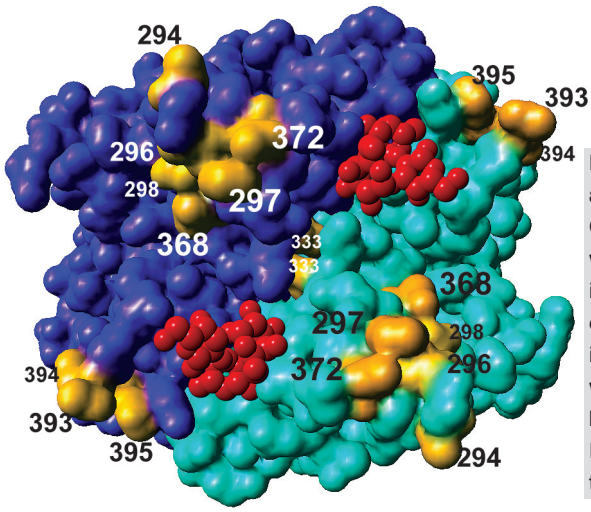


Figure 5.9. Sites identified by molecular adaptation analyses located on top of the GII.4 capsid dimer. Two parts of the dimer were given different shades of blue, the sites identified as under positive selection, co-evolving with other sites, or sites previously identified to be involved in host-interactions, were colored yellow. Ligands (B trisaccaride) bound in the binding pocket are shown in red. Protein structure 2OBT was used for generating this representation [32].

Table 5.4. Molecular adaptation in the capsid protein, detected by different maximum likelihood approaches.

aa number	REL	FEL	iFEL	DIR	Co-Ev
6	x	x	x		
9	x	x	x	x	
15	x				
47	x		x		
231					x (509)
238					x (504)
294				x	
297					x (372)
333				x	
352			x		
368					x (372, 394)
372			x		x (297, 368)
394					x (368, 407, 534)
395			x	x	
407			x		x (394)
504					x (238)
509					x (231)
534	x	x	x		x (394, 540)
540					x (534)

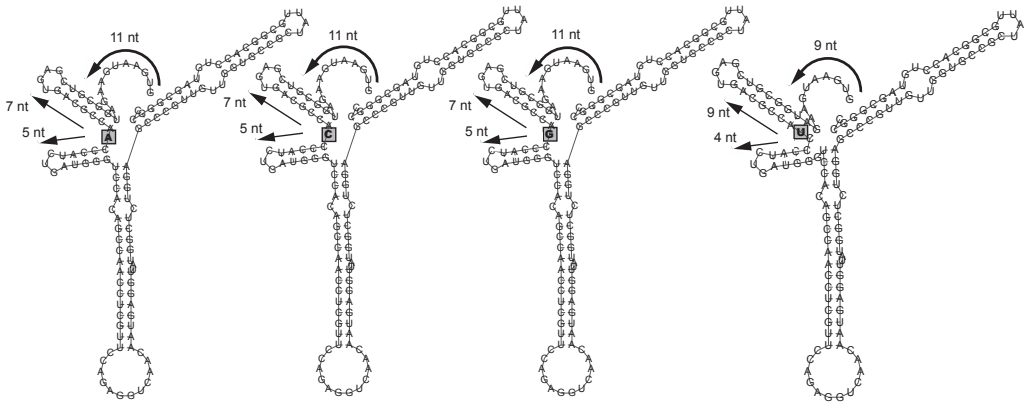


Figure 5.10. Predicted secondary structures in the 5'end of the ORF2 GII.4 RNA, with all possible nucleotides modeled at site 26 (boxed). Four nucleotides upstream of the first ATG were included, as these are conserved in the 5'end of ORF1 and ORF2 and likely to have a function in translation.

High sampling density is necessary for accurate estimate of the effective population size of cyclic pathogen

While Bayesian coalescent analyses of the large partial polymerase dataset reflected seasonal epidemic dynamics, the analysis of the longer capsid sequences offered little detail about the phylodynamic patterns of norovirus. To investigate the nature of this difference, we performed an analysis of polymerase sequences matching the capsid sequences genotypically as well as temporally, which showed that a lack of phylodynamic detail can generally be attributed to a lower sampling density (figure 5.3). This may not be so surprising as it was previously thought that coalescent analyses would not be so effective at capturing cyclical population dynamics [244]. Only recently, a comprehensive analysis of a large H3N2 influenza virus dataset (1302 taxa) was able to uncover seasonal dynamics from genetic data [248]. Interestingly, this data set was similar in size compared to the large dataset of partial norovirus polymerases (1383 taxa) presented here, although in the influenza study full gene segments were analyzed. The population bottlenecks in the norovirus GII.4 population history are to a large extent comparable to those seen for influenza and constitute repetitive large scale losses of genetic diversity. We note that viruses with seasonal dynamics do not necessarily have to exhibit such dynamics in genetic diversity. Short infections with strong cross-immunity, as seen for measles virus, allow many strains to co-circulate with frequencies contingent on neutral epidemiological processes [98]. For such viruses, seasonal epidemics may arise from repeated exhaustion of susceptible host populations [97]. Therefore, our phylodynamic analysis predicts that there may only be partial subsequent cross-immunity against GII.4 variants. It is important to note that sampling size impacts the resolution of phylodynamic inference, but the actual sampling scheme does not dictate a pattern of fluctuating population size. Rambaut *et al*.

[248] performed simulations using various demographic scenarios but with a sampling scheme used to obtain influenza genetic data from seasonal epidemics. In all cases, the simulated demographic history was accurately recovered. A sparse sampling prior to 2000 also makes it difficult to unequivocally conclude an increase in GII.4 infections. However, we note that the value for $N_e\tau$ before the first documented epidemic resulting of the emergence of a new genetic variant (the 1996-variant) in the BSP is lower than the estimates between subsequent epidemic peaks. This suggests an increase in the number of norovirus GII.4 infections, which is further reinforced by a recent study of archival stool samples from the Children's Hospital, Washington, DC (1974 to 1991) (Bok et al., 2009). Although this study identified GII.4 strains in the early seventies, this was not the predominant genotype before 1991 (Bok et al., 2009). An increase of the GII.4 variant therefore seems to provide a plausible explanation for the coincident increase in the number of norovirus infections.

Recently circulating GII.4 strains share a most recent common ancestor in the early 1980s

The estimated substitution rates (9×10^{-3}substitutions per site per year for the partial polymerase sequences and 5.3×10^{-3} substitutions per site per year for the capsid sequences) corresponded to the values recently reported for norovirus GII.4 [18] and were well within the range of what is commonly found for RNA viruses, e.g. between 3.5×10^{-3} and 8.5×10^{-4} for HMPV complete genomes [50] and for influenza A virus the highest rate, reported for the HA gene, was 5.72×10^{-3} substitutions per site per year [248]. A lower rate was observed for the polymerase subset matched to the capsid data set. Since the TMRCA estimates were consistent between these two polymerase data sets, the lower rate may be explained by differences in rate variation among sites, in particular for the proportion of invariant sites, which is sensitive to the number of taxa in the data set [282]. Dating the MRCA for these strains back to the early 1980's does not mean that the GII.4 lineage arose only then, but rather suggests that the strains that circulated during the past two decades share a common ancestor at that time. It seems not unlikely that the GII.4 lineage was less diverse before the 1980's, not comprising different variants as during the past decades. Alternatively, if multiple GII.4 variants did exist before the MRCA of the current GII.4 variants, the occurrence of a population bottleneck may have left progeny virus of only one variant. The strains detected in the 1970s reported recently seem to confirm that multiple GII.4 variants existed before the Camberwell cluster arose [18]. Analyzing data of multiple norovirus genotypes will provide a more detailed insight into the branching times of these different genotype clusters, and also in this case, including older sequences will be more elucidating.

Possible shortcomings of this study

Ideally we would have used full genome sequences. However, for (GII.4) norovirus these are only sparsely available. Instead, we showed that Bayesian coalescent demographic analyses of a large dataset containing very short sequences offered important and reliable

insights into GII.4 variant dynamics. For the GII.4 norovirus datasets presented here, the densely sampled but short polymerase sequences provided data that better defined the epidemic history than fewer longer capsid sequences. Perhaps additional to the limited size of the capsid dataset, strong selective pressure on the capsid protein confounded analysis of the capsid gene. We were also aware of the possibility that our capsid sequences dataset may have been biased in sampling. Sequencing of full capsid genes is not standard practice; the viruses of which sequences were available have all previously been selected by various researchers as 'interesting enough to sequence'. Thus, relatively many sequences of strains belonging to the 1996 variant are present in the dataset, especially compared to strains belonging to the younger variants, 2004, 2006a and 2006b.

Selective pressure shaped the population dynamics of norovirus GII.4 variants

Our genealogical test clearly rejected neutral evolution for the norovirus capsid dataset. Although P-values were somewhat higher using a Bayesian skyline plot model in the posterior predictive simulation compared to more restricted demographic functions, we arrived at the same conclusions for all analyses. This demonstrates that the neutrality test is not overly sensitive to the demographic detail in the analysis. Nevertheless, through the advances made here, we demonstrate that this test can now be performed under any complex demographic scenario, a generalization that may further promote its use. To investigate the selective forces in more detail, we fitted different evolutionary models to identify sites in the capsid under selective pressure. We examined an extension of previously performed dN/dS analyses [18,170] to detect positively selected sites involved in population level selection, avoiding the effect of within host evolution. Eight positively selected codons were identified at $p<0.05$ with the iFEL approach. Of these sites, four (352, 372, 395 and 407) are located in the protruding regions of the protein. These sites have also been identified previously as 'informative' sites (at least two shared an identical amino acid mutation in the alignment) [269]. Amino acid 395, that was also detected as directionally evolving, is located in a surface exposed loop of the capsid protein, that is part of a variable site of carbohydrate interaction (amino acids 393-394-395) that has been identified by a number of studies as a locus for ligand binding and specificity [6,32,170]. Codon 394, located in this same domain, is part of an intricate network of co-evolving positions, also containing codons 297, 368 and 372. Amino acid 297 is part of a site identified by Allen *et al*, (296-297-298) [6], predicted as another of two host ligand binding sites. This particular site was not identified by Cao *et al.* [32] who performed co-crystallization assays with P-particle dimers and A and B-trisaccharides, the host ligands of norovirus. It is structurally close to amino acids 368 and 372, on top of the protein, flanking the ligand binding pockets, and to 294, under directional evolution, which is located on the outside of the 296-298 loop, relative to the binding pocket.

Lindesmith *et al.* also identified sites under positive selection, using fewer sequences. They used three different methods, single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL), under the Tamura-Nei model of evolution. More codons under positive selection were identified in this study, but less strict nominal significance values were used. Bok *et al.* [18] applied SLAC analyses for detection of positive selection and found six amino acids under positive selection. We chose to identify sites under selective pressure for
internal branches in a phylogeny using iFEL because external branches are prone to deleterious mutational load. Such mutations are expected to be young and more likely fall on the external branches of a population-level phylogeny [245], where they can confound the identification of positively selected sites. To avoid this adverse effect we focused on internal branches only, on which advantageous mutations are more likely to fall.

Codon 333 was previously identified as an informative site [269] and our analyses found it to be under directional selection. It is located in the hydrophobic part of the P-dimer interface, just below the carbohydrate binding site described by Tan *et al.* [290], facing its counterpart in the other protomer of the same dimer (distance 4 Å) [32]. Changes in this amino acid are not likely to be involved in antigenic change but more likely structural compensation for other mutations.

Positive selection is not only detected for sites involved in immune escape; secondary RNA structures are important for efficient RNA translation

Codons 6 and especially 9 have consistently been identified as sites under strong selective pressure, and involved in defining the distinction between variants, not only in this study but in others as well [83,170]. Given their positions at the N-terminus of the protein, inside the shell, they seemed unlikely to have been under selective pressure through antibody recognition. This notion led us to investigate the RNA encoding the 5'end of ORF2. The nucleotide sequence upstream of codon 9 (nucleotide 26) is strongly conserved among all norovirus genotypes, and a highly similar sequence is found at the 5'end of ORF1 in norovirus. Mutations have rarely been detected here, apart from at nucleotide 26, and at nucleotide 17 (codon 6). We propose that secondary structure predictions of these RNA regions provide an explanation for this pattern (figure 5.10). Highly conserved stem-loop structures create the circumstances necessary for translation initiation of ORF1 and ORF2. Nucleotides A, C and G at position 26 all result in almost identical structures, in which 7 nucleotides from the start of ORF2, or 11 when including the 4 nucleotides upstream from the AUG, are left free. When theoretically a U/T is inserted here the predicted structure changes, resulting in a diminished length of the free nucleotide strand of 5 nucleotides. We are unaware of sequences with nucleotide T at position 26 present either in our or the

public databases (data not shown), leading us to believe that a length of at least 7 unpaired nucleotides counting from the first AUG is necessary for efficient RNA translation from the subgenomic RNA. Thus, while normal capsids could be formed from strains with (silent) mutations in this area, the replication process of the virus may be disrupted by altering the secondary RNA structures. This theory is further supported by the observation that no other synonymous or non-synonymous mutations are found in the 9th codon, e.g. at the third nucleotide position, nor at other nucleotides in this genomic area, apart from the previously mentioned nucleotide 17, in an RNA loop. Tentative analyses of other norovirus genotypes demonstrated that the 5' region is equally conserved within the different genotypes and yielded similar secondary RNA structures, which allow for point mutations in the loops of the structures (data not shown). To further substantiate this hypothesis site-directed mutagenesis studies are required, which go beyond the scope of this study.

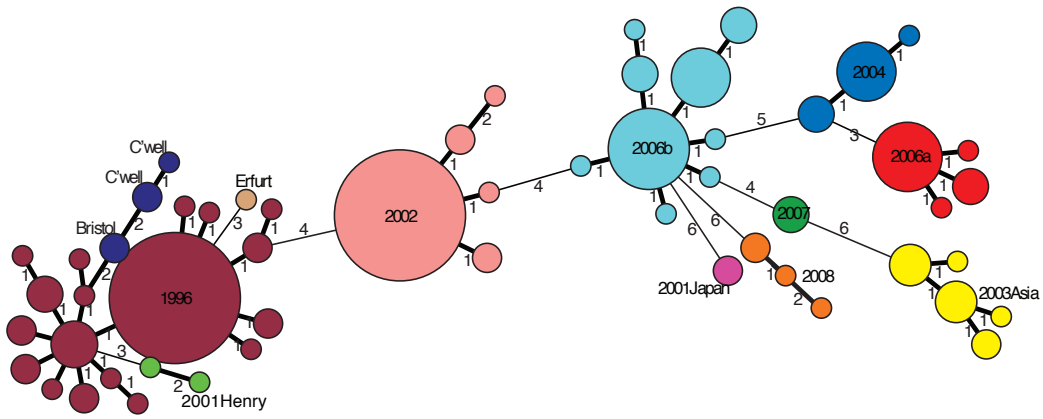


Figure 5.11. Minimum Spanning Tree (MST) of amino acids 6, 9, 294, 296-297-298, 333, 352, 368, 372, 393-394-395, 407, 534. Thick lines represent distances of 1 or 2 amino acids, thin lines 3 or more. An MST connects all samples in such a manner that the summed distance between all samples or branches is minimized. Different colors indicate the different variants. Different variants are separated by at least 2 amino acids.

Identifying future new GII.4 variants

All of the substitutions described above were associated with at least one variant transition; they appear at branches that give rise to new variants (figure 5.7 and figures 5.6, 5.8A and 5.8B). This indicates that these mutations include the molecular determinants of cluster replacement. Previously identified antigenic sites [6,170] did not enable distinction between all the different GII.4 variants (e.g. considering amino acids 296-298 and 393-395, the 2006b and 2007 variants that are clearly phylogenetically distinct, share identical amino acids).

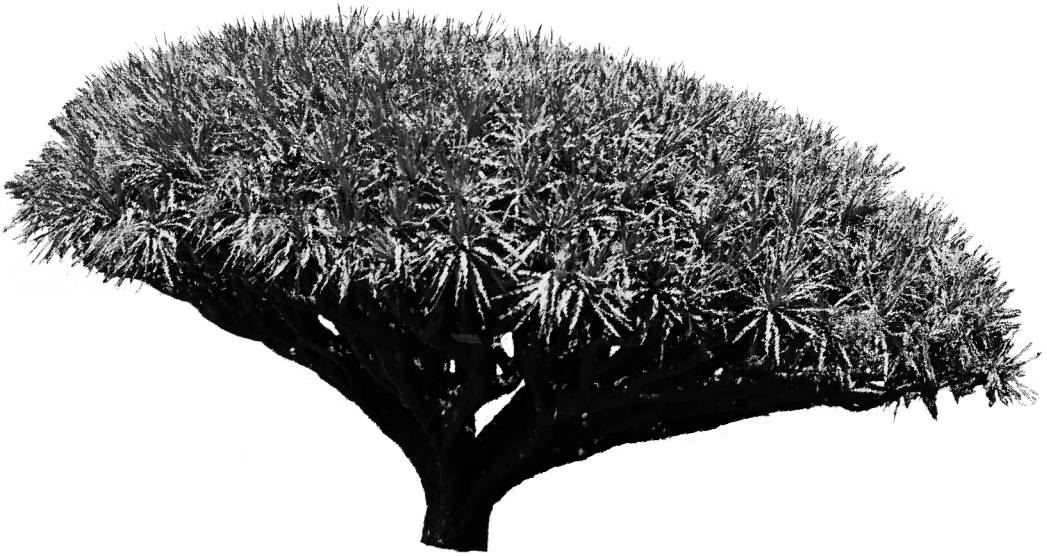
When amino acids 6, 9, 294, 333, 352, 368, 372, 407 and 534, identified here as under positive, directional or co-evolutionary pressure, are added to these six amino acids, all currently identified GII.4 variants (excluding the Bristol and Camberwell strains, that circulated

before the 1996 variant) are separated by at least two amino acid differences (figure 5.11). Thus, specific analysis of these sites will aid early recognition of novel variants in the future. The two most recent distinct GII.4 variants, that have both been detected throughout the world in both 2008 and 2009, albeit at low prevalence, the 2007- and 2008-variants, are identical to the still dominant 2006b variant in amino acids 296-298, that were identified by Allen *et al.*, and 2007 is also identical in site 393-395, whereas the 2008 variant has two substitutions on those sites (2006b: STT, 2008: D/NTA). Thus, considering the sites listed above, the 2008 variant would have the best chance of becoming the next dominant strain, whereas, when considering the full capsid sequence, the 2008 variant is more similar to the 2006b variant than is the 2007 variant.

Using Bayesian phylodynamic techniques we showed that since 2002 the number of GII.4 infections has experienced expansion dynamics. Additionally we further substantiated the evidence for signature sites for variant transition, which may aid in the early recognition of potential new epidemic variants, although we stress that examining pre-defined amino acids does not enable certain identification of GII.4 variants, for which full capsid sequences should be determined. We showed that it is important to select the genomic region to analyze by phylodynamic, coalescent methods with care, and our different datasets illustrated that for the phylodynamic analysis of pathogens undergoing repeated selective bottlenecks a considerable sampling density through time is required.

Acknowledgements

We thank the participants of NoroNet and the FBVE network, for sharing their norovirus sequences, (some of) which were used in these analyses. We especially thank Dr. R.M. Ratcliff (Institute of Medical and Veterinary Science and School of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia), Dr. N. Iritani (Osaka City Institute of Public Health and Environmental Sciences, Osaka, Japan), Drs. M. Jin and Z-J. Duan (Department of Viral Diarrhea, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China), Drs. M. Hoehne and E. Schreier (Robert Koch Institute, Berlin, Germany), Dr. Blenda Bottiger (Statens Serum Institute, Copenhagen, Denmark), Dr. A. Bosch (University of Barcelona, Spain), Dr. J. Buesa (University of Valencia), Dr. A. Sanchez Fauquier (Institute of Health Carlos III, Madrid, Spain), Dr. P. Pothier (University of Dijon, France), Dr. C-H. von Bonsdorff (University of Helsinki, Finland), Dr. G. Reuter (ÁNTSZ Regional Institute of State Public Health Service, Pécs, Hungary), Dr. K. Vainio (Norwegian Institute of Public Health), Dr. K-O. Hedlund (Swedish Institute for Infectious Disease Control, Solna, Sweden), and Dr. M. Poljsak-Prijatelj (University of Ljubljana, Slovenia).



Chapter 6

High Prevalence of Prolonged Norovirus Shedding and Illness among Hospitalized Patients: A Model for In Vivo Molecular Evolution

J. Joukje Siebenga[1,2], Mathias F. C. Beersma[2], Harry Vennema[1], Paula van Biezen[3], Nico J. Hartwig[4] and Marion Koopmans [1,2]

1 National Institute for Public Health and the Environment, Bilthoven
2 Department of Virology,
3 Department of Internal Medicine, and
4 Department of Pediatrics, Erasmus Medical Center, Rotterdam, the Netherlands

Published in The Journal of Infectious Diseases
• http://www.journals.uchicago.edu/toc/jid/current • 2008;198 (7):994–1001 •
DOI: 10.1086/591627

Abstract

During a 2-year survey in an academic hospital, 8 (8.4%) of all norovirus–positive patients showed prolonged norovirus illness and shedding (duration, 21–182 days). All patients had underlying illnesses, resulting in some level of immunodeficiency in 5. Four patients were admitted to the hospital with gastroenteritis, 2 acquired norovirus while hospitalized, and 2 were outpatients. Genotypes GII.4 and GIIb-GII.3 were found. Reinfection occurred in 3 patients. Full capsid sequences were determined from strains detected in sequentially collected stool specimens to study evolution. The greatest number of amino acid mutations in a given patient was 11; they were detected in norovirus isolates recovered over a 119-day period and were mapped to positions at or near putative antigenic sites. In the patient with most severe immune dysfunction, only 5 amino acids mutated over 182 days, suggesting immune-driven selection. The severe impact on patients and hospitals and the potential role of prolonged shedders as a reservoir for viral antigenic variants lead us to stress the importance of confinement of outbreaks of norovirus infection that occur in hospitals.

Introduction

Noroviruses are a genetically and antigenically diverse genus of the Caliciviridae family and are the leading cause of acute gastroenteritis in people of all ages [17,55,74,283]. Episodes last 12–60 h and mainly entail vomiting and diarrhea. Although norovirus illness is usually self-limiting, the burden of disease is considerable, with many people affected and a potentially great impact in healthcare settings [105,129,313]. Between 1994 and 2005, a total of 74% of all reported outbreaks of viral gastroenteritis in the Netherlands were caused by norovirus, and up to 68% of these were caused by GII.4 strains [268,283].

Successive GII.4 variants caused global epidemics, having accumulated mutations in a stepwise manner [27,175,212,265,268,269]. As we described elsewhere [269], the GII.4 variants are thought to result from an immune driven selection process known as epochal evolution. It is unclear where these variants arise. It was previously suggested in a Swedish study that they may evolve in chronically infected patients [210]. This study described a case in which, over the course of 1 year, 11 amino acid mutations accumulated in the capsid protein of a GII.3 strain (ARG320/1999/US-like) that was infecting an immunocompromized patient. The number of amino acid mutations reported by Nilsson *et al*. [210] is similar to the number that distinguish variants from each other (range, 8–25 mutations), and the mutations approximately mapped to the region where most variant distinguishing amino acids are located [210,269]. Only 2 other studies of prolonged norovirus shedding have been reported. One describes a child with T cell deficiency who was infected with a GII.3 strain (also ARG320/1999/US-like) [84]. In virus recovered from this patient over a 6-month period, no amino acid mutations were found in a partial capsid sequence of 277 nucleotides in the 5′ part of ORF2. A recent article described shedding of norovirus for 6–7 weeks in 3 of 71 children who presented to a pediatric clinic because of acute gastroenteritis; these children recovered clinically within a short period [205]. This study presented no information on norovirus genotypes or other genetic data. Little more is known about the prevalence of chronic norovirus infection.

This study was conducted as part of a larger survey of the role of norovirus infections in a tertiary care hospital [13]. We selected patients hospitalized during 2005–2006 who tested positive for norovirus for at least 3 weeks. We identified prolonged shedding and illness in 11 (8.4%) of all norovirus positive patients. GII.4 and GIIb–GII.3 recombinant strains were found. We determined changes in the gene encoding the norovirus capsid during the course of illness for 8 patients. The rate of mutation accumulation was lowest among norovirus recovered from patients who had the most severe immune impairment.

Patients, materials, and methods

Study population

Patients were retrospectively selected during a review of laboratory records from 2005–2006 at Erasmus Medical Center, an 1100-bed tertiary care hospital with 30,000 patients annually. Norovirus diagnostic tests had been requested for 899 samples from 502 patients admitted during 2005 and for 1084 samples from 571 patients admitted during 2006. Routine polymerase chain reaction (PCR) analysis was used for detection of genogroup I and II noroviruses; 38 patients in 2005 and 93 in 2006 had positive PCR results. Patients were included in this study if fecal samples recovered during a period of ≥3 weeks tested positive for norovirus and if ≥2 fecal samples were available. To compare the mutation accumulation rates among norovirus from patients with underlying illnesses with those among norovirus from otherwise healthy patients, we analyzed a series of specimens from a female child (age, 4 months) who had a 2-day episode of norovirus infection but no underlying illness.

RNA isolation and sequencing

We investigated stool specimens that were originally collected for diagnostic analysis, which included testing for adenovirus, rotavirus, astrovirus, norovirus, and enterovirus. Only a single patient (patient 4) tested positive for adenovirus, which was detected in a single sample. Unused portions of the specimens were subsequently stored at -80°C.

Fresh RNA was extracted from stool specimens in accordance with the methods of Svraka *et al*. [283]. Sequencing was done as previously described [49,269,322]. The MagNAPureLC total nucleic acid isolation kit (Roche Diagnostics GmbH) was used for extraction, in accordance with the recommendations of the manufacturer. Overlapping fragments of viral RNA were then reverse transcribed using AMV-RT (Invitrogen), yielding cDNA that was amplified and subsequently sequenced using the ABIPrism BigDye Terminator v3.0 Ready Reaction Cycle kit. The same primers were used for amplification and sequencing.

Data processing

DNA sequences were processed, aligned, and analyzed using Bionumerics software (Applied Maths BVBA [Sint-Martens-Latem]). Additional analyses used BioEdit Sequence Alignment Editor 7.0.1 (Isis Pharmaceuticals) and DNASP 4.10.

Nucleotide sequence accession numbers

Capsid sequences of the first and last norovirus isolate recovered from each patient were submitted to the DNA Database of Japan and assigned accession numbers AB385626 through AB385643.

Results

Characteristics of patients

For 11 (8.4%) of all norovirus-positive patients, stool samples recovered over a period of ≥3 weeks tested positive for norovirus. Three patients were excluded from detailed analyses because samples were no longer available or because samples failed to yield positive PCR results, leaving 8 patients for the study. Samples were obtained from 2 patients in 2005, and samples from 9 patients were collected in 2006. Demographic and clinical characteristics of the 8 study patients are shown in figure 6.1 and table 6.1. Five patients were aged <3 years, and 3 were aged 37, 67, and 69 years. All patients had underlying conditions of varying nature. Patients 1–4 and 7 were immunocompromized, of whom 3 patients had leukopenia, and 1 had severe lymphocyte dysfunction after allogenic stem cell transplantation (table 6.1). The 3 non-immunocompromized patients did not receive immunosuppressive medication.

Characteristics of stool samples and gastroenteritis

Samples were originally obtained from the patients for diagnostic purposes. Figure 6.1 shows the times of hospitalization and sampling and the course of gastroenteritis. Patients 2 and 4–6 were admitted to the hospital with gastroenteritis symptoms, patients 1 and 8 first showed gastroenteritis symptoms 6 and 4 days after admission, and patients 3 and 7 were outpatients (table 6.1).

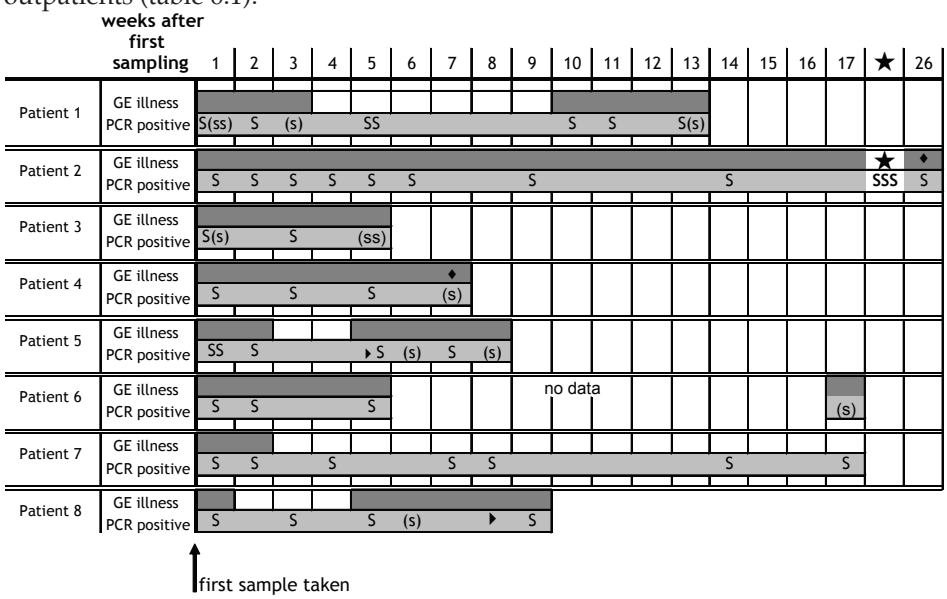


Figure 6.1. Course of gastroenteritis (GE) for study patients. PCR, polymerase chain reaction; S, available sample; (s), original norovirus-positive sample no longer available; diamond, death of patient; ★, break in time scale; ▸, another norovirus genotype detected.

Table 6.1. Demographic and clinical characteristics of patients from whom norovirus was recovered, 2005–2006.

Patient	Gender	Age	Underlying Illness (and Treatment)	Influence of Illness on Immune System	Leucocytes x 10^9/ liter	Lympho-cytes %	Immune suppressive therapy	Series length (days), # samples	Norovirus excretion (weeks)		GE symptoms	Onset of GE relative to hosp	Days after hosp 1st sample taken
									GIIb-GII.3	GII.4			
1	F	4m	Acute myeloid leukemia	Aplasia during chemotherapy	1.4 (± 0.21)	51-94	Chemotherapy	88, 7	13		Intermittent episodes, initial vomiting, diarrhea after 2 days	6 days after	6
2	M	67y	Churg Strauss Syndrome	Immunoglobulin deficiency (IgA <0.01 g/l; IgM < 0.3 g/l)	normal	normal	Dexamethason 3ddx1mg; mycophenolate; mofetil 2dd 750mg	182, 11		26	Diarrhea, no vomiting, at least throughout recorded time. Patient died.	Before	3
3	M	69y	T-cell non Hodgkin Lymphoma	Severe T cell depletion	2.3 - 5.0	3-5	Prednison 1dd 50mg; alemtuzumab	21, 2		3	Diarrhea, no vomiting, at least throughout recorded time	NA	NA
4	M	2y5m	Wishet Aldrich syndrome; Allogenic stemcell transplantation	Dysfunctional CD4/ CD8 lymphocytes	6.1 - 10.5 (normal)	20-48 (normal)	None	33, 3		5	Diarrhea, no vomiting, at least throughout recorded time. Patient died.	Before (adenovirus)	3
5	M	9.5m	Down Syndrome; Morbus Hirschprung; congenital defects	None	normal	normal	None	51, 5	1	2	Intermittent episodes of vomiting and diarrhea	Before	2
6	F	3.5m	Dilated cardiomyopathy	None	normal	normal	None	33, 3		5	Diarrhea, no vomiting, at least throughout recorded time	Before	17
7	M	37y	Seminal testis tumor; idiophathic trombo-cytopenic purpura (ITP)	CD4 / CD8 Lymphopenia	4.2 (± 2.7)	9-17	Prednisone 1dd 40mg	119, 8		17	Two days vomiting, followed by diarrhea, that cleared after two weeks	NA	NA
8	M	3m	Very long chain Acyl-CoA dehydrogenase deficiency; cardiomyopathy	None	normal	normal	none	65, 4	4	1	Intermittent episodes of vomiting and diarrhea, clearing and reappearing, intensified after 65 days.	After (parecho virus)	4

The shortest sample series involved patient 3, for whom 2 samples remained available for the current analysis. Five samples had originally been recovered over 32 days from this patient, and all tested positive for norovirus. Three specimens each were available from patients 4 and 6. Specimens were missing for both patients, and gastroenteritis symptoms lasted longer than the interval marked by the 3 available samples. One sample from patient 4 tested positive for adenovirus; this patient died as a result of poor health complicated by gastroenteritis. Patient 2 had the longest series, consisting of 11 samples recovered over 182 days. Patient 2 had severe gastroenteritis throughout this period and died possibly as a result of norovirus infection. Patient 8 was co-infected with parechovirus (data not shown) and had no symptoms after 2 weeks. Gastroenteritis complaints reappeared after 4 weeks, after which the patient had diarrhea and vomiting for a number of weeks; hospital records indicate "an increase of symptoms, initially with vomiting" 65 days after the first sample was taken.

