JOURNAL OF VIROLOGY, Sept. 1994, p. 5982–5990
0022-538X/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 68, No. 9

Sequence Diversity of Small, Round-Structured Viruses in the Norwalk Virus Group

JIANXIANG WANG,[1] XI JIANG,[1]† H. PAUL MADORE,[2]‡ JAMES GRAY,[3] ULRICH DESSELBERGER,[3] TAMIE ANDO,[4] YOSHIYUKI SETO,[5]§ ISAO OISHI,[6] JUDY F. LEW,[7] KIM Y. GREEN,[7] AND MARY K. ESTES[1]*

Division of Molecular Virology, Baylor College of Medicine, Houston, Texas 77030[1]; Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642[2]; Clinical Microbiology and Public Health Laboratory, Addenbrooke's Hospital, Cambridge CB2 2QW, United Kingdom[3]; Department of Virology, Tokyo Metropolitan Research Laboratory of Public Health, Shinjuku-ku, Tokyo 169,[4] Department of Virology, Osaka City Institute of Public Health and Environmental Sciences, Tennohji, Osaka 543,[5] and Laboratory of Virology, Osaka Prefectural Institute of Public Health, Higashinari-ku, Osaka, Osaka 537,[6] Japan; and Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Md 20892[7]

Received 27 September 1993/Accepted 9 June 1994

We have determined the nucleotide sequences of a highly conserved region of the RNA-dependent RNA polymerase of the prototype Snow Mountain agent (SMA) and of four other small, round-structured viruses (antigenically Norwalk virus [NV]-like or SMA-like) following reverse transcription-PCR amplification of viral RNA obtained from human stools. The stool samples were either from volunteers administered SMA or from sporadic cases and outbreaks of gastroenteritis that occurred in Japan and the United Kingdom between 1984 and 1992. The GLPSG and YGDD RNA polymerase motifs were in the proper locations in the sequences of the five SRSVs, but each sequence was distinct from the 8FIIa prototype NV sequence and from each other. Analysis of the sequences and reactivities in a new NV antigen enzyme-linked immunosorbent assay showed that the five viruses could be divided into two groups (serogroups) with NV and SMA, respectively, being the prototypes. The sequences of the capsid region and a nonstructural region (2C) were determined from one strain from each group. One virus (SRSV-KY-89/89/J), isolated in Japan and antigenically similar to the prototype NV (isolated 21 years earlier in Ohio), showed a remarkable level of sequence similarity to NV. KY-89 and the 8FIIa NV showed 87.2% nucleotide similarity over 2,516 continuous nucleotides amounting to 96 to 98.9% amino acid similarity in three distinct domains in two open reading frames. Between the prototype SMA and NV, the polymerase region showed 63% nucleotide and 59% amino acid similarity, respectively. Two other antigenically SMA-like isolates (SRSV-925/92/UK and SRSV-OTH-25/89/J), from the United Kingdom and Japan, showed 80% nucleotide and 88 to 92% amino acid similarity in the polymerase region to the prototype SMA isolated 16 and 13 years earlier in the United States. The capsid region of the antigenically SMA-like OTH-25 virus showed 53% nucleotide and 65% amino acid similarity to the prototype NV capsid region. Domains of sequence diversity and conservation were identified within the capsid protein of these two distinct prototype serotypes of virus. These results indicate that NV-like and SMA-like agents are still circulating, and sequence comparisons will be useful to identify and classify distinct viruses in the NV group.

Since the original report of Norwalk virus (NV) as the etiologic agent responsible for an outbreak of nonbacterial gastroenteritis in teachers, staff, and children in a school in Norwalk, Ohio (15), several viruses morphologically similar to NV have been detected and shown by serologic analysis to be associated with gastroenteritis. Because the NV and related agents cannot be cultivated in cell or organ culture or in a practical animal model, a simple and reliable method to characterize each newly identified agent has been lacking, and these agents often were named by the place where the outbreak of gastroenteritis occurred. Although some of the viruses were characterized antigenically by immune electron microscopy (IEM) and cross-challenge studies (5, 34), an interim classification system based on morphologic appearances classified these agents into a unique group designated small, round-structured viruses (SRSVs) (1).

By IEM and cross-challenge studies in human volunteers, at least three serotypes of SRSVs which are morphologically similar to NV were identified and well characterized (reviewed in reference 14). In addition to NV, two other viruses characterized as distinct serotypes were the Hawaii agent (35) and the Snow Mountain agent (SMA) (5) from the United States. Additional SRSVs include SRSVs UK 1 to UK 4 from the United Kingdom (19), SRSVs 1 to 9 from Japan (30), and human caliciviruses 1 to 4 from the United Kingdom and Japan (3). While the relationships among viruses typed by cross-challenge studies and IEM using reference human sera from volunteers allowed unequivocal type designation (5, 34), relationships among viruses studied by using other sera and methods such as solid-phase IEM (18, 19) and enzyme-linked immunoassays (ELISA) (6, 8, 21) have remained unclear.

* Corresponding author. Phone: (713) 798-3585. Fax: (713) 798-3586. BITNET address: mestes@bcm.tmc.edu.

† Present address: Center for Pediatric Research, Eastern Virginia Medical School, Norfolk, VA 23510-1001.

‡ Present address: Lederle-Praxis Biologicals, West Henrietta, NY 14586-9728.

§ Present address: Viral Gastroenteritis Section, Centers for Disease Control, Atlanta, GA 30333.

Typing of viruses with different human convalescent sera can lead to different conclusions because individual adults may respond distinctly to a second or subsequent infection by other viruses in this group (3, 20, 33).

Recently, the NV genome was cloned and characterized as a single-stranded RNA genome of about 7.7 kb that is predicted to encode three open reading frames (ORFs) (9, 13). The longest ORF (ORF1) may encode a 1,738-amino-acid polypeptide which has similarities to the polyprotein of picornaviruses, with domains of sequence identity to the picornavirus 2C-like protein, 3C-like cysteine protease, and 3D-like RNA-dependent RNA polymerase (13). ORF2 encodes the capsid protein (12). ORF3 is predicted to encode a 22,000-molecular-weight (22K) polypeptide with unknown function. The genomic organization of NV is similar to that of another SRSV from the United Kingdom called the Southampton (So) virus (16) and to other caliciviruses such as feline calicivirus (FCV) (2, 27, 32), San Miguel sea lion viruses (26), and rabbit hemorrhagic disease virus (RHDV) (23).

To gain more understanding of the genomic and antigenic relationships between NV and other SRSVs, we have determined the nucleotide sequence of the RNA polymerase region of the SMA and another four SRSVs (antigenically NV or SMA) isolated in different years and geographic locations. This region of the genome was chosen for analysis because it was expected to be relatively conserved. Therefore, the data obtained would yield information on overall genomic stability. We also extended the cDNA sequence to include the capsid protein and the 2C region from two of the five isolates that represent the NV and SMA groups. This report is the first to show that SMA has the same genome organization as NV and to describe the genetic heterogeneity and relationships between NV and other SRSVs, including a second prototype strain, SMA.

