Arch Virol (2001) 146: 2115–2132

Archives of Virology

© Springer-Verlag 2001
Printed in Austria

Genetic classification of "Sapporo-like viruses"

I. Schuffenecker[1], **T. Ando**[2], **D. Thouvenot**[1], **B. Lina**[1], and **M. Aymard**[1]

[1]Laboratoire de Virologie, Centre National de Référence pour les Entérovirus, Université Claude Bernard Lyon I, Lyon, France
[2]Viral Gastroenteritis Section, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.

Accepted June 25, 2001

Summary. "Sapporo-like viruses" (SLVs) and "Norwalk-like viruses" (NLVs) are an important cause of acute gastroenteritis in humans. While NLVs have been genetically classified into three major genetic groups consisting of 17 genetic subgroups, a classification of SLVs into comparable genetic groups remains to be determined. In an attempt to classify both SLVs and NLVs uniformly, the sequences of 2 SLV strains newly detected from French infants were analysed together with the published sequences of 9 SLV and 19 NLV strains. Distance and phylogenetic analyses were conducted on the sequences of the capsid gene, RNA polymerase gene, 3′ open reading frame (3′ORF), ORF overlapping the capsid gene, and 3′ untranslated region (3′UTR). The histogram showing frequency distribution of pairwise distances and the topology of the phylogenetic tree demonstrated that SLVs and NLVs could be classified uniformly on the basis of the entire capsid sequences and that the 11 SLV strains could be genetically classified into 3 major genetic groups, genogroups I, II and III, comprised of 5 genetic subgroups. The differentiation of the 11 SLV strains into these genetic groups was also maintained in the 4 remaining genome regions, while the sequences at the junction between the RNA polymerase and capsid genes were shown to be genogroup-specific.

Introduction

The members of the genera "Sapporo-like viruses" (SLVs) and "Norwalk-like viruses" (NLVs) in the family *Caliciviridae* [14, 32] are an important cause of gastroenteritis in humans [4, 8, 10, 12, 21, 28, 30, 39, 40, 43] and animals [7, 15, 26, 38]. In humans, the illnesses caused by SLVs and NLVs differ in epidemiological features and clinical symptoms. While NLVs are the major cause of outbreaks of nonbacterial gastroenteritis in all age groups, SLVs have been primarily associated

with paediatric gastroenteritis in disease-endemic settings [21]. In children under 2 years old, NLVs cause moderately severe disease with vomiting as predominant symptom, whereas SLVs cause mild diarrhoea [31].

The genomic organisation of SLVs is distinct from that of NLVs [24, 29]. NLVs possess three different open reading frames (ORFs), consisting of a 5′ ORF encoding for the non-structural proteins, including the RNA polymerase, a second ORF encoding for the major capsid protein and a 3′ ORF encoding for a basic minor structural protein [5, 6, 11]. In SLVs, the polyprotein gene is fused to and in frame with the capsid coding region [24, 29]. In addition, in all published SLV strains except a human strain, London/92 and a porcine strain, PEC/Cowden [5, 15, 18, 27, 29, 40], an additional ORF has been predicted in +1 frame, overlapping the N terminus of the capsid gene (capsid overlap). To date, the sequence of a ∼3-kb region extending from the RNA polymerase gene to the 3′ poly(A) tail is available in GenBank for 9 NLV strains, of which 6 have had their entire genome sequenced: Norwalk/68 [17], Southampton/91 [22], Lordsdale/93 [9], Camberwell/94 [3, 36], Chiba/87 (GenBank accession number No. AB042808) and Jena/80 [26]. Until this study, the sequence of the 3-kb region was available in GenBank for 8 SLV strains, of which 2 have had their entire genome sequenced: Manchester/93 [24, 25] and PEC/Cowden [15] (Table 1).

Recently, Ando et al. [1] have proposed a genetic classification of human NLVs into 2 major genetic groups, genogroups GI and GII, consisting of 5 and 10 genetic clusters, respectively. In addition, they described a third genogroup, including 2 bovine strains, the Newbury Agent-2 (NA-2) and Jena/80 (JV). Their genetic classification scheme of NLVs, based on the analysis of the partial capsid sequences of 101 strains, was supported by the analysis of the entire capsid sequences of 38 strains, representing the genetic diversity of the whole panel. The results of their analysis showed good concordance with the genetic classification scheme reported by Green et al. [13], based on the analysis of the entire capsid sequences of 35 NLV strains.

Human SLVs have also been reported as genetically diverse and classified into 4 genetic clusters, each of which is represented by Manchester/93 (and/or Sapporo/82), Houston/90 (and/or Parkville/94), Stockholm/97 and London/92, respectively [2, 18, 27, 40]. In addition, a porcine enteric calicivirus (PEC) has recently been shown to be related to the SLV genus and to belong to a fifth genetic cluster [15]. To date, human SLVs, within the family *Caliciviridae*, have often been considered to belong to a single genogroup, GIII, comparable with the two genogroups of NLVs, GI and GII [13, 29, 40, 41]. However, Jiang et al. have suggested the possible existence of two genogroups among human SLVs [18]. In addition, the phylogenetic relationship between the genetic cluster represented by the porcine strain PEC/Cowden and the 4 genetic clusters of human SLVs needs clarification.

The aim of the present study was to provide a genetic classification of SLV strains, using the same criteria and methodology as Ando et al. [1] for their classification of NLVs. We sequenced 3-kb of the 3′ half of the genome of 2 SLV strains detected in French infants and analysed these newly determined sequences,

Table 1. SLV and NLV strains analysed

Genus	Genogroup	Genetic cluster	Strain	GenBank accession number	Length of sequence before alignment				
					Pol (aa)[a]	Capsid (aa)[b]	3′ORF (aa)[b]	Capsid overlap (aa)[b]	3′UTR (nt)
SLVs	GI	1	Sapporo/82	U65427	300	561	165	161	81
	GI	1	Houston/86	U95643	292	561	165	161	80
	GI	1	Plymouth/92	X86559	209	561	165	161	82
	GI	1	Manchester/93	X86560	>500	561	165	161	82
	GI	1	**Lyon/30388/98**	**AJ251991**	**266**	**561**	**165**	**161**	**52[c]**
	GI	2	Houston/90	U95644	292	571	163	163	85
	GI	2	Parkville/94	U73124	264	571	163	163	89
	GI	3	Stockholm/97	AF194182	–[d]	565	–	162	–
	GII	1	London/92	U95645	292	557	166[e]	No[f]	130[e]
	GII	1	**Lyon/598/97**	**AJ271056**	**259**	**560**	**165**	**No**	**117[c]**
	GIII	1	PEC/Cowden	AF182760	>500	544	164	No	55
NLVs	GI	1	Norwalk/68	M87661	>500	530	212	No	66
	GI	1	KY/89	L23828	288	530	–	No	–
	GI	2	Southampton/91	L07418	>500	546	211	No	78
	GI	3	Desert Shield/90	U04469 AB022679	271	544	–	No	–
	GI	4	Chiba 407/87	D38547	–	544	208	No	91
	GII	1	Hawaii/71	U07611	267	535	–	No	–
	GII	2	Snow Mountain/76	L23831	–	542	–	No	–
	GII	2	Melksham/89	X81879	333	542	259	No	45
	GII	3	OTH 25/89	L23830	289	548	–	No	–
	GII	3	MX/89	U22498	289	548	254	No	39
	GII	3	Toronto/91	U02030	289	548	–	No	–
	GII	3	Carlton/42530/95	AF106021	–	–	254	No	37
	GII	3	Carlton/46444/95	AF106022	–	–	254	No	37
	GII	3	Mt Keira/96	AF106023	–	–	254	No	37
	GII	4	Bristol/93	X76716	475	539	268	No	45
	GII	4	Lordsdale/93	X86557	>500	539	267	No	45
	GII	4	Camberwell/94	AF145896	>500	539	268	No	45
	GII	unknown	G. Valley/84	AF106024	–	–	244	No	44
	GIII	1	Jena/80	AJ011099	>500	519	223	No	67

