Virome Analysis of Transfusion Recipients Reveals a Novel Human Virus That Shares Genomic Features with Hepaciviruses and Pegiviruses

Amit Kapoor,[a] Arvind Kumar,[a] Peter Simmonds,[b] Nishit Bhuva,[a] Lokendra Singh Chauhan,[a] Bohyun Lee,[a] Amadou Alpha Sall,[c] Zhezhen Jin,[d] Stephen S. Morse,[e] Beth Shaz,[f] Peter D. Burbelo,[g] W. Ian Lipkin[a]

Center for Infection and Immunity, Columbia University, New York, New York, USA[a]; Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom[b]; Institute Pasteur de Dakar, 36, Avenue Pasteur, Dakar, Senegal[c]; Department of Biostatistics, MSPH, Columbia University, New York, USA[d]; Department of Epidemiology, MSPH, Columbia University, New York, USA[e]; New York Blood Centers, New York, New York, USA[f]; Dental Clinical Research Core, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA[g]

ABSTRACT To investigate the transmission of novel infectious agents by blood transfusion, we studied changes in the virome composition of blood transfusion recipients pre- and posttransfusion. Using this approach, we detected and genetically characterized a novel human virus, human hepegivirus 1 (HHpgV-1), that shares features with hepatitis C virus (HCV) and human pegivirus (HPgV; formerly called GB virus C or hepatitis G virus). HCV and HPgV belong to the genera *Hepacivirus* and *Pegivirus* of the family *Flaviviridae*. HHpgV-1 was found in serum samples from two blood transfusion recipients and two hemophilia patients who had received plasma-derived clotting factor concentrates. In the former, the virus was detected only in the posttransfusion samples, indicating blood-borne transmission. Both hemophiliacs were persistently viremic over periods of at least 201 and 1,981 days. The 5′ untranslated region (UTR) of HHpgV-1 contained a type IV internal ribosome entry site (IRES), structurally similar to although highly divergent in sequence from that of HCV and other hepaciviruses. However, phylogenetic analysis of nonstructural genes (NS3 and NS5B) showed that HHpgV-1 forms a branch within the pegivirus clade distinct from HPgV and homologs infecting other mammalian species. In common with some pegivirus variants infecting rodents and bats, the HHpgV-1 genome encodes a short, highly basic protein upstream of E1, potentially possessing a core-like function in packaging RNA during assembly. Identification of this new human virus, HHpgV-1, expands our knowledge of the range of genome configurations of these viruses and may lead to a reevaluation of the original criteria by which the genera *Hepacivirus* and *Pegivirus* are defined.

IMPORTANCE More than 30 million blood components are transfused annually in the United States alone. Surveillance for infectious agents in the blood supply is key to ensuring the safety of this critical resource for medicine and public health. Here, we report the identification of a new and highly diverse HCV/GB virus (GBV)-like virus from human serum samples. This new virus, human hepegivirus 1 (HHpgV-1), was found in serum samples from blood transfusion recipients, indicating its potential for transmission via transfusion products. We also found persistent long-term HHpgV-1 viremia in two hemophilia patients. HHpgV-1 is unique because it shares genetic similarity with both highly pathogenic HCV and the apparently nonpathogenic HPgV (GBV-C). Our results add to the list of human viruses and provide data to develop reagents to study virus transmission and disease association and for interrupting virus transmission and new human infections.

Transfusion of blood or blood-derived products can save lives and improve health but requires safety measures for preventing bystander transmission of infectious agents. Exclusion of blood products contaminated with infectious agents includes rigorous testing for pathogenic viruses (1, 2). The unusually high infection prevalence of hepatitis C virus (HCV), HBV, and HIV in hemophilia patients and other transfusion recipients could have been prevented by earlier identification of these viruses and development of accurate diagnostic assays (3–8). Advances in sequencing platforms have revolutionized identification of new viruses (9) and have been widely adopted for identification and characterization of viruses that infect humans and animals (10–15). Genetic characterization of new viruses is a first crucial step toward their molecular and biological characterization. Sequence data of a new virus not only help in estimating its biological properties but also enable development of molecular reagents that can be used for molecular diagnostics and epidemiological and clinical investigations of their transmission and disease associations (9).

HCV infects ~200 million people worldwide and represents a genetically diverse group of human RNA viruses with similar biological properties. HCV typically persists, with only 20 to 30% of cases resolving after acute infection (16, 17). HCV is a global pathogen, as a large proportion of untreated individuals progress to severe liver disease, cirrhosis, and hepatocellular carcinoma. HCV therapies based on antivirals that directly target HCV replication are expensive, and prevention of new virus infections remains the priority (18). The identification of HCV in 1989 (19) was crucial for development of diagnostic assays to arrest its spread through transfusion and other parenteral routes of exposure. Other human viruses genetically related to HCV include human pegivirus (HPgV, formerly named GB virus C or hepatitis G virus) and a growing range of related viruses currently known to infect horses, rodents, bats, and monkeys (reviewed in references 20 and 21). HPgV causes widespread human infection (1 to 4% of healthy blood donors) that can persist for years with no clinical symptoms (22).

Identification of natural infection with nonprimate hepacivirus (NPHV), first in dogs (initially termed canine hepacivirus) and subsequently in horses, was the first indication of a wider host range for hepacivirus infections (23–25). More recently, identification of a diverse range of viruses found in bats, rodents, and Old World primates further expanded the host range of hepaciviruses (11, 12, 26–28). Similarly, several novel pegiviruses have recently been identified, with a single report of such viruses in bats in 2010 (29), followed by several recent reports of novel viruses in horses, rodents, and bats (11, 12, 26, 27, 30). These new discoveries have fundamentally revised our knowledge of viral diversity and host ranges of hepaciviruses and pegiviruses (20). However, so far, the only two viruses known to infect humans are HCV and HPgV. Here, we describe first identification of a highly divergent human virus, human hepegivirus 1 (HHpgV-1), that shares attributes of both HCV and HPgV.

RESULTS

Virome analysis of transfusion recipients. To characterize the changes in serum virome compositions of transfusion recipients, we studied pretransfusion and posttransfusion serum samples from subjects enrolled in the Transfusion-Transmitted Viruses Study (TTVS) from July 1974 through June 1980. One pre- and one posttransfusion serum sample from 44 individuals were enriched for viral nucleic acids and studied by unbiased amplification and Illumina sequencing (11). A total of 154 and 119 million reads were generated from the pretransfusion and posttransfusion samples, respectively. The descriptive statistics mean, standard deviation (SD), and median were calculated for viral abundance. The viral abundance at pretransfusion and the rate at posttransfusion were compared by paired Wilcoxon rank sum test. Comparison of posttransfusion samples with pretransfusion samples revealed significant increase in the numbers of total virus reads ($P < 0.009$) and anellovirus reads ($P < 0.007$) after transfusion (Fig. 1). To understand the nature and composition of increased anellovirus sequences, comparative analysis was done for all virus populations (virome) in pre- and posttransfusion samples from six TTVS subjects that showed nonsignificant increases in the numbers of normalized anellovirus reads and six TTVS subjects that showed >10-fold increases in the numbers of anellovirus virus reads in the posttransfusion samples. Our analysis shows that the majorities of increased anellovirus reads were new geno-

types or species that were genetically different from those that existed in individuals before transfusion. However, we also noticed persistence of a proportion of anellovirus sequences that were genetically similar among pre- and posttransfusion samples (Fig. 1). Together, these findings indicate that these subjects were infected by both stable and fluctuating anellovirus populations that can exist as coinfections. Of the total virus reads, approximately 40% of the reads remained unclassified due to their low similarity to known viruses (E value range, 10^{-2} to 10^{-6}). Further analysis of these unclassified reads revealed the presence of a novel flavivirus in two posttransfusion samples.

