Genomic Analyses of Human Sapoviruses Detected over a 40-Year Period Reveal Disparate Patterns of Evolution among Genotypes and Genome Regions

Kentaro Tohma [1,*] , **Michael Kulka** [2], **Suzie Coughlan** [3], **Kim Y. Green** [4] **and Gabriel I. Parra** [1,*]

1   Division of Viral Products, CBER, FDA, Silver Spring, MD 20993, USA
2   Division of Molecular Biology, CFSAN, FDA, Laurel, MD 20993, USA; Michael.Kulka@fda.hhs.gov
3   National Virus Reference Laboratory, UC-Dublin, Dublin 4, Ireland; suzie.coughlan@ucd.ie
4   Laboratory of Infectious Diseases, NIH, Bethesda, MD 20993, USA; kim.green@nih.gov
*   Correspondence: kentaro.tohma@fda.hhs.gov (K.T.); gabriel.parra@fda.hhs.gov (G.I.P.);
    Tel.: +1-240-4026935 (G.I.P.); Fax: +1-301-5951070 (G.I.P.)

Abstract: Human sapovirus is a causative agent of acute gastroenteritis in all age groups. The use of full-length viral genomes has proven beneficial to investigate evolutionary dynamics and transmission chains. In this study, we developed a full-length genome sequencing platform for human sapovirus and sequenced the oldest available strains (collected in the 1970s) to analyse diversification of sapoviruses. Sequence analyses from five major genotypes (GI.1, GI.2, GII.1, GII.3, and GIV.1) showed limited intra-genotypic diversification for over 20–40 years. The accumulation of amino acid mutations in VP1 was detected for GI.2 and GIV.1 viruses, while having a similar rate of nucleotide evolution to the other genotypes. Differences in the phylogenetic clustering were detected between RdRp and VP1 sequences of our archival strains as well as other reported putative recombinants. However, the lack of the parental strains and differences in diversification among genomic regions suggest that discrepancies in the phylogenetic clustering of sapoviruses could be explained, not only by recombination, but also by disparate nucleotide substitution patterns between RdRp and VP1 sequences. Together, this study shows that, contrary to noroviruses, sapoviruses present limited diversification by means of intra-genotype variation and recombination.

Keywords: sapovirus; calicivirus; genetic diversity; evolution; recombination; norovirus

1. Introduction

Human sapovirus is a causative agent of acute gastroenteritis in people from all age-groups. A member of the family *Caliciviridae*, sapoviruses are non-enveloped and their viral capsid has icosahedral symmetry. Their genomes are single-stranded, positive-sense, 3'-polyadenylated RNA molecules of approximately 7.4 kb in length. The genome is organized into two open reading frames (ORFs), which are flanked by a short 5'-end and a long 3'-end non-coding regions. An ORF3 has been predicted in some sapovirus strains, but its function is still unknown [1]. The ORF1 encodes the nonstructural proteins, including the RNA dependent RNA polymerase (RdRp), and the major viral capsid protein (VP1) [1,2]. Cryo-electron microscopy and homology modeling revealed that VP1 is structurally divided into a shell (S) and protruding (P) domains; the latter further divided into P1 and P2 subdomains [3]. Based on the genetic differences of the VP1-encoding sequences, human sapoviruses are phylogenetically clustered into four genogroups (GI, GII, GIV and GV), with each genogroup further clustered into multiple genotypes [1,4]. Discrepancies in the phylogenetic clustering can occur when using different genomic regions, particularly those encoding the RdRp and VP1, and this has led to the identification of intra- and inter-genogroup recombinant strains [1,5–13].

Indeed, recombination has been shown to be an important mechanism of diversification for norovirus and other caliciviruses [1,14,15]. Some of the recombination events reported for sapoviruses did not clearly describe both parental strains [5–7,10,13], thus the "recombination signal" might be confounded by the possibility of different evolutionary diversification patterns among sapovirus genomic regions [16].

Human sapovirus symptoms can include diarrhea, vomiting, nausea, stomach cramps, and myalgia. Sapovirus-related symptoms are usually mild, and can be resolved within a few days; however, severe disease has been reported in vulnerable individuals and can result in hospitalization [11,16]. Sapovirus transmission occurs by person-to-person contact and/or by consumption of contaminated water, soil, or food [1,17]. Similar to norovirus, gastroenteritis outbreaks associated to sapoviruses occur in semi-enclosed settings, such as childcare facilities, schools, nursing homes, and catering receptions [18,19].

Genomic analyses for evolutionary studies [20], epidemiological investigations [21], and transmission tracking of viral infections [22] have proven beneficial to control and prevent infections and mitigate the burden on patient care [23,24]. The use of whole, versus short, genome sequences can enhance those studies, particularly in viruses that can rapidly acquire mutations [25,26]. The application of metagenomic analyses using whole genome sequencing is increasing in use in clinical settings, because of the power and potential for obtaining longer nucleotide sequence reads of a pathogen's genome [27]. However, in most cases, the performance of these metagenomic analyses are greatly affected by the concentration of the viral contaminant and the complexity of the food and/or water matrix [28]. Different alternatives have been implemented in recent years, such as the enrichment of viral particles and random or targeted amplification of viral genomes, to improve the performance of the metagenomic approaches for viral diagnostics and outbreak investigations [29–32]. The aim of this study was to develop a simplified full-length viral genome sequencing platform for analyses of human sapovirus genomes. This new platform was used to sequence archival sapovirus-positive samples dating from the 1970s, and together with those sequences available in public databases were used to characterize the evolutionary dynamics of human sapoviruses. A better understanding of the evolutionary process that drives viral diversification could provide insights into the development of therapeutic and control strategies against these viruses.

2. Materials and Methods

2.1. Fecal Samples Positive for Human Sapovirus

Fifteen human fecal samples positive for human sapovirus were tested for full-length genome PCR amplification, sequencing, and evolutionary analysis. The samples were collected as part of a study conducted by the World Health Organization in different countries during 1976–1979 [33]. The retrospective research use of the samples in this study was approved by the NIAID Institutional Review Board. Samples were de-identified, and a waiver of informed consent was granted for this use. Four additional human fecal samples obtained from the National Virus reference Laboratory-UC Dublin confirmed positive for human sapovirus, and identified as GV.1 (GenBank accession MK291480), were tested for full-length genome PCR amplification.

2.2. Full-Length Viral Genome PCR and Sequencing

Sapovirus RNA genome was extracted from 10% stool suspensions using the MagMax Viral RNA Isolation Kit (Thermo Fisher Scientific, MA, USA). Viral RNA genome was retrotranscribed to complementary DNA (cDNA) using the Maxima Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The cDNA reaction was performed immediately after viral RNA extraction to reduce damage of the long viral RNA molecules by a freeze and thaw cycle. The annealing reaction (65 °C, 5 min then hold at 5 °C) contained 5 μL of extracted sample RNA, 2 μL 100 μM of the primer TX30SXN [34], 1 μL 10 mM dNTPs and volume adjusted to 14 μL final with DNase/RNase free (DNF)

water. Following the annealing step, a single 6 μL mixture containing 4 μL of 5× RT first-strand synthesis buffer, 0.1 μL Max H minus Enzyme Mix and 1.9 μL DNF water was added to each annealing reaction while on ice. The RT reaction was completed by incubation at 50 °C for 1 h, 85 °C for 5 min, then held at 4 °C and/or stored at −80 °C until use in the full-length viral genome PCR reaction.

The full-length viral genome amplification was performed using a similar protocol, as described for noroviruses [35]. Briefly, the SequalPrep Long PCR Kit (Thermo Fisher Scientific) was used with 5 ul of viral cDNA and 2 μL of each (10 μM) primer pSapo27 (5′ GTGATTGGTTAGTATGGCTTCCAAGCC 3′) and TX30SXN [34]. The complete PCR reaction was run in 1% agarose gel, and the resulting full-length viral genome amplicons (~7.4 kbp) were excised and purified using the Qiagen Gel Extraction Kit (Qiagen, CA, USA). The recovered amplicons were quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and subjected to next-generation sequencing (NGS). The library for NGS was prepared using the Nextera DNA Flex Library Prep Kit (Illumina, CA, USA), and the paired-end 2 × 150 bp sequence reads were obtained using the MiSeq system (Illumina). Reads were quality-filtered (base quality score ≥ 20) and mapped against reference genomes to reconstruct their consensus sequence using HIVE platform [36], as described previously [37]. Newly obtained nearly full-length sapovirus sequences were deposited in GenBank (accession numbers MN794205–MN794218).

