

Characterization of a novel human calicivirus that may be a naturally occurring recombinant

X. Jiang¹, C. Espul², W. M. Zhong¹, H. Cuello², and D. O. Matson¹

¹Center for Pediatric Research, Children's Hospital of The King's Daughters and Eastern Virginia Medical School, Norfolk, Virginia, U.S.A.

²Central Hospital, Mendoza, Argentina

Accepted July 15, 1999

Summary. We identified a Norwalk-like calicivirus (CV) whose genome likely was derived from naturally occurring recombination. This strain (Arg320) was detected by the EIA developed against recombinant Mexico virus (rMxV) capsids, but the viral RNA polymerase sequence was closer to Lordsdale virus, in a separate genetic cluster of Norwalk-like viruses. A 3.3 kb cDNA from the RNA polymerase region to the 3' end of the genome of Arg320 was cloned and sequenced. The sequence demonstrated that the capsid region of Arg320 shared 95% amino acid identity with MxV, but 68% identity with Lordsdale virus, while the RNA polymerase region shared 95% identity with Lordsdale virus, but 87% identity with MxV. Pair-wise sequence comparisons identified a potential recombination site at the polymerase/capsid junction. This is the first example of a naturally occurring recombinant in the CV family. Further studies to search for and characterize other strains may be necessary for understanding the genetic diversity of the family.

Introduction

Caliciviruses (CVs) are single-stranded, positive-sense RNA viruses that are closely related to picornaviruses. Four CV genera have been described, including "Norwalk-like viruses", "Sapporo-like viruses", *Vesivirus*, and *Lagovirus* [1]. CVs are genetically diverse but share a common genomic organization. All CVs have at least three genes, from 5' to 3' being a nonstructural polyprotein, a capsid gene, and a small protein of unknown function. In Sapporo-like viruses and *Lagoviruses*, the capsid gene is fused to the non-structural gene and most Sapporo-like viruses have an additional small ORF overlapping the capsid gene [10, 11, 19].

CVs share amino acid motifs with picornaviruses but have a different genomic organization. The capsid gene of CVs is switched in its 5'-3' position in the genome compared with that of picornaviruses [10, 11, 19], suggesting the two families might derive from one another by recombination. RNA recombination has been shown to occur in animal hosts by recovery of recombinant polioviruses from vaccinees [14, 15], however, the frequency of such recombination in nature remains unknown. RNA recombination has been suggested among CVs [5], but direct evidence remains lacking.

In this study, we describe a human CV (HuCV) whose genome likely was derived from recombination. A continuous genomic fragment spanning the RNA polymerase region to the 3' end of the genome was sequenced and the sequence demonstrated that the RNA polymerase and capsid genes of the strain shared closest sequence identities with distinct HuCV genetic clusters. Sequence alignment also suggested a potential recombination site.

Materials and methods

Stool specimens

Stool specimens from children hospitalized with diarrhea in Mendoza, Argentina, were collected in 1995 (3). A subset of 150 stool specimens that were negative for rotaviruses was selected for testing of HuCVs by recombinant EIAs and RT-PCR. Stool specimens were kept at -20°C before testing.

Detection of viral antigens in stool specimens by recombinant EIAs

EIAs specific for detection of Norwalk virus (NV) and Mexico virus (MxV) based on hyperimmune antisera against baculovirus-expressed recombinant NV (rNV) and MxV (rMxV) capsid proteins were used for detection of viral antigens in stool specimens [8, 9]. A positive result was defined by a P/N ratio ≥ 2.0 between the OD values in the wells coated with pre-immunized vs post-immunized hyperimmune antisera in each assay.

Detection of viral RNA by reverse transcription-polymerase chain reaction (RT-PCR)

Viral RNA was extracted from stool specimens using the Trizol method (GibcoBRL, Grand Island, NY). RT-PCR was performed to detect two regions of HuCV genomic RNA. The first RT-PCR used primers NV36/51 (8) to amplify the RNA-dependent RNA polymerase. The second RT-PCR used primer NV36 and oligo-dT to amplify a 3.3 kb cDNA fragment from the RNA polymerase region to the 3' end of the genome. Both RT-PCR used similar conditions except the RT-PCR for a long product (>3 kb) used a long (3–5 min) extension time [10].