Identification of HHpgV-1. Bioinformatic analysis of sequence data of two serum samples, TTVS-772 and TTVS-790, both collected after blood transfusion, indicated the presence of a novel flavivirus, showing only distant sequence homology to HCV and HPgV. PCR assays were used to confirm the presence of these flavivirus-like sequences in a fresh aliquot of the original serum samples. After initial analysis of data, both samples, TTVS-772 and TTVS-790, were subjected to more in-depth sequencing, generating >100 million single-end Illumina reads. All sequence reads that mapped to genomes of known hepaciviruses and pegiviruses and showed protein similarity (E value < 0.0001) with different genomic regions of these viruses were connected using PCRs of intervening genomic fragments and Sanger dideoxy sequencing. Finally, the complete polyprotein coding region of human hepegivirus 1 (HHpgV-1) present in TTVS-790 was sequenced at more than 4× coverage, and the genomic termini were acquired using random amplification of cDNA ends (RACE) (23), yielding a 9,538-nucleotide genome of the new virus, HHpgV-1.

Polyprotein-coding region of HHpgV-1. A continuous reading frame spanned most of the HHpgV-1 genome, with a termination codon at position 9501 (Fig. 2). Predicting the start of the virus polyprotein was more problematic because several in-frame ATG (methionine) triplets were present at the 5′ end of the HHpgV-1 sequence. Of these, those at positions 195 and 204 diverged from the optimal Kozak motif (GTTC|ATG|GAGG and AGGGC|ATG|CCCA, respectively). The codon at position 330 had an optimal sequence (GCAAC|AUG|GGGU), the fourth (position 519) was suboptimal (CCGGU|AUG|UGTT), and the fifth (position 587) fell within a region of identifiable protein sequence homology to the E1 sequence of pegiviruses. Determination of the RNA structure of the 5′ untranslated region (UTR) (see below) provided separate evidence that the (third) ATG triplet at position 330 was the genuine initiating codon, and this has been assumed in the analysis of protein cleavage sites below.

Evolutionary relatedness of HHpgV-1 with hepaciviruses and pegiviruses. The translated ORF of HHpgV-1 was aligned with pegivirus and hepacivirus sequences from humans and chimpanzees (HPgV, simian pegivirus infecting chimpanzees [SPgVtro], and HCV), Old World primates (SPgV-OWM and SHcV-OWM), viruses formerly described as GB virus A (GBV-A) and GBV-B from New World primates (SPfV-NWM and SHcV-NWM), rodent variants from rats and mice (RPgV and RHcV-1 to RHcV-3) and horses (EPgV-1, EPgV-2, and NPHV), and several different groups of pegiviruses and hepaciviruses from bats (BPgV-1 to BPgV-5 and BHcV-1 to BHcV-4).

Translated amino acid sequences could be readily aligned over most of the genome of pegiviruses and the nonstructural gene region of hepaciviruses. Evolutionary relationships of HHpgV-1 with other groups were determined by phylogenetic analysis of the
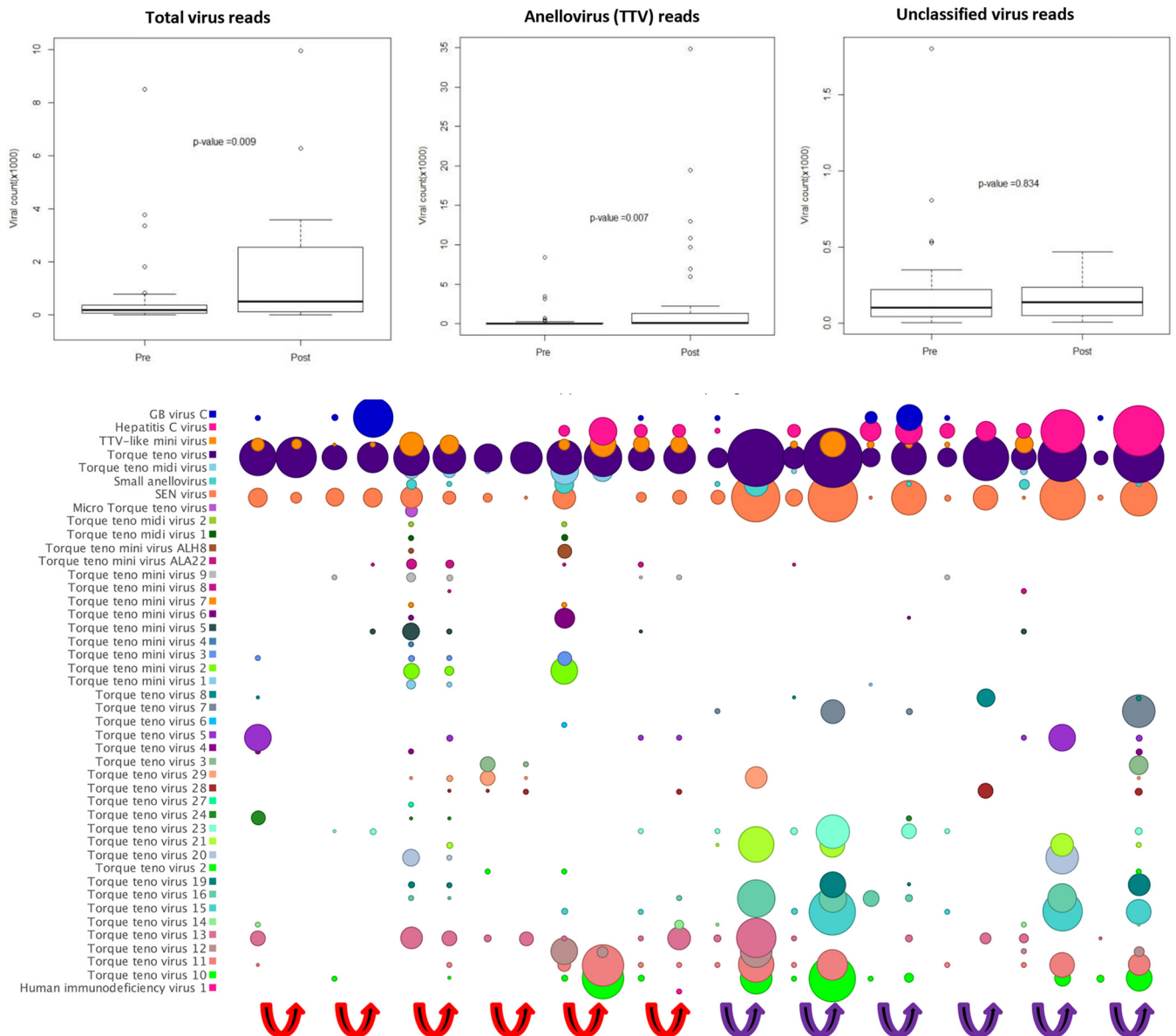
FIG 1 Virome analysis of TTVS pre- and posttransfusion samples. (Top) Statistical analysis of total virus sequences (left), anellovirus sequences (middle), and virus-like unclassified sequences (right) in pre- and posttransfusion samples. (Bottom) Metagenomic binning of all viruses sequences in pre- and posttransfusion samples from twelve TTVS subjects. Red arrows represent pre- and posttransfusion sample pairs that were similar in amount of total virus sequences, and purple arrows represent pre- and posttransfusion sample pairs where posttransfusion samples had significantly larger amounts of total virus sequences.