2.3. Sequence Analyses

To investigate the phylogenetic relationship and evolutionary dynamics of human sapoviruses, sequences from archival samples were analyzed with publicly available sequences. A dataset of nearly full-length genome sequences (>7000 nucleotide, *n* = 139), the VP1-encoding nucleotide sequences (>1600 nucleotide, *n* = 233), and the RdRp-encoding sequences (>650 nucleotide, *n* = 166) from human sapovirus genotypes were obtained from GenBank on January 31, 2019 (Table S1). Sequences from each genogroup/genotype were separately aligned using MUSCLE [38]. The phylogenetic relationship of the archival strains and reference strains from GenBank was analyzed using the RdRp- and the VP1-encoding nucleotide sequences using maximum-likelihood method implemented in MEGA7 [39]. The best substitution model of each dataset was selected, as per the Bayesian information criterion. The reconstructed phylogenetic trees were visualized using FigTree v1.4.3. The clock-like signal of the RdRp- and VP1-encoding nucleotide sequences were analyzed using inferred maximum-likelihood trees using TempEst v1.5 [40]. The root-to-tip divergence plot was constructed at the genotype level and the best-fitting root option was used to minimize the sum of the squared residuals. The evolutionary rate of the RdRp- and VP1-encoding nucleotide sequences was estimated using the Bayesian Markov chain Monte Carlo (MCMC) framework, as implemented in BEAST v1.8.4 [41]. All the analyzed genotypes showed strong clock-likeness and thus, we applied the strict clock model. The SRD06 model was used for estimating the substitution process of the sequences [42]. The population size was assumed to be constant throughout their evolutionary history. The MCMC runs were performed until the convergence of all the parameters was confirmed by visual inspection and effective sample size (>200 in all parameters), using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). The first 10% of the logs from the MCMC runs were removed as a burn-in, before summarizing the posterior values. The posterior values of the substitution rate were summarized using GraphPad Prism v7. To investigate the pattern of accumulation of amino acid substitutions in the VP1, the pairwise amino acid differences and the timespan of detection for each paired-strain were calculated using a Python script described elsewhere [35] and visualized using GraphPad Prism v7. Shannon entropy values were calculated using the Shannon Entropy-One tool as implemented in Los Alamos National Laboratory (www.hiv.lanl.gov). Entropy values for each amino acid position on the ORF1 and 2 were plotted and subjected to one-way ANOVA with Tukey's post-hoc multiple comparison test using GraphPad Prism v7. Entropy values for each nucleotide position at the RdRp-VP1 boundaries were calculated for human sapovirus and human norovirus for comparison. For human noroviruses, 1669 (nearly) full-length genome sequences from human norovirus GI and GII strains were obtained from GenBank on August 23, 2019. Sequences were multiple-aligned using MAFFT [43], and trimmed into the

VP1-encoding region or approximately 350 nucleotides of the RdRp-VP1 junction region for the entropy analyses. Finally, the recombination signal of the archival strains (GI.1/HK37/1977 and GII.4/T003/1976) was assessed using its nearly full-length genome sequences and SimPlot v3.5.1 [44]. The nucleotide similarity (window-size 200 nucleotide and step-size 20 nucleotide) was calculated using Kimura 2-parameter model and was plotted using GraphPad Prism v7.

3. Results and Discussion

The aim of this work was to develop a simplified full-length viral genome sequencing platform for analyses of human sapovirus genomes and evolution. To minimize the cost of sequence per sample, we decided to take advantage of a PCR amplicon-based platform that we developed to sequence human noroviruses [35]. To this end, we designed a primer that annealed to the first 27 nucleotide of the 5'-end of the sapovirus genome, which is highly conserved among strains, and together with a poly-T primer, would allow full-length viral RNA genome amplification (Figure 1a,b). The amplicons generated would serve as template for NGS analyses. The system showed successful amplification in 15 out of 19 sapovirus-positive samples tested. Representative GI, GII and GV amplicons are shown in Figure 1c. We successfully sequenced 14 full-length sapovirus genome amplicons ($\geq$ 7402 nucleotide) from the archival samples with an average depth of 6026–22516 reads per nucleotide position (Table S2, Figure 1d). The viruses sequenced presented the following genotypes: GI.1 ($n = 5$), GI.2 ($n = 1$), GII.1 ($n = 7$), and GII.4 ($n = 1$) (Figure 2). The amplified GV virus was omitted from the analyses, as it was previously sequenced (GenBank accession MK291480). One limitation of this system is the acquisition and/or selection of mutations during the RT-PCR amplification step. We have developed a similar system for noroviruses [35], and found that RT-PCR amplification of 1×10^5 genome copies of a synthesized norovirus RNA (length: 7550 nucleotide) resulted in $\leq$10 substitutions/indels arising at $\leq$12% of all reads, with a depth of coverage $\geq$1000 (Tohma et al. unpublished results). Despite the $\leq$12% false-positives derived from RT-PCR, the robustness of this system derived from the high depth of coverage, as compared with the metagenomic approaches that provides low depth of coverage [28]. Additionally, this system provides the opportunity of analyzing multiple samples per NGS run, without compromising the depth of coverage, and thus reducing the cost for large studies. This could be useful for intra-host analyses, as well as to provide accuracy for the consensus genome assembly [28,35].

To gain insight into the mechanisms of sapovirus evolution, we mined all full-length (or nearly full-length) capsid (VP1), partial viral RNA-dependent RNA polymerase (RdRp), and full-length (or nearly full-length) genomes of sapoviruses available in the public repositories. We were able to retrieve over 237 sequences from unique human sapoviruses, 125 corresponding to full-length (or nearly full-length) viral genomes, 219 corresponding to VP1, and 152 corresponding to RdRp. The majority of the public sequence data corresponded to strains detected since 2005 (186/237), and only three sequences corresponded to viruses (genotype GI.1) detected in the 1980s (Table 1). Thus, the viruses sequenced in this study provided the oldest sequence information for human sapoviruses. This is very important, as the lack of sequence information on historical viruses could lead to incorrect interpretation of the data [45], e.g., that the "oldest" sapovirus strain, GIII/Cowden/1979, detected in porcine could be the common ancestor of all sapoviruses [46]. The cleavage sites and size of the proteins from full-length genome dataset were predicted based on multiple alignment with a GII.2/Mc10/00 (AY237420) strain [47]. The cleavage sites on NS1/NS2 and NS6-NS7/VP1 were completely conserved, while others presented variation by genotypes (Table 2). The size of the proteins by genotypes are summarized in Table 3. Nonstructural and VP2 proteins showed a similar number of residues, while VP1 proteins showed a difference in its size (range 549–569 amino acids). To investigate the evolutionary pattern of human sapoviruses, we conducted phylogenetic and sequence analyses from five genotypes (GI.1, GI.2, GII.1, GII.3, and GIV.1) that presented $\geq$20 sequences of VP1 and/or RdRp. All five genotypes showed limited intra-genotype diversification, with mean nucleotide p-distance of 0.034–0.089 (RdRp) and 0.033–0.069 (VP1), and strong clock-like signals (evidenced by root-to-tip linear regression $R^2 \geq 0.56$ for

RdRp and $R^2 \geq 0.65$ for VP1, Figure 3a,b, respectively), with a rate of nucleotide evolution ranging from 1.32×10^{-3} to 3.38×10^{-3} nucleotide substitutions/site/year (Table 4). The accumulation of amino acid mutations in VP1 was detected for GI.2 and GIV.1 viruses (≤40 amino acid mutations), while minimal changes, ≤5 amino acid mutations over 20 years, were observed in VP1 from GI.1, GII.1, and GII.3 viruses (Figure 4). Despite showing accumulation of mutations, GI.2 and GIV.1 sapoviruses showed only <7% within-genotype differences in VP1 in last 20–40 years (Figure 4). Overall diversity within the VP1 protein showed that N-terminal region and P2 subdomain are the two of the most variable regions (Figure 5) [48]. This is in contrast to human noroviruses (Figure 5) and other caliciviruses [48], which present a conserved N-terminal region and a highly variable P2 subdomain.
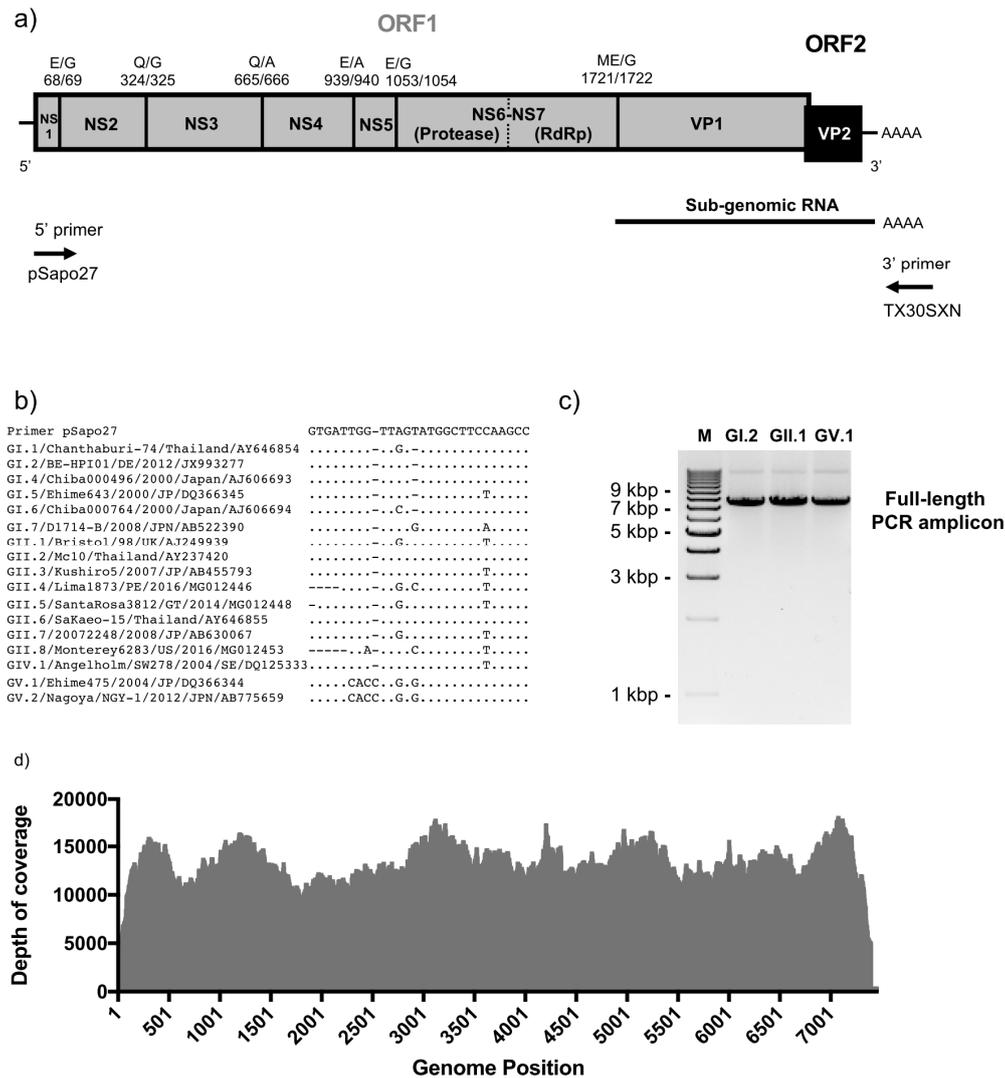

