

## Evolutionary Dynamics of GII.4 Noroviruses over a 34-Year Period<sup>†</sup>

Karin Bok,\* Eugenio J. Abente, Mauricio Realpe-Quintero,‡ Tanaji Mitra,  
Stanislav V. Sosnovtsev, Albert Z. Kapikian, and Kim Y. Green

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland

Received 29 April 2009/Accepted 2 September 2009

**Noroviruses are a major cause of epidemic gastroenteritis in children and adults, and GII.4 has been the predominant genotype since its first documented occurrence in 1987. This study examined the evolutionary dynamics of GII.4 noroviruses over more than three decades to investigate possible mechanisms by which these viruses have emerged to become predominant. Stool samples ( $n = 5,424$ ) from children hospitalized at the Children's Hospital in Washington, DC, between 1974 and 1991 were screened for the presence of noroviruses by a custom multiplex real-time reverse transcription-PCR. The complete genome sequences of five GII.4 noroviruses (three of which predate 1987 by more than a decade) in this archival collection were determined and compared to the sequences of contemporary strains. Evolutionary analysis determined that the GII.4 VP1 capsid gene evolved at a rate of  $4.3 \times 10^{-3}$  nucleotide substitutions/site/year. Only six sites in the VP1 capsid protein were found to evolve under positive selection, most of them located in the shell domain. No unique mutations were observed in or around the two histoblood group antigen (HBGA) binding sites in the P region, indicating that this site has been conserved since the 1970s. The VP1 proteins from the 1974 to 1977 noroviruses contained a unique sequence of four consecutive amino acids in the P2 region, which formed an exposed protrusion on the modeled capsid structure. This protrusion and other observed sequence variations did not affect the HBGA binding profiles of recombinant virus-like particles derived from representative 1974 and 1977 noroviruses compared with more recent noroviruses. Our analysis of archival GII.4 norovirus strains suggests that this genotype has been circulating for more than three decades and provides new ancestral strain sequences for the analysis of GII.4 evolution.**

Noroviruses are a major cause of epidemic gastroenteritis in children and adults. They are responsible for nearly half of all gastroenteritis cases and for more than 90% of nonbacterial gastroenteritis epidemics worldwide (38). In addition, noroviruses have emerged as the second most important cause of diarrhea in children, following rotaviruses (38). Noroviruses have been described as quasispecies on the basis of their population diversity, and the constant generation of genetic and antigenic heterogeneity may allow them to persist in human populations by evading the immune response (4, 12, 19). Although noroviruses were first detected in samples derived from an outbreak at a school in Norwalk, OH, in 1968 (30), little is known about the epidemiological dynamics responsible for the wide range of genetic and antigenic variation. A thorough understanding of the evolution and mechanism of persistence of these viruses is critical to developing adequate control strategies such as antiviral drugs and vaccines.

The genus *Norovirus* belongs to the family *Caliciviridae* along with three genera, *Sapovirus*, *Vesivirus*, and *Lagovirus*. Norovirus virions are nonenveloped and approximately 30 to 35 nm in

diameter. The icosahedral capsid surrounds a positive-sense single-stranded RNA genome covalently linked to VPg at the 5' end, polyadenylated at the 3' end, and approximately 7.7 kb in length. The RNA genome is organized into three open reading frames (open reading frame 1 [ORF1], ORF2, and ORF3). ORF1 (in genogroup II noroviruses) encodes a large nonstructural polyprotein that is processed by the virus-encoded proteinase into intermediate precursors and final nonstructural protein products (NS1-2<sup>N-term</sup>, NS3<sup>NTPase</sup>, NS4<sup>P20</sup>, NS5<sup>VPg</sup>, NS6<sup>Pro</sup>, and NS7<sup>Pol</sup>). ORF2 and ORF3 encode major (VP1) and minor (VP2) capsid proteins, respectively. Noroviruses bind saccharides of the histoblood group antigens (HBGAs) as part of the mechanism that mediates viral entry into the epithelial cells of the gastrointestinal tract (25). The protruding (P) domain of the VP1 protein has been cocrystallized with HBGAs to identify the amino acids involved in this interaction, and two sites (interaction site 1 and 2) have been mapped that participate in trisaccharide A and B binding (8).

The genus *Norovirus* has been divided into five major genogroups (G), designated genogroup I (GI) through GV (49). GI and GII contain the majority of norovirus strains associated with human disease. In order to facilitate studies of the molecular epidemiology of the noroviruses, a genetic typing system was developed by Zheng et al. that further grouped strains in GI and GII into 8 and 17 genotypes, respectively (49). A single genotype, GII.4, has been associated with the majority of global outbreaks since the mid-1990s, when active surveillance was initiated using molecular diagnostic techniques (32, 36). The Foodborne Viruses in Europe network reported that the majority of gastroenteritis outbreaks in 13 countries during 5

\* Corresponding author. Mailing address: Norovirus Gastroenteritis Unit, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 50 South Drive, Bldg. 50, Room 6316, Bethesda, MD 20892. Phone: (301) 594-1633. Fax: (301) 480-5031. E-mail: bokk@niaid.nih.gov.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

‡ Present address: Research & Development Department, Boehringer-Ingelheim Vetmedica, Guadalajara, Jalisco, Mexico.

<sup>§</sup> Published ahead of print on 16 September 2009.

years of surveillance (July 2001 to June 2006) were due to this single genotype (31). Genotype II.4 has also played a predominant role in sporadic gastroenteritis occurring in both developed and developing countries, as shown in recent reports from Italy and Nicaragua, where it was responsible for 81% and 68% of norovirus gastroenteritis cases, respectively (6, 42).

The predominance of GII.4 noroviruses, along with the periodic emergence of GII.4 variants associated with a sharp increase in the number of reported illnesses, has prompted intense study of this genotype since the mid-1990s, when it was first recognized as a major epidemic strain (1, 12, 32, 36, 43, 45). Lindesmith et al. analyzed the temporal evolution and HGBA binding characteristics of representative GII.4 noroviruses circulating since 1987 and proposed the existence of five clusters that have likely evolved through antigenic drift and that are serially replaced over time with intermediate periods of stasis (32). The evolutionary pattern described in their study began with a cluster represented by the Camberwell virus, associated with gastroenteritis in Australia in 1994 (10), and the MD145 virus, associated with gastroenteritis in a Maryland nursing home in 1987 (21). Little is known about the GII.4 noroviruses that may have circulated prior to this time, limiting the understanding of the evolution of this important genotype. In order to investigate whether the GII.4 genotype had been an important cause of gastroenteritis earlier than presently documented and to study its evolution over time, we analyzed one of the oldest epidemiologic studies of infantile gastroenteritis available. The Children's Hospital National Medical Center in Washington, DC (CHDC) study was initiated in 1974 in order to investigate the etiology of severe gastroenteritis in infants and young children (5). From the analysis of CHDC archival stool specimens, we identified GII.4 noroviruses dating back to 1974, which enabled a high-resolution phylogenetic analysis of GII.4 noroviruses spanning 34 years (1974 to 2007). Our data support the emergence of an ancestral GII.4 genotype in the 1960s that has evolved at a rate of approximately  $4.3 \times 10^{-3}$  nucleotide substitutions/site/year. In addition, modeling predictions of the GII.4 capsid showed evidence for structural shifts in the capsid architecture over time which did not affect HGBA binding but may have contributed to the emergence of new epidemic strains.

#### MATERIALS AND METHODS

**Study design.** A study was conducted at the Children's Hospital National Medical Center in Washington, DC (CHDC, previously referred to as CHNMC in reference 5), from 1974 to 1991 to examine the etiology of gastroenteritis in infants and young children (5). One or more stool specimens, diaper scrapings, and/or rectal swab samples were collected from control cases (infants and young children hospitalized for reasons other than diarrhea) and diarrhea cases from infants and young children hospitalized with diarrhea and/or vomiting of 4 days duration or less. Samples were stored as undiluted stool samples or 2% fecal filtrate samples at  $-80^{\circ}\text{C}$  (29). Samples collected from 1974 to 1981 and from 1987 to 1991 were available for analysis in the present study.

The study at the Children's Hospital National Medical Center in Washington, DC, was originally supported by several grants (including AI01528-17/10) and contracts (NIH-NIAID-71-2091) with the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH). The NIH Institutional Review Board was notified of the research plan, and the samples were considered exempt from board approval for this study.

**RNA extraction.** Stool samples were prepared as 10% suspensions in phosphate-buffered saline (PBS). Stool sample material contained in diaper scrapings and rectal swabs was resuspended in PBS. Available fecal filtrate samples were processed without further manipulation. For diagnostic purposes, the RNA

TABLE 1. Primers and probes designed for the one-step multiplex real-time RT-PCR

Primer or probe	Sequence (5'-3')	Nucleotide positions <sup>a</sup>
MGBGI_F	ATGTTCCGCTGGATGC	5285-5300
MGBGI_R	CCTTAGACGCCATCWTTCAT	5360-5376
MGBGI_Probe	VIC-TGGACAGGRGATC-MGB	5321-5333
MGBGII_F	ATGTTYAGRTGGATGAGRTT	5012-5031
MGBGII_R	ACGCCATCTTCATTCACA	5080-5097
MGBGII_Probe	FAM-TGGGAGGGCGATC-MGB	5048-5060

<sup>a</sup> GI primer and probe nucleotide positions are based on those of Hu/NoV/GI.1/Norwalk virus/1968/US; GII primer and probe nucleotide positions are based on those of Hu/NoV/GII.4/MD145-12/1987/US.

present in 5,424 samples was extracted using Trizol following the manufacturer's instructions (Invitrogen, Carlsbad, CA). Further RNA extraction was required for samples used to sequence the complete norovirus genome as described previously (21). Briefly, 400  $\mu\text{l}$  of the 10% dilution or filtration material was extracted in an equal volume of Genetron (1,1,2-trichloro-1,2,2-trifluoroethane; Aqua Solutions, Deer Park, TX). The RNA present in 100  $\mu\text{l}$  of the Genetron-extracted sample was further purified according to instructions in the RNeasy kit (Qiagen, Valencia, CA). The purified RNA was stored at  $-70^{\circ}\text{C}$ .