Cloning and sequencing RT-PCR products

Viral cDNA fragments with expected sizes determined by agarose gel electrophoresis were cloned into pGEM-T (Promega, Madison, WI). For sequencing the cDNA (400 bp) amplified by primer pair NV36/51, the reverse and forward primers flanking the cDNA in the plasmid vector were used. Multiple steps were used to determine the sequence of the entire 3.3 kb cDNA. The end sequences of the cDNA were determined using the same forward and reverse primers. Internal sequences were determined using subclones generated by restriction enzyme digestion and re-ligation when convenient restriction enzyme sites were available. The

remaining sequences of the cDNAs were determined by genomic walking with primers based upon newly obtained sequences adjacent to those regions. All sequences were confirmed by sequencing from both strands and at least twice.

Sequence analysis of cloned viral cDNAs

The sequences of the cloned viral cDNA were analyzed using PC/GENE software, version 6.90 (IntelliGenetics, Mountain View, CA). Multiple sequence alignment and a dendrogram predicting genetic relationships were generated using the CLUSTAL program, based upon the method of Higgins et al. [7]. Pair-wise sequence comparison between different strains was performed manually on the alignment and numbers of sequence identities between strains were calculated in a moving frame of 25 nt. Sequences used in the analysis included Lordsdale virus (X86557) and the MxV (U22498) from the GenBank. The RNA polymerase sequences of additional strains were derived from either the GenBank or our database. The GenBank sequence number for strain Arg320 is AF190818 for the RNA polymerase region and AF190817 for the 3.3 kb cDNA at the 3' end of the genome.

Results

Identification of a group of HuCVs potentially being natural recombinants

In the collection of diarrhea stools from children in Mendoza, seven (5%) stool specimens from seven hospitalized children were positive for HuCVs by the rMxV EIA and by RT-PCR with RNA polymerase region primers. None of the seven specimens was positive by the rNV EIA. Sequence analysis of the RT-PCR products showed that the seven strains belonged to two genetic clusters of Norwalk-like viruses, containing the prototype Snow Mountain virus (SMV, 2). Two of the seven strains belonged to the MxV cluster and the other five belonged to the Lordsdale cluster. The two MxV-like strains revealed 95% nucleotide and amino acid identities with the prototype MxV. The five "Lordsdale like" strains showed 85% nt identity with the prototype Lordsdale virus, but only 75% nt identity with MxV. In our previous studies of HuCV variation, if a strain was positive by the rMxV EIA, it usually shared a high sequence identity (>95%) in the RNA polymerase region with the prototype MxV. We hypothesized the five "Lordsdale-like" strains might be recombinants.

Characterization of the potential recombinant by cloning and sequencing both the RNA polymerase and capsid genes

One Mendoza stool (Arg320) containing a "Lordsdale-like" virus was selected to clone a cDNA from the RNA polymerase region to the 3' end of the genome. A 3.3 kb cDNA was amplified using a one-step RT-PCR with primer NV36 and oligo-dT. The cDNA then was cloned and sequenced. To ensure the sequence was not derived from contamination by another virus, the two ends of the cDNA were re-sequenced on purified plasmid DNAs from a single bacterial colony. Sequences identical with previously determined sequences at the two ends were obtained from the re-sequencing. The resulting sequence contained 3 312 bp plus a poly-A tail and revealed typical HuCV genomic features. The 5' end of the cDNA encoded the C-terminus of ORF1, including the RNA polymerase gene. This

RNA polymerase sequence matched the sequence obtained earlier. The second ORF started at nt 853 and ended at nt 2 496, encoding a protein of 548 aa which is predicted to be the viral capsid protein. The third ORF started at nt 2 499 and ended at nt 3 259, encoding a protein of 252 aa with unknown function. A non-coding region of 55 nt was found between the third ORF and the poly-A tail. There was no sequence match in the GenBank or in our library of unpublished CV sequences.