Figure 1. Full-length sapovirus genome PCR amplification and sequencing. (**a**) Sapovirus genome structure and primers pSapo27 and TX30SXN designed for one-step RT-PCR. Open reading frames (ORFs) and predicted cleavage sites are indicated on top. Amino acid positions are based on GI.1/Manchester/93 (X86560) strain (**b**) Alignment of 5'-primer (pSapo27) with human sapovirus genomes available in public repositories. (**c**) Agarose gel electrophoresis of full-length genome PCR amplicons of sapoviruses representing each of the genogroups GI, GII and GV. (**d**) Graph showing the depth of reads coverage in nearly-full length sapovirus genome sequences obtained from one of the archival samples (C102) in our study.
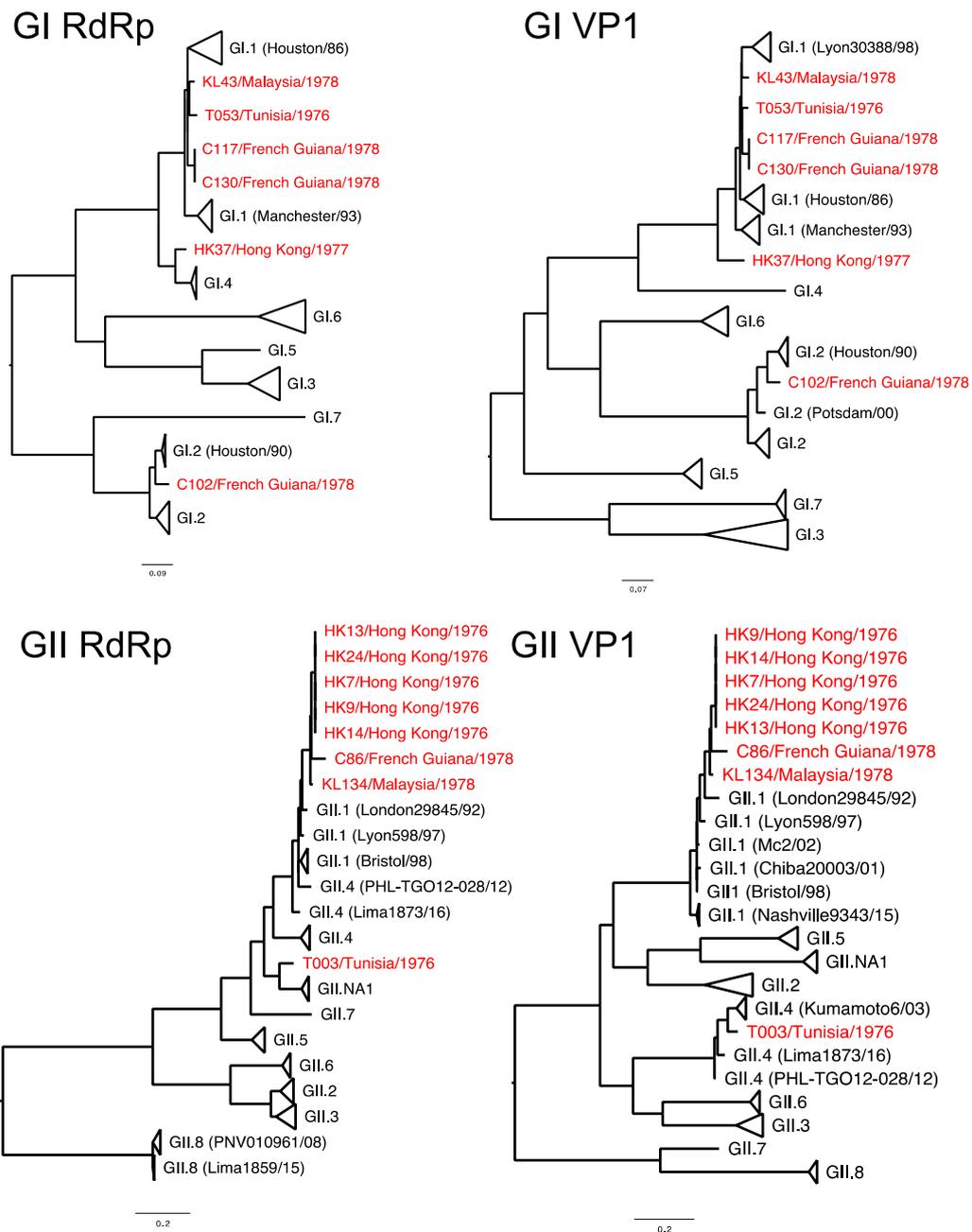
Figure 2. Phylogenetic tree of sapovirus GI and GII strains, circulating since 1976. The maximum likelihood phylogenetic trees were constructed using Kimura-2 parameter model with bootstrap values of ≥70. Samples analyzed in this study are highlighted in red. Sequences for partial RdRp (viral polymerase, >650 nucleotide length) and nearly full-length VP1 (≥1575 nucleotide length) were separately analyzed for genotyping and detection of possible recombinants.

Table 1. The number of sapovirus sequences newly obtained or retrieved from public repositories for this study, for each genotype on the corresponding year of their detection.

Year	GI.1	GI.2	GI.3	GI.4	GI.5	GI.6	GI.7	GII.1	GII.2	GII.3	GII.4	GII.5	GII.6	GII.7	GII.8	GII.NA	GIV.1	GV.1	GV.2
1976	1 *							5 *			1 *								
1977	1 *																		
1978	3 *	1 *						2 *											
1982	2																		
1986	1																		
1990		1							1										
1992								1											
1993	1																		
1994		1																	
1995																		1	
1997								1											
1998	1							1											
1999									1	2							3		
2000		1		1	2	1			1			1					2		
2000/01	1																		
2001	2		1					1										1	
2002								1									1		
2002/03	2																	1	
2003											1						1		
2004	2	1		1			1			2			1				2	1	
2005	1	3		1		2				2						1			
2006			1		1				1										
2007	1	1								1							13		
2008	7	1	1			1	1			5	1		2	1	1	1	7		
2009	1	2								1					2				
2010		1								2		1					1	1	
2011											2								
2012	1	5						1			1								1
2013	4	8							1	1									
2014	8							4	2	2		1			3		10	2	
2015	2	1				1		4	1	7		3			1		5	5	
2016	3	10							1	1		1			3		7		2
2018																	1		
Total	45	37	3	3	3	5	2	21	9	26	7	7	3	1	10	2	51	13	3

* Sequences obtained in this study.

Table 2. Number of sequences by patterns of cleavage sites on the ORF1-encoding proteins.

Genotype	NS1/NS2		NS2/NS3		NS3/NS4				NS4/NS5			NS5/NS6-NS7			NS6-NS7/VP1
	E/G	E/A	E/G	E/S	Q/G	Q/A	Q/G	Q/S	E/A	E/G	Q/A	E/A	E/G	E/S	ME/G
GI.1	23			23	23				23			8	8	7	23
GI.2	11			11				11	10	1		9	2		11
GI.3	2			2				2	2				2		2
GI.4	1			1			1		1			1			1
GI.5	1			1				1	1				1		1
GI.6	3		3					3	3			3			3
GI.7	1			1			1			1		1			1
GII.1	17			17				17	17			17			17
GII.2	6			6				6	6			6			6
GII.3	12			12				12	12			11			12
GII.4	5			5				5	5			5			5
GII.5	6			6				6	6			6			6
GII.6	2			2				2	2			2			2
GII.7	1			1				1	1			1			1
GII.8	8			8				8		8		8			8
GII.NA1	2			2				2	2			2			2
GIV.1	27			27				27	27			27			27
GV.1	8	8						8			8	8			8
GV.2	3				3				3			3	3			3

Table 3. Length of nonstructural and structural proteins [†].