**Norovirus diagnosis: one-step multiplex real-time RT-PCR design and optimization.** A total of 5,424 samples were screened for the presence of norovirus GI and GII RNA using a multiplex real-time reverse transcription-PCR (RT-PCR) in a one-step format. Because of the sequence diversity among norovirus genotypes, only short stretches of conserved sequences were available for probe design. Two 13-nucleotide probes were designed on the basis of homologous sequences in the polymerase region within GI and GII sequences, respectively. Each probe was then conjugated to minor groove binding groups (MGB; Applied Biosystems, Foster City, CA) in order to increase the specificity and annealing temperature of the probes and to two different fluorophores, 5-carboxyfluorescein (FAM) and VIC to allow for GI and GII discrimination in a single reaction (Applied Biosystems) (Table 1). The reaction was optimized for  $\text{MgCl}_2$ , primer and probe concentration, and annealing temperature. The Brilliant II core one-step RT-PCR kit (Stratagene, La Jolla, CA) was used to prepare 800-sample reaction mix batches containing the following:  $1 \times$  core RT-PCR buffer, 3 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  of each deoxynucleoside triphosphate, 300 nM ROX reference dye, 1.25 U of SureStart *Taq* DNA polymerase, 1  $\mu\text{l}$  of reverse transcriptase, 800 nM each of primer MGBGI\_F and MGBGI\_R, 500 nM primer MGBGII\_F, 300 nM primer MGBGII\_R, and 250 nM of each MGB probe (GI and GII; Table 1). Each reaction mix batch was tested by establishing a triplicate standard curve with known concentrations of plasmids containing the full-length genome of GI and GII noroviruses (20, 21), together with known positive RNA samples representing both genogroups. The reaction mix was then aliquoted in 1-ml portions and tested for stability after one freeze-thawing cycle. Finally, 5  $\mu\text{l}$  of RNA was mixed with 20  $\mu\text{l}$  of reaction mix and then incubated at  $45^{\circ}\text{C}$  for 30 min and then at  $95^{\circ}\text{C}$  for 10 min, followed by 45 cycles of PCR, with 1 cycle consisting of 30 s at  $95^{\circ}\text{C}$ , 1 min at  $52^{\circ}\text{C}$ , and 30 s at  $72^{\circ}\text{C}$ , in an ABI 7900 HT instrument (Applied Biosystems).

Samples that showed an amplification signal in the multiplex real-time RT-PCR were further evaluated by a regular RT-PCR using primer pairs Mon431/Mon432 and Mon 433/Mon434 as previously described (21). Amplicons obtained from norovirus-positive stool samples were isolated in agarose gels and further purified with a gel extraction kit (Qiagen). The nucleotide sequence was determined directly from the purified DNA amplicon using the Big Dye terminator cycle sequencing ready reaction kit (Applied Biosystems), and the sequencing products were resolved on an ABI PRISM 3730 automated DNA sequencer (Applied Biosystems). Noroviruses were provisionally genotyped by submitting the partial polymerase sequence to a BLAST search (35).

**Complete genome sequencing.** Five out of eight GII.4-positive samples were selected for complete genome sequencing on the basis of the quality and quantity of the available RNA: Hu/NoV/GII.4/CHDC2094/1974/US (CHDC2094/1974), Hu/NoV/GII.4/CHDC5191/1974/US, Hu/NoV/GII.4/CHDC4871/1977/US, Hu/NoV/GII.4/CHDC4108/1987/US, and Hu/NoV/GII.4/CHDC3967/1988/US. The oldest GII.4 strains were CHDC2094/1974, collected at the end of the 1973 to 1974 winter season, and Hu/NoV/GII.4/CHDC5191/1974, collected at the beginning of the next winter season (1974 to 1975). Consensus sequences were obtained for each sample via primer walking every  $\sim 500$  nucleotides. The 5' and 3'

prime sequences were obtained using 5' or 3' rapid amplification of cDNA end kits (Roche Applied Science, Indianapolis, IN) according to the manufacturer's instructions. The sequence fragments were aligned into a contig to obtain the final sequence with SeqMan v8.0 (Lasergene, Madison, WI). Complete genome sequences were determined independently two times for each sample: once from the RNA in the original Trizol extraction and once from RNA purified again from the original stool specimen by using the RNeasy kit as described above.

**Sequence alignment.** The five GII.4 genome sequences from this study were aligned to norovirus genome sequences available in the GenBank database that represented all norovirus genogroups (for a total of 56 norovirus genomes). Sequences were aligned using Clustal X 1.8, and alignment was manually edited with MegAlign v8.0 (Lasergene, Madison, WI).

The evolutionary analysis was performed on a total of 185 VP1 nucleotide sequences corresponding to GII.4 noroviruses from this study together with sequences available in the GenBank database as of August 2008 (see Table S1 in the supplemental material). The alignment was created and edited as described above, and the corresponding dates were assigned using the BEAUTI application, part of the BEAST package (14, 15).

**Phylogeny.** A phylogenetic tree was inferred by maximum likelihood reconstruction based on the nucleotide alignment of the whole-genome sequences as implemented in the PhyML software (23). The statistical significance of phylogenies constructed was estimated by bootstrap analysis with 100 pseudoreplicate data sets. The tree was displayed with the Treeview program (37).

**Evolutionary analysis of GII.4 sequences.** The parameter values for the best-fit model of nucleotide substitutions were determined using Akaike information criterion as implemented in MODELTEST (41). Rates of nucleotide substitution per site and the time to the most recent common ancestor (MRCAs) were estimated using Bayesian Markov chain Monte Carlo as implemented in the BEAST software package (14, 15). The model of substitution was the general time reversible, and the data set was partitioned by codon positions. Since no demographic model is available from organisms showing cyclic annual/seasonal behavior, the Bayesian Skyline model for population growth was selected (16). BEAST files were run three times each (for a total of 9 runs), either assuming a constant rate of evolution across the tree (constant molecular clock), or using relaxed-clock models (lognormal and exponential) which model a molecular rate that varies among lineages (13). In all cases, statistical uncertainty in parameter values across the sampled trees is given by the 95% highest probability density values. For all models, chains were run until convergence was achieved (as assessed using the TRACER program, part of The BEAST package). The BEAST analysis was also used to infer a maximum credibility clade (MCC) tree, where the branch length was calibrated to reflect temporal patterns.

**Positive selection and recombination analysis in GII.4 sequences.** Site-specific positive selection was also evaluated as rates of nonsynonymous or synonymous substitutions among 185 sequences present in the GII.4 alignment. Our analysis was performed using the single likelihood ancestor counting method as available in the HyPhy package (40). The general time reversible substitution model was chosen, and a neighbor-joining tree was used as an input, with a limit of significance level of 0.25. The presence of recombination was evaluated using the Recombination Detection Program (34).

**Protein structure modeling.** The amino acid differences between Hu/NoV/CHDC2094/GII.4/1974/US and VA387 were inserted into the pdb files for the corresponding P-domain structure (2obr: crystal structure of the P domain of norovirus VA387; 2obt: crystal structure of the P domain of norovirus VA387 in complex with blood group trisaccharide type B) (8). The new structure files were loaded and prepared for minimization using SYBYL8.0 (Tripos International, St. Louis, MO). Protein preparation for minimization consisted of adding AMBER7 ff99 atom types to all atoms. For standard residues, SYBYL8.0 used the AMBER7 ff99 parameter file. For the polysaccharides, SYBYL8.0 used an atom-typing algorithm to assign AMBER7 ff99 atom types; adding AMBER 7 ff99 charges to the protein and Gasteiger-Marsili charges to the polysaccharides; optimizing H-bonding of side chain amides.

The modified protein files were then minimized using the Powell conjugate gradient method with the AMBER 7 ff99 force field and terminated at a gradient of 0.05 kcal/mol (SYBYL8.0). Minimization was done with an effective dielectric of ~32 to reproduce the shielding of charges. A cutoff of 8 Å for nonbonded interactions was used.

**Expression and purification of rVLPs.** In order to express recombinant virus-like particles (rVLPs), the ORF2 and ORF3 genes of norovirus strains Hu/NoV/GII.4/CHDC5191/1974/US (nucleotides 5085 to 7510), Hu/NoV/GII.4/CHDC 4871/1977/US (nucleotides 5085 to 7510), Hu/NoV/GI.1/Norwalk virus/1968/US [nucleotides 5358 to 7654 plus poly(A)<sub>10</sub>], and Hu/NoV/GII.1/Hawaii/1971/USA (nucleotides 5085 to 7471) were RT-PCR amplified and cloned into either the FastBac or the Gateway vector, pENTR (Invitrogen). Recombination of plasmid

DNA with baculovirus DNA was performed using the Baculodirect kit (Invitrogen), and a baculovirus stock was obtained following transfection of the recombination product into Sf-9 cells as recommended by the manufacturer (serum-free Sf9 cells; Invitrogen). The baculovirus stock was used to infect Sf-9 suspension cultures for VLP production. Culture medium from baculovirus-infected cells was layered onto a 25% sucrose cushion and subjected to centrifugation in a SW28 rotor at 76,200 × g for 4 h at 4°C (22). The resulting pellets were dissolved in PBS (pH 7.5) and further purified through a cesium chloride (CsCl) gradient by centrifugation in a SW55 rotor at 218,400 × g overnight. The collected fractions (~1.3 g/ml density) were dialyzed against PBS, and the protein concentration was determined with a commercial Bradford assay kit (Pierce, Rockford, IL). The yields ranged from 2 to 3 mg of each VLP/150 ml of cell culture infected at a multiplicity of infection of ≥5. The presence of VLPs was confirmed by electron microscopy.