Sequence alignment of Arg320 with MxV and Lordsdale virus

Sequence alignment of the 3.3 kb Arg320 cDNA with corresponding regions of MxV and Lordsdale virus showed that Arg320 shared 86% and 71% nt identities with MxV and Lordsdale virus, respectively, however, the relative sequence identities differed between the RNA polymerase and capsid sequences. The capsid region of Arg320 shared high (95%) aa identity with MxV, confirming the rMxV EIA result, but low (68%) identity with Lordsdale virus. The RNA polymerase region of Arg320 shared high (95%) aa identity with Lordsdale virus, but lower (87%) aa identity with MxV. Multiple sequence alignment of the capsid sequences placed Arg320 in the MxV cluster, but multiple sequence alignment of the RNA polymerase genes placed it in the Lordsdale cluster (Fig. 1).

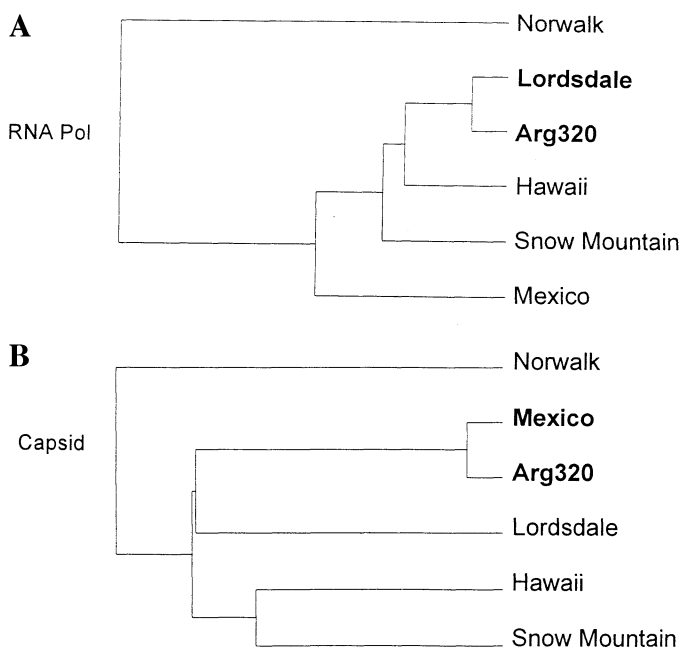


Fig. 1. Dendrogram of the amino acid sequences of the RNA polymerase (**A**) and capsid (**B**) regions of Arg320, MxV, Lordsdale virus and other prototypes of human CVs. Partial amino acid sequences of the RNA polymerase (from primer 36 to the end of the gene) and the entire capsid gene were aligned using CLUSTAL in the PC/GENE package

Pair-wise sequence comparison between Arg320, MxV, and Lordsdale virus

To localize the potential recombination site and to understand a possible recombination mechanism, we performed pair-wise sequence comparisons between Arg320, MxV, and Lordsdale virus. The plotted relative nt identities showed that Arg320 shared a constantly low level of sequence identity in the RNA polymerase region and a constantly high level of nt identity in the capsid region with MxV, and a constantly high level of sequence identity in the polymerase region but a constantly low level of nt identity in the capsid region with Lordsdale virus. A single cross of the two curves was observed at the ORF1/capsid junction (Fig. 2). The identity levels found in the capsid genes extended to ORF3 and the 3' non-coding region. When MxV and Lordsdale virus were compared, no significant change of sequence identity between the RNA polymerase and capsid regions was observed and the curves did not cross with the two curves described above (Fig. 2).

Although MxV and Lordsdale virus shared 70% nt identity over the 3.3 kb cDNA, a small region with significantly higher nt identity was observed around the ORF1/capsid junction (Figs. 2 and 3). This highly conserved region starts at nt 750, 103 nt upstream from the capsid, and ends at nt 875, including 23 nt of the capsid gene. Of 126 nt in the region, 116 (92%) were identical between the two