more highly conserved NS3- and NS5B-encoding regions. Trees from the two regions were congruent in which the HHpgV-1 sequence was distinct from all pegivirus and hepacivirus sequences published to date (Fig. 3). In both regions, sequences split into separate clades corresponding to classifications as hepaciviruses and pegiviruses. Within the latter clade, two genetically distinct lineages were apparent (groups 1 and 2); group 1 comprised variants from humans, primates, horses, and bat lineages 1, 4, and 5, while the HHpgV-1 sequence and variants from rodents and bat lineages 2 and 3 fell into group 2.

Prediction of HHpgV-1 polyprotein cleavage sites. The positions of gene boundaries of HHpgV-1 were predicted using two approaches. Translated amino acid sequences in the 5′ structural gene region were submitted to the Signal P 4.0 server to identify signalase cleavage sites used by other pegiviruses to process structural genes (Fig. 2). This revealed three potential cleavage sites comparable in position to those previously identified for equine pegiviruses and in part to those in the annotation of SPgV-NWM. However, in addition to the E1, E2, and X proteins, HHpgV-1 was predicted to include an additional 65-amino-acid, cysteine-rich protein upstream of E1. Additional coding regions have been previously identified in rodent PgV (12) and labeled "VR" in the group 2 bat pegiviruses (lineages 2 and 3) (26). Y or VR proteins were highly variable in size and showed no identifiable sequence similarity to each other. The sequence from HHpgV-1 did, however, share the property of being highly basic with several posi-
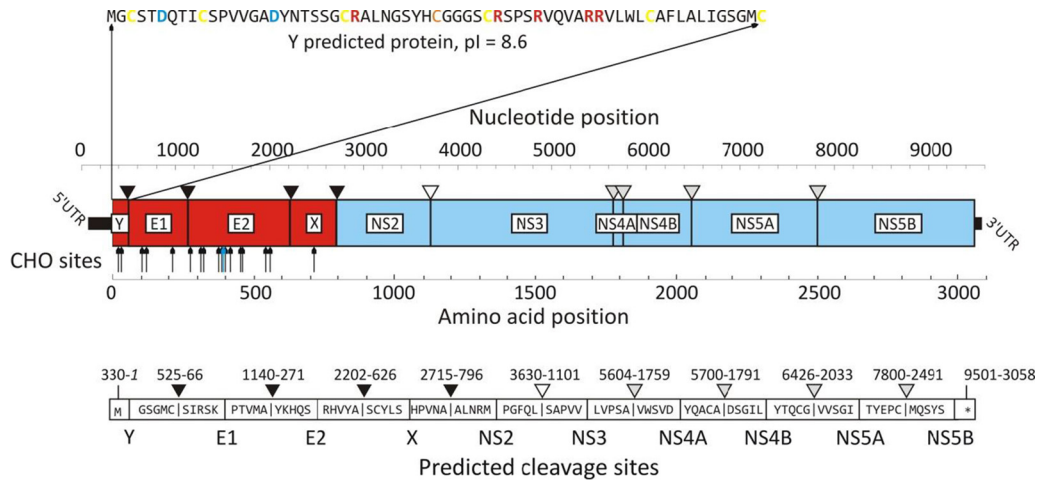
FIG 2 Genome organization and polyprotein cleavage sites of HHpgV-1. The positions of cleavage sites of the viral NS2 (white triangles) and NS3 protease (grey triangles) were predicted by alignment and homology to sites in the nonstructural genes previously characterized in other pegiviruses and hepaciviruses. Cleavage sites in structural gene regions (black triangles) were independently predicted using the SignalP 4.1 server. Predicted N-linked glycosylation sites in envelope proteins are depicted by vertical arrows; the single predicted O-linked glycosylation site is indicated with a blue arrow. The amino acid sequence of the predicted Y protein is shown above the genome diagram. Charged residues and cysteines are in color.

tively charged residues (pI 8.6) (Fig. 2), comparable to previously analyzed bat sequences (pI 8.2 to 9.1) and RPgV (pI 10.1). The Y and VR proteins, differed, however, in their predicted locations, with rodent and bat Y/VR proteins possessing a predicted signalase cleavage site at the N-terminus that may re-locate the protein through the endoplasmic reticulum (ER). In contrast, the Y protein of HHpgV-1 possessed no such initial translocating sequence, which potentially indicates a cytoplasmic location for the protein. This processing resembles that of the core proteins of hepaciviruses, which functions in RNA packaging and virion formation. Although considerably shorter than the processed HCV protein, the Y protein of the HHpgV-1 sequence may have a similar function.

The E2 protein contained 11 potential N-linked glycosylation sites, a greater number than recorded for other pegiviruses; additional sites were detected in E1 ($n = 3$) and X ($n = 1$). E2 additionally contained a potential O-linked site at position 422 (Fig. 2). This pattern of heavy glycosylation, comparable in extent to that of HCV and other hepaciviruses, was distinct from the infrequent sites recorded for human and other group 1 pegiviruses. For prediction of cleavage sites between nonstructural proteins, there was sufficient similarity between the HHpgV-1 and other pegivirus sequences with annotated sites for homologous positions to be identified (Fig. 2). Proteins predicted in the NS region of HHpgV-1 were comparable in position and size to those of other pegiviruses and hepaciviruses (Fig. 2).

Sequence divergence between sequences of HHpgV-1 and other pegiviruses aligned over the genome was high, ranging from 61 to 65% with group 1 pegivirus sequences and 58% to 60% with group 2 in the structural gene region (Table 1). This corresponded to amino acid distance ranges of 68 to 79% and 69 to 71% in this genome region. The nonstructural gene regions were better conserved, with group 1 and 2 nucleotide distance ranges of 57 to 59% and 55% between HHpgV-1 and group 1 and 2 sequences, respectively, and 68 to 69% and 62 to 64% amino acid distances for those two groups (Table 2). As implied by these results, sequence divergence was distributed throughout the pegivirus genome, with

≥40% minimum amino acid sequence divergence values of windowed fragment between HHpgV-1 and each lineage of other pegiviruses (Fig. 4). Highest divergence was observed in the structural gene regions and NS4B/NS5A, while minimum values were found in association with functionally conserved polymerase and helicase motifs in NS5B and NS3.