The SLV strains sequenced in this study are highlighted in bold. Other SLV and NLV sequences were available in GenBank. Genogroups and genetic clusters of SLVs were determined in this study; those of NLVs refer to results of previous [35, 1] and the current studies

[a] Length of the C-terminal portion available in GenBank. >500 refers to strains entirely sequenced

[b] Length of the entire predicted protein

[c] Sequence of the 3′ UTR was not entirely determined (see more details in the text)

[d] Sequence not available in GenBank

[e] Length of the revised sequence (see more details in the text)

[f] Predicted open reading frame does not exist

together with the published sequences of 9 SLV and 19 NLV strains. Distance and phylogenetic analyses were conducted on the sequences of the capsid, RNA polymerase, 3′ ORF, capsid overlap (for SLVs only), 3′UTR, and junction of the RNA polymerase and capsid protein, in an attempt to classify uniformly SLVs and NLVs and to compare clustering of SLV sequences in distinct regions of their genomes.

Materials and methods

Detection and sequencing of Lyon/598/97 and Lyon/30388/98 SLV strains

Virus sources

Stool samples, collected in the Lyon area (France) between July 1997 and June 1999, from 28 paediatric outpatients involved in gastroenteritis outbreaks and 178 infants hospitalized for gastroenteritis, were screened for SLVs by reverse transcription-polymerase chain reaction (RT-PCR) directed against a 307 base region in the RNA polymerase gene. Five samples were found positive for SLVs. Two of the 5 detected strains, Lyon/598/97 and Lyon/30388/98, which were identified in a 11 month-old male infant involved in a day-care centre outbreak and in a 2 month-old male hospitalized for sporadic gastroenteritis, respectively, were sequenced in the 3′ half of their genome.

SLV genome detection

RNA was extracted from 100 µl of 10% stool suspension using TRI-REAGENT (Sigma) and amplified by RT-PCR using SR33 (5′ TGTCACGATCTCATCATCACC 3′) and SR80 (5′ TGGGATTCTACACAAAACCC 3′) primers [27], according to the following procedure. The RT step was carried out in 20 µl of AMV-RT buffer (Promega) containing 20 pmol of SR33, 10 U of AMV-RT (Promega), 30 U of RNAsine (Promega) and 1 mM of dNTPs (Eurogentec). The PCR step was carried out in a final volume of 100 µl by mixing with 80 µl of PCR buffer (Promega) containing 30 pmol of SR33 and SR80, 4 mM of $MgCl_2$, 200 mM of dNTPs and 5 U of Taq polymerase (Promega). The thermocycle format was as follows: one cycle of RT at 55 °C for 1 h; 40 amplification cycles with denaturation for 30 sec at 94 °C, annealing for 30 sec at 42 °C and extension for 30 sec at 72 °C.

Amplification and sequencing of a 3-kb region at the 3′ end of the genome

We amplified overlapping fragments of 441–623 nucleotides, which covered a 3-kb stretch extending from within the RNA polymerase to within the 3′UTR region (Fig. 1a). The primers were designed on the basis of the sequence of London/92 and Plymouth/92 for Lyon/598/97 and Lyon/30388/98, respectively. The RT-PCR procedure was the same as described above, except for the annealing temperatures which were adjusted to the melting temperature of primers. The RT-PCR products were purified (High pure PCR product purification kit, Roche Diagnostics) and sequenced, without cloning, on an automated sequencer (ABI 377™, Applied Biosystems) using the Taq dye-deoxy terminator chemistry (Applied Biosystems). All RT-PCR products were sequenced on both strands. Each region was sequenced twice from 2 amplification products prepared by independent RT-PCRs.

Sequence analysis

Wisconsin Package version 10.0 (Genetics Computer Group, USA) was used to generate alignments, calculate pairwise sequence distances and reconstruct phylogenetic trees. Initial alignment of amino acid sequences was made by using the PileUp program with the following settings: gap weight = 4–8 and gap length weight = 1–2. The final alignment was made by using the same program with the default setting (gap weight = 8 and gap length weight = 2), after manual adjustment of the initial alignment to account for optimal alignment among closely related sequences and highly conserved amino acid motifs. The available sequences of the RNA polymerase, capsid, 3′ORF, capsid overlap, and 3′UTR of 10–11 SLV strains were aligned separately (Table 1). Similarly, 13–15 NLV sequences were aligned in the same genome regions except the capsid overlap which does not exist in NLVs (Table 1). Both alignments of SLV and NLV sequences were used to generate a combined alignment (Table 2). In addition, 27 nucleotides and the corresponding 9 amino acids (including a stop signal for NLVs) locating at the junction between the RNA polymerase and the capsid protein genes were aligned separately (Fig. 1b). The sequence alignments used in this study are available upon request.

Uncorrected (unweighted) distances of pairwise amino acid sequences were calculated with the program Distances, and matrixes of diagonal rectangle showing the pairwise distances were converted to histograms showing frequency distribution of the pairwise distances using the program Histogram of the Microsoft Excel software.

Unrooted trees were reconstructed by using the PAUPSearch program [37] with the optimality criterion of minimum evolution [33] and the mean character distance as the distance correction method. Bootstrap analysis based on 100 replicates was used to estimate the statistical significance of branches in the best tree. The best tree was obtained by using the heuristic tree search method based on Farris' simple algorithm, for which initial tree was created by neighbour joining method [34] and negative branch lengths were reset to zero. The phylograms thus obtained were converted to radial trees using the program Treetool version 2.0.1 (Ribosomal RNA Database Project, University of Illinois board of treetools, rdp@phylo.life.uiuc.edu).