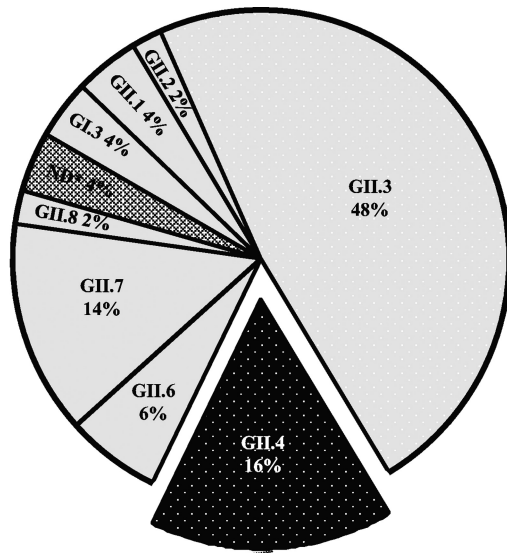
**Binding of archival norovirus rVLPs with synthetic ABH histoblood group antigens.** The rVLPs were screened for binding to a panel of HBGA-associated oligosaccharides as described elsewhere (8, 32). Briefly, 96-well plates (polyvinyl microtiter plates, Dynatech, Chantilly, VA; neutravidin-coated plates, Pierce) were coated for 2 h at room temperature with either human serum albumin coupled to specific synthetic oligosaccharides or synthetic oligosaccharide-biotin conjugates containing human HBGA epitopes at a concentration of 20 and 10 µg/ml, respectively. After blocking with 5% Biotin, 1.25 µg/ml of norovirus rVLPs was added and incubated for 1 h at room temperature. The binding of captured rVLPs was determined by incubation with genotype-specific hyperimmune sera generated in rVLP-immunized guinea pigs (1:10,000 dilution), followed by incubation with a peroxidase-conjugated goat anti-guinea pig serum (KPL, Gaithersburg, MD). The peroxidase substrate 2,2'-azino-bis(3-ethylbenzothiazolinesulfonic acid) (ABTS) (KPL) was used, and plates were read at 405 nm in a Dynex Technologies Revelation 4.25 plate reader (Dynatech). The oligosaccharides examined in this study were biotin conjugates type A, type B, Le<sup>a</sup>, Le<sup>b</sup>, Le<sup>x</sup>, Le<sup>y</sup>, H type 1, H type 3, and human serum albumin conjugate H type 2 (GlycoTech Corporation, Rockville, MD).

**Nucleotide sequence accession numbers.** The complete genome sequences were submitted to the GenBank database under the following accession numbers: FJ537134 (Hu/NoV/GII.4/CHDC5191/1974/US), FJ537135 (Hu/NoV/GII.4/CHDC2094/1974/US), FJ537136 (Hu/NoV/GII.4/CHDC3967/1988/US), FJ537137 (Hu/NoV/GII.4/CHDC4108/1987/US), and FJ537138 (Hu/NoV/GII.4/CHDC4871/1977/US).

## RESULTS

**Identification of noroviruses in the archival stool sample collection from the Children's Hospital, Washington, DC, study (1974 to 1991).** A total of 5,424 stool samples from infants and young children hospitalized with gastroenteritis or other conditions (controls) in the CHDC study were screened for the presence of norovirus RNA by a multiplex one-step real-time RT-PCR designed to detect GI and GII noroviruses (for primer and probe sequences, see Table 1). Three hundred five (5.6%) samples yielded a positive signal in the real-time RT-PCR, but only 50 samples could be confirmed as norovirus positive by RT-PCR amplification of a portion of the polymerase gene and direct sequence analysis of the PCR products. It was possible that long-term storage or multiple freeze-thaw cycles of the samples decreased the efficiency of the diagnostic RT-PCR, and an accurate assessment of the incidence of norovirus-associated illness was not established in this study.

The norovirus-positive stool samples were obtained from children hospitalized with diarrhea, with the exception of two strains detected in stool specimens from the two control cases hospitalized for other conditions. Rotavirus was not detected in the norovirus-positive samples by RT-PCR (data not shown). Sequence analysis of norovirus-specific RT-PCR products showed evidence for the presence of several genotypes in the collection, with the distribution as follows: GII.3 (48%), GII.4 (16%), GII.7 (14%), GII.6 (6%), GII.1 (4%), GI.3 (4%), GII.2 (2%), and GII.8 (2%) (Fig. 1). The viruses from the two con-



**Hu/NoV/GII.4/CHDC2094/1974/US**  
**Hu/NoV/GII.4/CHDC5191/1974/US**  
 Hu/NoV/GII.4/CHDC1555/1975 /US  
**Hu/NoV/GII.4/CHDC4871/1977/US**  
 Hu/NoV/GII.4/CHDC2691/1978/US  
 Hu/NoV/GII.4/CHDC3311/1980/US  
**Hu/NoV/GII.4/CHDC4108/1987/US**  
**Hu/NoV/GII.4/CHDC3967/1988/US**

FIG. 1. Distribution of genotypes in the norovirus-positive samples of the study of Children’s Hospital in Washington, DC. A specific norovirus genotype was assigned to 48 of the 50 norovirus-positive samples on the basis of partial polymerase coding sequence. The GII.4 norovirus strains selected for complete genome sequencing in this study are indicated in boldface type. ND\*, a specific norovirus genotype could not be determined on the basis of the sequence available.

rol cases belonged to GII.3, and two samples (4%) could not be assigned to a known genotype, although they clustered most closely with human GII.3 noroviruses (data not shown). These data indicate that GII noroviruses have likely been the predominant genogroup for several decades in the human population and in addition show that GII.4 noroviruses were associated with disease as early as 1974.

**Archival GII.4 norovirus whole-genome analysis.** The detection of GII.4 noroviruses in the archival CHDC specimens allowed the evaluation of five new complete GII.4 genome sequences for comparison with those of approximately 50 noroviruses representing GI (human), GII (human and swine), GIII (bovine), and GV (murine) strains. A complete genome sequence for human and feline noroviruses in GIV was not available in the public database. The CHDC GII.4 viruses clustered most closely with other GII.4 strains, confirming that they belong to this genotype (Fig. 2). The CHDC samples isolated in the 1970s clustered as the ancestors of all GII.4 viruses detected subsequently, with CHDC2094/1974 representing the oldest GII.4 virus characterized thus far. In addition, the two CHDC samples isolated during the 1980s (CHDC 4108/1987 and CHDC3967/1988) grouped together with cocirculating noroviruses in that time period, such as the MD145-12 virus, associated with a 1987 gastroenteritis outbreak in a nearby Maryland nursing home (21). Norovirus sequences

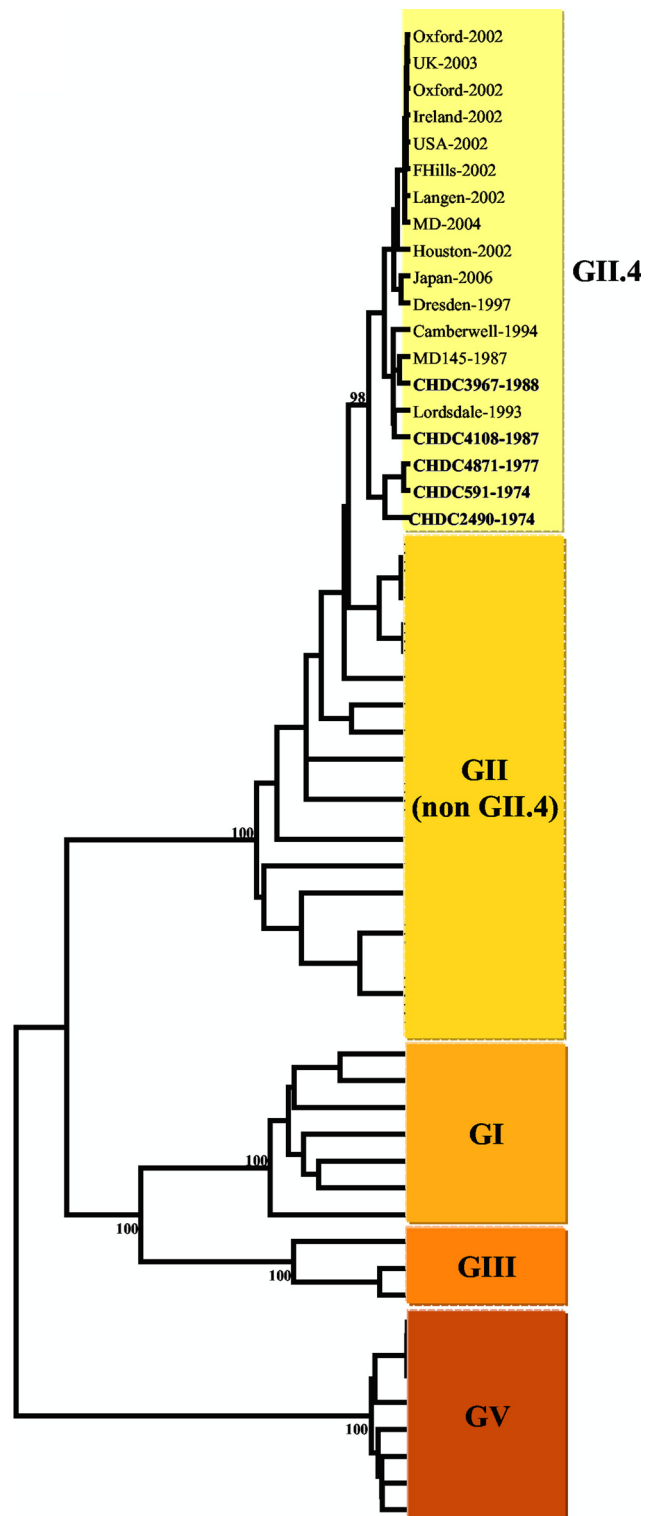


FIG. 2. Maximum likelihood analysis of complete genome sequences from noroviruses of different genogroups. Complete genomic sequences of norovirus genogroups I, II, III, and V obtained from the GenBank database (see Table S1 in the supplemental material) were compared with five GII.4 CHDC samples from this study (shown in boldface type). Bootstrap values are shown at the nodes of each selected phylogenetic group only. FHills-2002, Hu/NoV/GII.4/Farmington Hills/2002/US; Camberwell-1994, Hu/NoV/GII.4/Camberwell/1994/AUS; CHDC3967-1988, Hu/NoV/GII.4/CHDC 3967/1988/US; CHDC4108, Hu/NoV/GII.4/CHDC4108/1987/US.