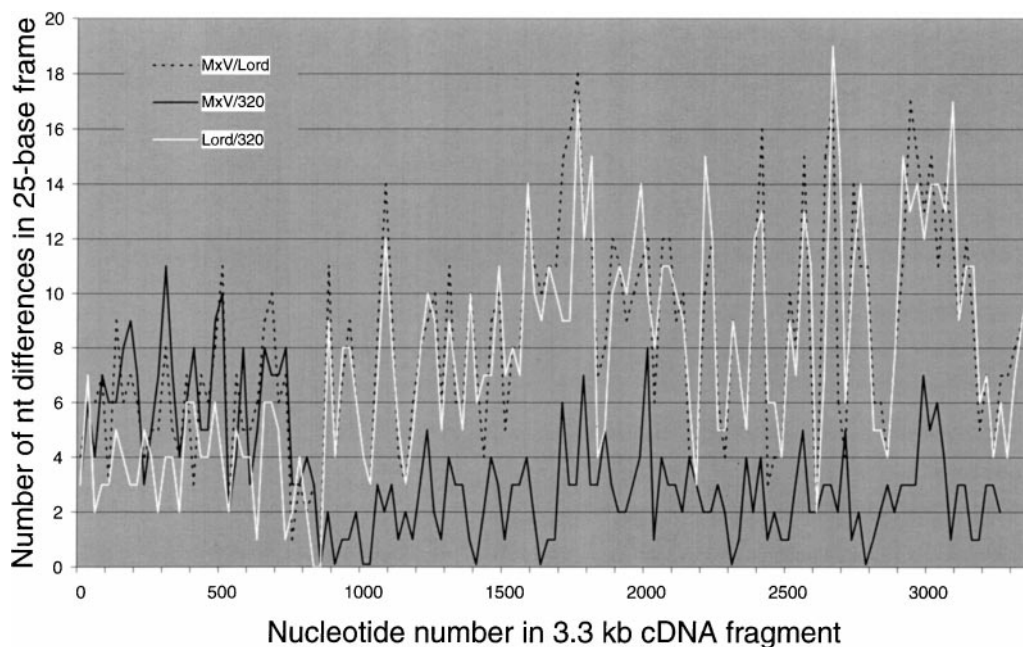


Fig. 2. Pair-wise sequence comparisons between Arg320, MxV, and Lordsdale virus. The nucleotide sequences of the 3.3 kb cDNA of Arg320 and the corresponding sequences of MxV and Lordsdale virus were aligned pair-wise using PC/GENE. The numbers of different nucleotides in a moving frame of 25 nucleotides were plotted according to the position in the consensus sequence after the alignment

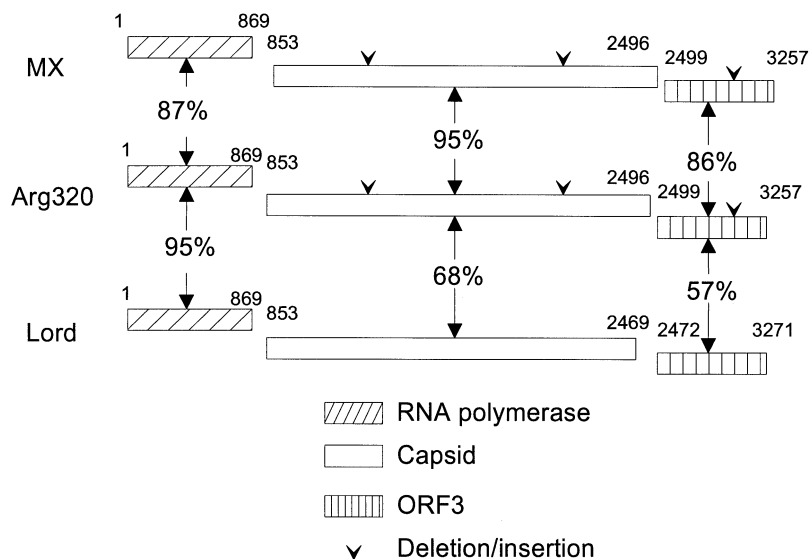


Fig. 4. Comparison of sequence identity and genomic organization of Arg320, MxV, and Lordsdale virus. Insertion/deletions were identified based upon amino acid sequence alignment of the capsid and ORF3, respectively. The two insertion/deletion sites in capsid gene shared between Arg320 and MxV were found at amino acids 295 and 401 for the two capsid genes. The single insertion/deletion site in ORF3 shared by Arg320 and MxV was found at amino acid 42 of the two viruses

strains. The predicted subgenomic RNAs of the two strains start in this region (nt 850). The first 28 nt at the 5' end of the viral subgenomic RNA were identical between MxV and Lordsdale virus (Fig. 3).