Nucleotide sequence accession numbers

The sequences of SLV/Lyon/598/97 and SLV/Lyon/30388/98 were deposited in the EMBL/GenBank database under accession numbers AJ271056 and AJ251991. The GenBank accession numbers of other SLV and NLV strains used in this study are listed in Table 1.

Table 2. Aligned SLV and NLV sequences subjected to analysis

	Polymerase (aa)[a]	Capsid (aa)[b]	3′ORF (aa)[b]	Capsid overlap (aa)[b]	3′UTR (nt)[c]
SLV sequences	210	583	173	165	52
NLV sequences	265	576	292	No[d]	36
SLV and NLV sequences	215	603	288	No[d]	Not done

[a]Length of the alignment of the C-terminal sequences
[b]Length of the alignment of the entire predicted proteins
[c]Portion of the alignment of the 3′UTR sequences (see Fig. 4)
[d]Capsid-overlap ORF does not exist in NLV strains

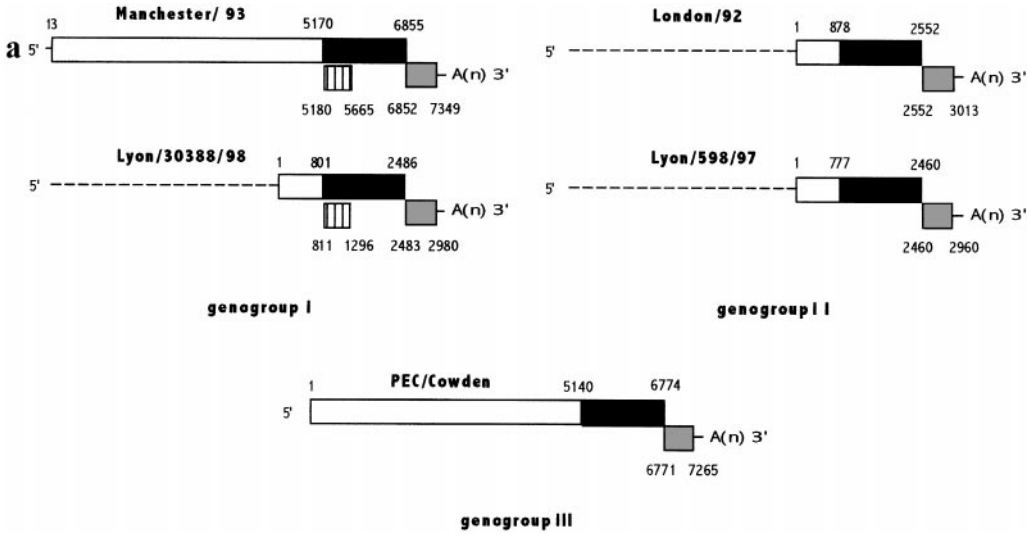
Results

Sequences of Lyon/598/97 and Lyon/30388/98

Lyon/598/97 and Lyon/30388/98 sequences had 3077 and 3032 nucleotides in length, respectively, extending from within the predicted RNA polymerase gene to within the 3′UTR. The RNA polymerase gene in these sequences was contiguous to the capsid gene and in the same reading frame (Fig. 1a). The predicted proteins of the entire capsid gene and the 3′ORF had molecular sizes comparable to those reported for other SLV strains available in GenBank. The 3′UTR of Lyon/598/97 and Lyon/30388/98 was not entirely sequenced, because the reverse primers used for RT-PCR amplification and sequencing were located upstream the 3′ terminus of the 3′UTR (Table 1). The predicted ORF overlapping the capsid gene in +1 frame (capsid overlap) was present in Lyon/30388/98 but not in Lyon/598/97 (Fig. 1 and Table 1).

Alignment of sequences

Overall, both the separate and combined alignments of the SLV and NLV sequences showed relatively good concordance with those previously reported [3, 13, 16, 18, 27, 35, 42]. However, the alignment of 10 SLV sequences indicated marked differences in the amino acid and nucleotide sequences of the 3′ end of the 3′ORF and the 5′ end of the 3′UTR of London/92 strain compared to the remaining 9 SLV strains. Comparison of London/92 and Lyon/598/97 sequences disclosed a nucleotidic difference, that could explain this marked differences between London/92 and other SLV sequences; while London/92 sequence showed a stretch of 7 cytosines between nt 2955 and 2961, Lyon/598/97 sequence showed a stretch of 8 cytosines in the corresponding stretch located between nt 2863 and 2870. Since the sequence of Lyon/598/97 was verified twice in two independent experiments, we considered that our observation was more likely to be ascribed to a deletion in London/92 sequence than a C insertion in Lyon/598/97 sequence. The frame-shift resulting from this presumed deletion explained the substitution of 16 amino acids and the C-terminal deletion of 12 amino acids in the 3′ORF of London/92 and the addition of 35 nucleotides to the 5′ end of its 3′UTR. We considered that these dramatic amino acid differences were more likely to be ascribed to an uncertainty in nucleotidic sequence identification or reporting of London/92. We therefore replaced the presumably deleted nucleotide with a symbol X, indicating an as yet undetermined nucleotide, and used the revised London/92 sequence (Table 1) to generate new alignments of 3′ORF and 3′UTR sequences. To generate the alignment in the 3′UTR, we also excluded 12 nucleotides located at the 5′ end of Houston/86, Houston/90, and Manchester/93 sequences in the alignment reported by Jiang [18], because these nucleotides (including a stop codon) encoded the C-terminal amino acids of the 3′ORF. In the revised alignment of the 3′UTR sequences, Lyon/598/97 and London/92 strains showed, at the 5′ end, 45 and 44 additional nucleotides, which were absent in other SLV strains (Fig. 4).



