JOURNAL OF VIROLOGY, Aug. 2008, p. 7336–7345 Vol. 82, No. 15
0022-538X/08/$08.00+0 doi:10.1128/JVI.02371-07

Genetic Analysis of the Capsid Gene of Genotype GII.2 Noroviruses▽

Nobuhiro Iritani,[1,2]* Harry Vennema,[1] J. Joukje Siebenga,[1] Roland J. Siezen,[3] Bernadet Renckens,[3] Yoshiyuki Seto,[4] Atsushi Kaida,[2] and Marion Koopmans[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[1]; Department of Microbiology, Osaka City Institute of Public Health and Environmental Sciences, 8-34 Tojo-cho, Tennoji-ku, Osaka 543-0026, Japan[2]; Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, P.O. Box 9101, 6500HB Nijmegen, The Netherlands[3]; and Laboratory of Microbiology, Osaka Prefecture University, 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan[4]

Received 2 November 2007/Accepted 5 May 2008

Noroviruses (NoVs) are considered to be a major cause of acute nonbacterial gastroenteritis in humans. The NoV genus is genetically diverse, and genotype GII.4 has been most commonly identified worldwide in recent years. In this study we analyzed the complete capsid gene of NoV strains belonging to the less prevalent genotype GII.2. We compared a total of 36 complete capsid sequences of GII.2 sequences obtained from the GenBank ($n = 5$) and from outbreaks or sporadic cases that occurred in The Netherlands ($n = 10$) and in Osaka City, Japan ($n = 21$), between 1976 and 2005. Alignment of all capsid sequences did not show fixation of amino acid substitutions over time as an indication for genetic drift. In contrast, when strains previously recognized as recombinants were excluded from the alignment, genetic drift was observed. Substitutions were found at five informative sites (two in the P1 subdomain and three in the P2 subdomain), segregating strains into five genetic groups (1994 to 1997, 1999 to 2000, 2001 to 2003, 2004, and 2005). Only one amino acid position changed consistently between each group (position 345). Homology modeling of the GII.2 capsid protein showed that the five amino acids were located on the surface of the capsid and close to each other at the interface of two monomers. The data suggest that these changes were induced by selective pressure, driving virus evolution. Remarkably, this was observed only for nonrecombinant genomes, suggesting differences in behavior with recombinant strains.

Noroviruses (NoVs) are an important cause of acute nonbacterial gastroenteritis in adults and children worldwide (13). NoVs are members of the family *Caliciviridae*, having a positive-sense single-stranded RNA genome. Their genome is organized into three open reading frames (ORFs). ORF1 encodes nonstructural proteins including the RNA-dependent RNA polymerase (RdRp), ORF2 encodes a major structural capsid protein including a shell (S) domain and a protruding (P) domain, and ORF3 encodes a minor structural protein (13, 18, 41). The S domain forms the inner part (shell) of the viral capsid, and the P domain forms the arch-like structures that protrude from the virion. The P domain is further divided into P1 and P2 subdomains that correspond to the sides and the top of the arch-like capsomeres, respectively (13, 31).

Based on the genetic analysis of the RdRp and capsid regions, human NoVs can be divided into three genogroups (Gs), GI, GII, and GIV (2, 14, 39), which further segregate into distinct lineages called genotypes (2, 20, 36, 37). Recently, Kageyama et al. (20) proposed that at least 31 genotypes could be distinguished within GI and GII. The GII.4 genotype, which is represented by the Lordsdale/93/UK strain, has been the most commonly identified genotype worldwide in recent years. Genetic characterization of strains belonging to this genotype have shown a sequence of variants that have arisen over time, suggesting that rapid genetic evolution of GII.4 NoVs may in part explain their successful spread and impact on people of all ages (5, 9, 27, 29, 30, 33, 40).

Viruses belonging to other NoV genotypes are found less consistently, causing sporadic outbreaks or temporary epidemics in a limited geographic region or time period (5, 17, 23, 26). As a result, far less is known about the population structure of these variants (28). The genetic analysis of other genotype NoVs excluding GII.4 will improve our understanding of genetic evolution and its relevance for the epidemiology of NoVs.

During the spring of 2004, an epidemic of GII.2 NoV (which is represented by the Melksham/94/UK strain [Melksham]) occurred in Osaka City, Japan. Our previous study of this regional epidemic described the molecular epidemiology of these GII.2 strains (17). Here, we describe the genetic characterization of GII.2 strains from those outbreaks in comparison with viruses detected over a 12-year period in the GenBank, The Netherlands, and Japan.

MATERIALS AND METHODS

GII.2 strains. A total of 36 NoV strains that had been characterized as GII.2 genotype were used for this study (Table 1). The capsid sequence data for five GII.2 strains were obtained from the GenBank. Of these, the Melksham strain and the Chesterfield/434/1997/US strain (11, 29) have been characterized as belonging to the GII.2 genotype on the basis of RdRp as well as capsid regions. The Snow Mountain/76/US strain has been characterized as a recombinant NoV, with a distinct (non-GII.2) RdRp region and a GII.2 capsid region (4, 16). The other two strains from the GenBank (Ina/02/JP and Buds/02/US) were characterized as GII.2 genotype on the basis of the capsid region, but their sequences of the RdRp region were unknown. The capsid sequence data of Buds/02/US strain lacked the first 6 nucleotides (nt) from the 5′ end of the capsid gene. Twenty-one GII.2 strains were obtained from outbreaks or sporadic cases detected in Osaka City, Japan, between April 1996 and March 2005. These were 21

* Corresponding author. Mailing address: Department of Microbiology, Osaka City Institute of Public Health and Environmental Sciences, 8-34 Tojo-cho, Tennoji-ku, Osaka 543-0026, Japan. Phone: 81 6 6771 3147. Fax: 81 6 6772 0676. E-mail: n-iritani@city.osaka.lg.jp.

▽ Published ahead of print on 14 May 2008.

TABLE 1. GII.2 NoV strains used in this study

Source	Strain (abbreviation)[a]	RdRp sequence type[b]	Accession no.	Note[c]
GenBank	Snow Mountain/76/US (SM)	GII-NA	AY134748	
	Melksham/94/UK (Melksham)	GII.2	X81879	
	Chesterfield/434/97/US (CF434)	GII.2	AY054300	
	Ina/02/JP (Ina)	Unknown	AB195225	
	Buds/02/US (Buds)	Unknown	AY660568	
Osaka City, Japan	OC97049/97/JP (97049)	GII-NA	AB279553	O
	OC01243/01/JP (01243)	GII-NA	AB279554	O
	OC02012/02/JP (02012)	GII.2	AB279555	O
	OC02022/02/JP (02022)	GII-NA	AB279556	O
	OCS020289/02/JP (S020289)	GII.b	AB279570	S
	OC04038/04/JP (04038)[d]	GII.2	AB279557	O*
	OC04042/04/JP (04042)[d]	GII.2	AB279558	O*
	OC04043/04/JP (04043)[d]	GII.2	AB279559	O*
	OCS030697/04/JP (S030697)[d]	GII.2	AB279571	S*
	OC04056-1/04/JP (04056-1)[d]	GII.2	AB279560	O*
	OC04056-2/04/JP (04056-2)[d]	GII.2	AB279561	O*
	OC04059/04/JP (04059)	GII.2	AB279562	O*
	OCS040035/04/JP (S040035)	GII.2	AB279572	S*
	OC04067/04/JP (04067)[d]	GII.2	AB279563	O*
	OC04071/04/JP (04071)	GII.2	AB279564	O*
	OC04073/04/JP (04073)	GII.2	AB279565	O*
	OC04075/04/JP (04075)	GII.2	AB279566	O*
	OC04076/04/JP (04076)[d]	GII.2	AB279567	O*
	OCS040100/04/JP (S040100)	GII.2	AB279573	S*
	OC04169/04/JP (04169)	GII-NA	AB279568	O
	OC05010/05/JP (05010)	GII.b	AB279569	O
The Netherlands	Sensor99-191/99/NL (S99-191)	GII.2	AB281081	S
	OB0037-246/00/NL (OB0037)	GII.2	AB281082	O
	OB0048-318/00/NL (OB0048)	GII.2	AB281083	S
	OB0115-195/01/NL (OB0115)	GII.2	AB281084	O
	EP0125-006/01/NL (EP0125)	GII.2	AB281085	O
	EP0207-001/02/NL (EP0207)	GII.2	AB281086	O
	EP0239-001/02/NL (EP0239)	GII.2	AB281087	O
	OB0371-459/03/NL (OB0371)	GII.2	AB281088	O
	OB0528-158/05NL (OB0528)	GII.2	AB281089	O
	OB0587-470/05/NL (OB0587)	GII.2	AB281090	O