Genotype	ORF1							ORF2
	NS1	NS2	NS3	NS4	NS5	NS6-NS7	VP1	VP2
GI.1	68	256	341	274	114	668 [*]	559	166
GI.2	67	256	341	274	115	668	569	164
GI.3	68	256	341	274	114	668	564	166
GI.4	68	256	341	274	114	668	559	166
GI.5	68	256	341	274	114	668	565	166
GI.6	67	256	341	274	114	668	563	166
GI.7	67	256	341	274	115	668	566	166
GII.1	69	256	341	274	115	667	558	167
GII.2	69	256	341	274	115	667	556	167
GII.3	69	256	341	274	115	667	559	167
GII.4	69	256	341	274	115	667	557	167
GII.5	69	256	341	274	115	667	557	167
GII.6	69	256	341	274	115	667	559	167
GII.7	69	256	341	274	115	667	556	167
GII.8	70	256	341	275	115	667	554	167
GII.NA1	69	256	341	274	115	667	557	167
GIV.1	69	256	341	274	115	667	549	168
GV.1	67	257	341	279	123	667	567	167
GV.2	65	257	341	279	123	667	567 [#]	168

[†] Length is based on the number of amino acids. [*] HM002617/GI.1/Sapporo_MT-2010 strain has 669 amino acids.
[#] AB775659/GV.2/NGY-1 strain has 569 amino acids.
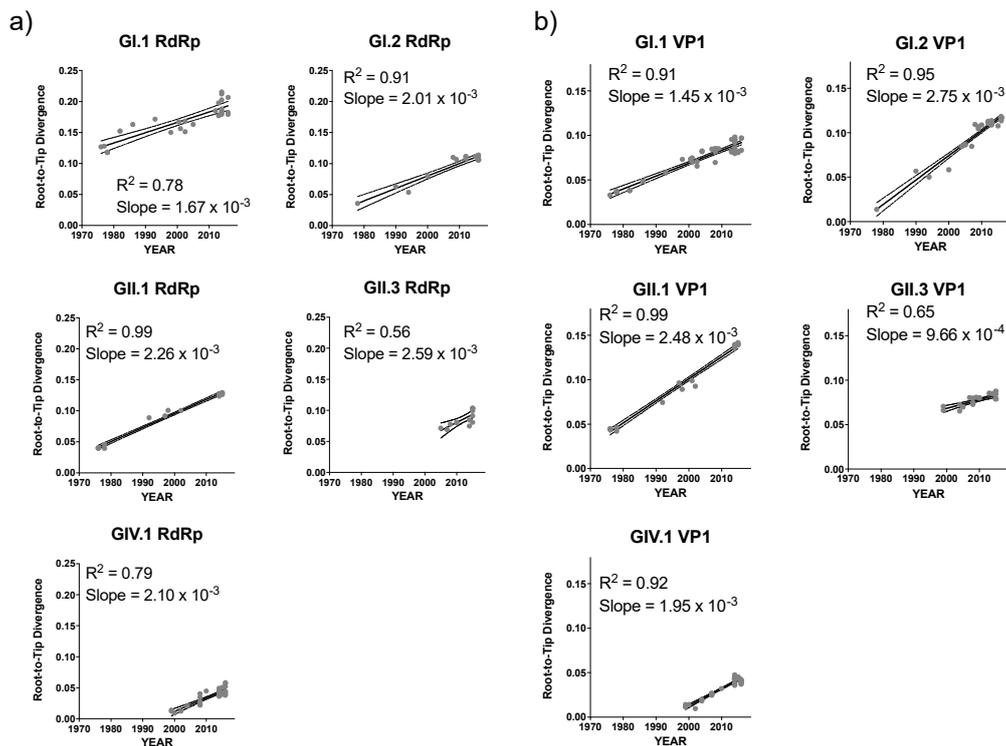
Figure 3. The time-ordered genetic divergence pattern of the RdRp- and VP1-encoding nucleotide sequences of major sapovirus genotypes. Root-to-Tip linear regression analyses were performed to investigate the association between genetic divergence of the (**a**) RdRp- and (**b**) VP1-encoding nucleotide sequences and collection years. The x-axis indicates the collection year and the y-axis shows the root-to-tip divergence on the maximum-likelihood tree. Linear lines indicate the best-fit linear regression of root-to-tip divergence, based on the collection years. Strains are represented by grey dots.

Table 4. Evolutionary rate of the RdRp- and VP1-encoding nucleotide sequences for major five genotypes of human sapoviruses.

Genotype	Mean Substitution Rate, Subs/Site/Year (95% HPD Interval)	
	RdRp	VP1
GI.1	$2.25 \ (1.82\text{–}2.69) \times 10^{-3}$	$1.38 \ (1.14\text{–}1.65) \times 10^{-3}$
GI.2	$3.38 \ (2.60\text{–}4.16) \times 10^{-3}$	$1.32 \ (0.95\text{–}1.70) \times 10^{-3}$
GII.1	$2.90 \ (2.33\text{–}3.48) \times 10^{-3}$	$2.75 \ (2.21\text{–}3.29) \times 10^{-3}$
GII.3	$2.82 \ (1.54\text{–}4.11) \times 10^{-3}$	$2.31 \ (1.69\text{–}2.94) \times 10^{-3}$
GIV.1	$2.82 \ (2.29\text{–}3.36) \times 10^{-3}$	$2.00 \ (1.50\text{–}2.51) \times 10^{-3}$

HPD: highest posterior density.
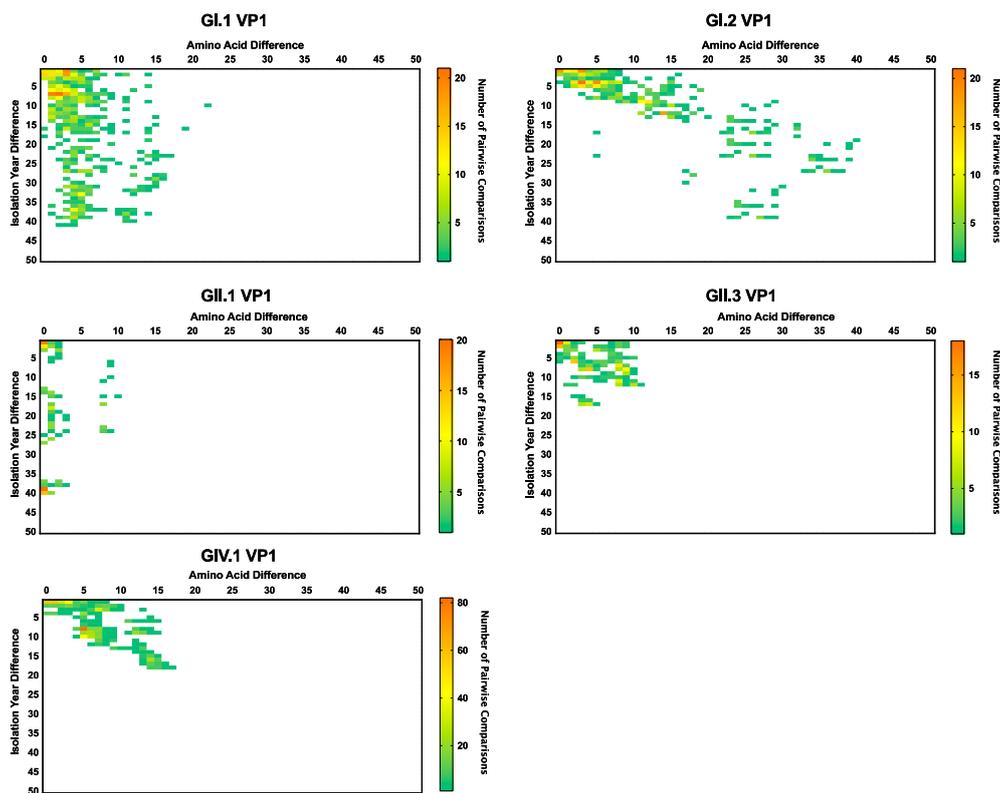
Figure 4. Temporal amino acid diversity patterns of VP1 capsid protein from major sapovirus genotypes. Comparison of patterns reveal differences in VP1 diversification. Over a 40-year period, GI.1 strains differ by 15-20 amino acids, whereas GI.2 strains differ by up to 40 amino acids. The heat map represents the number of pairwise comparisons among available sequences, red being the highest and green the lowest number of pairwise comparisons.

