Arch Virol (2004) 149: 1309–1323
DOI 10.1007/s00705-004-0296-9

Archives of Virology
Printed in Austria

Genetic diversity among sapoviruses

T. Farkas[1,3], **W. M. Zhong**[1], **Y. Jing**[4], **P. W. Huang**[1], **S. M. Espinosa**[1,6], **N. Martinez**[5], **A. L. Morrow**[2,3], **G. M. Ruiz-Palacios**[6], **L. K. Pickering**[1,3,7], and **X. Jiang**[1,3]

[1]Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, U.S.A.
[2]Center for Epidemiology and Biostatistics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, U.S.A.
[3]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, U.S.A.
[4]Eastern Virginia Medical School, Norfolk, Virginia, U.S.A.
[5]Virology Laboratory, Central Hospital, Mendoza, Argentina
[6]Department of Infectious Diseases, Institute of Medical Services and Nutrition, Mexico City, Mexico
[7]National Immunization Program, Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.

Received October 3, 2003; accepted January 13, 2004
Published online March 17, 2004

Summary. *Norovirus* and *Sapovirus* are two genera of the family *Caliciviridae* that contain viruses that can cause acute gastroenteritis in humans. Noroviruses (NOR) are genetically highly diverse but limited studies of the genetic diversity of sapoviruses (SAP) have been reported. In this study we characterized twenty-five SAP detected in our laboratory from outbreaks or sporadic cases of acute gastroenteritis in children from different geographical locations and in adults involved in a cruise ship outbreak investigation and a nursing home outbreak. Based on significant differences of partial RNA polymerase sequences (278–286 nt), the 25 strains were grouped into 12 genetic clusters, including 9 potential new clusters. Extended sequence analysis of the capsid gene of selected strains representing five potential new clusters supported this grouping. Four strains (Hou7-1181/90, Mex340/90, Cruise ship/00 and Argentina39) had <84% amino acid (aa) identity to each other and to the published sequences in the GenBank. Mex14917/00 was almost identical to Stockholm/97/SE whose RNA polymerase sequence was unknown. Phylogenetic and distance analyses of the capsid region of the four new strains showed that Hou7-1181/90 and Argentina39 represent two new genogroups and Mex340/90 and Cruise ship/00 belong to two new clusters within the London/92 genogroup. Thus, based on the capsid sequences we

propose to classify the currently known SAP into nine genetic clusters within five genogroups, including one genogroup that is represented by an animal calicivirus, the porcine enteric calicivirus (PEC).

Introduction

Norovirus and *Sapovirus*, previously known as "Norwalk-like viruses" and "Sapporo-like viruses", respectively, are two genera of *Caliciviridae* [2, 9], that contain viruses that can cause acute gastroenteritis in humans. Virions of sapoviruses (SAP) have typical calicivirus morphology with a "Star of David" appearance, while virions of noroviruses (NOR) have an atypical surface appearance and are referred to as small round structured viruses (SRSV). The NOR genome contains three open reading frames (ORF) encoding the nonstructural proteins (ORF1), the capsid protein (ORF2) and a minor structural protein (ORF3), respectively [7, 13]. In contrast, the SAP capsid gene is fused to the non-structural gene, which is similar to that of the *Rabbit haemorrhagic disease virus* (RHDV) [11, 18, 21]. All known human SAP, except for the London/92-like strains, contain an ORF overlapping with the 5′ end of the capsid gene, which is not found in the NOR genome. The presence of a conserved translation initiation motif GCAAUGG at the 5′ end of this ORF suggests that a functional protein may be encoded in this ORF [29].

NOR and SAP also differ in their epidemiology and host range. NOR infect all ages and commonly cause outbreaks of gastroenteritis frequently associated with contaminated food or water. SAP mainly infect infants and young children [10, 25], although SAP gastroenteritis outbreaks in adults have been described [23]. NOR cause moderate to severe diarrhea. SAP associated diarrhea is milder, but severe cases can occur [26–28].

The prototype Sapporo virus was found in an infant home in Sapporo, Japan in 1977 [3]. Compared with NOR, SAP are less well studied due to lack of proper diagnostic assays. Since the cloning of the prototype Norwalk and Sapporo viruses, at least fifteen NOR genetic clusters in three genogroups have been described [1, 8], while only five genetic clusters in three genogroups (GI–GIII) of SAP, including an animal enteric calicivirus, PEC/Cowden, have been reported [29].

We recently developed a primer pair (p289/p290) that detects both NOR and SAP [15]. Using this primer pair in a cohort study of acute gastroenteritis in children in Mexico, we found that SAP caused approximately 40% of the human calicivirus (HuCV)-associated diarrhea in children less than 2 years of age [4]. Extended studies using this primer pair as well as our updated primer sets of p289/p290 resulted in discovery of new strains of SAP including a new animal enteric calicivirus, the mink enteric calicivirus (MEC) [12]. The current study reports the genetic variation of SAP based on sequence analyses of a 286 bp RNA polymerase region of the genomes of 25 SAP obtained in our laboratory. We also sequenced the capsid genes of five SAP strains and demonstrated that based on capsid sequence analysis, at least nine distinct genetic clusters within five genogroups of SAP exist.