MATERIALS AND METHODS

Virus samples and outbreaks. The stool samples containing virus used in this study were initially characterized and supplied by different laboratories. All samples were positive for SRSV by electron microscopy or IEM, and no other type of viral particles was observed. To facilitate description of the viruses in these stools, we have adopted a method that describes the viruses isolated from the samples by using the following cryptogram: SRSV-strain designation/year of isolation/country of isolation (10). Once the viruses have been defined in this manner, we refer to them by strain designation/year of isolation/abbreviation for country of isolation or simply by strain designation. The viruses studied were SRSV-Sa-1283/84/Japan (Sa-1283/84/J or Sa-1283) from a sporadic case of diarrhea in a 1-year-old female (30), SRSV-KY-89/89/Japan from a 37-year-old male who had ingested uncooked oysters, and SRSV-OTH-25/89/Japan from a 2-year-old female infant (29), as well as SRSV-925/92/UK, SRSV-916/92/UK, and SRSV-982/92/UK from a diarrheal outbreak that occurred at Peterhouse College, Cambridge, United Kingdom, after oysters were consumed at a dinner on February 7, 1992. In this outbreak, 50 of 90 to 100 participants were clinically ill. The predominant symptoms were diarrhea, nausea, vomiting, abdominal cramps, and headache, and the duration of symptoms was 2 to 3 days. The oysters were harvested from coastal waters near Pembrokeshire in Wales and smoked at 60°C for 12 h before being sold. Analyses of questionnaires completed by the diners indicated that the oysters were the probable source of infection. Only those who ate oysters, including one person who ate six oysters and nothing else during the meal, were ill.

TABLE 1. Sequences of primers used for RT-PCR

Primer name	Position of 5′ nucleotide	Polarity	Primer sequence, 5′ to 3′
36	4475	+	ATAAAAGTTGGCATGAACA
35	4944	−	CTTGTTGGTTTGAGGCCATAT
69	4721	+	GGCCTGCCATCTGGATTGCC
39	4878	−	GTTGACACAATCTCATCATC
78	1670	+	GGGCCCCCTGGTATAGGTAA
80	1958	−	TGGTGATGACTATAGCATCAGACACAAA
56	4903	+	ACTCACCCAAATCCTCAA
23	5247	−	TCCAATCCACCAGTCTTG
42	5595	+	AGTTTGGGTCCCCATCTTAATCCTTT
55	5747	−	TGAACCAAAACCAGGGGG
58	5210	+	AGCAAAGTCATACATGAAAT
59	5650	−	CCATTATACATTTGTGA
60	5712	+	ATAATAGTTTCCTGCATA
61	6134	−	CACACTCTGGACATTGTCTG
76	6095	+	TAGTGGCATGGGTATTTC
77	6333	−	TATGCCAATCACAGCCAC
72	6296	+	CATTGGGTTTCCAGACCT
63	6530	−	ATAATTGGGGATCTTCCAAA
64	6491	+	GTCTGGCTCCCAAGTTGACC
75	6744	−	CGGTATCAGGGTCAACAT
74	6707	+	TGAGGCTGCCCTGCTCCA
3	7027	−	GAGGGCCGCTTCGCCACC
82	4544	+	CACTATGATGCAGATTA
81	4869	−	ATCTCATCATCACCATA
86[a]		+	TGGGAGGGCGATCGCAATCT
95[a]		−	CGCTCAGATTGTCAATAGG
96[a]		+	GGAAATGCGTTTACAGCAGGA
97[a]		−	AAAGCACATATTTGGCTCGG
98[a]		+	AATGGGCGCGTCACTCTTG
99[a]		−	CACCAGGAAAATTGGGGGC
94[a]		+	TGGTCCTTACCCAACTATT
93[a]		−	CTATAAAAGCTCCAGCCAT

[a] Selected on the basis of the sequence of the Toronto agent (formerly designated minireovirus). The precise location within the genome is not yet known (17).

SRSVs were seen by direct electron microscopy in 6 (54.5%) of 11 fecal samples collected and in each of the samples used for sequence analysis. The sample containing SMA (SRSV-SMA/76/US) was from a human volunteer (NY12) experimentally infected with a virus pool that originated from individuals ill during a waterborne outbreak of gastroenteritis at a resort camp (25). The KY-89 and OTH-25 SRSVs were chosen for study because they had been found to have different patterns of protein reactivity by Western blot (immunoblot) analysis (29). KY-89 contained two proteins of 62K and 33K, while OTH-25 contained only a single 33K protein by Western blot analysis (29). Sa-1283 originally was classified as NV related by IEM (30).

Primers used in RT-PCR. Three sets of primers based on the sequence of the NV genome initially were designed for the reverse transcription (RT)-PCR detection of SRSVs in the NV group (Table 1). The primer pair designated 36/35 is located in a highly conserved region (RNA polymerase region) of the NV genome, and it is predicted to produce a product of 470 bases. Primers 39 and 69 are a nested set of primers located within the 36/35 region, and they produce a fragment of 158 bases. Primers 78 and 80 are located in the nonstructural 2C region, and they produce a 289-base product. All primers were synthesized and purified by The Midland Certified Reagent Company, Midland, Tex. Other primers were used for the cloning and sequence analysis of the capsid protein of KY-89

and OTH-25. Some of these primers were selected on the basis of the sequence of the minireovirus (17).

Sample preparation for detection of SRSVs by RT-PCR. Viral RNA was extracted from stools for RT-PCR testing by the cetyltrimethylammonium bromide method described previously (11). Stool suspensions (10%) were extracted with an equal volume of genetron (1,1,2-trichloro-1,2,2-trifluoroethane; Dupont Co., Wilmington, Del.), and after clarification, virus in the supernatant was concentrated by precipitation with 8% polyethylene glycol 6000 (BDH Chemicals, Gallard-Schlesinger, Carle Place, N.Y.) and 0.4 M NaCl and then centrifugation. The pellets were suspended and digested with proteinase K, and the viral RNA was purified by phenol-chloroform extraction in the presence of cetyltrimethylammonium bromide (Sigma Chemical Co., St. Louis, Mo.) followed by extraction with chloroform. Viral RNA then was precipitated with ethanol, and pelleted viral RNA was suspended in a final volume of 20 μl of Milli-Q water and used directly for RT-PCR or precipitated with ethanol and stored at −20°C until used.

Conditions used for RT-PCR to detect virus in stools. For each sample, RT-PCR was performed as described previously (11). Briefly 5 μl of purified viral RNA was reverse transcribed for 1 h at 42°C with 10 U of avian myeloblastosis virus reverse transcriptase in 25 μl of PCR buffer containing 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 3.3 mM each dATP, dCTP, dGTP, and dTTP, 1.0 μM primer, and 40 U of RNasin. PCR (40 cycles) was performed by adding 70 μl of PCR buffer containing the second primer and 5 U of *Taq* polymerase (Perkin-Elmer Cetus Corp., Norwalk, Conn.) to the RT reaction mixture. Amplification consisted of denaturation for 1 min at 94°C, annealing for 1 min 20 s at 55°C, and extension for 1 min at 72°C. The first denaturation time was increased to 4 min, and the final extension time was increased to 15 min. The PCR products were analyzed by agarose gel electrophoresis and visualized by staining with ethidium bromide and UV illumination. RT-PCR first was performed with each sample at 55°C. If no RT-PCR product was obtained, the same sample was tested again, using a lower (37°C) annealing temperature in the amplification cycle to try and detect samples containing significant sequence mismatches between the primers and target nucleic acid. A significant number of positive results leading to new sequence information was obtained with this second lower-temperature run.