When analyzing the full-length ORFs of all the human sapovirus strains, amino acid mutations were equally distributed among the proteins, but were relatively higher at the NS1, followed by VP1 and VP2 proteins (Figure 6a, $p < 0.05$ in NS1 vs. all the other proteins except VP2, $p < 0.05$ in VP1 vs. NS1–7, and VP2 vs. NS2-7, from a one-way ANOVA with Tukey's post-hoc multiple comparison test). The ratio of sites with entropy > 0 (i.e., sites with ≥1 amino acid mutation(s)) was calculated for each protein from the five genotypes analyzed here. NS1 still presented the higher number of mutations at genotype level (Figure 6b). GI.2 and GIV.1 viruses presented a slightly higher number of mutations on VP1 as compared with other genotypes, as was expected from their evolving pattern (Figure 4), but, overall, mutations were very limited.
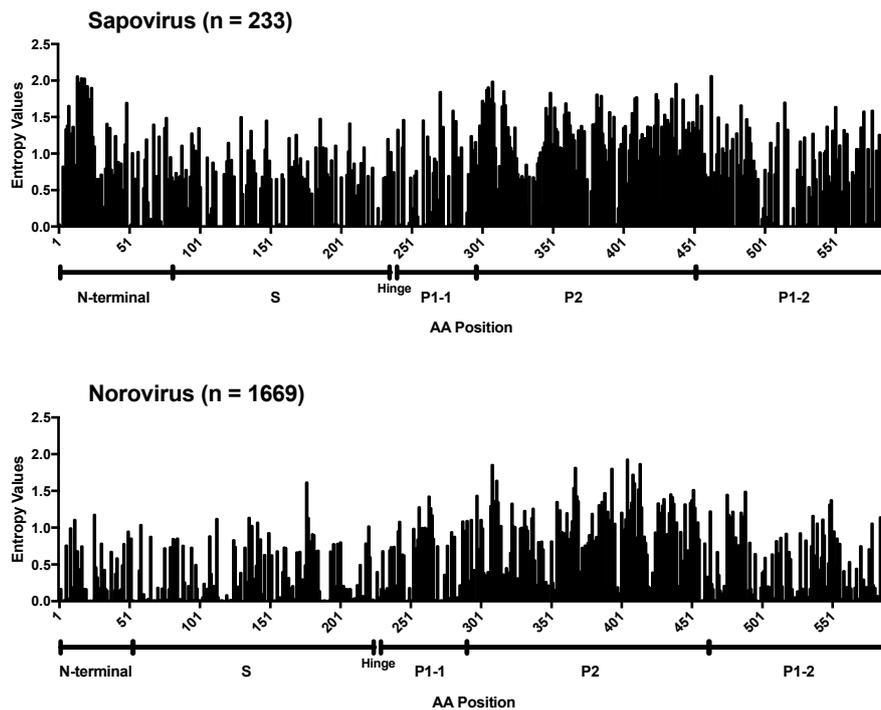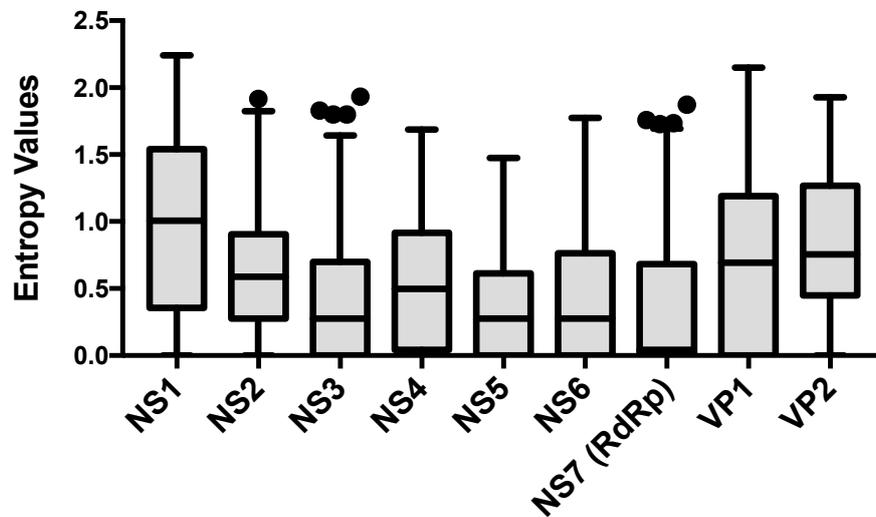
Figure 5. Conservation analyses of the VP1 capsid protein from sapovirus. Shannon entropy values were calculated to quantify the amino acid variation for the capsid protein VP1 from human sapovirus GI, GII, GIV, and GV strains. Sapovirus dataset includes human sapovirus GI, GII, GIV, and GV strains (*n* = 233). Norovirus dataset includes human norovirus GI and GII strains (*n* = 1669). Position of amino acid residues was based on the multiple sequence alignments including indels. Structural domains were assigned based on the sapovirus strain GI.1/Mc114 (AY237422) [3] and norovirus strain GI.1/8FIIa (M87661) [49].

The contrast on the evolutionary dynamics between norovirus and sapovirus is noteworthy. While different norovirus genotypes could present multiple variants [35]; sapoviruses showed limited diversification within the genotypes in both nucleotide and amino acid level (Figures 3 and 4), and none of them presented defined variants [1]. In addition, noroviruses presented a single dominant genotype, GII.4, that showed the chronological-emergence of variants; however, no globally dominant genogroups/genotypes has been reported in sapoviruses [50–57]. Thus, although the sapovirus is prevailing in low- and middle-income countries [56,58], limitations on their diversification might restrict their overall prevalence in the human population, as compared to norovirus [52,57,59–61].

Differences in the phylogenetic clustering was detected between RdRp and VP1 sequences for two of the viruses, HK37/Hong Kong/1977 and T0003/Tunisia/1976, that we sequenced (Figure 2), suggesting potential recombinant strains. We analysed the similarity across the genome of those viruses and representative strains from each of the clusters, and noticed the lack of substantial cross-similarity of the putative recombinant strains with potential parental strains (Figure 7a). We then analysed other recombinant strains reported in the literature [6,9], and found similar patterns on their similarity across the genome with the putative parental strains (Figure 7a). Thus, for example, the GIV strain Ehime1107, which was previously described as recombinant [6], shows the highest similarity (~68% in average) with GII strains at NS proteins, but major differences (0–40%) in the similarity of VP1 and VP2 with all strains (Figure 7a). Moreover, the GIII.3 strain p2, isolated from a piglet and also reported as a recombinant [9], showed high similarity with both parental strains (GIII.2/JJ59 and GIII.3/CH430) at NS proteins, but differences on the similarity at VP1 and VP2 against those parental strains (Figure 7a). Because of the recurrent lack of both parental strains, we examined the inter-genotype/genogroup nucleotide substitution differences among the two genomic regions, i.e., RdRp and VP1, and found that the VP1-encoding region presented a higher number of nucleotide differences between the different

genotypes/genogroups (Figure 7b). These differences on the number of substitutions and the lack of substantial cross-over of those putative recombinant strains suggest that recombination events reported for sapoviruses might be the result of differential phylogenetic clustering. False-positives of recombination derived from discordance of tree topologies among genes were similarly addressed in the phylogeny of mammals [62], animal mitochondrial DNA [63], influenza viruses [64], as well as tick-borne encephalitis viruses [65]. Recombination between the RdRp- and VP1-encoding genes has been extensively shown for noroviruses, and proposed to play a major role in the emergence of novel strains [14,66,67]. In sapoviruses, the RdRp and VP1 proteins are encoded in the same ORF1, while in noroviruses, those proteins are encoded by two different ORFs [15]; ORF1 and ORF2, respectively. The high frequency of RdRp-VP1 recombinant norovirus strains has been explained by the fact that the ORF1/2 junction region is highly conserved across norovirus strains, which facilitates template switching between the genomic and/or sub-genomic RNA that encodes VP1 and VP2 [67]. However, while a sub-genomic RNA that includes VP1- and VP2-encoding regions has been detected for sapovirus [68], the RdRp-VP1 junction region of sapovirus showed less conservation compared to noroviruses. These differences were mostly present at the N-terminal encoding region of VP1, as the 5' end of the sub-genomic RNA was highly conserved in both viruses (Figure 8). Although most DNA and positive-stranded RNA viruses are regarded as prone for recombination [69,70], restrictions on recombination due to differences in the replication mechanisms have been reported for different viruses. Thus, a very low signal for homologous recombination was detected for negative stranded RNA viruses [71] and some positive stranded RNA viruses, e.g., West Nile viruses and tick-borne encephalitis viruses [72]. Therefore, differences in genome organization, the less-conserved RdRp-VP1 boundary, and/or replication mechanisms could limit diversification of sapoviruses by means of RdRp-VP1 recombination. While recombinant and putative parental sapovirus strains have been reported for multiple strains [8,11,12], parental donors for viruses have not yet been detected in nature in many cases [6,7,9]. Further surveillance and phylogenetic analyses will be needed to establish whether recombination is a major driver of sapovirus evolution.
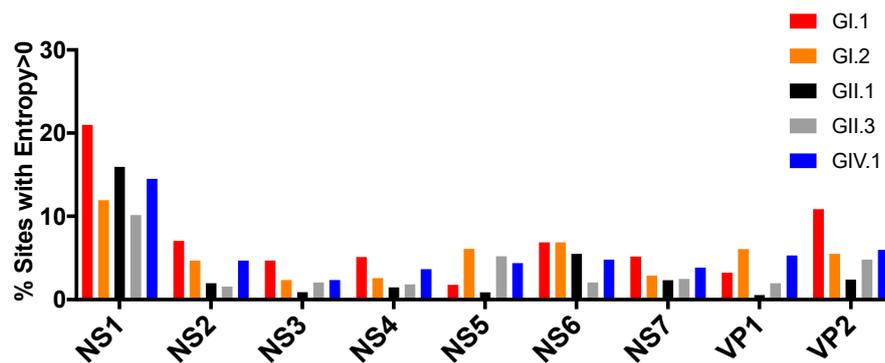
(a)

(b)