[a] NoV strains are arranged in chronology of detection from top (oldest) to bottom (most recent) for each source.
[b] NA, not assigned.
[c] O, outbreak; S, sporadic case; *, spring epidemic in 2004.
[d] These strains have identical amino acid sequences in the complete capsid gene and only one strain (OC04038/04/JP) has been used for long-term genetic analysis.

of the 23 GII.2 strains identified during a 9-year study period out of a total of 238 outbreaks and 200 sporadic cases of NoV infection. From the genetic analysis across the junction between the RdRp and the capsid regions, 6 of these 21 GII.2 strains have been characterized as recombinants, which have non-GII.2 RdRp regions and GII.2 capsid regions (Fig. 1) (17).

The strains from The Netherlands were collected from a 12-year study period. Between 1994 and 2005, GII.2 NoVs were detected in 13 (1.7%) out of 745 NoV-associated outbreaks and three sporadic cases in The Netherlands. Initially, these GII.2 NoVs were characterized by the comparison of sequences in the RdRp region (Fig. 1). The detection method and criteria for genotyping at the RdRp region have been previously described (8, 36). The complete capsid genes of 10 strains from eight outbreaks and two sporadic cases were amplified by reverse transcription-PCR (RT-PCR) and were used for this analysis.

Amplification and sequencing for the complete capsid gene of GII.2 strains. Viral RNA was extracted from stool suspensions by using a QIAamp viral RNA Mini kit (Qiagen, Hilden, Germany). RT-PCR was carried out with the reaction mixtures and enzymes as previously described (8). RT was performed at 42°C for 2 to 3 h with reverse primer, N235Rex (17), and enzyme was inactivated at 95°C for 5 min. PCR was performed using several pairs of PCR primers (Table 2) with a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA) under the following conditions: denaturation at 95°C for 1 min; 40 cycles of 95°C for 15 s, 50°C for 30 s, and 72°C for 1 min; and a final cycle of incubation at 72°C for 5 min. When a PCR failed to produce strong products, we performed nested PCR. The amplified fragments were sequenced directly with a Big Dye terminator cycle sequencing kit and ABI 3700 sequencer (Applied Biosystems, Foster City, CA). The nucleotide sequences were determined in both orientations using the PCR primers. DNA sequences were edited using SeqManII (DNAStar Inc., Konstanz, Germany).

Sequence analysis. Nucleotide or amino acid sequence alignments were performed with BioEdit (version 7.052) (15), Clustal X (version 1.81) (35), or MUSCLE (version 3.51) (10). The extraction of the informative sites from nucleotide or amino acid sequence alignments was performed with ProSeq (version 2.91) (12). The rate of change for different domains was compared using chi-square statistics. In this analysis, a site was designated as an informative site when at least two strains had an identical amino acid in the alignment that differed from the other sequences. A phylogenetic tree with 1,000 bootstrap replications was constructed by the neighbor-joining method, and the genetic distances were calculated according to the Kimura two-parameter method (21). We performed additional phylogenetic analysis by the Bayesian method using MrBayes (version 3.1.2) (32). Location of specific domains of the GII.2 NoV capsid gene was done according to Chen et al. (7). For computational predictions of the structure of the GII.2 NoV capsid protein, we used the X-ray crystal structures of the capsid protein of Norwalk/68/US (GI.1 genotype, Protein Data Bank identifier [PDB ID] 1IHM, consisting of a complete trimer) (31) and VA387/98/US (GII.4 genotype, PDB ID 2OBR, consisting of only a monomeric P domain) (6) as templates to build homology models. The sequence alignments for the structure and the three-dimensional (3D) models for GII.2 NoV capsid proteins were made by using the WHAT IF program (38) and the 3D-Jigsaw (3)