Materials and methods

Stool samples and viruses

Twenty-five SAP sequences were analyzed. Specimens were obtained from our surveillance of HuCV-associated acute gastroenteritis in children less than 5 years of age in Mexico (Mex), Houston (Hou) and Argentina (Arg), from healthy adults during a cruise ship outbreak investigation and from a nursing home outbreak in Virginia (Table 1).

RNA extraction and RT-PCR

Viral RNA was extracted from 10–20% stool specimens using the Trizol reagent (Gibco BRL, Gaithersburg, MD) according to the manufacturer's instructions. The precipitated RNA was dissolved in 20 μL of molecular biology grade water (Eppendorf AG, Hamburg, Germany) and 1–2 μL of the sample was used in the RT-PCR as template. The remainder of the RNA was stored at −70 °C for repeated assays.

Primer pair p289/p290 was used to amplify the RNA polymerase region [15]. A modified degenerate primer set in the same region of the two original primers containing 2 reverse

Table 1. Isolates with available RNA polymerase sequences used in the study

Study of origin	Date of collection	Age of subjects (years)	Strain	Accession number
Mexico cohort I [4]	1989–91	<2	Mex223	AY157854
			Mex240	AY157867
			Mex270	AY157851
			Mex315	AY157870
			Mex318	AY157853
			Mex335	AY157869
			Mex339	AY157865
			Mex340	AF435809
			Mex347	AY157852
Mexico cohort II	1999–2000	<2	Mex11628	AY157859
			Mex11645	AY157860
			Mex11690	AY157864
			Mex11718	AY157866
			Mex11859	AY157857
			Mex11880	AY157856
			Mex14296	AY157858
			Mex14917	AF435810
Houston child care	1989–92	<5	Hou7-1181	AF435811
			Hou7-7114	AY157868
			Hou9-2792	AY157855
			Hou10-4548	AY157861
Argentina [19]	1995–98	<2	Arg39	AF405717
			Arg54	AF405716
Virginia	1998	adult	VA98326	AY157862
Cruise ship	2000	adult	Cruise ship	AY157863

Table 2. RT-PCR primers utilized in the study

Primer	Sense	Location*	Sequence
p290	positive	4354–4376	GATTACTCCAAGTGGGACTCCAC
p290H	positive	4354–4376	GATTACTCCAGGTGGGACTCCAC
p290I	positive	4354–4376	GATTACTCCAGGTGGGACTCAAC
p290J	positive	4354–4376	GATTACTCCAGGTGGGATTCAAC
p290K	positive	4354–4376	GATTACTCCAGGTGGGATTCCAC
p337	positive	4525–4544	CCCTCTGGCATGCCATTCAC
p289	negative	4663–4684	TGACAATGTAATCATCACCATA
p289H	negative	4663–4684	TGACGATTTCATCATCACCATA
p289I	negative	4663–4684	TGACGATTTCATCATCCCCGTA
p302**	positive	5161–5178	ATCGCGGATCCCGGGTGTTTGAGATGGAGGGC
p255***	negative	oligo-dT	AGTAGCCTCGAGCGGCCGCTT-$(T)_{23}$

*numbering given according to positions in the Manchester virus genome (X86560)
**p302 with *Bam*HI, *Sma*I and *Xma*I overhangs (capsid initiation codon is underlined)
***p255 with *Not*I and *Xho*I overhangs

(289H,I) and 4 forward (290H,I,J,K) primers was used in later studies. Both the original primer pair and the new primer set produce a 331 bp product for SAP. To amplify the 3′ genomic sequences containing the capsid region of the viral genome, an oligo-dT primer with primers 302 or 337 was used. Primer 302 was designed based on a conserved region of the 5′ end of capsid sequences of known SAP. Primer 337 was designed based on the internal sequence of the RT-PCR product in the p289/p290 region of Hou7-1181/90 strain. All primer sequences used in this study are listed in Table 2.

To synthesize long cDNAs in the capsid region, SuperScript™ II RNA polymerase was used according to the manufacturer's protocol (Life Technologies, Inc. Gaithersburg, MD). The synthesized first strand cDNAs were then amplified using the TaKaRa Ex Taq™ (TaKaRa Bio Inc., Shiga, Japan) for 35 cycles with denaturation for 30 sec at 94 °C, annealing for 1 min at 49 °C and extension for 2 min and 30 sec at 72 °C.

Sequence analysis

RT-PCR products were cloned into pGEM-T vector (Promega, Madison, WI) according to the manufacturer's protocol. Positive clones were identified by PCR screening. The cloned cDNAs were sequenced using M13 forward and reverse primers by the chain termination method on an automated sequencer (ALFexpress™, Pharmacia, Uppsala, Sweden). To sequence clones with long amplicons, deletion clones for both orientations were generated using the Kilo-Sequencing Deletion Kit (TaKaRa Bio Inc., Shiga, Japan) or primer walking was applied. For each strain at least two clones from two independent RT-PCR reactions were sequenced.