Prediction of 5′ UTR RNA structure. The separate grouping of pegiviruses into two main clades in the coding region was reproduced in sequence relationships of the 5′ UTR. All sequences of group 1 viruses could be aligned and possessed predicted RNA secondary structures closely matching that of GBV-A and the recently predicted structure of equine pegivirus (11). Surprisingly, 5′ UTR sequences of group 2 pegiviruses showed no homology to those of group 1 but nevertheless could be aligned with each other. Although highly divergent in sequence from all other known viral 5′ UTR sequences, certain motifs matched those of members of the *Hepacivirus* genus, including the highly conserved sequence TACAGCCTGATAGGGT at position 274. The sequence lies within domain IIIe (pseudoknot region) in the hepacivirus internal ribosome entry site (IRES). Accordingly, the sequence of HHpgV-1 was structurally aligned to hepacivirus 5′ UTR sequences, an analysis that revealed the presence of a typical type IV IRES (Fig. 5) and upstream stem-loops corresponding in position, shape and size to domains I and II of HCV. However, the miR-122 seed match sequence (ACACUCCA) was absent in the 5′ UTR sequence of HHpgV-1 and all other known pegiviruses. As anticipated based on their sequence homology, the other group 2 pegiviruses have similarly structured type IV IRES elements despite their substantial sequence divergence from each other (data not shown). In this RNA structure model of the HHpgV-1 5′ UTR, the third ATG at position 330 that possesses the optimal Kozak sequence is located 11 bases downstream from the pseudoknot in domain IV and, by analogy, is the most plausible initiating codon for HHpgV-1.

GORS. Thermodynamic folding analysis of the HHpgV-1 sequence revealed a mean 7.6% free energy difference between its minimum folding energy and that of sequence order-randomized
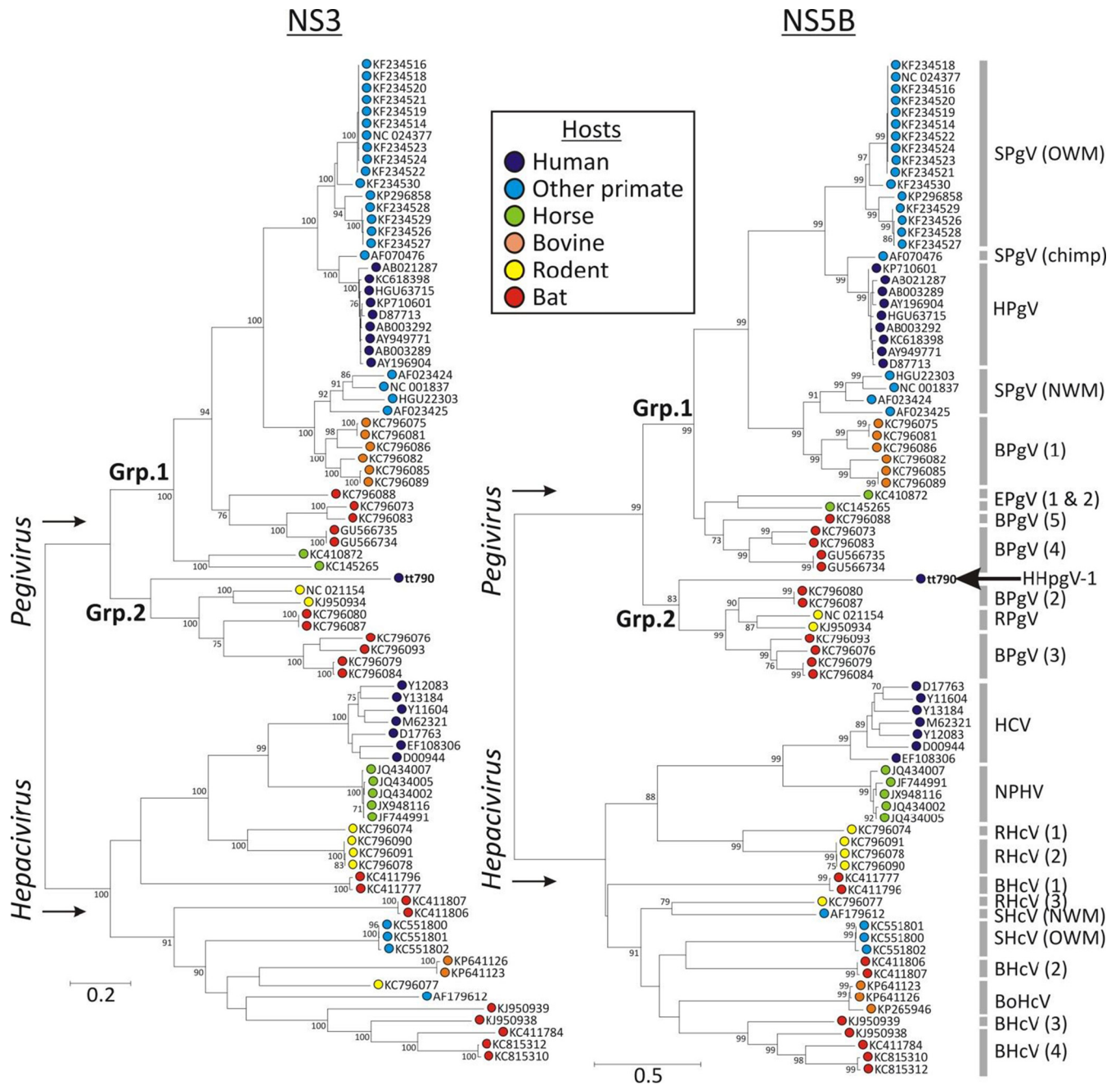
FIG 3 Maximum-likelihood phylogenetic analysis of the NS3 and NS5B genes of HHpgV-1 (tt790) and available complete genome sequences of other hepaciviruses and pegiviruses infecting different mammalian species. Phylogenetic analysis of each data set used 100 bootstrap resamplings to determine robustness of grouping; values are shown on branches. Abbreviations: HcV, hepacivirus; PgV, pegivirus; OWM, Old World monkey; NWM, New World monkey; R, rodent; B, bat; Bo, bovine; NPHV, nonprimate hepacivirus. Lineage numbers are shown in parentheses.

controls (MFED), an observation consistent with the presence of genome-scale ordered RNA structure (GORS) in the HHpgV-1 sequence (23). However, this MFED value was lower than those of the group 1 viruses HPgV (mean, 11.7%; standard deviation, 0.7%), SPgV-OWM (10.9% ± 1.3%), SPgV-NWP (11.8% ± 0.4%), EPgV-1 (10.4%), and EPgV-2 (11.9%). However, the MFED value of HHpgV-1 was comparable to MFED values of group 2 pegiviruses (7.9% to 8.8%) (Table 3).