Cloning and sequencing of RT-PCR-amplified products. Because direct sequencing of the RT-PCR-amplified products sometimes gave inconsistent results, we cloned the RT-PCR product first and then sequenced the cloned cDNA. Cloning was done using the pCR II vector in the TA cloning kit (Strategene, La Jolla, Calif.), and sequencing was performed by using Sequenase and the dideoxynucleotide chain termination method (DNA sequencing kit; U.S. Biochemical Corp., Cleveland, Ohio).

NV antigen ELISA. An antigen ELISA (7) was used to detect viruses antigenically related to NV in the stool samples collected from different geographic areas. Hyperimmune rabbit serum and guinea pig serum produced against the recombinant NV particles produced in insect cells (10) were used as the capture and detector antibodies, respectively. NV-positive stool samples from volunteers infected with NV were included in each plate as a positive control. Samples with an optical density at 414 nm of >0.5 and positive/negative ratio of >2 were considered to be positive for NV. A prechallenge and convalescent serum pair specific for SMA (33) was used in IEM to determine if any of the viruses negative in the NV ELISA were antigenically related to SMA.

TABLE 2. Detection of SRSVs by ELISA and RT-PCR with different primer pairs

SRSV strain	Reactivity			
	ELISA	Primers 36 and 35	Primers 69 and 39	Primers 78 and 80
NV-8FIIa/68/US	+	+	+	+
KY-89/89/J	+	+	+	+
Sa-1283/84/J	+	+	+	+
SMA/76/US	−	+[a]	+[a]	−
925/92/UK	−	+	+[a]	−[b]
OTH-25/89/J	−	+	+[a]	+[a,c]

[a] RT-PCR positive at low temperature.
[b] RT-PCR negative at low temperature.
[c] Size of PCR product smaller than predicted.

Computer sequence analysis of the amplified cDNA products. Computer analyses of the amplified cDNA products were performed by using the Molecular Biology Computational Resource computer network at Baylor College of Medicine. Multiple alignment of the sequences was done using the PILEUP program of the GCG package developed by the University of Wisconsin.

Nucleotide sequence accession numbers. Sequence data from this study have been deposited with the EMBL/GenBank data libraries. The accession numbers for the various viruses are L23831 (SMA/76/US), L23829 and L23930 (OTH-25/89/J), L23826 (925/92/UK), L23832 (Sa-1283/84/J), and L23827 and L23828 (KY-89/89/J).

RESULTS

Design of primers and RT-PCR amplification of the 3D-like RNA polymerase and 2C-like regions from SRSVs. Three primer pairs were used to detect a variety of SRSVs (Tables 1 and 2). An initial primer set (36/35) to detect SRSVs genomically related to NV was selected because it contained a higher number of base matches with other caliciviruses (FCV and RHDV) than with picornaviruses. For example, the sequence of these primers was relatively well conserved (18-of-21-bp and 16-of-19-bp similarity of primers 35 and 36, respectively) compared with the F9 FCV strain. Therefore, this primer pair was expected to detect only caliciviruses and not picornaviruses. In control experiments, primer set 36/35 did not detect poliovirus type 1 and hepatitis A virus.

A second pair of primers (69/39), located inside the region of 36/35, was designed after sequence information was accumulated within the 36/35 region from field isolates of SRSVs. Primer set 69/39 contained a higher number of base matches with the sequences of the newly detected SRSVs, and studies of field samples found that primer set 69/39 detected more positive samples than did primer set 36/35 (unpublished data). The smaller product (158 bp) from primer set 69/39 also resulted in higher sensitivity of virus detection (unpublished observations). Primer set 78/80 was designed as a supplement to the two sets of primers described above. This set of primers is located in the highly conserved 2C region of the NV genome and detected a subset of the viruses (Table 2).

Regardless of whether a sample was positive in the NV antigen ELISA, primer set 36/35 was able to amplify one band of the expected size (470 bp), using an annealing temperature of 55°C to perform the RT-PCR (Table 2). The SMA sample yielded a RT-PCR product only when the annealing temperature was 37°C. Some samples showed extra bands other than the major 470-bp band when the amplification was performed

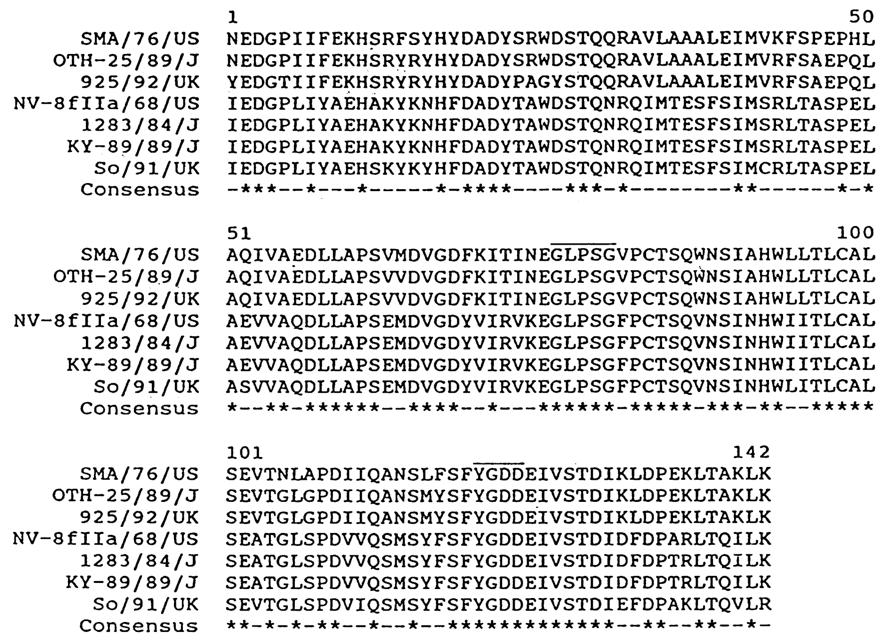

FIG. 1. Alignment of the predicted amino acid sequences of the RNA-dependent RNA polymerase region of seven different SRSVs. The predicted sequences of the RT-PCR-amplified products obtained by using primers 36 and 35 are aligned from amino acids 1451 to 1592 of the predicted ORF1 of the NV genome (13) (EMBL/GenBank databank accession number M87661). The sequence of the So virus was from GenBank accession number L07418. This alignment was done by using the PILEUP program of the GCG package, and the asterisks below the alignment show the consensus sequences. The highly conserved RNA polymerase motifs are indicated by lines.

at lower temperature (37°C). These positive results indicating the viruses in all the stool samples were SRSVs genomically related to NV were confirmed by sequence analysis of the amplified DNA fragments (see below).