Collection of follow-up samples from the patients after the causative agent of gastroenteritis complaints had been established was generally poor, especially after hospital discharge. The exception was patient 7, who did not report gastroenteritis symptoms but was still PCR positive while the current study was being performed. At our request, 2 additional samples were obtained from him during outpatient clinic visits, lengthening the series duration from 56 to 119 days.

No spatial or temporal relation could be identified among the gastroenteritis episodes in the patients described here. Those who were admitted to the same ward were hospitalized at different time points (at least 5 months apart) and had different viral strains.

Findings of sequence analyses

Patients 2–4, 6, and 7 were infected by strains typed as GII.4 throughout their infection. Patient 1 had a strain with polymerase type GIIb during the course of infection, and patients 5 and 8 were initially infected by GIIb strains before being infected by GII.4 strains. For all GIIb strains, the capsid genotypes belonged to GII.3 (ARG320/1999/US-like). The observation that patient 5 had 2 separate episodes of norovirus illness and shedding, caused by 2 different norovirus strains, excludes this patient from our definition of prolonged shedding, leaving 7 patients who shed the same virus for >3 weeks and showed prolonged illness. Patient 8 showed prolonged illness during his GIIb episode, but this was not confirmed for the GII.4 episode.

GII.4 strains belonged to variants 2004 (recovered from patients 2 and 5), 2006a (recovered from patients 4, 6, 7 and 8), and 2006b (recovered from patient 3). Figures 6.2A and 6.2B

show the neighbor-joining trees of the GII.4 and GII.3 strains. Reference strains of the GII.4 variants have been included. In sequential strains from all 8 patients, mutations were found in the capsid (figure 6.2). The sequence patterns for a number of nucleotides showed the presence of ambiguous sites before complete substitution (data not shown). Fixation of these mutations was seen in the 5 long series but could not be confirmed for the short series because of low sample numbers. The virus isolated from stool specimens from patient 4, of which only 3 were available, had no nucleotide mutations after 14 days but had 15, resulting in 8 amino acid changes in the capsid, after 33 days. No difference was found in the partial polymerase sequences used for genotyping the strain.

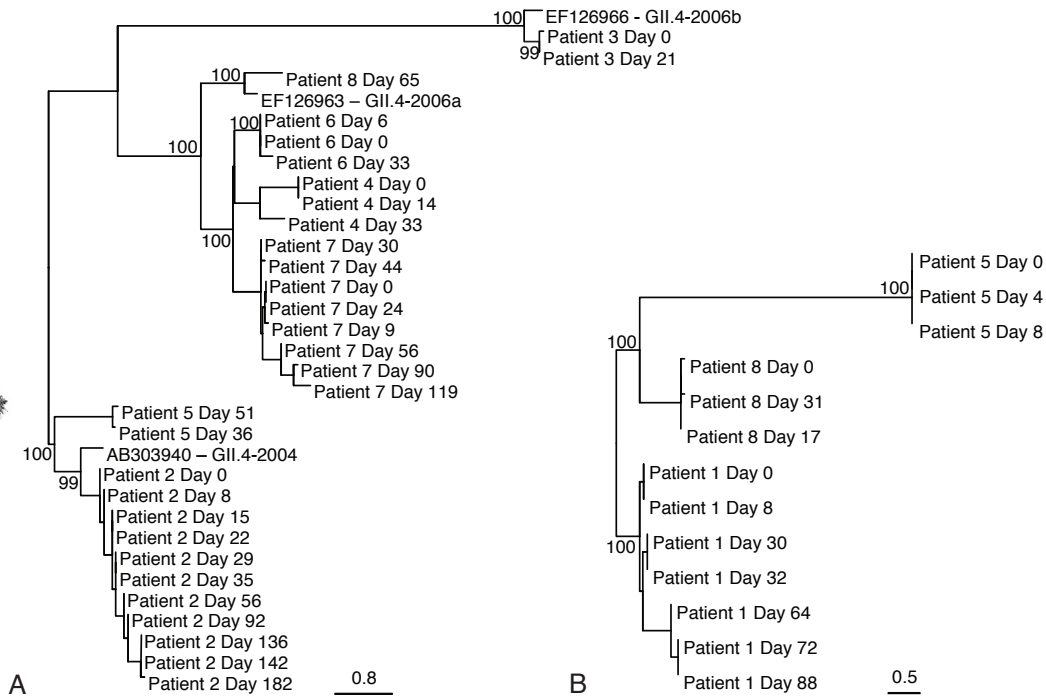


Figure 6.2. Neighbor-joining trees of complete amino acid sequences. Trees were obtained by means of the Jukes and Cantor correction, for GII.4 (A) and GII.3 (B) norovirus strains. Bootstrap values are specified as a percentage of 1000 iterations.

Figure 6.3 shows rates of mutation accumulation and trend lines for patients with the longest series (i.e., patients 1, 2, and 7). We also determined the mutation-fixation rate of norovirus shed by a child who had gastroenteritis symptoms for 2 days but no underlying illnesses (data are not shown in detail; 9 samples that were PCR positive for GIIb-GII.3 recombinants were recovered over 34 days, with a total accumulation of 6 nucleotide and 4 amino acid mutations in the capsid). Norovirus from this otherwise healthy child had a fixation rate of 0.13 amino acids/day. Viruses from patients 1 and 7, with moderate cellular

immune deficiency and normal levels of serum immunoglobulins, had a fixation rate of 0.07 amino acid mutations/day; norovirus from patient 2 had a fixation rate of 0.03 amino acid mutations/day. It is interesting to note that patient 2 had severe humoral immunodeficiency and no detectable serum levels of IgA and IgM. When the first and last samples of each series were compared, the ratio of nonsynonymous mutations in norovirus from these samples was sometimes far greater than 1 (data not shown).

In the GII.4 strains, 15 of 34 nonsynonymous nucleotide mutations were located in the P1 domain, affecting 13 amino acids (244, 246, 248, 258, and 268 in the N-terminus and 407, 409, 411, 413, 479, 486, 497, and 519 in the C-terminus); 14 were located in P2, affecting 9 amino acids (295, 317, 327, 340, 375, 378, 389, 393, and 397); 1 was in the N domain (amino acid 33); and 4 were in the S domain (amino acids 93, 98, 130, and 174). Amino acids 174, 244, 340, 378, 389, 393, 397, 407, 413, and 497 were previously identified as informative sites in GII.4-variant transitions, and amino acids 340 and 407 changed with every variant transition [269]. Only in the P1 and P2 domains were identical amino acid mutations observed in different patients. Patients 4 and 7 both had amino acid substitutions at positions 248 and 258. In patients 2 and 6, G295 mutated to N295. In patients 2, 3, and 5, R340 mutated; in patients 2 and 5, it mutated to G340 in the last sample, turning RRD into RGD. In patient 3, E340 changed into G340, turning KED into KGD. Similar to the RGD motif, KGD involved a cell-recognition and -binding motif [47].
Of the GII.3 strains, fewer samples were available and fewer mutations were seen, compared with the GII.4 strains. Strikingly, 5 of 8 mutations occurred in the C terminal part of the P1 domain, and only 2 mutations were located in the P2 domain.

Structural implications of amino acid polymorphisms
Figure 6.4 shows the affected amino acids in the previously modeled GII.4 capsid for each patient separately [269]. Amino acid 340, that changed in 3 patients, establishing the presence of an extra RGD or RGD-like motif, is positioned in a loop on top of the P2 domain. A number of mutating amino acids clustered on the surface of the protein. Amino acids 317, 327, 407, 409, 411, and 413 were located at the side of the protein of intradimeric interaction in an upward oriented fold of the P1 domain next to the P2 domain. The residues in 1 protein were in steric proximity to the residues in the neighboring protein of another dimer. Residues 295, 375, and 378 were close together in loops on top of P2, as were 393 and 397 with 389. Amino acids 244, 246, and 248 were all in the same β-sheet in P1. However, amino acids 93, 98, 268, 497, and 519 were located inside the protein at various locations, and 98, 497, and 519 were situated in β-sheets. The substitutions that took place here were not drastic: A to S, S to G, T to I, V to I, and S to A.

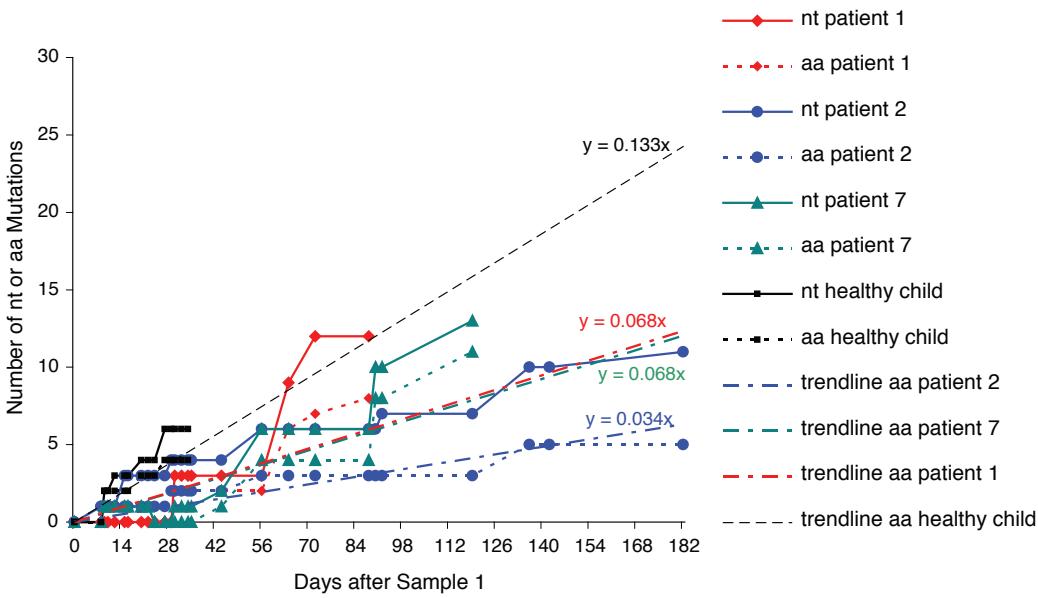


Figure 6.3. Rate of accumulation of mutations. Amino acid (aa) and nucleotide (nt) mutations among norovirus isolates recovered from patients with the 3 longest sample series and 1 otherwise healthy child, shown with a time axis. Day 0 corresponds to the first day of sampling. Trend lines were added for the aa mutation-accumulation rates.

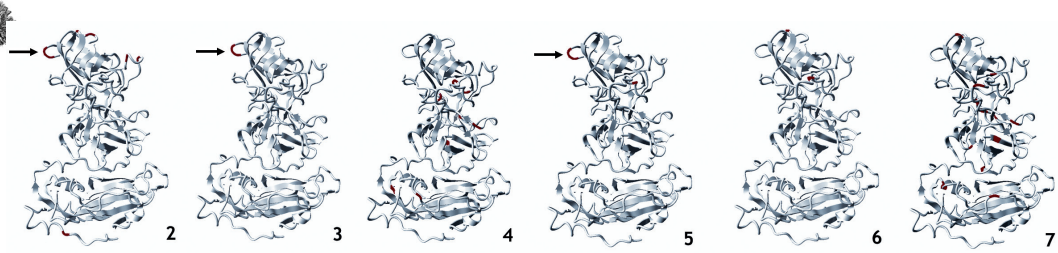


Figure 6.4. Locations of mutations in the capsid protein. Mutating amino acids (red) in the GII.4 capsid of norovirus recovered from patients 2–7. The arrows indicate the inserted RGD motif (amino acid 340) in the loop on top of the protein.

Discussion

We showed that significantly prolonged periods of gastrointestinal illness due to norovirus infection, combined with shedding over unusually long durations, are more common than previously recognized. Eleven patients (8.4% of the total number of norovirus positive patients in the hospital) were identified in a 2 year period at an 1100-bed hospital. For 8, a sufficient quantity of sample material was available for analysis in the current study. Of the 11 patients initially included, 2 had samples obtained in 2005, and 9 had samples

obtained in 2006. From 2006 onward, infection prevention protocols were intensified, which led to collection of follow-up samples from immunocompromized patients after the first diagnosis, resulting in the availability of more series. The greater number of patients during 2006 can also be explained by the high number of off-season outbreaks and the subsequent winter epidemic, caused by the emergence of 2 new variants of GII.4 in the spring of 2006 [269]and by an increase of GIIb norovirus in the pediatric wards in 2006 (unpublished data).

All 7 patients identified in this study as prolonged shedders of norovirus had underlying illnesses. Five patients (patients 1–4 and 7) had impaired immunity, and 5 had prolonged gastrointestinal illness (patient 7 did not have prolonged gastrointestinal illness, and the course of illness for patient 8 was not clear from the laboratory records). Symptoms of prolonged illness mainly entailed diarrhea; no persistent vomiting was recorded. Patient 8 had co-infection with parechovirus, and patient 4 tested positive for adenovirus once. No other common gastroenteritis viruses were found. Thus, we cannot exclude that some of the gastroenteritis symptoms were caused by factors other than norovirus, such as medication or other infections.

Strains belonging to 2 different norovirus genotypes were found in ensuing samples from 3 of the initial 11 patients. These patients were probably reinfected with new strains, but it cannot be ruled out that the mixed strains were present at the beginning of their infection. The GII.4–2006a strain in patient 4 showed no mutations in the first 2 weeks but showed 15 nucleotide mutations between days 14 and 33. Although the strain still had the characteristic genetic make-up of a 2006a variant, the occurrence of such a large number of mutations in such a short period probably reflects a new infection with a different strain. Thus, 4 of the initial 11 patients had reinfections with 2 distinct viral strains, all likely nosocomial.

Persistent infections with common respiratory viruses in immunocompromized patients have been described for parainfluenza 3, influenza, respiratory syncytial virus, human rhinovirus, and others [123,132,292,340]. The clinical impact of such viruses in this risk group might be severe and includes an increased risk of influenza viruses resistant to treatment [123]. Similarly, chronic infections were reported for rotavirus (duration of infection, >450 days) [333] and astrovirus (duration, >225 days) [46,333]. It was described previously that patients who were already sick got sicker because of norovirus infection, even those with self-limiting infections [196]. Comparable to these observations, we found that norovirus causes severe disease in hospitalized patients at high risk for acquisition of nosocomial infection.

Additionally, we sought to address the possibility that chronic shedders were sources of

new norovirus variants. The accumulation and fixation of mutations in 5 patients indicates that, although norovirus could not be cleared, an immune response imposed pressure on the virus, causing it to modify its capsid protein to evade immune recognition. The rate of mutation fixation in the viral capsids seemed linked to the level of impairment of both cellular and humoral immunity. Norovirus from patient 2, who had no humoral immunity and had cellular immunity that was weakened by medication, showed the lowest rate (0.03 amino acids/day). Patients 1 and 7 had mildly impaired immunity and had virus that accumulated 0.07 amino acid mutations/day, and norovirus from the otherwise healthy child had a fixation rate of 0.13 amino acid mutations/day. Properly functioning immunity thus seemed to induce a higher mutation-fixation rate. If the data from the article by Nilsson *et al*. [210] that described the Swedish chronic shedder are plotted similarly (data not shown), 0.04 amino acid mutations accumulated daily, fitting with our data. This patient received immunosuppressive medication, resulting in low total lymphocyte counts, low CD4+ cell counts, but normal immunoglobulin concentrations [210]. Interestingly, patient 2, who had virus with the lowest fixation rate, had no detectable IgA (concentration, <0.01 g/L), which is normally excreted in the intestine. The role assumed for immune pressure in directing the mutational pattern in these viruses is confirmed by our observation of high ratios of nonsynonymous to synonymous mutations found among viruses from all patients.

In our previous study, informative sites were determined by comparing subsequent epidemic variants of GII.4 strains. Ten of the changing amino acids recognized in present study were informative sites, and 4 were direct neighbors of informative sites [269]. Of these, amino acids 340 and 407 changed with every variant transition. Amino acid 340 is part of a putative RGD motif that appears and disappears from the sequence. Although the biological relevance of these individual sites remains to be determined, it is striking that 340 changed in 3 patients and 407 changed in 1 patient investigated here. This strengthens the idea that these amino acids are involved in immune evasion, as was also suggested by Lindesmith *et al*. [170]. When comparing the GII.3 amino acid substitutions found in this study with the substitutions reported by Nilsson *et al*. [210], the following 3 matches were detected: 310, 404, and 415. This consistence suggests a role in immune evasion for these amino acids.

The norovirus genotypes that were found - GIIb-GII.3 recombinant and GII.4 (variants 2004, 2006a and 2006b) - were also found in 2005 and 2006 during population surveillance that is ongoing in the Netherlands (2.8% and 60%, respectively, of all outbreak-associated norovirus). It is remarkable that, both in this study and in the previously described cases of chronic norovirus infection, strains with GII.3 ARG320/1999/US-type capsids were found so predominantly. A large proportion of the patients were very young; at the onset of illness,

5 of 8 were <3 years of age. The 3 patients who were initially selected but later excluded from the study were <1 year of age. All patients with GIIb-GII.3 strains were young infants. This discrepancy in genotype distribution between age groups might be explained by the prevalence of strains going around in pediatric wards in the hospital, where outbreaks of both GII.4 norovirus infection and GIIb-GII.3 norovirus infection were detected. Alternatively, GII.3 strains might be better equipped to establish long-term infections or infections in young children. This illustrates our limited understanding of norovirus epidemiology.

Although the available evidence provides no definite proof, it is conceivable that similar chronic shedders are the reservoir in which antigenic variants of norovirus arise; GII.4 variants accumulated 8–25 amino acid mutations in the capsid over periods of ≥ 2 years [269]. The numbers of mutations found here are similar to that number. However, even though we identified a surprisingly high number of patients with prolonged shedding durations in this study, this finding bears no relation to the much higher frequency of infection in the population, where shedding of virus may last up to 3 or 4 weeks after clearance of symptoms. We have shown here that norovirus in healthy individuals accumulates mutations more rapidly than in immunocompromized patients. Therefore, healthy persons may be of more importance than chronic shedders. In addition, although chronic shedders remain ill (with gastroenteritis symptoms) and PCR positive for norovirus, the data do not enable us to conclude that the virus shed by the patients is infectious. Still, considering that a single specific amino acid mutation has been shown to be sufficient to confer resistance to immunity in murine norovirus [172]or to increase the virulence of West Nile virus [21], it is clear that the changes observed in this study require further study.

Because of the frequency at which these prolonged infections occur, the possibly severe impacts they have both on patients and hospital alike, and the possibility that these patients are reservoirs in which antigenic variants evolve, we cannot overstress the importance of adequately containment of outbreaks of norovirus infection in health-care settings.

Chapter 7

Chronic Shedders as Reservoir for Nosocomial Transmission of Norovirus

Faizel H.A. Sukhrie[1,2], J. Joukje Siebenga [1,2], Matthias F.C. Beersma[2], and Marion Koopmans [1,2]

1 Laboratory for Infectious Diseases and Perinatal Screening, National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, the Netherlands
2 Erasmus Medical Center Department of Virology, Rotterdam, the Netherlands

Published in adaped form in: The Journal of Clinical Microbiology
• http://jcm.asm.org/ • DOI: 10.1128/JCM.01308-10

Abstract

Norovirus infection in immunocompromized patients may lead to prolonged norovirus shedding. Here, we provide combined epidemiological and molecular evidence that proves norovirus transmission by a long term chronic shedder over 17 days after the first diagnosis. Chronically infected patients should be considered as a significant source of infection for norovirus outbreaks within healthcare institutions.

Introduction

Norovirus is a leading cause of acute gastro-enteritis affecting people of all age groups [13,71]. Outbreaks with norovirus occur often and worldwide [270]. In the Netherlands, large numbers of outbreaks are reported each winter, particularly from health care institutions siebenga[268]. In a systematic evaluation of newly diagnosed patients with norovirus in a large tertiary care hospital, we found that nosocomial norovirus transmission is common, and may lead to chronic infection, disease and shedding in at least 6% of patients [13]. The case histories of the chronic patients showed various underlying illnesses resulting in impaired immunity followed by prolonged norovirus shedding, in some cases for periods longer than 1 year [266]. The question arose whether these chronic shedders were possible sources for nosocomial infections within the hospital setting, also after they had been infected for a number of weeks. Because norovirus can not be cultured in vitro [69], it has remained impossible to assess whether the viruses shed by such patients are still infectious. The finding that noroviruses evolved within chronic patients suggested that detailed molecular virological data in combination with epidemiological data could be used to track possible routes of norovirus transmission within the hospital [337,338]. This study of the role of chronically infected patients in nosocomial spread of norovirus was performed as part of a larger study, which will be described elsewhere.

Methods

Epidemiological records for all norovirus cases reported in the period 2002-2007 were retrieved from the ErasmusMC data bank, including admission dates, sampling dates and departments. Fecal samples associated with these cases had been stored at -80ºC Patient samples were sequenced as previously described [337]. Briefly, the P2 domain of the ORF2 with a length of approximately 700 nucleotides was sequenced in both directions using the ABI Prism BigDye Terminator v3.0 ready reaction cycle sequencing kit. Strain sequences from patients with chronic norovirus had been described previously [266]. Norovirus positive patients hospitalized in the same period (defined as six months before to six months after

the first sampling of all chronic shedders) were selected, and their stool samples used for analyses. This selection was made to represent background diversity of norovirus strains circulating within the hospital. To identify patients who were nosocomially infected by chronic shedders [266], strain sequences obtained from regular hospitalized patients with 100 % identity to sequences previously obtained from chronic shedders over a minimum fragment length 600 nucleotides were identified. The sequences were subsequently analyzed using the BEAST program package (BEAST version 1.4.8) that employs a Bayesian Markov Chain Monte Carlo (MCMC) approach [63], and allows for the inclusion of a time factor in the phylogenetic analysis. The resulting samples of trees were summarized using Tree-Annotator (distributed in the BEAST program package), into Maximum Clade Credibility trees, which are scaled to a time-scale shown on the X-axis. The tips are aligned to detection dates of the samples, whereas the node heights represent genetic distances.

Results

During the study period, we found three molecular clusters containing sequences of patients who had been recognized as chronic shedders patients 4, 6 and 8 (numbering corresponds to numbering used in [266]), and other hospitalized patients; two with strains of genotype GII.4-2006a and one of GII.b/GII.3 strains. Chronic patient 6 was admitted to the hospital multiple times, while chronic patients 4 and 8 stayed in the hospital, mainly in the same location during their norovirus infections. They were sampled and tested for norovirus a number of times during their admissions or visits, and sequences identical to theirs were detected among other admitted patients. Based on molecular information combined with demographic data the transmission links were assumed from the chronic patient to the hospitalized patients. An overview of the corresponding clusters is shown in table 1.

Table 1. Transmission of norovirus from three chronically infected patients.

	Cluster size [a]	Available samples	Symptoms and diagnosis of the chronic patients during and after admission	Estimated transmission delay from chronic shedders after their first diagnosis [b]	Genotype
1	5	Chr. Pat. 4/ 3 NT Pat.	Chr. Pat. 4 admitted with chronic diarrhea, tested pos. for norovirus after 3 days	After 4 days	GII.4-2006a
2	3	Chr. Pat. 6/ 1 NT Pat.	Chr. Pat. 6 admitted with infant regurgitation, pos. for norovirus after 13 days	After 4 daysΩ	GII.4-2006a
3	9	Chr. Pat. 8/ 6 NT Pat.	Chr. Pat. 8 admitted with diarrhea and vomiting, tested pos. for norovirus after 2 days	After 2, 3, 6, 7 and 17 days	GIIb/ GII.3

[a] Each cluster contains one chronic patient (indicated as Chr. Pat. 4, 6 and 8) and other hospitalized patients indicated as NT (Noro Transmission) patients.

[b] Transmission delay was defined by the number of days between the date of diagnosis of the chronic patients and the onset of illness of the other patients in the cluster.

The transmissions between chronic patients 4 and 6, and the other patients in the two GII.4 clusters occurred shortly after the chronic patients were first diagnosed. In the GIIb/GII.3 cluster, transmission was detected both shortly after initial diagnosis of the chronic patient (involving at least five other patients), and also after a longer interval (involving one patient; NT23) (figure 7.1).

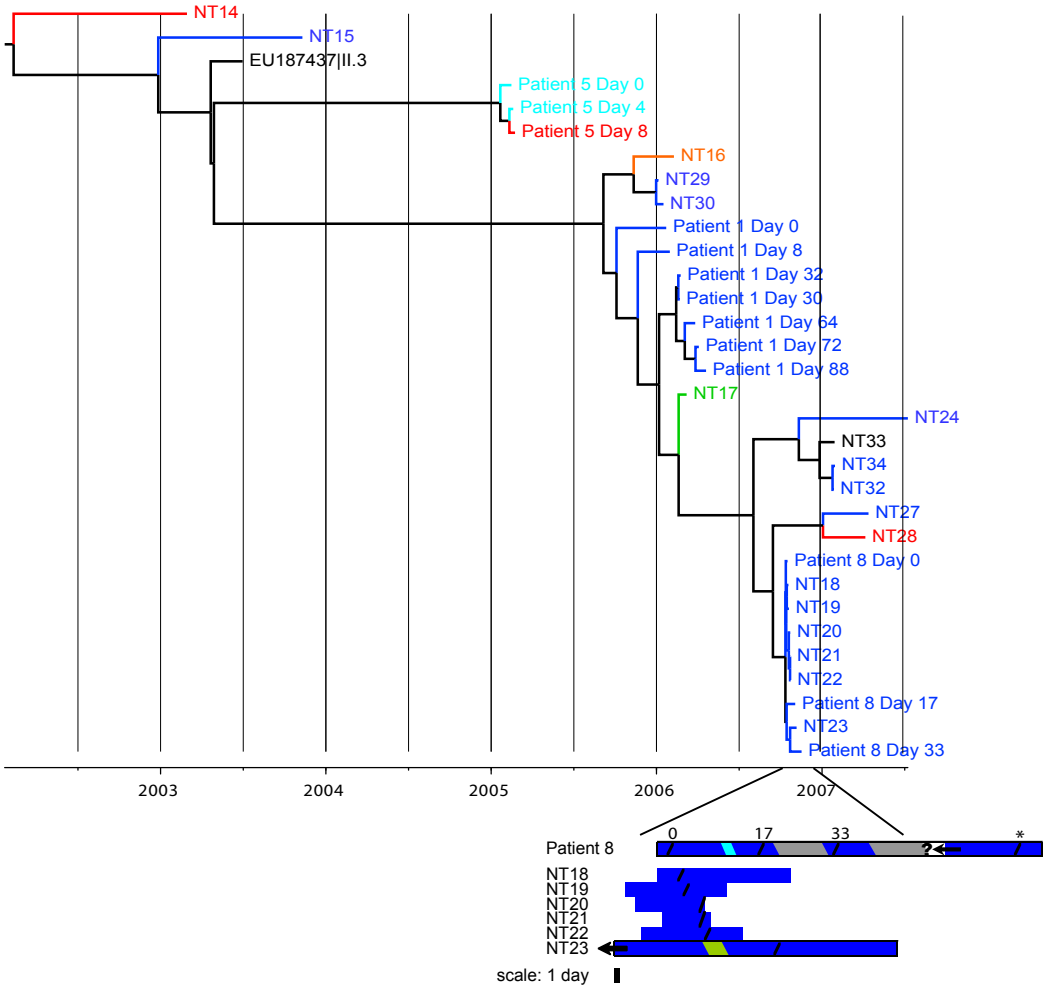


Figure 7.1. Maximum Clade Credibility tree of sequences of the P2 domain of GIIb/GII.3 sequences (top), and transmission scheme of cluster 3 (bottom). Patients identified in this study are numbered with NT (Noro Transmission) numbers; the chronically infected patients included in this tree are patients 1 and 8. Below the tree detailed information describing the admission times and locations of the patients involved in the outbreak with patient 8 are shown compared to patient 8. The colors in the figure and in the tree indicate departments in the hospital; grey indicates that the patient was discharged from the hospital. The black italic lines show the sampling dates, while the arrows indicate that no admission or discharge information was available.

The patients who were infected during the first week of this hospital outbreak, all shared identical sequences in the genomic region analyzed. The sequence of the norovirus strain detected on day 17 in chronic patient 8 showed one nucleotide difference compared to the strain detected on day 0, and was identical to that of patient NT23, whose onset of disease occurred 20 days after the onset of disease of chronic patient 8. This strongly indicates that this patient was infected by chronic patient 8, at least 17 days after the first diagnosis.

Discussion

In this study we show plausible evidence for transmission of norovirus infection from a chronically infected patient to another hospitalized patient, at least 17 days after the initial diagnosis of the chronic patient, who had shown symptoms of norovirus illness for several days prior to diagnosis. Two other clusters that included chronic shedders remained unresolved with respect to the direction of transmission, because all cases were diagnosed within a few days. However, patient 4 already had chronic diarrhea prior to hospitalization, which was resistant to treatment and coincided with chronic shedding of norovirus. Therefore, it is plausible that this is a second example of transmission from a chronic shedder.

Evidence was most convincing for patient number 8, for whom the second sample (taken at day 17) showed a unique mutation that was identified in another patient hospitalized in the same ward. As this sequence was unique in the entire dataset, a link with the chronic shedder is highly likely. However, sources of norovirus in the hospital may vary from patients to staff, contaminated environments and food items and despite extensive outbreak investigations the exact modes of transmission often remain unclear. This study, however, shows that chronically infected patients may contribute to the spread of norovirus in hospitals. To our knowledge this is the first study that provides evidence for this hypothesis. This points at an important aspect of infection control: contrary to earlier beliefs, patients who had norovirus illness may shed the viruses for weeks, and recent data suggest that chronic shedding is relatively common in persons with impaired immune functions who contract the illness. Given the high incidence of norovirus infections and the increasing size of the population that is immunocompromized, this problem is likely to increase in the years to come. Therefore, as part of infection control policy in the hospital, the possible contribution of such patients to nosocomial spread should be considered.

Chapter 8

Unexplained Gastroenteritis Illness and Deaths in the Elderly Are Associated with Norovirus Epidemics

Liselotte van Asten[1], J. Joukje Siebenga[1,4], Cees van den Wijngaard[1], Robert Verheij[2], Hans van Vliet[1], Mirjam Kretzschmar[1], Hendriek Boshuizen[3], Wilfrid van Pelt[1], Marion Koopmans[1, 4]

1 Centre for Infectious Disease Control Netherlands,
National Institute for Public Health and the Environment, Bilthoven, the Netherlands
2 Netherlands Institute of Health Services Research (NIVEL), Utrecht, the Netherlands
3 National Institute for Public Health and the Environment, Bilthoven, the Netherlands
Expertise Centre for Methodology and Information Services
4 Erasmus Medical Center, Rotterdam, the Netherlands

Submitted for publication

Abstract

In recent years new variant strains of norovirus emerged worldwide, that seem to evolve by mutation much like influenza viruses, and which have been associated with many more annual outbreaks than before. However, the impact of such increased norovirus activity on morbidity and mortality is not clear.

We studied trends of unexplained gastroenteritis (because norovirus infection is rarely specifically registered) in medical registrations (i.e. GP-visits, hospitalizations and deaths) and their association with known trends in norovirus outbreaks. Using weekly counts in the elderly (aged 65+) from 1999-2006, we applied Poisson regression analyses adjusted for other pathogens and seasonal trends (linear, sine and cosine terms).

In the elderly, norovirus outbreak activity was significantly associated with unexplained gastroenteritis (uGE) in all three medical registrations: an estimated 25.7 (16.5-33.5) uGE_GP-visits, 2.15 (95%CI: 1.57-2.74) uGE_hospitalizations, and 0.14 (95%CI: 0.08-0.21) uGE_deaths were associated with each norovirus outbreak. For the heaviest norovirus season (2004/2005), these models attributed up to 3804 uGE_GP consultations, 318 uGE_hospitalizations, and 21 uGE_deaths yearly to norovirus outbreaks on a total elderly population of 2.3 million.

In the elderly, extreme norovirus outbreak activity is associated with increases of unexplained gastroenteritis morbidity and even deaths, and thus should not be regarded as an infection with trivial health risks.

Introduction

Diarrheal disease has been linked to increased mortality among the elderly especially in winter [167]. This seasonal rise suggests that a putative cause may be infections with norovirus, which typically peak in the winter, in regions of temperate climate [204]. Norovirus infection, or 'winter vomiting disease', is the most common cause of community-acquired sporadic cases and of outbreaks of acute gastroenteritis, with vomiting and diarrhea being the characteristic symptoms [53,329].

Noroviruses are highly infectious and notorious for causing outbreaks in semi-closed communities such as nursing homes [268]. Symptomatic infection is generally regarded as a mild and trivial disease, perhaps needing medical attention but causing no serious long-term effects. However, the elderly may be particularly vulnerable to gastroenteritis complications and even death [87,176,196]. The extent of the morbidity and mortality risk posed by norovirus infections is unclear and probably underestimated in this age group [278]. A study in England and Wales found 1 death for every 50 norovirus outbreaks in 1990-2000 [176]. More recently, since the appearance of new norovirus variants [268], smaller studies in the U.S.A. and Japan showed significantly higher mortality [220,335], indicating that norovirus outbreaks are potentially life-threatening in this age group. At the population level, our data from the Netherlands and data from the UK also suggest that there is an association between norovirus outbreaks and increased morbidity and mortality in the elderly [110,306].