TABLE 2. Nucleotide substitution rates and divergence times over 34 years for the VP1 gene of GII.4 norovirus strains

Molecular clock	Nucleotide substitution rate <sup>a</sup> (10 <sup>-3</sup> substitution/site/yr (95% HPD <sup>b</sup> ))	TMRCA <sup>c</sup> (95% HPD) by:	
		No. of yr	Date
Strict clock	4.3 (3.85–4.76)	41 (39–43)	1966 (1964–1968)
UCED <sup>d</sup>	5.6 (4.7–6.4)	40 (34–50)	1967 (1957–1973)
UCLN <sup>e</sup>	5.1 (4.4–6)	40 (34–48)	1967 (1959–1973)

<sup>a</sup> Mean rate for three individual determinations.

<sup>b</sup> 95% HPD, 95% highest probability density.

<sup>c</sup> TMRCA, time of most recent common ancestor.

<sup>d</sup> UCED, uncorrelated exponential deviation.

<sup>e</sup> UCLN, uncorrelated log normal.

showed an average of 20% and not more than 40% nucleotide diversity within a genotype, while differences between genogroups were over 40% (measured as percent nucleotide distances [data not shown]).

**Evolutionary dynamics of GII.4 noroviruses.** One hundred eighty-five GII.4 VP1 sequences from noroviruses spanning a 34-year period were submitted to an evolutionary analysis (see Table S1 in the supplemental material). A Bayesian coalescent method was used to infer the rate of evolutionary change expressed as nucleotide substitutions per site per year, using three different molecular clock models (Table 2). The most conservative method (strict clock) calculated that GII.4 noroviruses evolved at a rate of  $4.3 \times 10^{-3}$  nucleotide substitutions/site/year. This rate of evolution was comparable to the relaxed-clock estimations (uncorrelated exponential deviation and uncorrelated log normal), varying from  $4.4 \times 10^{-3}$  to  $6.4 \times 10^{-3}$  nucleotide substitutions/site/year when considering the highest probability density intervals. The age of the analyzed population was also estimated using the same Bayesian approach, expressed as the time to the most recent common ancestor. The mean age of the GII.4 population was 40 to 41 years considering the three molecular clock estimations. The strict-clock estimation dated the ancestor to 1966, which was similar to that of the relaxed-clock model, 1967.

On the basis of the evolutionary analysis of the GII.4 noroviruses, an MCC tree was inferred, showing the phylogenetic relationship of capsid sequences (Fig. 3). The MCC tree showed the CHDC GII.4 sequences from the 1970s forming a cluster that likely evolved into the GII.4 clusters previously described by Lindesmith et al. (32) that are indicated in the tree. Camberwell is the cluster with the highest intra-amino acid and nucleotide divergence value (2.8% and 7%, respectively), followed by the CHDC, Den Haag, and Hunter clusters (4.7%, 3.3%, and 3.2% nucleotide distance, respectively) (Fig. 3). The largest sequence clusters, Grimsby and Farmington Hills, showed the most conserved nucleotide distance values (2.5% and 0.8%, respectively) and minimal amino acid variation within each cluster (1.1% and 0.4%, respectively). The previous assignment of Grimsby and Farmington Hills to independent clusters might simply be due to the larger numbers of outbreak strain sequences available for analysis at that time because of the predominance of these viruses in a period of high disease prevalence and not to the actual presence of two separate clusters (32).

**Amino acid variation and positive selection in the viral structural proteins, VP1 and VP2.** The VP1 amino acid sequences from the five CHDC GII.4 noroviruses in this study were aligned with representative viruses from the other GII.4 clusters (spanning 1974 to 2005). Seventy-three variable sites were detected, representing 13.5% of the total VP1 amino acid sequence of 540 residues, and their locations in the structure of the capsid protein are shown in Fig. 4. Approximately half the variable sites (36 positions) were located in the P2 region of the capsid. Fifteen amino acid residues (indicated by an asterisk in the sequence alignment) were present in one or more CHDC viruses and were not present in viruses from subsequent years. Of these, four consecutive amino acids between positions 292 and 295 (RVGI) represented a conserved sequence that was found exclusively in the CHDC strains isolated in the 1970s. The RVGI sequence was replaced with the sequence HIVG after 1978 and has undergone further evolution at residue 294. An analysis of the VP1 amino acid residues with evidence for positive selection pressure identified four residues in the shell domain (amino acids [aa] 6, 9, 15, and 47), one in the P1 domain (aa 534), and one in the P2 region (aa 395). Four of these residues (aa 6, 9, 395, and 534) were identical to those identified previously by Lindesmith et al. (32). The CHDC GII.4 noroviruses lacked an amino acid insertion at position 394, first detected in GII.4 noroviruses of the Farmington Hills cluster (11). According to a recombination detection analysis, none of the CHDC sequences showed evidence for recombination in the polymerase-VP1 region with other genotypes (7).

A comparable rate of amino acid substitutions was observed in an alignment of sequences corresponding to the minor structural protein, VP2 (Fig. 5). The highest number of overall variations was observed between CHDC VP2 sequences and those of the clusters described in the VP1 phylogeny (26.5% [71/268 aa]). Four sites (aa 4, 155, 158, and 187) likely evolved under positive selection pressure. CHDC2094/1974 presented unique amino acid sequences in approximately half the variable sites. A possible recombination point was identified immediately upstream of the ORF3 at nucleotide 6564 in CHDC 2094/1974 by the recombination detection program. The possibility of intragenotypic recombination between ORF2 and ORF3 had been previously described for GII.4 noroviruses (28).

**Positive selection in the viral nonstructural polyprotein encoded by ORF1.** The lower number of full-length genomic sequences for GII.4 noroviruses hampered detailed evolutionary analysis of ORF1. The same positive selection detection algorithm was employed to analyze the GII.4 whole-genome sequences available. The single likelihood ancestor counting analysis indicated that amino acid residues 79 (NS1<sup>Nterm</sup>), 773 (NS4<sup>P20</sup>), 846 (NS4<sup>P20</sup>), and 1091 (NS5<sup>Pro</sup>) evolved under positive selection. The significance of variation at these residues (if any) will require further investigation of norovirus nonstructural protein structure and function.

**Capsid protein structural modeling.** It has been suggested that noroviruses bind HBGA molecules as part of the mechanism that mediates viral entry into the epithelial cells of the gastrointestinal tract (25). The structural basis of this recognition has been recently described through the cocrystallization of a GII.4 strain (VA387) with trisaccharides A and B. The first

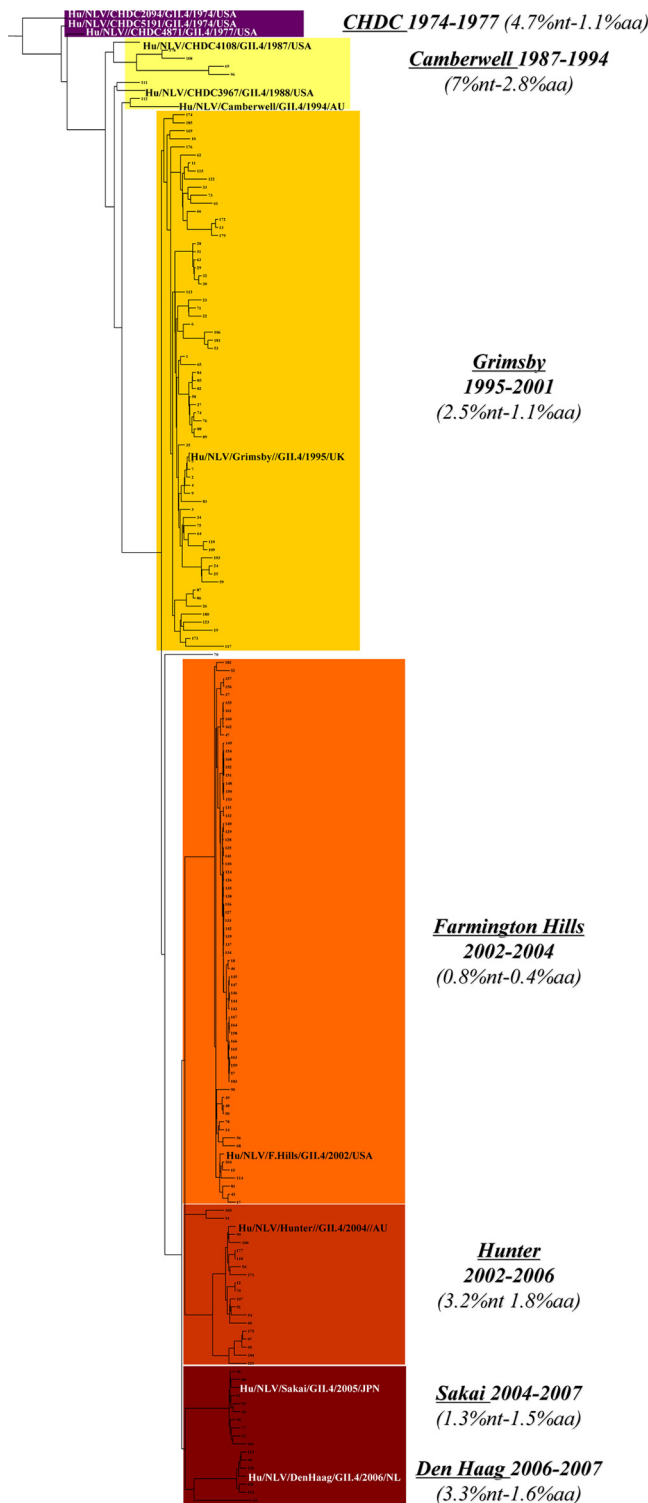


FIG. 3. Maximum credibility cladogram of GII.4 norovirus VP1 sequences, as a result of Bayesian estimation using the BEAST package. GII.4 sequences from 1987 to 2007 are shown colored in the different clusters previously proposed for GII.4 evolution (32). The CHDC sequences from the 1970s are shown in a separate cluster. The percent nucleotide (nt) and percent amino acid (aa) identities are shown in parentheses. Camberwell-1994, Hu/NoV/GII.4/Camberwell/1994/AUS; Grimsby, Hu/NoV/GII.4/Grimsby/1995/UK; Farmington Hills, Hu/NoV/GII.4/Farmington Hills/2002/US; Hunter, Hu/NoV/GII.4/Hunter/2004/AUS; Sakai, Hu/NoV/GII.4/Sakai/2005/JPN.