Amplification of the RNA from the stool samples with primer set 69/39 produced the same results as those obtained with primer set 36/35. Expected products of 158 bases were obtained for the viruses tested; however, it was necessary to perform RT-PCR at lower temperature to obtain products from the three viruses that were negative in the NV antigen ELISA. Sequence analysis of the RT-PCR products confirmed the results previously obtained with primer set 36/35. Primer set 69/39 also was shown to be useful for nested RT-PCR following an initial amplification of stool samples with primer set 36/35 (data not shown). However, because of the increased potential for contamination during nested PCR, this method was not used further.

Amplification of the viruses by using primer set 78/80 was successful for samples that were closely related to NV and for the OTH-25 sample when the reaction was performed at a low temperature (although the size of the PCR product was smaller than with NV and KY-89).

Comparison of the sequences of two nonstructural regions of different SRSVs. Figure 1 shows an alignment of the deduced amino acid sequences of the five SRSVs tested in this study compared with the sequences of the prototype NV and the So/91/UK isolate recently reported to be in the SMA serotype (16). Two other SRSVs from the same outbreak as 925/92/UK had the same nucleotide sequence as 925/92/UK (data not shown). The high degree of similarity observed in these viruses suggests that these SRSVs are closely related. The highly conserved motifs (GLPSG and YGDD) of the RNA-dependent RNA polymerases of RNA viruses were observed in all of the SRSVs.

The percentages of the nucleotide and amino acid sequence similarities between pairs among these viruses are shown in Table 3. On the basis of the level of sequence similarity, the SRSVs fell into one of two groups of viruses, with the prototype NV and the SMA being representative strains of each group. These two viruses previously had been reported to be antigenically distinct from each other on the basis of IEM studies (5). SRSVs So/91/UK, KY-89/89/J, and SA-1283/84/J would belong to the NV group, as determined from their higher levels of nucleotide similarities (78, 87, and 87%, respectively) and amino acid similarities (92, 99, and 99%, respectively) to NV and lower levels of nucleotide similarities (64, 63, and 62%, respectively) and amino acid similarities (64, 59, and 59%, respectively) to SMA. SRSVs 925/92/UK and OTH-25/89/J would belong to the SMA group because they share 97 and 95% nucleotide similarity with each other and

TABLE 3. Relationships of the sequences of the RNA polymerase region among SRSVs[a]

	NV-8FIIa	KY-89	Sa-1283	So	SMA	925	OTH-25
NV-8FIIa/68/US		87	87	78	63	62	62
KY-89/89/J	99		96	77	63	61	61
Sa-1283/84/J	99	100		77	62	62	61
So/91/UK	92	91	91		64	65	65
SMA/76/US	59	59	59	64		80	80
925/92/UK	60	60	60	63	88		97
OTH-25/89/J	61	61	61	65	92	95	

[a] The sequences from nucleotides 4494 to 4923 of the sequence of the polymerase region of ORF1 of NV-8FIIa (13) were compared with the similar sequences among the indicated pairs of virus strains. The numbers in the upper triangle show the percent identity of the aligned nucleic acid sequences. The numbers in the lower triangle show the percent identity of the aligned amino acids.

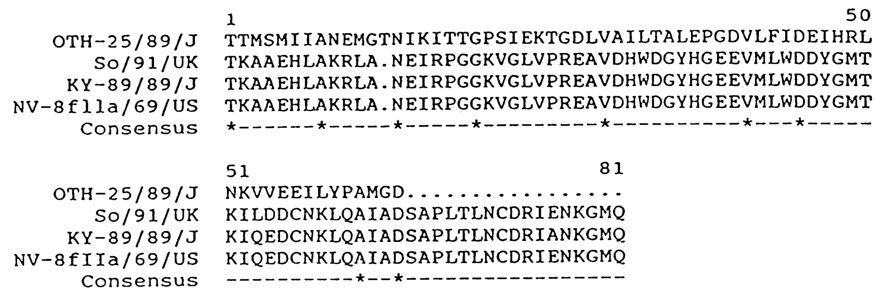

FIG. 2. Alignment of the predicted amino acid sequence of a portion of the 2C-like protein within ORF1 of the different SRSVs. The sequences of the SRSVs were from the RT-PCR products obtained by using primers 78 and 80, and the amino acid sequences are aligned with amino acids 516 to 597 of the NV ORF1. The origin of the So virus sequence is as described in the legend to Fig. 1.

80% nucleotide similarity with SMA but only 62% nucleotide similarity with NV. This grouping was consistent with the antigenic behavior of these viruses in the newly developed antigen ELISA. The two viruses in the NV group available for testing (KY-89 and Sa-1283), and none of the viruses in the SMA group, were positive when tested by the NV antigen test (Table 2). The So virus in the NV group was not available for antigen testing. The OTH-25 and UK-925 viruses also were tested by IEM and found to react with a convalescent sera (but not with the prechallenge sera) specific for SMA (data not shown). These results validated the use of the OTH-25 virus as a surrogate for SMA.

Comparison of the partial sequences of the 2C region of these viruses showed that this region was not as well conserved among the NV and SMA-like viruses as was the polymerase region (Fig. 2). The 2C region of the OTH-25/89/J virus was 16 amino acids shorter than the same region of the other viruses.

Comparisons of the sequences of the structural region of a second NV-like SRSV and of a SMA-like SRSV. RT-PCR, cloning, and sequencing were used to extend the DNA sequence from the 36/35 polymerase region to other parts of the viral genome. KY-89/89/J, the first nonprototype virus studied, was chosen for two reasons. First, KY-89 was closely related to NV, since it was detected in the NV antigen ELISA. Second, KY-89 was isolated in Japan 21 years after the prototype 8FIIa NV was isolated in Norwalk, Ohio. The primers used for RT-PCR with KY-89 were designed on the basis of the 8FIIa NV sequence, and our efforts were focused on obtaining sequence between the polymerase region and the 3′ end of the genome so that the sequence of the capsid region would be obtained. Individual pairs of primers crossing the 3′ end of the NV genome were randomly selected to amplify KY-89 (Table 1). Because this virus is closely related to NV, most of these primers yielded RT-PCR products. However, regions close to the 3′ end of the genome were difficult to amplify with the randomly selected NV primers. Therefore, primers specific for KY-89 were selected on the basis of the newly obtained viral sequence. A total of nine overlapping clones (ranging in size between 153 and 441 bp) covering 3 kb of the KY-89 viral genome were obtained.

The genomic organization of the extended 3′ end of the viral genome of KY-89 was identical to that of NV. Specifically, the viral RNA polymerase (from primer 36 to the end of ORF1) was found at the 5′ end of the sequenced cDNAs, ORF2 was in the center, and ORF3 was at the 3′ end of the extended region (data not shown). Comparison of the sequences of the KY-89 and NV showed that most of the nucleotide changes (84%) in KY-89 were third-base changes. A similar degree of sequence divergence was observed between both the nonstructural protein (2C and RNA polymerase) and capsid regions of the KY-89 virus and NV (Table 4). Thus, the KY-89 virus, isolated in Japan 21 years after the prototype was isolated in the United States, showed relatively minor sequence divergence.