Norovirus epidemiology changed remarkably in recent years. In 2002, a genetic shift in circulating strains was observed with the global emergence of the GII.4-2002 norovirus variant, which coincided with an unusually high number of outbreaks worldwide in winter 2002/2003, preceded by off-seasonal norovirus activity in spring 2002 [1,175]. Reported outbreaks more than doubled in most countries participating in a European network that monitors norovirus epidemiology and strain diversity [146]. In 2004, another genetic variant (GII.4-2004 strain) emerged, followed by winters with high norovirus activity, and in 2006 two novel variants emerged (GII.4-2006a and -2006b), which again coincided with an exceptionally high number of outbreaks starting early spring 2006 [27,265,268] .

The information available from our ongoing norovirus outbreak surveillance, which included three recent winters with heavy norovirus activity and information on genotypes, permitted unique exploration of norovirus outbreak activity and its association with gastroenteritis morbidity and mortality based on medical registrations (GP-consultations, hospitalizations and deaths) for an 8 year period (1999-2006). Because norovirus laboratory

testing is uncommon and the health burden of norovirus infections is unknown, studying medical diagnoses and causes of death of unexplained gastroenteritis (i.e. of unknown causative pathogen) allowed us to estimate the proportion that is attributable to norovirus activity, in particular for the heavy outbreak seasons of 2002/2003, 2004/2005 and 2005/2006.

Methods

To study the activity of norovirus (explanatory variable) independent of other agents (covariables) in association with morbidity and mortality (outcome variables) at the population level, we analyzed time series from 5 data sources (table 8.1). These series reflect weekly counts of selected gastrointestinal complaints or diagnoses derived from registries in the Netherlands. Since norovirus testing [189] is not common in general practices (GP) or in hospitals (their coding systems do not refer specifically to norovirus infection), cases are rarely registered per se, even if recognized, but are coded into a category for infectious or viral gastroenteritis (probably most labeled to be of unknown etiology). We have grouped such categories under 'unexplained gastroenteritis' (uGE), as seen in three of the five data sources that follow.

1) Norovirus outbreak surveillance
Information for 13 years of norovirus outbreaks (1994-2006) was acquired from our national norovirus outbreak surveillance system, initiated in 1994. A cluster of two or more epidemiologically linked cases was considered an outbreak. In this surveillance a minimal set of data is collected for reported outbreaks and combined with results from molecular-biological detection and typing techniques [268,314,322].

2) uGE in general practitioner (GP) consultations
A diagnosis of a 'suspected gastro-intestinal infection' (ICPC D73) was defined as uGE. Vomiting and diarrhea (both common in norovirus infection, International Classification of Primary Care [ICPC] codes D10 and D11) were considered separately. GP consultation data came from a sentinel network of GPs (LINH) covering 2% of the Dutch population between 2001 and 2006. Virtually all Dutch citizens, including individuals in homes for the elderly, are registered with a GP (i.e. family physician), their first point of contact with the Dutch health care system. Individuals in nursing homes are alternatively grouped with a nursing home GP.

3) uGE in hospitalizations
A hospital discharge diagnosis (primary or secondary) of gastroenteritis designated to be of viral or of infectious etiology but not clearly specified was defined as uGE (International

Classification of Diseases, Ninth Revision, or ICD9, codes 0086, 0088, 0090-0091, 0059, and 5589). We also included discharge diagnoses registered as 'Other and unspecified noninfectious gastroenteritis' (ICD9: 5589) as this code is known to consist of a large viral GE infectious component, not registered or recognized as such [54]. Data were obtained from the Dutch National Medical Register (LMR) which covered 99% of the Dutch population (16.3 million) between 1999 and 2006.

4) uGE in mortality data

Causes of death attributed to viral gastroenteritis (A081-A089) and infectious gastroenteritis (A091-A099, International Classification of Diseases and Related Health Problems, or ICD10) were defined as uGE deaths, with data from 1999 to 2006. Both primary and secondary causes of death were considered. The code A081 designates norovirus as a cause of death but was applied to only one individual during the whole study period (table 8.2). Our mortality statistics originate from Statistics Netherlands (CBS) and cover the total Dutch population.

5) Classical laboratory surveillance

Infectious pathogens besides norovirus that can cause gastrointestinal complaints were considered (positive results of diagnostic testing) using time series for: influenza (coverage 73%), rotavirus (38%), *salmonella* (64%), *campylobacter* (50%), and *shigella* (100%), available from laboratory surveillance for 1999-2006 [309] (for viruses: Weekly Sentinel Surveillance System of the Dutch Working Group on Clinical Virology). The time series are representative for the Dutch population of all ages.

Study population

Graphic representation and analyses of trends were restricted to persons 65 years and older, except for laboratory surveillance data and norovirus outbreak data which do not provide data on age. However most norovirus outbreaks occur in nursing homes and thus in the elderly. The gastroenteritis morbidity pattern in the elderly differs from that in children, whose gastroenteritis hospitalization rates are dominated by rotavirus activity [54,307].

Statistical analyses

Monthly time series of uGE trends in the elderly were plotted. On weekly data we used Poisson regression models (which included linear and periodic components) to characterize various trends:

Model 1) uGE GP consultations explained by a combination of: a linear trend, a seasonal trend, norovirus outbreak activity, and other pathogens from laboratory surveillance; model 2) uGE hospitalizations explained by the same variables as given for model 1; and model 3) uGE deaths explained by the same variables as given for model 1.

We used the following equation:

$UGE(t) \sim Poisson[\lambda(t)]$

$\lambda(t) = b_0 + b_1*t + b_2* \sin(2*pi*t/52) + b_3*\cos(2*pi*t/52) + b_4*P1(t+lagx) + b_5*P2(t+lagx) + b_6*P3(t+lagx) \ldots + R(t)$

(See below, in the appendix to the chapter, for the list of definitions of all terms).

This model is similar to previously used models by others who have modeled topics such as hospital admissions attributable to rotavirus activity [54,257], and deaths attributable to influenza and RSV [88]. In this model we also incorporated seasonal trends as these need to be removed from the data since many health variables show systematic variation over the course of a year even if these variables are not causally related [261]. We used a generalized linear model with a Poisson distributed error and an identity link and adjusted the model for overdispersion in the data. For the model we assumed that each weekly pathogen count was associated with a constant number of UGE cases throughout the study period (i.e. regression coefficients: b_4, b_5, b_6 etc. not varying with time). b0 is the estimated constant level of weekly registered UGE which is not explained by the variation in pathogen trends nor the seasonal trends that were included in our model. We did not include the population at risk in the model, as the population increase in this period is moderate. Also any increasing trend will be taken into account by the linear term in our model (b1). For each parameter we computed Wald 95% confidence intervals.

For each model we first checked 1) whether a significant increasing or decreasing linear trend with time was present and 2) whether a significant seasonal trend (sine and cosine terms), was present. Then, using a forward stepwise selection, we checked which additional explanatory variables (norovirus outbreaks, and other pathogen counts) contributed significantly to the trends in the outcome variables (which were defined either as uGE_GP consultations, or uGE_hospitalizations, or uGE_mortality). We also evaluated the association with the lagged values of the explanatory variable (up to 4 weeks backwards in time), building each increment in the model by adding all possible lags of all pathogens and selecting the lag with the best fit (assessed with the deviance), until no more pathogens contributed significantly to the model. In the final model, each pathogen, appropriately lagged, was included in the model only once. Negative associations were not included to avoid over-modeling of the data, with the underlying consideration that pathogens can cause disease but generally do not decrease disease burden.

Results

Norovirus outbreaks on the rise

Data on the frequency and starting dates of reported norovirus outbreaks, available since 1994, is detailed elsewhere [268]. Outbreaks occur mostly in the winter season (October-March). The three recent winter seasons of 2002/2003, 2004/2005, 2006/2007, stand out in these time series, with well over 100 confirmed norovirus outbreaks in each season. Although only the beginning of the 2006/2007 winter was included in our study period, the weekly number of reported outbreaks were also relatively high (figure 8.1A). The norovirus time series did not coincide with known trends in any other monitored enteric pathogen nor with influenza, which occasionally causes gastroenteritis symptoms (figure 8.1A). For most (95%) norovirus outbreaks, the month of occurrence was known; for 72% information on week was also available. As the missing weeks were distributed equally over the observed years, the trend is similar whether based on monthly or weekly data (data not shown).

Trends in unexplained gastroenteritis illness (at GP's and hospitalizations)

Marked elevations of uGE among the elderly were seen in both GP consultations and in hospitalizations, coinciding with all the three heavy epidemic norovirus peaks (winter 2002/2003, 2004/2005 and 2005/2006 figures 8.1B and 8.1C). The average monthly incidence of uGE registered by sentinel physicians increased a marked 70%-240% above average at the height of the three unusual norovirus peaks (overall average: 91/100,000 elderly individuals, increasing to 150, 171, and 217 respectively) (figure 8.1B). A similar trend, albeit a bit less extreme, was seen for hospitalizations, with incidence of uGE increasing more than 30% above average during the three epidemic norovirus outbreaks (the overall average monthly incidence of uGE was 19/100,000, but peaking at 24, 27 and 26 per 100,000 elderly in the epidemic seasons) (figure 8.1C). Such high peaks in unexplained morbidity were seen at no other time in uGE hospitalizations and at only one additional other time in uGE GP consultations: April 2004, coinciding with rotavirus season.

In regression models, norovirus outbreaks were a significant predictor of both the number of uGE GP consultations as well as the number uGE hospitalizations. Added to a model composing of a linear and seasonal trend, norovirus activity (lag 0), rotavirus (lag 0), Shigella (2 weeks previously), and campylobacter (lag 0), were significant predictors of uGE GP consultations in a multivariate model (p-values for all coefficients <0,02) (table 8.3). For hospitalizations: norovirus activity (3 weeks previously), rotavirus (lag 0), *shigella* (lag 4), *campylobacter* (lag 0), influenza (lag 0) were significant predictors of uGE (p-values for all coefficients <0,02). In the model of uGE GP consultations, the beta for norovirus was 0.51 (95%CI: 0.33-0.67; p<0.0001) (table 8.3). With the GP-sentinel data representing 2% of

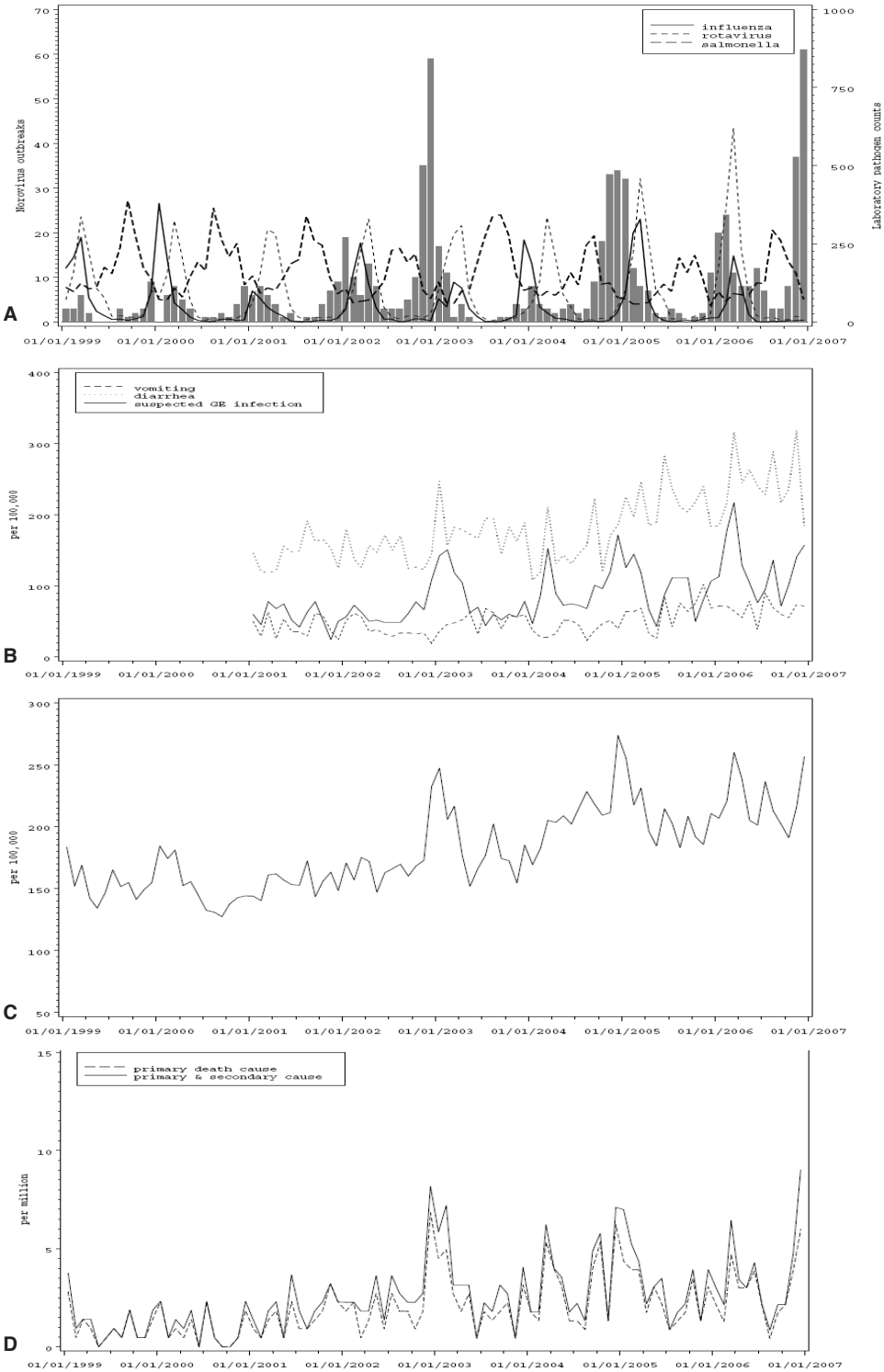

A
Norovirus outbreaks
Laboratory pathogen counts
influenza
rotavirus
salmonella
B
per 100,000
vomiting
diarrhea
suspected GE infection
C
per 100,000
D
per million
primary death cause
primary & secondary cause
01/01/1999
01/01/2000
01/01/2001
01/01/2002
01/01/2003
01/01/2004
01/01/2005
01/01/2006
01/01/2007

Figure 8.1. Trenddata. A.) Monthly trends in norovirus outbreaks (grey bars), and laboratory-based trends in influenza, rotavirus, and salmonella, in the Netherlands from 1999 through 2006. Shigella and campylobacter not shown; like salmonella, they peak in the summer. B.) Monthly incidence of unexplained gastroenteritis consultations with general practitioners (sentinel network) in the elderly (aged 65+). * Unexplained gastroenteritis: suspected GE infection, ICPC code D73. C.) Monthly incidence of unexplained viral gastroenteritis hospitalizations in Dutch elderly (aged 65+). Unexplained gastroenteritis of possible viral origin: ICD9 codes 0086-0093 + 0059 + 5589. D.) Monthly incidence of unexplained gastroenteritis deaths* in Dutch elderly (aged 65+). Unexplained gastroenteritis of possible viral origin: ICD10 codes: A08-A09. Black line: registered as either primary or secondary death cause, Dashed line: registered as primary death cause.

the Dutch population, we estimate this beta to be approximately 50 times higher for the total population (50*0.51 = 25.5, i.e. approximately 25 uGE GP visits for every norovirus outbreak), thus attributing over 11,000 uGE GP visits to norovirus activity over a 5 year period (the majority occurring in the three epidemic years: e.g. 3804 attributed to the epidemic year of 2004/2005) (table 8.4).
The beta for uGE hospitalizations was 2.15 (95%CI: 1.57-2.74), thus attributing from 215 up to 318 uGE hospitalizations to norovirus activity in the 3 epidemic years, and lower numbers in the years with milder norovirus outbreak activity (41 up to 129 hospitalizations) (table 8.4).

Whilst the uGE trend in both GP and hospital data clearly coincided and was associated with norovirus activity, the GP consultations of vomiting and of diarrhea (both symptoms occurring with numerous illnesses) showed no clear trends.

Trends in unexplained gastroenteritis deaths
A total of 551 deaths with uGE as primary or secondary cause were registered in 1999-2006 in the elderly, or an average monthly incidence of 2.6 deaths per million inhabitants 65 years and older. As with uGE morbidity, uGE deaths also rose remarkably during the three norovirus epidemics (figure 8.1D), with the monthly rate rising to above 6.0 per million during the norovirus seasons (even peaking at 8.2 per million in the 2002/2003 winter). UGE was coded as the primary cause of death (N=437) more often than as secondary cause (N=114, table 8.2), but both showed the same temporal trend. The number of deaths in the two sub-classifications of uGE deaths (A08: 'viral intestinal infection' and A09: 'gastroenteritis of presumed infectious origin') were roughly equal, with 26% more deaths coded in the latter (280 vs 354 for all age groups combined) during the 8 year study period.

Visually, the peaks in mortality during the two norovirus epidemics did not seem to coincide with known seasons of influenza or infections with rotavirus, *salmonella*, *shigella*, or *campylobacter* (figure 8.1A). Indeed, when modeling uGE deaths by laboratory counts of enteric pathogens and correcting for a linear and seasonal trend with time, norovirus

activity (1 weeks previously) remained as the single pathogen with significant predictive value (figure 8.2). The beta for norovirus was 0.14 (95%CI: 0.08-0.21; p<0.0001; table 8.3), i.e. one death registered as uGE for every 7 outbreaks (79 of all 551 uGE deaths in the elderly were thus attributed to norovirus over the 8 year period) (table 8.4 & figure 8.2). The remaining gastro-pathogens did not significantly improve the model, nor did influenza.

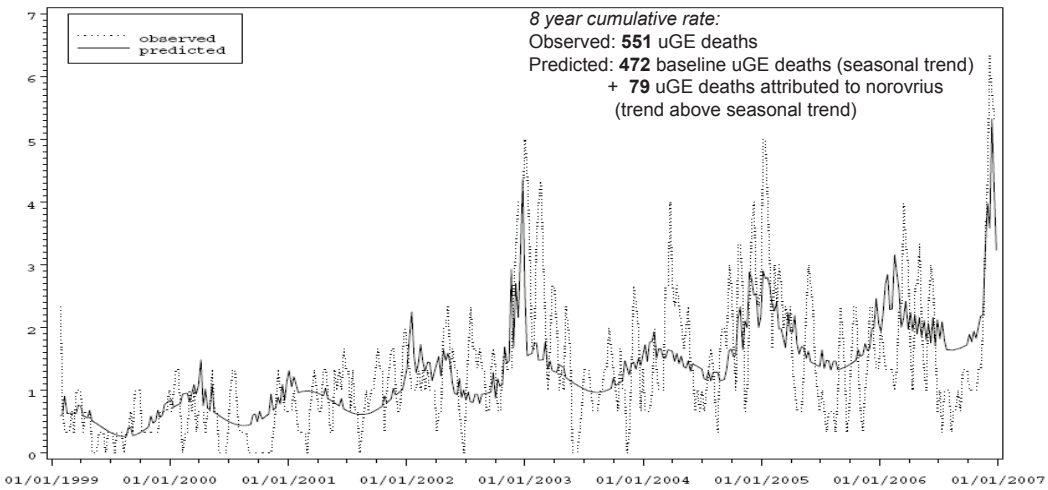


Figure 8.2. Modeled weekly deaths due to unexplained gastroenteritis in Dutch elderly (65+ years of age), 1999-2006. Observed (dashed line): Based on the Dutch national death registry: any death cause (primary or secondary) coded as viral, intestinal (ICD10: A08) or gastroenteritis of presumed infectious origin (ICD10: A09). Plotted as the 3-week moving average. Predicted (black line): based on coefficients obtained from a regression model, which included an intercept, a linear increasing trend with time, a cyclical seasonal trend (a sine and cosine term), and total norovirus outbreaks (1 week previously) as explanatory variables (norovirus outbreaks not available stratified by age).

Discussion

While norovirus gastroenteritis is generally viewed as a trivial illness of short duration, this study using population-based databases, shows that to date, its public health consequences may have been underestimated. Extreme outbreak activity (in the 2002/2003, 2004/2005, and 2005/2006 winters) due to the introduction of new strains [175] coincided with sharp increases of uGE morbidity, hospitalizations and even deaths in the elderly.

To date there is no other published data to our knowledge that estimates the impact of norovirus outbreaks at different public health levels: from milder illness (presenting to general practitioners), to severe illness (presenting at or during hospitalizations) and even deaths. That norovirus is significantly associated not only with morbidity but even with deaths in the elderly is remarkable considering that norovirus is regarded as an illness

with a low case-fatality rate [223]. A few recent – mostly anecdotal – reports have already suggested that norovirus infections may be more severe for the elderly. For example 17 Japanese elderly died in nursing homes after being struck by norovirus starting December 2004 [220], and nursing homes in the U.S. and in Israel reported unexpected numbers of deaths during a single norovirus outbreak, also in the absence of any influenza activity [2,29,335]. In October 2008, outbreaks in nursing homes both in the Netherlands and in New Zealand lead to media reports of deaths among inhabitants. An older, larger study in England and Wales estimated only one death for every 50 outbreaks [176], but these data were acquired before the worldwide emergence of the recent epidemic norovirus variants since 2002 [175] and only considered persons within the outbreaks and not in the exposed and infected population as a whole. In the Netherlands we noted that after the striking epidemic norovirus season of 2002/2003 (the first since 1996, a year which also coincided with a novel GII.4 variant), gastroenteritis deaths suddenly increased [306]. Our data in the current study, which includes three recent epidemic years over an 8 year study period, indicate that one in 7 reported norovirus outbreaks was associated with a fatality in the elderly population for which uGE was registered as a cause of death. In total, our models estimated up to 21 yearly uGE deaths, 318 hospitalizations, and 3804 GP visits, to be attributable to norovirus activity in the heavy 2004/2005 norovirus season alone. The numbers of deaths are comparable to a recent result from the UK (with a population size roughly 3 times larger than the Netherlands) attributing an average 80 yearly deaths to norovirus infection [110]. To date no other studies have published estimates of norovirus associated illness and deaths. With 3 to 8 uGE deaths attributed to norovirus even in non-epidemic norovirus years, our study suggests that in the past, prior to extreme epidemic seasons, norovirus associated deaths have also been underestimated or overlooked.

Although our study shows significant morbidity and mortality attributable to norovirus activity, we expect our estimates to be an underestimation. Illness and deaths due to infectious gastroenteritis are poorly recognized and underreported [78], with Frenzen even suggesting a roughly 12-fold underestimation [78,199,273]. Reason for this is that it may be difficult to ascertain which illnesses and underlying diseases in the elderly contribute exactly to death. As a result, coding causes of death as is done for registration purposes may underestimate the contribution of complications such as a viral infection. A recent in-depth Dutch study to better identify direct causes of deaths, focused on the last two days of life of nursing home patients and showed that dehydration (i.e. disorders of electrolyte and fluid balance), while reversible, is among the three most common causes of death [20]. Such symptoms or consequences, which can also come hand in hand with a norovirus infection, apparently remain largely unidentified or unregistered as a (contributing) cause of death. In our model, the highest norovirus correlation with uGE deaths was found when

a lag of 1 week was used, suggesting that norovirus outbreaks preceded uGE deaths by 1 week. Such a delay may also contribute to the poor recognition and reporting of uGE as a primary, secondary, or contributing cause of death. For hospitalizations the optimal delay was longer, 3 weeks, making recognition (and thus registration) of a gastrointestinal cause of hospitalization even less likely. In the community gastroenteritis complaints seem to present simultaneously with the emergence of norovirus outbreaks as illustrated by an optimal delay of 0 weeks in the model of uGE GP-visits. To explain the difference in delay duration between hospitalizations and deaths (1 week versus 3 weeks), we hypothesize that whilst norovirus infection may trigger death relatively rapidly in the very frail, in the less frail elderly population the virus may be introduced later (nursing homes, with the frailest populations, are notorious for outbreaks), or it may trigger a slower deterioration of health or a slower aggravation of underlying illnesses.

That epidemic norovirus years show more outbreaks and thus more associated illness and death might suggest that new viral strains also manifest with increased pathogenicity, but this can not be proven on the basis of the observations in this study. New variants could be expected to cause greater numbers of disease simply due to a larger pool of susceptibles in which immunity has not yet developed. Recent evidence suggests that host immunity does play a role in norovirus evolution with the GII.4 variants being antigenic variants [156,170,175,268,269].

Surveillance artifacts could play a role in an ecological analysis [310] such as ours, but the consistency of the trends across all the different medical registrations (norovirus surveillance, GP-consultations, hospitalizations, and deaths), supports the hypothesis that norovirus epidemics are associated with morbidity and death in the elderly, and that these trends may be influenced by viral evolution. Furthermore, for the norovirus outbreak registration, it must be noted that other countries also registered high peaks of norovirus activity in the same epidemic years and the number of outbreaks in the Netherlands seem to be clearly increasing [317], supporting our belief that the increases seen in our registration are not spurious. However, for the medical registries bias can not be ruled out entirely. For example, physicians might be more prone to coding gastroenteritis as a cause of illness or death when they know that it is the norovirus season – but we have no a priori reason to believe that gastro-enteritis would otherwise be ignored outside the norovirus season. Other potential confounder of our results are *Clostridium difficile* ribotype 027 infections, which are associated with high mortality. However the emergence of this type in the Netherlands was first noted in 2005 [312], long after the first epidemic season (2002/2003) of norovirus that was included in our data. Furthermore *Clostridium difficile* infections in the Netherlands are not known to exhibit a clear seasonal trend.

There are no known indications of coding practices changing over time, and coding systems were consistent over the study years within each registry (ICPC codes for GP data, ICD9 codes Dutch version for hospitalizations, and ICD10 codes for the death registry).

In conclusion, our investigation indicates that norovirus infection is a significant contributor to severe morbidity and mortality in the elderly and that in fact the potentially severe consequences of norovirus infections have long been underestimated. Because the relationship between norovirus activity and morbidity and mortality is poorly understood [329], and because of the probable underestimation of the health burden of this virus, even in our study, this issue needs further research, especially at the individual level, i.e. following elderly patients from GPs and nursing home physicians to hospitalization and death. Such studies are difficult to conduct and to our knowledge there are no estimates available that include data from such follow-up studies in the modeling of the disease burden of norovirus infections. That heavy norovirus seasons - epidemics driven by viral evolution [175]- result in more morbidity and mortality is of public health concern, and even more so if the mortality is premature and could have been prevented – either by hygiene and isolation measures or by treatment. Thirteen years of norovirus surveillance in the Netherlands illustrates the frequent emergence of norovirus variants causing epidemic seasons in 1996, 2002, and 2004, and even with two new variants emerging in 2006 causing the greatest number of outbreaks recorded in a single season (winter of 2006/2007) [155,175,269,320]. With such a rapidly evolving virus, stricter hygiene protocols in nursing homes are therefore called for [112], and vigilant monitoring of outbreaks and molecular evolution remains crucial.

Table 8.2. Deaths registered as unexplained gastroenteritis, 1999-2006 (viral and infectious), in the total Dutch population.

ICD10 code	Description	As primary cause	As primary or secondary cause
A08	Intestinal infections, viral & specified	215	280
(A08.0	*Enteritis due to rotavirus)*	(5)	(6)
(A08.1	*Enteritis due to norwalk-virus)*	(2)	(2)
(A08.2	*Enteritis due to adenovirus)*	(1)	(3)
(A08.3	*Viral enteritis, other)*	(6)	(12)
(A08.4	*Viral enteritis, not elsewhere specified)*	(201)	(257)
(A08.5	*Intestinal infections)*	(0)	(0)
A09	Diarrhea & gastroenteritis of infectious origin	283	354
Total		498	634
In elderly[a]		[a]437 (88%)	[a]551 (87%)

[a] Restricted to those aged 65 years and older.

Appendix to Chapter 8

Poisson regression models (with an identity link and Poisson error) explaining the variation in the number of unexplained gastro-enteritis cases per week were constructed. A separate model was constructed for each data source: UGE registered in 1) general practitioner sentinel data, 2) hospitalizations, and 3) mortality.

$UGE(t) \sim Poisson[\lambda(t)]$

$\lambda(t) = b_0 + b_1*t + b_2* \sin(2*pi*t/52) + b_3*\cos(2*pi*t/52) + b_4*P1(t+lagx) + b_5*P2(t+lagx) + b_6*P3(t+lagx) \ldots + R(t)$

UGE (t) The number of unexplained gastroenteritis cases at time t.
t Time in weeks.
λ the intensity or expected value
b_0 Regression coefficient describing the expected baseline numbers of UGE unexplained by any of the model parameters.
b_1: Regression coefficient for linear trend in time.
b_2 and b_3: Regression coefficients for sine and cosine terms. This models seasonal trends in which the period of the cyclical trend is defined as recurring each 52 weeks (i.e. yearly).
b4, b5, b6, etc. Regression coefficients for each of the pathogens in the model that significantly explain variation in UGE at time t.
Pi(t) number of cases of pathogen i at time t (from laboratory surveillance).
lagx Number of weeks for which the respective pathogen trends lag behind UGE trend ($0 \leq lagx \leq 5$ weeks).
R(t) a term accounting for overdispersion

Table 8.1. Characteristics of data sources.

	Time-period	National Coverage[a]	Description	Source	Codes or pathogens
Norovirus surveillance data	1994 - 2006	National (but underreported)	Passive reporting of outbreaks by health services and food inspectorates	Norovirus Outbreak Surveillance System	Norovirus
General practitioner data	2001 – 2006	2%	Signs, symptoms and prescribed medication at: surgery visits, telephone, calls and home visits	Netherlands Information Network of General Practice (LINH, a sentinel system)	ICPC-1[b]: D10, D11, D73 Vomiting; diarrhea; suspected gastro-intestinal infection
Hospital discharge diagnoses	1999 – 2006	>95%	Main discharge diagnosis and secondary diagnoses, with hospitalization and discharge dates	Dutch National Medical Register (LMR)	ICD9[c]: 0086-0093 + 0059 + 5589: 0086 Enteritis due to specified virus 0089 Enteritis viral, not elsewhere classified 009 Ill-defined intestinal infections: 0090 Infectious colitis, enteritis, and gastroenteritis 009.1 Colitis, enteritis, and gastroenteritis of presumed infectious origin 009.2 Infectious diarrhea 009.3 Diarrhea of presumed inf origin 0059 Food poisoning, unspecified 5589 Other and unspecified noninfectious gastroenteritis and colitis
Causes of death	1999 – 2006	100%	Date and cause of death (primary and secondary causes)	Statistics Netherlands (CBS)	ICD10[d]: A08, A09 Viral and other specified intestinal infections[e]; diarrhea and gastroenteritis of presumed infectious origin
Classical laboratory surveillance	1999 –2006	Common viruses: 38-73% Common bacteria: 50-100%	Positive results of diagnostic testing	Weekly Sentinel Surveillance System of the Dutch Working Group on Clinical Virology	rotavirus, salmonella, shigella, campylobacter, influenza, respiratory syncytial virus

[a] Percentage of the total population in the Netherlands, [b] ICPC: International Classification of Primary Care, [c] ICD9: International Classification of Diseases, Ninth Revision, [d] ICD10 International Classification of Diseases and Related Health Problems, tenth revision, [e] A specific description suggesting that they are explained by a pathogen, but the majority was 'viral enteritis not elsewhere specified'

Table 8.3. Norovirus outbreaks and other significant predictors of unexplained gastroenteritis in GP consultations, hospitalizations and deaths: results from regression models in persons 65 years and older.

Outcome variable		Predictive pathogens in multivariate[a] models			
		Norovirus outbreaks			Other significant pathogens in the same multivariate model
Unexplained gastroenteritis	**Time period**	**Optimal lag for norovirus outbreak**	**Parameter estimate (beta)[c]**	**95% CI**	***Pathogen (optimal lag[d])***
uGE GP consultations[b]	2001-2006	0	25.7[e]	(16.5-33.5)[e]	Rotavirus (0), shigella (2), campylobacter (4),
uGE hospitalizations[b]	1999-2006	3 weeks previously	2.15	(1.57-2.74)	Rotavirus (0), shigella (4), campylobacter (0), Influenza (0)
uGE deaths[b]	1999-2006	1 week previously	0.14	(0.08-0.21)	None

[a] Shown are pathogens that are significant predictors in a multivariate regression model that also includes a constant, a linear, and a seasonal trend (sine and cosine terms).

[b] uGE: unexplained gastroenteritis of possible viral origin (for GP consultations: ICPC code D73, for hospitalizations: ICD9 code 0086-0093 + 0059 + 5589, for deaths: ICD10 codes: A08-A09 registered as either primary or secondary death cause).

[c] Represents the number of patients or deaths associated with one registered norovirus outbreak.

[d] Given in number of weeks previous to the outcome variable.

[e] Beta: 0.51 based on GP sentinel data with a coverage of 2% of the total population (beta and CI are multiplied by 50 to estimate a beta for the total population)

Table 8.4. Estimated unexplained gastroenteritis diagnoses in the elderly (GP consultations, hospitalizations, and deaths) that are attributed to norovirus outbreaks.