Phylogenetic analysis

The following capsid sequences published in the GenBank were used in the phylogenic analysis: Norovirus: Alphatron (AF195847), Amsterdam (AF195848), BS5 (AF093797), Chiba (AB022679), Desert Shield (U04469), Erfurt (AF427118), Hawaii (U07611), Hillingdon (AJ277607), Jena (AJ011099), Leeds (AJ277608), Lordsdale (X86557), Mexico

(U22498), Melksham (X81879), Musgrove (AJ277614), Norwalk/68 (M87661), Sindlesham (AJ277615), Southampton (L07418), Sw918 (AB074893), Winchester (AJ277609); Sapovirus: Bristol/98 (AJ249939), Houston/86 (U95643), Houston/90 (U95644), London/92 (U95645), Lyon598/97 (AJ271056), Lyon30388/98 (AJ251991), Manchester/93 (X86560), Parkville/94 (U73124), Plymouth/92 (X86559), porcine enteric CV (PEC/Cowden) (AF182760), Sapporo/82 (U65427) and Stockholm/97 (AF194182). Multiple alignments of nucleotide (nt) and amino acid (aa) sequences were created using the Omiga v2.0 software (Oxford Molecular Ltd, Oxford, UK). Aligned sequences were edited in GeneDoc v2.5. [22]. Dendograms based on nt and aa alignments were constructed by the UPGMA clustering method of Molecular Evolutionary Genetics Analysis (MEGA version 2.1) with Jukes-Cantor distance or Poisson correction distance calculations, respectively [17]. The confidence values of the internal nodes were obtained by performing 125 bootstrap analyses. Pairwise distances and mean within (intra-) and between (inter-) genogroups and clusters were also calculated. To compare the data obtained by the MEGA, nucleotide alignments were also analyzed by maximum likelihood algorithm (program DNAML) of the PHYLIP package (v3.52c), with global rearrangement and ten times randomized sequence input order [6]. The GenBank accession numbers for RNA polymerase sequences of strains described in our laboratory are listed in Table 1. The 3′ genomic sequences including the capsid for Mex340/90, Mex14917/00, Hou7-1181/90, Argentina39 and Cruise ship/00 were submitted to the GenBank under accession numbers AF435812, AF435813, AF435814, AY289803 and AY289804, respectively.

Results

Collection of SAP strains from different populations

A cohort study of acute gastroenteritis in children 0–2 years of age was conducted in Mexico City between 1989 and 1991 from which 115 diarrheal and 66 non-diarrheal stool samples were tested for the presence of HuCVs. From these specimens, 6 (5.2%) diarrheal samples and 2 (3%) non-diarrheal samples yielded SAP strains [4]. Of the eight SAP strains, four were closely related to Houston/86, one to London/92 and three belonged to a potentially new cluster (Mex340) within the SAP genus. A second potential new cluster (Mex339) was identified later from the same cohort.

A second cohort study was undertaken of children 0–2 years of age in the same neighborhood of Mexico City, between 1999 and 2000. A total of 254 diarrheal stool samples were tested and 8 (3.1%) SAP were identified. Of the 8 SAP strains, two had high nt identity with Houston/86 (87% and 94%), one with Houston/90 (94%) and the remaining five revealed unique sequences (<80%) distinct from known SAP.

We also studied 44 gastroenteritis outbreaks occurring in child care centers in Houston, Texas between 1989–1991 by testing 5 randomly selected stool samples from each outbreak [5]. Prototype Houston/86 and Houston/90 SAP were identified from the same child care centers earlier [14, 20]. SAP sequences were obtained from 4 outbreaks (9%). Each of three strains had high nt identity (94–98%) with London/92, Houston/86 and Houston/90, respectively, and the fourth strain (Hou7-1181) revealed 96% nt homology to one of the unique strain (Mex11718) identified from the second cohort in Mexico.