The predicted amino acid sequence of the capsid of OTH-25, a SMA-like virus, also was obtained and compared with the KY-89 sequence and other available sequences (Fig. 3 and data not shown). The predicted capsid of OTH-25 was 548 amino acids in length, 18 amino acids longer than the predicted NV capsid (Fig. 3). The predicted OTH-25 capsid contained a large insertion between amino acids 296 and 320. This site is of interest as it is in a region that is distinct among the So virus, OTH-25, and NV capsid sequences. Three additional small insertions were found in the OTH-25 and So virus capsid sequences at amino acids 366 to 380, 393 to 395, and 460 to 463 but not in the capsid sequences of NV and KY-89 (Fig. 3). The

TABLE 4. Nucleotide and amino acid sequence identities among two SRSVs (KY-89 and OTH-25) and NV

Virus	Location (nucleotides) within the NV genome	Nucleotides		Amino acids	
		Length	% Identity with NV	Length	% Identity with NV
KY-89/89/J					
	ORF1 2C (1690–1930)	241	87.5	80	98.8
	ORF1 3D (4494–5359)	866	87.9	288	99.7
	ORF2 (5346–6935)	1,590	87.0	530	98.3
	ORF3 (6938–7009)	72	84.7	24	100.0
OTH-25/89/J					
	ORF1 2C (1690–1880)	194	36.1	64	50.0
	ORF1 3D (4494–5359)	869	62.1	289	79.4
	ORF2 (5346–6935)	1,644	53.4	548	65.4

```
                 1                                      ┌          50
   OTH-25/89/J   MKMASNDAAPSNDGAAG...LVPEIN.NEAMALDPVAGAAIAAPLTGQQN
      So/91/UK   MMMASKDAPQSADGASGAGQLVPEVNTADPLPMEPVAGPTTAVATAGQVN
     KY-89/89/J   MMMASKDATSSVDGASASVQLVPEVNASDPLAMDPVAGSSTAVATAGQVN
NV-8fIIa/68/US   MMMASKDATSSVDGASGAGQLVPEVNASDPLAMDPVAGSSTAVATAGQVN  50
     Consensus   *-***-**--*-***-----****-*---------****---*-----**-*
                                                       └

                 51                                                100
   OTH-25/89/J _ IIDPWIMNNFVQAPGGEFTVSPRNSPGEVLLNLELGPEINPYLAHLARMY
      So/91/UK   MIDPWIVNNFVQSPQGEFTISPNNTPGDILFDLQLGPHLNPFLSHLSQMY
     KY-89/89/J   PIDPWIINNFVQAPQGEFTISPNNTPGDVLFDLSLGPHLNPFLLHLSQMY
NV-8fIIa/68/US   PIDPWIINNFVQAPQGEFTISPNNTPGDVLFDLSLGPHLNPFLLHLSQMY  100
     Consensus   -*****-*****-*-****-**-*-**--*-*-***--**-*-**--**

                 101                                               150
   OTH-25/89/J   NGYAGGFEVQVVLAGNAFTAAKVIFAAIPPNFPIDNLSAAQITMCPHVIV
      So/91/UK   NGWVGNMRVRILLAGNAFSAGKIIVCCVPPGFTSSSLTIAQATLFPHVIA
     KY-89/89/J   NGWVGNMRVRIMLAGNAFTAGKIIVSCIPPGFGSQQLTIAQATLFPHVIA
NV-8fIIa/68/US   NGWVGNMRVRIMLAGNAFTAGKIIVSCIPPGFGSHNLTIAQATLFPHVIA  150
     Consensus   **--*---*---******-*-*-*-----**-*-----*--**-*--****-

                 151                                               200
   OTH-25/89/J   DVRQLEPINLPMPDVRNNFFHYNQGSDSRLRLIAMLYTPLRANNSGD..D
      So/91/UK   DVRTLEPIEMPLEDVRNVLYHTND.NQPTMRLVCMLYTPLRTGGGSGNSD
     KY-89/89/J   DVRTLDPIEVPLEDVRNVLFHNNDRNQQTMRLVCMLYTPLSTGGGTG..D
NV-8fIIa/68/US   DVRTLDPIEVPLEDVRNVLFHNNDRNQQTMRLVCMLYTPLRTGGGTG..D  198
     Consensus   ***-*-**--*--****---*-*-------**--*****---------*

                 201                                               250
   OTH-25/89/J   VFTVSCRVLTRPSPDFSFNFLVPSTMESKTKPFTLPILTISEMSNSRFPV
      So/91/UK   SFVVAGRVLTAPSSDFSFLFLVPPTIEQKTRAFTVPNIPLQTLSNSRFPS
     KY-89/89/J   SFVVAGRVMTCPSPDFNFLFLVPPTVEQKTRPFTLPNLPLSSLSNSRAPL
NV-8fIIa/68/US   SFVVAGRVMTCPSPDFNFLFLVPPTVEQKTRPFTLPNLPLSSLSNSRAPL  248
     Consensus   -*-*--**-*-**-**-*-****-*-*-**-**-*------****-*-
                              ^

                 251                                      ┐        300
   OTH-25/89/J   PIDSLHTSPTESIVVQCQNGRVTLDGELMGTTQLLPNQICAFRGTLTRST
      So/91/UK   LIQGMILSPDASQVVQFQNGRCLIDGQLLGTTPATSGQLFRVRGK.....
     KY-89/89/J   PISGMGISPDNVQSVQFQNGRCTLDGRLVGTTPVSLSHVAKIRGT.....
NV-8fIIa/68/US   PISSIGISPDNVQSVQFQNGRCTLDGRLVGTTPVSLSHVAKIRGT.....  293
     Consensus   -*-----**------**-****----**-*-***-----------**------
                                                          ┘

                 301                                               350
   OTH-25/89/J   NRASDQADTATPRLFNHHWHIQLDNLNGTPYDPAEDIPAPLGTPDFRGKV
      So/91/UK   ...INQGART..........LNLTEVDGKPFMAFDS.PAPVGFPDFGKCD
     KY-89/89/J   ...SN..GTV..........INLTELDGTPFHPFEG.PAPIGFPDLGGCD
NV-8fIIa/68/US   ...SN..GTV..........INLTELDGTPFHPFEG.PAPIGFPDLGGCD  327
     Consensus   ---------------------*-----*-*--------***-*-**-----

                 351                                               400
   OTH-25/89/J   FGVA...GQRNPDSTTRAHEAKVDTTSGRFTPKLGSLEITTESDDFDPNQ
      So/91/UK   WHMRISKTPNNTGSGDPMRSVSVQTNVQGFVPHLGSIQFDEVFNHPTGDY
     KY-89/89/J   WHINMTQFGHSSQTQ.....YDVDTTPDTSVPHLGSIQANGI...GSGNY
NV-8fIIa/68/US   WHINMTQFGHSSQTQ.....YDVDTTPDTFVPHLGSIQANGI...GSGNY  369
     Consensus   ----------------------*-*-------*-***---------------

                 401                                               450
   OTH-25/89/J   STKFTPVG.IGVDNEADFQQWSLPDYSGQFTHNMNLAPAVAPNFPGEQLL
      So/91/UK   IGTIEWISQPSTPPGTDINLWEIPDYGSSLSQAANLAPPVFPPGFGEALV
     KY-89/89/J   IGVLSWVSPPSHPSGSQVDLWKIPNYGSSITEATHLAPSVYSPGFGEVLV
NV-8fIIa/68/US   VGVLSWISPPSHPSGSQVDLWKIPNYGSSITEATHLAPSVYPPGFGEVLV  419
     Consensus   --------------------*--*-*---------***-*------**-*-

                 451                                               500
   OTH-25/89/J   FFRSQLPSSGGRSNGI.LDCLVPQEWVQHFYQESAPAQTQVALVRYVNPD
      So/91/UK   YFVSAFPGPNNRSAPNDVPCLLPQEYITHFVSEQAPTMGDAALLHYVDPD
     KY-89/89/J   FFMSKIPGP....GGDSLPCLLPQGYISHLASEQAPTVGEGPLLHYVDPD
NV-8fIIa/68/US   FFMSKMPGP....GAYNLPCLLPQEYISHLASEQAPTVGEAALLHYVDPD  465
     Consensus   -*-*--*--------------**-**-----*----*-**-------*--**-**

                 501                                               550
   OTH-25/89/J   TGRVLFEAKLHKLGFMTIAKNGDS..PITVPPNGYFRFESWVNPFYTLAP
      So/91/UK   TNRNLGEFKLYPGGYLTCVPNGVGAGPQQLPLNGVFLFVSWVSRFYQLKP
     KY-89/89/J   TDRNLGEFKAYPDGFLTCVPNGASSGPQQLPINGVFVFVSWVSRFYQLKP
NV-8fIIa/68/US   TGRNLGEFKAYPDGFLTCVPNGASSGPQQLPINGVFVFVSWVSRFYQLKP  515
     Consensus   *-*-*-*-*----*--*---**----*---*-**-*-*-***--**-*-*

                 551           566
   OTH-25/89/J   MGTGN.GRRRIQ....
      So/91/UK   VGTASTARGRLGVRRI
     KY-89/89/J   VGTASTARGRLGLRR.
NV-8fIIa/68/US   VGTASSARGRLGLRR. 530
     Consensus   -**----*-*------
```