		Morbidity and mortality attributed to norovirus outbreak activity		
	Norovirus activity	**uGE[a] GP**	**uGE[a] hospitalizations**	**uGE[a] deaths**
Season[b]	**Nr. of outbreaks**	**Attributed to Norovirus**	**Attributed to Norovirus**	**Attributed to Norovirus**
1999 / 2000	27	-	58	4
2000 / 2001	19	-	41	3
2001 / 2002	60	1542	129	8
2002 / 2003	128	3290	275	18
2003 / 2004	30	771	65	4
2004 / 2005	148	3804	318	21
2005 / 2006	100	2570	215	14

[a] uGE: unexplained gastroenteritis of possible viral origin (for GP consultations: ICPC code D73, for hospitalizations: ICD9 codes 0086-0093 + 0059 + 5589, for deaths: ICD10 codes: A08-A09 registered as either primary or secondary death cause).

[b] Including only complete years running from July-June. 1998/1999 and 2006/2007 seasons not shown as both only include half a year of data (data was available from 1/1/1999 - 31/12/2006)

Chapter 9

General Discussion and Conclusions

Partly taken from: Cpt. 1 Norovirus Epidemiology
In: Caliciviruses: Molecular and Cellular Virology
By: J. Joukje Siebenga, Erwin Duizer and Marion P.G. Koopmans
Edited by: Grant S. Hansman, Jason Jiang and Kim Y. Green

At the start of this thesis project in March 2005 the third recorded global epidemic of norovirus GII.4 illness was just coming to its end (winter season 2004-2005), and the realization that norovirus infections contribute to a potentially severe burden of disease had settled in the minds of people working in the field of norovirus research. During the course of this PhD norovirus outbreak reports peaked once more during the winter of 2006-2007. The number of winter-outbreaks in the Netherlands has stayed elevated since then. The realization that norovirus is a pathogen with a substantial public health impact is still growing. Besides the increasing occurrence of norovirus infections, this is largely a result of increasing knowledge of the virus, its epidemiology and clinical impact and the mechanisms underlying these phenomena. The recent progress in norovirus epidemiology and evolution research will be discussed here, in relation to the work presented in this thesis.

Norovirus diversity and prevalence

Many genotypes exist and circulate, GII.4 dominates

Although many genotypes circulate in the global population, the GII.4 viruses have been dominant globally during the past decade [270]. The genetic diversity and prevalence of the GI, GII and GIV norovirus strains is presented in figure 9.1. Some of the represented genotypes have not been reported in the FBVE network since 1990, among which genotypes GII.11 and GII.19. These last two genotypes are genotypes that infect swine, rather than humans.

Viruses detected in sporadic cases are generally of more diverse genotypes, indicating that the GII.4 strains are especially good at spreading and causing symptomatic infections, thus at causing outbreaks. Strains belonging to other genotypes cause outbreaks and sometimes severe problems as well. Notably the GII.3 viruses have been reported often in relation to illness in children, hospital outbreaks, and prolonged infections. Besides the regular GII.3 strains, recombinant strains with GIIb ORF1 sequences circulated, which is why the percentage of detected GII.3 polymerases is lower than the percentage of GII.3 capsids in figure 9.1.

Among the GII.4 strains separate variants are distinguished, that have displaced each other through time

Characterization of a systematic sampling of outbreak strains showed that the GII.4 strains cluster into separate lineages [269], which each displaced the previously dominant variant (figures 9.2 and 9.3). The first recorded epidemic occurred in the winter of 1995-96 [212,320], and from 2002 onward, after what was in hindsight a relatively long period of genetic stasis, a series of epidemic winters followed [27,175,330]. In the spring of 2006 two genetically distinct variants emerged. These strains co-circulated during the next winter, after which the 2006b variant became dominant over the other, 2006a variant. In 2008 and 2009 new

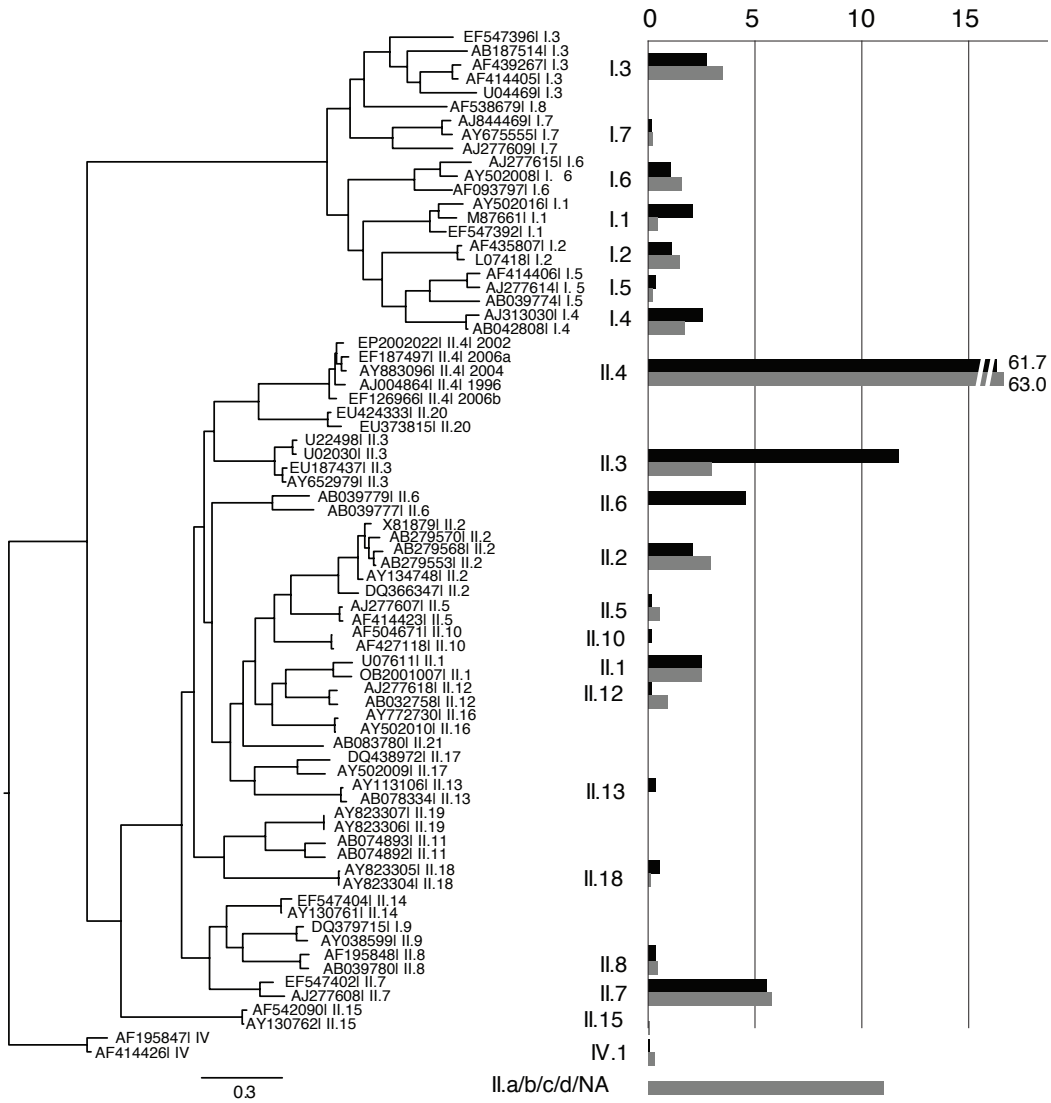


Figure 9.1. Norovirus phylogenetic tree and genotype prevalence. The left panel shows a phylogenetic tree based on complete capsid nucleotide sequences, constructed using PHYML. The sequences selected to represent the genotypes have been derived from the sequence alignment used in the new norovirus typing tool. This alignment contains sequences aimed at covering optimal diversity within each genotype to enable reliable genotyping. Nomenclature refers to the current genotype nomenclature for which a consensus discussion is taking place (http://www.rivm.nl/mpf/norovirustypingtool). To the right prevalence levels as percentage of all reported norovirus outbreaks are shown per genotype. These are based on the FBVE database, containing sequences of the outbreaks reported to the FBVE network, detected between 1990 and 2008 (described in [128]). Typing results were obtained either by partial capsid sequencing (1472 sequences, represented by black bars) or partial polymerase sequencing (5089 sequences, grey bars). NA: not assigned.

lineages of GII.4 strains were identified, and detected at several locations around the world, but the 2006b variant remained dominant for three seasons in a row. During the 2009-2010 winter the 2008 variant took over as the dominant strain in several countries, causing the majority of reported norovirus outbreaks (data not shown).

Most GII.4 variants, but not all, have a global distribution

Sequence comparison in our global prevalence study showed that four of the GII.4 variants had a true global distribution, namely the 1996, 2002, 2004 and 2006b variants. The 2006a variant was scarcely detected in Asia, and the 2003Asia variant, conversely, was rarely detected outside Asia [270]. Besides the major variants that caused significant numbers of outbreaks, several less successful GII.4 strains circulated. These minor variants of GII.4, e.g. the 2001Japan and the 2001Henry variants, have been widely detected geographically, but at low prevalence. The cause for such differences in spread and prevalence currently remains open for speculation. Structural differences in the capsid protein may limit the host range of some variants, although based on the assessment of the currently identified ligand binding sites [32,289] no differences could be identified that might cause this [267,270]. The amino acids that are directly involved in binding types A and B antigens, namely T_{344}, R_{345}, A_{346}, D_{374}, C_{441} and G_{443}, are all identical for the 2006a, 2006b and 2003Asia variants. This is not surprising, as the binding interface is known to be conserved [289]. The reported binding pocket, formed by the residues 390 – 395 and 444, also involved in ligand binding, does show a lot of variation, but the sequences of 2006a and 2006b are identical (QDGSTT), therefore not providing an explanation for differing host binding properties. Moreover, although the majority of the 2003Asia variant strains have a slightly different sequence (QDGSSA) some 2003Asia strains were reported with the same sequence (e.g. AB294791 and AB294786). Alternatively, the lack of a successful seeding event may have contributed to the lack of success of these variants in all geographic areas. However, this seems an unlikely explanation in light of the observed very rapid and efficient spread of other norovirus strains as well as other viruses across the globe [122,270]. Competition between different variants may have played a role in limiting the spread of certain variants, if they were antigenically similar. Previous immunity to an older variant may provide partial cross-protection against a particular newly emerging one, as is reported for influenza A [231] and many other viruses, provided they were not encountered with too much time in between, given the poor long term immunity mounted against norovirus infections. Nonetheless, GII.4-2006a and -2006b co-circulated in Europe, America and Oceania without restraints.

Norovirus surveillance has a relatively limited global coverage, and especially from developing countries not much data is available. Thus, it remains largely unknown which strains circulate in Africa, South America and large parts of Asia. The still relatively sparse data from other continents do not provide a dense enough sampling to pinpoint a geographic

origin for emerging variants, although Oceania and Asia seem to have detected three of the global variants first, against one in Europe [270].

The evolutionary pattern of GII.4 is shaped by population immunity

While during epidemics the percentage of GII.4 strains detected in outbreaks could be well over 80%, during inter-epidemic years GII.4 prevalence was lower, and strains of other genotypes were reported in higher numbers, both in percentage as well as in absolute numbers [268]. For example, strains with GIIb ORF1 sequences were reported frequently in the winter of 2003-2004 in Europe [168].

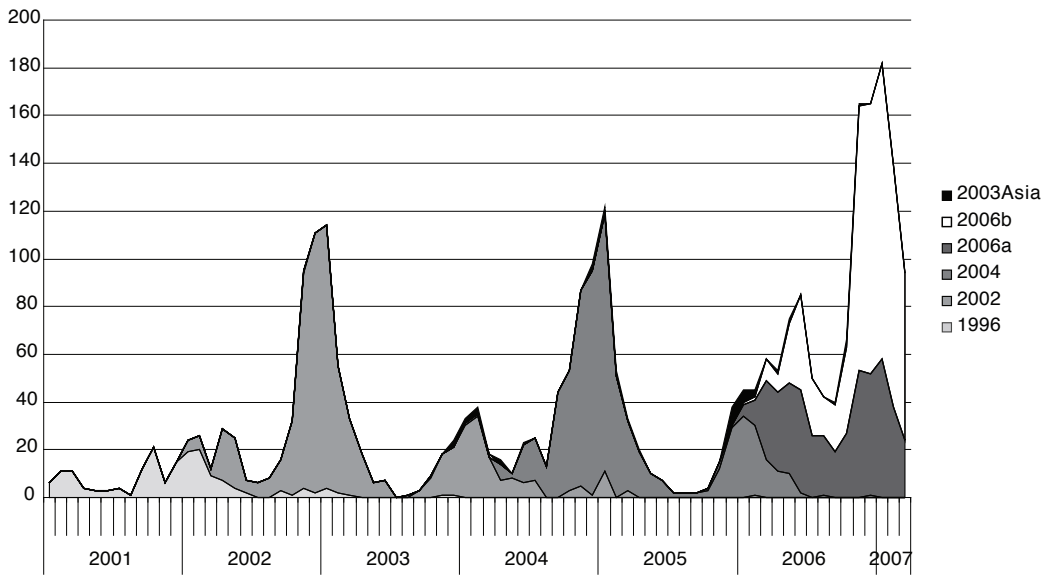


Figure 9.2. Norovirus GII.4 variants displaced each other through time. Data obtained from the first global NoroNet GII.4 prevalence overview [270]. Strains (n=3089) detected between January 2001 and March 2007, contributions from Hungary, Hong Kong, Japan, Australia, Germany, Netherlands, USA, Canada and New Zealand.

The leading explanation for the observed pattern of variant emergence, circulation and displacement is that the high prevalence of GII.4 in the population leads to population immunity against the circulating dominant variant [6,30,170,269,270]. Although no long-term immunity develops in individuals [128,227,336], population immunity against norovirus would be based on short-term immunity that may occasionally be boostered by encounters with the same virus-variant while it still circulates. The development of population immunity implies that the subsequent GII.4 variants must be antigenically distinct, or at least the ones that directly succeeded each other in time. This, in the absence of

a method for performing neutralization assays, was confirmed by ligand-binding blocking experiments using virus-like particles of different GII.4 variants [30,170], showing that subsequent GII.4 variants are indeed antigenic variants. Although complex crossreactivity patterns of sera were reported in a number of studies [30,107,170], these crossreacting antibodies were shown not be able to block binding of VLPs to the putative ligands, HBGAs [30], suggesting that these crossreactive antibodies are not protective.
Changes in host-ligand binding properties were demonstrated between different recent GII.4 variants [59], suggested to enable the virus to switch between host populations. For the 2004 variant and the 2003Asia variant (or the 2005 variant in other literature) only weak binding to phenotyped saliva samples could be found, and similarly only weak binding to synthetic HBGAs [170]. These results may imply that noroviruses utilize different ligands than the HBGAs, additional to, or in place of the HBGAs.
Although the capsid genes of epidemic variants did evolve during the time they circulated in the population, they were surprisingly stable during the stasis periods, as compared to the larger differences between different variants. For Influenza A virus, the continuous circulation of virus in the tropics, more precisely in densely populated cities in Asia, is seen as a driving factor for strain evolution [256]. Whether this is also the case for noroviruses, remains to be seen. Outbreaks caused by noroviruses are seasonal in zones of temperate climate, but it is unclear whether noroviruses are seasonal in the tropics, due to lacking structural surveillance in these regions. The rationale given for the seasonality of norovirus circulation in temperate regions, including decreased stability of virus particles in environments of higher relative humidity, and social factors such as crowding indoors due to cold weather [95,174,204], does not make it plausible that noroviruses are equally seasonal in tropical climate regions. So, continuous circulation in the tropical areas, combined with low levels of circulation (sporadic cases) in the population in the temperate climates [53], may be important drivers for norovirus evolution.

What explains this dominance of GII.4 strains, compared to the other genotypes?
It is as yet unknown why the GII.4 genotype is the genotype that is most successful at causing illness in humans, although it is not difficult to speculate upon possible causes. Some of such speculative hypotheses have not been or cannot be tested. For example, a study assessing the processivity of the polymerases of different genotypes and variants is currently ongoing in a laboratory in Germany; higher polymerase activity could give the GII.4 genotype a key advantage. Alternatively, higher stability of the virus particles, higher infectivity or a lower infectious dose could contribute to the relative success of GII.4, but these properties are currently impossible to measure, since no infectivity assay is available for norovirus. Also, it could be that the amount of virus particles shed by GII.4 infected individuals lies higher than that amount for individuals infected by strains of different genotypes. Although some

studies have quantitatively determined the amount of virus shed by norovirus infected persons [12,165,180,302], no structured assessment has been made striated for genotype. Other (bases for) hypotheses have been tested. Noroviruses of different genotypes are known to have different host specificities, based on their specific binding patterns to host histo blood group antigen (HBGA) types. Several studies have shown that GII.4 has the broadest range for binding of these HBGAs [120,287]; GII.4 therefore has the broadest host range. An additional important property of GII.4 noroviruses, contributing to their continued success, is their propensity to come up with an antigenically distinct shell every couple of years, as it has done during the past decade, thus keeping the pool of susceptibles as large as possible. No other norovirus genotype has been shown to do this at the same scale. In the end the true explanation for the continued success of GII.4 likely lies in a combination of the above mentioned factors, and perhaps additional ones. There is of course an element to this that is self-fulfilling; when the numbers of circulating viruses are high (due to high prevalence, and high levels of shedding), there is more chance for new infections, which means more replication cycles and thus more chance for genetically different strains to emerge.

The amount of norovirus infections has increased during the two past decades

A topic for debate has long been whether norovirus infections were truly increasing in numbers or whether the rise observed in the reported outbreaks was merely the result of a surveillance artifact, caused by increasing awareness, more tests being performed and improvement of detection methods. We argue, based on findings presented in this thesis (chapter 5), and a study done by others of historical feces collections [18], that noroviruses are indeed gaining territory.

Using phylodynamic methods we assessed the number of effective GII.4 norovirus infections, using the genetic diversity of circulating GII.4 as a proxy (chapter 5). Although sampling in the earlier period of this study period was relatively poor, an increasing trend in the amount of infections was shown. Recently, Bok *et al.* studied a historic feces collection (samples collected between 1974 and 1991 in a children's hospital) and tested samples for the presence of norovirus [18]. Although in samples taken as early as 1974 GII.4 strains were detected, they were only present in 16% of all successfully genotyped samples. GII.3 strains were found in 48% of genotyped samples, an interesting finding, as the sample collection was the result of a study to determine the etiology of gastroenteritis in hospitalized children. To date GII.3 is still relatively often reported in relation to illness in children and in outbreaks in hospital settings. From the results of the historical study it may be cautiously concluded that although GII.4 strains circulated in the 1970s (and likely before then), GII.4 was not the dominant genotype at that time, and no other genotype was as dominant in causing illness as GII.4 is now. Combining that with the rise observed in GII.4 infections, it follows that likely not only GII.4 infections have increased, but that resulting of this increase the total

number of norovirus infections has gone up concurrently (i.e., GII.4 strains did not take the place of a previously dominant, other genotype). This corresponds with anecdotal reports of nursing staffs in hospitals and nursing homes, and with suspicions of researchers around the world.

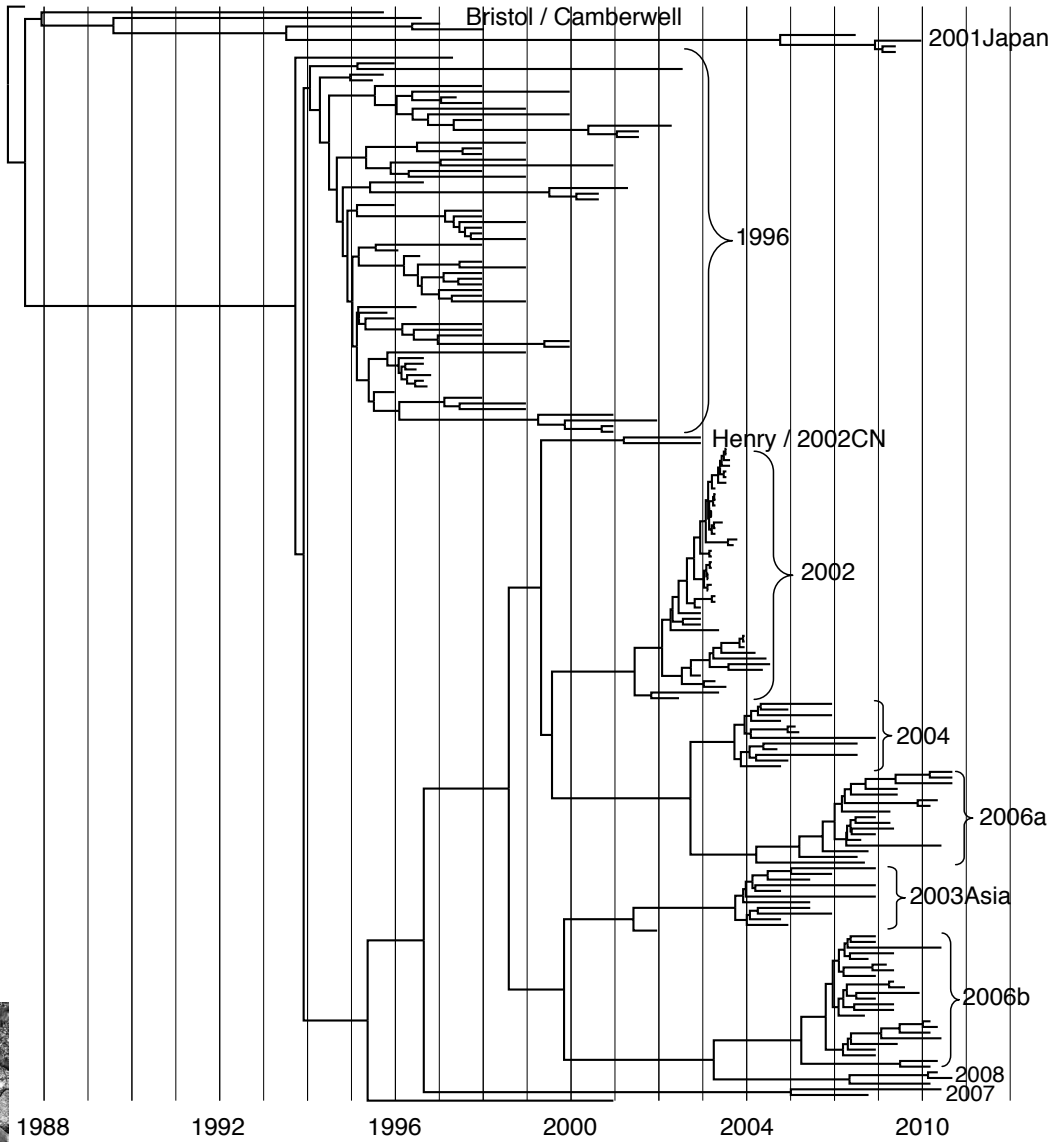


Figure 9.3. Maximum Clade Credibility tree of 194 norovirus GII.4 full capsid nucleotide sequences, detected worldwide. The tree is shown on a time scale, with tips positioned on the detection dates and has been generated with BEAST. Variant clusters are indicated in the figure.

The origin of emerging GII.4 variants remains unknown

The genetic diversity discerning between different antigenic variants of the GII.4 genotype may be large, up to 25 amino acid residues in the 541 amino acid capsid protein (~5%) [269]. Subsequent variants emerged in a stepwise manner, with periods of phenotypic stasis separated by the emergence of phenotypically, in this case genetically and antigenically, distinct strains. This is strongly reminiscent of the evolutionary pattern displayed by Influenza A viruses, and is also known as epochal evolution [142]. No intermediate strains, bridging between subsequent variants, have been detected, which may point to introductions from a non-human reservoir, since surveillance in animals is very rare. On the other hand the relative low intensity of surveillance worldwide, and the lack of screening of asymptomatic patients among the human population does not make detection of low prevalence intermediate strains likely.

Immunocompromized individuals may be a possible source from which new variant strains evolve. Prolonged infections and accompanying shedding of noroviruses have been reported in several studies, including in this thesis. Immunocompromized patients are at especially high risk for such infections. In patients who showed prolonged shedding viruses belonging to GII.2, GII.3, GII.b-GII.3 and GII.4 have been found [84,164,180,210,266]. Viruses detected in these patients accumulated mutations in their capsid proteins rapidly, up to 11 amino acid mutations during a 4 month period [266]. Patients with prolonged shedding were also shown to be infectious to others after having carried the infection for several weeks (this thesis). Nevertheless patients with most severe immune-dysfunction accumulated mutations at the lowest rate, and especially low general antibody levels in the blood seemed to be associated with low mutation accumulation rates [266], likely relating to a lower immune pressure. Conversely, in otherwise healthy children who showed no prolonged symptoms, shedding of norovirus was shown until 6 weeks after infection and long after clearance of symptoms [205]. Viruses shed in the stools of an otherwise healthy child were shown to accumulate mutations at the highest rate reported; 0.13 amino acid mutations were accumulated in the complete capsid daily (~4 amino acids monthly) [266]. Among the general and immunocompetent population the number of infections that occur is magnitudes larger than the number of prolonged infections in immunocompromized patients, so if indeed the mutation accumulation rate is significantly higher in immunocompetent people than in immunocompromized patients, the general population is a more likely source for new variant strains.

Further diversification: recombination

As commonly described for other RNA viruses, genetic diversity of noroviruses of all genotypes is increased not only through genetic drift, but also through recombination.

Recombination most often involves strains within a genogroup [25,26], but has also been reported between strains belonging to different genogroups [207,232]. For a number of norovirus strains only ORF1 sequences have been detected with no known 'own' capsid genes. These 'promiscuous' ORF1 genes have been found associated with capsids of different genotypes. Since genotype assignment is based on capsid sequences, no genotype number but a letter has been assigned: GIIa, b and c. Especially strains with GIIb ORF1 have been found consistently in norovirus outbreak surveillance systems since their first detection, in France in 2000 (F. Bon, P. Pothier, C. Hemery, M. Cournot, C. Castor, H. de Valk, P. Gourier-Frery, P. Benhamida, R. Roques, L. Le Coustumier, F. Villeneuve, P. Megraud, P. Le Cann, E. Kohli, and A. Gallay, 21st Réunion Interdisciplinaire de Chimiothérapie Anti-infectieuse, Paris, France, 2001). It is unclear what the 'source' of these non-structural genes has been, as their parent full length viruses have never been detected through human disease surveillance. This may mean that the parent strains did not result in disease, originated from poorly sampled geographic areas, were present at low frequency, or –theoretically- were from another reservoir than humans. However, for such interspecies transmission events there is no evidence at present.

It is difficult to determine how important recombination within genotypes is. Recombination is a mechanism that has been suggested to have contributed to the GII.4 variant diversity. Given the high prevalence of GII.4 strains, and the co-circulation of different variants, especially during the periods of transition from one to the subsequent variant, recombination, requiring a double infection within a patient of GII.4 strains of distinct variants, is likely to occur. Nevertheless, detection of recombination between different variants is difficult, given the relatively high similarities between the GII.4 variants, the propensity of mutations that discern variants from each other to occur preferentially at a limited set of sites and the observation that such mutations regularly go back and forth between a limited range of codons.

The biological importance of recombination events, within or between genotypes or -groups, is unclear. Virulence factors for norovirus are unknown; as discussed previously it is unknown what makes GII.4 more successful than other genotypes. Thus, what part of the genome is relevant for the success, or virulence, of a strain, is unknown, although many - but not all - of the factors that would likely affect success of the virus are governed by the capsid protein (stability, host range, binding avidity, antigenicity), located downstream of the commonly identified recombination site. Thus, an ORF1 of any genotype coupled to an ORF2/3 of GII.4 would likely gain fitness compared to the nonGII.4 parent, but not necessarily compared to the GII.4 parent.

New perspectives on severity of illness

Predominance of GII.4 leads to more severe illness, more hospital infections and more person to person transmission?

Not only does genotype GII.4 cause most infections and outbreaks, it appears that GII.4 causes more severe illness as well. A systematic study of symptoms reported by nursing homes where norovirus outbreaks took place, revealed that attack rates for residents were higher in GII.4 outbreaks, they vomited more often and also suffered from other symptoms more often, compared to outbreaks caused by nonGII.4 viruses [80]. Increased vomiting may contribute to more efficient transmission of the virus [34]. Additionally, our study of 12 years norovirus surveillance in the Netherlands revealed that outbreaks in healthcare settings were more often than expected associated with GII.4 strains, and that GII.4 viruses are more likely to spread via the rapid person to person transmission route (chapter 2).

It should be noted that this 'preference' of the GII.4 viruses may lead to a reporting bias in our surveillance system; hospital and nursing homes are required to report outbreaks of infectious illnesses. The increase in the reported GII.4 norovirus related outbreaks can't be explained by an increasing number of people living in nursing homes, at least not in The Netherlands. The website of Statistics Netherlands reports a 24% decrease in the number of people living in nursing homes between 1995 and 2009 (156.482 to 118.945 persons, www.cbs.nl, accessed 17-02-2010).

Chronic illness

An ever increasing amount of papers reports on chronic norovirus illness establishing in patients with underlying illnesses, including in this thesis [33,84,141,164,180,205,210,266, 328]. Our retrospective study of hospital records [266] showed that over 8% of norovirus positive patients shedding virus for a prolonged period of time. We additionally showed in this thesis that these chronically infected patients, who are usually symptomatic with gastrointestinal illness for prolonged periods of time, are continued sources of infection, as they remain infectious to others (chapter 7).

Excess morbidity and mortality

Whereas different studies have reported on individual cases of deaths resulting from norovirus infections [220,255,305,317], no record is kept of norovirus related deaths. Norovirus is practically never registered as a cause of death, or morbidity, because testing for norovirus is relatively uncommon. Two separate studies in the Netherlands and in England and Wales (this thesis chapter 8 and [110]), using population based databases describing morbidity and mortality reports, and norovirus outbreak reports, gave a much needed insight into the role of norovirus in disease and death in the elderly. Through syndrome

surveillance it was shown that especially during seasons when many norovirus outbreaks occur, excess morbidity and mortality resulting of norovirus infections occurs. Additionally, recent research suggests that the use of the drug statin, commonly used by persons at risk for more severe norovirus infections to lower cholesterol levels, may be a risk factor for contracting norovirus disease [255]. Combined with the chronic illness described above, and the great impact outbreaks may have in hospitals or other (healthcare) settings, the impact of norovirus related illness is bigger than was previously assumed.

Future perspectives

Increasing awareness and a need for internationalization of surveillance

So, for reasons not yet entirely understood, the number of reported outbreaks in many parts of the world has increased steeply since 2002 [23,121,268,330]. Our syndrome surveillance study of norovirus illness showed that simultaneously with this rise both morbidity and mortality due to norovirus have increased (this thesis: chapter 8), observations that are supported by anecdotal reports from e.g., the nursing staff of hospitals and daycare centers. Such findings indicate that norovirus illness has become more severe during the past decade, likely resulting from changes in the predominant norovirus GII.4 strain.

Studying surveillance data, using combined virological, lab-generated data and epidemiological background has taught us much about noroviruses infectivity and epidemiology. Especially in view of the lacking of possibilities to culture norovirus, the importance of long running systematic surveillance can't be overestimated. Comparison of local surveillance datasets has shown that circulating strains are generally highly uniform across the world, with the globally epidemic GII.4 strains predominant [270]. Less highly prevalent strains are similarly spread globally, such as the GII.2 genotype [122]. These observations indicate that norovirus illness and epidemiology should be regarded as an international global health problem. Yet, norovirus surveillance data have been difficult to compare among countries because case- and outbreak definitions and molecular detection techniques are not standardized [154]. PCR-based or EIA detection techniques vary in their abilities to detect different norovirus genotypes. Also, different population subgroups, for example children, elderly or hospital populations, are targeted as a result of local differences in reporting routes, e.g., directly to a central laboratories, through local laboratories or municipal health services, and differences in priorities, e.g., focusing on food-related outbreaks or health care related outbreaks or aiming to provide coverage of the complete population. This is a common problem, also known for other multinational disease reporting networks (http://www.ecdc.europa.eu/en/Publications/AER_report.aspx) [8], which will hopefully be overcome by increasing the effectiveness of international collaboration schemes. The initiation of several successful multi-institute and international

joint surveillance efforts in the past few years has been an encouraging development towards this goal (www.fbve.nl, www.noronet.nl and ozfoodnet.org.au).
At the same time, at the national level more and more regional laboratories have or are developing the techniques and the equipment to perform decentralized diagnostic testing. Although this constitutes a reduction of the workload for the central laboratories, and perhaps results may be generated more quickly, this development may form a threat to the maintenance of national databases. Regional labs may not have a need for genotyping data, or may not report their data to central labs, unless there is a mandate for this.