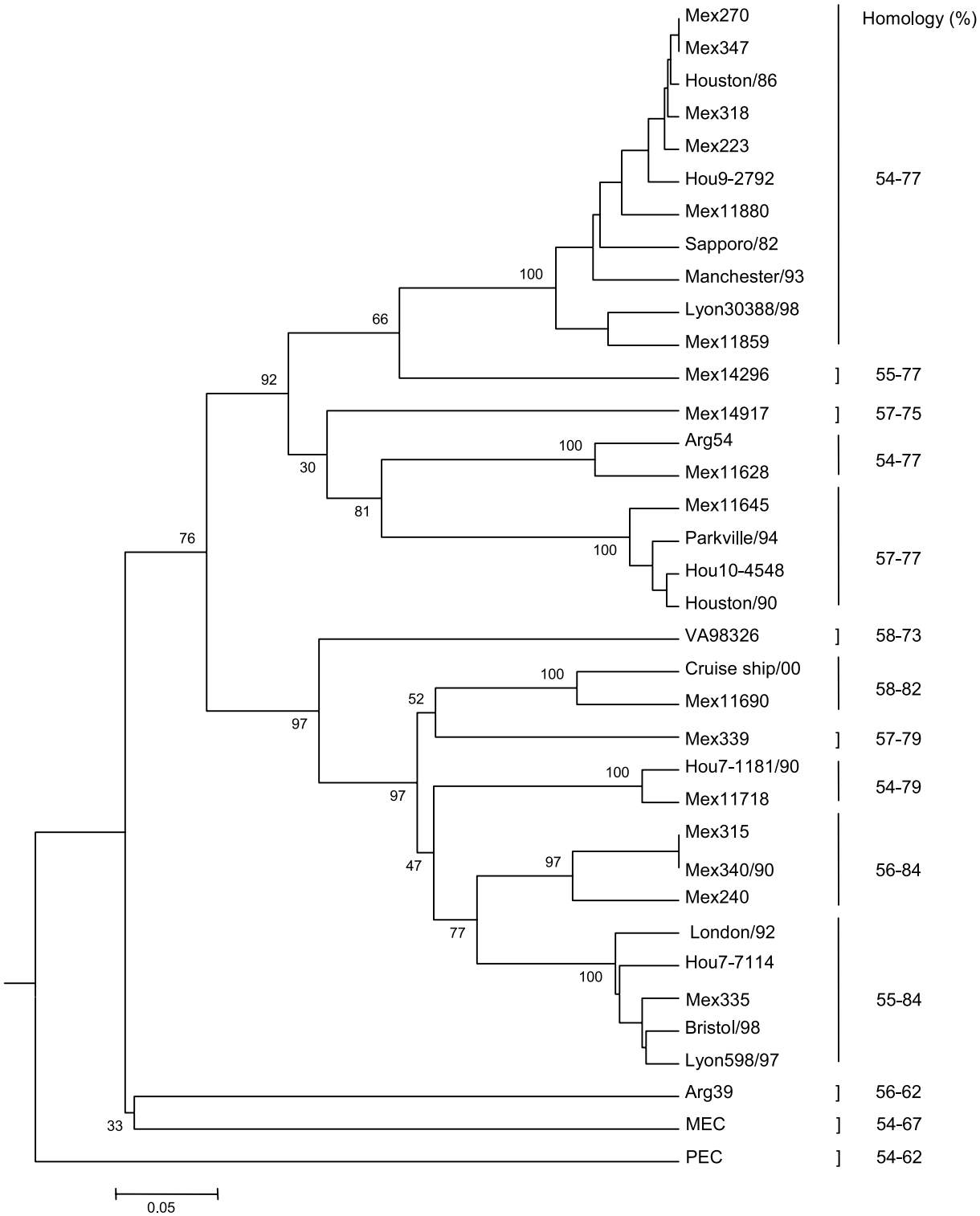

Fig. 1. Phylogenetic tree based on partial RNA polymerase gene nucleic acid sequences of selected SAP. The tree was constructed by the UPGMA clustering method (MEGA v2.1). Scale bar represents the phylogenetic distances expressed as units of expected nt substitutions per site. Bootstrap values are indicated as % of 125 replicates. Inter-cluster homology range is shown as %

Four additional unique SAP sequences were included in the study. Two (Arg39 and Arg54) were identified in our surveillance of acute gastroenteritis in hospitalized children in Argentina, [19] one (Cruise ship/00) was isolated from a healthy adult during a cruise ship outbreak investigation in 2000, and one (VA98326) originated from an outbreak of acute gastroenteritis in a nursing home in Virginia in 1998.

Genetic variation of SAP based on the RNA polymerase sequences

The RNA polymerase sequences of the 25 strains were analyzed based on cloned RT-PCR products by primer pair 289/290 and derivatives. These amplicons contain 286 bp virus specific sequences. The two strains derived from Argentina were amplified by different primers that resulted in 278 bp virus specific sequences [19]. The same region of the RNA polymerase of the PEC and MEC animal SAP also were included in the analysis [11, 12]. All strains contained the GLPSG motif when translated in the second frame. According to pairwise alignment scores of the sequences, 14 distinct clusters, of which 9 represented possible new clusters of human origin, were identified at a level of $\leq$84% nt identity with known cluster representative strains or strains from other groups included in the alignment (Fig. 1). Although the calculated mean inter-cluster distances obtained by MEGA2.1 or maximum likelihood analysis for the new clusters were less (except for Arg39) than those between established SAP clusters, they correlated well with mean distances among established NOR clusters in the same RNA polymerase region (data not shown).

Cloning and sequencing the capsid cDNA of five SAP

To further characterize the genetic relationships among SAP, we amplified and sequenced the 3′ end of the genome (containing the capsid gene) of five strains, representing 5 of the 9 candidate new clusters. Amplification of strains representing the remaining new clusters were either unsuccessful or is ongoing.

Mex14917/00, Mex340/90, Arg39 and Cruise ship/00 were amplified by oligo-dT and p302 and the resulting cDNAs contained 2272 bp, 2291 bp, 2290 bp and 2275 bp viral sequences without the poly-A tail, respectively. Hou7-1181/90 was amplified by oligo-dT and p337 and the cDNA contained a 3048 bp viral sequence. In addition to the capsid gene, all five cDNAs contained the small ORF2 sequence

Table 3. The length of ORFs (aa) and 3′UTR (nt) of the five newly described SAP

Strain	Capsid	3′ORF	Capsid overlap	3′UTR
Mex14917/00	566	165	162	77
Mex340/90	558	166	None	108
Cruise ship/00	559	166	None	98
Hou7-1181/90	553	167	161	93
Arg39	569	166	155	86

Table 4. Nucleotide and amino acid [ ] sequence identity (%) among sapoviruses in the capsid region