HHpgV-1 infections, prevalence, and coinfections. To determine the prevalence and nature of HHpgV-1 infection in humans,

we tested pretransfusion and posttransfusion serum samples from 46 and 116 TTVS subjects, respectively. We also tested 1 to 5 longitudinally collected serum samples from 106 Multicenter Hemophilia Cohort Study (MHCS) subjects. Only two posttransfusion samples from TTVS subjects were positive for HHpgV-1 RNA, TTVS-772 and TTVS-790. The HHpgV-1-positive serum sample TTVS-772 was collected from a subject 17 days posttransfusion. A sample from the same individual collected one day before transfusion (TTVS-0927) and another collected 281 days after transfusion (TTVS-1095) tested negative for the virus. The

TABLE 1 Nucleotide sequence divergence of structural and nonstructural genes of HHpgV-1 and other pegiviruses

Gene region and group	Lineage	Nucleotide distance (%)[a]											
		Group 1									Group 2		
		HHpgV-1	Ape	OWM	NWM	Bat_	Bat_4	Bat_5	Equine_1	Equine_2	Bat_2	Bat_3	Rodent
Structural													
1	Ape	63.0	**18.3**										
	OWM	61.3	43.1	**18.9**									
	NWM	63.0	52.5	53.0	**43.0**								
	Bat_	63.4	51.6	51.4	47.5	**38.4**							
	Bat_4	64.1	55.2	55.5	56.9	57.5	**41.2**						
	Bat_5	65.2	52.5	51.7	56.5	56.4	55.4	—					
	Equine_1	63.9	58.3	58.2	61.5	61.1	62.7	60.1	—				
	Equine_2	63.3	57.0	56.9	58.5	57.4	59.5	58.0	61.3	—			
2	Bat_2	59.5	60.9	60.9	62.7	61.5	62.9	60.2	62.4	59.6	**9.0**		
	Bat_3	59.3	59.5	60.4	62.9	62.5	62.8	61.3	61.7	61.3	54.8	**41.9**	
	Rodent	58.0	59.2	61.0	61.6	62.3	62.4	59.8	60.9	59.8	52.2	54.2	**49.2**
Nonstructural													
1	Ape	59.1	**15.4**										
	OWM	59.0	37.1	**16.1**									
	NWM	59.1	46.8	47.9	**36.7**								
	Bat_	58.6	46.6	47.0	40.0	**31.4**							
	Bat_4	58.1	51.7	51.8	51.6	50.8	**32.7**						
	Bat_5	57.4	49.8	51.2	50.5	50.6	47.2	—					
	Equine_1	58.7	52.9	53.6	52.3	52.0	50.8	51.6	—				
	Equine_2	57.8	52.6	53.1	52.5	51.9	50.8	51.0	49.4	—			
2	Bat_2	54.9	54.9	55.3	55.4	54.4	54.4	54.4	55.7	55.4	**5.5**		
	Bat_3	55.3	55.2	56.3	55.8	54.9	55.3	55.4	55.4	56.4	45.4	**30.8**	
	Rodent	54.9	54.8	55.3	55.4	54.4	54.4	54.4	55.9	55.2	45.2	47.7	**42.7**

[a] Bold indicates within-group comparisons. "—", No within-group comparison made (only one group member exists).

HHpgV-1-positive serum sample TTVS-790 was collected at 7 days posttransfusion. A sample from the same individual collected one day before transfusion (TTVS-1608) and another collected at 241 days after transfusion (TTVS-791) tested negative for HHpgV-1 RNA. These results indicate transient viremia of HHpgV-1 in posttransfusion samples followed by virus clearance in both the infected patients.

Of the 106 MHCS subjects, multiple samples from only two subjects, M3127 and M4287, were positive for HHpgV-1 RNA. The two adjacent samples from subject M3127 positive for HHpgV-1 were collected 201 days apart, while a sample collected 144 days before the first positive sample and two other samples collected at days 560 and 3461 after the second positive sample remained negative for HHpgV-1 infection. For the second HHpgV-1-positive MHCS subject (M4287), four longitudinal samples were available for this study and all were positive for HHpgV-1. The latter three of these four samples were collected at 399, 742, and 1,981 days after the first sample collection day. These results indicate persistence of HHpgV-1 viremia for more than 201 and 1,981 days (5.4 years) in these two MHCS subjects.

All PCR products were sequenced from both directions, to confirm results and to study the genetic diversity among different HHpgV-1 variants. Phylogenetic analysis confirmed the close genetic relatedness among HHpgV-1 variants within their respective MHCS and TTVS subjects (Fig. 6). Sequences were highly conserved, with substitutions being restricted largely to synonymous sites. However, by analogy to other pegiviruses and hepaciviruses, greater variability is likely elsewhere in the genome. Coinfections with HCV and HPgV were determined by viremia screening of HHpgV-1-positive samples for for HPgV and HCV. Three of the four HHpgV-1-infected subjects were found to be coinfected with HCV (genotype 1a in two subjects and 1b in one subject). All HHpgV-1 subjects and their respective samples remained negative for HPgV coinfection.

DISCUSSION

We used an unbiased (sequence-independent) virus discovery approach (9) to characterize known as well as unknown viruses that can pose a risk of transfusion transmission. We studied longitudinal pre- and posttransfusion samples from TTVS subjects for comparative analysis of changes in virome compositions after transfusion. The number of virus reads, particularly of anelloviruses, significantly increased after transfusion (Fig. 1). TTV species were studied in transfusion recipients earlier (31, 32), but the cross section design and low-depth sequencing approach used in previous studies were not suitable for characterizing the compositional changes in TTV populations over time. There can be several reasons for the changes we observed in the virome of transfusion recipients, such as deteriorating health of subjects and new nosocomial infections, including transfusion transmission of viruses. The presence of a novel human virus, HHpgV-1, in two transfusion recipients was our most interesting observation. The availability of longitudinal serum samples for TTVS subjects was helpful in determining the nature of HHpgV-1 infection; the virus infection was present only in the posttransfusion samples and was followed by clearance of viremia in both subjects. Although our

TABLE 2 Translated amino acid sequence divergence of structural and nonstructural genes of HHpgV-1 and other pegiviruses

Gene region and group	Lineage	Amino acid distance (%)											
		Group 1									Group 2		
		HHpgV-1	Ape	OWM	NWM	Bat_	Bat_4	Bat_5	Equine_1	Equine_2	Bat_2	Bat_3	Rodent
Structural													
1	Ape	74.9	**10.4**										
	OWM	74.7	44.0	**14.7**									
	NWM	76.3	61.2	61.7	**43.8**								
	Bat_	76.3	59.5	59.4	53.4	**37.9**							
	Bat_4	79.2	64.9	64.8	68.1	68.9	**44.7**						
	Bat_5	78.3	62.3	60.1	67.6	66.9	66.4	**0.0**					
	Equine_1	78.9	67.6	68.7	75.4	75.1	77.1	72.6	**0.0**				
	Equine_2	76.8	66.1	65.7	71.4	69.3	71.4	68.9	73.5	**0.0**			
2	Bat_2	71.4	74.6	74.0	77.8	76.6	77.5	75.8	77.9	73.9	**1.0**		
	Bat_3	70.6	73.0	72.8	77.4	76.8	75.9	75.6	76.5	74.7	66.4	**46.0**	
	Rodent	69.2	71.7	73.9	76.6	77.1	77.0	75.1	75.1	75.2	61.9	65.2	**56.4**
Nonstructural													
1	Ape	68.5	**5.4**										
	OWM	68.6	32.7	**10.4**									
	NWM	68.9	49.2	50.5	**31.5**								
	Bat_	69.0	49.2	49.2	37.6	**25.4**							
	Bat_4	68.6	57.0	56.3	57.8	56.6	**30.1**						
	Bat_5	67.9	55.8	56.4	56.3	56.4	51.6	**0.0**					
	Equine_1	68.6	58.9	59.6	59.3	59.0	57.1	57.4	**0.0**				
	Equine_2	68.7	58.7	59.2	60.4	59.8	56.6	57.3	53.5	**0.0**			
2	Bat_2	62.2	62.9	64.6	65.2	64.3	63.3	64.8	64.4	65.0	**0.5**		
	Bat_3	63.3	63.8	65.4	64.9	63.9	63.7	64.8	64.8	66.0	48.1	**27.9**	
	Rodent	63.6	63.6	64.4	64.8	63.6	62.7	63.7	64.6	64.4	48.3	50.6	**42.0**