FIG. 3. Alignment of the deduced amino acid sequences of the capsid regions of the different SRSVs. The sequences shown align the complete predicted amino acids of the four different SRSVs. This alignment was done by using the PILEUP program of the GCG package. The sequence of OTH-25/89/J is shown first, as it has the longest predicted capsid protein. The highly conserved region (called region B by Neill [26]) is bracketed. The conserved PP is highlighted by a bar, and the cleavage site detected for NV in insect cells is shown by an arrow.

predicted cleavage site between amino acids 218 and 219 (F^ LVPP detected in insect cells expressing NV capsids [12]) was conserved among So and KY-89 but conserved only partially in the OTH-25 virus, which contained FLVPS instead of FLVPP. Only the PP of the PPG triplet noted previously to be present in VP3 of picornaviruses and the NV capsid were conserved among all of these viruses, with OTH-25 again having the divergent sequence. Between OTH-25 and NV, the highest sequence similarity was in the polymerase region followed by the capsid region (Table 4).

Analysis of the relationships of NV and SMA by sequence comparisons. We also examined the relatedness of the predicted polymerase and capsid regions of these viruses in relation to other caliciviruses and picornaviruses. When the polymerase regions were compared, the two Japanese NV isolates and the So virus clustered with NV and the OTH-25/89/J and 925/92/UK isolates clustered with SMA (Fig. 4). Similarly, comparisons of the capsid sequences indicated that the So virus was closer to NV than to OTH-25, the SMA-like virus (Fig. 5).

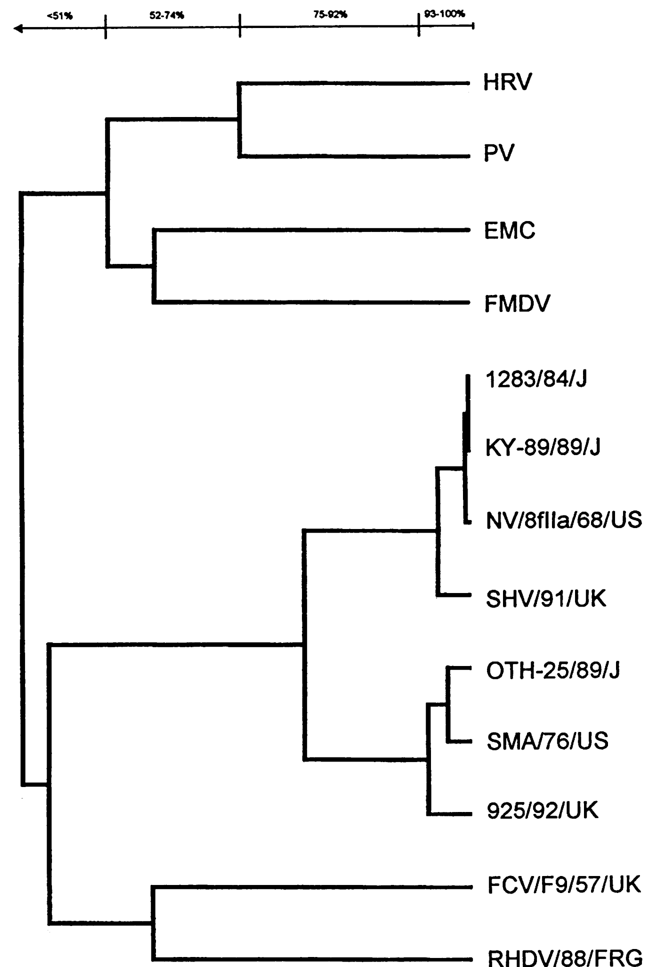

FIG. 4. Dendrogram of predicted genetic relationships among the different SRSVs by comparison of a part of the polymerase region within ORF1. Sequences from amino acids 1451 to 1592 of the NV genome were used to determine the order of the pairwise alignments. The horizontal branch lengths are proportional to the similarities of the sequences. The sequences were compared with those of two caliciviruses (feline calicivirus [FCV] and rabbit hemorrhagic disease virus [RHDV]) and four picornaviruses. The accession numbers for the sequences used for comparison were M86370 (FCV), M67473 (RHDV), V01150 (poliovirus [PV] type 1), K02121 (human rhinovirus [HRV] type 14]), M33457 (encephalomyocarditis virus [EMC]), and X00871 foot-and-mouth disease virus [FMDV]).

DISCUSSION

RT-PCR combined with sequence analysis has been found to be a powerful technique for study of sequence variation in many virus families. This study demonstrated the usefulness of this technique in studying genetic diversity of SRSVs in the NV group. This technique is particularly important for this group of viruses and other related SRSVs, because these viruses still cannot be grown in either cell culture or animal models.