Figure 6. Genome conservation analyses of human sapovirus. (**a**) Shannon entropy values were calculated to quantify the amino acid variation for each site in the nonstructural and capsid proteins in ORF1 and 2 from human sapovirus GI, GII, GIV, and GV strains. Only the sequences with complete length available were included. Boxes represent 25th and 75th percentiles, and whiskers represent 1.5× interquartile ranges from the box. Outliers outside the whiskers were shown by circles. (**b**) Shannon entropy values were calculated for major sapovirus genotypes, and summarized as a ratio (%) of sites with entropy > 0 (i.e., sites with ≥1 mutation(s)) at each of the proteins. Genomic regions were predicted based on the sapovirus strain GI.1/Manchester (X86560) [1].
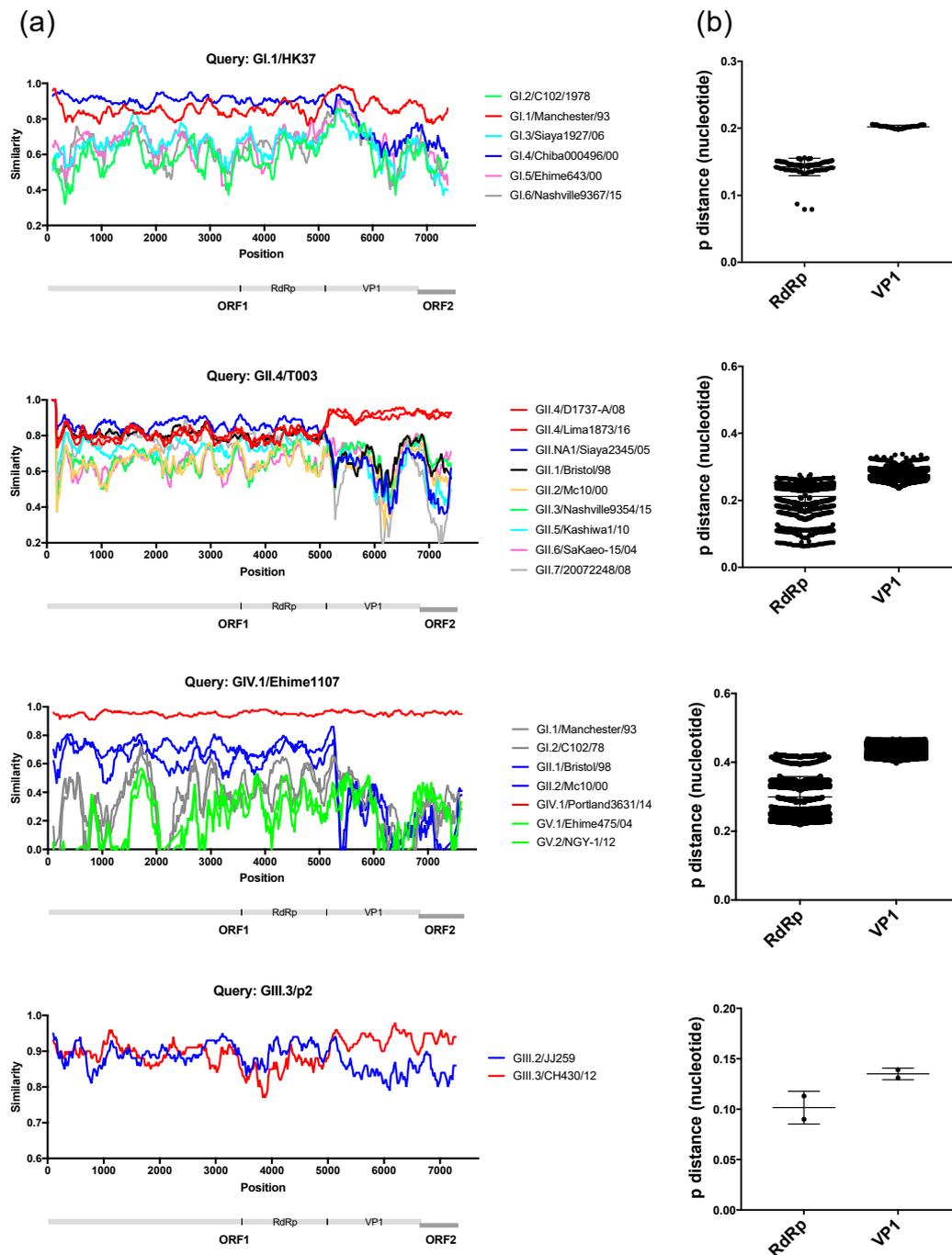
Figure 7. Genomic analyses of putative recombinant sapoviruses present limited evidence of recombination at the RdRp-VP1 boundaries. (**a**) Site-by-site nucleotide similarity analyses of four strains (GI.1/HK7 [MN794208], GII.4/T003 [MN794218], GIV.1/Ehime1107 [DQ058829] [6], and GIII.3/p2 [KX688107] [9]) that presented differences in the phylogenetic clustering when using RdRp- or VP1-encoding regions. (**b**) Inter-genotype/genogroup nucleotide substitution differences among the two genomic regions; RdRp and VP1. The VP1-encoding region presented a higher number of substitutions between the different genotypes/genogroups. Note the lack of both parental strains for the putative recombinant strains.
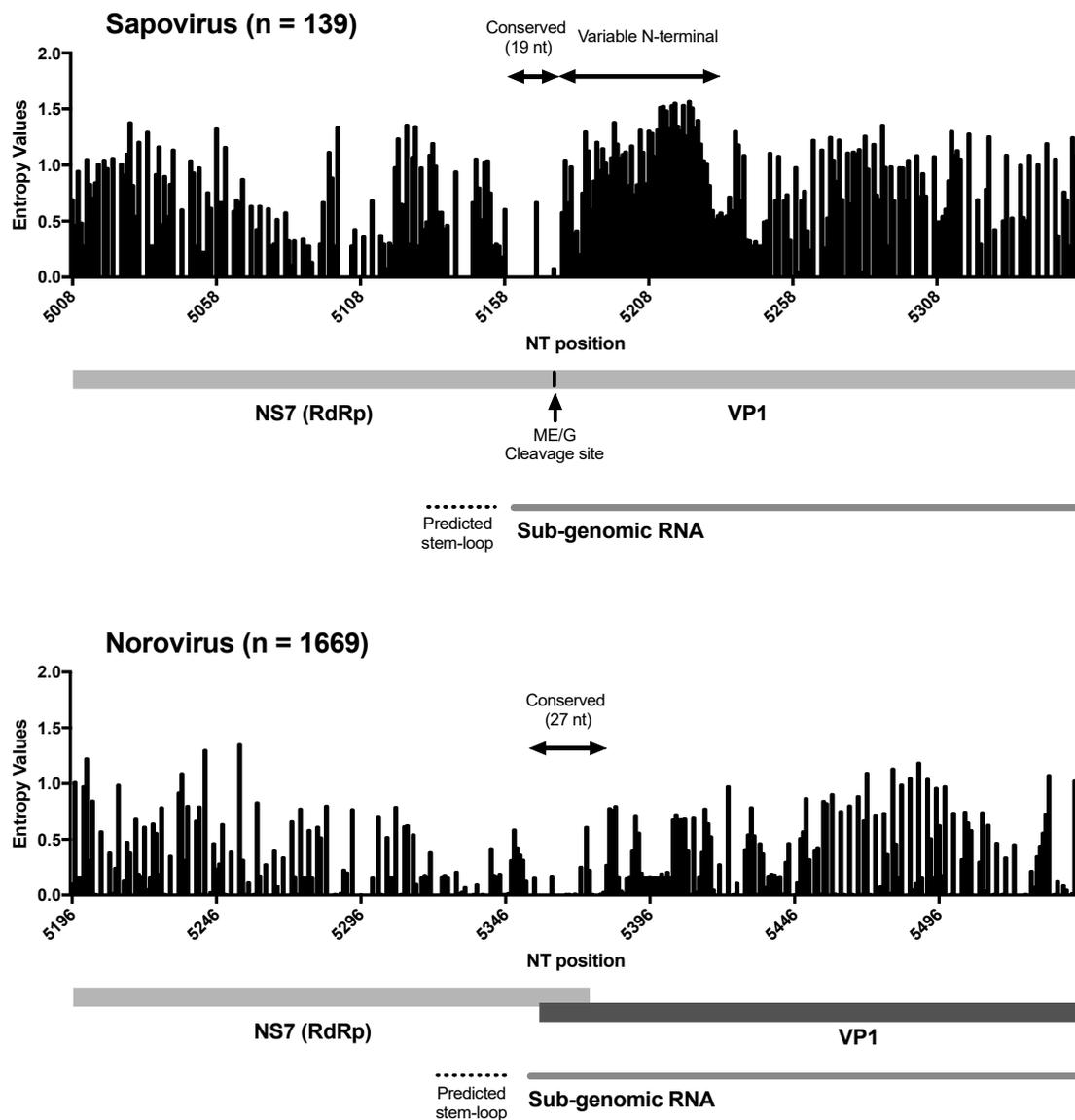
Figure 8. Variation at the RdRp-VP1 boundaries could explain limited recombination in human sapovirus. Shannon entropy values were calculated to quantify the nucleotide conservation for each site in the RdRp-VP1 boundaries. Sapovirus dataset includes human sapovirus GI, GII, GIV, and GV strains (*n* = 139), and norovirus dataset includes human norovirus GI and GII strains (*n* = 1669). Position of nucleotides was based on the multiple sequence alignments, including indels. Stem-loop sequences on predicted sub-genomic RNA promoter and sub-genomic RNA was determined based on analyses from Simmonds et al. [73].

This study has filled historical gaps in the sapovirus sequence database, as well as provided a genome sequencing platform that can be easily adapted and implemented toward the studies of virus evolution, intra-host dynamics in chronically infected individuals, tracking, traceback and the intervention of clinical and foodborne/waterborne outbreaks of illness due to sapovirus [74]. Sapoviruses present limited intra-genotype diversification by means of amino acid mutations over 20–40 years. Moreover, we presented evidence that human sapoviruses may be less prone to recombination at the RdRp-VP1 boundary as compared to the noroviruses. Understanding the mechanisms of human sapovirus diversification would provide valuable information on the natural history of sapovirus infection that could be used to develop better strategies of prevention and control.