Additional features of Arg320's genome revealed different relationships with MxV and Lordsdale virus. For example, the capsid protein of Arg320 (548 aa) was larger than that of Lordsdale virus (539 aa), but was the same as that of MxV (Fig. 4). Two identical deletion/insertion regions were identified in the capsid gene of Arg320 and MxV, but not for Lordsdale virus. ORF3 of Arg320 also had the same size (252 aa) as that of MxV, which is smaller than that of Lordsdale virus (276 aa). One deletion/insertion was identified in ORF3 of MxV and Arg320, but not for Lordsdale virus.

Discussion

This study describes a HuCV strain that likely was derived from naturally occurring recombination. The viral RNA polymerase gene shares closest sequence with one HuCV strain but its capsid gene shares closest sequence identity with another HuCV strain. Spontaneous point mutation would not result in such sequence changes between two regions next to each other in a genome. The cDNA used to obtain the sequence was generated from one RT-PCR reaction covering the entire region. Therefore, the cDNA was from a single virus, not from multiple viruses due to contamination or dual infection of the child. The sequences of both ends of the cDNA were confirmed by simultaneously sequencing of purified

cDNA from a single bacterial colony, excluding the possibility of contamination during the sequencing. Finally, the two potential parental strains are in the range of genetic diversity among strains currently co-circulating [4, 20, 24, 25], therefore, the recombination event could have occurred recently, but not necessarily during the infection of this child.

Naturally occurring recombination of HuCVs was suggested when the capsid and RNA polymerase regions of SMV were characterized [6]. The capsid nucleotide sequences of SMV and Melksham virus (MeV) were almost identical (94%) but the RNA polymerase sequence were significantly different (79%). When MeV was compared with MxV and a Japanese strain, Oth-25, there seemed no significant difference of sequence identities between the capsid and RNA polymerase regions. Therefore SMV likely is a recombinant virus. However, the capsid and RNA polymerase region amplicons of SMV were generated separately and that fact does not exclude the possibility of different source of strains, which was not the case for Arg320 described in this study.

The change of relative sequence identity occurred at the ORF1/capsid junction, indicating the recombination site occurred in this region. This site also is suggested to be the break and rejoin site between caliciviruses and picornaviruses, although direct evidence of recombination between the two families remains lacking. The high sequence conservation in the ORF1/capsid junction among the two strains representing the putative parent strains and the putative recombinant suggests the recombination might be driven by homologous RNA interaction. The "copy choice" model of RNA recombination has been preferred for single-stranded RNA viruses, including picornaviruses and coronaviruses [2, 13, 17, 18, 21, 23]. Co-infection of different HuCV strains in one patient has been observed in our surveillance of HuCV-associated outbreaks of acute gastroenteritis in Virginia [12] and mixed infection of multiple strains in one outbreak has been reported [5]. Therefore, high sequence conservation would increase the interaction between RNA molecules from different strains and recombination could occur as a result of such homologous interaction. In the copy-choice model, recombination could occur during RNA replication with the involvement of viral RNA polymerase, which may switch templates of two different strains at the highly conserved region of the genomes of the two strains.

CVs contain subgenomic RNA that covers the entire 3' end of the genome from the capsid gene to the 3' end. Therefore, the CV subgenomic RNA could act as an independent unit participating in the recombination event. It remains unclear how CV subgenomic RNA is involved in virus replication, but it is clear that both CV genomic and subgenomic RNAs share highly conserved 5' end sequence and both RNAs are assembled into virions. The high conserved 5' end of the subgenomic RNAs between MxV and Lordsdale virus also is included in the highly conserved ORF1/capsid region. Therefore, interaction between genomic and subgenomic RNAs could occur by the same mechanisms as that of genomic-genomic interaction, which would significantly increase the chance of recombination events that otherwise result only from genomic-genomic RNA interaction. Furthermore, because both genomic and subgenomic RNA are

assembled into virions, co-infection of different strains may result in reassortment of different genomic RNAs from different strains, as found in the influenza virus [26], which could result in recombination in future generations, although direct evidence of reassortment of CV RNAs remains lacking.