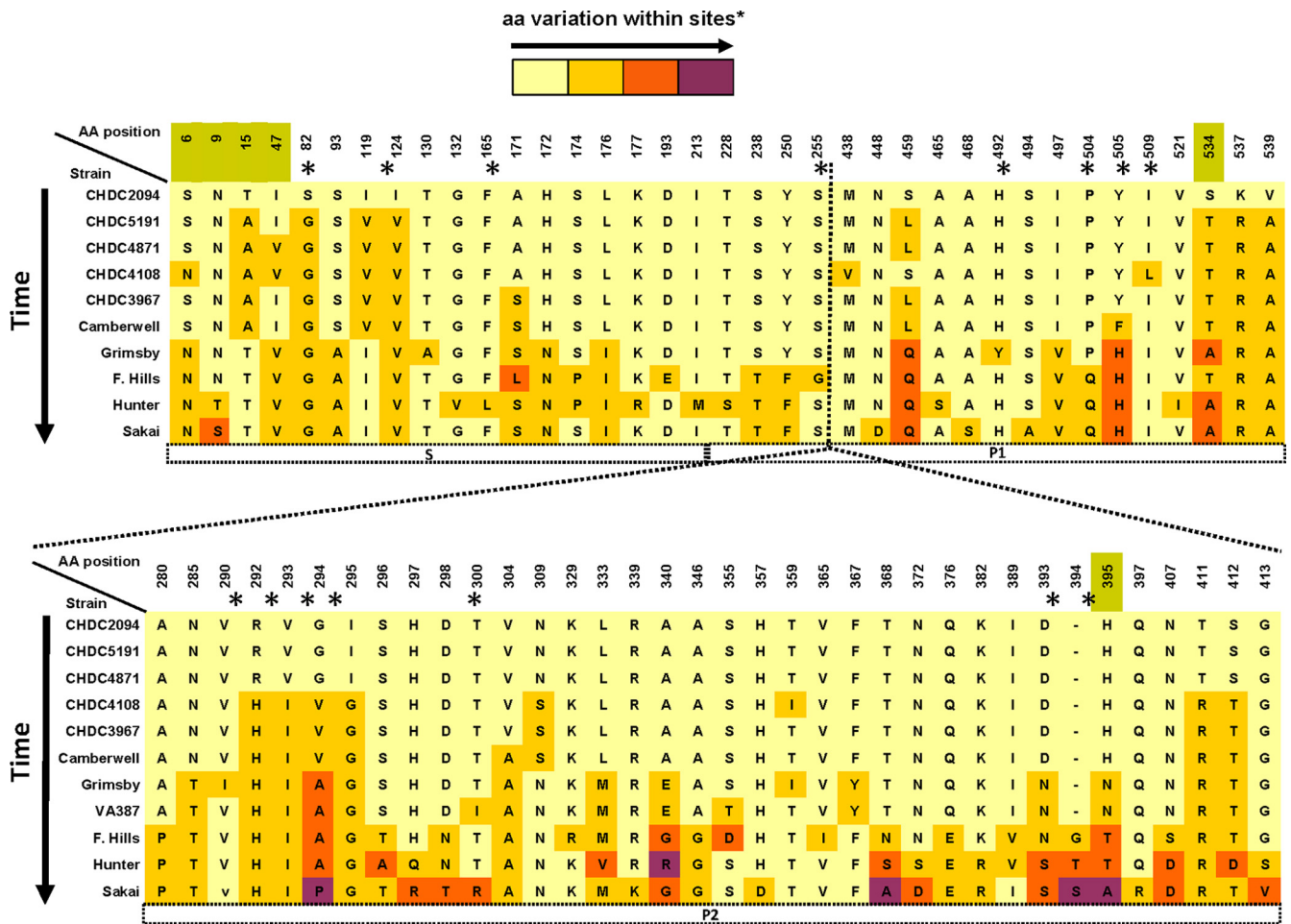
interaction site was located in the P2 region of the capsid, where the trisaccharide interacted via hydrogen bonding with the P2 domain of the viral capsid (aa 343, 344, 345, 374, 441, 442, and 443) (8). A second site also located in the P2 domain was thought to stabilize the interaction between the ligands (aa 390, 391, 392, 393, and 395). The interaction sites remained the same either with trisaccharide A or B (8). The amino acid substitutions unique to the oldest CHDC sequence (CHDC 2094/1974) were modeled onto the solved capsid protein structure of the GII.4 norovirus complexed with trisaccharide B (VA387 [Fig. 6A]). With the exception of positions 534, 537, and 539, which were not resolved in the original particle structure, all CHDC2094 substitutions were found to be surface exposed, but not directly adjacent to the mapped HBGA binding sites (8). The unique RVGI sequence formed a protrusion from the P region neighboring HBGA binding site 1. The difference in conformation and the relationship between the RVGI protrusion and the trisaccharide B binding site is illustrated in Fig. 6B, in which the two predicted protein structures (VA387 and CHDC2094) are superimposed. The RVGI sequence does not appear to overlap with the HBGA binding site, but it does show a change in structure from the HIAG sequence present in VA387.

**Archival norovirus binding to HBGA carbohydrates.** Several studies have proposed that human noroviruses are under selective pressure for binding to a wide diversity of HBGA-containing receptors in the gut (12, 32). In order to compare the HBGA binding properties of the GII.4 historical samples with those of more recently circulating strains, we examined the binding of rVLPs derived from the oldest GII.4 norovirus strains to a panel of HBGA oligosaccharides. rVLPs were generated for the CHDC5191/1974 and CHDC4871/1977 noroviruses and studied in direct HBGA binding assays (Fig. 7). Norwalk virus (GI.1) and Hawaii virus (GII.1) rVLPs were included as controls. The rVLPs from both archival GII.4 viruses exhibited binding patterns similar to those described previously for the Camberwell GII.4 cluster (32), recognizing H type 3, Le<sup>y</sup>, and B antigens. These results confirmed that the archival norovirus carbohydrate binding properties were conserved over time. In addition, these data indicate that the unique RVGI sequence does not affect the carbohydrate binding pattern.

## DISCUSSION

Noroviruses are increasingly recognized as important pathogens of gastroenteritis worldwide, but the mechanisms responsible for the emergence of successful predominant strains are poorly understood (24, 27). A one-tube multiplex real-time RT-PCR assay was developed for a large-scale analysis of RNA extracted from samples obtained up to 34 years ago from children hospitalized with diarrhea. This study reports identification of the earliest known strains of several predominant norovirus genotypes belonging to genogroups I and II. The diversity of these viruses and their ancestral relatedness to contemporary noroviruses suggest that several norovirus genotypes, including GII.4, have been circulating concurrently for at least the last few decades.

The GII.4 genotype has been the predominant pathogen in global gastroenteritis outbreaks among the 25 or more noro-

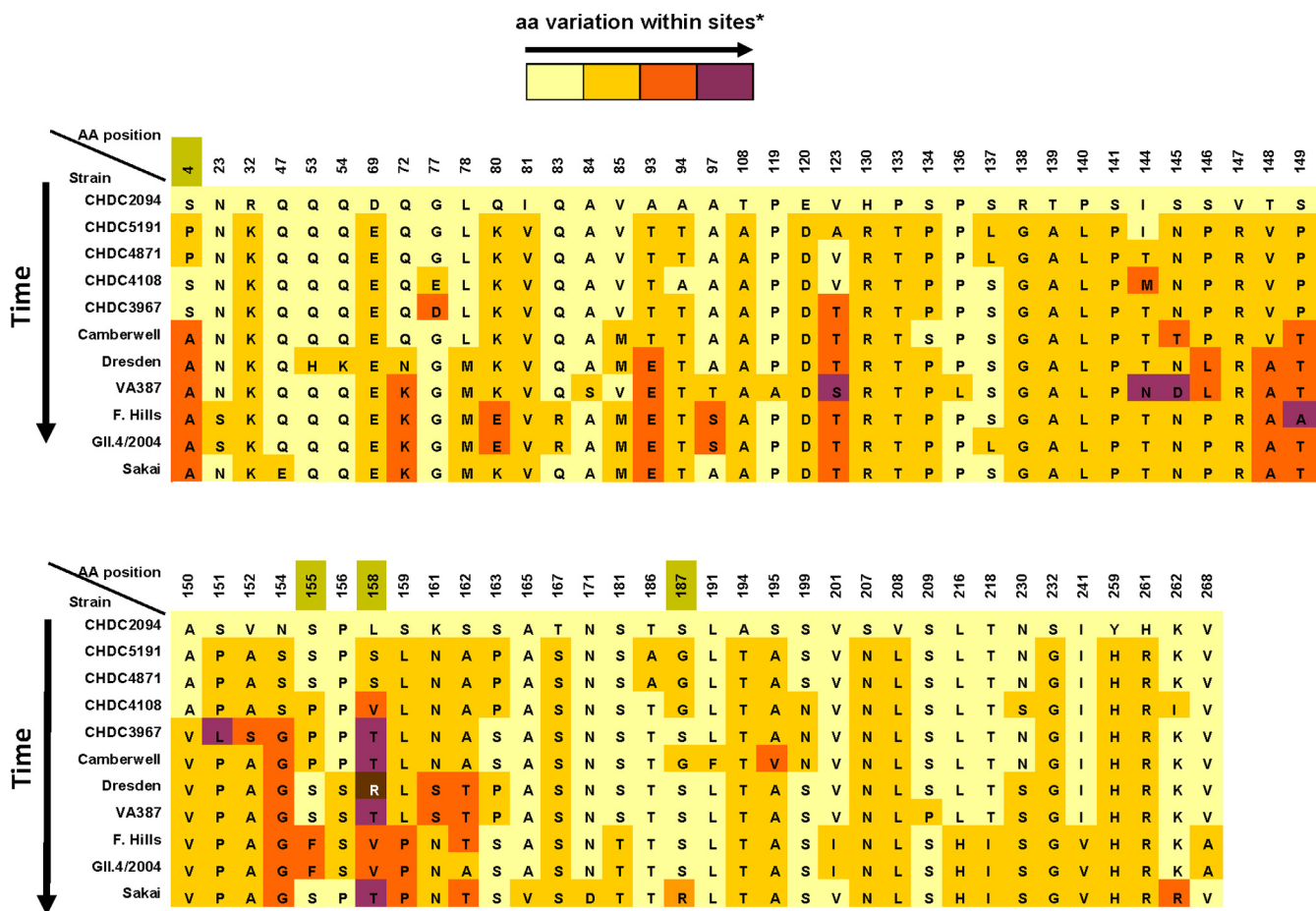


\*: Amino acid variation within individual sites compared to the oldest GII.4 strain available Hu/NoV/GII.4/CHDC2094/1974/USA.