References

1. Oka, T.; Wang, Q.; Katayama, K.; Saif, L.J. Comprehensive review of human sapoviruses. *Clin. Microbiol. Rev.* **2015**, *28*, 32–53. [CrossRef] [PubMed]

2. Clarke, I.N.; Lambden, P.R. Organization and expression of calicivirus genes. *J. Infect. Dis.* **2000**, *181* (Suppl. 2), S309–S316. [CrossRef]

3. Miyazaki, N.; Taylor, D.W.; Hansman, G.S.; Murata, K. Antigenic and Cryo-Electron Microscopy Structure Analysis of a Chimeric Sapovirus Capsid. *J. Virol.* **2015**, *90*, 2664–2675. [CrossRef] [PubMed]

4. Oka, T.; Mori, K.; Iritani, N.; Harada, S.; Ueki, Y.; Iizuka, S.; Mise, K.; Murakami, K.; Wakita, T.; Katayama, K. Human sapovirus classification based on complete capsid nucleotide sequences. *Arch. Virol.* **2012**, *157*, 349–352. [CrossRef] [PubMed]

5. Diez-Valcarce, M.; Castro, C.J.; Marine, R.L.; Halasa, N.; Mayta, H.; Saito, M.; Tsaknaridis, L.; Pan, C.Y.; Bucardo, F.; Becker-Dreps, S.; et al. Genetic diversity of human sapovirus across the Americas. *J. Clin. Virol.* **2018**, *104*, 65–72. [CrossRef] [PubMed]

6. Hansman, G.S.; Takeda, N.; Oka, T.; Oseto, M.; Hedlund, K.O.; Katayama, K. Intergenogroup recombination in sapoviruses. *Emerg. Infect. Dis.* **2005**, *11*, 1916–1920. [CrossRef]

7. Katayama, K.; Miyoshi, T.; Uchino, K.; Oka, T.; Tanaka, T.; Takeda, N.; Hansman, G.S. Novel recombinant sapovirus. *Emerg. Infect. Dis.* **2004**, *10*, 1874–1876. [CrossRef]

8. Kuroda, M.; Masuda, T.; Ito, M.; Naoi, Y.; Doan, Y.H.; Haga, K.; Tsuchiaka, S.; Kishimoto, M.; Sano, K.; Omatsu, T.; et al. Genetic diversity and intergenogroup recombination events of sapoviruses detected from feces of pigs in Japan. *Infect. Genet. Evol.* **2017**, *55*, 209–217. [CrossRef]

9. Li, J.; Shen, Q.; Zhang, W.; Zhao, T.; Li, Y.; Jiang, J.; Yu, X.; Guo, Z.; Cui, L.; Hua, X. Genomic organization and recombination analysis of a porcine sapovirus identified from a piglet with diarrhea in China. *Virol. J.* **2017**, *14*, 57. [CrossRef]

10. Kumthip, K.; Khamrin, P.; Ushijima, H.; Chen, L.; Li, S.; Maneekarn, N. Genetic recombination and diversity of sapovirus in pediatric patients with acute gastroenteritis in Thailand, 2010–2018. *PeerJ* **2020**, *8*, e8520. [CrossRef]

11. Liu, X.; Yamamoto, D.; Saito, M.; Imagawa, T.; Ablola, A.; Tandoc, A.O., 3rd; Segubre-Mercado, E.; Lupisan, S.P.; Okamoto, M.; Furuse, Y.; et al. Molecular detection and characterization of sapovirus in hospitalized children with acute gastroenteritis in the Philippines. *J. Clin. Virol.* **2015**, *68*, 83–88. [CrossRef] [PubMed]

12. Dos Anjos, K.; Lima, L.M.; Silva, P.A.; Inoue-Nagata, A.K.; Nagata, T. The possible molecular evolution of sapoviruses by inter- and intra-genogroup recombination. *Arch. Virol.* **2011**, *156*, 1953–1959. [CrossRef] [PubMed]

13. Wang, Q.H.; Han, M.G.; Funk, J.A.; Bowman, G.; Janies, D.A.; Saif, L.J. Genetic diversity and recombination of porcine sapoviruses. *J. Clin. Microbiol.* **2005**, *43*, 5963–5972. [CrossRef] [PubMed]

14. Bull, R.A.; Tanaka, M.M.; White, P.A. Norovirus recombination. *J. Gen. Virol.* **2007**, *88*, 3347–3359. [CrossRef] [PubMed]

15. Desselberger, U. Caliciviridae Other Than Noroviruses. *Viruses* **2019**, *11*, 286. [CrossRef]

16. Pietsch, C.; Liebert, U.G. Intrahost viral evolution during chronic sapovirus infections. *J. Clin. Virol.* **2019**, *113*, 1–7. [CrossRef]

17. Hergens, M.P.; Nederby Ohd, J.; Alm, E.; Askling, H.H.; Helgesson, S.; Insulander, M.; Lagerqvist, N.; Svenungsson, B.; Tihane, M.; Tolfvenstam, T.; et al. Investigation of a food-borne outbreak of gastroenteritis in a school canteen revealed a variant of sapovirus genogroup V not detected by standard PCR, Sollentuna, Sweden, 2016. *Eurosurveillance* **2017**, *22*. [CrossRef]

18. Pang, X.L.; Lee, B.E.; Tyrrell, G.J.; Preiksaitis, J.K. Epidemiology and genotype analysis of sapovirus associated with gastroenteritis outbreaks in Alberta, Canada: 2004–2007. *J. Infect. Dis.* **2009**, *199*, 547–551. [CrossRef]

19. Johansson, P.J.; Bergentoft, K.; Larsson, P.A.; Magnusson, G.; Widell, A.; Thorhagen, M.; Hedlund, K.O. A nosocomial sapovirus-associated outbreak of gastroenteritis in adults. *Scand. J. Infect. Dis.* **2005**, *37*, 200–204. [CrossRef]

20. Rambaut, A.; Pybus, O.G.; Nelson, M.I.; Viboud, C.; Taubenberger, J.K.; Holmes, E.C. The genomic and epidemiological dynamics of human influenza A virus. *Nature* **2008**, *453*, 615–619. [CrossRef]

21. Fraser, C.; Donnelly, C.A.; Cauchemez, S.; Hanage, W.P.; Van Kerkhove, M.D.; Hollingsworth, T.D.; Griffin, J.; Baggaley, R.F.; Jenkins, H.E.; Lyons, E.J.; et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* **2009**, *324*, 1557–1561. [CrossRef] [PubMed]

22. Lemey, P.; Rambaut, A.; Bedford, T.; Faria, N.; Bielejec, F.; Baele, G.; Russell, C.A.; Smith, D.J.; Pybus, O.G.; Brockmann, D.; et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* **2014**, *10*, e1003932. [CrossRef] [PubMed]

23. Armstrong, G.L.; MacCannell, D.R.; Taylor, J.; Carleton, H.A.; Neuhaus, E.B.; Bradbury, R.S.; Posey, J.E.; Gwinn, M. Pathogen Genomics in Public Health. *N. Engl. J. Med.* **2019**, *381*, 2569–2580. [CrossRef] [PubMed]

24. Gwinn, M.; MacCannell, D.; Armstrong, G.L. Next-Generation Sequencing of Infectious Pathogens. *JAMA* **2019**, *321*, 893–894. [CrossRef]

25. Biek, R.; Pybus, O.G.; Lloyd-Smith, J.O.; Didelot, X. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* **2015**, *30*, 306–313. [CrossRef]

26. Kao, R.R.; Haydon, D.T.; Lycett, S.J.; Murcia, P.R. Supersize me: How whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.* **2014**, *22*, 282–291. [CrossRef]

27. Timme, R.E.; Strain, E.; Baugher, J.D.; Davis, S.; Gonzalez-Escalona, N.; Sanchez Leon, M.; Allard, M.W.; Brown, E.W.; Tallent, S.; Rand, H. Phylogenomic Pipeline Validation for Foodborne Pathogen Disease Surveillance. *J. Clin. Microbiol.* **2019**, *57*. [CrossRef]

28. Houldcroft, C.J.; Beale, M.A.; Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **2017**, *15*, 183–192. [CrossRef]

29. Capobianchi, M.R.; Giombini, E.; Rozera, G. Next-generation sequencing technology in clinical virology. *Clin. Microbiol. Infect.* **2013**, *19*, 15–22. [CrossRef]

30. Djikeng, A.; Halpin, R.; Kuzmickas, R.; Depasse, J.; Feldblyum, J.; Sengamalay, N.; Afonso, C.; Zhang, X.; Anderson, N.G.; Ghedin, E.; et al. Viral genome sequencing by random priming methods. *BMC Genom.* **2008**, *9*, 5. [CrossRef]

31. Brown, J.R.; Roy, S.; Ruis, C.; Yara Romero, E.; Shah, D.; Williams, R.; Breuer, J. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: A Robust and Sensitive Method. *J. Clin. Microbiol.* **2016**, *54*, 2530–2537. [CrossRef] [PubMed]

32. Yang, Z.; Leonard, S.R.; Mammel, M.K.; Elkins, C.A.; Kulka, M. Towards next-generation sequencing analytics for foodborne RNA viruses: Examining the effect of RNA input quantity and viral RNA purity. *J. Virol. Methods* **2016**, *236*, 221–230. [CrossRef] [PubMed]

33. National Institute of Allergy and Infectious Diseases. *Annual Report of Program Activities*; Government Printing Office: Washington, DC, USA, 1979.

