Genomic Analyses of Human Sapovirus Reveal Disparate Patterns of Evolution Among Genotypes and Genome Regions

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

Division of Molecular Biology, CFSAN, FDA[1], Laurel, MD, USA; Division of Viral Products, CBER, FDA[2], Silver Spring, MD, USA; National Virus Reference Laboratory, UC-Dublin[3], Dublin, Ireland; Laboratory of Infectious Diseases, NIH[4], Bethesda, MD, USA.

Aims

- To develop a simple (cost-efficient) genome sequencing platform for human sapoviruses
- To investigate the evolutionary dynamics of human sapoviruses by characterizing archival samples

Abstract

Sapovirus is a genetically diverse genus of the Caliciviridae family. Human sapovirus is a causative agent of acute gastroenteritis and spread occurs by person-to-person contact and/or contaminated water, soil, or food. Epidemiologic tracking and evolutionary study are enhanced when using full, versus short, genome sequences. In this investigation we sought to develop a full-genome sequencing platform for human sapovirus genomic studies. Stools positive for sapovirus were tested for single-amplicon full-length genome RT-PCR using primers annealing at the 5'-/3'-end of the viral genome. Amplicons were sequenced using the Illumina next-generation-sequencing platform. A local data base containing >180 sequences representing five genotypes (≥25 sequences and spanning ~40 years) was created for analyses of coding regions for viral polymerase (vPOL) and capsid (VP1). Phylogenetic trees and diversity plots were generated using different algorithm/analytics packages. The newly developed full-length RT-PCR platform was very robust for the genome amplification of GI, GII, and GV human sapoviruses. Thus, 14/15 sapovirus-positive samples collected during 1976-78 were positive for full genome amplicon and sequenced. These viruses provide the oldest sequence information for sapoviruses. Phylogenetic and sequence analyses from five genotypes (GI.1, GI.2, GII.1, GII.3, and GIV.1) showed limited intra-genotype diversification and a rate of nucleotide (nt) evolution ranging from 1.3-3.4×10-3 nt substitutions/site/year. Differences in the phylogenetic clustering was detected between vPOL and VP1 sequences, but site-by-site similarity among genotypes/genogroups suggested false-positive signal for recombination. Accumulation of amino acid mutations in VP1 was detected for GI.2 and GIV.1 viruses, while similar in rate of nt evolution to the other genotypes. Differences on the genetic robustness of the VP1 was detected among sapovirus genotypes; with GI.2 and GIV.1 presenting higher flexibility. Contrary to noroviruses, sapoviruses present limited diversification by means of recombination. The full-genome platform presented could facilitate tracking and intervention of sapovirus outbreaks, and traceback investigation of foodborne illness.

FDA Mission Relevance

The research presents a full-genome platform approach to the sequence identification and subsequent genetic and amino acid comparative analyses of sapovirus genogroups, an emerging group of caliciviruses associated with outbreaks of human acute gastroenteritis and linked to food/water borne illnesses. The workflow described may provide an important approach toward the tracking and intervention of sapovirus outbreaks, and traceback investigation of foodborne illness. Understanding the mechanisms of human sapovirus diversification could provide valuable information on the natural history of sapovirus infection, evolution, and virus/disease transmission.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration (FDA), or the U.S. Government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the FDA.

Universal genome sequencing platform

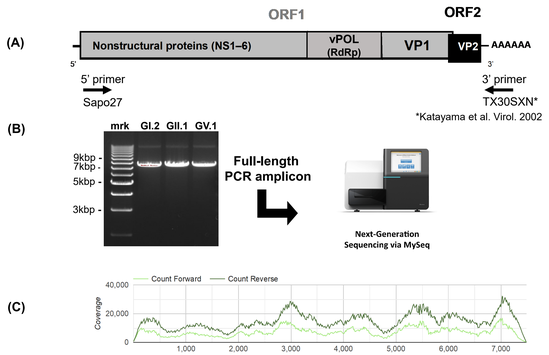


Figure 1. (A) Human sapovirus genome structure and primers designed for RT and PCR. vPOL is also known as RdRp. (B) Agarose gel electrophoresis of full-length amplicons representing genogroups GI, GII, and GV. Amplicons are excised and purified prior to whole genome sequencing. (C) Typical depth of read coverage for both strands of an amplicon.