Although recombination was predicted to occur at the ORF1/capsid junction, the precise recombination site was difficult to pinpoint because of the high conservation in this region between the two strains representing putative parents. Furthermore, because the putative recombination event may be predicted to have occurred in this part, the sequences at both sides of the recombination site of the progeny strain may have changed since the recombination event due to naturally occurring point mutations. The parent strains also would have changed due to the same mechanism. Thus, we may never observe 100% sequence identity of corresponding regions that would identify the parent strains. Because the potential recombination event described in this paper may have occurred recently, the sequences adjoining the potential recombination site maintained a high level of identity, which allowed us to identify the recombinant. It would not be difficult to imagine many CV strains that currently are circulating could have derived from recombination events in the remote past. The existence of a recombinant within the CV family also provides understanding of the relationship between picornaviruses and caliciviruses which belong to a superfamily [22] that probably was divided into two families by recombination.

This study extends our knowledge of the genetic diversity of CVs. If the RNA recombination described in this study is a common phenomenon, we would see more diversity of the family, which would make genotyping more difficult. For example, many reports on CV genotyping have been based upon sequences of the RNA polymerase region, due to its relatively high sequence conservation and ease of designing PCR primers. In comparison, only a few capsid genes have been characterized. The viral capsid protein is responsible for antigenicity and possibly protective immunity. Genotyping of CVs based upon the RNA polymerase sequences clearly is not the best choice if ORF1/capsid gene recombination is common. In addition, it remains unclear whether additional recombination sites exist. Continued searches for Arg320-like and other recombinant strains are necessary.

The significance of this study has been emphasized by the occurrence of another four strains in Mendoza that are apparently similar (sharing >98% amino acid sequence identities in the RNA polymerase region with Arg320) and by a recent identification of an Arg320-like strain in an outbreak of gastroenteritis in Virginia [12]. These findings indicate that such recombinants are genetically stable, are still circulating, cause illness, and possibly have a wide distribution. The fact that strains similar to both putative parents of the recombinant are currently co-circulating not only allowed us to identify the recombinant, but also provide indirect evidence that natural recombination is still occurring. In another words, if the parental strains had less survival advantage than the recombinant and the recombination event occurred long ago, we would not be able to recognize the recombinant because representatives of the parental genetic clusters would no

longer exist. Therefore, molecular recombination could markedly increase the genetic diversity of currently circulating strains in the CV family.

Acknowledgements

This study was supported by the U.S. Public Health Service (AI37093). We thank Tamas Berke for critical discussion on this paper. This study was presented in part at the International Workshop on Human Caliciviruses, Atlanta, Georgia, March 1999.