34. Katayama, K.; Shirato-Horikoshi, H.; Kojima, S.; Kageyama, T.; Oka, T.; Hoshino, F.; Fukushi, S.; Shinohara, M.; Uchida, K.; Suzuki, Y.; et al. Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **2002**, *299*, 225–239. [CrossRef] [PubMed]

35. Parra, G.I.; Squires, R.B.; Karangwa, C.K.; Johnson, J.A.; Lepore, C.J.; Sosnovtsev, S.V.; Green, K.Y. Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity. *PLoS Pathog.* **2017**, *13*, e1006136. [CrossRef]

36. Simonyan, V.; Chumakov, K.; Dingerdissen, H.; Faison, W.; Goldweber, S.; Golikov, A.; Gulzar, N.; Karagiannis, K.; Vinh Nguyen Lam, P.; Maudru, T.; et al. High-performance integrated virtual environment (HIVE): A robust infrastructure for next-generation sequence data analysis. *Database (Oxford)* **2016**, *2016*. [CrossRef]

37. Tohma, K.; Saito, M.; Mayta, H.; Zimic, M.; Lepore, C.J.; Ford-Siltz, L.A.; Gilman, R.H.; Parra, G.I. Complete Genome Sequence of a Nontypeable GII Norovirus Detected in Peru. *Genome Announc.* **2018**, *6*. [CrossRef]

38. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]

39. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]

40. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2016**, *2*, vew007. [CrossRef]

41. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [CrossRef]

42. Shapiro, B.; Rambaut, A.; Drummond, A.J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **2006**, *23*, 7–9. [CrossRef] [PubMed]

43. Kuraku, S.; Zmasek, C.M.; Nishimura, O.; Katoh, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* **2013**, *41*, W22–W28. [CrossRef] [PubMed]

44. Lole, K.S.; Bollinger, R.C.; Paranjape, R.S.; Gadkari, D.; Kulkarni, S.S.; Novak, N.G.; Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **1999**, *73*, 152–160. [CrossRef] [PubMed]

45. Tohma, K.; Lepore, C.J.; Ford-Siltz, L.A.; Parra, G.I. Evolutionary dynamics of non-GII genotype 4 (GII.4) noroviruses reveal limited and independent diversification of variants. *J. Gen. Virol.* **2018**, *99*, 1027–1035. [CrossRef] [PubMed]

46. Barry, A.F.; Duraes-Carvalho, R.; Oliveira-Filho, E.F.; Alfieri, A.A.; Van der Poel, W.H.M. High-resolution phylogeny providing insights towards the epidemiology, zoonotic aspects and taxonomy of sapoviruses. *Infect. Genet. Evol.* **2017**, *56*, 8–13. [CrossRef] [PubMed]

47. Oka, T.; Yamamoto, M.; Katayama, K.; Hansman, G.S.; Ogawa, S.; Miyamura, T.; Takeda, N. Identification of the cleavage sites of sapovirus open reading frame 1 polyprotein. *J. Gen. Virol.* **2006**, *87*, 3329–3338. [CrossRef]

48. Okada, M.; Yamashita, Y.; Oseto, M.; Ogawa, T.; Kaiho, I.; Shinozaki, K. Genetic variability in the sapovirus capsid protein. *Virus Genes* **2006**, *33*, 157–161. [CrossRef]

49. Prasad, B.V.; Hardy, M.E.; Dokland, T.; Bella, J.; Rossmann, M.G.; Estes, M.K. X-ray crystallographic structure of the Norwalk virus capsid. *Science* **1999**, *286*, 287–290. [CrossRef]

50. Gallimore, C.I.; Iturriza-Gomara, M.; Lewis, D.; Cubitt, D.; Cotterill, H.; Gray, J.J. Characterization of sapoviruses collected in the United Kingdom from 1989 to 2004. *J. Med. Virol.* **2006**, *78*, 673–682. [CrossRef]

51. Johnsen, C.K.; Midgley, S.; Bottiger, B. Genetic diversity of sapovirus infections in Danish children 2005–2007. *J. Clin. Virol.* **2009**, *46*, 265–269. [CrossRef]

52. Kumthip, K.; Khamrin, P.; Maneekarn, N. Molecular epidemiology and genotype distributions of noroviruses and sapoviruses in Thailand 2000-2016: A review. *J. Med. Virol.* **2018**, *90*, 617–624. [CrossRef] [PubMed]

53. Lasure, N.; Gopalkrishna, V. Epidemiological profile and genetic diversity of sapoviruses (SaVs) identified in children suffering from acute gastroenteritis in Pune, Maharashtra, Western India, 2007-2011. *Epidemiol. Infect.* **2017**, *145*, 106–114. [CrossRef] [PubMed]

54. Mann, P.; Pietsch, C.; Liebert, U.G. Genetic Diversity of Sapoviruses among Inpatients in Germany, 2008–2018. *Viruses* **2019**, *11*, 726. [CrossRef] [PubMed]

55. Murray, T.Y.; Nadan, S.; Page, N.A.; Taylor, M.B. Diverse sapovirus genotypes identified in children hospitalised with gastroenteritis in selected regions of South Africa. *J. Clin. Virol.* **2016**, *76*, 24–29. [CrossRef]

56. Sanchez, G.J.; Mayta, H.; Pajuelo, M.J.; Neira, K.; Xiaofang, L.; Cabrera, L.; Ballard, S.B.; Crabtree, J.E.; Kelleher, D.; Cama, V.; et al. Epidemiology of Sapovirus Infections in a Birth Cohort in Peru. *Clin. Infect. Dis.* **2018**, *66*, 1858–1863. [CrossRef]

57. Svraka, S.; Vennema, H.; van der Veer, B.; Hedlund, K.O.; Thorhagen, M.; Siebenga, J.; Duizer, E.; Koopmans, M. Epidemiology and genotype analysis of emerging sapovirus-associated infections across Europe. *J. Clin. Microbiol.* **2010**, *48*, 2191–2198. [CrossRef]

58. Platts-Mills, J.A.; Liu, J.; Rogawski, E.T.; Kabir, F.; Lertsethtakarn, P.; Siguas, M.; Khan, S.S.; Praharaj, I.; Murei, A.; Nshama, R.; et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: A reanalysis of the MAL-ED cohort study. *Lancet Glob. Health* **2018**, *6*, e1309–e1318. [CrossRef]

59. Grant, L.R.; O'Brien, K.L.; Weatherholtz, R.C.; Reid, R.; Goklish, N.; Santosham, M.; Parashar, U.; Vinje, J. Norovirus and Sapovirus Epidemiology and Strain Characteristics among Navajo and Apache Infants. *PLoS ONE* **2017**, *12*, e0169491. [CrossRef]

60. Hassan, F.; Kanwar, N.; Harrison, C.J.; Halasa, N.B.; Chappell, J.D.; Englund, J.A.; Klein, E.J.; Weinberg, G.A.; Szilagyi, P.G.; Moffatt, M.E.; et al. Viral Etiology of Acute Gastroenteritis in <2-Year-Old US Children in the Post-Rotavirus Vaccine Era. *J. Pediatric Infect. Dis. Soc.* **2019**, *8*, 414–421. [CrossRef]

61. Torner, N.; Martinez, A.; Broner, S.; Moreno, A.; Camps, N.; Dominguez, A.; Working Group for the Study of Acute Viral Gastroenteritis Outbreaks in Catalonia. Epidemiology of Acute Gastroenteritis Outbreaks Caused by Human Calicivirus (Norovirus and Sapovirus) in Catalonia: A Two Year Prospective Study, 2010–2011. *PLoS ONE* **2016**, *11*, e0152503. [CrossRef]

62. Holmes, E.C. Different rates of substitution may produce different phylogenies of the eutherian mammals. *J. Mol. Evol.* **1991**, *33*, 209–215. [CrossRef]

63. Sun, S.; Evans, B.J.; Golding, G.B. "Patchy-tachy" leads to false positives for recombination. *Mol. Biol. Evol.* **2011**, *28*, 2549–2559. [CrossRef]

64. Worobey, M.; Rambaut, A.; Pybus, O.G.; Robertson, D.L. Questioning the evidence for genetic recombination in the 1918 "Spanish flu" virus. *Science* **2002**, *296*. [CrossRef]

65. Bertrand, Y.J.; Johansson, M.; Norberg, P. Revisiting Recombination Signal in the Tick-Borne Encephalitis Virus: A Simulation Approach. *PLoS ONE* **2016**, *11*, e0164435. [CrossRef]

66. Parra, G.I. Emergence of norovirus strains: A tale of two genes. *Virus Evol.* **2019**, *5*. [CrossRef]

67. Bull, R.A.; Hansman, G.S.; Clancy, L.E.; Tanaka, M.M.; Rawlinson, W.D.; White, P.A. Norovirus recombination in ORF1/ORF2 overlap. *Emerg. Infect. Dis.* **2005**, *11*, 1079–1085. [CrossRef] [PubMed]

68. Chang, K.O.; Sosnovtsev, S.V.; Belliot, G.; Kim, Y.; Saif, L.J.; Green, K.Y. Bile acids are essential for porcine enteric calicivirus replication in association with down-regulation of signal transducer and activator of transcription 1. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8733–8738. [CrossRef]

69. Bentley, K.; Evans, D.J. Mechanisms and consequences of positive-strand RNA virus recombination. *J. Gen. Virol.* **2018**, *99*, 1345–1356. [CrossRef] [PubMed]

70. Perez-Losada, M.; Arenas, M.; Galan, J.C.; Palero, F.; Gonzalez-Candelas, F. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infect. Genet. Evol.* **2015**, *30*, 296–307. [CrossRef] [PubMed]

71. Han, G.Z.; Worobey, M. Homologous recombination in negative sense RNA viruses. *Viruses* **2011**, *3*, 1358–1373. [CrossRef] [PubMed]

72. Taucher, C.; Berger, A.; Mandl, C.W. A trans-complementing recombination trap demonstrates a low propensity of flaviviruses for intermolecular recombination. *J. Virol.* **2010**, *84*, 599–611. [CrossRef] [PubMed]

73. Simmonds, P.; Karakasiliotis, I.; Bailey, D.; Chaudhry, Y.; Evans, D.J.; Goodfellow, I.G. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res.* **2008**, *36*, 2530–2546. [CrossRef] [PubMed]

74. Timme, R.E.; Sanchez Leon, M.; Allard, M.W. Utilizing the Public GenomeTrakr Database for Foodborne Pathogen Traceback. *Methods Mol. Biol.* **2019**, *1918*, 201–212. [CrossRef] [PubMed]