Genomic sequences from archival samples

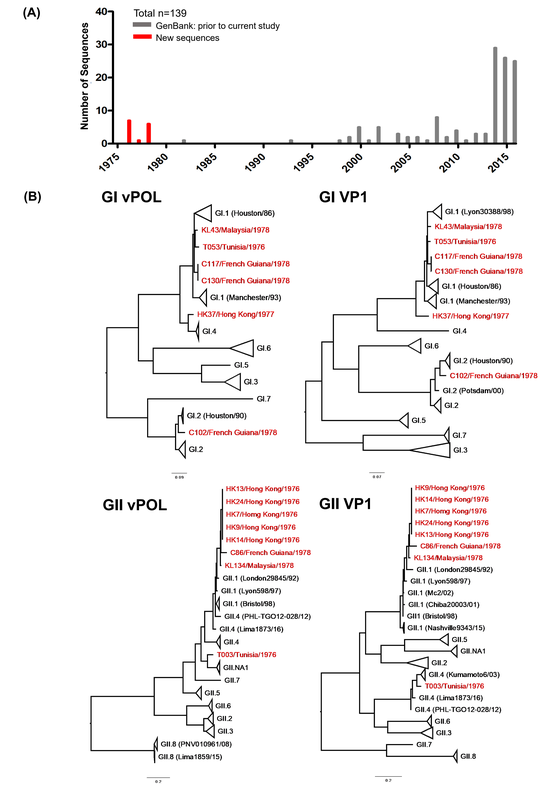
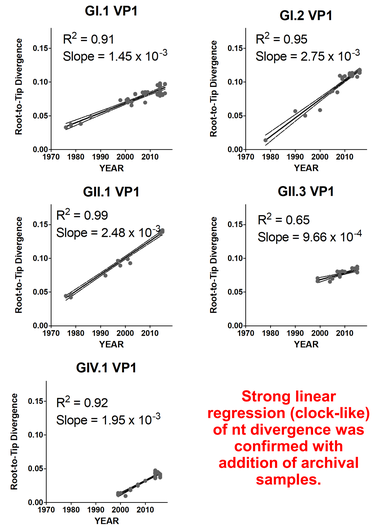


Figure 2. (A) Year assignment for sequences used in this study. (B) Phylogenetic tree of human sapovirus strains GI and GII, circulating since 1976. Study samples are highlighted in red, sequences for vPOL and VP1 were analyzed separately for genotyping and detection of possible recombinants.

Linear evolution of VP1-encoding nucleotide sequences



Strong linear regression (clock-like) of nt divergence was confirmed with addition of archival samples.

Figure 3. Root-to-Tip linear regression analyses to investigate the association between genetic divergence of the VP1-encoding sequences and collection years, for human sapovirus genotypes GI.1, GI.2, GII.1, GII.3, and GIV.1.

Similar rate of evolution among human sapovirus genotypes

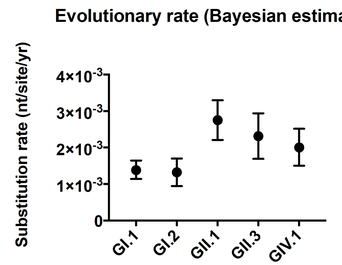


Figure 4. Mean substitution rates of the VP1-encoding nucleotide sequences for human sapovirus genotypes GI.1, GI.2, GII.1, GII.3, and GIV.1.

Amino acid diversity of VP1 reveals different patterns of diversification in sapovirus genotypes

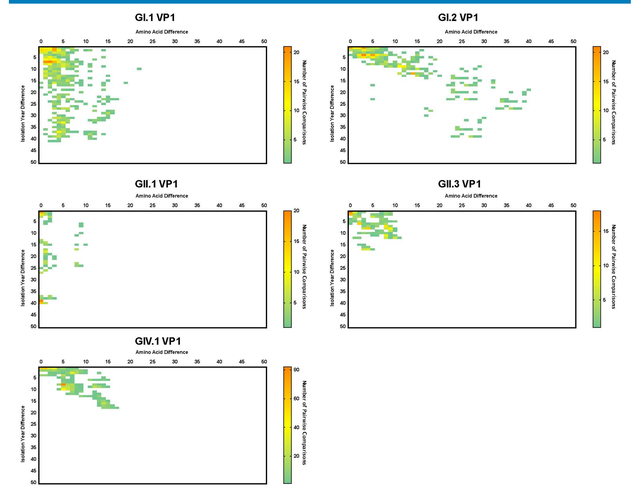


Figure 5. Temporal amino acid diversity patterns of VP1 capsid protein from human sapovirus genotypes GI.1, GI.2, GII.1, GII.3, and GIV. The heat map represents the number of pairwise comparisons; red and green represent the highest and lowest number of pairwise comparisons, respectively.

"False-signal" of recombination in human sapoviruses can be explained by disparate patterns of evolution among genome regions

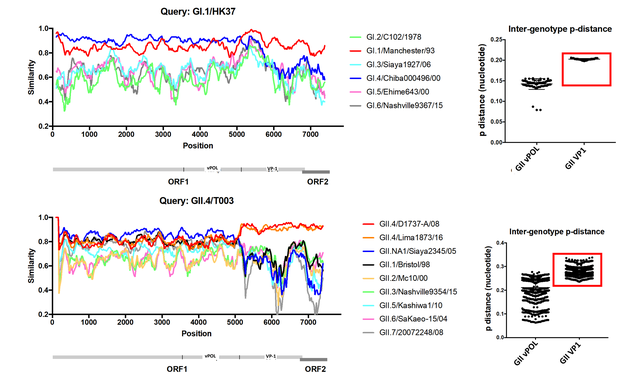


Figure 6. Genomic analyses of putative recombinant human sapoviruses HK37 and T003 showing (left panels) a site-by-site nucleotide similarity analyses across their genomes, and (right panels) the inter-genotype/genogroup nucleotide substitution differences among the genomic regions for vPOL and VP1.

Conclusions

- A simple genome sequencing platform for human sapoviruses that could facilitate epidemiological studies was developed
- Genome sequences from historical sapoviruses were newly obtained
- Despite presenting different patterns of amino acid divergence, human sapovirus genotypes presented similar rates of evolution