results provide evidence for transfusion-mediated transmission of HHpgV-1, further studies are necessary to verify the presence of HpgV-1 in the transfused blood and to rule out other possible iatrogenic infection sources over this period. Although the presence of a new viral genome in the human serum samples confirms infection, more evidence should be sought, including seroconversion of infected patients; therefore, we are developing a serological assay for HHpgV-1 (23, 33). Considering that processing clinical samples in a laboratory, as well as contaminated nucleic acid extraction reagents, can be a potential source of divergent viruses (9, 34), we always included appropriate negative controls in all experiments of this study. The control samples included in this study were water and PBS extracted with each batch of serum samples and also nontemplate reagent controls for reverse transcription and screening PCR assays. Moreover, comparison of HHpgV-1 sequences amplified from different samples collected from differ-
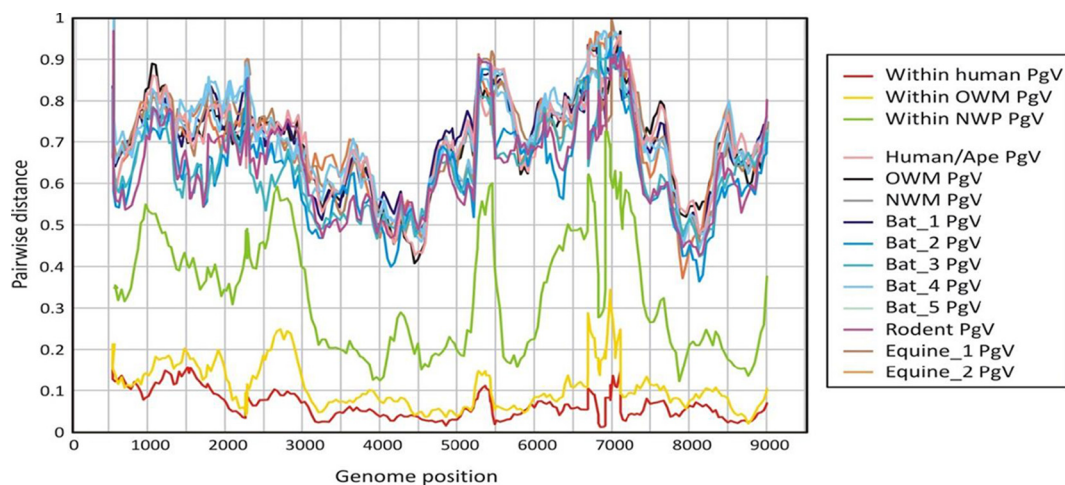

FIG 4 Amino acid sequence divergence between HHpgV-1 and other pegiviruses using a 240-nucleotide fragments with 24-nucleotide increments across the virus alignment (midpoint plotted on the x axis). Within-species distances for human and simian pegiviruses were included for comparison. The divergence scan was numbered using the AF121950 HPgV reference sequence. OWM, Old World monkey; NWP, New World primate; NWM, New World monkey.
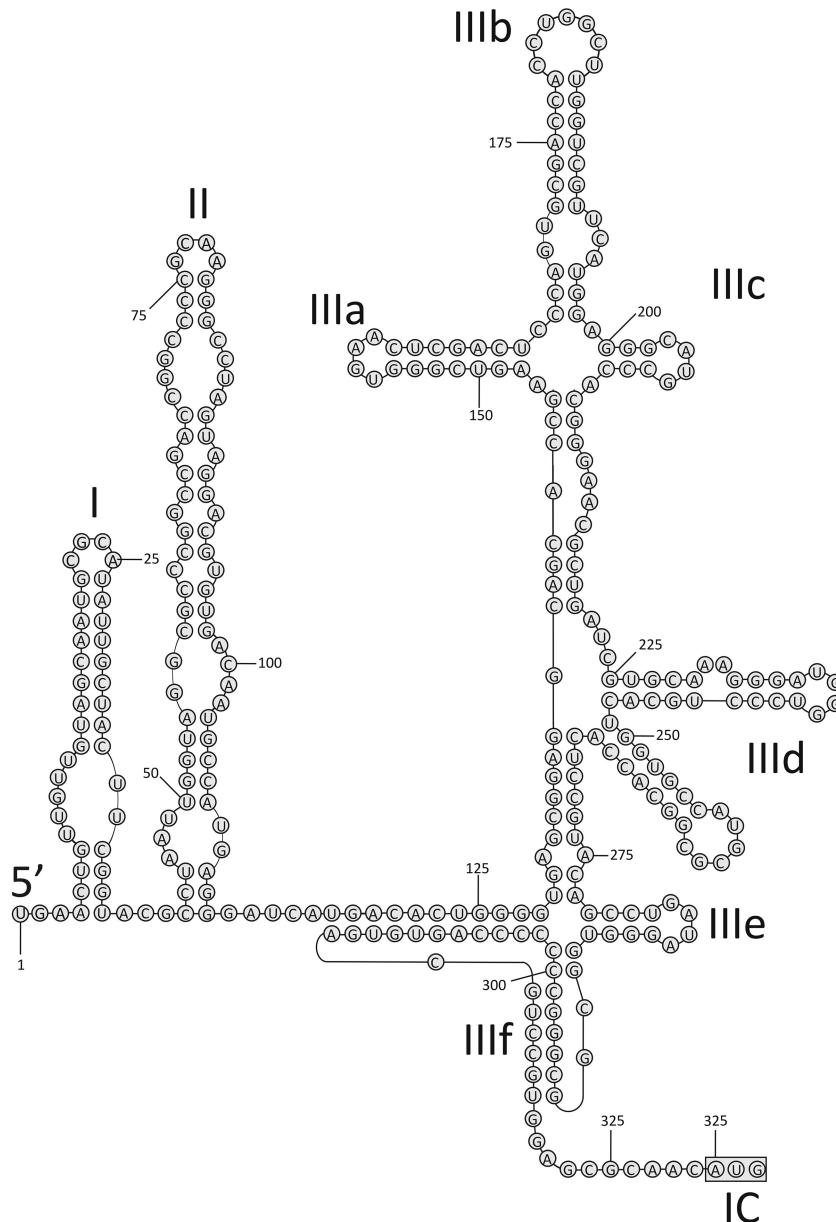
FIG 5 Predicted RNA secondary structure of the 5′ UTR, based on structural mapping of the HHpgV-1 sequence to hepacivirus sequences based on an initial seed match in domain IIIe. Stem-loops for domains I and II were predicted by Mfold.

ent individuals and at different time points revealed the existence of some diversity, concentrated at synonymous sites; the finding of patient-specific groupings of sequences (Fig. 6) similarly argues strongly against laboratory or reagent contamination as the HHpgV-1 sequences.