Clarifying nomenclature to streamline communication about outbreak strains
An important reason for assigning noroviruses to genotypes, and for some genotypes even to the more detailed variants, is that these names simplify communication about outbreak strains and epidemiology, and thus tracking of the spread of the virus. In recent years, as more has become known about norovirus molecular epidemiology, the role of recombination among noroviruses has been increasingly recognized. For surveillance purposes, usually only a small fragment of the norovirus genome is typed, located on either side of the main recombination hot spot (on the ORF1-2 junction [25,26]); in either ORF1 (the polymerase) or in ORF2 (the capsid gene). To promote clear communication with regards to detected strains, we suggest that double typing should be adopted in norovirus nomenclature. When both the ORF1 and ORF2 genotypes are known, they can both be indicated, and when only one region has been typed, information might be added to the typing result stating which region, or part of which ORF, was sequenced.
An ongoing discussion is what constitutes a variant, or how to define one. In line with the definitions for genogroups and genotypes, it could be natural to define the variants based on genetic divergence cut-off rules. We share the concern expressed by Bok *et al.* [18], that it is tricky to base a scheme to identify genetic markers for discerning cluster transitions between variants on limited sets of sequences. The tendency of such genetic markers is to 'switch back and forth' between a limited set of nucleotides, or amino acids (as shown in chapter 3 and 5) because secondary RNA structures, or tertiary protein structures only allow for a limited set of nucleotides or amino acids for certain sites. We strongly advocate to use a sequence set that is as complete as possible, as Bok does, but moreover, to include knowledge of epidemiology in the consideration, especially while we still can't surely point out which residues play a role in defining antigenicity of the variants. For that reason we strongly argue against the suggestion made by Bok *et al.* that the 1996 and 2002 variants (or the Grimsby and Farmington Hills clusters) have wrongly been separated into two different clusters, merely because a large quantity of sequences were available from each of these clusters; the epidemic peak observed globally in the winter of 2002-2003 resulted of the emergence of an antigenically different variant. Similarly, although the 2004 and 2006a

variant are relatively closely related at the genetic level, the epidemiology of the viruses shows that the 2006a variant belongs to a different cluster.
A difficulty in using phylogenetics as a starting point for finding new variants, is that strains that are present in low quantities in the alignment, as will initially be the case for newly emerging variants, tend to 'disappear' in other clusters. Until more is known about sites in the virus that determine antigenicity, we therefore suggest a combination of our other knowledge is used; phylogenetic analyses of the full capsid sequence, and assessment of the amino acid sites identified in chapter 5 of this thesis (also listed below), ánd epidemiological data.

Noroviruses as a proxy for other infectious agents
The highly infectious nature of norovirus combined with its high prevalence and relative mild disease make it a unique proxy for studying the spread of viruses or other infectious agents. People working in the control of infectious diseases are quite aware of the threat of possibly highly virulent emerging pathogens, and are continuously vigilant in preventing their spread. The ease with which norovirus has shown to spread through contaminated foodstuffs, or the difficulties we are confronted with in controlling large (hospital) outbreaks, are clear signs that we may face serious problems when a pathogens emerges with similar infectiousness but higher virulence. Norovirus and norovirus outbreaks can be used in studies assessing the effectiveness of control measures.

Prevention: Hygiene measures
Since no therapeutic drug or vaccine for curing norovirus infection, or preventing it, is currently available, the only real preventive measures we can take at the moment lie in the increasing of hygiene. By doing so, the person to person spread of the virus can be limited [79,96,112]. Control measures were shown to be most effective when they were implemented within three days after onset of disease of the first patient [79]. Rapid laboratory confirmation of norovirus as the etiological agent of an outbreak is essential for that.

Prevention: Therapeutic drugs


Although no drugs to treat norovirus infections are available, work is being done to develop such therapeutics, which would be of special importance for patients with immune disorders, who may contract chronic norovirus infections, and who may not benefit from future vaccines. At present treatment is limited to rehydration by replacement of fluids and electrolytes, and when necessary antisecretory and antimotility agents. Work is being done to develop methods for identifying antivirals that may help clear norovirus infections. Using norovirus replicon-bearing cells, ribavirin and interferons were identified as good therapeutic options for treating norovirus infections [198].

Prevention: Vaccine development

While norovirus infections are generally not life threatening, the high infection rates, the rapidity of spread and the debilitating symptoms, which sometimes necessitate closure of the outbreak environment for disinfection purposes, make norovirus an interesting target for vaccine producing companies. At a number of locations work is being done towards the development of a vaccine, and at least one company has a cellculture derived VLP-based vaccine in clinical trial phase (http://www.ligocyte.com/downloads/Noro.pdf, accessed 28-1-2010). Obvious hurdles in the development of an effective vaccine are the high genetic and antigenic diversity of the virus, the rapid turnover of the dominant GII.4 variants, and the poor immune response to norovirus infections. The general approach is to include multiple genotypes and variants in the vaccine, and to include the possibility to update the vaccine with newly identified variants. Nonetheless, when real infections already produce a poor, short-lived immune response, it is challenging to get a proper response using a non infectious vaccine, boostered by an adjuvant.

As with the new rotavirus vaccines, that protect against one or a limited subset of the (most important) circulating strains, a risk lies in creating a niche that may be filled by other, currently less fit genotypes or variants. Bok *et al.* showed that before the 1980s, or perhaps even later, the GII.4 strains were not dominant, but strains belonging to genotype GII.3 were detected more often. As GII.4 as a genotype has shown us, it is very well possible for a genotype to evolve to a highly successful clade of viruses within a matter of years. The GII.3 viruses may well become dominant again, as may any other genotype of norovirus.

Prevention: Prediction of epidemics

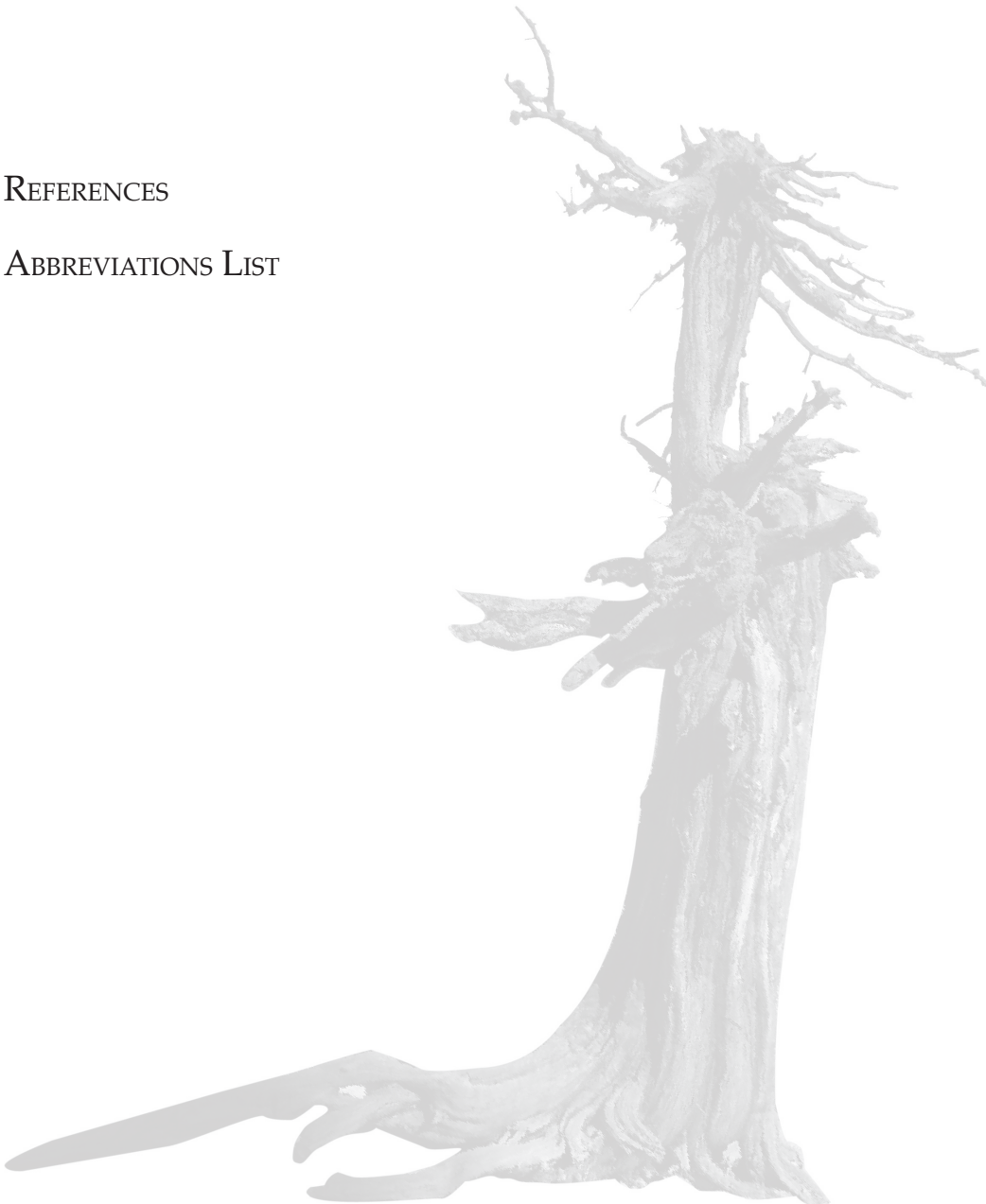
Having an international surveillance system in place is an important step towards enabling better prediction of upcoming epidemics. Changing molecular epidemiology, the shift to a new predominant GII.4 variant, or perhaps another genotype, can be detected more quickly with larger datasets. Using the knowledge of the amino acid sites identified in this thesis (chapter 5), strains with epidemic potential can be identified early on, and their spread can be easily assessed. The sites to look for are amino acids 6, 9, 294, 296-297-298, 333, 352, 368, 372, 393-394-395, 407, 534, in the capsid protein, where a potential new epidemic variant should differ at a number of these sites from the previously circulating one (figure 5.11). Possibly, these sites can be updated when in the future we have more detailed information describing neutralizing epitopes on the norovirus partcle. With better surveillance coverage, especially in Asia, Africa, and South America, and (semi-) real-time linked databases, future studies may be able to pinpoint geographic origins of epidemic strains. Additionally, the Southern hemisphere countries form a good mirror for the Northern hemisphere countries, with regards to the amount of norovirus activity; when the surveillance systems in New Zealand and Australia report increased numbers of norovirus outbreaks, this has in the past

generally meant that the number of outbreaks in the ensuing Northern hemisphere winter was likewise elevated. Reliable prediction of epidemic norovirus winters can contribute to containment and prevention of outbreaks, by enabling preparation for potential outbreaks, and informing relevant authorities and institutes.

Other reasons to keep noroviruses under close watch

Viruses closely related to norovirus, namely Feline Calicivirus and Rabbit Hemorrhagic Disease Virus (RHDV), belonging to the genera *Vesiviruses* and *Lagoviruses*, are known to cause severe illnesses in their hosts. Feline Calicivirus normally causes ulceration on the mouth of the infected animal, but occasionally outbreaks of virulent systemic disease (VSD) occur. In these cases the virus seems to expand its cell tropism, infecting many tissues. Genetic analysis of VSD outbreak strains suggest that the genetic mutations required for causing severe disease arise independently in each outbreak, although no virulence markers have yet been identified [3]. For RHDV it was recently shown that the high virulence likely evolved once, and not on separate occasions [139]. Nonetheless, the notion that viruses closely related to noroviruses may dramatically increase in virulence, by mutation of a limited number of amino acids, is worrying. The more distant West Nile Virus needs but one amino acid change for increased virulence [21]; in the light of the high prevalence and mutation accumulation rate of noroviruses, this should incur the notion that vigilant surveillance of noroviruses and more research in the virulence factors of RHDV, FCV and human noroviruses are called for.

The research in this thesis adds to our knowledge of norovirus molecular epidemiology. The results may contribute to targeting our efforts to control and minimize the impact of norovirus illness, attempts which are validated by our new understanding that norovirus can't always be considered a 'mild illness', especially not among the elderly and hospitalized people. Population surveillance remains vital to monitor which strains circulate, and to continue to generate questions and hypotheses for further research.



1. **Norovirus activity--United States, 2002.** MMWR Morb Mortal Wkly Rep 2003 Jan 24;52(3):41-5.

2. **Norovirus activity--United States, 2006-2007.** MMWR Morb Mortal Wkly Rep 2007 Aug 24;56(33):842-6.

3. Abd-Eldaim M, Potgieter L, Kennedy M. **Genetic analysis of feline caliciviruses associated with a hemorrhagic-like disease.** J Vet Diagn Invest 2005 Sep;17(5):420-9.

4. Adler JL, Zickl R. **Winter vomiting disease.** J Infect Dis 1969 Jun;119(6):668-73.

5. Agus SG, Dolin R, Wyatt RG, Tousimis AJ, Northrup RS. **Acute infectious nonbacterial gastroenteritis: intestinal histopathology. Histologic and enzymatic alterations during illness produced by the Norwalk agent in man.** Ann Intern Med 1973 Jul;79(1):18-25.

6. Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-Gomara M. **Analysis of Amino Acid Variation in the P2 Domain of the GII-4 Norovirus VP1 Protein Reveals Putative Variant-Specific Epitopes.** PLoS ONE 2008;3(1):e1485.

7. Allwood PB, Malik YS, Hedberg CW, Goyal SM. **Effect of temperature and sanitizers on the survival of feline calicivirus, Escherichia coli, and F-specific coliphage MS2 on leafy salad vegetables.** J Food Prot 2004 Jul;67(7):1451-6.

8. Amato-Gauci A, Ammon A. **ECDC to launch first report on communicable diseases epidemiology in the European Union.** Euro Surveill 2007 Jun;12(6):E070607.2.

9. Ando T, Noel JS, Fankhauser RL. **Genetic classification of „Norwalk-like viruses.** J Infect Dis 2000 May;181 Suppl 2:S336-48.

10. Armah GE, Gallimore CI, Binka FN, Asmah RH, Green J, Ugoji U, Anto F, Brown DW, Gray JJ. **Characterisation of norovirus strains in rural Ghanaian children with acute diarrhoea.** J Med Virol 2006 Nov;78(11):1480-5.

11. Atmar RL, Estes MK. **The epidemiologic and clinical importance of norovirus infection.** Gastroenterol Clin North Am 2006 Jun;35(2):275-90, viii.

12. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, Graham DY. **Norwalk virus shedding after experimental human infection.** Emerg Infect Dis 2008 Oct;14(10):1553-7.

13. Beersma MF, Schutten M, Vennema H, Hartwig NG, Mes TH, Osterhaus AD, van Doornum GJ, Koopmans M. **Norovirus in a Dutch tertiary care hospital (2002-2007): frequent nosocomial transmission and dominance of GIIb strains in young children.** J Hosp Infect 2009 Jan 13.

14. Belliot G, Lavaux A, Souihel D, Agnello D, Pothier P. **Use of murine norovirus as a surrogate to evaluate resistance of human norovirus to disinfectants.** Appl Environ Microbiol 2008 May;74(10):3315-8.

15. Bertolotti-Ciarlet A, Crawford SE, Hutson AM, Estes MK. **The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein.** J Virol 2003 Nov;77(21):11603-15.

16. Billgren M, Christenson B, Hedlund KO, Vinje J. **Epidemiology of Norwalk-like human caliciviruses in hospital outbreaks of acute gastroenteritis in the Stockholm area in 1996.** J Infect 2002 Jan;44(1):26-32.

17. Blanton LH, Adams SM, Beard RS, Wei G, Bulens SN, Widdowson MA, Glass RI, Monroe SS. **Molecular and Epidemiologic Trends of Caliciviruses Associated with Outbreaks of Acute Gastroenteritis in the United States, 2000-2004.** J Infect Dis 2006 Feb 1;193(3):413-21.

18. Bok K, Abente EJ, Realpe-Quintero M, Mitra T, Sosnovtsev SV, Kapikian AZ, Green KY. **Evolutionary Dynamics of GII.4 Noroviruses over a Thirty-four Year Period.** J Virol 2009 Sep 16.

19. Bon F, Ambert-Balay K, Giraudon H, Kaplon J, Le Guyader S, Pommepuy M, Gallay A, Vaillant V, de Valk H, Chikhi-Brachet R, et al. **Molecular epidemiology of caliciviruses detected in sporadic and outbreak cases of gastroenteritis in france from december 1998 to february 2004.** J Clin Microbiol 2005 Sep;43(9):4659-64.

20. Brandt HE, Deliens L, van der Steen JT, Ooms ME, Ribbe MW, van der Wal G. **The last days of life of nursing home patients with and without dementia assessed with the palliative care outcome scale.** Palliat Med 2005 Jun;19(4):334-42.

21. Brault AC, Huang CY, Langevin SA, Kinney RM, Bowen RA, Ramey WN, Panella NA, Holmes EC, Powers AM, Miller BR. **A single positively selected West Nile viral mutation confers increased virogenesis in American crows.** Nat Genet 2007 Sep;39(9):1162-6.

22. Bruen TC, Philippe H, Bryant D. **A simple and robust statistical test for detecting the presence of recombination.** Genetics 2006 Apr;172(4):2665-81.

23. Bruggink L, Marshall J. **Molecular changes in the norovirus polymerase gene and their association with incidence of GII.4 norovirus-associated gastroenteritis outbreaks in Victoria, Australia, 2001-2005.** Arch Virol 2008 Jan 29.

24. Bucardo F, Nordgren J, Carlsson B, Paniagua M, Lindgren PE, Espinoza F, Svensson L. **Pediatric norovirus diarrhea in Nicaragua.** J Clin Microbiol 2008 Aug;46(8):2573-80.

25. Bull RA, Hansman GS, Clancy LE, Tanaka MM, Rawlinson WD, White PA. **Norovirus recombination in ORF1/ORF2 overlap.** Emerg Infect Dis 2005 Jul;11(7):1079-85.

26. Bull RA, Tanaka MM, White PA. **Norovirus recombination.** J Gen Virol 2007 Dec;88(Pt 12):3347-59.

27. Bull RA, Tu ET, McIver CJ, Rawlinson WD, White PA. **Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis.** J Clin Microbiol 2006 Feb;44(2):327-33.

28. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. **Predicting the evolution of human influenza A.** Science 1999 Dec 3;286(5446):1921-5.

29. Calderon-Margalit R, Sheffer R, Halperin T, Orr N, Cohen D, Shohat T. **A large-scale gastroenteritis outbreak associated with Norovirus in nursing homes.** Epidemiol Infect 2005 Feb;133(1):35-40.

30. Cannon JL, Lindesmith LC, Donaldson EF, Saxe L, Baric RS, Vinje J. **Herd immunity to GII.4 noroviruses is supported by outbreak patient sera.** J Virol 2009 Jun;83(11):5363-74.

31. Cannon JL, Papafragkou E, Park GW, Osborne J, Jaykus LA, Vinje J. **Surrogates for the study of norovirus stability and inactivation in the environment: A comparison of murine norovirus and feline calicivirus.** J Food Prot 2006 Nov;69(11):2761-5.

32. Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, Zhang XC, Jiang X, Li X, Rao Z. **Structural basis for the recognition of blood group trisaccharides by norovirus.** J Virol 2007 Jun;81(11):5949-57.

33. Carlsson B, Lindberg AM, Rodriguez-Diaz J, Hedlund KO, Persson B, Svensson L. **Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection.** J Gen Virol 2009 Feb;90(Pt 2):432-41.

34. Caul EO. **Hyperemesis hiemis--a sick hazard.** J Hosp Infect 1995 Jun;30 Suppl:498-502.

35. CDC-WHO. **History and Epidemiology of Global Smallpox Eradication**. [Web Page]; http://www.bt.cdc.gov/agent/smallpox/training/overview/pdf/eradicationhistory.pdf. [Accessed 29 Apr 2009].

36. Chakravarty S, Hutson AM, Estes MK, Prasad BV. **Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity.** J Virol 2005 Jan;79(1):554-68.

37. Chan MC, Sung JJ, Lam RK, Chan PK, Lee NL, Lai RW, Leung WK. **Fecal viral load and norovirus-associated gastroenteritis.** Emerg Infect Dis 2006 Aug;12(8):1278-80.

38. Cheetham S, Souza M, McGregor R, Meulia T, Wang Q, Saif LJ. **Binding patterns of human norovirus-like particles to buccal and intestinal tissues of gnotobiotic pigs in relation to A/H histo-blood group antigen expression.** J Virol 2007 Apr;81(7):3535-44.

39. Cheetham S, Souza M, Meulia T, Grimes S, Han MG, Saif LJ. **Pathogenesis of a genogroup II human**

norovirus in gnotobiotic pigs. J Virol 2006 Nov;80(21):10372-81.

40. Chen R, Neill JD, Estes MK, Prasad BV. **X-ray structure of a native calicivirus: Structural insights into antigenic diversity and host specificity.** Proc Natl Acad Sci U S A 2006 May 15.

41. Chen R, Neill JD, Noel JS, Hutson AM, Glass RI, Estes MK, Prasad BV. **Inter- and intragenus structural variations in caliciviruses and their functional implications.** J Virol 2004 Jun;78(12):6469-79.

42. Clarke IN, Lambden PR. **Organization and expression of calicivirus genes.** J Infect Dis 2000 May;181 Suppl 2:S309-16.

43. Clay S, Maherchandani S, Malik YS, Goyal SM. **Survival on uncommon fomites of feline calicivirus, a surrogate of noroviruses.** Am J Infect Control 2006 Feb;34(1):41-3.

44. Colomba C, De Grazia S, Giammanco GM, Saporito L, Scarlata F, Titone L, Arista S. **Viral gastroenteritis in children hospitalised in Sicily, Italy.** Eur J Clin Microbiol Infect Dis 2006 Aug 18.

45. Colomba C, Saporito L, Giammanco GM, De Grazia S, Ramirez S, Arista S, Titone L. **Norovirus and gastroenteritis in hospitalized children, Italy.** Emerg Infect Dis 2007 Sep;13(9):1389-91.

46. Coppo P, Scieux C, Ferchal F, Clauvel J, Lassoued K. **Astrovirus enteritis in a chronic lymphocytic leukemia patient treated with fludarabine monophosphate.** Ann Hematol 2000 Jan;79(1):43-5.

47. D'Souza SE, Ginsberg MH, Plow EF. **Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif.** Trends Biochem Sci 1991 Jul;16(7):246-50.

48. G. Daniels. **Human Blood Groups.** 2002.

49. de Bruin E, Duizer E, Vennema H, Koopmans MP. **Diagnosis of Norovirus outbreaks by commercial ELISA or RT-PCR.** J Virol Methods 2006 Nov;137(2):259-64.

50. de Graaf M, Osterhaus AD, Fouchier RA, Holmes EC. **Evolutionary dynamics of human and avian metapneumoviruses.** J Gen Virol 2008 Dec;89(Pt 12):2933-42.

51. De Roda Husman AM, Bijkerk P, Lodder W, Van Den Berg H, Pribil W, Cabaj A, Gehringer P, Sommer R, Duizer E. **Calicivirus inactivation by nonionizing (253.7-nanometer-wavelength.** Appl Environ Microbiol 2004 Sep;70(9):5089-93.

52. de Wit MA, Koopmans MP, Kortbeek LM, van Leeuwen NJ, Bartelds AI, van Duynhoven YT. **Gastroenteritis in sentinel general practices,The Netherlands.** Emerg Infect Dis 2001 Jan-2001 Feb 28;7(1):82-91.

53. de Wit MA, Koopmans MP, Kortbeek LM, Wannet WJ, Vinje J, van Leusden F, Bartelds AI, van Duynhoven YT. **Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology.** Am J Epidemiol 2001 Oct 1;154(7):666-74.

54. de Wit MA, Koopmans MP, van der Blij JF, van Duynhoven YT. **Hospital admissions for rotavirus infection in the Netherlands.** Clin Infect Dis 2000 Sep;31(3):698-704.

55. de Wit MA, Kortbeek LM, Koopmans MP, de Jager CJ, Wannet WJ, Bartelds AI, van Duynhoven YT. **A comparison of gastroenteritis in a general practice-based study and a community-based study.** Epidemiol Infect 2001 Dec;127(3):389-97.

56. de Wit MA, Widdowson MA, Vennema H, de Bruin E, Fernandes T, Koopmans M. **Large outbreak of norovirus: The baker who should have known better.** J Infect 2007 Jun 27.

57. Delport W, Scheffler K, Seoighe C. **Models of coding sequence evolution.** Brief Bioinform 2009 Jan;10(1):97-109.

58. Dolin R, Levy AG, Wyatt RG, Thornhill TS, Gardner JD. **Viral gastroenteritis induced by the Hawaii agent. Jejunal histopathology and serologic response.** Am J Med 1975 Dec;59(6):761-8.

59. Donaldson EF, Lindesmith LC, Lobue AD, Baric RS. **Norovirus pathogenesis: mechanisms of**

persistence and immune evasion in human populations. Immunol Rev 2008 Oct;225(1):190-211.

60. Dove W, Cunliffe NA, Gondwe JS, Broadhead RL, Molyneux ME, Nakagomi O, Hart CA. **Detection and characterization of human caliciviruses in hospitalized children with acute gastroenteritis in Blantyre, Malawi.** J Med Virol 2005 Dec;77(4):522-7.

61. Doyle TJ, Stark L, Hammond R, Hopkins RS. **Outbreaks of noroviral gastroenteritis in Florida, 2006-2007.** Epidemiol Infect 2008 Apr 23;1-9.

62. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. **Relaxed phylogenetics and dating with confidence.** PLoS Biol 2006 May;4(5):e88.

63. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees.** BMC Evol Biol 2007;7:214.

64. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. **Bayesian coalescent inference of past population dynamics from molecular sequences.** Mol Biol Evol 2005 May;22(5):1185-92.

65. Drummond AJ, Suchard MA. **Fully Bayesian tests of neutrality using genealogical summary statistics.** BMC Genet 2008;9:68.

66. Duizer E, Bijkerk P, Rockx B, De Groot A, Twisk F, Koopmans M. **Inactivation of caliciviruses.** Appl Environ Microbiol 2004 Aug;70(8):4538-43.

67. Duizer, E.; Koopmans, M. **Tracking foodborne viruses: lessons from noroviruses.** Emerging Foodborne Pathogens.CRC Press; 2006. pp. p. 77-110.

68. Duizer E, Pielaat A, Vennema H, Kroneman A, Koopmans M. **Probabilities in norovirus outbreak diagnosis.** J Clin Virol 2007 Sep;40(1):38-42.

69. Duizer E, Schwab KJ, Neill FH, Atmar RL, Koopmans MP, Estes MK. **Laboratory efforts to cultivate noroviruses.** J Gen Virol 2004 Jan;85(Pt 1):79-87.

70. Edgar RC. **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** BMC Bioinformatics 2004 Aug 19;5:113.

71. Estes MK, Prasad BV, Atmar RL. **Noroviruses everywhere: has something changed?** Curr Opin Infect Dis 2006 Oct;19(5):467-74.

72. Evans MR, Meldrum R, Lane W, Gardner D, Ribeiro CD, Gallimore CI, Westmoreland D. **An outbreak of viral gastroenteritis following environmental contamination at a concert hall.** Epidemiol Infect 2002 Oct;129(2):355-60.

73. Falkenhorst G, Krusell L, Lisby M, Madsen SB, Bottiger B, Molbak K. **Imported frozen raspberries cause a series of norovirus outbreaks in Denmark, 2005.** Euro Surveill 2005 Sep;10(9):E050922.2.

74. Fankhauser RL, Monroe SS, Noel JS, Humphrey CD, Bresee JS, Parashar UD, Ando T, Glass RI. **Epidemiologic and molecular trends of "Norwalk-like viruses" associated with outbreaks of gastroenteritis in the United States.** J Infect Dis 2002 Jul 1;186(1):1-7.

75. Fankhauser RL, Noel JS, Monroe SS, Ando T, Glass RI. **Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United States.** J Infect Dis 1998 Dec;178(6):1571-8.

76. Farkas T, Nakajima S, Sugieda M, Deng X, Zhong W, Jiang X. **Seroprevalence of noroviruses in swine.** J Clin Microbiol 2005 Feb;43(2):657-61.

77. Farkas T, Sestak K, Wei C, Jiang X. **Characterization of a rhesus monkey calicivirus representing a new genus of Caliciviridae.** J Virol 2008 Jun;82(11):5408-16.

78. Frenzen PD. **Deaths due to unknown foodborne agents.** Emerg Infect Dis 2004 Sep;10(9):1536-43.

79. Friesema IH, Vennema H, Heijne JC, de Jager CM, Morroy G, van den Kerkhof JH, de Coster EJ, Wolters BA, Ter Waarbeek HL, Fanoy EB, et al. **Norovirus outbreaks in nursing homes: the evaluation of**

infection control measures. Epidemiol Infect 2009 May 11;1-12.

80. Friesema IH, Vennema H, Heijne JC, de Jager CM, Teunis PF, van der Linde R, Duizer E, van Duynhoven YT. **Differences in clinical presentation between norovirus genotypes in nursing homes.** J Clin Virol 2009 Dec;46(4):341-4.

81. Gallimore CI, Cheesbrough JS, Lamden K, Bingham C, Gray JJ. **Multiple norovirus genotypes characterised from an oyster-associated outbreak of gastroenteritis.** Int J Food Microbiol 2005 Jun 18.

82. Gallimore CI, Green J, Lewis D, Richards AF, Lopman BA, Hale AD, Eglin R, Gray JJ, Brown DW. **Diversity of noroviruses cocirculating in the north of England from 1998 to 2001.** J Clin Microbiol 2004 Apr;42(4):1396-401.

83. Gallimore CI, Iturriza-Gomara M, Xerry J, Adigwe J, Gray JJ. **Inter-seasonal diversity of norovirus genotypes: Emergence and selection of virus variants.** Arch Virol 2007 Jul;152(7):1295-303.

84. Gallimore CI, Lewis D, Taylor C, Cant A, Gennery A, Gray JJ. **Chronic excretion of a norovirus in a child with cartilage hair hypoplasia (CHH).** J Clin Virol 2004 Jun;30(2):196-204.

85. Gallimore CI, Taylor C, Gennery AR, Cant AJ, Galloway A, Iturriza-Gomara M, Gray JJ. **Environmental monitoring for gastroenteric viruses in a pediatric primary immunodeficiency unit.** J Clin Microbiol 2006 Feb;44(2):395-9.

86. Gallimore CI, Taylor C, Gennery AR, Cant AJ, Galloway A, Xerry J, Adigwe J, Gray JJ. **Contamination of the Hospital Environment with Gastroenteric Viruses: Comparison of Two Paediatric Wards over a Winter Season.** J Clin Microbiol 2008 Jul 9.

87. Gangarosa RE, Glass RI, Lew JF, Boring JR. **Hospitalizations involving gastroenteritis in the United States, 1985: the special burden of the disease among the elderly.** Am J Epidemiol 1992 Feb 1;135(3):281-90.

88. Gay NJ, Andrews NJ, Trotter CL, Edmunds WJ. **Estimating deaths due to influenza and respiratory syncytial virus.** JAMA 2003 May 21;289(19):2499; author reply 2500-2.

89. Gehrke C, Steinmann J, Goroncy-Bermes P. **Inactivation of feline calicivirus, a surrogate of norovirus (formerly Norwalk-like viruses), by different types of alcohol in vitro and in vivo.** J Hosp Infect 2004 Jan;56(1):49-55.

90. Glass PJ, White LJ, Ball JM, Leparc-Goffart I, Hardy ME, Estes MK. **Norwalk virus open reading frame 3 encodes a minor structural protein.** J Virol 2000 Jul;74(14):6581-91.

91. Gomes KA, Stupka JA, Gomez J, Parra GI. **Molecular characterization of calicivirus strains detected in outbreaks of gastroenteritis in Argentina.** J Med Virol 2007 Nov;79(11):1703-9.

92. Gray JJ, Kohli E, Ruggeri FM, Vennema H, Sanchez-Fauquier A, Schreier E, Gallimore CI, Iturriza-Gomara M, Giraudon H, Pothier P, et al. **European multicenter evaluation of commercial enzyme immunoassays for detecting norovirus antigen in fecal samples.** Clin Vaccine Immunol 2007 Oct;14(10):1349-55.

93. Green, K. Y. **Caliciviridae: The noroviruses.** Fields Virology 5 ed. Philadelphia: Lippincott Williams & Wilkins ; 2007. pp. p. 949-80.

94. Green, K. Y.; Chanock, R. M.; Kapikian, A. Z. **Human Caliciviruses.** Fields Virology 4 ed. Philadelphia: Lippincott Williams & Wilkins; 2001. pp. 841-74.

95. Greer AL, Drews SJ, Fisman DN. **Why "Winter" Vomiting Disease? Seasonality, Hydrology, and Norovirus Epidemiology in Toronto, Canada.** Ecohealth 2010 Feb 12.

96. Greig JD, Lee MB. **Enteric outbreaks in long-term care facilities and recommendations for prevention: a review.** Epidemiol Infect 2008 May 12;1-11.

97. Grenfell BT, Bjornstad ON, Kappey J. **Travelling waves and spatial hierarchies in measles epidemics.** Nature 2001 Dec 13;414(6865):716-23.

98. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. **Unifying the epidemiological and evolutionary dynamics of pathogens.** Science 2004 Jan 16;303(5656):327-32.

99. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. **The Vienna RNA websuite.** Nucleic Acids Res 2008 Jul 1;36(Web Server issue):W70-4.

100. Guindon S, Gascuel O. **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** Syst Biol 2003 Oct;52(5):696-704.

101. Gulati BR, Allwood PB, Hedberg CW, Goyal SM. **Efficacy of commonly used disinfectants for the inactivation of calicivirus on strawberry, lettuce, and a food-contact surface.** J Food Prot 2001 Sep;64(9):1430-4.