Genogroup/ Cluster	Strain	GI/1					GI/2		GI/3		GII/1			GII/2 13	GII/3 14	GIII 15	GIV 16
		1	2	3	4	5	6	7	8	9	10	11	12				
GI/1	1. Sapporo/82	–															
	2. Houston/86	94[97]	–														
	3. Manchester/93	94[99]	93[98]	–													
	4. Plymouth/92	94[98]	93[97]	94[99]	–												
	5. Lyon30388/98	93[98]	92[97]	93[99]	96[99]	–											
GI/2	6. Houston/90	72[77]	72[77]	72[78]	72[78]	73[78]	–										
	7. Parkville/94	73[79]	73[77]	73[79]	73[79]	73[79]	97[97]	–									
GI/3	8. Stockholm/97	73[74]	73[71]	73[73]	73[73]	73[73]	72[72]	73[73]	–								
	9. Mex14917/00	73[80]	73[77]	73[78]	73[78]	73[78]	72[77]	73[78]	97[91]	–							
GII/1	10. London/92	54[44]	53[43]	54[44]	54[44]	54[43]	52[42]	52[42]	51[40]	52[41]	–						
	11. Bristol//98	55[46]	54[45]	54[45]	55[46]	55[45]	52[45]	52[45]	52[43]	52[45]	94[92]	–					
	12. Lyon598/97	55[45]	54[44]	54[45]	55[45]	54[45]	52[44]	52[44]	52[43]	52[45]	93[92]	96[99]	–				
GII/2	13. Mex340/90	54[46]	53[45]	54[46]	54[46]	54[46]	52[44]	52[44]	53[46]	53[46]	73[79]	74[84]	73[83]	–			
GII/3	14. Cruise ship/00	54[44]	53[43]	53[44]	54[45]	54[44]	54[44]	53[44]	53[41]	53[44]	72[75]	73[81]	73[81]	72[82]	–		
GIII	15. PEC	49[40]	50[39]	50[40]	51[40]	51[40]	50[39]	49[40]	49[37]	50[40]	50[37]	50[39]	50[39]	50[37]	50[38]	–	
GIV	16. Hou7-1181/90	56[48]	56[47]	56[48]	56[48]	56[49]	56[46]	57[47]	58[46]	58[47]	52[39]	52[42]	52[42]	53[44]	52[43]	50[37]	–
GV	17. Arg39	56[49]	57[48]	57[49]	57[49]	57[49]	55[48]	56[49]	56[48]	56[50]	54[43]	55[47]	55[47]	54[48]	55[49]	52[41]	57[47]

and a short non-translated region upstream of the poly-A tail. A third ORF, overlapping with the 5′ region of the capsid gene, was identified in Mex14917/00, Hou7-1181/90 and Arg39, but not in Mex340/90 and Cruise ship/00 (Table 3). This ORF contained the **G**CA**AUG**G sequence motif at the 5′ end, which is favored for translation initiation in mammalian cells, although the −3 purine was part of the amplification primer in the case of Mex14917/00 and Arg39 (p302). This motif is conserved among GI SAP strains [29].

The capsid sequence of Mex14917/00 had 97% nt and 91% aa identity with Stockholm/97 (Table 4) whose RNA polymerase sequence is not available from public databases [30]. The capsid alignments showed that there were 3 potentially missing nt in the Stockholm/97 capsid at nt 1040, 1144 and 1172 compared with Mex14917/00 (data not shown), resulting in an abrupt reduction in aa identity in this short region of Stockholm/97 with known SAP. When the 3 nt were inserted into the Stockholm/97 sequence at the corresponding positions, the alignment disruptions were eliminated and the aa identity of Stockholm/97 increased to 98% with Mex14917/00. Based on this finding we hypothesize that the published Stockholm/97 sequence might contain errors.

The other four capsid sequences showed significant variation compared with known SAP. Mex340/90 and Cruise ship/00 had the highest identity with Bristol/98 [74/84% and 73/81% (nt/aa), respectively] and the two strains shared a 72/82% (nt/aa) homology to each other. Hou7-1181/90 and Arg39 had less than 58% nt and 50% aa identities with any of the published SAP capsid sequences (Table 4).

Determination of the genetic relationships of SAP based on capsid sequences

Phylogenetic analysis of the capsid sequences and the calculated distances obtained by either the UPGMA or DNAML methods based on either nt or aa alignments suggested that SAP could be divided into five genogroups and nine genetic clusters, represented by 16 human and 1 animal SAP (Fig. 2).

NOR genetic cluster representatives (GI/1-7, GII/1-10 and GIII/1) were included in the alignment and distance calculation (data not shown). Phylogenetic distances based on amino acid alignments showed a range of 0.60 to 0.90 or 0.56 to 0.79 between Hou7-1181/90 or Arg39 and any of the other SAP, respectively. These distances, in both cases, were significantly longer than any of the inter-cluster distances (0.19–0.27) between the previously described 3 clusters of GI SAP and correlated well with the distances between established SAP (0.67–0.94) or NOR (0.63–0.76) genogroups. Therefore, we assigned Hou7-1181/90 and Arg39 as the representatives of two new genogroups of SAP.