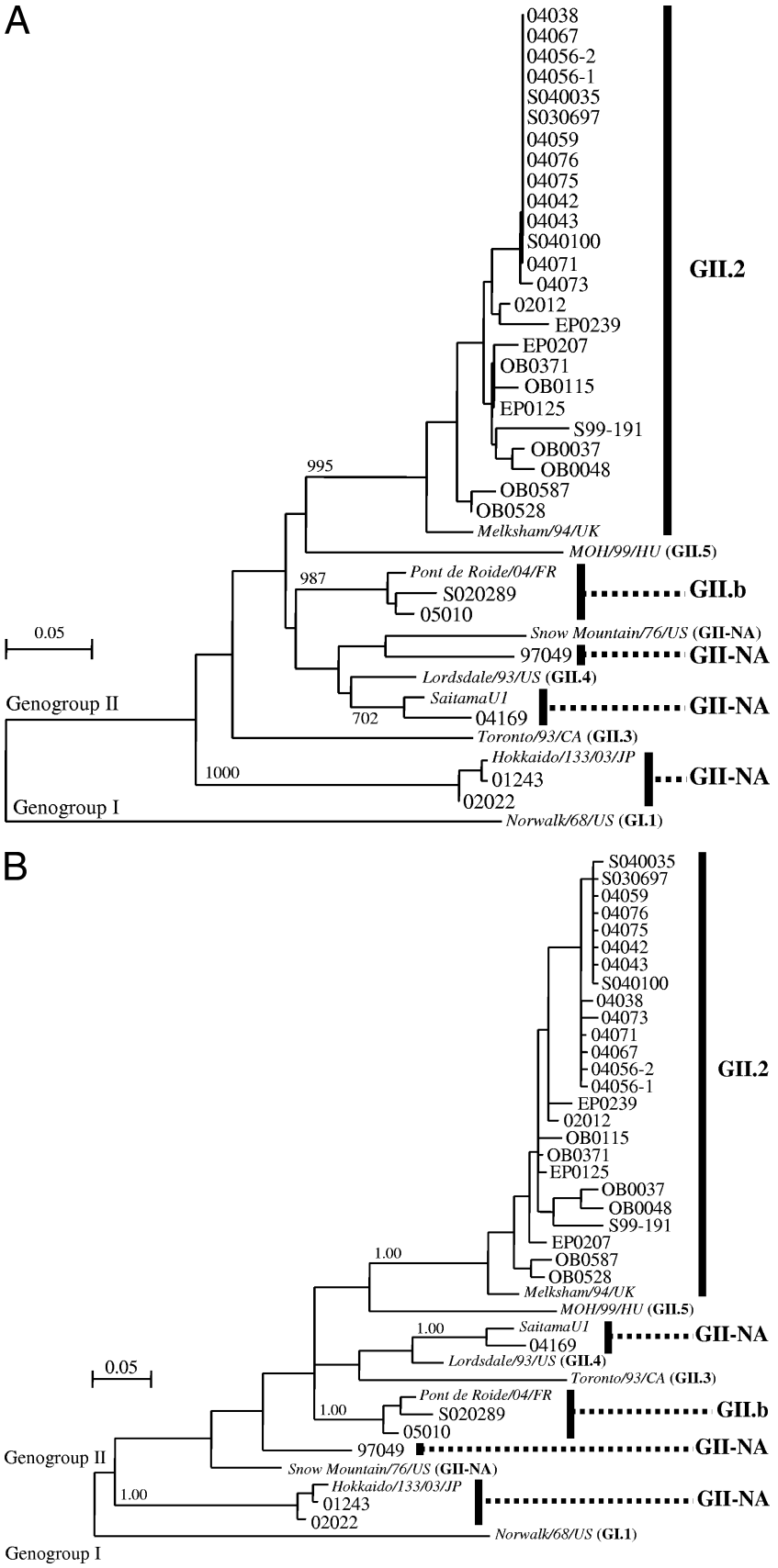
A
04038
04067
04056-2
04056-1
S040035
S030697
04059
04076
04075
04042
04043
S040100
04071
04073
02012
EP0239
EP0207
OB0371
OB0115
EP0125
S99-191
OB0037
OB0048
OB0587
OB0528
Melksham/94/UK
MOH/99/HU (GII.5)
Pont de Roide/04/FR
S020289
05010
Snow Mountain/76/US (GII-NA)
97049
Lordsdale/93/US (GII.4)
SaitamaU1
04169
Toronto/93/CA (GII.3)
Hokkaido/133/03/JP
01243
02022
Norwalk/68/US (GI.1)
995
987
702
1000
0.05
Genogroup II
Genogroup I
GII.2
GII.b
GII-NA
GII-NA
GII-NA
B
S040035
S030697
04059
04076
04075
04042
04043
S040100
04038
04073
04071
04067
04056-2
04056-1
EP0239
02012
OB0115
OB0371
EP0125
OB0037
OB0048
S99-191
EP0207
OB0587
OB0528
Melksham/94/UK
MOH/99/HU (GII.5)
SaitamaU1
04169
Lordsdale/93/US (GII.4)
Toronto/93/CA (GII.3)
Pont de Roide/04/FR
S020289
05010
97049
Snow Mountain/76/US (GII-NA)
Hokkaido/133/03/JP
01243
02022
Norwalk/68/US (GI.1)
1.00
1.00
1.00
1.00
0.05
Genogroup II
Genogroup I
GII.2
GII-NA
GII.b
GII-NA
GII-NA

TABLE 2. Primers used to amplify the capsid gene of GII.2 NoV strains

Primer	Sequence (5′ to 3′)[a]	Polarity	Location (nt)	Reference or source
COG2F	CA**R**GA**RBC**N**A**TGTT**Y**AG**R**TGGATGAG	+	5003[b]	19
G2SKF	C**N**TGGGAGGGCGATCGCAA	+	5058[b]	22
G2SKR	CC**R**CC**N**GCAT**RH**CC**R**TT**R**TACAT	−	5401[b]	22
N235Rex	GC**W**A**N**R**A**AAGCTCC**W**GCCAT	−	6273[b]	17
MKcap508-524F	CAGAAAGATGATCCCAA	+	508[c]	This study
MKcap524-508R	TTGGGATCATCTTTCTG	−	524[c]	This study
MKcap662-649R	TCCACTGTTGGTGG	−	662[c]	This study
MKcap1109-1125F	TGGGTCAGATTCAAATT	+	1109[c]	This study
MKcap1125-1109R	AATTTGAATCTGACCCA	−	1125[c]	This study
MKcap1304-1290R	AAGAGCAGGCGCTCC	−	1304[c]	This study
MK35R	CAAAAGCTCCAGCCAT	−	1644[c]	This study

[a] Abbreviations for residues in boldface are as follows: B, not A; H, not G. In addition, N is A, C, G, or T; R is A or G; W is A or T; Y is C or T.
[b] Location of the 5′ end of the primer corresponding to the nucleotide position of Lordsdale/93/UK (X86557).
[c] Location of the 5′ end of the primer corresponding to the nucleotide position of ORF2 and ORF3 of Melksham/94/UK (X81879).

and EsyPred3D (25) servers. WHAT IF could not model residues 342, 344, and 345 based on the Norwalk/68/US capsid protein template (PDB ID 1IHM); as they are present in an inserted loop of the P2 subdomain relative to the template structure, but they are predicted to be in the same position as in the 3D model based on the GII.4 genotype capsid protein template (PDB ID 2OBR). A dimer of the GII.2 NoV capsid protein was modeled by superimposing two predicted monomers onto the trimeric template of the Norwalk/68/US capsid protein. The 3D models were visualized by the YASARA view program (version 6.813, http://www.yasara.org/) (24).

Nucleotide sequence accession numbers. The nucleotide sequences determined in this study have been deposited in the DDBJ/EMBL/GenBank with the accession numbers AB281081 to AB281090 (Table 1).

RESULTS

Capsid gene sequence analysis of GII.2 strains collected over 30 years. Sequence data from a total of 36 GII.2 strains showed that the capsid gene was 1,629 nt long and coded for a protein of 542 amino acids. There were no deletions or insertions in the capsid gene among these strains collected over a 30-year period. Sequence comparison showed ≥83.6% nucleotide and ≥93.5% amino acid identities among these GII.2 strains. The phylogenetic tree based on the 1,629 nucleotide sequences of the complete capsid gene confirmed that all strains were characterized as GII.2 genotype (Fig. 2).

Genetic analysis of GII.2 strains from a regional outbreak in Japan in the spring of 2004. Of the 21 GII.2 strains detected in Osaka City, Japan, 14 strains were detected in samples from cases in outbreaks or sporadic cases in the spring epidemic between March and May 2004 (Table 1). These strains were closely related to each other (≥99.2% nucleotide and ≥ 99.1% amino acid identities). To find out if these viruses changed genetically during circulation in the community for a short period of time (about 3 months), the complete nucleotide and amino acid sequences of the capsid gene from these 14 GII.2 strains were compared. A total of 45 nucleotide substitutions were observed (Table 3). The majority were third-base position changes (77.8%) and synonymous substitutions (82.2%). These nucleotide changes resulted in amino acid changes in eight codons, five of which were located in the P2 subdomain. In the alignment of the spring epidemic strains over the 3-month period, of the 45 nucleotide substitutions, only one nucleotide position was fixed (nt 594 in the S domain) at the end of this epidemic. This did not result in an amino acid change. All amino acid changes were sporadic. Eight of the 14 strains had an identical amino acid sequences (Table 1). Of these, strain 04038 was used for further genetic analysis.