102. Halperin T, Yavzori M, Amitai A, Klement E, Kayouf R, Grotto I, Huerta M, Hadley LA, Monroe SS, Cohen D, et al. **Molecular analysis of noroviruses involved in acute gastroenteritis outbreaks in military units in Israel, 1999-2004.** Eur J Clin Microbiol Infect Dis 2005 Oct;24(10):697-700.

103. Hamano M, Kuzuya M, Fujii R, Ogura H, Yamada M. **Epidemiology of acute gastroenteritis outbreaks caused by Noroviruses in Okayama, Japan.** J Med Virol 2005 Oct;77(2):282-9.

104. Han MG, Smiley JR, Thomas C, Saif LJ. **Genetic recombination between two genotypes of genogroup III bovine noroviruses (BoNVs) and capsid sequence diversity among BoNVs and Nebraska-like bovine enteric caliciviruses.** J Clin Microbiol 2004 Nov;42(11):5214-24.

105. Hansen S, Stamm-Balderjahn S, Zuschneid I, Behnke M, Ruden H, Vonberg RP, Gastmeier P. **Closure of medical departments during nosocomial outbreaks: data from a systematic analysis of the literature.** J Hosp Infect 2007 Apr;65(4):348-53.

106. Hansman GS, Doan LT, Kguyen TA, Okitsu S, Katayama K, Ogawa S, Natori K, Takeda N, Kato Y, Nishio O, et al. **Detection of norovirus and sapovirus infection among children with gastroenteritis in Ho Chi Minh City, Vietnam.** Arch Virol 2004 Sep;149(9):1673-88.

107. Hansman GS, Natori K, Shirato-Horikoshi H, Ogawa S, Oka T, Katayama K, Tanaka T, Miyoshi T, Sakae K, Kobayashi S, et al. **Genetic and antigenic diversity among noroviruses.** J Gen Virol 2006 Apr;87(Pt 4):909-19.

108. Hansman GS, Oka T, Li TC, Nishio O, Noda M, Takeda N. **Detection of human enteric viruses in Japanese clams.** J Food Prot 2008 Aug;71(8):1689-95.

109. Harrington PR, Vinje J, Moe CL, Baric RS. **Norovirus capture with histo-blood group antigens reveals novel virus-ligand interactions.** J Virol 2004 Mar;78(6):3035-45.

110. Harris JP, Edmunds WJ, Pebody R, Brown DW, Lopman BA. **Deaths from norovirus among the elderly, England and Wales.** Emerg Infect Dis 2008 Oct;14(10):1546-52.

111. Hedlund KO, Rubilar-Abreu E, Svensson L. **Epidemiology of calicivirus infections in Sweden, 1994-1998.** J Infect Dis 2000 May;181 Suppl 2:S275-80.

112. Heijne JC, Teunis P, Morroy G, Wijkmans C, Oostveen S, Duizer E, Kretzschmar M, Wallinga J. **Enhanced hygiene measures and norovirus transmission during an outbreak.** Emerg Infect Dis 2009 Jan;15(1):24-30.

113. Herbert TP, Brierley I, Brown TD. **Detection of the ORF3 polypeptide of feline calicivirus in infected cells and evidence for its expression from a single, functionally bicistronic, subgenomic mRNA.** J Gen Virol 1996 Jan;77 (Pt 1):123-7.

114. Hewitt J, Bell D, Simmons GC, Rivera-Aban M, Wolf S, Greening GE. **Gastroenteritis outbreak caused by waterborne norovirus at a New Zealand ski resort.** Appl Environ Microbiol 2007 Dec;73(24):7853-7.

115. Ho EC, Cheng PK, Wong DA, Lau AW, Lim WW. **Correlation of norovirus variants with epidemics of acute viral gastroenteritis in Hong Kong.** J Med Virol 2006 Nov;78(11):1473-9.

116. Hoebe CJ, Vennema H, de Roda Husman AM, van Duynhoven YT. **Norovirus outbreak among primary**

schoolchildren who had played in a recreational water fountain. J Infect Dis 2004 Feb 15;189(4):699-705.

117. Hohne M, Schreier E. **Detection and characterization of norovirus outbreaks in Germany: application of a one-tube RT-PCR using a fluorogenic real-time detection system.** J Med Virol 2004 Feb;72(2):312-9.

118. Hsu CC, Riley LK, Livingston RS. **Molecular characterization of three novel murine noroviruses.** Virus Genes 2007 Apr;34(2):147-55.

119. Huang P, Farkas T, Marionneau S, Zhong W, Ruvoen-Clouet N, Morrow AL, Altaye M, Pickering LK, Newburg DS, LePendu J, et al. **Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns.** J Infect Dis 2003 Jul 1;188(1):19-31.

120. Huang P, Farkas T, Zhong W, Tan M, Thornton S, Morrow AL, Jiang X. **Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns.** J Virol 2005 Jun;79(11):6714-22.

121. Hutson AM, Atmar RL, Estes MK. **Norovirus disease: changing epidemiology and host susceptibility factors.** Trends Microbiol 2004 Jun;12(6):279-87.

122. Iritani N, Vennema H, Siebenga JJ, Siezen RJ, Renckens B, Seto Y, Kaida A, Koopmans M. **Genetic analysis of the capsid gene of genotype GII.2 noroviruses.** J Virol 2008 Aug;82(15):7336-45.

123. Ison MG, Gubareva LV, Atmar RL, Treanor J, Hayden FG. **Recovery of drug-resistant influenza virus from immunocompromised patients: a case series.** J Infect Dis 2006 Mar 15;193(6):760-4.

124. Iturriza Gomara M, Simpson R, Perault AM, Redpath C, Lorgelly P, Joshi D, Mugford M, Hughes CA, Dalrymple J, Desselberger U, et al. **Structured surveillance of infantile gastroenteritis in East Anglia, UK: incidence of infection with common viral gastroenteric pathogens.** Epidemiol Infect 2008 Jan;136(1):23-33.

125. Jiang X, Graham DY, Wang KN, Estes MK. **Norwalk virus genome cloning and characterization.** Science 1990 Dec 14;250(4987):1580-3.

126. Jiang X, Wang M, Graham DY, Estes MK. **Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein.** J Virol 1992 Nov;66(11):6527-32.

127. Jiang X, Wang M, Wang K, Estes MK. **Sequence and genomic organization of Norwalk virus.** Virology 1993 Jul;195(1):51-61.

128. Johnson PC, Mathewson JJ, DuPont HL, Greenberg HB. **Multiple-challenge study of host susceptibility to Norwalk gastroenteritis in US adults.** J Infect Dis 1990 Jan;161(1):18-21.

129. Johnston CP, Qiu H, Ticehurst JR, Dickson C, Rosenbaum P, Lawson P, Stokes AB, Lowenstein CJ, Kaminsky M, Cosgrove SE, et al. **Outbreak management and implications of a nosocomial norovirus outbreak.** Clin Infect Dis 2007 Sep 1;45(5):534-40.

130. Kageyama T, Kojima S, Shinohara M, Uchida K, Fukushi S, Hoshino FB, Takeda N, Katayama K. **Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR.** J Clin Microbiol 2003 Apr;41(4):1548-57.

131. Kageyama T, Shinohara M, Uchida K, Fukushi S, Hoshino FB, Kojima S, Takai R, Oka T, Takeda N, Katayama K. **Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to Norovirus in Japan.** J Clin Microbiol 2004 Jul;42(7):2988-95.

132. Kaiser L, Aubert JD, Pache JC, Deffernez C, Rochat T, Garbino J, Wunderli W, Meylan P, Yerly S, Perrin L, et al. **Chronic rhinoviral infection in lung transplant recipients.** Am J Respir Crit Care Med 2006 Dec 15;174(12):1392-9.

133. Kanerva M, Maunula L, Lappalainen M, Mannonen L, von Bonsdorff CH, Anttila VJ. **Prolonged norovirus outbreak in a Finnish tertiary care hospital caused by GII.4-2006b subvariants.** J Hosp Infect 2009 Jan 19.

134. Kapikian AZ. **The discovery of the 27-nm Norwalk virus: an historic perspective.** J Infect Dis 2000 May;181 Suppl 2:S295-302.

135. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. **Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis.** J Virol 1972 Nov;10(5):1075-81.

136. Kaplan JE, Feldman R, Campbell DS, Lookabaugh C, Gary GW. **The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis.** Am J Public Health 1982 Dec;72(12):1329-32.

137. Karst SM, Wobus CE, Lay M, Davidson J, Virgin HW 4th. **STAT1-dependent innate immunity to a Norwalk-like virus.** Science 2003 Mar 7;299(5612):1575-8.

138. Kelly RJ, Ernst LK, Larsen RD, Bryant JG, Robinson JS, Lowe JB. **Molecular basis for H blood group deficiency in Bombay (Oh) and para-Bombay individuals.** Proc Natl Acad Sci U S A 1994 Jun 21;91(13):5843-7.

139. Kerr PJ, Kitchen A, Holmes EC. **Origin and phylodynamics of rabbit hemorrhagic disease virus.** J Virol 2009 Dec;83(23):12129-38.

140. Khamrin P, Takanashi S, Chan-It W, Kobayashi M, Nishimura S, Katsumata N, Okitsu S, Maneekarn N, Nishio O, Ushijima H. **Immunochromatography test for rapid detection of norovirus in fecal specimens.** J Virol Methods 2009 May;157(2):219-22.

141. Kirkwood CD, Streitberg R. **Calicivirus shedding in children after recovery from diarrhoeal disease.** J Clin Virol 2008 Sep 10.

142. Koelle K, Cobey S, Grenfell B, Pascual M. **Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans.** Science 2006 Dec 22;314(5807):1898-903.

143. Koopmans M. **Food-borne norovirus outbreaks: a nuisance or more than that?** Wien Klin Wochenschr 2005 Dec;117(23-24):789-91.

144. Koopmans M, Duizer E. **Foodborne viruses: an emerging problem.** Int J Food Microbiol 2004 Jan 1;90(1):23-41.

145. Koopmans M, Harris J, Verhoef L, Depoortere E, Takkinen J, Coulombier D. **European investigation into recent norovirus outbreaks on cruise ships: update.** Euro Surveill 2006 Jul;11(7):E060706.5.

146. Koopmans M, Vennema H, Heersma H, van Strien E, van Duynhoven Y, Brown D, Reacher M, Lopman B. **Early identification of common-source foodborne virus outbreaks in Europe.** Emerg Infect Dis 2003 Sep;9(9):1136-42.

147. Koopmans M, von Bonsdorff CH, Vinje J, de Medici D, Monroe S. **Foodborne viruses.** FEMS Microbiol Rev 2002 Jun;26(2):187-205.

148. Kosakovsky Pond SL, Frost SD. **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** Mol Biol Evol 2005 May;22(5):1208-22.

149. Kosakovsky Pond SL, Poon AF, Leigh Brown AJ, Frost SD. **A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus.** Mol Biol Evol 2008 Sep;25(9):1809-24.

150. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. **GARD: a genetic algorithm for recombination detection.** Bioinformatics 2006 Dec 15;22(24):3096-8.

151. Kosakovsky Pond SL, Scheffler K, Gravenor MB, Poon AF, Frost SD. **Evolutionary Fingerprinting of Genes.** Mol Biol Evol 2009 Oct 28.

152. Kosek M, Bern C, Guerrant RL. **The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000.** Bull World Health Organ 2003;81(3):197-204.

153. Krisztalovics K, Reuter G, Szucs G, Csohan A, Borocz K. **Increase in norovirus circulation in Hungary**

in October-November 2006. Euro Surveill 2006;11(12):E061214.2.

154. Kroneman A, Harris J, Vennema H, Duizer E, van Duynhoven Y, Gray J, Iturriza M, Bottiger B, Falkenhorst G, Johnsen C, 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 Mar;30(1):82-90.

155. Kroneman A, Vennema H, Harris J, Reuter G, von Bonsdorff CH, Hedlund KO, Vainio K, Jackson V, Pothier P, Koch J, et al. **Increase in norovirus activity reported in Europe.** Euro Surveill 2006 Dec;11(12):E061214.1.

156. Kroneman A, Vennema H, Van Duijnhoven Y, Duizer E, Koopmans M. **High number of norovirus outbreaks associated with a GGII.4 variant in the Netherlands and elsewhere: does this herald a worldwide increase?** Eurosurveillance Weekly 2004 Dec;Eurosurveillance Weekly 2004;8 (52): 041223.

157. Kroneman A, Verhoef L, Harris J, Vennema H, Duizer E, van Duynhoven Y, Gray J, Iturriza M, Bottiger B, Falkenhorst G, et al. **Analysis of integrated virological and epidemiological reports of norovirus outbreaks collected within the foodborne Viruses in Europe network from 1-7-2001 to 30-6-2006.** J Clin Microbiol 2008 Sep;46(9):2959-65.

158. Kuritsky JN, Osterholm MT, Greenberg HB, Korlath JA, Godes JR, Hedberg CW, Forfang JC, Kapikian AZ, McCullough JC, White KE. **Norwalk gastroenteritis: a community outbreak associated with bakery product consumption.** Ann Intern Med 1984 Apr;100(4):519-21.

159. Kuusi M, Aavitsland P, Gondrosen B, Kapperud G. **Incidence of gastroenteritis in Norway--a population-based survey.** Epidemiol Infect 2003 Aug;131(1):591-7.

160. Kuusi M, Nuorti JP, Maunula L, Minh Tran NN, Ratia M, Karlsson J, von Bonsdorff CH. **A prolonged outbreak of Norwalk-like calicivirus (NLV) gastroenteritis in a rehabilitation centre due to environmental contamination.** Epidemiol Infect 2002 Aug;129(1):133-8.

161. L'Homme Y, Sansregret R, Plante-Fortier E, Lamontagne AM, Ouardani M, Lacroix G, Simard C. **Genomic characterization of swine caliciviruses representing a new genus of Caliciviridae.** Virus Genes 2009 Aug;39(1):66-75.

162. Le Guyader F, Loisy F, Atmar RL, Hutson AM, Estes MK, Ruvoen-Clouet N, Pommepuy M, Le Pendu J. **Norwalk virus-specific binding to oyster digestive tissues.** Emerg Infect Dis 2006 Jun;12(6):931-6.

163. Le Guyader FS, Bon F, Demedici D, Parnaudeau S, Bertone A, Crudeli S, Doyle A, Zidane M, Suffredini E, Kohli E, et al. **Detection of multiple noroviruses associated with an international gastroenteritis outbreak linked to oyster consumption.** J Clin Microbiol 2006 Nov;44(11):3878-82.

164. Lee BE, Pang XL, Robinson JL, Bigam D, Monroe SS, Preiksaitis JK. **Chronic norovirus and adenovirus infection in a solid organ transplant recipient.** Pediatr Infect Dis J 2008 Apr;27(4):360-2.

165. Lee N, Chan MC, Wong B, Choi KW, Sin W, Lui G, Chan PK, Lai RW, Cockram CS, Sung JJ, et al. **Fecal viral concentration and diarrhea in norovirus gastroenteritis.** Emerg Infect Dis 2007 Sep;13(9):1399-401.

166. Lemey P, Lott M, Martin DP, Moulton V. **Identifying recombinants in human and primate immunodeficiency virus sequence alignments using quartet scanning.** BMC Bioinformatics 2009 Apr 27;10(1):126.

167. Lew JF, Glass RI, Gangarosa RE, Cohen IP, Bern C, Moe CL. **Diarrheal deaths in the United States, 1979 through 1987. A special problem for the elderly.** JAMA 1991 Jun 26;265(24):3280-4.

168. Lindell AT, Grillner L, Svensson L, Wirgart BZ. **Molecular epidemiology of norovirus infections in Stockholm, Sweden, during the years 2000 to 2003: association of the GGIIb genetic cluster with infection in children.** J Clin Microbiol 2005 Mar;43(3):1086-92.

169. Lindesmith L, Moe C, Lependu J, Frelinger JA, Treanor J, Baric RS. **Cellular and humoral immunity following Snow Mountain virus challenge.** J Virol 2005 Mar;79(5):2900-9.

170. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, Baric RS. **Mechanisms of GII.4**

Norovirus Persistence in Human Populations. PLoS Med 2008 Feb 12;5(2):e31.

171. Liu G, Ni Z, Yun T, Yu B, Chen L, Zhao W, Hua J, Chen J. **A DNA-launched reverse genetics system for rabbit hemorrhagic disease virus reveals that the VP2 protein is not essential for virus infectivity.** J Gen Virol 2008 Dec;89(Pt 12):3080-5.

172. Lochridge VP, Hardy ME. **A single amino acid substitution in the P2 domain of VP1 of murine norovirus is sufficient for escape from antibody neutralization.** J Virol 2007 Sep 5.

173. Lochridge VP, Jutila KL, Graff JW, Hardy ME. **Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions.** J Gen Virol 2005 Oct;86(Pt 10):2799-806.

174. Lopman B, Armstrong B, Atchison C, Gray JJ. **Host, weather and virological factors drive norovirus epidemiology: time-series analysis of laboratory surveillance data in England and Wales.** PLoS One 2009;4(8):e6671.

175. Lopman B, Vennema H, Kohli E, Pothier P, Sanchez A, Negredo A, Buesa J, Schreier E, Reacher M, Brown D, et al. **Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant.** Lancet 2004 Feb 28;363(9410):682-8.

176. Lopman BA, Adak GK, Reacher MH, Brown DW. **Two epidemiologic patterns of norovirus outbreaks: surveillance in England and wales, 1992-2000.** Emerg Infect Dis 2003 Jan;9(1):71-7.

177. Lopman BA, Gallimore C, Gray JJ, Vipond IB, Andrews N, Sarangi J, Reacher MH, Brown DW. **Linking healthcare associated norovirus outbreaks: A molecular epidemiologic method for investigating transmission.** BMC Infect Dis 2006 Jul 11;6(1):108.

178. Lopman BA, Reacher MH, Van Duijnhoven Y, Hanon FX, Brown D, Koopmans M. **Viral gastroenteritis outbreaks in Europe, 1995-2000.** Emerg Infect Dis 2003 Jan;9(1):90-6.

179. Lopman BA, Reacher MH, Vipond IB, Sarangi J, Brown DW. **Clinical manifestation of norovirus gastroenteritis in health care settings.** Clin Infect Dis 2004 Aug 1;39(3):318-24.

180. Ludwig A, Adams O, Laws HJ, Schroten H, Tenenbaum T. **Quantitative detection of norovirus excretion in pediatric patients with cancer and prolonged gastroenteritis and shedding of norovirus.** J Med Virol 2008 Jun 12;80(8):1461-7.

181. Luttermann C, Meyers G. **A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA.** J Biol Chem 2007 Jan 9.

182. Lyman WH, Walsh JF, Kotch JB, Weber DJ, Gunn E, Vinje J. **Prospective Study of Etiologic Agents of Acute Gastroenteritis Outbreaks in Child Care Centers.** J Pediatr 2008 Sep 9.

183. Lynch M, Painter J, Woodruff R, Braden C. **Surveillance for foodborne-disease outbreaks--United States, 1998-2002.** MMWR Surveill Summ 2006 Nov 10;55(10):1-42.

184. Maguire AJ, Green J, Brown DW, Desselberger U, Gray JJ. **Molecular epidemiology of outbreaks of gastroenteritis associated with small round-structured viruses in East Anglia, United Kingdom, during the 1996-1997 season.** J Clin Microbiol 1999 Jan;37(1):81-9.

185. Makary P, Maunula L, Niskanen T, Kuusi M, Virtanen M, Pajunen S, Ollgren J, Tran Minh NN. **Multiple norovirus outbreaks among workplace canteen users in Finland, July 2006.** Epidemiol Infect 2008 Apr 4;1-6.

186. Malek M, Barzilay E, Kramer A, Camp B, Jaykus LA, Escudero-Abarca B, Derrick G, White P, Gerba C, Higgins C, et al. **Outbreak of Norovirus Infection among River Rafters Associated with Packaged Delicatessen Meat, Grand Canyon, 2005.** Clin Infect Dis 2009 Jan 1;48(1):31-7.

187. Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoen N, Clement M, Le Pendu J. **ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world.** Biochimie 2001 Jul;83(7):565-73.

188. Marks PJ, Vipond IB, Carlisle D, Deakin D, Fey RE, Caul EO. **Evidence for airborne transmission of**

Norwalk-like virus (NLV) in a hotel restaurant. Epidemiol Infect 2000 Jun;124(3):481-7.

189. Marshall JA, Bruggink LD. **Laboratory diagnosis of norovirus.** Clin Lab 2006;52(11-12):571-81.

190. Martella V, Campolo M, Lorusso E, Cavicchio P, Camero M, Bellacicco AL, Decaro N, Elia G, Greco G, Corrente M, et al. **Norovirus in captive lion cub (Panthera leo).** Emerg Infect Dis 2007 Jul;13(7):1071-3.

191. Martella V, Lorusso E, Decaro N, Elia G, Radogna A, D'Abramo M, Desario C, Cavalli A, Corrente M, Camero M, et al. **Detection and molecular characterization of a canine norovirus.** Emerg Infect Dis 2008 Aug;14(8):1306-8.

192. Martin DP. **Recombination detection and analysis using RDP3.** Methods Mol Biol 2009;537:185-205.

193. Martinez N, Espul C, Cuello H, Zhong W, Jiang X, Matson DO, Berke T. **Sequence diversity of human caliciviruses recovered from children with diarrhea in Mendoza, Argentina, 1995-1998.** J Med Virol 2002 Jun;67(2):289-98.

194. Matsuura Y, Tohya Y, Mochizuki M, Takase K, Sugimura T. **Identification of conformational neutralizing epitopes on the capsid protein of canine calicivirus.** J Gen Virol 2001 Jul;82(Pt 7):1695-702.

195. Mattison K, Shukla A, Cook A, Pollari F, Friendship R, Kelton D, Bidawid S, Farber JM. **Human noroviruses in swine and cattle.** Emerg Infect Dis 2007 Aug;13(8):1184-8.

196. Mattner F, Sohr D, Heim A, Gastmeier P, Vennema H, Koopmans M. **Risk groups for clinical complications of norovirus infections: an outbreak investigation.** Clin Microbiol Infect 2006 Jan;12(1):69-74.

197. Maunula L, Kalso S, Von Bonsdorff CH, Ponka A. **Wading pool water contaminated with both noroviruses and astroviruses as the source of a gastroenteritis outbreak.** Epidemiol Infect 2004 Aug;132(4):737-43.

198. McCormack GP, Clewley JP. **The application of molecular phylogenetics to the analysis of viral genome diversity and evolution.** Rev Med Virol 2002 Jul-2002 Aug 31;12(4):221-38.

199. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV. **Food-related illness and death in the United States.** Emerg Infect Dis 1999 Sep-1999 Oct 31;5(5):607-25.

200. Medici MC, Martinelli M, Abelli LA, Ruggeri FM, Di Bartolo I, Arcangeletti MC, Pinardi F, De Conto F, Izzi G, Bernasconi S, et al. **Molecular epidemiology of norovirus infections in sporadic cases of viral gastroenteritis among children in Northern Italy.** J Med Virol 2006 Nov;78(11):1486-92.

201. Mills CE, Robins JM, Lipsitch M. **Transmissibility of 1918 pandemic influenza.** Nature 2004 Dec 16;432(7019):904-6.

202. Monica B, Ramani S, Banerjee I, Primrose B, Iturriza-Gomara M, Gallimore CI, Brown DW, M F, Moses PD, Gray JJ, et al. **Human caliciviruses in symptomatic and asymptomatic infections in children in Vellore, South India.** J Med Virol 2007 May;79(5):544-51.

203. Motomura K, Oka T, Yokoyama M, Nakamura H, Mori H, Ode H, Hansman GS, Katayama K, Kanda T, Tanaka T, et al. **Identification of monomorphic and divergent haplotypes in the 2006-2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history.** J Virol 2008 Nov;82(22):11247-62.

204. Mounts AW, Ando T, Koopmans M, Bresee JS, Noel J, Glass RI. **Cold weather seasonality of gastroenteritis associated with Norwalk-like viruses.** J Infect Dis 2000 May;181 Suppl 2:S284-7.

205. Murata T, Katsushima N, Mizuta K, Muraki Y, Hongo S, Matsuzaki Y. **Prolonged norovirus shedding in infants <or=6 months of age with gastroenteritis.** Pediatr Infect Dis J 2007 Jan;26(1):46-9.

206. Nakagomi T, Correia JB, Nakagomi O, Montenegro FM, Cuevas LE, Cunliffe NA, Hart CA. **Norovirus infection among children with acute gastroenteritis in Recife, Brazil: disease severity is comparable to rotavirus gastroenteritis.** Arch Virol 2008;153(5):957-60.

207. Nayak MK, Balasubramanian G, Sahoo GC, Bhattacharya R, Vinje J, Kobayashi N, Sarkar MC, Bhattacharya MK, Krishnan T. **Detection of a novel intergenogroup recombinant Norovirus from Kolkata, India.** Virology 2008 Jul 20;377(1):117-23.

208. Neill JD, Sosnovtsev SV, Green KY. **Recovery and altered neutralization specificities of chimeric viruses containing capsid protein domain exchanges from antigenically distinct strains of feline calicivirus.** J Virol 2000 Feb;74(3):1079-84.

209. Ng TL, Chan PP, Phua TH, Loh JP, Yip R, Wong C, Liaw CW, Tan BH, Chiew KT, Chua SB, et al. **Oyster-associated outbreaks of Norovirus gastroenteritis in Singapore.** J Infect 2005 Dec;51(5):413-8.

210. Nilsson M, Hedlund KO, Thorhagen M, Larson G, Johansen K, Ekspong A, Svensson L. **Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype.** J Virol 2003 Dec;77(24):13117-24.

211. Noda M, Fukuda S, Nishio O. **Statistical analysis of attack rate in norovirus foodborne outbreaks.** Int J Food Microbiol 2008 Feb 29;122(1-2):216-20.

212. Noel JS, Fankhauser RL, Ando T, Monroe SS, Glass RI. **Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution.** J Infect Dis 1999 Jun;179(6):1334-44.

213. Nordgren J, Kindberg E, Lindgren PE, Matussek A, Svensson L. **Norovirus gastroenteritis outbreak with a secretor-independent susceptibility pattern, Sweden.** Emerg Infect Dis 2010 Jan;16(1):81-7.

214. Nuanualsuwan S, Mariam T, Himathongkham S, Cliver DO. **Ultraviolet inactivation of feline calicivirus, human enteric viruses and coliphages.** Photochem Photobiol 2002 Oct;76(4):406-10.

215. Nygard K, Torven M, Ancker C, Knauth SB, Hedlund KO, Giesecke J, Andersson Y, Svensson L. **Emerging genotype (GGIIb) of norovirus in drinking water, Sweden.** Emerg Infect Dis 2003 Dec;9(12):1548-52.

216. O'Reilly CE, Bowen AB, Perez NE, Sarisky JP, Shepherd CA, Miller MD, Hubbard BC, Herring M, Buchanan SD, Fitzgerald CC, et al. **A waterborne outbreak of gastroenteritis with multiple etiologies among resort island visitors and residents: Ohio, 2004.** Clin Infect Dis 2007 Feb 15;44(4):506-12.

217. Ochoa M, Barcena J, de la Luna S, Melero JA, Douglas AR, Nieto A, Ortin J, Skehel JJ, Portela A. **Epitope mapping of cross-reactive monoclonal antibodies specific for the influenza A virus PA and PB2 polypeptides.** Virus Res 1995 Aug;37(3):305-15.

218. Okada M, Ogawa T, Kaiho I, Shinozaki K. **Genetic Analysis of Noroviruses in Chiba Prefecture, Japan, between 1999 and 2004.** J Clin Microbiol 2005 Sep;43(9):4391-401.

219. Okada M, Ogawa T, Yoshizumi H, Kubonoya H, Shinozaki K. **Genetic variation of the norovirus GII-4 genotype associated with a large number of outbreaks in Chiba prefecture, Japan.** Arch Virol 2007 Sep 14;(152(12)):2249-52.

220. Okada M, Tanaka T, Oseto M, Takeda N, Shinozaki K. **Genetic analysis of noroviruses associated with fatalities in healthcare facilities.** Arch Virol 2006 Aug;151(8):1635-41.

221. Oliver SL, Asobayire E, Charpilienne A, Cohen J, Bridger JC. **Complete genomic characterization and antigenic relatedness of genogroup III, genotype 2 bovine noroviruses.** Arch Virol 2007 Feb;152(2):257-72.

222. Oliver SL, Asobayire E, Dastjerdi AM, Bridger JC. **Genomic characterization of the unclassified bovine enteric virus Newbury agent-1 (Newbury1) endorses a new genus in the family Caliciviridae.** Virology 2006 Jun 20;350(1):240-50.

223. Olsen SJ, MacKinnon LC, Goulding JS, Bean NH, Slutsker L. **Surveillance for foodborne-disease outbreaks--United States, 1993-1997.** MMWR CDC Surveill Summ 2000 Mar 17;49(1):1-62.

224. Page RD. **TreeView: an application to display phylogenetic trees on personal computers.** Comput Appl Biosci 1996 Aug;12(4):357-8.

225. Papaventsis DC. **Norovirus infection in children with acute gastroenteritis, madagascar, 2004-2005.** Emerg Infect Dis 2007 Jun;13(6):908-11.

226. Parker TD, Kitamoto N, Tanaka T, Hutson AM, Estes MK. **Identification of Genogroup I and Genogroup II broadly reactive epitopes on the norovirus capsid.** J Virol 2005 Jun;79(12):7402-9.

227. Parrino TA, Schreiber DS, Trier JS, Kapikian AZ, Blacklow NR. **Clinical immunity in acute gastroenteritis caused by Norwalk agent.** N Engl J Med 1977 Jul 14;297(2):86-9.

228. Patel MM, Hall AJ, Vinje J, Parashar UD. **Noroviruses: a comprehensive review.** J Clin Virol 2009 Jan;44(1):1-8.

229. Patel MM, Widdowson MA, Glass RI, Akazawa K, Vinje J, Parashar UD. **Systematic literature review of role of noroviruses in sporadic gastroenteritis.** Emerg Infect Dis 2008 Aug;14(8):1224-31.

230. Pawa N, Vanezis AP, Tutton MG. **Spontaneous bowel perforation due to norovirus: a case report.** Cases J 2009;2:9101.

231. Perez-Losada M, Porter ML, Tazi L, Crandall KA. **New methods for inferring population dynamics from microbial sequences.** Infect Genet Evol 2007 Jan;7(1):24-43.

232. Phan TG, Kaneshi K, Ueda Y, Nakaya S, Nishimura S, Yamamoto A, Sugita K, Takanashi S, Okitsu S, Ushijima H. **Genetic heterogeneity, evolution, and recombination in noroviruses.** J Med Virol 2007 Jul 2;79(9):1388-400.

233. Phan TG, Kuroiwa T, Kaneshi K, Ueda Y, Nakaya S, Nishimura S, Yamamoto A, Sugita K, Nishimura T, Yagyu F, et al. **Changing distribution of norovirus genotypes and genetic analysis of recombinant GIIb among infants and children with diarrhea in Japan.** J Med Virol 2006 Jul;78(7):971-8.

234. Phan TG, Takanashi S, Kaneshi K, Ueda Y, Nakaya S, Nishimura S, Sugita K, Nishimura T, Yamamoto A, Yagyu F, et al. **Detection and genetic characterization of norovirus strains circulating among infants and children with acute gastroenteritis in Japan during 2004-2005.** Clin Lab 2006;52(9-10):519-25.

235. Pond SK, Muse SV. **Site-to-site variation of synonymous substitution rates.** Mol Biol Evol 2005 Dec;22(12):2375-85.

236. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ. **Adaptation to different human populations by HIV-1 revealed by codon-based analyses.** PLoS Comput Biol 2006 Jun 23;2(6):e62.

237. Pond SL, Frost SD, Muse SV. **HyPhy: hypothesis testing using phylogenies.** Bioinformatics 2005 Mar 1;21(5):676-9.

238. Poon AF, Lewis FI, Frost SD, Kosakovsky Pond SL. **Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models.** Bioinformatics 2008 Sep 1;24(17):1949-50.

239. Poon AF, Lewis FI, Pond SL, Frost SD. **An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope.** PLoS Comput Biol 2007 Nov;3(11):e231.

240. Prasad BV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. **X-ray crystallographic structure of the Norwalk virus capsid.** Science 1999 Oct 8;286(5438):287-90.

241. Prasad BV, Hardy ME, Jiang X, Estes MK. **Structure of Norwalk virus.** Arch Virol Suppl 1996;12:237-42.

242. Pringle CR. **Virus taxonomy--1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998.** Arch Virol 1999;144(2):421-9.

243. Promed-mail**Viral Gastroenteritis Update 2006 (02).** 2006 Jun 26; 20060626.1776.

244. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. **The epidemic behavior of the hepatitis C virus.** Science 2001 Jun 22;292(5525):2323-5.

245. Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. **Phylogenetic evidence**

for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol Biol Evol 2007 Mar;24(3):845-52.

246. Rachakonda G, Choudekar A, Parveen S, Bhatnagar S, Patwari A, Broor S. **Genetic diversity of noroviruses and sapoviruses in children with acute sporadic gastroenteritis in New Delhi, India.** J Clin Virol 2008 Sep;43(1):42-8.