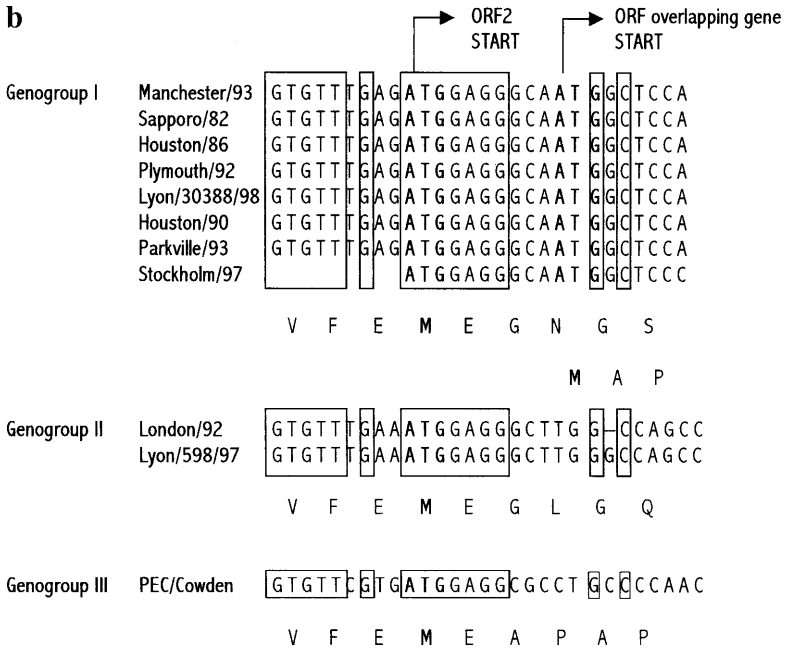


Fig. 1. Genomic organisation of SLV strains and their genogroup-specific sequences at the RNA polymerase-capsid junction. **a** Genomic organisation of Lyon/30388/98 and Lyon/598/97 was predicted from the 3′-terminal 3-kb nucleotide sequences of their corresponding cDNAs and is represented below that of Manchester/93 and London/92, prototype strains of their respective genogroups. The ORF overlapping the capsid gene is predicted only in strains of genogroup I. **b** Aligned nucleotide and predicted amino-acid sequences at the junction between the RNA polymerase and capsid genes. In strains of genogroup II and III, the ATG codon initiating the ORF overlapping the capsid gene is absent.
□ Polyprotein gene, ■ capsid gene, ▭ 3′ORF, ▥ ORF overlapping capsid gene

Analysis of sequences

Capsid protein

Based on the combined alignment of SLV and NLV sequences, the frequency distribution of the pairwise uncorrected distances between capsid protein sequences showed 4 well differentiating peaks (Fig. 2, top histogram), corresponding to, from right to left, the distances between members of different genera

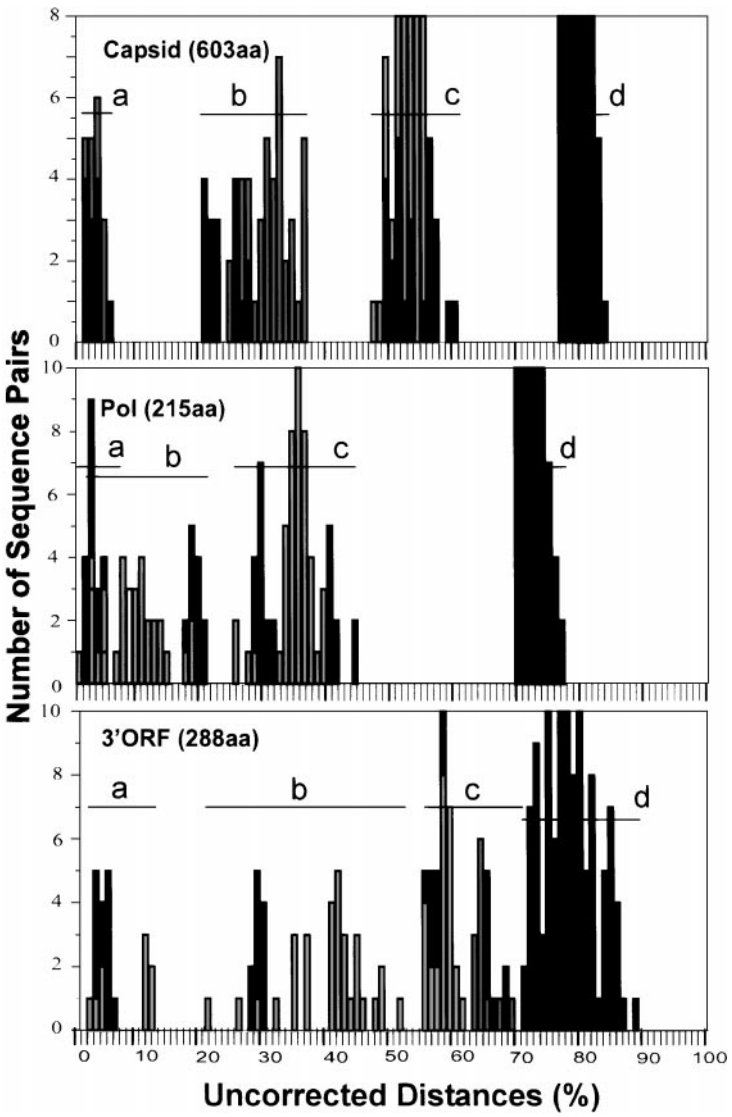


Fig. 2. Results of the distance analysis conducted in 3 different genome regions of SLVs and NLVs. Top, middle and bottom histograms show the frequency distribution of the uncorrected pairwise distances between capsid protein, RNA polymerase and 3′ORF aligned sequences. *a*, *b*, *c* and *d* indicate the distribution of the intra-cluster, inter-cluster, inter-genogroup and inter-genus distances, respectively. Black and gray bars indicate the distribution of the pairwise distances within SLVs and NLVs, respectively

(inter-genus distances), members of different genogroups (inter-genogroup distances), members of different genetic clusters within the same genogroup (inter-cluster distances) and members within the same genetic cluster (intra-cluster distances). It should be stated that, in the present study, genogroups and genetic clusters refered to major genetic groups and genetic subgroups comprized into the major genetic groups, respectively. Inter-genus distances distributed between 84 and 86%. Inter-genogroup, inter-cluster and intra-cluster distances between SLV sequences were comprised between 49% and 55%, 19% and 25%, 1% and 5%, respectively. Those between NLV sequences were comprised between 47% and 56%, 24% and 37%, 1% and 4%, respectively. Identification of the pairwise SLV sequences attributed to the inter-genogroup, inter-cluster and intra-cluster distances indicated that the 11 SLV sequences segregated into 3 genogroups, comprised of 5 genetic clusters. Similarly, identification of the pairwise NLV sequences indicated that the 15 NLV sequences segregated into 3 genogroups comprised of 9 genetic clusters.