FIG. 4. Amino acid variation of the VP1 capsid region of representative strains over time. The GII.4 strains from the study of the Children’s Hospital in Washington, DC, were compared with representative strains from the clusters previously proposed for GII.4 evolution (32) as follows: CHDC2094, Hu/NoV/GII.4/CHDC2094/1974/US; CHDC5191, Hu/NoV/GII.4/CHDC5191/1974/US; CHDC4871, Hu/NoV/GII.4/CHDC4871/1977/US; CHDC4108, Hu/NoV/GII.4/CHDC4108/1987/US; CHDC3967, Hu/NoV/GII.4/CHDC3967/1988/US; Camberwell, Hu/NoV/GII.4/Camberwell/1994/AUS; Grimsby, Hu/NoV/GII.4/Grimsby/1995/UK; F. Hills, Hu/NoV/GII.4/Farmington Hills/2002/US; Hunter, Hu/NoV/GII.4/Hunter/2004/AUS; Sakai, Hu/NoV/GII.4/Sakai/2005/JP. VA387, Hu/NoV/GII.4/VA387/1998/US was included in the P2 region alignment for comparison (see Table S1 in the supplemental material for GenBank accession numbers). Amino acids labeled with an asterisk indicate changes that are unique to the sequences of the CHDC study. Positions predicted to evolve under positive selection are highlighted in green.

virus genotypes associated with human disease. It has been postulated that the GII.4 Camberwell cluster noroviruses emerged in the 1980s, when these viruses encountered a naïve population that allowed them to establish a niche in individuals that expressed H type 3 and Le<sup>y</sup> antigens (12, 32). Proposed mechanisms for the persistence of the GII.4 noroviruses in the human population have included receptor switching and antigenic variation or a combination of the two (12). Our study showed that GII.4 noroviruses were circulating and associated with severe illness more than a decade prior to the emergence of the Camberwell cluster and that they already had the ability to bind H type 3 and Le<sup>y</sup> antigens, a property that remained intact until the appearance of the Farmington Hills cluster. This finding argues against a major shift in the HBGA binding profile of GII.4 noroviruses in the 1980s. We further examined

the mechanisms by which these viruses may have evolved into the present GII.4 clusters by structural modeling. An evolutionary trend analysis was performed on the basis of current structural knowledge of the HBGA binding sites in the viral capsid (8, 47, 48). The unique amino acid residues that were present only in sequences from the 1970s were all located in surface-exposed areas of the P particle that have not been implicated in carbohydrate binding. Moreover, the HBGA binding sites and surrounding amino acids have remained constant since the 1970s, with the exception of positions 393 to 395 that showed variation in 1997, when the Camberwell cluster was replaced with the Grimsby cluster (11). Tan et al. have performed mutational analysis of the HBGA binding sites of VA387 (47, 48) and demonstrated that binding of VA387 to HBGAs could be inhibited by mutation of position 338 from



\*: Amino acid variation within individual sites compared to the oldest GII.4 strain available Hu/NoV/GII.4/CHDC2094/1974/USA.

FIG. 5. Amino acid variation in the VP2 minor capsid protein of representative strains over time. The GII.4 strains from the study of the Children’s Hospital in Washington, DC, were compared with representative strains from the clusters previously proposed for GII.4 VP1 evolution (32) as follows: CHDC2094, Hu/NoV/GII.4/CHDC2094/1974/US; CHDC5191, Hu/NoV/GII.4/CHDC5191/1974/US; CHDC4871, Hu/NoV/GII.4/CHDC4871/1977/US; CHDC4108, Hu/NoV/GII.4/CHDC4108/1987/US; CHDC3967, Hu/NoV/GII.4/CHDC3967/1988/US; Camberwell, Hu/NoV/GII.4/Camberwell/1994/AUS; Dresden, Hu/NoV/GII.4/Dresden/1997/GER; F. Hills, Hu/NoV/GII.4/Farmington Hills/2002/US; GII.4/2004, Hu/NoV/GII.4/MD-2004/2004/US; Sakai, Hu/NoV/GII.4/Sakai/2005/JP. VA387, Hu/NoV/GII.4/VA387/1998/US was included for comparison (see Table S1 in the supplemental material for GenBank accession numbers). Positions predicted to evolve under positive selection are highlighted in green.

Thr to Ala, which also was unchanged in the archival GII.4 samples. In addition, they examined amino acids in the capsid that are sterically close and likely involved in interaction with the HBGAs and found that changes in certain positions (aa 331, 348, 346, and 389) altered the binding pattern (47, 48). Comparison of the CHDC VP1 sequences with these critical HBGA binding residues showed that these amino acids have remained conserved over the decades, with the exception of position 389 where an isoleucine was replaced with valine in the Farmington Hills and Hunter clusters, only to revert to isoleucine in the Sakai cluster. This observation suggests that the HBGA binding sites remained constant, but it does not rule out a role for evolution at other sites that could affect the overall conformation of binding to carbohydrate ligands or the receptor. In addition, a major limitation in the study of norovirus evolution is the inability to analyze functional neutralization sites in the VP1 protein. When these assays become avail-

able, it may be possible to visualize the mechanisms of immune selection in light of the modeled structural changes described in this study.

A key component in establishing evolutionary mechanisms of the noroviruses is knowledge of the rate at which noroviruses generate genetic diversity that leads to fixed mutations in the viral population. Most RNA viruses evolve at a rate of approximately  $10^{-3}$  nucleotide substitutions/site/year (17). We found that the capsid region of GII.4 noroviruses evolved at a comparable rate over the last 34 years, which might have allowed the virus to generate replacement clusters and constantly evade the immune response, within the same genotype. However, in many viruses, an immune-driven evolution pattern correlates with the presence of positively selected sites in exposed regions of the capsid due to the selective pressure of neutralizing antibodies (44). Of interest, four of the six positively selected positions of the capsid protein identified in our



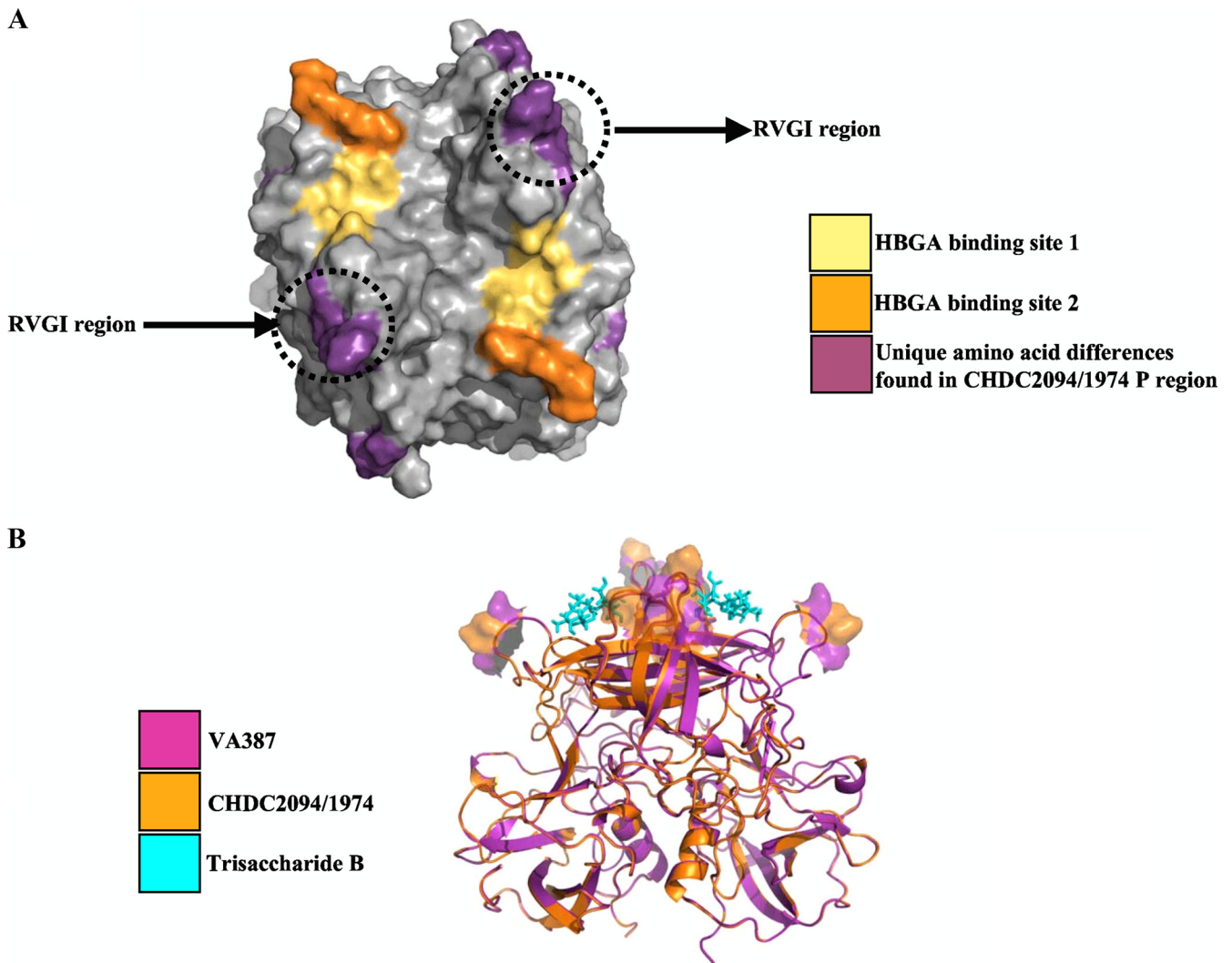


FIG. 6. (A) Minimized structure of the P region of Hu/NoV/GII.4/CHDC2094/1974/US based on the protein structure of Hu/NoV/GII.4/VA387/1998/US. The corresponding sequence changes were added to the original structure, and the resulting P-region structure was minimized. The unique amino acid changes found in CHDC2094/1974 (marked with an asterisk in Fig. 4) were mapped into the new structure (purple). The two carbohydrate binding sites are shown in yellow (site 1) and orange (site 2). (B) Superimposed structures of CHDC2094/1974 and VA387/1998. The experimentally determined structure of VA387/1998 (pink) and the minimized structure of CHDC2094/1974 (orange) were superimposed and shown in ribbon format. The amino acid differences between the two structures and surrounding the carbohydrate binding site are highlighted in surface format. The trisaccharide molecule is colored in cyan blue.