247. Radford AD, Willoughby K, Dawson S, McCracken C, Gaskell RM. **The capsid gene of feline calicivirus contains linear B-cell epitopes in both variable and conserved regions.** J Virol 1999 Oct;73(10):8496-502.

248. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. **The genomic and epidemiological dynamics of human influenza A virus.** Nature 2008 May 29;453(7195):615-9.

249. Reimerink J, Stelma F, Rockx B, Brouwer D, Stobberingh E, van Ree R, Dompeling E, Mommers M, Thijs C, Koopmans M. **Early-life rotavirus and norovirus infections in relation to development of atopic manifestation in infants.** Clin Exp Allergy 2009 Feb;39(2):254-60.

250. Reuter G, Krisztalovics K, Vennema H, Koopmans M, Szucs G. **Evidence of the etiological predominance of norovirus in gastroenteritis outbreaks-emerging new-variant and recombinant noroviruses in Hungary.** J Med Virol 2005 Aug;76(4):598-607.

251. Reuter G, Pankovics P, Szucs G. **Genetic drift of norovirus genotype GII-4 in seven consecutive epidemic seasons in Hungary.** J Clin Virol 2008 Jun;42(2):135-40.

252. Reuter G, Vennema H, Koopmans M, Szucs G. **Epidemic spread of recombinant noroviruses with four capsid types in Hungary.** J Clin Virol 2006 Jan;35(1):84-8.

253. Rockx B, De Wit M, Vennema H, Vinje J, De Bruin E, Van Duynhoven Y, Koopmans M. **Natural history of human calicivirus infection: a prospective cohort study.** Clin Infect Dis 2002 Aug 1;35(3):246-53.

254. Rodriguez-Guillen L, Vizzi E, Alcala AC, Pujol FH, Liprandi F, Ludert JE. **Calicivirus infection in human immunodeficiency virus seropositive children and adults.** J Clin Virol 2005 Jun;33(2):104-9.

255. Rondy M, Koopmans M, Rotsaert C, VAN Loon T, Beljaars B, VAN Dijk G, Siebenga J, Svraka S, Rossen JW, Teunis P, et al. **Norovirus disease associated with excess mortality and use of statins: a retrospective cohort study of an outbreak following a pilgrimage to Lourdes.** Epidemiol Infect 2010 May 24;1-11.

256. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, et al. **The global circulation of seasonal influenza A (H3N2) viruses.** Science 2008 Apr 18;320(5874):340-6.

257. Ryan MJ, Ramsay M, Brown D, Gay NJ, Farrington CP, Wall PG. **Hospital admissions attributable to rotavirus infection in England and Wales.** J Infect Dis 1996 Sep;174 Suppl 1:S12-8.

258. Schouls LM, van der Ende A, van de Pol I, Schot C, Spanjaard L, Vauterin P, Wilderbeek D, Witteveen S. **Increase in genetic diversity of Haemophilus influenzae serotype b (Hib) strains after introduction of Hib vaccination in The Netherlands.** J Clin Microbiol 2005 Jun;43(6):2741-9.

259. Schouls LM, van der Heide HG, Vauterin L, Vauterin P, Mooi FR. **Multiple-locus variable-number tandem repeat analysis of Dutch Bordetella pertussis strains reveals rapid genetic changes with clonal expansion during the late 1990s.** J Bacteriol 2004 Aug;186(16):5496-505.

260. Schreier E, Doring F, Kunkel U. **Molecular epidemiology of outbreaks of gastroenteritis associated with small round structured viruses in Germany in 1997/98.** Arch Virol 2000;145(3):443-53.

261. Schwartz J, Spix C, Touloumi G, Bacharova L, Barumamdzadeh T, le Tertre A, Piekarksi T, Ponce de Leon A, Ponka A, Rossi G, et al. **Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions.** J Epidemiol Community Health 1996 Apr;50 Suppl 1:S3-11.

262. Scott FW. **Virucidal disinfectants and feline viruses.** Am J Vet Res 1980 Mar;41(3):410-4.

263. Sdiri-Loulizi K, Gharbi-Khelifi H, de Rougemont A, Chouchane S, Sakly N, Ambert-Balay K, Hassine

M, Guediche MN, Aouni M, Pothier P. **Acute infantile gastroenteritis associated with human enteric viruses in Tunisia.** J Clin Microbiol 2008 Apr;46(4):1349-55.

264. Seto Y, Iritani N, Kubo H, Kaida A, Murakami T, Haruki K, Nishio O, Ayata M, Ogura H. **Genotyping of Norovirus strains detected in outbreaks between April 2002 and March 2003 in Osaka City, Japan.** Microbiol Immunol 2005;49(3):275-83.

265. Siebenga J, Kroneman A, Vennema H, Duizer E, Koopmans M. **Food-borne viruses in Europe network report: the norovirus GII.4 2006b (for US named Minerva-like, for Japan Kobe034-like, for UK V6) variant now dominant in early seasonal surveillance.** Euro Surveill 2008 Jan 10;13(2).

266. Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. **High Prevalence of Prolonged Norovirus Shedding and Illness among Hospitalized Patients: A Model for In Vivo Molecular Evolution.** J Infect Dis 2008 Oct 1;198(7):994-1001.

267. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M. **Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants.** PLoS Pathog 2010;6(5):e1000884.

268. Siebenga JJ, Vennema H, Duizer E, Koopmans MP. **Gastroenteritis caused by norovirus GGII.4, The Netherlands, 1994-2005.** Emerg Infect Dis 2007 Jan;13(1):144-6.

269. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, Koopmans M. **Epochal Evolution of GGII.4 Norovirus Capsid Proteins from 1995 to 2006.** J Virol 2007 Sep;81(18):9932-41.

270. Siebenga JJ, Vennema H, Zheng DP, Vinje J, Lee BE, Pang XL, Ho EC, Lim W, Choudekar A, Broor S, et al. **Norovirus Illness Is a Global Problem: Emergence and Spread of Norovirus GII.4 Variants, 2001-2007.** J Infect Dis 2009 Sep 1;200(5):802-12.

271. Simmons G, Garbutt C, Hewitt J, Greening G. **A New Zealand outbreak of norovirus gastroenteritis linked to the consumption of imported raw Korean oysters.** N Z Med J 2007;120(1264):U2773.

272. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. **Mapping the antigenic and genetic evolution of influenza virus.** Science 2004 Jul 16;305(5682):371-6.

273. Smith Sehdev AE, Hutchins GM. **Problems with proper completion and accuracy of the cause-of-death statement.** Arch Intern Med 2001 Jan 22;161(2):277-84.

274. Sorhannus U, Kosakovsky Pond SL. **Evidence for positive selection on a sexual reproduction gene in the diatom genus Thalassiosira (Bacillariophyta).** J Mol Evol 2006 Aug;63(2):231-9.

275. Sosnovtsev SV, Belliot G, Chang KO, Onwudiwe O, Green KY. **Feline calicivirus VP2 is essential for the production of infectious virions.** J Virol 2005 Apr;79(7):4012-24.

276. Souza M, Azevedo MS, Jung K, Cheetham S, Saif LJ. **Pathogenesis and immune responses in gnotobiotic calves after infection with the genogroup II.4-HS66 strain of human norovirus.** J Virol 2008 Feb;82(4):1777-86.

277. Spada E, Sagliocca L, Sourdis J, Garbuglia AR, Poggi V, De Fusco C, Mele A. **Use of the minimum spanning tree model for molecular epidemiological investigation of a nosocomial outbreak of hepatitis C virus infection.** J Clin Microbiol 2004 Sep;42(9):4230-6.

278. Strausbaugh LJ, Sukumar SR, Joseph CL. **Infectious disease outbreaks in nursing homes: an unappreciated hazard for frail elderly persons.** Clin Infect Dis 2003 Apr 1;36(7):870-6.

279. Suffredini E, Corrain C, Arcangeli G, Fasolato L, Manfrin A, Rossetti E, Biazzi E, Mioni R, Pavoni E, Losio MN, et al. **Occurrence of enteric viruses in shellfish and relation to climatic-environmental factors.** Lett Appl Microbiol 2008 Sep 12.

280. Sugieda M, Nagaoka H, Kakishima Y, Ohshita T, Nakamura S, Nakajima S. **Detection of Norwalk-like virus genes in the caecum contents of pigs.** Arch Virol 1998;143(6):1215-21.

281. Sugieda M, Nakajima S. **Viruses detected in the caecum contents of healthy pigs representing a new**

genetic cluster in genogroup II of the genus "Norwalk-like viruses". Virus Res 2002 Aug;87(2):165-72.

282. Sullivan J, Swofford DL, Naylor GJP. **The Effect of Taxon Sampling on Estimating Rate Heterogeneity Parameters of Maximum-Likelihood Models .** Mol Biol Evol 1999 Oct;16(1347-1356).

283. Svraka S, Duizer E, Vennema H, de Bruin E, van der Veer B, Dorresteijn B, Koopmans M. **Etiological Role of Viruses in Outbreaks of Acute Gastroenteritis in The Netherlands from 1994 through 2005.** J Clin Microbiol 2007 May;45(5):1389-94.

284. Tajiri H, Kiyohara Y, Tanaka T, Etani Y, Mushiake S. **Abnormal computed tomography findings among children with viral gastroenteritis and symptoms mimicking acute appendicitis.** Pediatr Emerg Care 2008 Sep;24(9):601-4.

285. Takanashi S, Hashira S, Matsunaga T, Yoshida A, Shiota T, Tung PG, Khamrin P, Okitsu S, Mizuguchi M, Igarashi T, et al. **Detection, genetic characterization, and quantification of norovirus RNA from sera of children with gastroenteritis.** J Clin Virol 2009 Feb;44(2):161-3.

286. Tan M, Huang P, Meller J, Zhong W, Farkas T, Jiang X. **Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket.** J Virol 2003 Dec;77(23):12562-71.

287. Tan M, Jiang X. **Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle.** Trends Microbiol 2005 Jun;13(6):285-93.

288. Tan M, Jin M, Xie H, Duan Z, Jiang X, Fang Z. **Outbreak studies of a GII-3 and a GII-4 norovirus revealed an association between HBGA phenotypes and viral infection.** J Med Virol 2008 Jul;80(7):1296-301.

289. Tan M, Xia M, Cao S, Huang P, Farkas T, Meller J, Hegde RS, Li X, Rao Z, Jiang X. **Elucidation of strain-specific interaction of a GII-4 norovirus with HBGA receptors by site-directed mutagenesis study.** Virology 2008 Aug 7.

290. Tan M, Xia M, Chen Y, Bu W, Hegde RS, Meller J, Li X, Jiang X. **Conservation of carbohydrate binding interfaces: evidence of human HBGA selection in norovirus evolution.** PLoS ONE 2009;4(4):e5058.

291. Tan M, Zhong W, Song D, Thornton S, Jiang X. **E. coli-expressed recombinant norovirus capsid proteins maintain authentic antigenicity and receptor binding capability.** J Med Virol 2004 Dec;74(4):641-9.

292. Taylor CE, Craft AW, Kernahan J, Millman R, Reid MM, Scott R, Toms GL. **Local antibody production and respiratory syncytial virus infection in children with leukaemia.** J Med Virol 1990 Apr;30(4):277-81.

293. Teunis PF, Moe CL, Liu P, E Miller S, Lindesmith L, Baric RS, Le Pendu J, Calderon RL. **Norwalk virus: How infectious is it?** J Med Virol 2008 Jun 12;80(8):1468-76.

294. Thurston-Enriquez JA, Haas CN, Jacangelo J, Riley K, Gerba CP. **Inactivation of feline calicivirus and adenovirus type 40 by UV radiation.** Appl Environ Microbiol 2003 Jan;69(1):577-82.

295. Tian P, Bates AH, Jensen HM, Mandrell RE. **Norovirus binds to blood group A-like antigens in oyster gastrointestinal cells.** Lett Appl Microbiol 2006 Dec;43(6):645-51.

296. Tian P, Engelbrektson AL, Jiang X, Zhong W, Mandrell RE. **Norovirus recognizes histo-blood group antigens on gastrointestinal cells of clams, mussels, and oysters: a possible mechanism of bioaccumulation.** J Food Prot 2007 Sep;70(9):2140-7.

297. Tohya Y, Yokoyama N, Maeda K, Kawaguchi Y, Mikami T. **Mapping of antigenic sites involved in neutralization on the capsid protein of feline calicivirus.** J Gen Virol 1997 Feb;78 (Pt 2):303-5.

298. Torner N, Dominguez A, Ruiz L, Martinez A, Bartolome R, Buesa J, Ferrer MD, Catalan Viral Gastroenteritis Study Group. **Acute gastroenteritis outbreaks in Catalonia, Spain: norovirus versus Salmonella.** Scand J Gastroenterol 2008;43(5):567-73.

299. Troeger H, Loddenkemper C, Schneider T, Schreier E, Epple HJ, Zeitz M, Fromm M, Schulzke JD.

Structural and functional changes of the duodenum in human norovirus infection. Gut 2009 Aug;58(8):1070-7.

300. Tsang OT, Wong AT, Chow CB, Yung RW, Lim WW, Liu SH. **Clinical characteristics of nosocomial norovirus outbreaks in Hong Kong.** J Hosp Infect 2008 Jun;69(2):135-40.

301. Tu ET, Bull RA, Greening GE, Hewitt J, Lyon MJ, Marshall JA, McIver CJ, Rawlinson WD, White PA. **Epidemics of gastroenteritis during 2006 were associated with the spread of norovirus GII.4 variants 2006a and 2006b.** Clin Infect Dis 2008 Feb 1;46(3):413-20.

302. Tu ET, Bull RA, Kim MJ, McIver CJ, Heron L, Rawlinson WD, White PA. **Norovirus excretion in an aged-care setting.** J Clin Microbiol 2008 Jun;46(6):2119-21.

303. Tu ET, Nguyen T, Lee P, Bull RA, Musto J, Hansman G, White PA, Rawlinson WD, McIver CJ. **Norovirus GII.4 strains and outbreaks, Australia.** Emerg Infect Dis 2007 Jul;13(7):1128-30.

304. Turcios RM, Widdowson MA, Sulka AC, Mead PS, Glass RI. **Reevaluation of epidemiological criteria for identifying outbreaks of acute gastroenteritis due to norovirus: United States, 1998-2000.** Clin Infect Dis 2006 Apr 1;42(7):964-9.

305. Turcios-Ruiz RM, Axelrod P, St John K, Bullitt E, Donahue J, Robinson N, Friss HE. **Outbreak of necrotizing enterocolitis caused by norovirus in a neonatal intensive care unit.** J Pediatr 2008 Sep;153(3):339-44.

306. Van Asten, L.; Van den Wijngaard, C.; Siebenga, J. J.; Van Pelt, W.; Van Vliet, H.; Koopmans, M**Greater pathogenicity of norovirus strains in 2003? A syndromic approach.** Advances in Disease Surveillance2:175(Abstract).

307. Van Damme P, Giaquinto C, Huet F, Gothefors L, Maxwell M, Van der Wielen M. **Multicenter prospective study of the burden of rotavirus acute gastroenteritis in Europe, 2004-2005: the REVEAL study.** J Infect Dis 2007 May 1;195 Suppl 1:S4-S16.

308. Van de Peer Y, De Wachter R. **TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment.** Comput Appl Biosci 1994 Sep;10(5):569-70.

309. Van den Brandhof WE, Kroes ACM, Bosman A, Peeters MF, Heijnen MLA. **[Reporting virus diagnostics in the Netherlands: representativeness of the virological weekly reports].** Infectieziekten Bulletin 2002;13:110-3.

310. van Duynhoven YT, de Jager CM, Kortbeek LM, Vennema H, Koopmans MP, van Leusden F, van der Poel WH, van den Broek MJ. **A one-year intensified study of outbreaks of gastroenteritis in The Netherlands.** Epidemiol Infect 2005 Feb;133(1):9-21.

311. G. Van Houtven,A. Honeycutt,B. Gilman,N. McCall,W. Throneburg**Costs of Illness for Environmentally Related Health Effects in Older Americans .** 2005 Apr.

312. van Steenbergen J, Debast S, van Kregten E, van den Berg R, Notermans D, Kuijper E. **Isolation of Clostridium difficile ribotype 027, toxinotype III in the Netherlands after increase in C. difficile-associated diarrhoea.** Euro Surveill 2005 Jul;10(7):E050714.1.

313. Vardy J, Love AJ, Dignon N. **Outbreak of acute gastroenteritis among emergency department staff.** Emerg Med J 2007 Oct;24(10):699-702.

314. Vennema H, de Bruin E, Koopmans M. **Rational optimization of generic primers used for Norwalk-like virus detection by reverse transcriptase polymerase chain reaction.** J Clin Virol 2002 Aug;25(2):233-5.

315. Verhoef L, Boxman IL, Duizer E, Rutjes SA, Vennema H, Friesema IH, de Roda Husman AM, Koopmans M. **Multiple exposures during a norovirus outbreak on a river-cruise sailing through Europe, 2006.** Euro Surveill 2008 Jun 12;13(24).

316. Verhoef L, Depoortere E, Boxman I, Duizer E, van Duynhoven Y, Harris J, Johnsen C, Kroneman A, Le Guyader S, Lim W, et al. **Emergence of new norovirus variants on spring cruise ships and prediction of winter epidemics.** Emerg Infect Dis 2008 Feb;14(2):238-43.

317. Verhoef L, Duizer E, Vennema H, Siebenga J, Swaan C, Isken L, Koopmans M, Balay K, Pothier P, McKeown P, et al. **Import of norovirus infections in the Netherlands and Ireland following pilgrimages to Lourdes, 2008--preliminary report.** Euro Surveill 2008 Oct 30;13(44):pii: 19025.

318. Verhoef LP, Kroneman A, van Duynhoven Y, Boshuizen H, van Pelt W, Koopmans M. **Selection tool for foodborne norovirus outbreaks.** Emerg Infect Dis 2009 Jan;15(1):31-8.

319. Vidal R, Solari V, Mamani N, Jiang X, Vollaire J, Roessler P, Prado V, Matson DO, O'Ryan ML. **Caliciviruses and foodborne gastroenteritis, Chile.** Emerg Infect Dis 2005 Jul;11(7):1134-7.

320. Vinje J, Altena SA, Koopmans MP. **The incidence and genetic variability of small round-structured viruses in outbreaks of gastroenteritis in The Netherlands.** J Infect Dis 1997 Nov;176(5):1374-8.

321. Vinje J, Hamidjaja RA, Sobsey MD. **Development and application of a capsid VP1 (region D) based reverse transcription PCR assay for genotyping of genogroup I and II noroviruses.** J Virol Methods 2004 Mar 15;116(2):109-17.

322. Vinje J, Koopmans MP. **Molecular detection and epidemiology of small round-structured viruses in outbreaks of gastroenteritis in the Netherlands.** J Infect Dis 1996 Sep;174(3):610-5.

323. Vinje J, Vennema H, Maunula L, von Bonsdorff CH, Hoehne M, Schreier E, Richards A, Green J, Brown D, Beard SS, et al. **International collaborative study to compare reverse transcriptase PCR assays for detection and genotyping of noroviruses.** J Clin Microbiol 2003 Apr;41(4):1423-33.

324. Vriend G. **WHAT IF: a molecular modeling and drug design program.** J Mol Graph 1990 Mar;8(1):52-6, 29.

325. Wang QH, Han MG, Cheetham S, Souza M, Funk JA, Saif LJ. **Porcine noroviruses related to human noroviruses.** Emerg Infect Dis 2005 Dec;11(12):1874-81.

326. Wang QH, Souza M, Funk JA, Zhang W, Saif LJ. **Prevalence of noroviruses and sapoviruses in swine of various ages determined by reverse transcription-PCR and microwell hybridization assays.** J Clin Microbiol 2006 Jun;44(6):2057-62.

327. Wei C, Farkas T, Sestak K, Jiang X. **Recovery of infectious virus by transfection of in vitro-generated RNA from tulane calicivirus cDNA.** J Virol 2008 Nov;82(22):11429-36.

328. Westhoff TH, Vergoulidou M, Loddenkemper C, Schwartz S, Hofmann J, Schneider T, Zidek W, van der Giet M. **Chronic norovirus infection in renal transplant recipients.** Nephrol Dial Transplant 2009 Mar;24(3):1051-3.

329. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, Hudson MJ, Roderick PJ. **Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive.** BMJ 1999 Apr 17;318(7190):1046-50.

330. Widdowson MA, Cramer EH, Hadley L, Bresee JS, Beard RS, Bulens SN, Charles M, Chege W, Isakbaeva E, Wright JG, et al. **Outbreaks of acute gastroenteritis on cruise ships and on land: identification of a predominant circulating strain of norovirus--United States, 2002.** J Infect Dis 2004 Jul 1;190(1):27-36.

331. Widdowson MA, Sulka A, Bulens SN, Beard RS, Chaves SS, Hammond R, Salehi ED, Swanson E, Totaro J, Woron R, et al. **Norovirus and foodborne disease, United States, 1991-2000.** Emerg Infect Dis 2005 Jan;11(1):95-102.

332. Wolf S, Williamson W, Hewitt J, Lin S, Rivera-Aban M, Ball A, Scholes P, Savill M, Greening GE. **Molecular detection of norovirus in sheep and pigs in New Zealand farms.** Vet Microbiol 2008 Jul 5.

333. Wood DJ, David TJ, Chrystie IL, Totterdell B. **Chronic enteric virus infection in two T-cell**

immunodeficient children. J Med Virol 1988 Apr;24(4):435-44.

334. Wu FT, Oka T, Katayama K, Wu HS, Donald Jiang DS, Miyamura T, Takeda N, Hansman GS. **Genetic diversity of noroviruses in Taiwan between November 2004 and March 2005.** Arch Virol 2006 Feb 7.

335. Wu HM, Fornek M, Schwab KJ, Chapin AR, Gibson K, Schwab E, Spencer C, Henning K. **A norovirus outbreak at a long-term-care facility: the role of environmental surface contamination.** Infect Control Hosp Epidemiol 2005 Oct;26(10):802-10.

336. Wyatt RG, Dolin R, Blacklow NR, DuPont HL, Buscho RF, Thornhill TS, Kapikian AZ, Chanock RM. **Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers.** J Infect Dis 1974 Jun;129(6):709-14.

337. Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ. **Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses.** J Clin Microbiol 2008 Mar;46(3):947-53.

338. Xerry J, Gallimore CI, Iturriza-Gomara M, Gray JJ. **Tracking the transmission routes of genogroup II noroviruses in suspected food-borne or environmental outbreaks of gastroenteritis through sequence analysis of the P2 domain.** J Med Virol 2009 Jul;81(7):1298-304.

339. Yee EL, Palacio H, Atmar RL, Shah U, Kilborn C, Faul M, Gavagan TE, Feigin RD, Versalovic J, Neill FH, et al. **Widespread outbreak of norovirus gastroenteritis among evacuees of Hurricane Katrina residing in a large "megashelter" in Houston, Texas: lessons learned for prevention.** Clin Infect Dis 2007 Apr 15;44(8):1032-9.

340. Zambon M, Bull T, Sadler CJ, Goldman JM, Ward KN. **Molecular epidemiology of two consecutive outbreaks of parainfluenza 3 in a bone marrow transplant unit.** J Clin Microbiol 1998 Aug;36(8):2289-93.

341. Zheng DP, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. **Norovirus classification and proposed strain nomenclature.** Virology 2005 Dec 8.

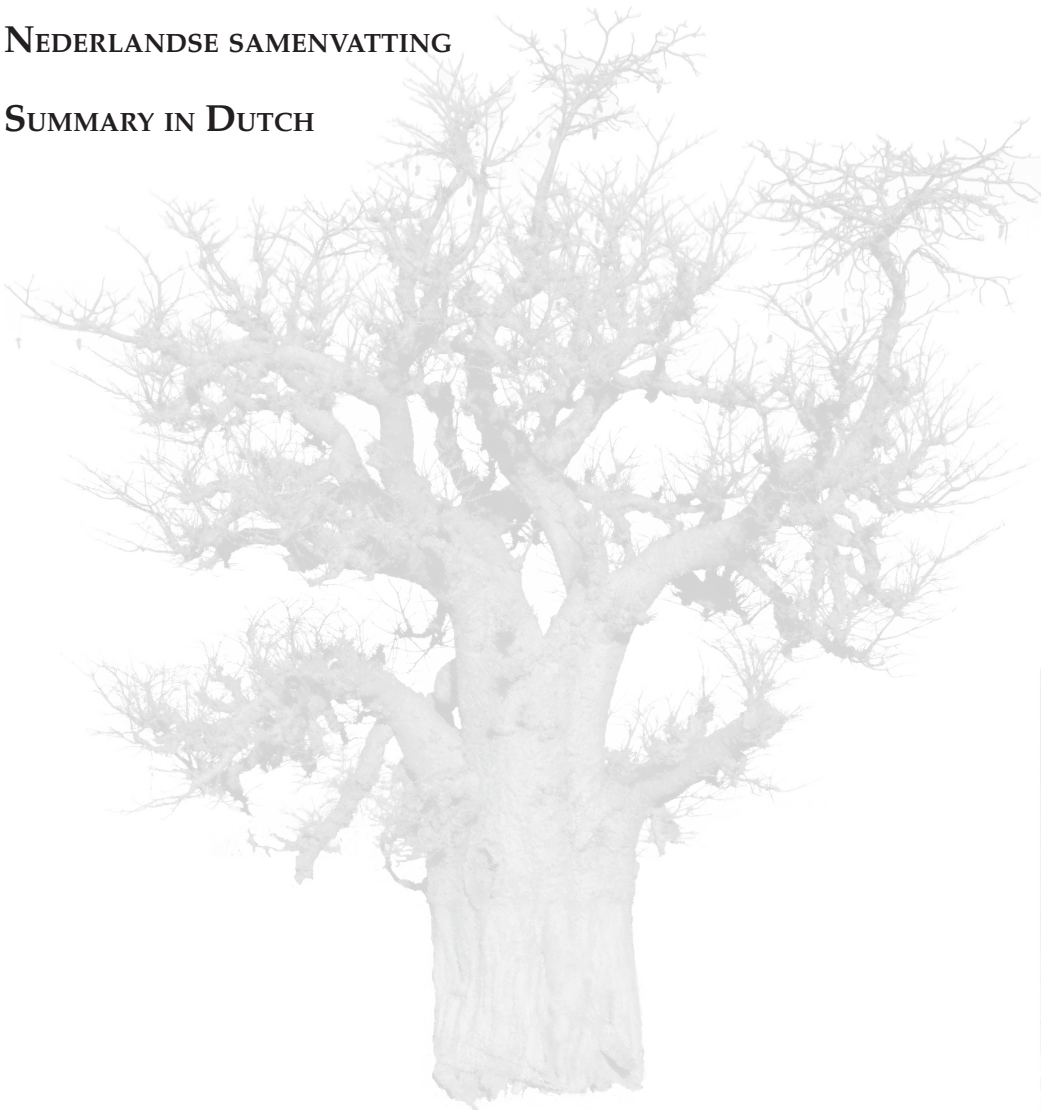
342. Zingg W, Colombo C, Jucker T, Bossart W, Ruef C. **Impact of an outbreak of norovirus infection on hospital resources.** Infect Control Hosp Epidemiol 2005 Mar;26(3):263-7.

Abbreviations List

aa	amino acid
BSP	Bayesian Skyline Plot
DNA	Deoxynucleic Acid
EIA	Enzyme Immuno Assay
FCV	Feline CaliciVirus
G	Genogroup (GI, GII etc)
GE	Gastro Enteric
HBGA	Histo Blood Group Antigens
HPD	Highest Probability Density
kb	kilo base(pair)
MCC	Maximum Clade Credibility
MCMC	Markov Chain Monte Carlo
MRCA	Most Recent Common Ancestor
MST	Minimum Spanning Tree
$N_e \tau$	population size (N_e) multiplied by generation time (τ), a measure for genetic diversity
nt	nucleotide
NV	Norwalk Virus
ORF	Open Reading Frame
P	Protruding (Domain of the capsid)
PCR	Polymerase Chain Reaction
R0	Reproduction number
RdRp	RNA dependent RNA polymerase
RHDV	Rabbit Hemorrhagic Disease Virus
RNA	RiboNucleic Acid
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
S	Shell (Domain of the capsid)
VESV	Vesicular Exanthema of Swine Virus
VP	Viral Protein

Nederlandse samenvatting

Summary in Dutch



Norovirussen zijn wereldwijd de belangrijkste veroorzakers van zowel uitbraken als sporadische gevallen van gastro-enteritis. Normaal gesproken veroorzaken norovirussen zelflimiterende, en milde infecties bij mensen, die 1 tot 3 dagen duren, met de belangrijkste symptomen diarree en (projectiel) braken. Een combinatie van structurele uitbraak surveillance met gedegen virologisch onderzoek heeft ons, bij het ontbreken van een eenvoudig kweeksysteem voor norovirus, nuttige zaken geleerd over de moleculaire epidemiologie van de norovirussen, die mogelijk gebruikt kunnen worden in het verminderen van de impact en de omvang van toekomstige uitbraken en epidemieën.

Sinds 1996 hebben een aantal wereldwijde norovirus epidemieën plaatsgevonden, die allen volgden op het opkomen van nieuwe varianten van het dominante genotype, GII.4. In **hoofdstuk 2** worden de resultaten van de langlopende passieve norovirus surveillance in Nederland beschreven. Tussen 1994 en 2006 werden gemiddeld 68% van gerapporteerde uitbraken door genotype GII.4 veroorzaakt. Dit genotype wordt ook vaker in gezondheidszorg gerelateerde plaatsen aangetroffen, en persoon tot persoon transmissie wordt het vaakst gerapporteerd als transmissie route. In **hoofdstuk 4** worden resultaten van verschillende surveillance programma's verspreid over de wereld gecombineerd; 62% van de uitbraken gerapporteerd in deze wereldwijde studie werden veroorzaakt door GII.4 stammen. Hoewel er geen langdurige immuniteit in individuen tegen norovirus infecties gevonden wordt, ook niet tegen homologe infecties, duidt de moleculaire epidemiologie van het virus op het ontstaan van populatie immuniteit tegen de epidemische varianten van GII.4; gedurende de winterseizoenen die volgden op epidemische seizoenen was het percentage van GII.4 altijd veel lager dan gedurende de epidemische seizoenen.

Ook wanneer we gedetailleerder naar de virussen zelf kijken, zoals in **hoofdstuk 3** is beschreven, zien we hiervoor bewijs: volgens een stapsgewijs patroon vindt evolutie plaats, en de eerdere overheersende varianten worden steeds vervangen door nieuwe, genetisch verschillende varianten van het zelfde genotype. We kunnen dus aannemen dat deze genetische varianten ook antigene varianten zijn, hoewel dit niet direct is getest. Het merendeel van de mutaties die verschillende varianten van elkaar onderscheiden, liggen in het uitstekende, buitenste gedeelte van het capside eiwit. Dit gedeelte speelt een belangrijke rol bij binding aan gastheer cellen, en waarschijnlijk ook bij neutralisatie door een immuun respons.

Hoofdstuk 4 gaat verder in op de geografische distributie van de GII.4 varianten. Tot en met het jaar 2007 zijn er acht verschillende te onderscheiden, waarvan er vier een wereldwijde verspreiding hadden, de 1996-, 2002-, 2004- en 2006b varianten. Andere varianten veroorzaakten in beperkte geografische gebieden epidemieën; de 2006a- en de 2003Asia

varianten. Weer anderen werden wel wereldwijd aangetroffen, maar veroorzaakten slechts kleine aantallen uitbraken. Een verklaring voor verschillen in verspreiding van de varianten kan liggen in de genetische vatbaarheid van specifieke bevolkingsgroepen, die onder meer bepaald wordt door de zogenaamde Histo BloodGroup Antigens (HBGA's), die per individu verschillend tot expressie komen. Het is bekend dat mensen met verschillende HBGA profielen verschillen vertonen in vatbaarheid voor de verschillende geno*typen* van norovirus. Het is goed mogelijk dat zulk onderscheid ook op sub-genotype niveau te maken is; voor varianten. Met de gegevens over de binding van virus deeltjes aan deze HBGA's, de liganden, die momenteel bekend zijn, kan zo'n verschil echter niet verklaard worden.

Hoewel het aantal gerapporteerde norovirus uitbraken wereldwijd sterk is gestegen tijdens de laatste twee decennia, was het niet duidelijk of dit een werkelijke stijging representeerde, of dat er sprake was van toenemende bewustheid over norovirus, en groeiende mogelijkheden in, en toegankelijkheid tot diagnostische testen. In **hoofdstuk 5** worden de resultaten van de phylodynamica analyses van een grote set sequentie data met bijbehorende detectiedata gepresenteerd. Er kon worden vastgesteld dat de hoeveelheid norovirus infecties sinds 1996, toen de eerste wereldwijde epidemie van GII.4 stammen begon, waarschijnlijk inderdaad is toegenomen. De bestudeerde GII.4 stammen evolueerden met een snelheid van 4.3 tot 9.0 x 10^{-3} mutaties per site per jaar, en een meest recente gemeenschappelijke voorouder van de stammen die in de studie geïncludeerd waren, circuleerde in de vroege 1980er jaren. Dit betekent echter niet dat er voor die tijd geen GII.4 stammen circuleerden, zoals ook in een recent paper beschreven is.