Following cloning and sequencing of the amplified RT-PCR products, a sequence database can be obtained. This information is valuable for understanding the relationships between different isolates and their evolution over time. Accumulation of such information first from well-characterized SRSVs is essential for understanding and developing a logical typing system for these viruses. This information also will be important for understanding the natural history, epidemiology, and importance of specific viruses in causing human disease and for designing the most optimal primers for virus detection and characterization by RT-PCR. For example, sequence analysis recently showed that the minireovirus is a calicivirus that is closely related to SMA (17).

We have found that viruses responsible for acute epidemic gastroenteritis of humans show a spectrum of different nucleotide sequences. This study shows that a subset of these viruses can be put into two genogroups which correspond to two of the prototype serotypes (NV and SMA) described in previous IEM studies (5). We propose designating these genogroups 1 and 2 for NV and SMA, respectively. In a number of recent collaborative studies, primers 36 and 35 have been used to obtain viral cDNA sequences from different SRSV isolates. In these studies, viruses closely and distantly related to NV were detected and sequenced (4, 17, 22, 24, 28). Because additional characterization of these strains is necessary for final classification, we have proposed and used an interim nomenclature system for SRSVs. This system designates each virus by the term SRSV followed by the strain designation, year of isolation, and country in which the virus was isolated. Because our recent studies have found that some human virus strains with typical calicivirus morphology contain nucleotide sequences very similar to that of NV or SMA (4), ultimately, all of these viruses should be classified as human caliciviruses instead of as SRSVs. When this is done, we propose naming these agents HuCV followed by the virus strain name, the year of isolation, and the place of isolation. As serotypes are clarified, the serotype designation can be added. This proposal initially was described in a summary of our collaborative studies (10).

The two groups of viruses (NV and SMA) identified in this study on the basis of their genetic variation agree with the early characterization of these agents as being serologically distinct. They also are antigenically distinct in our newly developed antigen ELISA for NV. Development of a specific antigen ELISA for the viruses in the SMA group and two-way antigenic testing of previously classified and unclassified SRSV isolates should confirm the validity of these groups and determine whether intermediate types exist between these two groups. For example, such testing should resolve the classification of the So SRSV. Lambden et al. (16) noted that the So virus was a serotype 3 SMA-like virus on the basis of testing with the solid-phase IEM typing system of Lewis (19). Com-

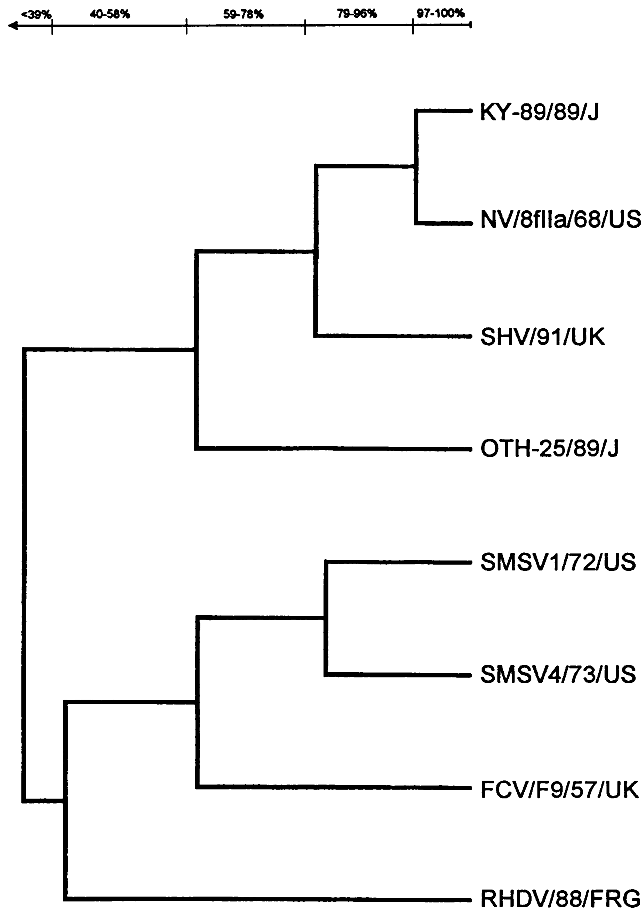

FIG. 5. Dendrogram of predicted genetic relationships among the different SRSVs by comparison of the capsid regions. The 530 amino acids of the predicted capsid protein of the NV genome were used to determine the order of the pairwise alignments. The horizontal branch lengths are proportional to the similarities of the sequences. The heterologous viruses used for comparison are the same as in Fig. 4.

parison of our sequence data of SMA with the sequence of the So virus indicates the So virus sequence is more closely related to that of NV than to that of SMA (Table 3 and Fig. 3 to 5). This finding suggests that the solid-phase IEM typing system may not always be correct, possibly because it relies on the use of human serum reagents whose specificity can never be ensured. Sera from humans may show cross-reactivities with NV-like viruses in more than one serotype, as documented by showing that the recombinant NV can detect cross-reactive seroresponses in volunteers given the Hawaii agent or SMA (34). It will be important to determine how the So SRSV reacts in the new NV antigen ELISA, because this virus had the lowest similarity with viruses in the NV group. We would predict that this virus will react in the NV antigen assay because testing of other samples to date has shown NV antigen-positive reactivities with viruses with polymerase amino acid sequence similarities as low as 89% (10). Obtaining sequence in the capsid region of additional distinct virus serotypes (such as the Hawaii agent) is needed to permit a clearer correlation and definition of regions of sequence diversity that will discriminate different serotypes. Accumulation of sequence information will facilitate our ability to understand the evolution of these RNA viruses.

ACKNOWLEDGMENTS

This work was supported in part by Public Health Service cooperative agreement AI30448 from the National Institute of Allergy and Infectious Diseases, cooperative agreement NA17FQ0550 from the National Shellfish Indicator Study funded by the National Oceanic and Atmospheric Administration, and the Public Health Laboratory Service, United Kingdom.

We thank A. Kapikian for thoughtful comments, K. Yamazaki, S. Sekine, and Y. Hayashi for help in the initial identification and characterization of the SRSVs from Japan, and Michele Hardy and Judy Ball for help in revising the figures and performing the immune electron microscopy, respectively.