study were located in the S domain of the capsid, while only one position (aa 395), was found in the P2 region exposed on the surface of the virus. Amino acid residue 395 was identified also as under positive selection by Lindesmith et al. (32) in an independent analysis. Considering the recognized role of antibody pressure in the evolution of many RNA viruses (2, 18), it was striking that so few positively selected sites were observed on the predicted GII.4 capsid surface over the 34-year period. Little is known about norovirus immunity, although antibodies do not correlate with resistance to illness in adult volunteers challenged with Norwalk virus, and immunity is at best short term (34). The role of immune-driven evolution in the emergence and persistence of the GII.4 noroviruses may be complex, and factors beyond (but not ruling out) antibody pressure may be important.

The most striking difference between the CHDC GII.4 sequences from the 1970s and the Camberwell cluster from the 1980s was the presence of a conserved 4-aa RVGI sequence in the archival strains that was replaced with the sequence HIVG. The RVGI sequence resulted in a detectable change in the predicted surface of the capsid, creating a unique protrusion. However, the presence of this predicted protrusion did not affect the HBGA binding pattern of the archival noroviruses compared to those of the Camberwell cluster, suggesting that other selective pressures may be involved in the evolution of GII.4 noroviruses. Additional studies will be needed to establish whether the GII.4 cluster replacement in the 1980s was associated with a change in this exposed 4-aa sequence.

The evolution of noroviruses has focused primarily on phylogenetic inference of the coding sequence of the VP1 protein

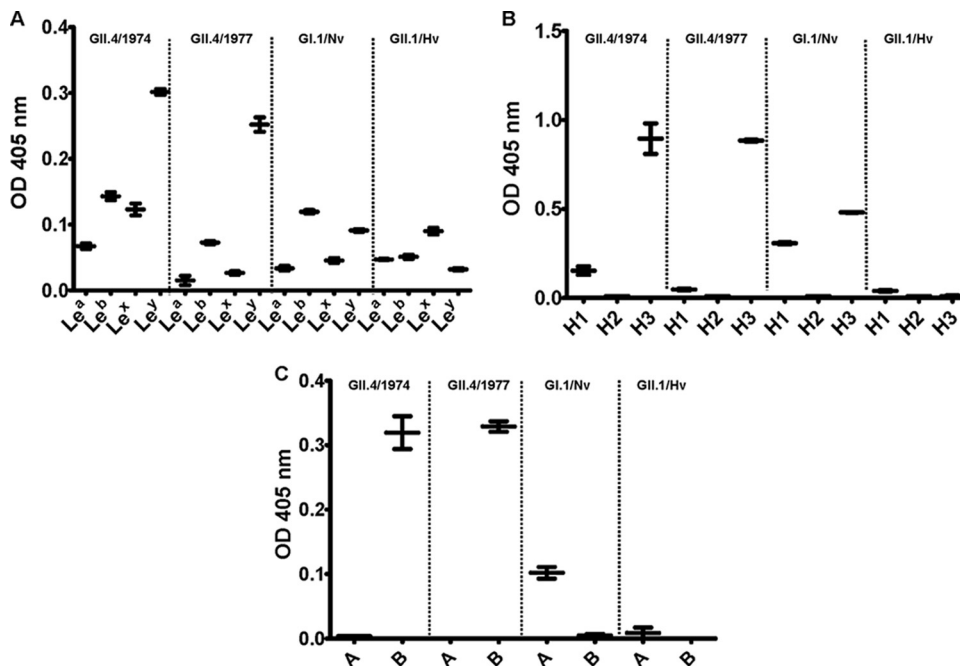


FIG. 7. Binding of 1974 and 1977 norovirus recombinant VLPs with synthetic HBGAs. The binding of rVLPs from archival noroviruses Hu/NoV/GII.4/CHDC5191/1974/US (GII.4/1974) and Hu/NoV/GII.4/CHDC4871/1977/US (GII.4/1977) to synthetic HBGA oligosaccharides was determined by optical density at 405 nm (OD 405 nm) at room temperature. Hu/NoV/GI.1/Norwalk virus/1968/US (GI.1/NV) and Hu/NoV/GII.1/Hawaii/1971/USA (GII.1/HV) rVLPs were utilized as controls. (A) rVLP binding assays to Lewis antigens a, b, x, and y. (B) rVLP binding assay to H1, H3, and H3 oligosaccharides. (C) rVLP binding assay to blood group A and B synthetic antigens.

(9, 26, 45). The VP2 protein, a minor component of the capsid, has been less studied (10, 28). The VP2 protein is essential for generating infectious particles during feline calicivirus replication, and its presence confers greater stability to recombinant norovirus VLPs (3, 46). The VP2 protein of the CHDC noroviruses presented higher amino acid variation than VP1 since the 1970s, with amino acid substitutions distributed across the entire VP2 sequence. No conclusions could be made on the locations of these mutations in the viral particle because a norovirus VP2 structure is not yet available. One possibility is that the high degree of diversity in this region might reflect a role for VP2 in population adaptation by improving viral fitness as the mutations become fixed in the VP1 sequence. Alternatively, VP2 may itself have a role in replication subject to selective pressure.

Several epidemiological studies have reported the use of genetic markers to track the emergence of new GII.4 norovirus variants that were responsible for an increase in the number of outbreaks diagnosed (1, 33), and it was of interest to examine whether these genetic markers existed in the archival strains. Lopman et al. detected an unusual set of mutations that they noted as present only in GII.4 polymerase sequences analyzed worldwide after 2002 (33). The “new variant” changed from AACTTG to AATCTG during that year, and it was proposed that the AATCTG sequence might be used as a genetic marker for the shift in prevalence, although the amino acid sequence (NL) remained the same. The analysis of our archival specimens showed that the CHDC4108/1987 genome contained the AATCTG set of mutations, suggesting that this sequence was present in viruses circulating several years prior to 2002. The

three samples sequenced from the 1970s contained GATCTC at this site, which results in one nonsynonymous amino acid substitution (DL). Allen et al. in an analysis of the amino acid variation in the P2 domain of capsid GII.4 sequences identified two sites (aa 296 to 298 and aa 393 to 395) as “hot spots” that impacted the biochemical properties of P2 and changed between epidemic waves (1). Both sites remained unchanged from 1974 until 1997, when the Grimsby cluster became predominant, indicating that these hot spots were not involved in earlier cluster replacements. These two examples show that the inference of new genetic markers based on limited sequencing data should be interpreted with caution. On the other hand, the archival norovirus analysis confirmed that some predictions were consistent over time. A study by Phan et al. described a classification system for noroviruses circulating in children in Japan (39), where four amino acids at positions 101 to 104 of the S domain and two amino acids at positions 522 to 523 of the C terminus represented an identification code specific for each genotype of virus. The GII.4 CHDC samples shared the “NGYA-NQ” signature of this genotype, showing that this genetic marker remained consistent for over 3 decades.

In summary, this study provides insight into the complex evolutionary patterns of GII.4 noroviruses since 1974. We report the first evolution rate of the GII.4 norovirus capsid, including the oldest available strains that were circulating at the time of the discovery of noroviruses. This study demonstrates that GII.4 noroviruses emerged prior to the 1980s and suggests that the ability of GII.4 noroviruses to predominate over time may involve unrecognized virus and host interactions beyond antibody selection and HBGA binding properties. Ef-

forts are in progress to identify norovirus strains in even earlier archival collections (1960s) and to compare the evolutionary characteristics of the other genotypes identified in the CHDC study with those of the GII.4 noroviruses. The success of a potential norovirus vaccine or antiviral drug may depend on our understanding of the evolutionary dynamics that allows the GII.4 genotype to persist over time.

#### ACKNOWLEDGMENTS

We thank Etsuko Utagawa, Norma Santos, and Taka Hoshino (Laboratory of Infectious Diseases, NIAID) for their contributions to this study. We also thank Michael Dolan, Ana P. Goncalvez, and German Añez-Gutierrez for their assistance with structure modeling. We also thank Lashanda Long-Croal for assistance in managing the RNA samples and the database and the Children's Hospital National Medical Center Research Group lead by H. W. Kim, which originally obtained the specimens and the clinical data (5).

This research was supported by the Intramural Research Program of the NIH, NIAID.