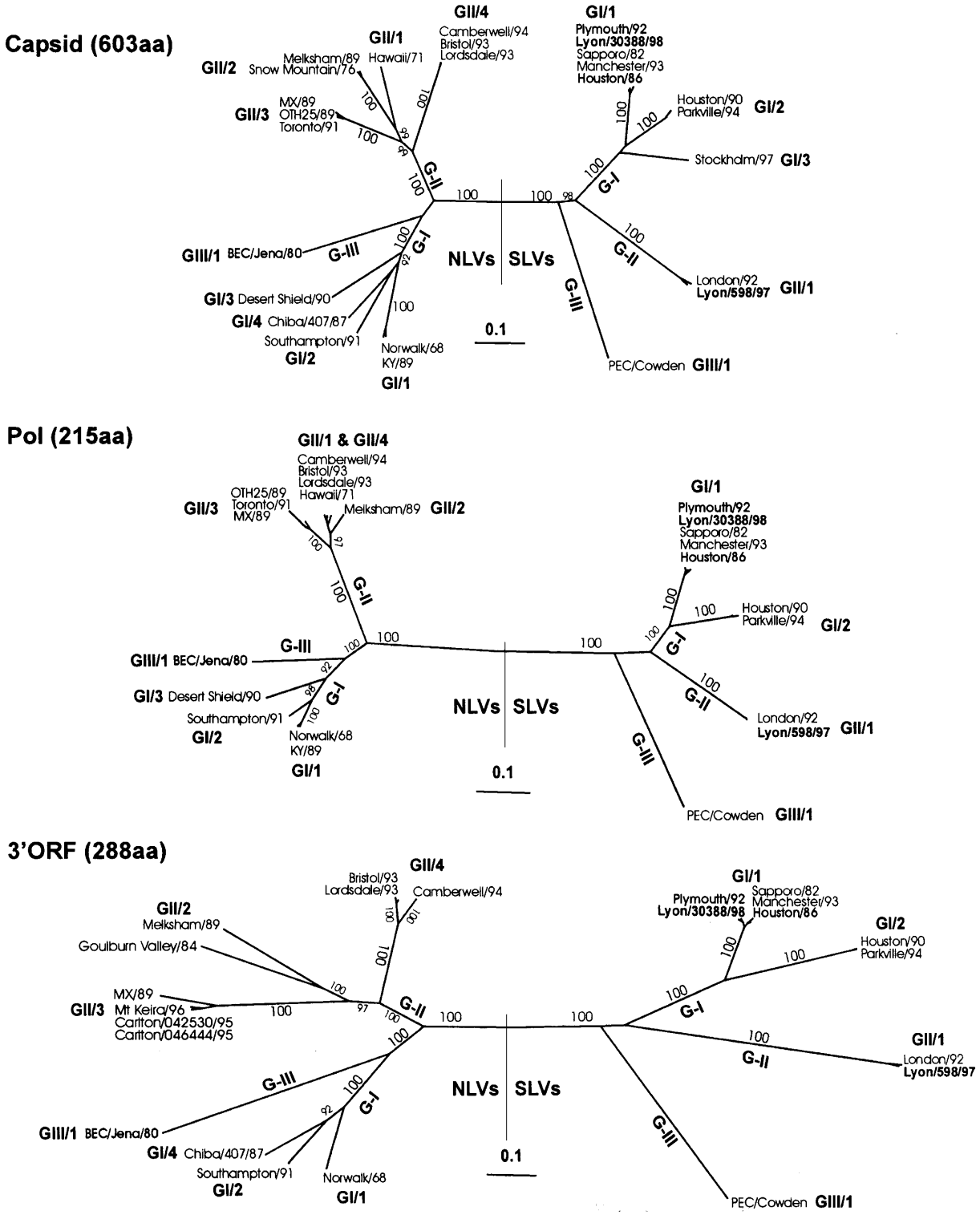
The phylogenetic tree of SLVs and NLVs demonstrated a branching pattern that was consistent with the pattern of frequency distribution of the uncorrected distances (Fig. 3, top). The confidence level of the branching pattern in the phylogenetic tree was high, as indicated by the bootstrap values of 92–100% repeatability. Moreover, the same branching patterns were indicated by the phylogenetic trees reconstructed separately using each of the alignments of SLV and NLV sequences (data not shown).

These results permitted us to conclude that the 3 genogroups and 5 genetic clusters of SLV strains were consistent with the 3 genogroups and 9 genetic clusters of NLVs used in this study. We therefore designated the major genetic groups and genetic subgroups of SLVs in accordance with the naming criteria of the genogroups and genetic clusters of NLVs [1]: GI/1 (Sapporo/82, Houston/86, Plymouth/92, Manchester/93, and Lyon/30388/98), GI/2 (Houston/90 and Parkville/93), GI/3 (Stockholm/97), GII/1 (London/92 and Lyon/598/97) and GIII/1 (PEC/Cowden) (Fig. 3, top).

RNA polymerase

The sequence of the RNA polymerase coding region was more conserved than that of the capsid coding region, but the intra-cluster, inter-cluster, and inter-genogroup distances (black bars) of the 10 available SLV sequences (note: RNA polymerase sequence was not available for Stockholm/97 strain) were still well differentiated (Fig. 2, middle histogram). In contrast, distinctness of differentiation of the 13 available NLV sequences by the intra-cluster (2–6%) and inter-cluster (2–14%) distances (gray bars) dropped markedly.

The topology of the phylogenetic tree indicated a good concordance with the pattern of the frequency distribution of the uncorrected pairwise distances (Fig. 3, middle). While the topologies were well consistent between trees based on the RNA polymerase and capsid sequences for the 10 SLV strains, there was an apparent discrepancy between those based on the RNA polymerase and capsid


Capsid (603aa)
GII/4
Camberwell/94
Bristol/93
Lordsdale/93
GII/1
Hawaii/71
GII/2
Melksham/89
Snow Mountain/76
GII/3
MX/89
OTH25/89
Toronto/91
GI/1
Plymouth/92
Lyon/30388/98
Sapporo/82
Manchester/93
Houston/86
Houston/90
Parkville/94
GI/2
Stockholm/97
GI/3
G-II
G-I
G-III
NLVs
SLVs
GIII/1 BEC/Jena/80
GI/3 Desert Shield/90
GI/4 Chiba/407/87
Southampton/91
GI/2
Norwalk/68
KY/89
GI/1
London/92
Lyon/598/97
GII/1
PEC/Cowden
GIII/1
0.1
Pol (215aa)
GII/1 & GII/4
Camberwell/94
Bristol/93
Lordsdale/93
Hawaii/71
GII/3
OTH25/89
Toronto/91
MX/89
Melksham/89
GII/2
GI/1
Plymouth/92
Lyon/30388/98
Sapporo/82
Manchester/93
Houston/86
Houston/90
Parkville/94
GI/2
G-II
G-I
G-III
GIII/1 BEC/Jena/80
GI/3 Desert Shield/90
Southampton/91
GI/2
Norwalk/68
KY/89
GI/1
NLVs
SLVs
London/92
Lyon/598/97
GII/1
PEC/Cowden
GIII/1
0.1
3'ORF (288aa)
GII/4
Bristol/93
Lordsdale/93
Camberwell/94
GII/2
Melksham/89
Goulburn Valley/84
GII/3
MX/89
Mt Keira/96
Carlton/042530/95
Carlton/046444/95
GI/1
Sapporo/82
Plymouth/92
Lyon/30388/98
Manchester/93
Houston/86
GI/2
Houston/90
Parkville/94
G-II
G-I
G-III
GII/1
London/92
Lyon/598/97
NLVs
SLVs
GIII/1 BEC/Jena/80
GI/4 Chiba/407/87
Southampton/91
GI/2
Norwalk/68
GI/1
PEC/Cowden
GIII/1
0.1

sequences for the 13 NLV strains. In particular, the branches leading to GII/1 and GII/4 clusters grouped together in the tree based on the RNA polymerase sequences. The histograms and trees based on the separate alignments of the SLV and NLV sequences indicated identical patterns of distribution as those based on the combined alignment (data not shown).