Genetic analysis of all GII.2 strains collected between 1976 and 2005. The nucleotide and amino acid sequences of the complete capsid genes were compared for 29 GII.2 strains collected between 1975 and 2005 (30 years), excluding 7 GII.2 strains detected in Osaka City with identical amino acid sequences (10 strains from The Netherlands, 14 from Osaka City, and 5 from GenBank). A total of 488 nucleotide changes were observed (Table 4). Again, the majority of these were third-base changes (83.8%) and synonymous (85.9%). In total, 59 nucleotide changes resulted in amino acid changes (34 in the P2 subdomain) (Table 4). Twenty-five of these were informative changes (19 in the P2 subdomain), but none appeared to be fixed in the genome over time. Of the 25 informative sites, the amino acid position 345 was the most variable (Fig. 3). Nevertheless, statistical analysis showed a significantly higher rate of mutation in the P2 subdomain than in P1 and S, suggesting selective pressure ($P = 0.0018$, chi-square 5.63; and $P < 0.0001$, chi-square 12.9).

Genetic analysis of Melksham-like strains detected between 1994 and 2005. In order to understand the apparent discrepancy between selective changes in the P2 subdomain and the absence of fixation of these mutations, we repeated our anal-

FIG. 1. Phylogenetic analysis of partial RdRp gene sequences of GII.2 NoV strains by the neighbor-joining method (A) and the Bayesian method (B). The genotypes at the RdRp regions that are not assigned numbers are represented as GII-NA. The scale indicates the number of substitutions per site. Reference strains of NoV used in this analysis are given in italics. (A) The bootstrap values are indicated on each branch. (B) Tree topology was evaluated on the basis of 1,500,000 generations. The posterior probabilities are indicated on each branch (≥0.95 of the posterior probability means that the branch has high credibility). The GenBank accession numbers of the additional reference strains used in this analysis are as follows: Hokkaido/133/03/JP, AB212306; Lordsdale/93/UK, X86557; MOH/99/HU, AF397256; Norwalk/68/US, M87661; Pont de Roide 673/04/FR, AY682549; Saitama U1, AB039775; Toronto/93/CA, U02030. For strain abbreviations, see Table 1.

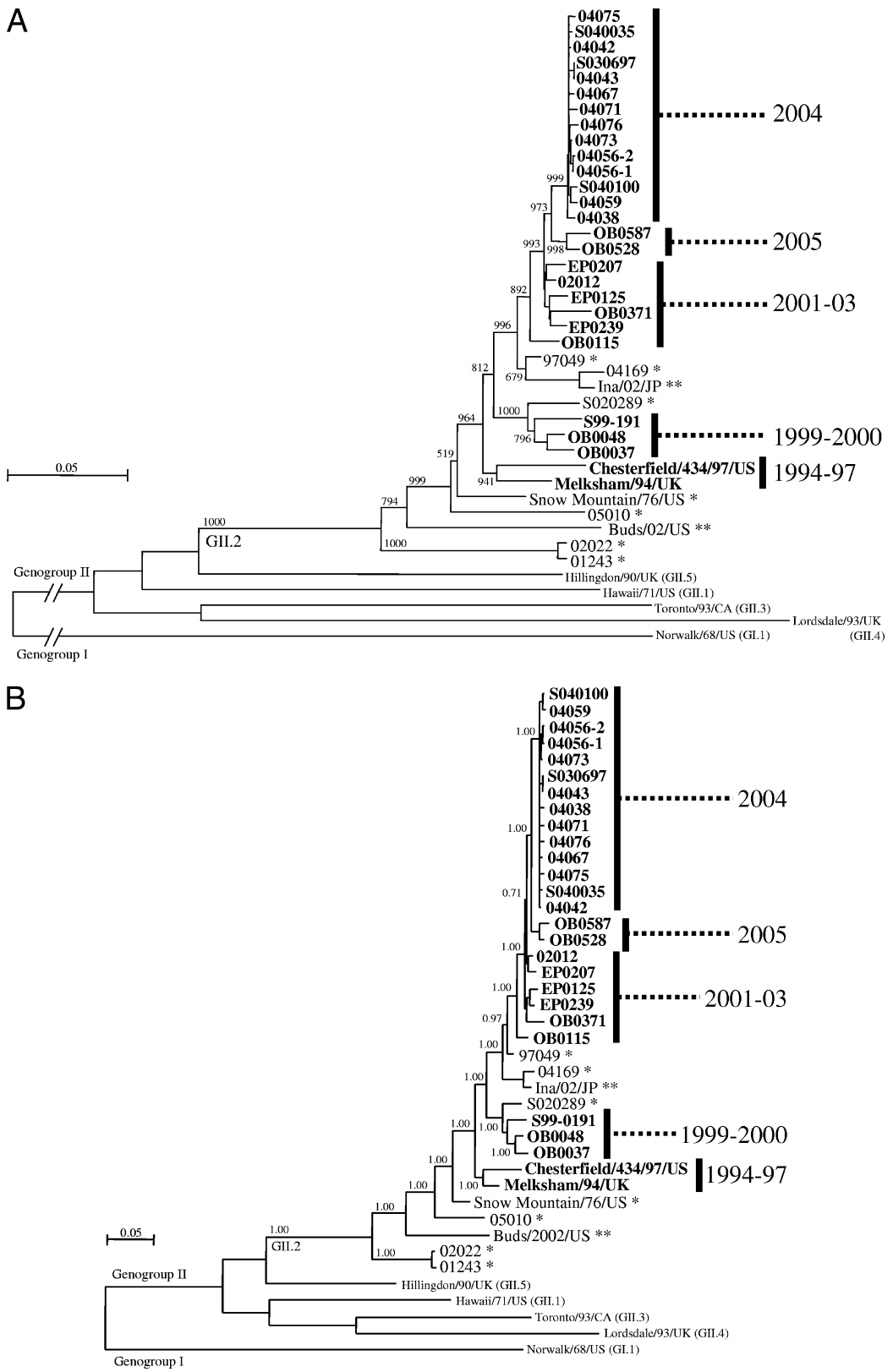

FIG. 2. Phylogenetic analysis of complete capsid gene sequences of GII.2 NoV strains by the neighbor-joining method (A) and the Bayesian method (B). Melksham-like strains, which have a matching (GII.2) RdRp sequence, are shown in boldface. Asterisks indicate the GII.2 strains that have different (*) or unknown (**) genotypes in the RdRp region. The scale indicates the number of substitutions per site. (A) The bootstrap values are indicated on each branch. (B) Tree topology was evaluated on the basis of 500,000 generations. The posterior probabilities are indicated on each branch. The GenBank accession numbers of the additional reference strains used in this analysis are as follows: Hawaii/71/US, U07611; Hillingdon/90/UK, AJ277607; Lordsdale/93/UK, X86557; Norwalk/68/US, M87661; Toronto/93/CA, U02030. For strain abbreviations, see Table 1.