References

1. Berke T, Golding B, Jiang X, Cubitt WD, Wolfaardt M, Smith AW, Matson DO (1997) A phylogenetic analysis of the caliciviruses. *J Med Virol* 52: 419–424
2. Copper PD (1977) Genetics of picornaviruses In: Fraenkel-Conrat H, Wagner RR (eds) *Comprehensive virology*, vol 9: Regulation and genetics. Plenum Press, New York, pp 133–207
3. Espul C, Martinez N, Jiang X, Zhong WM, Cuello H, Matson DO (1999) Sequence diversity of human caliciviruses in a province of Argentina (in preparation)
4. Fankhauser RL, Noel JS, Monroe SS, Ando T, Glass RI (1998) Molecular epidemiology of “Norwalk-like viruses” in outbreaks of gastroenteritis in the United States. *J Infect Dis* 178: 1 571–1 578
5. Gray JJ, Green J, Cunliffe C, Gallimore C, Lee JV, Neal K, Brown DW (1997) Mixed genogroup SRSV infection among a party of canoeists exposed to contaminated recreational water. *J Med Virol* 52: 425–429
6. Hardy ME, Kramer SF, Treanor JJ, Estes MK (1997) Human calicivirus genogroup II capsid sequence diversity revealed by analyses of the prototype Snow Mountain agent. *Arch Virol* 142: 1 469–1 479
7. Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. *Comput Appl Biosci* 8: 189–191
8. Jiang X, Wang JX, Estes MK (1995) Characterization of SRSVs using RT-PCR and a new antigen ELISA. *Arch Virol* 140: 363–374
9. Jiang X, Cubitt D, Hu J, Treanor JJ, Dai X, Matson DO, Pickering LK (1995) Development of a type-specific EIA for detection of Snow Mountain agent-like human caliciviruses in stool specimens. *J Gen Virol* 76: 2 739–2 747
10. Jiang X, Cubitt DW, Berke T, Dai XM, Zhong WM, Pickering LK, Matson DO (1997) Sapporo-like human caliciviruses are genetically and antigenically diverse. *Arch Virol* 142: 1 813–1 827
11. Jiang X, Wang M, Wang K, Graham DY, Estes MK (1993) Sequence and genomic organization of Norwalk virus. *Virology* 195: 51–61
12. Jiang X, Wilton N, Zhong WM, Farkas T, Barrett E, Guerrero M, Ruiz-Palacios G, Green KY, Green J, Hale AD, Estes MK, Pickering LK, Matson DO (1999) Diagnosis of human caliciviruses by enzyme immune assays. International Workshop on Human Caliciviruses, Atlanta, Georgia, March 29–31
13. Keck JG, Stohlman SA, Soe LH, Makino S, Lai MMC (1987) Multiple recombination sites at the 5′-end of murine coronavirus RNA. *Virology* 156: 331–341
14. Kew OM, Nottay BK (1984) Evolution of the oral polio vaccine strains occurs by both mutation and intramolecular recombination. In: Chanock RM (ed) *Modern approaches to vaccines: molecular and chemical basis of virus virulence*. Cold Spring Harbor Laboratory New York, pp 357–362
15. Kew OM, Nottay BK, Hatch MH, Nakanomo JH, Obijeski JF (1981) Multiple genetic changes can occur in the oral poliovaccines upon replication in humans. *J Gen Virol* 56: 337–347

16. King AM, Slade WR, Neuman JWI, McCahon D (1980) Temperature sensitive mutants of foot-and-mouth disease virus with altered structural polypeptides. II Comparisons of recombination and biochemical maps. *J Virol* 34: 67–72
17. King AM, McCahon D, Slade WR, Neuman JWI (1982) Recombination in RNA. *Cell* 29: 921–928
18. Kirkegaard K, Baltimore D (1986) The mechanism of RNA recombination of poliovirus. *Cell* 47: 433–441
19. Liu BL, Clarke IN, Caul EO, Lambden PR (1995) Human enteric caliciviruses have a unique genome structure and are distinct from the Norwalk-like viruses. *Arch Virol* 140: 1–12
20. Maguire AJ, Green J, Brown DW, Desselberger U, Gray JJ (1999) Molecular epidemiology of outbreaks of gastroenteritis associated with small round-structured viruses in East Anglia, United Kingdom, during the 1996–1997 season. *J Clin Microbiol* 37: 81–89
21. Makino S, Keck JG, Stohlman SA, Lai MMC (1986) High-frequency RNA recombination of murine coronaviruses. *J Virol* 57: 729–737
22. Strauss JH, Strauss EG (1988) Evolution of RNA viruses. *Annu Rev Microbiol* 42: 657–683
23. Tolskaya EA, Romanova LA, Kolesniova MS, Agol VI (1983) Intertypic recombination in poliovirus: Genetic and biochemical studies. *Virology* 124: 121–132
24. Vinjie J, Koopmans MP (1996) Molecular detection and epidemiology of small round-structured viruses in outbreaks of gastroenteritis in the Netherlands. *J Infect Dis* 174: 610–615
25. Vinjie J, Altena SA, Koopmans MPG (1997) The incidence and genetic variability of small round-structured viruses in outbreaks of gastroenteritis in The Netherlands. *J Infect Dis* 176: 1 374–1 378
26. Webster RG, Laver WG, Air GM, Schild GC (1982) Molecular mechanisms of variation in influenza viruses. *Nature* 296: 115–121

Authors' address: Dr. X. Jiang, Center for Pediatric Research, 855 West Brambleton Avenue, Norfolk, VA23510-1001, U.S.A.

Received April 26, 1999