HHpgV-1 was identified by Illumina sequencing in the post-transfusion samples from two of the 46 originally enrolled TTVS subjects. Subsequent specific PCR screening of posttransfusion samples from 70 TTVS subjects failed to identify any further positive samples. HHpgV-1 was also detected only infrequently in hemophiliacs (2 of 106 subjects). The infrequency of detection of HHpgV-1 compared to HCV and HPgV in this risk group may be explained by the fact that (i) HHpgV-1 is a much less prevalent human infection, (ii) this virus establishes persistence less fre-

quently, or (iii) the PCR-based screening method used for viremia detection was unable to amplify sequences of more genetically divergent variants of HpgV-1. Historically, the latter explanation led to an underestimate of the prevalence of infection of HCV (35) that was only resolved when the full extent of the genotype diversity of HCV was identified. Determining the incidence and prevalence of human HHpgV-1 infection will require additional molecular and serological studies.

The genomic organization of HHpgV-1 revealed several novel genetic features in this new human virus. First, the 5′ UTR was nonhomologous to that of most other pegiviruses, forming a type IV IRES, hitherto described only for members of the genera *Hepacivirus* and *Pestivirus* of *Flaviviridae*. HHpgV-1 shared sequence motifs with other group 2 pegiviruses, despite their sequence di-

TABLE 3 Mean folding energy values of different pegivirus lineages/groups

Group	Lineage	No.[a]	MFE (kcal/mol)	MFED (%)[b]
1	HHpgV-1		−65.06	7.6
	Human_PgV	10	−82.78	11.7 ± 0.7
	OWM_PgV	8	−73.55	10.9 ± 1.3
	NWM_PgV	3	−76.08	11.8 ± 0.4
	Bat_1_PgV	6	−80.54	10.6 ± 1.6
	Bat_4_PgV	4	−72.99	10.2 ± 0.7
	Bat_5_PgV	1	−77.15	9.7
	Equine_1_PgV	1	−76.08	10.4
	Equine_2_PgV	1	−72.31	11.9
2	Bat_2_PgV	2	−77.12	7.9 ± 0.0
	Bat_3_PgV	4	−72.67	8.5 ± 0.9
	Rodent_PgV	2	−79.91	8.8 ± 0.8

[a] Number of sequences compared.
[b] Mean of 225-base fragments, in 24-base increments, over the complete coding region. Values are means and standard deviations.

vergence from each other, allowing the retrospective identification of this IRES type in all members of this pegivirus subgroup, undocumented in the original publications (12, 26). This finding is complementary to the earlier description of pegivirus-like IRES sequences in three bat-associated hepaciviruses (27). As with picornaviruses (36), exchange of the translation modules between otherwise distantly related viruses evidently occurs relatively often in the evolutionary history of RNA viruses, although the forces driving such modular exchanges and effects of different IRES types on virus replication strategies and interaction with host cells are largely unknown. However, variable dependence of different IRES types on host translation initiation factors, such as eIF3 and eIF-2α (37, 38), clearly influences their ability to withstand shutoff of host cell translation mediated through stress responses or induction of interferon-stimulated genes (39). Another genomic feature of HHpgV-1 is the presence of short coding sequences upstream of E1 (Fig. 3). This coding region encodes a predicted cytoplasmic basic protein that, although considerably shorter than the HCV core protein, may perform similar functions in RNA binding and packaging during virus assembly. HHPgV-1 similarly resembles HCV and several other hepaciviruses in possessing numerous sites for potential N-linked glycosylation in the predicted envelope genes, particularly E2, in which 11 N-linked and one predicted O-linked sites are located. This pattern of heavy glycosylation, which contrasts with the paucity of sites in HPgV and most other pegiviruses, is believed to shield viral envelopes
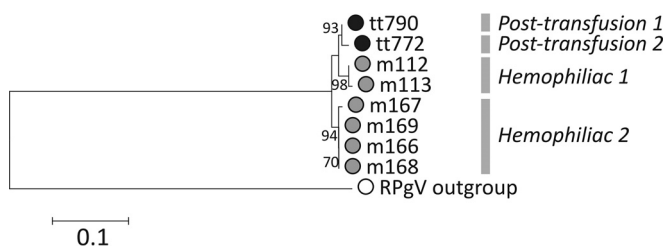

FIG 6 ML phylogenetic analysis of partial NS3 sequences of HHpgV-1 variants found in different human serum samples using a rodent pegivirus sequence as the outgroup. Phylogenetic analysis used 100 bootstrap resamplings to determine robustness of grouping. HHpgV-1 variants found in serum samples from the same patients are labeled on the right (tt, TTVS; m, MHCS).

from antibody-mediated neutralization (40, 41) and to participate in virus entry into hepatocytes (42). It is conceivable that HHpgV-1 virions display similar host interactions.

The finding of persistent viremia in hemophiliacs in sequentially collected samples and the time interval between sampling and their previous exposure to potentially infectious clotting factor concentrates (1989 to 1990) demonstrates the propensity of HHpgV-1 for long-term host persistence, an attribute shared with most, if not all, hepaciviruses and pegiviruses. HHpgV-1 resembles other viruses in these genera in possessing a large-scale structured RNA genome (GORS) that is tightly associated with host persistence throughout different families of positive-stranded mammalian RNA viruses (43, 44). Although this association is poorly characterized functionally (45), HHpgV-1 shares midrange MFED values with other group 2 pegiviruses and with HCV (Table 3). Future bioinformatics characterization of the configuration of RNA structures in such viruses, and potentially experimentally in RPgV, will be informative in furthering our understanding about this generic property of persistent RNA viruses.

The recent documentation of substantially greater sequence diversity of hepaciviruses and pegiviruses and the existence of mosaic features and variability in genome organizations may lead to a reevaluation of the original criteria by which the genera *Hepacivirus* and *Pegivirus* were originally defined (22). As reviewed above, possession of a type IV IRES does not define hepaciviruses, as this element is present among group 2 pegiviruses (Fig. 4). Similarly, some hepaciviruses have pegivirus-like 5′ UTR sequences. Possession of a core protein may turn out to be a similarly variable attribute, with several pegiviruses possessing coding sequences upstream of E1 and, in the case of HHpgV-1, a potential cytoplasmic location that may confer a biological function analogous to the hepacivirus core protein. Nevertheless, phylogenetic analysis of conserved genes in the nonstructural genome region clearly maintains separate, bootstrap-supported groupings of pegiviruses and hepaciviruses, while the E1/E2/X structural gene block of pegiviruses shows little or no homology with the E1 and E2 protein genes of hepaciviruses. Future characterization of these viruses in a wider range of hosts will provide more information on the value and durability of current criteria for *Hepacivirus* and *Pegivirus* genus assignments.