TABLE 3. The numbers of nucleotide and amino acid substitutions in the complete capsid gene among 14 GII.2 strains detected in the spring epidemic of 2004 in Osaka City, Japan

Domain or subdomain[a]	Nucleotide substitutions						Amino acid substitutions	
	No. of substitutions	No. at the indicated position of the changed codon			Synonymous changes (no.)	Nonsynonymous changes (no.)	No. of substitutions	Informative changes (no.)
		1st	2nd	3rd				
N	1	0	0	1	1	0	0	0
S	18	3	2	13	16	2	2	0
P1	7	0	0	7	6	1	1	0
P1-1	4	0	0	4	4	0	0	0
P1-2	3	0	0	3	2	1	1	0
P2	19	3	2	14	14	5	5	0
Total	45	6	4	35	37	8	8	0

[a] Each domain of GII.2 NoV capsid gene was determined according to Chen et al. (7). The region and abbreviation of each domain are as follows: N, 5′ end of ORF2 and N terminal domain (residues 1 to 45); S, shell domain (residues 46 to 216); P1, P1 subdomain (P1-1, residues 217 to 274; P1-2, residues 421 to 542); P2, P2 subdomain (residues 275 to 420).

ysis after removing recombinant genomes from the alignment (Table 5) (nucleotide and amino acid sequences of the complete capsid gene from 20 Melksham-like strains: 10 strains from The Netherlands, 8 from Osaka City, and 2 from GenBank). Sequence comparison showed ≥91.5% nucleotide and ≥97.4% amino acid identities among these Melksham-like strains. A total of 301 nucleotide changes were observed (Table 5), the majority of which were third-base changes (87%) and synonymous (88.7%). These nucleotide changes resulted in amino acid changes in 32 codons, half of which were located in the P2 subdomain. Twelve of 32 amino acid positions were informative (9 in the P2 subdomain) (Fig. 3). In contrast with the previous finding, several mutations were fixed: of the 12 informative sites, 2 amino acid substitutions (amino acid positions 245 and 440) in the P1 subdomain and 3 amino acid substitutions (amino acid positions 342, 344, and 345) in P2 subdomain were cumulative (Table 6), segregating the strains into five genetic groups (1994 to 1997, 1999 to 2000, 2001 to 2003, 2004, and 2005) by the neighbor-joining method (Fig. 2A) and Bayesian method (Fig. 2B).The strains detected in the spring epidemic in Osaka City had a unique sequence, with S or P residues at amino acid position 364 (Fig. 3 and Table 6). The other six informative sites were not fixed.

The 3D structure of the P domain of a monomer of the NoV capsid protein was predicted by WHAT IF, 3D-Jigsaw, and EsyPred3D, based on the known 3D structure of the VA387/98/US GII.4 genotype capsid protein, which has 55% amino acid sequence identity in the P domain to the Melksham capsid protein. A comparison of the positions of the six fixed mutations to the predicted 3D structure indicated that all six residues were predicted to be located at the surface of the capsid protein, with three residues (342, 344, and 345) close to each

TABLE 4. The numbers of nucleotide and amino acid substitutions in the complete capsid gene among 29 GII.2 strains collected in the GenBank, The Netherlands, and Japan over a 30-year period

Domain or subdomain[a]	Nucleotide substitutions						Amino acid substitutions	
	No. of substitutions	No. at the indicated position of the changed codon			Synonymous changes (no.)	Nonsynonymous changes (no.)	No. of substitutions	Informative changes (no.)
		1st	2nd	3rd				
N	20	0	2	18	17	3	3	1
S	137	9	3	125	131	6	6	1
P1	162	17	7	138	143	19	16	4
P1-1	53	4	2	47	49	4	3	2
P1-2	109	13	5	91	94	15	13	2
P2	169	28	13	128	128	41	34	19
Total	488	54	25	409	419	69	59	25

[a] See note to Table 3.

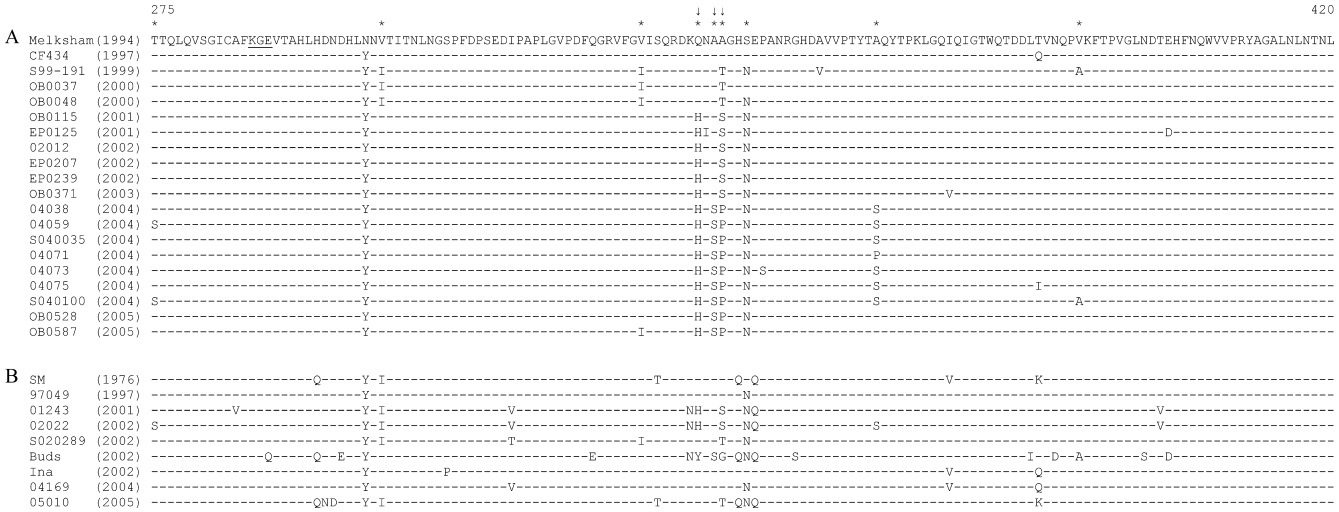

FIG. 3. Amino acid alignment of the P2 subdomains of GII.2 NoV strains showing mutations along the aligned sequences. The upper sequence alignment group (A) includes the Melksham-like strains and the lower group (B)includes the other GII.2 strains, which were recombinant genomes. In each group, sequences are arranged chronologically from top (oldest) to bottom (most recent). The detection years of the strains are indicated in parentheses. The asterisk indicates informative sites among Melksham-like strains. The arrow denotes cumulatively changing amino acid positions (342, 344, and 345) among Melksham-like strains. The numbers above the sequences indicate the sequence position relative to the position in the capsid protein of strain Melksham. Underlined sequences indicate the KGE motif that corresponds to the RGD-like motif of other NoVs and was determined with the amino acid sequence alignment of other NoVs according to Tan et al. (34). For strain abbreviations, see Table 1.

other in the P2 subdomain (Fig. 4). Furthermore, residues 342, 344, 345 of the P2 subdomain and residues 245 and 440 of the P1 subdomain were grouped closely together on the predicted 3D structure of a dimer (3D-modeling by WHAT IF) (Fig. 5).

DISCUSSION

In this study, we analyzed the complete capsid gene of GII.2 NoVs collected over a 30-year period. The collection included a relatively large sample from an epidemic that was observed only in Japan in the spring period of 2004. In this 3-month period, no evolutionary changes were observed, but compared with other GII.2 Melksham-like strains these variants had unique amino acid sequences (S or P) at position 364. One of the recombinant genomes, strain 02022, had the same amino acid residues, suggesting that the presence of this mutation was not causally related to the epidemic pattern. The sequence analysis confirmed that the spring epidemic was an outbreak.