3′ORF

While the intra-cluster, inter-cluster, and inter-genogroup distances between the SLV sequences (Fig. 2, bottom histogram, black bars) were highly distinct, the distinctness of these distances almost disappeared for the NLV sequences (gray bars). The distribution of the inter-cluster distances of the NLV sequences consisted of non-overlapping two parts attributed to the GI (19–27%) and GII (30–48%) sequences.

These differences in the frequency distribution of the distances between SLV and NLV sequences were also noticeable in the phylogenetic tree (Fig. 3, bottom); although both the SLV and NLV sequences were differentiated into the expected genogroups and genetic clusters in the tree, lengths of the 3 branches leading into the 3 GI clusters of NLV sequences were shortened as compared with the branches leading into the 4 GII clusters of the NLV sequences. Similar analyses of the separate alignments of the SLV and NLV sequences provided strong support to the analysis of the combined alignment (data not shown).

Capsid overlap

This frame could be predicted only in the GI strains of SLVs (Table 1, Fig. 1). In all the GI strains, the predicted initiator codon of this frame was maintained in a conserved region of 7-nt, GCAAUGG, referred to be strongly favoured for translation initiation [6, 24]. Unequivocal differentiation of the 8 GI capsid overlap sequences into the 3 genetic clusters GI/1, GI/2 and GI/3 was evident by both the marked difference between the intra-cluster (0–7%) and inter-cluster (37–44%) distances (Table 3) and the 3 distinct branches leading from a bifurcation to these 3 genetic clusters in a phylogenetic tree (data not shown).

3′UTR

The 10 aligned 52-nt sequences of SLVs were clearly differentiated into GI/1, GI/2, GII/1 and GIII/1 genetic clusters. The differentiability of the genogroups I and II from the genetic clusters was not as high as in other genome regions described above: the intra-cluster, inter-cluster, and inter-genogroup distances

Fig. 3. Results of the phylogenetic analysis conducted in 3 different genome regions of SLVs and NLVs. Capsid, RNA polymerase and 3′ORF aligned sequences were analysed using the PAUP program. Bootstrap values are indicated as % of 100 replicates. Genogroups and genetic clusters were determined in the present study, based on the analysis of the entire capsid sequences; those of NLVs refer to results of the previous [1, 35] and the current studies

Table 3. Pairwise distances among capsid overlap ORF sequences from genogroup I SLV strains. GI/1, GI/2 and GI/3 indicate the genetic clusters of SLV strains, as determined in the present study (see also Table 1). The pairwise distances were determined using the program Distances

	GI/2		GI/1					GI/3
	Houston/90	Parkville/94	Plymouth/92	Lyon/30388/98	Manchester/93	Sapporo/82	Houston/86	Stockholm/97
Houston/90	0	3.66	41.36	41.36	40.74	41.36	40.74	44.72
Parkville/94		0	38.27	38.27	37.65	38.27	37.65	42.86
Plymouth/92			0	0	1.23	4.32	4.32	39.62
Lyon/30388/98				0	1.23	4.32	4.32	39.62
Manchester/93					0	4.32	4.94	40.25
Sapporo/82						0	7.41	41.51
Houston/86							0	39.62
Stockholm/97								0

distributed 0–8%, 18–22%, and 24–28%, respectively. In contrast, the pairwise distances between PEC/Cowden and human SLV strains distributed between 50 and 64%. The phylogenetic tree demonstrated 4 distinct branches leading to the 4 genetic clusters, which were grouped into the 3 genogroups (data not shown). It should be noted that the 5′-terminal highly conserved nucleotides of the GII/1 strains of SLVs were not included in these analyses, because the currently available programs were not well suited for the analysis of the overhanging sequences in the alignment (Fig. 4).

Regarding the 13 NLV sequences, the inter-cluster distances (33–56%) were well differentiated from the intra-cluster distances (0–3%), but were not differentiated from the inter-genogroup distances (40–70%). Similarly, the phylogenetic tree demonstrated differentiation of the 13 sequences into the 8 genetic clusters, but these clusters could not be grouped into the 3 genogroups by the branch lengths (data not shown). These results, together with the absence of the highly conserved 52-nucleotide stretch indicated that the sequences of 3′UTR were more divergent in NLVs than in SLVs.

Junction of RNA polymerase and capsid genes

The analysis of the aligned SLV sequences showed that a 27-nucleotide stretch located at the junction of the RNA polymerase and capsid genes was genogroup-specific (Fig. 1b). Nucleotide differences between GI, GII and GIII sequences correlated with the presence of the capsid overlap frame in GI strains and its absence in GII and GIII strains (Fig. 1b).

The 13 NLV sequences also maintained this genogroup-dependent property of the highly conserved 27 nucleotides at this junction. These 27 nucleotides included a stop codon at the 3′ terminus of ORF1 and an octa-peptide at