Middels *in silico* analyses van een set capside sequenties, ook beschreven in **hoofdstuk 5**, konden aminozuren met een rol in adaptatie worden gevonden. Een interessante vinding hierbij was dat ook secundaire structuren in het RNA dat voor het capside eiwit codeert, een sturende rol blijken te hebben in deze adaptatie. Nucleotiden in het 5' uiteinde van Open Reading Frame 2, dat codeert voor het capside eiwit, zijn zeer geconserveerd, en de spaarzame mutaties die in dit gebied kunnen optreden zijn gelimiteerd door de secundaire RNA structuur, die een belangrijke rol heeft in transcriptie en / of translatie processen.
Een set aminozuren bestaande uit aminozuren die geïdentificeerd zijn in de moleculaire adaptatie studies, en in ligand bindings analyses, kan worden gebruikt om mogelijke toekomstige epidemische varianten te onderscheiden.

In de zoektocht naar een mogelijke bron voor nieuwe varianten werd onderzoek gedaan naar patiënten met een chronische norovirus infectie, beschreven in **hoofdstuk 6**. Dergelijke langdurige infecties bleken relatief veel voor te komen. Tot in ruim 8% van alle norovirus positieve patiënten in een retrospectieve 2-jarige studie in een ziekenhuis scheidde gedurende

langere periodes (21 tot 182 dagen) norovirus uit, en had daarbij bijbehorende symptomen. Deze chronische patiënten hadden allen immuunstoornissen, die het onmogelijk maakten de infectie te klaren. De virussen in deze patiënten accumuleerden mutaties met een snelheid die omgekeerd evenredig was met de mate van immunologische beperking van de patiënt. Virussen die aangetroffen werden in de faeces van een verder gezond kind, accumuleerden mutaties met een hogere snelheid. Hoewel chronische patiënten dus een mogelijke bron van antigene varianten zijn, is de kans dat deze in de gezonde populatie tot stand komen wellicht groter, omdat deze veel groter is dan de immuungecompromitteerde populatie.

Chronische norovirus infecties kunnen een fatale uitkomst hebben in reeds zwakke patiënten. Twee van de acht chronisch geïnfecteerde patiënten in onze studie overleden ten dele als gevolg van de norovirus infectie. Er is geen therapie of vaccin beschikbaar. Daarnaast liet ons onderzoek in **hoofdstuk 7** zien dat chronisch geïnfecteerde patiënten in de ziekenhuisomgeving een bron van nosocomiale infecties kunnen vormen. Deze bevindingen onderstrepen nogmaals het belang van goede hygiëne in ziekenhuizen, en van het voorkomen of beperken van uitbraken in omgevingen waar zwakke personen verblijven.

Niet alleen immuungecompromitteerden lopen risico's bij norovirus infecties, ook bij ouderen verloopt een norovirus infectie niet altijd zonder problemen. Door trendanalyses van medische registratie systemen, en de norovirus uitbraak database (**hoofdstuk 8**), konden we achterhalen dat norovirus activiteit significant geassocieerd is met onverklaarde gastro-enteritis, in huisarts bezoek, ziekenhuisopnames én in sterfte. Met een gerapporteerde norovirus uitbraak is 0.14 sterfgeval wegens gastro-enteritis geassocieerd. Voor één van de recente epidemische jaren (2004-05) komt dat op 21 sterfgevallen per 2,3 miljoen 65+-ers.

Het onderzoek beschreven in dit proefschrift draagt bij aan de kennis over norovirus moleculaire epidemiologie. De resultaten kunnen bijdragen aan de gerichte bestrijding van norovirus infecties, en maken duidelijk dat we norovirus niet langer altijd af kunnen doen als een 'milde infectie', zeker niet in de ziekenhuis populatie en bij ouderen. Surveillance in de bevolking blijft belangrijk om te monitoren welke genotypen en varianten circuleren, en om hypotheses voor verder onderzoek te blijven genereren.

Dankwoord

Word of Gratitude

Toen ik een jaar of 11, 12 was, hield ik een spreekbeurt over bomen. Het was natuurlijk een razend interessant verhaal, over duizenden jaren oude woudreuzen, jaarringen en turgor. Geen idee had ik dat die keer op de ZSV niet de laatste keer zou zijn dat ik over bomen zou presenteren... Om het onverwachte in ere te houden, vindt u in dit boekje een kleine hommage aan de bomen van vlees en bloed.

Lange tijd was het schrijven van dit dankwoord iets wat in de categorie Lange-Termijn-Planning thuishoorde. Normaal gesproken doe ik voor mezelf niet zo aan LTP, maar dit was toch wel iets om naar uit te kijken. Er zijn vele mensen geweest die hebben bijgedragen en actief of passief deelgenomen aan mijn promotietraject, en hoewel dat niet altijd duidelijk zal zijn geweest, ik ben jullie dankbaar! Op deze plek neem ik mijn kans dat uit te spreken, op papier en ook nog gedrukt, dus *waar*.

Marion, word ik -ijs en weder dienende- dan je eerste echte eigen promovendus? Hoewel je soms lastig te bereiken was, had ik altijd mazzel met je niet onder stoelen of banken gestoken interesse voor noro's en noro-onderzoek, waar ik natuurlijk lang niet genoeg gebruik van gemaakt heb. Je vorsende blik ('wat wil dat meisje nou??') zal ik nooit vergeten, net als je hulp te ontdekken wat het nu is (wat ik wil). In de tussentijd zijn en blijven noro's prachtige virussen, en niet aflatende bronnen van verwondering.

Harry, jij als echte Fries, en ik als reserve-Fries, hadden nooit uitgebreide gesprekken, maar er is toch uitgebreid geboomd. Met vragen over BN, fylogenie, en de zin en onzin van verschillende methoden kon ik altijd bij je terecht.

Erwin, jou ben ik ook veel dank verschuldigd, voor inhoudelijke, procedurele en morele bijstand op allerlei momenten. Ik heb de roets nog steeds niet gehuurd... maar dat gaat zeker nog wel eens gebeuren.

Ook ben ik dank verschuldigd aan mijn studenten, Suus en Marjan. Hoewel jullie werk niet 1 op 1 in dit boekje te vinden is, hebben jullie natuurlijk wel een belangrijke bijdrage geleverd aan het norovirus onderzoek. Marjan, de ORF1 eiwitten worden in Duitsland verder uitgeplozen.

There were many people with whom I had the privilege and pleasure to collaborate, and learn from, in EVENT, DIVINE, NoroNet, and beyond. Special mention goes to a number of people. Roland Siezen en Bernadette Renckens hebben bijgedragen aan de mooie grafische weergave van het GII.4 capside, op basis van de Norwalk Virus kristal structuur. Ian Clarke was a gracious host to me for a couple of weeks in his lab in Southampton. Unfortunately I

never got far with the VLPs back home... Rodney Ratcliff and Julie were extremely generous and flew me from Sydney to Adelaide, let me stay in their wonderful home, showed me and Merijn around in South Australia and gave me the opportunity to present my work to an interested and large crowd. Thank you for your many motivating and kind words, then and in the more recent history! Philippe Lemey, dankzij jou is het fylodynamica verhaal een mooi geheel geworden.
All co-authors of papers included in this thesis have -obviously- contributed to the realization of the end result. On several occassions I had the opportunity to call upon experts in different fields to expand and strengthen my work.

Hoewel ik eerst eventjes de enige (echte, want is Johan nu ook AIO of niet?) AIO was op Virologie, kreeg ik al snel versterking, en toenemende gezelligheid. Sanela blijft natuurlijk mijn maatje, voor het dansje, voor reisjes en voor vrouwengeklets. Verdere uitbreiding bleef niet uit. Marieke, nu mijn buuv, Marcel, ook al van 'vroeger', en nu ook paranimf, Sabine vd S / Sabin-like Person, Linda, Era, en Faizel. Onze AIO etentjes hadden natuurlijk veel frequenter moeten zijn. Sabine D, al ben je helemaal geen AIO, ik voeg je toch aan dit rijtje toe. Wat-oh-wat zouden wij allemaal toch doen volgend jaar rond deze tijd? Laten we in ieder geval nog eens afspreken voor een biertje, en ik wil al jullie verdedigingen meemaken!

De gehele Virologie afdeling heeft natuurlijk een stempel gedrukt op mijn proefschrift, en op mijn tijd bij het RIVM. Dank gaat uiteraard vooral naar de Gastro (Enteraal) groep, met Bas, Erwin dB, Gokhan, Edin, database koningin Annelies en alle andere mensen in het lab voor de praktische hulp en adviezen, en natuurlijk nooit te vergeten, Beheer voor bestellingen en het Secretariaat, voor het regelen van afspraken met Marion, het boeken van reizen en het beantwoorden van vele vragen.

Toen ik een eindje op weg was in mijn promotie bedachten Ivette en ik Proneri, en tot onze blijdschap viel ons plan in bijzonder vruchtbare aarde. Dank aan de enthousiastelingen van het eerste uur, en de mensen die onze stokjes met verve hebben overgenomen.

Hoewel ik zelden tot nooit in Rotterdam was, moeten daar natuurlijk ook mensen bedankt worden. Wim, het secretariaat, voor hulp en advies. Miranda, je hebt me goed op weg geholpen met de fylodynamica. Thijs dank voor het opsporen van die mooie sample-reeksen voor me!

De KNSRB en Minerva zorgden voor de nodige en onnodige afwisseling van het onderzoek, voor fietstochtjes naar Houten, ouderwetse diners in den lande en vliegreisjes naar Boston. Sander, hier is ie dan, eindelijk!

Wageningen, die tijd raak ik nooit meer kwijt! Argo, de mooiste club van Nederland, waar ik zelfs na mijn afstuderen nog machtig mooie momenten mee mocht maken. Dames van Chiquane, Kasjé, Bestuur Siebenga, het is geweldig om te zien dat we nog steeds met lief en leed bij elkaar terecht kunnen. Almanak: hier is boek nummer 2!

Jelle, ik kan voor jou (en Noortje) heel kort zijn: Ik ben blij dat je ook in Utrecht woont :-), niet alleen omdat ik anders de mosselen en asperges altijd in mijn eentje moet opeten. Fijn dat er vrienden zoals jij bestaan!

Els en Barend, een hele nieuwe vriendschap! We maakten al prachtige dingen mee, zoals de onherkenbaar verbouwde C2+ in Friesland, en de mooie mooie reis in West-Afrika! Pleegouders voor tomaten-kindertjes, wat volgt nog?

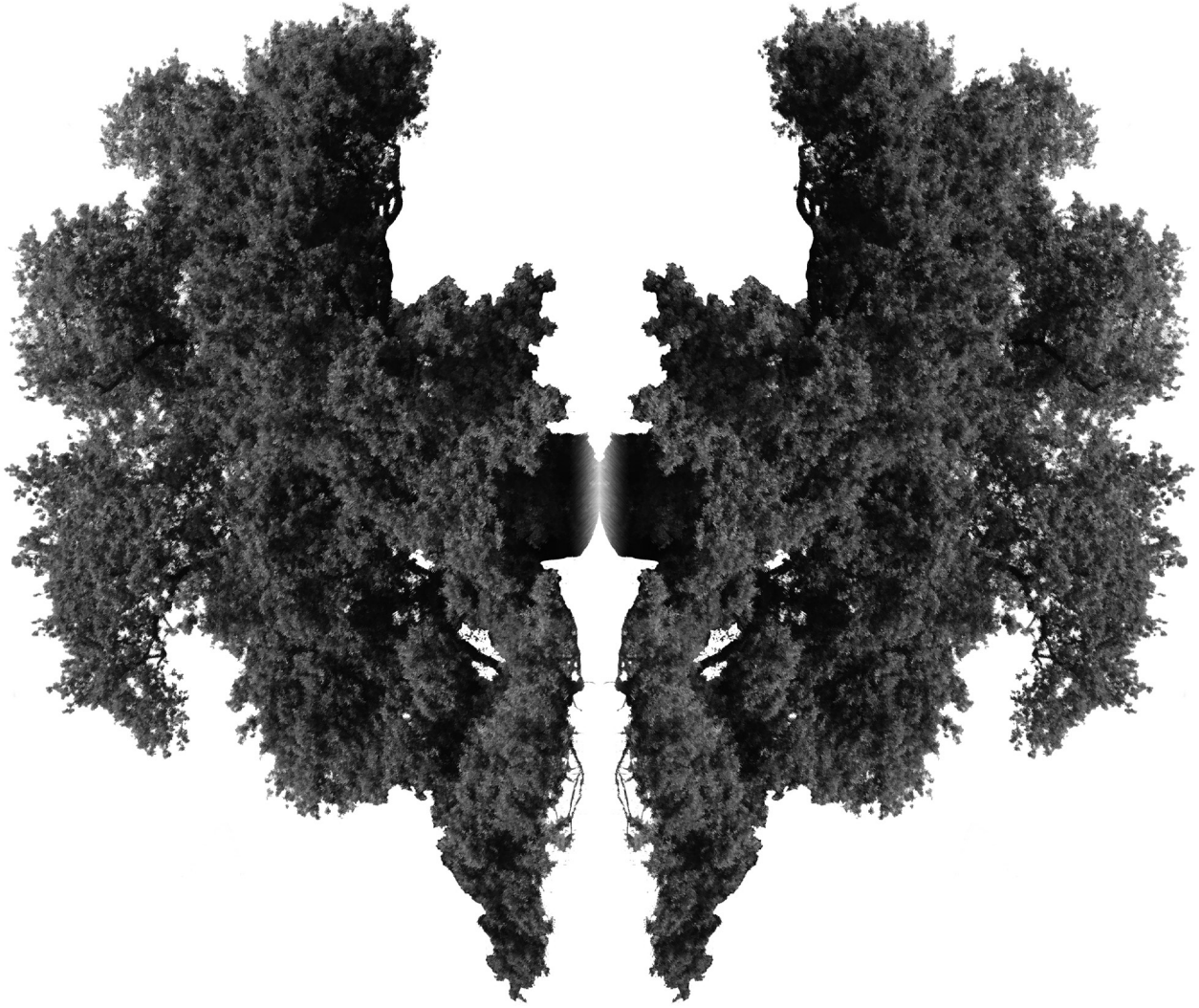
Myke, voor biertjes, verhalen, niksen, vakantie, wijntjes, meer verhalen en altijd een stukje roeien kunnen we bij elkaar terecht. Waar een goede brei-partij al niet toe kan leiden!

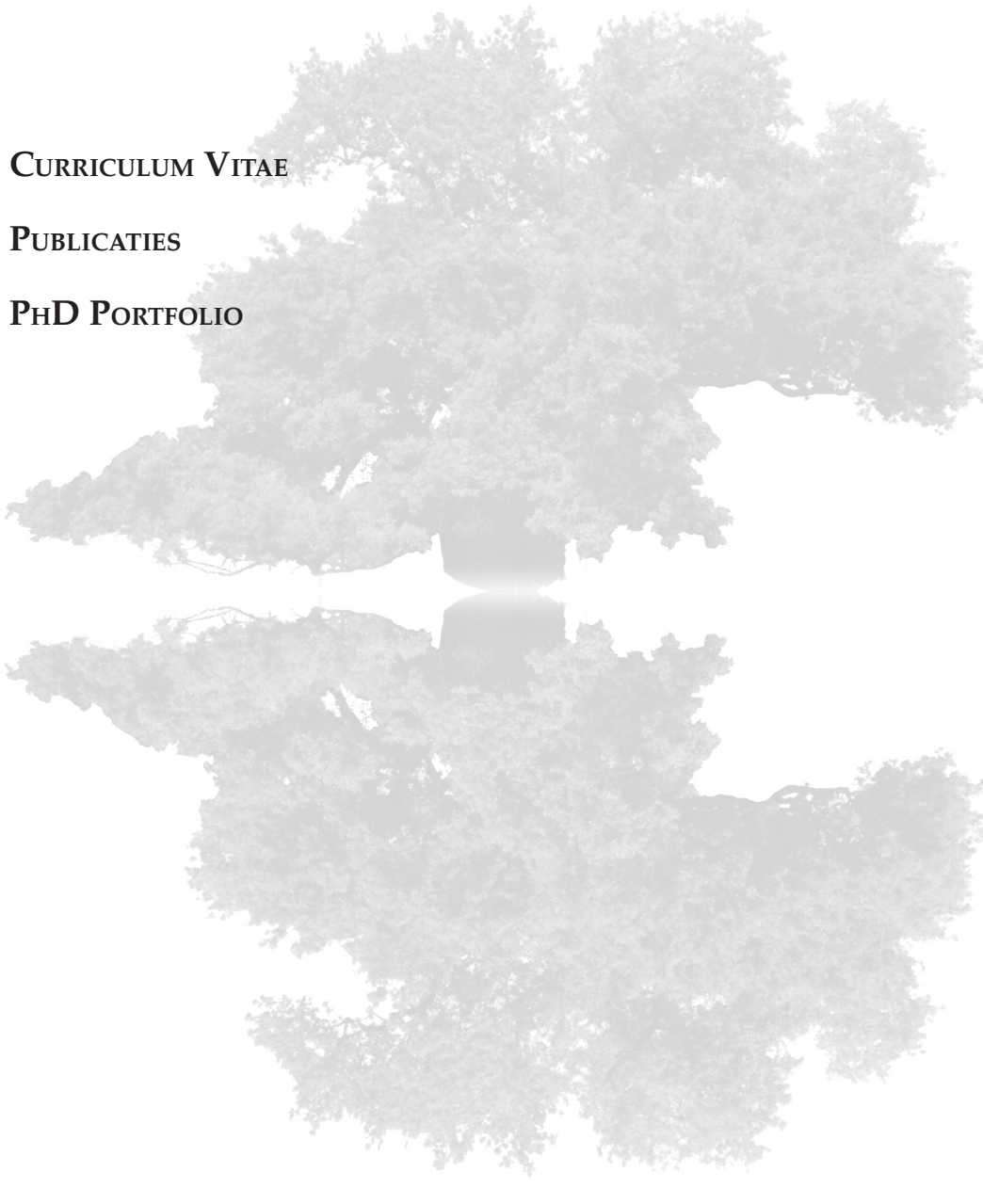
Mendelt, ik vind het geweldig dat je mijn paranimf wilt zijn. Je bent een top-broer, die trouw aanbiedt onhebbelijke (ex-)vriendjes voor me te slaan (dat kan ik zelluf wel), die precies weet hoe ik in elkaar steek, en die altijd klaarstaat met morele steun en computerhulp. Maar je bent ook 10 keer zo slim als ik en was dus feilloos in staat de virologie als hobby/interesse onderwerp op te pakken. Zo kon je me herhaaldelijk op ontwikkelingen wijzen waar ik zelf nog nauwelijks van op de hoogte was. Dank voor je TWIV tip! Met Fenneke mag je niet ontbreken op mijn verdediging en feestje, en dus ook niet in het dankwoord. Als paranimf ben je met Marcel een ideaal koppel; Marcel bracht me op de hoogte van de oorspronkelijke, en dus ware functie van de nimfen. Ik kan me niet veiliger wanen dan met jullie aan mijn zijde.

Papa en mama, altijd daar, in crisis en blije dagen, soms met een gerichte trap onder de kont, vaak met een kritische vraag (of 10) maar altijd constructief, en meestal natuurlijk gewoon gezellig en lief! Met een portie afhaalchinees of met een taart, maar liever met een gezonde maaltijd. Dank voor jullie geslaagde opvoeding (al zeg ik het zelf), de behoorlijk gelukte genetische samenstelling, en niet aflatende steun. Daar ik jullie product ben, is dit boekje ook van, en voor, jullie!

Merijn, we hebben het elkaar niet gemakkelijk gemaakt de afgelopen jaren, maar daar zijn we de mensen ook niet naar. Ik ben heel, heel blij dat ik deze laatste regels van mijn boekje, op jouw Appel getypt, aan jou mag richten. Waar de toekomst me ook heenbrengt, of jou, ik hoop met mijn hele wezen dat we er samen heen gaan. Ik heb wel zin in dat avontuur!

«Projectile vomiting», by Erik Kriek, the illustrator for the book Bacteriënjagers.





Curriculum Vitae

Publicaties

PhD Portfolio

Curriculum Vitae

Joukje Siebenga werd geboren op 21 juni 1978 in Emmeloord, Noordoostpolder. Het gymnasium diploma behaalde zij in 1996, aan het Stedelijk Gymnasium te Utrecht. Hierna werd begonnen met de studie Bioprocestechnologie, aan de Universiteit Wageningen. Tijdens deze studie werden drie afstudeervakken gedaan. Het grote afstudeervak werd gedaan bij DSM Food Specialties, in Delft, en de Genetica vakgroep van Wageningen Universiteit, onder begeleiding van Dr. Fons Debets en Dr. Noël van Peij. Het onderwerp was een klassiek genetische benadering van stamverbetering en genetische karakterisering van *Aspergillus niger.* De kleine afstudeervakken werden aan de Wageningen Universiteit uitgevoerd. Bij de vakgroep Celbiologie en Immunologie werd onderzoek gedaan onder begeleiding van Dr. René J. Stet, naar de karakterisering van nieuwe immunoglobuline-achtige transcripten in meerval, karper en forel. Bij Virologie werd onder begeleiding van Dr. Hendrik Marks onderzoek gedaan naar polymorfismen in White Spot Syndrome Virus. De stage werd uitgevoerd aan het California Institute of Technology (CalTech) in Pasadena, California, USA, onder begeleiding van Prof. Dr. James H. Strauss. Het betrof onderzoek naar Alphavirus eiwit expressie. Na het behalen van het doctoraal diploma werd begonnen met het promotie onderzoek, aan de afdeling Virologie van het Erasmus Medisch Centrum, uitgevoerd aan de Virologie afdeling van het Centrum voor Infectieziekten bestrijding van het RIVM, onder begeleiding van Prof. Dr. Marion P.G. Koopmans en Dr. Harry Vennema.

Publicaties

1. Dieu BT, Marks H, **Siebenga JJ**, Goldbach RW, Zuidema D, Duong TP, Vlak JM. Molecular epidemiology of white spot syndrome virus within Vietnam. J Gen Virol 2004 Dec;85(Pt 12):3607-18.

2. **Siebenga JJ**, Vennema H, Duizer E, Koopmans MP. Gastroenteritis caused by norovirus GGII.4, The Netherlands, 1994-2005. Emerg Infect Dis 2007 Jan;13(1):144-6.

3. **Siebenga JJ**, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, Koopmans M. Epochal Evolution of GGII.4 Norovirus Capsid Proteins from 1995 to 2006. J Virol 2007 Sep;81(18):9932-41.

4. **Siebenga J**, Kroneman A, Vennema H, Duizer E, Koopmans M. Food-borne viruses in Europe network report: the norovirus GII.4 2006b (for US named Minerva-like, for Japan Kobe034-like, for UK V6) variant now dominant in early seasonal surveillance. Euro Surveill 2008 Jan 10;13(2).

5. Verhoef L, Depoortere E, Boxman I, Duizer E, van Duynhoven Y, Harris J, Johnsen C, Kroneman A, Le Guyader S, Lim W, Maunula L, Meldal H, Ratcliff R, Reuter G, Schreier E, **Siebenga J**, Vainio K, Varela C, Vennema H, Koopmans M.. Emergence of new norovirus variants on spring cruise ships and prediction of winter epidemics. Emerg Infect Dis 2008 Feb;14(2):238-43.

6. Duizer E, Kroneman A, **Siebenga J**, Verhoef L, Vennema H, Koopmans M. Typing database for noroviruses. Euro Surveill 2008 May 8;13(19).

7. Iritani N, Vennema H, **Siebenga JJ**, Siezen RJ, Renckens B, Seto Y, Kaida A, Koopmans M. Genetic analysis of the capsid gene of genotype GII.2 noroviruses. J Virol 2008 Aug;82(15):7336-45.

8. **Siebenga JJ**, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. High Prevalence of Prolonged Norovirus Shedding and Illness among Hospitalized Patients: A Model for In Vivo Molecular Evolution. J Infect Dis 2008 Oct 1;198(7):994-1001.

9. Verhoef L, Duizer E, Vennema H, **Siebenga J**, Swaan C, Isken L, Koopmans M, Balay K, Pothier P, McKeown P, et al. Import of norovirus infections in the Netherlands and Ireland following pilgrimages to Lourdes, 2008--preliminary report. Euro Surveill 2008 Oct 30;13(44):pii: 19025.

10. **Siebenga JJ**, Vennema H, Zheng DP, Vinje J, Lee BE, Pang XL, Ho EC, Lim W, Choudekar A, Broor S, Halperin T, Rasool NB, Hewitt J, Greening, GE, Jin M, Duan ZJ, Lucero Y, O'Ryan M, Hoehne M, Schreier E, Ratcliff RM, White PA, Iritani N, Reuter G, Koopmans M. Norovirus Illness Is a Global Problem: Emergence and Spread of Norovirus GII.4 Variants, 2001-2007. J Infect Dis 2009 Sep 1;200(5):802-12.

11. **Siebenga JJ**[*], Lemey P[*], Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. PLoS Pathog 2010;6(5):e1000884.
[*]JJ Siebenga and P Lemay contributed equally to this study

12. Svraka S, Vennema H, van der Veer B, Hedlund KO, Thorhagen M, **Siebenga J**, Duizer E, Koopmans M. Epidemiology and genotype analysis of emerging sapovirus-associated infections across Europe. J Clin Microbiol 2010 Jun;48(6):2191-8..

13. Rondy M, Koopmans M, Rotsaert C, Van Loon T, Beljaars B, Van Dijk G, **Siebenga J**, Svraka S, Rossen JW, Teunis P, et al. Norovirus disease associated with excess mortality and use of statins: a retrospective cohort study of an outbreak following a pilgrimage to Lourdes. Epidemiol Infect 2010 May 24;1-11.

14. Sukhrie FHA, **Siebenga JJ**, Beersma MFC, Koopmans M. Chronic shedders as reservoir for nosocomial transmission of norovirus. Accepted for publication.

PhD portfolio Joukje Siebenga

RIVM
Center for Infectious Disease Control, (CIb)
Laboratory for Infectious Diseases and Perinatal Screening, (LIS)
Virology Laboratory (VIR)

Erasmus MC
Department of Virology

Research School:	Postgraduate School Molecular Medicine
Research Period:	2005-2010
Promotor:	Prof.dr. Marion P.G. Koopmans DVM
Co-promotor:	Dr. Harry Vennema

Scientific Presentations

2009 Attended Designing the **Microbial Research Commons**: An International Symposium, Washington DC, USA.

2009 **Scientific Spring Meeting NVMM**, Papendal, the Netherlands. Norovirus strain evolution, impact on pathogenesis. (oral presentation).

2009 5th and Final Annual Meeting of the **Foodborne Viruses in Europe Network, Uppsala, Sweden**. The global epidemiology of the GII.4 strains. (oral presentation) What phylodynamic analyses can tell us about norovirus history. (oral presentation)

2008 **Epidemics**[1], Asilomar, CA, USA. Population dynamics of epidemic norovirus lineages, the GII.4 genotype further explored. (oral presentation)

2008 **Dutch Annual Virology Symposium**, Amsterdam, the Netherlands. Noroviruses; Epochal Evolution and Prolonged Shedding in High Risk Patients; a role for immunity. (oral presentation)

2008 **Institute of Medical and Veterinary Science**, Adelaide, Australia. Norovirus; global strains with local consequences. (oral presentation, on invitation of Dr. Rodney Ratcliff)

2007 **3rd International Calicivirus Meeting**, Cancun, Mexico. Global molecular epidemiology of GII.4 variants, 2001 – 2007. (oral presentation) Prolonged NoV Infections: Prevalence and in vivo Evolution. (poster presentation)

2007 4th Annual Meeting of the **Foodborne Viruses in Europe Network, Pécs, Hungary**. Prolonged NoV Infections: Prevalence and in vivo Evolution. (poster presentation) Global molecular epidemiology of GII.4 variants, 2001 – 2007. (oral presentation)

2006 3rd Annual Meeting of the **Foodborne Viruses in Europe Network, Rome, Italy**.

Genetic drift in the norovirus GGII4 capsid protein. (poster presentation)

2006 **American Society for Virology Meeting**, Madison, WI, USA. Genetic drift in the norovirus GGII4 capsid protein. (oral presentation).

2005 2nd Annual Meeting of the **Foodborne Viruses in Europe Network, Ljubljana**, Slovenia. Studying biological properties of emerging norovirus variants. (oral presentation)

Courses

2010 **BioBusiness Summer School.** Covering a.o. product development, patents & licenses, from start-up to IPO, biopharma business models, new market opportunities, entrepreneurship and emerging diseases.

2008-present **Journal club meetings**, bi-monthly. Discussions on phylogeny, phylodynamics, molecular adaptation.

2008 **Course in phylogeny**. International training course in Pathogen Phylogeny, provided by the Postgraduate School Molecular Medicine and the Department of Virology, Erasmus MC.

2007 **English Biomedical Writing and Communication**, provided by the Postgraduate School Molecular Medicine.

2006-2009 Several courses provided by **Proneri**, RIVM, including Mindmapping, Speedreading, Poster-making and Presenting, Career Planning, Time Management.

2006 **Course in Virology**. One-week international course in General Virology provided by the Postgraduate School Molecular Medicine and the Department of Virology, Erasmus MC

2005 International training workshop on the use of **GelCompar II and BioNumerics**.

2005-present **Department meetings** of the Virology Department every two weeks.

2005-present Monthly meetings of **LIS**.

Symposia

2010 **Dutch Annual Virology Symposium**.

2008 **Norovirus know-how**: Ontwikkelingen binnen en buiten het laboratorium, provided by the CIb RIVM

2007 Symposium **Post-infectious Diseases**, provided by the Postgraduate School Molecular Medicine

Collaborations

- **Foodborne Viruses in Europe Network**. Collaboration on norovirus strain

diversity and prevalence in Europe.

- **Noronet**. Setting up of collaboration on norovirus strain diversity and prevalence in Europe.
- **Dr. P. Lemey**. REGA institute, Leuven, Belgium. Collaboration on Norovirus Evolution and Phylodynamics.
- **Prof. Dr. R. Siezen**. Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands. Collaboration on structure predictions of norovirus capsid protein.
- **Prof. Dr. K.O. Hedlund**. Swedish Institute for Infectious Disease Control, Solna, Sweden. Norovirus evolution in chronically infected patients.
- **Dr. Jacques Rohayem**. Expression and Activity of Norovirus non-structural proteins.
- **Prof. Dr. I.N. Clarke**. University of Southampton Medical School, Southampton, United Kingdom. Collaboration on Expression of Norovirus VLPs.
- **Dr. M.F.C. Beersma**. Erasmus Medical Centre, Rotterdam. Collaboration on the study of chronic norovirus shedding in hospital patients.
- **Dr. L.C. Lindesmith**. University of North Carolina, Chapel Hill, NC, USA. Collaboration on norovirus VLP-ligand specificity.

Miscellaneous

- Organized **international expert meeting GESTURE**: Global exchange of microbial sequences to underpin response to health threats (6-8 December 2009). Wrote background documentation for this meeting, and presented this.
- Awarded **CIb Young Researcher Prize 2008** (€ 2500)
- Founding president of the **PhD candidate Network for RIVM, NVI and MNP, Proneri**. Founding of Proneri, positioning it within the organization, organizing several PhD Days, comprising workshops and informal moments.
- **Reviewer** for Journal of Infectious Disease, Journal of Medical Virology, Future Virology, Journal of Clinical Virology, Infection, Genetics and Evolution
- Recipient of **travel grant** for the 3rd International Calicivirus Meeting, Cancun, Mexico (US$ 1000).
- **Supervision** of MSc and Higher Laboratory Education student-interns.
- Board Member of **Royal Dutch Student Rowing Association (KNSRB)** (2005-2008). A.o. organization of the Varsity rowing competition.
- Board Member of **Minerva Rowing Club** (2005-2009). A.o. annual organization of a team of young rowers to compete at the Head of the Charles competition, Boston, USA.




Phylogenetic tree *Minimum Spanning tree* Neighbor Joining tree *Ladder-like tree* Maximum Clade Credibility tree *tree topologies and branch lengths* SYM+I+Γ4 *TreeCon* Treeview *Posterior distribution of trees*





Cpt 1 Royal Botanical Gardens, Sydney, Australia. Queensland Bottle Tree, *Brachychiton rupestris.*

Cpt 2 Dogon Vallei, Mali. Baobab tree, *Adansonia digitata.*

Cpt 3 Royal Botanical Gardens, Sydney, Australia. Gum tree, *Eucalyptus.*

Cpt 4 Dogon Vallei, Mali. Two trees on one trunk.

Cpt 5 Royal Botanical Gardens, Sydney, Australia. Ponytail Palm, *Beaucarnea recurvata.*

Cpt 6 Royal Botanical Gardens, Sydney, Australia. Dragon's Blood tree, *Dracaena draco.*

Cpt 7 Yosemite National Park, California, USA.

Cpt 8 Royal Botanical Gardens, Sydney, Australia. Moreton Bay Fig, *Ficus macrophylla.*

Cpt 9 Royal Botanical Gardens, Sydney, Australia. Port Jackson Fig, *Ficus rubiginosa.*



Notes