Molecular characterization of the GII.2 capsids over the full study period showed an interesting difference between recombinant genomes and nonrecombinant (Melksham-like) genomes. The Melksham-like strains clearly evolved over time, with accumulation of mutations particularly in the P2 subdomain. In contrast, GII.2 capsids from recombinant genomes did not fit this pattern and had a seemingly erratic pattern of mutations. Melksham-like strains are occasionally observed in molecular surveillance data from The Netherlands and Osaka

TABLE 5. The numbers of nucleotide and amino acid substitutions in the complete capsid gene among 20 Melksham-like strains over a 12-year period

Domain or subdomain[a]	Nucleotide substitutions						Amino acid substitutions	
	No. of substitutions	No. at the indicated position of the changed codon			Synonymous changes (no.)	Nonsynonymous changes (no.)	No. of substitutions	Informative changes (no.)
		1st	2nd	3rd				
N	10	0	2	8	8	2	2	0
S	81	4	3	74	76	5	6	1
P1	98	8	3	87	89	9	8	2
P1-1	34	2	1	31	32	2	2	1
P1-2	64	6	2	56	57	7	6	1
P2	112	13	6	93	94	18	16	9
Total	301	25	14	262	267	34	32	12

[a] See note to Table 3.

TABLE 6. Amino acid substitutions at six positions in the P domain among Melksham-like strains

Strain description			Amino acid and codon at the indicated position of the capsid protein[a]											
			245		342		344		345		364		440	
Name[b]	Country[c]	Detection date (mo/day/yr)	Residue	Codon	Residue	Codon	Residue	Codon	Residue	Codon	Residue	Codon	Residue	Codon
Melksham	UK	1994	P	CCC	Q	CAG	A	GCC	A	GCT	A	GCA	L	CTC
CF434	US	1997	—	CCC	—	CAG	—	GCT	—	GCT	—	GCA	—	CTT
S99-191	NL	1/1999	S	TCT	—	CAA	—	GCC	T	ACT	—	GCA	I	ATC
OB0037	NL	9/2000	S	TCT	—	CAA	—	GCC	T	ACT	—	GCA	I	ATC
OB0048	NL	12/2000	S	TCT	—	CAA	—	GCC	T	ACT	—	GCA	I	ATC
OB0115	NL	2/2001	S	TCT	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
EP0125	NL	11/2001	S	TCC	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
02012	JP	1/2002	S	TCC	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
EP0207	NL	1/2002	S	TCC	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
EP0239	NL	2/2002	S	TCC	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
OB0371	NL	10/2003	S	TCC	H	CAC	—	GCC	S	TCT	—	GCA	I	ATC
04038	JP	3/11/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
04059	JP	4/18–30/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
S040035	JP	4/21/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
04071	JP	5/16–23/2004	S	TCC	H	CAC	S	TCC	P	CCT	P	CCA	I	ATC
04073	JP	5/22/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
04075	JP	5/25/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
S040100	JP	5/31/2004	S	TCC	H	CAC	S	TCC	P	CCT	S	TCA	I	ATC
OB0528	NL	1/2005	S	TCC	H	CAC	S	AGC	P	CCT	—	GCA	I	ATT
OB0587	NL	4/2005	S	TCC	H	CAC	S	AGC	P	CCT	—	GCA	I	ATT

[a] The position number corresponds to the capsid sequence of the Melksham/94/UK strain; a dash indicates sequence identity with this strain. Residues are indicated by their single-letter codes.
[b] For strain abbreviations, see Table 1.
[c] UK, United Kingdom; US, United States; NL, The Netherlands; JP, Japan.

City, and strains from Japan and The Netherlands could be consistently grouped in the same alignment, suggesting continuous widespread circulation and an ability to cause disease. The recombinant genomes with GII.2 capsids, however, occur sporadically and did not show evidence for evolution, suggesting that their circulation is limited (1, 4, 17). The finding that strains from such widely separated geographic regions were similar shows that evolution of GII.2 NoVs is a global phe-

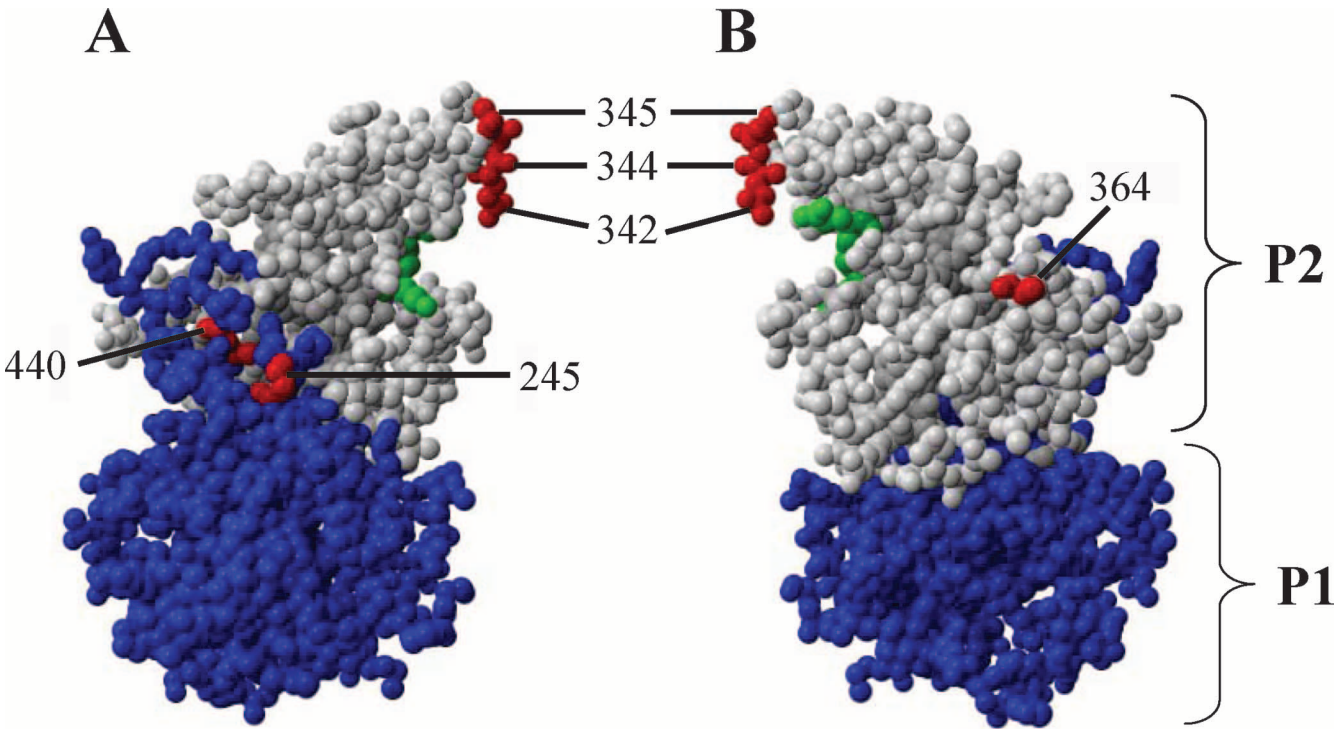

FIG. 4. Location of six fixed amino acid residues (positions 245, 342, 344, 345, 364, and 440, shown in red) on the monomer of the capsid protein. This 3D structure for the monomer P domain of the GII.2 NoV capsid protein was made by WHAT IF. The P1 and P2 subdomains are shown in blue and gray, respectively. The S domain is not shown. The KGE motif in the P2 domain is shown in green. (A) Predicted location of amino acid residues 342, 344, 345, and 364 in the P2 subdomain. (B) A view identical to panel A rotated 180° horizontally showing the location of amino acid residues 245 and 440 in the P1 subdomain.