To investigate whether Mex340/90 and Cruise ship/00 represent new genetic clusters, we compared the intra- and inter-cluster distances between the 17 published SAP capsid sequences. Intra-cluster distances ranged between 0.00 and 0.06 or with omitting the possibly incorrect Stockholm/97 sequence between 0.00 and 0.05. The inter-cluster distances between Mex340/90 or Cruise ship/00 and the 3 strains of the London/92 cluster (0.13–0.16 or 0.15–0.19, respectively) were

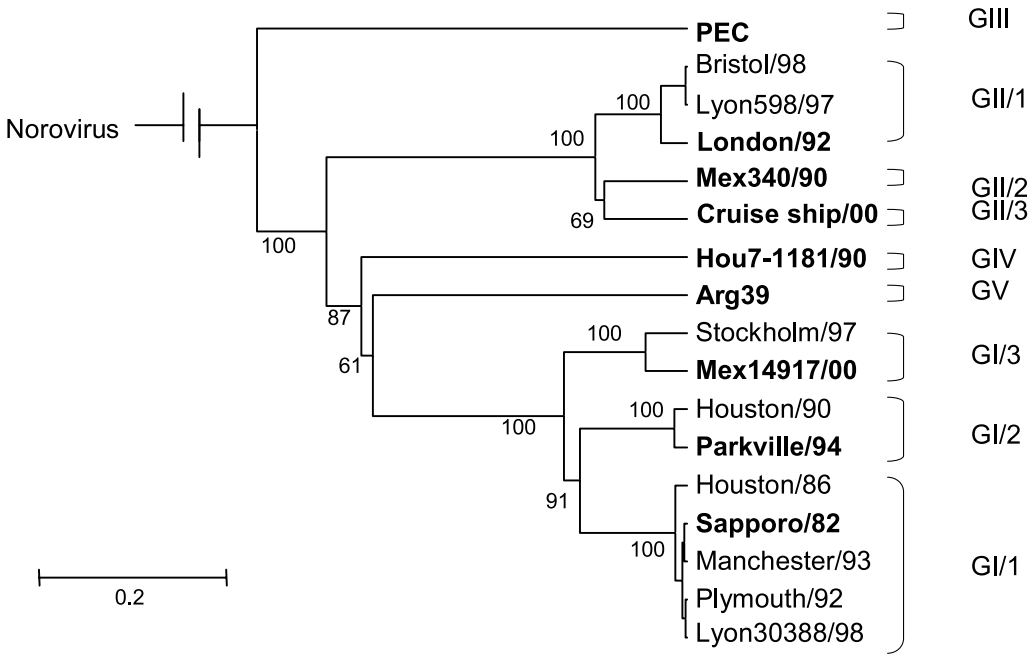

Fig. 2. Phylogenetic tree based on available SAP capsid gene amino acid sequences. The tree was constructed by the UPGMA clustering method (MEGA v2.1). Scale bar represents the phylogenetic distances expressed as units of expected aa substitutions per site. Bootstrap values are indicated as % of 125 replicates. Genogroups and genetic clusters are indicated. Prototype and candidate prototype strains for each cluster are highlighted in bold

shorter than or equal to those among the 9 strains of the 3 GI SAP clusters (0.19 and 0.27), but they were significantly longer than any of the intra-cluster distances (0.00–0.05). These distances were also comparable with distances between established NOR cluster representatives, as the distance between BS5 and Musgrove was 0.17 and between Melksham and Hillingdon was 0.19. Phylogenetic distances and trees based on nt sequence alignments placed Mex340/90 or Cruise ship/00 slightly further from the London/92 cluster than those based on the aa alignments. Furthermore, distances based on nt alignments between Mex340/90 or Cruise ship/00 and the London/92-like viruses (0.30–0.32) correlated well with the distances between established NOR cluster representatives, as the distance between Melksham and Hillingdon was 0.31 (data not shown). These data supports the assignment of Mex340/90 and Cruise ship/00 as representatives of 2 new genetic clusters within genogroup II of SAP.

Discussion

In this report we have extended the genetic variation of SAP by analyzing partial RNA polymerase sequences obtained in our laboratory followed by characterization of the capsid sequences of five strains. Pairwise nt sequence and phyloge-

netic distance analysis of partial RNA polymerase sequences provided useful information for identification of potential new genetic clusters of SAP. However, the calculated inter-cluster distances based on established genetic clusters of SAP were more stringent in this region, compared to that of NOR. Among the 25 SAP obtained in our laboratory, 12 distant genetic clusters (9 were new) were identified and the extended capsid sequence analysis of five strains, representing five of the 9 new clusters, supported this result. Phylogenetic trees and inter- and intra-group (both for genetic clusters and genogroups) distance comparisons of capsid sequences allowed us to identify two new genetic clusters as well as two new genogroups of SAP.

Mex14917/00 grouped with Stockholm/97. The alignments of the two strains revealed possible errors in the published Stockholm/97 sequence. There are three missing nucleotides causing 3 frameshifts in a 132 bp region of the Stockholm/97 capsid sequence which has a significant effect on the 44 deduced amino acids encoded by this region. Thus Mex14917/00 sequences are more likely to accurately represent this cluster as a reference strain.

Hou7-1181/90 and Arg39 clearly segregated from the previously described SAP genogroups and from each other (Fig. 2). Analysis of the intra-cluster and inter-genogroup distances placed these strains as representatives of two new genogroups of SAP. Schuffenecker et al., recently classified SAP into 3 genogroups: GI (Sapporo/82-like viruses), GII (London/92-like viruses) and GIII (PEC/Cowden) [29]. Okada et al. [24] by the analysis of 130 aa N-terminal capsid sequences of several Japanese SAP isolates, described a possible fourth genogroup, represented by Chiba/00067T/1999, which had 96% aa identity with Hou7-1181/90. According to these classification schemes, we assign Hou7-1181/90 and Arg39 as representatives of GIV and GV SAP, respectively. Sapporo/82-like viruses (GI), Hou7-1181/90 (GIV) and Arg39 (GV) possess the overlapping N-terminal ORF of the capsid sequences, while London/92-like strains (GII) lack this ORF. Hou7-1181/90 and Arg39 are also more related to GI than GII strains according to phylogenetic distance analysis. The distance between Hou71181/90 and the 9 GI strains or the 5 GII strains analyzed is between 0.60 and 0.68 or 0.74 and 0.81, respectively. The distance between Arg39 and the 9 GI strains or the 5 GII strains analyzed is between 0.56 and 0.63 or 0.64 and 0.72, respectively. Arg39 grouped significantly further from other SAP by analyzing the partial RdRp region compared to the whole capsid sequence analysis, indicating that it may be a recombinant virus [16]. However, since we did not amplify the RdRp and the capsid region in a continuous RT-PCR product, we can not exclude the possibility that the two sequences obtained for the two regions are from two different strains.