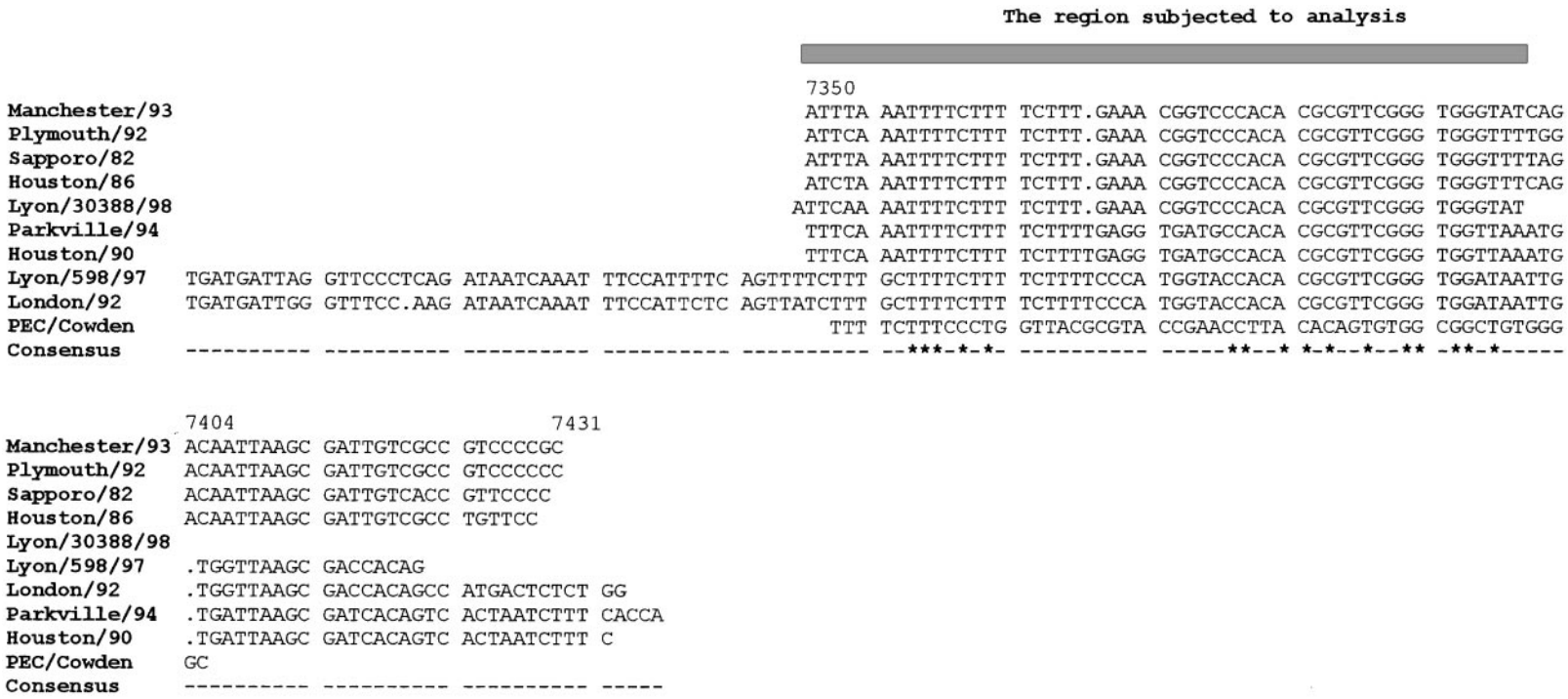


Fig. 4. Alignment of the sequences in the 3′UTR of 10 SLV strains. The region subjected to pairwise distance and phylogenetic analyses is highlighted by a gray bar. We indicated nucleotide numbers of Manchester/93 genomic sequence. *indicates the nucleotides conserved among all SLV sequences

the N-terminus of ORF2: MMMASKDA in GI, MKMASNDA in GII, and MKMTDKDV in GIII.

Discussion

In the present study, we performed distance and phylogenetic analyses in 5 genome regions of 11 SLV strains, including 2 newly sequenced, and 19 NLV published strains. Our main goal was to provide a genetic classification of SLVs comparable to the genetic classification schemes of NLVs reported recently [1, 14], based on the analysis of the capsid sequences. We also analysed 4 other genome regions, namely RNA polymerase, the 3′ORF, the ORF overlapping gene and the 3′UTR to provide a more comprehensive comparison of genomic relationships among SLVs, comparatively to NLVs.

Based on the analysis of the entire capsid sequences, we could uniformly estimate the clustering of NLVs and SLVs and therefore classify SLVs and NLVs into comparable genogroups and genetic clusters. The 11 SLV strains analysed in the present study were classified into 5 genetic clusters and 3 genogroups, 2 genogroups comprising the 10 human strains and the latter comprising the porcine PEC/Cowden strain. Classification of human SLV strains into 2 major genetic groups had been previously suggested by Jiang et al. [18]. However, their analysis was limited to only 4 SLV strains (Manchester/93, Houston/86, Houston/90, and London/92). Based on the analysis of an increased number of strains (including the new Lyon/97/598 strain belonging to the London/92 cluster) and the use of uniform criteria for the classification of SLVs and NLVs, we could unambiguously classify the human SLVs into 2 genogroups and show that these were equivalent to the genogroups of NLVs. In addition, we confirmed that the porcine strain, PEC/Cowden, formed a separate cluster in SLVs, as suggested by Guo et al. [15] and provided new evidence for a third genogroup of SLVs. Additional sequencing data on PEC isolates should further predict their relationship with PEC/Cowden and human caliciviruses.

The most noticeable characteristic of the SLV sequences was their highly distinct differentiation into the same genogroups and genetic clusters maintained throughout all the genome regions examined. By contrast, the property of the NLV sequences to be differentiated into genogroups and genetic clusters markedly dropped in the RNA polymerase, 3′ORF and 3′UTR, suggesting that SLVs could be evolutionarily more stable than NLVs. Additional data should further predict the comparative molecular evolution of SLVs and NLVs.

Previously, the genetic classification of NLVs based on the RNA polymerase sequences had been reported as being essentially identical to that based on the capsid protein sequences [41], and the strains which did not agree with this rule were considered as possible recombinants [16, 19]. Our results may provide an alternative explanation to discrepant genetic grouping patterns, i.e grouping of RNA polymerase sequence of individual strains may not be necessarily concordant with that of the capsid sequence because of a difference in genetic characteristics

between NLVs and SLVs. The questions concerning the recombinants of NLVs should be examined in future studies.

An additional finding in this study was an apparent genogroup-specificity of the SLV and NLV sequences located at the junction of the RNA polymerase and capsid genes. Previously, highly conserved motifs have been reported at the 5′ terminus of the genome and the junction between RNA polymerase and capsid genes in FCV, RHDV, certain NLVs, and Manchester/93 SLV strain [5, 23, 25] and have been predicted to have a regulatory role such as providing a signal for packaging, replication or transcription of the genomic and subgenomic RNAs. However, their genogroup-specific property has not yet been reported. Of note, the difference of sequence between GI, GII and GIII strains of SLVs in this region correlated with the genogroup-dependent presence or absence of the capsid overlap frame. The question whether a biologically active protein is produced from the capsid overlap frame has been a matter of controversy among scientists [6, 27]. Our observation that the sequence motif GCAAUGG, strongly favoured for translation initiation, is conserved among all 8 GI strains examined suggests that a biologically active protein may be expressed in GI SLV strains during replication, as already ascertained by Clarke and Lambden [6]. This predicted protein could be associated with a property specific for GI strains.

It should be also noted that the property of the 3′UTR sequences of SLVs to be differentiated into the same genogroups and genetic clusters as the capsid sequences has not yet been reported.

The goals of the genetic classification of SLVs and comparative analysis of SLV and NLV sequences, as attempted in this study, were both to establish a commonly acceptable genetic classification scheme of the SLV genus and to elucidate distinguishing characteristics of the SLV sequences that may help to understand the differences in epidemiological and clinical features between SLV and NLV infections. Achieving such goals will require analysis of more sequences that adequately represent the genetic diversity of SLV strains. Although SLVs have been shown to circulate world-wide, the number of detected and characterised SLV strains remains very low compared to that of NLV strains, partly due to the lack of sensitivity of the detection tests [40]. Thus, molecular epidemiology has a long way to go before it provides sufficient information on the genetic diversity of SLV strains. Interestingly, we report here, for the first time, circulation of SLVs of distinct genogroups in France. The consistency of the differentiation of SLV sequences into genogroups and genetic clusters, whatever the region considered, suggests that RT-PCR amplification of a 307 bp RNA polymerase fragment currently used for diagnosis is sufficient to properly classify new SLV strains. This observation may be valuable for epidemiological studies, as biological material is often lacking. However, it is likely more diversity may appear as new strains are found and characterised.

Acknowledgement

We would like to thank Martine Poulet for technical assistance.