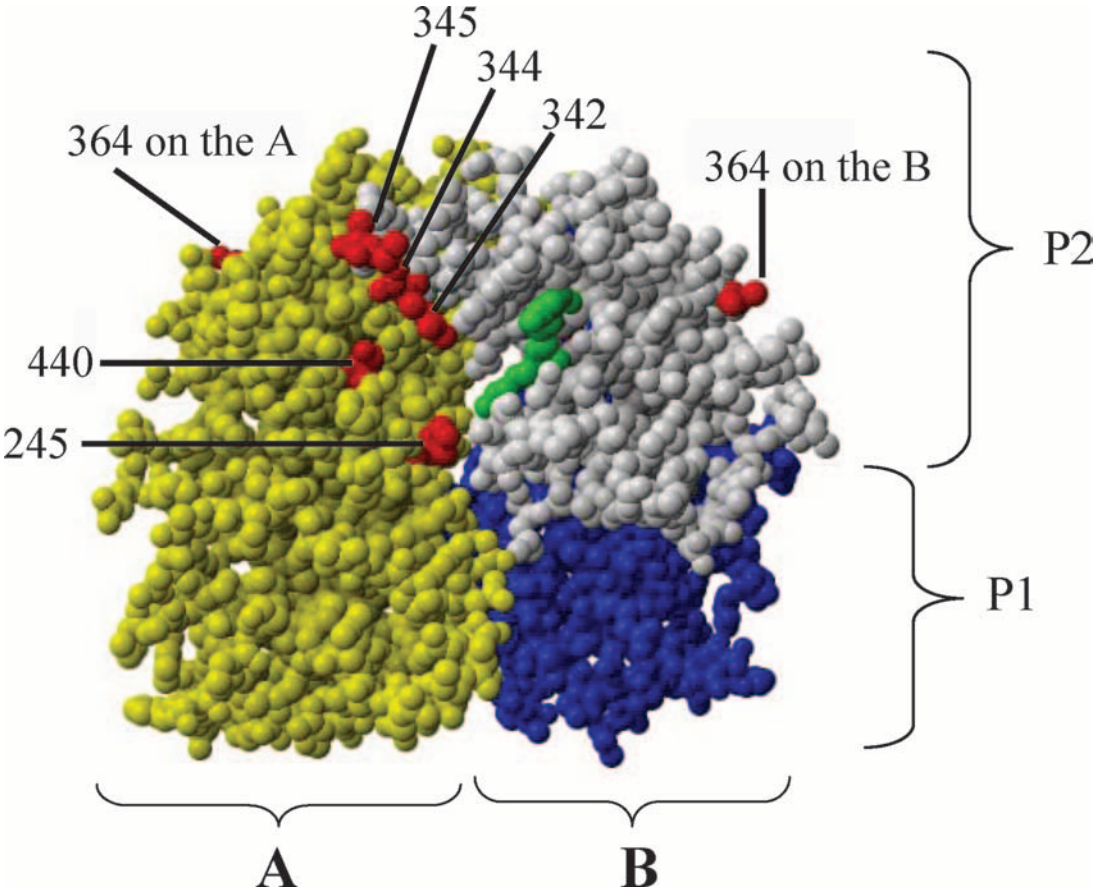

FIG. 5. The 3D structure of a dimer of the GII.2 NoV capsid protein, as predicted by WHAT IF, showing the location of the six informative amino acid residues (red). Monomer A is shown in yellow, while the monomer B is shown in blue (P1) and gray (P2). The S domains are not shown. The KGE motif (green) in the P2 subdomain on monomer B is also shown.

nomenon, similar to what has been described for GII.4 NoVs (5, 27, 33).

The molecular characterization of Melksham-like strains showed accumulation of mutations over a 12-year period, segregating them into five genetic groups. Three of these were supported in phylogenetic analysis with high bootstrap values. Two others (Fig. 2, 1999 to 2000 and 2001 to 2003) were not distinguishable in phylogenetic analysis, possibly because of the small number of isolates in the comparison. The present collection, however, contains all GII.2 strains detected in the two countries in the past 12 years of surveillance. Amino acid position 345 was the most variable in all GII.2 strains. This is suggestive of immune pressure, particularly because the amino acid is predicted to be located on a surface-exposed part of the capsid. The fixed amino acid changes were topologically clustered, judging from the 3D structure prediction, indicating that combined they may form an epitope. Immunity to NoVs has not been studied very systematically, but recent work on GII.4 NoVs has suggested that these viruses evolved under immune pressure (9, 30, 33). The lack of a tissue culture model makes it difficult to corroborate these findings by using cross-neutralization tests with hyperimmune serum directed against specific variants. Evolution of strains in an immunocompromised patient with chronic shedding of NoVs has been demonstrated by Nilsson et al. (28), who suggested that the cumulative amino acid substitutions appearing in the P2 subdomain were immune response driven. To clarify the relations between these amino acid mutations and receptor or antibody binding further immunological studies are needed.

Recently, new variants of the predominant NoV, the GII.4 genotype, appeared in 2002 (27), 2004 to 2005 (5), and 2006 (33) with global distribution. These new GII.4 variants were characterized into different clusters by the phylogenetic analysis based on the capsid gene (5, 9, 30). Okada et al. (30) reported that GII.4 strains had cumulative amino acid changes in the P domain and 5% to 18% amino acid diversity in the P2 subdomain between GII.4 subtypes for 5 years study. Bull et al. (5) also described more than 5% amino acid diversity in the complete capsid gene between past GII.4 strains and variants. Siebenga et al. (33) found 9% fixed amino acid mutations across the capsid gene over a 12-year period, but only five of these consistently delineated subsequent epidemic strains. In contrast, here we found only 2.6% amino acid diversity across the complete capsid gene of Melksham-like strains in a 12-year period, showing a clearly lower rate of change. The difference between these genotypes is intriguing and suggests clear differences in epidemiology that are not easy to explain. This study has been done using a limited number of strains in a rare genotype from a limited number of locations. It is necessary to

continue molecular surveillance of NoV infections to further the evolutional analysis of each genotype.

In conclusion, the present study showed that NoVs belonging to a rare genotype evolved by the accumulation of mutations in the surface-exposed parts of the capsid, suggesting immune response-driven evolution. The pattern of change similar to what has been observed for GII.4 is a global one, suggesting that these viruses circulate much more than is apparent from their presence in surveillance data. The data also showed for the first time that recombinant genomes of NoVs behave differently, even when they share the same capsid genes as nonrecombinant genomes. Understanding how NoVs evolve is necessary for finding more effective ways to control this disease and particularly its impact in healthcare settings.

ACKNOWLEDGMENTS

We are grateful to Erwin Duizer, Erwin de Bruin, and Bas van de Veer (RIVM, Bilthoven, The Netherlands) for supporting our work.

This work was supported by the European Commission, DG Research Quality of Life Program, 6th Framework (EVENT, SP22-CT-2004-502571), and by a grant for Research on Food Safety of the Ministry of Health, Labour and Welfare, Japan.