Mex340/90 and Cruise ship/00 grouped with London/92, but with a distance of 0.16 and 0.19, respectively, comparable to that between Mex14917 (GI/3) and members of the Sapporo (GI/1) or Parkville (GI/2) clusters (0.19–0.21). Including the newly described Bristol/98 and Lyon598/97 capsid sequences in the analysis shortened this distance (0.13), however it was still almost 3 times more than the highest intra-cluster distance (0.05) among SAP and correlated

Table 5. Classification scheme of sapoviruses

Genogroup/cluster	Prototype strain	Other strains
GI/1	Sapporo/82	Houston/86 Plymouth/92 Manchester/93 Lyon30388/98
GI/2	Parkville/94	Houston/90
GI/3	Stockholm/97	Mexico14917/00
GII/1	London/92	Lyon598/97 Bristol/98
GII/2	Mexico340/90	
GII/3	Cruise ship/00	
GIII	PEC/Cowden	
GIV	Houston7-1181/90	
GV	Argentina39	

with distances among NOR cluster representatives. Considering these we assigned Mex340/90 and Cruise ship/00 as representatives of two new clusters within GII of SAP.

Schuffenecker et al. reported the distinct differentiation of SAP sequences into the same genetic groups and genetic clusters regardless what genomic regions examined [29]. We also analyzed the phylogenetic relationship based on the 3′ ORF sequences (except for Stockholm/97) and found similar clustering to that of the capsid analysis (data not shown).

In summary, we added 5 new SAP capsid sequences to the previously published 12 [24, 29]. Based on distance and phylogenetic analysis of the 17 SAP capsid sequences, SAP could be classified into 9 genetic clusters within five genogroups. Genogroups I, II, IV and V consist of human strains and genogroup III consists of an animal strain (Table 5).

We demonstrated that new members of SAP still could be found, and SAP are genetically more diverse than previously thought. SAP mainly cause illness in children, while NOR cause disease in all age groups. One previous hypothesis to explain this difference is that SAP are genetically less diverse than NOR and most individuals are likely to be exposed to all circulating types of SAP during childhood and probably develop life time immunity. Our data suggest that SAP may be as diverse as NOR and the introduction of a new antigenically different strain commonly occurs in a community. Therefore, alternative hypotheses need to be developed to explain this difference between SAP and NOR.

Our previous surveillance studies demonstrated that multiple genetic clusters of SAP could be found in a single community, which is similar to that found for

NOR. This is in contrast to the literature where only limited genetic diversity of SAP has been described in most reports from different countries. One possible explanation is that the primers used in our studies were designed for detection of both NOR and SAP. Using this primer pair, we previously demonstrated that the ratio between NOR and SAP was approximately 60% to 40% of HuCV associated diarrhea, respectively, in children under 2 years of age in a Mexico cohort study [4]. This primer pair also resulted in detection of the mink enteric calicivirus (MEC), an animal calicivirus related to SAP [12], which also shows that these primers are not only able to detect both NOR and SAP, but also are able to detect the widest diversity of strains. Recently we developed a new generation of primers in the same region but with higher sensitivity. We believe with the continuation of this work, we will have a better understanding of the genetic variation of SAP as well as the genetic and antigenic classification in the near future.

Acknowledgements

This study was supported by the National Institutes of Health of the United States (RO1 AI37093 and PO1 HD13021).

References

1. Ando T, Noel JS, Fankhauser RL (2000) Genetic classification of Norwalk-like viruses. J Infect Dis 181 [Suppl 2]: S336–S348
2. Berke T, Matson DO (2000) Reclassification of the Caliciviridae into distinct genera and exclusion of hepatitis E virus from the family on the basis of comparative phylogenetic analysis. Arch Virol 145: 1421–1436
3. Chiba S, Sakuma Y, Kogasaka R, Akihara M, Horino K, Nakao T, Fukui S (1979) An outbreak of gastroenteritis associated with calicivirus in an infant home. J Med Virol 4: 249–254
4. Farkas T, Jiang X, Guerrero ML, Zhong W, Wilton N, Berke T, Matson DO, Pickering LK, Ruiz-Palacios G (2000) Prevalence and genetic diversity of human caliciviruses (HuCVs) in Mexican children. J Med Virol 62: 217–223
5. Farkas T, Pennington D, Jiang X, Pickering LK (2001) Human calicivirus (HuCV) gastroenteritis (GE) outbreaks in child care centers (CCC) in Houston in 1989–1991. Am Society for Virology, 20th Annual Meeting, University of Wisconsin, Madison, WI, July 21–25, W47–W48
6. Felsenstein J (1993) PHYLIP (Phylogeny Interference Package). Department of Genetics, University of Washington, Seattle
7. Glass PJ, White LJ, Ball JM, Leparc-Goffart I, Hardy ME, Estes MK (2000) Norwalk virus open reading frame 3 encodes a minor structural protein. J Virol 74: 6581–6591
8. Green J, Vinje J, Gallimore CI, Koopmans M, Hale A, Brown DW (2000) Capsid protein diversity among Norwalk-like viruses. Virus Genes 20: 227–236
9. Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, Matson DO, Nakata S, Neill JD, Studdert MJ, Thiel HJ (2000) Taxonomy of the caliciviruses. J Infect Dis 181 [Suppl 2]: S322–S330
10. Green KYCR, Kapikian AZ (2001) Human Caliciviruses. In: Knipe DM, Howley PM, Griffin DE et al. (eds) Fields Virology, 4th ed, vol 1. Lippincott & Wilkins, Philadelphia, pp 841–874