#### REFERENCES

- Allen, D. J., J. J. Gray, C. I. Gallimore, J. Xerry, and M. Iturriza-Gomara. 2008. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. *PLoS One* 3:e1485.
- Banke, S., M. R. Lillemark, J. Gerstoft, N. Obel, and L. B. Jørgensen. 2009. Positive selection pressure introduces secondary mutations at the Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J. Virol.* 83:8916–8924.
- Bertolotti-Ciarlet, A., S. E. Crawford, A. M. Hutson, and M. K. Estes. 2003. 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.* 77:11603–11615.
- Biebricher, C. K., and M. Eigen. 2006. What is a quasispecies? *Curr. Top. Microbiol. Immunol.* 299:1–31.
- Brandt, C. D., H. W. Kim, W. J. Rodriguez, J. O. Arrobio, B. C. Jeffries, E. P. Stallings, C. Lewis, A. J. Miles, R. M. Chanock, A. Z. Kapikian, and R. H. Parrott. 1983. Pediatric viral gastroenteritis during eight years of study. *J. Clin. Microbiol.* 18:71–78.
- Bucardo, F., J. Nordgren, B. Carlsson, M. Paniagua, P. E. Lindgren, F. Espinoza, and L. Svensson. 2008. Pediatric norovirus diarrhea in Nicaragua. *J. Clin. Microbiol.* 46:2573–2580.
- Bull, R. A., M. M. Tanaka, and P. A. White. 2007. Norovirus recombination. *J. Gen. Virol.* 88:3347–3359.
- Cao, S., Z. Lou, M. Tan, Y. Chen, Y. Liu, Z. Zhang, X. C. Zhang, X. Jiang, X. Li, and Z. Rao. 2007. Structural basis for the recognition of blood group trisaccharides by norovirus. *J. Virol.* 81:5949–5957.
- Carlsson, B., A. M. Lindberg, J. Rodriguez-Diaz, K. O. Hedlund, B. Persson, and L. Svensson. 2009. Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection. *J. Gen. Virol.* 90:432–441.
- Cauchi, M. R., J. C. Doultree, J. A. Marshall, and P. J. Wright. 1996. Molecular characterization of Camberwell virus and sequence variation in ORF3 of small round-structured (Norwalk-like) viruses. *J. Med. Virol.* 49:70–76.
- Dingle, K. E. 2004. Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. *J. Clin. Microbiol.* 42:3950–3957.
- Donaldson, E. F., L. C. Lindesmith, A. D. Lobue, and R. S. Baric. 2008. Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. *Immunol. Rev.* 225:190–211.
- Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307–1320.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–1192.
- Duffy, S., L. A. Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9:267–276.
- Duvvuri, V. R., B. Duvvuri, W. R. Cuff, G. E. Wu, and J. Wu. 2009. Role of positive selection pressure on the evolution of H5N1 hemagglutinin. *Genomics Proteomics Bioinformatics* 7:47–56.
- Escarmis, C., E. Lazaro, and S. C. Manrubia. 2006. Population bottlenecks in quasispecies dynamics. *Curr. Top. Microbiol. Immunol.* 299:141–170.
- Fernandez-Vega, V., S. V. Sosnovtsev, G. Belliot, A. D. King, T. Mitra, A. Gorbalenya, and K. Y. Green. 2004. Norwalk virus N-terminal nonstructural protein is associated with disassembly of the Golgi complex in transfected cells. *J. Virol.* 78:4827–4837.
- Green, K. Y., G. Belliot, J. L. Taylor, J. Valdesuso, J. F. Lew, A. Z. Kapikian, and F. Y. Lin. 2002. A predominant role for Norwalk-like viruses as agents of epidemic gastroenteritis in Maryland nursing homes for the elderly. *J. Infect. Dis.* 185:133–146.
- Green, K. Y., A. Z. Kapikian, J. Valdesuso, S. Sosnovtsev, J. J. Treanor, and J. F. Lew. 1997. Expression and self-assembly of recombinant capsid protein from the antigenically distinct Hawaii human calicivirus. *J. Clin. Microbiol.* 35:1909–1914.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
- Hoehne, M., and E. Schreiber. 2006. Detection of norovirus genogroup I and II by multiplex real-time RT-PCR using a 3'-minor groove binder-DNA probe. *BMC Infect. Dis.* 6:69.
- Huang, P., T. Farkas, S. Marionneau, W. Zhong, N. Ruvoen-Clouet, A. L. Morrow, M. Altaye, L. K. Pickering, D. S. Newburg, J. LePendu, and X. Jiang. 2003. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. *J. Infect. Dis.* 188:19–31.
- Iritani, N., A. Kaida, H. Kubo, N. Abe, T. Murakami, H. Vennema, M. Koopmans, N. Takeda, H. Ogura, and Y. Seto. 2008. Epidemic of genotype GII.2 noroviruses during spring 2004 in Osaka City, Japan. *J. Clin. Microbiol.* 46:2406–2409.
- Iturriza-Gomara, M., J. Xerry, C. I. Gallimore, C. Dockery, and J. Gray. 2008. Evaluation of the Loopamp (loop-mediated isothermal amplification) kit for detecting norovirus RNA in faecal samples. *J. Clin. Virol.* 42:389–393.
- Kamel, A. H., M. A. Ali, H. G. El-Nady, A. de Rougemont, P. Pothier, and G. Belliot. 2009. Predominance and circulation of enteric viruses in the region of Greater Cairo, Egypt. *J. Clin. Microbiol.* 47:1037–1045.
- Kapikian, A. Z., H. W. Kim, R. G. Wyatt, W. L. Cline, J. O. Arrobio, C. D. Brandt, W. J. Rodriguez, D. A. Sack, R. M. Chanock, and R. H. Parrott. 1976. Human reovirus-like agent as the major pathogen associated with "winter" gastroenteritis in hospitalized infants and young children. *N. Engl. J. Med.* 294:965–972.
- Kapikian, A. Z., R. G. Wyatt, R. Dolin, T. S. Thornhill, A. R. Kalica, and R. M. Chanock. 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J. Virol.* 10:1075–1081.
- Kroneman, A., L. Verhoef, J. Harris, H. Vennema, E. Duizer, Y. van Duynhoven, J. Gray, M. Iturriza, B. Böttiger, G. Falkenhorst, C. Johnsen, C.-H. von Bonsdorff, L. Maunula, M. Kuusi, P. Pothier, A. Galloway, E. Schreiber, M. Höhne, J. Koch, G. Szücs, G. Reuter, K. Krisztalovics, M. Lynch, P. McKeown, B. Foley, S. Coughlan, F. M. Ruggeri, I. Di Bartolo, K. Vainio, E. Isakbaeva, M. Poljsak-Prijatelj, A. H. Grom, J. Z. Mijovski, A. Bosch, J. Buesa, A. S. Fauquier, G. Hernández-Pezzi, K.-O. Hedlund, and M. Koopmans. 2008. Analysis of integrated virological and epidemiological reports of norovirus outbreaks collected within the Foodborne Viruses in Europe Network from 1 July 2001 to 30 June 2006. *J. Clin. Microbiol.* 46:2959–2965.
- Lindesmith, L. C., E. F. Donaldson, A. D. Lobue, J. L. Cannon, D. P. Zheng, J. Vinje, and R. S. Baric. 2008. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5:e31.
- Lopman, B., H. Vennema, E. Kohli, P. Pothier, A. Sanchez, A. Negro, J. Buesa, E. Schreiber, M. Reacher, D. Brown, J. Gray, M. Iturriza, C. Gallimore, B. Böttiger, K. O. Hedlund, M. Torven, C. H. von Bonsdorff, L. Maunula, M. Poljsak-Prijatelj, J. Zimsek, G. Reuter, G. Szucs, B. Melegh, L. Svensson, Y. van Duynhoven, and M. Koopmans. 2004. Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* 363:682–688.
- Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21:260–262.
- McGinnis, S., and T. L. Madden. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32:W20–W25.
- Noel, J. S., R. L. Fankhauser, T. Ando, S. S. Monroe, and R. I. Glass. 1999. Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *J. Infect. Dis.* 179:1334–1344.
- Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357–358.
- Patel, M. M., M. A. Widdowson, R. I. Glass, K. Akazawa, J. Vinje, and U. D. Parashar. 2008. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg. Infect. Dis.* 14:1224–1231.
- Phan, T. G., K. Kaneshi, Y. Ueda, S. Nakaya, S. Nishimura, A. Yamamoto, K. Sugita, S. Takanaishi, S. Okitsu, and H. Ushijima. 2007. Genetic heterogeneity, evolution, and recombination in noroviruses. *J. Med. Virol.* 79:1388–1400.
- Pond, S. L., and S. D. Frost. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

41. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
42. **Ramirez, S., G. M. Giammanco, S. De Grazia, C. Colomba, V. Martella, and S. Arista.** 2009. Emerging GII.4 norovirus variants affect children with diarrhea in Palermo, Italy in 2006. *J. Med. Virol.* **81**:139–145.
43. **Reuter, G., P. Pankovics, and G. Szucs.** 2008. Genetic drift of norovirus genotype GII-4 in seven consecutive epidemic seasons in Hungary. *J. Clin. Virol.* **42**:135–140.
44. **Shih, A. C., T. C. Hsiao, M. S. Ho, and W. H. Li.** 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Natl. Acad. Sci. USA* **104**:6283–6288.
45. **Siebenga, J. J., H. Vennema, B. Renckens, E. de Bruin, B. van der Veer, R. J. Siezen, and M. Koopmans.** 2007. Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* **81**:9932–9941.
46. **Sosnovtsev, S. V., G. Belliot, K. O. Chang, O. Onwudiwe, and K. Y. Green.** 2005. Feline calicivirus VP2 is essential for the production of infectious virions. *J. Virol.* **79**:4012–4024.
47. **Tan, M., P. Huang, J. Meller, W. Zhong, T. Farkas, and X. Jiang.** 2003. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. *J. Virol.* **77**:12562–12571.
48. **Tan, M., M. Xia, S. Cao, P. Huang, T. Farkas, J. Meller, R. S. Hegde, X. Li, Z. Rao, and X. Jiang.** 2008. Elucidation of strain-specific interaction of a GII-4 norovirus with HBGA receptors by site-directed mutagenesis study. *Virology* **379**:324–334.
49. **Zheng, D. P., T. Ando, R. L. Fankhauser, R. S. Beard, R. I. Glass, and S. S. Monroe.** 2006. Norovirus classification and proposed strain nomenclature. *Virology* **346**:312–323.