REFERENCES

1. **Ambert-Balay, K., F. Bon, F. L. Guyader, P. Pothier, and E. Kohli.** 2005. Characterization of new recombinant noroviruses. J. Clin. Microbiol. **43:** 5179–5186.
2. **Ando, T., J. S. Noel, and R. L. Fankhauser.** 2000. Genetic classification of "Norwalk-like viruses. J. Infect. Dis. **181**(Suppl. 2)**:**S336–S348.
3. **Bates, P. A., L. A. Kelley, R. M. MacCallum, and M. J. E. Sternberg.** 2001. Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins **45**(Suppl. 5)**:**39–46.
4. **Bull, R. A., G. S. Hansman, L. E. Clancy, M. M. Tanaka, W. D. Rawlinson, and P. A. White.** 2005. Norovirus recombination in ORF1/ORF2 overlap. Emerg. Infect. Dis. **11:**1079–1085.
5. **Bull, R. A., T. V. Elise, C. J. McIver, W. D. Rawlinson, and P. A. White.** 2006. Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis. J. Clin. Microbiol. **44:**327–333.
6. **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.
7. **Chen, R., J. D. Neil, J. S. Noel, A. M. Hutson, R. I. Glass, and B. V. Prasad.** 2004. Inter- and intragenus structural variations in caliciviruses and their functional implications. J. Virol. **78:**6469–6479.
8. **de Bruin, E., E. Duizer, H. Vennema, and M. P. G. Koopmans.** 2006. Diagnosis of norovirus outbreaks by commercial ELISA or RT-PCR. J. Virol. Methods **137:**259–264.
9. **Dingle, K. E., and Norovirus Infection Control in Oxfordshire Communities Hospitals.** 2004. Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. J. Clin. Microbiol. **42:**3950–3957.
10. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.
11. **Fankhauser, R. L., J. S. Noel, S. S. Monroe, T. Ando, and R. I. Glass.** 1998. Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United states. J. Infect. Dis. **178:**1571–1578.
12. **Filatov, D. A.** 2002. ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. Mol. Ecol. Notes **2:**621–624.
13. **Green, K. Y.** 2007. *Caliciviridae*: The noroviruses, p. 949–979. *In* B. N. Fields, D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology, 5th ed. Lippincott, Williams & Wilkins, Philadelphia, PA.
14. **Green, S. M., K. E. Dingle, P. R. Lambden, E. O. Caul, C. R. Ashley, and I. N. Clarke.** 1994. Human enteric *Caliciviridae*: a new prevalent small round-structured virus group defined by RNA-dependent RNA polymerase and capsid diversity. J. Gen. Virol. **75:**1883–1888.
15. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:**95–98.
16. **Hardy, M. E., S. F. Kramer, J. J. Treanor, and M. K. Estes.** 1997. Human calicivirus genogroup II capsid sequence diversity revealed by analyses of the prototype Snow Mountain agent. Arch. Virol. **142:**1469–1479.
17. **Iritani, N., A. Kaida, H. Kubo, N. Abe, T. Murakami, H. Venemma, M. Koopmans, N. Takeda, H. Ogura, and Y. Seto.** 21 May 2008. An epidemic of GII.2 genotype noroviruses in the spring of 2004 in Osaka City, Japan. J. Clin. Microbiol. doi:10.1128/JCM.01993-07.
18. **Jiang, X., M. Wang, K. Wang, and M. K. Estes.** 1993. Sequence and genomic organization of Norwalk virus. Virology **195:**51–61.
19. **Kageyama, T., S. Kojima, M. Shinohara, K. Uchida, S. Fukushi, F. B. Hoshino, N. Takeda, and K. Katayama.** 2003. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. J. Clin. Microbiol. **41:**1548–1557.
20. **Kageyama, T., M. Shinohara, K. Uchida, S. Fukushi, F. B. Hoshino, S. Kojima, R. Takai, T. Oka, N. Takeda, and K. Katayama.** 2004. Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to norovirus in Japan. J. Clin. Microbiol. **42:**2988–2995.
21. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
22. **Kojima, S., T. Kageyama, S. Fukushi, F. B. Hoshino, M. Shinohara, K. Uchida, K. Natori, N. Takeda, and K. Katayama.** 2002. Genogroup-specific PCR primers for detection of Norwalk-like viruses. J. Virol. Methods **100:** 107–114.
23. **Koopmans, M., J. Vinje, M. de Wit, I. Leenen, W. van der Poel, and Y. van Duynhoven.** 2000. Molecular epidemiology of human enteric caliciviruses in The Netherlands. J. Infect. Dis. **181**(Suppl. 2)**:**S262–S269.
24. **Krieger, E., G. Koraimann, and G. Vriend.** 2002. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins **47:**393–402.
25. **Lambert, C., N. Leonard, X. De Bolle, and E. Depiereux.** 2002. ESyPred3D: prediction of proteins 3D structures. Bioinformatics **18:**1250–1256.
26. **Lewis, D. C., A. Hale, X. Jiang, R. Eglin, and D. W. G. Brown.** 1997. Epidemiology of Mexico virus, a small round-structured virus in Yorkshire, United Kingdom, between January 1992 and March 1995. J. Infect. Dis. **175:**951–954.
27. **Lopman, B., H. Vennema, E. Kohli, P. Pothier, A. Sanchez, A. Negredo, J. Buesa, E. Schreier, M. Reacher, D. Brown, J. Gray, M. Iturriza, C. Gallimore, B. Bottiger, K. Hedlund, M. Torven, C. Bonsdorff, L. Maunula, M. Poljsak-Prijatelj, J. Zimsek, G. Reuter, G. Szucs, B. Melegh, L. Svennson, Y. Duijnhoven, and M. Koopmans.** 2004. Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. Lancet **363:**682–688.
28. **Nilsson, M., K. O. Hedlund, M. Thorhagen, G. Larson, K. Johansen, A. Ekspong, and L. Svensson.** 2003. 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 phenotype. J. Virol. **77:**13117–13124.
29. **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.
30. **Okada, M., T. Ogawa, I. Kaiho, and K. Shinozaki.** 2005. Genetic analysis of noroviruses in Chiba Prefecture, Japan, between 1999 and 2004. J. Clin. Microbiol. **43:**4391–4401.
31. **Prasad, B. V., M. E. Hardy, T. Dokland, J. Bella, M. G. Rossmann, and M. K. Estes.** 1999. X-ray crystallographic structure of the Norwalk virus capsid. Science **286:**287–290.
32. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572–1574.
33. **Siebenga, J. J., H. Vennema, B. Renckens, E. de Bruin, B. van der Veer, R. J. Siezen, and M. Koopmans.** 2007. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. J. Virol. **81:**9932–9941.
34. **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.
35. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
36. **Vinjé, J., J. Green, D. C. Lewis, C. I. Gallimore, D. W. Brown, and M. P. Koopmans.** 2000. Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses." Arch. Virol. **145:**223–241.
37. **Vinjé, J., R. A. Hamidjaja, and M. D. Sobsey.** 2004. 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 **116:**109–117.
38. **Vriend, G.** 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. **8:**52–56.
39. **Wang, J., X. Jiang, H. P. Madore, J. Gray, U. Desselberger, T. Ando, Y. Seto, I. Oishi, J. F. Lew, K. Y. Green, et al.** 1994. Sequence diversity of small, round-structured viruses in the Norwalk virus group. J. Virol. **68:**5982–5990.
40. **Widdowson, M. A., E. H. Cramer, L. Hadley, J. S. Bresse, R. S. Beard, S. N. Bulens, M. Charles, W. Chege, E. Isakbaeva, J. G. Wright, E. Mintz, D. Forney, J. Massey, R. I. Glass, and S. S. Monroe.** 2004. Outbreaks of acute gastroenteritis on cruise ships and on land: identification of a predominant circulating strain of norovirus—United States, 2002. J. Infect. Dis. **190:**27–36.
41. **Xi, J. N., D. Y. Graham, K. N. Wang, and M. K. Estes.** 1990. Norwalk virus genome cloning and characterization. Science **250:**1580–1583.