11. Guo M, Chang KO, Hardy ME, Zhang Q, Parwani AV, Saif LJ (1999) Molecular characterization of a porcine enteric calicivirus genetically related to Sapporo-like human caliciviruses. J Virol 73: 9625–9631
12. Guo M, Evermann JF, Saif LJ (2001) Detection and molecular characterization of cultivable caliciviruses from clinically normal mink and enteric caliciviruses associated with diarrhea in mink. Arch Virol 146: 479–493
13. Jiang X, Graham D, Wang K, Estes M (1990) Norwalk virus genome cloning and characterization. Science 250: 1580–1583
14. Jiang X, Cubitt WD, Berke T, Zhong W, Dai X, Nakata S, Pickering LK, Matson DO (1997) Sapporo-like human caliciviruses are genetically and antigenically diverse. Arch Virol 142: 1813–1827
15. Jiang X, Huang PW, Zhong WM, Farkas T, Cubitt DW, Matson DO (1999) Design and evaluation of a primer pair that detects both Norwalk- and Sapporo-like caliciviruses by RT-PCR. J Virol Methods 83: 145–154
16. Jiang X, Espul C, Zhong WM, Cuello H, Matson DO (1999) Characterization of a novel human calicivirus that may be a naturally occurring recombinant. Arch Virol 144: 2377–2387
17. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA
18. Lambden PR, Caul EO, Ashley CR, Clarke IN (1994) Human enteric caliciviruses are genetically distinct from small round structured viruses. Lancet 343: 666–667
19. Martinez N, Espul C, Cuello H, Zhong W, Jiang X, Matson DO (2002) Sequence diversity of human caliciviruses recovered from children with diarrhea in Mendoza, Argentina, 1995–1998. J Med Virol 67: 289–298
20. Matson DO, Estes MK, Glass RI, Bartlett AV, Penaranda M, Calomeni E, Tanaka T, Nakata S, Chiba S (1989) Human calicivirus-associated diarrhea in children attending day care centers. J Infect Dis 159: 71–78
21. Meyers G, Wirblich C, Thiel HJ (1991) Rabbit hemorrhagic disease virus-molecular cloning and nucleotide sequencing of a calicivirus genome. Virology 184: 664–676
22. Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: Analysis and visualization of genetic variation. EMBNEW NEWS 4: 14
23. Noel JS, Liu BL, Humphrey CD, Rodriguez EM, Lambden PR, Clarke IN, Dwyer DM, Ando T, Glass RI, Monroe SS (1997) Parkville virus: a novel genetic variant of human calicivirus in the Sapporo virus clade, associated with an outbreak of gastroenteritis in adults. J Med Virol 52: 173–178
24. Okada M, Shinozaki K, Ogawa T, Kaiho I (2002) Molecular epidemiology and phylogenetic analysis of Sapporo-like viruses. Arch Virol 147: 1445–1451
25. Pang XL, Honma S, Nakata S, Vesikari T (2000) Human caliciviruses in acute gastroenteritis of young children in the community. J Infect Dis 181 [Suppl 2]: S288–S294
26. Pang XL, Zeng SQ, Honma S, Nakata S, Vesikari T (2001) Effect of rotavirus vaccine on Sapporo virus gastroenteritis in Finnish infants. Pediatr Infect Dis J 20: 295–300
27. Robinson S, Clarke IN, Vipond IB, Caul EO, Lambden PR (2002) Epidemiology of human Sapporo-like caliciviruses in the South West of England: Molecular characterisation of a genetically distinct isolate. J Med Virol 67: 282–288
28. Sakai Y, Nakata S, Honma S, Tatsumi M, Numata-Kinoshita K, Chiba S (2001) Clinical severity of Norwalk virus and Sapporo virus gastroenteritis in children in Hokkaido, Japan. Pediatr Infect Dis J 20: 849–853

29. Schuffenecker I, Ando T, Thouvenot D, Lina B, Aymard M (2001) Genetic classification of "Sapporo-like viruses". Arch Virol 146: 2115–2132
30. Vinje J, Deijl H, van der Heide R, Lewis D, Hedlund KO, Svensson L, Koopmans MP (2000) Molecular detection and epidemiology of Sapporo-like viruses. J Clin Microbiol 38: 530–536

Author's address: Tibor Farkas, D.V.M., Ph.D., Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, U.S.A., e-mail: Tibor.Farkas@cchmc.org