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, Golding B, Jiang X, Cubitt DW, Wolfaardt M, Smith AW, Matson DO (1997) Phylogenetic analysis of the caliciviruses. J Med Virol 52: 419–424
3. Cauchi MR, Doultree JC, Marshall JA, Wright PJ (1996) Molecular characterization of Camberwell virus and sequence variation in ORF3 of small round-structured (Norwalk-like) viruses. J Med Virol 49: 70–76
4. Chiba S, Nakata S, Numata-Kinoshita K, Honma S (2000) Sapporo virus: history and recent findings. J Infect Dis 181[Suppl 2]: S303–S308
5. Clarke IN, Lambden PR (1997) The molecular biology of caliciviruses. J Gen Virol 78: 291–301
6. Clarke IN, Lambden PR (2000) Organization and expression of calicivirus genes. J Infect Dis 181[Suppl 2]: S309–S316
7. Dastdjerdi AM, Green J, Gallimore CI, Brown DWJ, Bridger JC (1999) The bovine Newbury agent-2 is genetically more closely related to human SRSVs than to animal caliciviruses. Virology 254: 1–5
8. Dedman D, Laurichesse H, Caul EO, Wall PG (1998) Surveillance of small round-structured virus (SRSV) infections in England and Wales, 1990–1995. Epidemiol Infect 121: 139–140
9. Dingle KE, Lambden PR, Caul EO, Clarke IN (1995) Human enteric Caliciviridae: the complete genome sequence and expression of virus-like particles from a genetic group II small round structured virus. J Gen Virol 76: 2349–2355
10. Fankhauser RL, Noel JS, Monroe SS, Glass RI (1998) Molecular epidemiology of small round-structured viruses in outbreaks of gastroenteritis in the United States. J Infect Dis 178: 1571–1578
11. 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
12. Glass RI, Noel J, Ando T, Fankhauser R, Belliot G, Mounts A, Parashar UD, Bresee JS, Monroe SS (2000) The epidemiology of enteric caliciviruses from humans: a reassessment using new diagnostics. J Infect Dis 181[Suppl 2]: S254–S261
13. Green J, Vinje J, Gallimore CI, Koopmans A, Hale A, Brown DW (2000) Capsid diversity among Norwalk-like viruses. Virus Genes 20: 227–236
14. Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, Matson DO, Nakata S, Neill JD, Studdert MJ, Thiel H-J (2000) Taxonomy of the caliciviruses. J Infect Dis 181[Suppl 2]: S322–S330
15. 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
16. Hardy ME, Kramer SF, Treanor JJ, Estes MK (1997) Human calicivirus genogroup II capsid sequence diversity revealed by the analyses of the prototype Snow Mountain Agent. Arch Virol 142: 1469–1479
17. Jiang X, Graham DY, Wang K, Estes MK (1990) Norwalk virus genome cloning and characterization. Science 250: 1580–1583
18. 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
19. Jiang X, Espul C, Zhong M, Cuello H, Matson DO (1999) Characterization of a novel human calicivirus that may be a naturally occurring recombinant. Arch Virol 144: 2377–2387

20. 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
21. Kapikian AZ, Estes MK, Chanock RM (1996) Norwalk group of viruses. In: Fields BN, Knipe DM, Howley PM, et al. (eds) Fields Virology, 3rd ed, vol 1. Lippincott-Raven, Philadelphia, pp 783–810
22. Lambden PR, Caul EO, Ashley CR, Clarke IN (1993) Sequence and genome organization of a human small round-structured (Norwalk-like) virus. Science 259: 516–519
23. Lambden PR, Liu BL, Clarke IN (1995) A conserved sequence motif at the 5′ terminus of the Southhampton virus genome is characteristic of the Caliciviridae. Virus Genes 10: 149–152
24. 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: 1345–1356
25. Liu BL, Clarke IN, Caul EO, Lambden PR (1997) The genomic 5′ terminus of Manchester calicivirus. Virus Genes 15: 25–28
26. Liu BL, Lambden PR, Gunther H, Otto P, Elschner M, Clarke IN (1999) Molecular characterization of a bovine enteric calicivirus: relationship to the Norwalk-like viruses. J Virol 73: 819–825
27. 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
28. Noel JS, Fankhauser RL, Ando T, Monroe SS, Glass RI (1999) Identification of a distinct common strain of ⟪Norwalk-like viruses⟫ having a global distribution. J Infect Dis 179: 1334–1344
29. Numata K, Hardy ME, Nakata S, Chiba S, Estes MK (1997) Molecular characterization of morphologically typical human calicivirus Sapporo. Arch Virol 142: 1537–1552
30. Pang X-L, Joensuu J, Vesikari T (1999) Human calicivirus-associated sporadic gastroenteritis in Finnish children less than two years of age followed prospectively during a rotavirus vaccine trial. Pediatr Infect Dis J 18: 420–426
31. Pang X-L, 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
32. Pringle CR (1999) Virus taxonomy. Arch Virol 144: 421–429
33. Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum evolution method of phylogenetic inference. Mol Biol Evol 10: 1073–1095
34. Saitou A, Nei M (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
35. Seah EL, Gunesekere IC, Marshall JA, Wright PJ (1999) Variation in ORF3 of genogroup 2 Norwalk-like viruses. Arch Virol 144: 1007–1014
36. Seah EL, Marshall JA, Wright PJ (1999) Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells. J Virol 73: 10531–10535
37. Swofford DL, Olsen GJ, Waddel PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics, 2nd ed. Sinauer Associates, Sunderland, pp 407–462
38. van der Poel WH, Vinjé J, van der Heide R, Herrera MI, Vivo A, Koopmans MP (2000) Norwalk-like caliciviruses genes in farm animals. Emerg Infect Dis 6: 36–41

39. Vinjé 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: 1374–1378
40. Vinjé J, Deijl H, van der Heide R, Lewis D, Hedlund KO, Svensson L, Koopmans MPG (2000) Molecular detection and epidemiology of Sapporo-like viruses. J Clin Microbiol 38: 530–536
41. Vinjé J, Green J, Lewis DC, Gallimore CI, Brown DW, Koopmans MP (2000) Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses". Arch Virol 145: 223–241
42. Wang J, Jiang X, Madore HP, Gray J, Desselberger U, Ando T, Seto Y, Oishi I, Lew JF, Green KY, Estes MK (1994) Sequence diversity of small round-structured viruses in the Norwalk group. J Virol 68: 5982–5990
43. Wright PJ, Gunesekere IC, Doultree JC, Marshall JA (1998) Small round-structured (Norwalk-like) viruses and classical human caliciviruses in Southeastern Australia, 1980–1996. J Med Virol 55: 312–320

Author's address: Dr. I. Schuffenecker, Laboratoire de Virologie, Centre National de Référence pour les Entérovirus, Université Claude Bernard Lyon I, 69373 Lyon Cedex 08, France; e-mail: i.schuffenecker@free.fr

Received February 6, 2001