REFERENCES

1. **Appleton, H.** 1987. Small round viruses: classification and role in food-borne infections. CIBA Found. Symp. **128:**108–119.
2. **Carter, M. J., I. D. Milton, J. Meanger, M. Bennett, R. M. Gaskell, and P. C. Turner.** 1992. The complete nucleotide sequence of a feline calicivirus. Virology **190:**443–448.
3. **Cubitt, W. D., N. R. Blacklow, J. E. Herrmann, N. A. Nowak, S. Nakata, and S. Chiba.** 1987. Antigenic relationships between human caliciviruses and Norwalk virus. J. Infect. Dis. **156:**806–813.
4. **Cubitt, W. D., X. Jiang, J. Wang, and M. K. Estes.** Sequence similarity of human caliciviruses and small round structured viruses. J. Med. Virol., in press.
5. **Dolin, R., R. C. Reichman, K. D. Roessner, T. S. Tralker, R. T. Schooley, W. Gary, and D. Morens.** 1982. Detection by immune electron microscopy of the Snow Mountain Agent of acute viral gastroenteritis. J. Infect. Dis. **146:**184–189.
6. **Gary, G. W. Jr., J. E. Kaplan, S. E. Stine, and L. J. Anderson.** 1985. Detection of Norwalk virus antibodies and antigen with a biotin-avidin immunoassay. J. Clin. Microbiol. **22:**274–278.
7. **Graham, D. Y., X. Jiang, T. Tanaka, A. Opekun, P. Madore, and M. K. Estes.** 1994. Norwalk virus infection of volunteers: new insights based on improved assays. J. Infect. Dis. **170:**34–43.
8. **Herrmann, J. E., N. A. Nowak, and N. R. Blacklow.** 1985. Detection of Norwalk virus in stools by enzyme immunoassay. J. Med. Virol. **17:**127–133.
9. **Jiang, X., D. Y. Graham, K. Wang, and M. K. Estes.** 1990. Norwalk virus genome cloning and characterization. Science **250:**1580–1583.
10. **Jiang, X., J. Wang, and M. K. Estes.** Characterization of SRSVs using RT-PCR and a new antigen ELISA, a short communication. Arch. Virol., in press.
11. **Jiang, X., J. Wang, D. Y. Graham, and M. K. Estes.** 1992. Detection of Norwalk virus in stool by polymerase chain reaction. J. Clin. Microbiol. **30:**2529–2534.
12. **Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes.** 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J. Virol. **66:**6527–6532.
13. **Jiang, X., M. Wang, K. Wang, and M. K. Estes.** 1993. Sequence and genomic organization of Norwalk virus. Virology **195:**51–61.
14. **Kapikian, A. Z., and M. K. Estes.** 1993. The Norwalk and related viruses, p. 924–933. *In* R. G. Webster and A. Granoff (ed.), Encyclopedia of virology. Academic Press Ltd., London.
15. **Kapikian, A. Z., R. G. Wyatt, R. Dolin, R. 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.
16. **Lambden, P. R., E. O. Caul, C. R. Ashley, I. N. Clarke.** 1993. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. Science **259:**516–519.
17. **Lew, J. F., M. Petric, A. Z. Kapikian, X. Jiang, M. K. Estes, and K. Y. Green.** 1994. Identification of minireovirus as a Norwalk-like virus in pediatric patients with gastroenteritis. J. Virol. **68:**3391–3396.
18. **Lewis, D. C.** 1990. Three serotypes of Norwalk-like virus demonstrated by solid-phase immune electron microscopy. J. Med. Virol. **30:**77–81.
19. **Lewis, D. C.** 1991. Norwalk agent and other small-round structural viruses in the U.K. J. Infection **23:**220–222.
20. **Madore, H. P., J. J. Treanor, R. Buja, and R. Dolin.** 1990.

Antigenic relationships among the Norwalk-like agents by serum antibody rises. J. Med. Virol. **32:**96–101.

21. **Madore, H. P., J. J. Treanor, K. A. Pray, and R. Dolin.** 1986. Enzyme-linked immunosorbent assays for Snow Mountain and Norwalk agents of viral gastroenteritis. J. Clin. Microbiol. **24:**456–459.
22. **Matson, D. O., W. M. Zhong, S. Nakata, X. Jiang, L. K. Pickering, and M. K. Estes.** Unpublished data.
23. **Meyers, G., C. Wirblich, and H.-J. Thiel.** 1991. Rabbit hemorrhagic disease virus—molecular cloning and nucleotide sequencing of a calicivirus genome. Virology **184:**664–676.
24. **Moe, C. L., J. Gentsch, T. Ando, G. Grohmann, S. S. Monroe, X. Jiang, J. Wang, M. K. Estes, Y. Seto, C. Humphrey, S. Stine, and R. I. Glass.** 1994. Application of PCR to detect Norwalk virus in fecal specimens from outbreaks of gastroenteritis. J. Clin. Microbiol. **32:**642–648.
25. **Morens, D. M., R. M. Zweighaft, T. M. Vernon, G. W. Gary, J. J. Eslien, B. T. Wood, R. C. Holman, and R. Dolin.** 1979. A waterborne outbreak of gastroenteritis with secondary person-to-person spread. Lancet **ii:**964–966.
26. **Neill, J. D.** 1992. Nucleotide sequence of the capsid protein gene of two serotypes of San Miguel sea lion virus: identification of conserved and non-conserved amino acid sequences among calicivirus capsid proteins. Virus Res. **24:**211–222.
27. **Neill, J. D., I. M. Reardon, and R. L. Heinrikson.** 1991. Nucleotide sequence and expression of the capsid protein gene of feline calicivirus. J. Virol. **65:**5440–5447.
28. **Numata, K., S. Nakata, X. Jiang, M. K. Estes, and S. Chiba.** 1994. Epidemiological study on Norwalk virus infection in Japan and Southeast Asia using enzyme-linked immunosorbent assays with baculovirus-expressed Norwalk virus capsid protein. J. Clin. Microbiol. **32:**121–126.
29. **Oishi, I., K. Yamazaki, T. Kimoto, and Y. Minekawa.** 1992. Demonstration of low molecular weight polypeptides associated with small, round-structured viruses by Western immunoblot analysis. Microbiol. Immunol. **36:**1105–1112.
30. **Okada, S., S. Sekine, T. Ando, Y. Hayashi, M. Murao, K. Yabuuchi, T. Miki, and M. Ohashi.** 1990. Antigenic characterization of small, round-structured viruses by immune electron microscopy. J. Clin. Microbiol. **28:**1244–1248.
31. **Thornhill, T. S., R. G. Wyatt, A. R. Kalica, R. Dolin, R. M. Chanock, and A. Z. Kapikian.** 1977. Detection by immune electron microscopy of 26-27nm virus-like particles associated with two family outbreaks of gastroenteritis. J. Infect. Dis. **135:**20–27.
32. **Tohya, Y., Y. Taniguchi, E. Takahashi, E. Utagawa, N. Takeda, K. Miyamura, S. Yamazaki, and T. Mikami.** 1991. Sequence analysis of the 3′-end of feline calicivirus genome. Virology **183:**810–814.
33. **Treanor, J. J., X. Jiang, H. P. Madore, and M. K. Estes.** 1993. Subclass-specific serum antibody responses to recombinant Norwalk capsid antigen (rNV) in adults infected with Norwalk, Snow Mountain, or Hawaii viruses. J. Clin. Microbiol. **31:**1630–1634.
34. **Wyatt, R. G., R. Dolin, N. R. Blacklow, H. L. DuPont, R. F. Buscho, T. S. Thornhill, A. Z. Kapikian, and R. M. Chanock.** 1974. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. J. Infect. Dis. **127:**709–714.