Comprehensive virome analysis helped us in the identification of HHpgV-1, a new human virus. We note that virome analysis also identified approximately 40% sequence reads that remained unclassified due to their low sequence similarity to known viruses or to any other known sequence. Viruses show remarkable diversity in terms of replication-expression strategy and genomic complexity (46). After acquiring the HHpgV-1 genome, we reverse mapped it to the unclassified virome reads and noticed that several unclassified sequence reads were part of the HHPgV-1 genome (core, E1, E2, NS4A, and NS4B genes). These findings suggest that other fractions of these unclassified reads may represent unrecognized highly diverse viruses or the most rapidly evolving part of their genomes. The genomic sequence data of HHpgV-1 will be helpful in designing molecular reagents to study its prevalence, persistence, transmission, and disease association in humans. Besides providing information regarding a new human virus infection, identification of HHpgV-1 expands our knowledge of the origin, genetic diversity, and evolution of hepaciviruses and pegiviruses.

MATERIALS AND METHODS

Human samples. All human serum samples used this study were obtained from the NHLBI sample repository and were collected as part of two studies, the Transfusion-Transmitted Viruses Study (TTVS) and the Multicenter Hemophilia Cohort Study (MHCS). The study plan and research were approved by Columbia University Medical Center, no. IRB-AAAN5157. TTVS samples were collected at 4 participating blood centers distributed across the United States from July 1974 through June 1980. The MHCS-I evaluated and prospectively followed patients with hemophilia or a related coagulation disorder form 1982 to 1986. Adults and children who had a congenital coagulation disorder (hemophilia A or B [congenital factor VIII or IX deficiency]), von Willebrand's disease, or another disorder were enrolled from 8 collaborating hemophilia centers in the United States between 1982 and 1985. Four additional centers from the United States and 4 centers from Europe joined the study between 1987 and 1990. In MHCS-I, subjects were evaluated semiannually with a standardized physical examination, abstraction of medical records, and phlebotomy (4–8).

Virome amplification, sequencing, and bioinformatic analysis. Serum samples were filtered through a 0.45-μm filter and treated with nucleases to digest free nucleic acids (NAs) for enrichment of viral NA and then extracted in NucliSens buffer using the automated easyMAG system (bio-Mérieux, United States). Total RNA extracts were reverse transcribed using a SuperScript III kit (Invitrogen Life Technologies) with random hexamer primers. The cDNA was RNase H treated prior to second-strand DNA synthesis using Klenow fragment (3′–5′ exonuclease negative) (New England Biolabs). The double-stranded cDNA was sheared to a 200-bp average fragment length using a Covaris E210 focused ultrasonicator. Sheared DNA was purified and used for Illumina library construction using a Kapa library preparation kit (KK8234; Kapa Biosystems) and SeqCap EZ Library SR (Nimblegen, Roche). The sequencing libraries were quantified using an Agilent Bioanalyzer 2100. Samples with low concentrations were amplified by increasing PCR cycle numbers from 9 to 14. All sequencing was done on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA), yielding an average of 150 million reads per sequencing lane. Virome analysis included only those sequences that were most closely related to vertebrate viruses and was done using MEGAN version 5.10.5 (47). Sequence data were demultiplexed using Illumina software to generate FastQ files for individual samples. Sequences were filtered using Q30 and mapped against reference genomes from GenBank with Bowtie2 mapper 2.0.6 (http://bowtie-bio.sourceforge.net). SAMtools (v 0.1.19) were used to generate the consensus genomes and coverage statistics. Integrative Genomics Viewer (v 2/3/55; Broad Institute) was used to generate the sequence coverage plots. Host-derived sequences were identified using Bowtie2-based sequence mappings against the reference host genomes downloaded from the NCBI database. Sequencing data obtained from the clinical samples were preprocessed using PRINSEQ (v 0.20.2) software, and primer-trimmed, quality score-filtered reads were aligned against the host reference databases to remove the host background. The host-subtracted sequence reads were *de novo* assembled using MIRA (v 4.0) or SOAPdenovo2 (v 2.04) assemblers, and then contigs and unique single sequences were subjected to a homology search using MegaBlast against the GenBank nucleotide database. Sequences that showed poor or no homology at the nucleotide level were subjected to a search with BLASTx against the viral GenBank protein database. Viral sequences from BLASTx analysis were subjected to another round of BLASTx homology search against entire GenBank protein database to correct for biased E values due to the smaller size of the virus-only database, and taxonomy was reassigned. Based on the contigs identified for different viral strains, GenBank sequences were downloaded and used for mapping the whole data set to recover partial or complete genomes. The descriptive statistics mean, standard deviation (SD), and median were calculated for viral count and viral rate. The pretransfusion viral rate and the posttransfusion rate were compared by paired Wilcoxon rank sum test.

Screening assays of HHpgV-1. The polyprotein and UTR of HHpgV-1 were aligned to all known hepaciviruses and pegiviruses. Nucleotide and amino acid motifs showing relative conservation among different virus lineages were used to make primers for screening of samples for HHpgV-1 and related variants. All PCR mixtures used AmpliTaq Gold 360 master mix (catalog no. 4398881; Applied Biosystems) and 2 μl of cDNA. The HHpgV-1 helicase gene heminested PCR assay used primer pair HHpgV-ak1 (5′-GTTGTATTCGCCACAGCCAC-3′) and HHpgV-ak2 (5′-TCAAAGTTTCCTGTGTAGCCTGT-3′) in the first round of PCR and the pair HHpgV-ak3 (5′-GTATTCGCCACAGCCACTCC-3′) and HHpgV-ak2 in the second round of PCR. For the first round, the PCR cycle included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 56°C for 1 min, and 72°C for 1 min, 30 cycles of 95°C for 30 s, 55°C for 1 min, and 72°C for 1 min, and a final extension at 72°C for 5 min. In first 10 cycles, the annealing temperature was ramped down by 0.5°C each cycle to allow mutation tolerance during primer hybridization. For the second round, PCR conditions included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 60°C for 1 min, and 72°C for 1 min, 30 cycles of 95°C for 30 s, 56°C for 1 min, and 72°C for 1 min, and a final extension at 72°C for 5 min. The hepatitis C virus PCR assay used the primer pair HCV-ak-F1 (5′-GCGCCCATCACGGCITAYGC-3′) and HCV-ak-R1 (5′-GTCTTGGTCCACRTTGGTRTACAT-3′) in the first round of PCR and the pair HCV-ak-F2 (5′-GCCCATCACGGCGTAYGCNCARCA-3′) and HCV-ak-R2(5′-CTTGGTCCACRTTGGTRTACATYTG-3′) in the second round of PCR. For the first round, the PCR cycle included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 58°C for 1 min, and 72°C for 45 s, 30 cycles of 95°C for 30 s, 55°C for 1 min, and 72°C for 45 s, and a final extension at 72°C for 5 min. In the first 10 cycles, the annealing temperature was ramped down by 0.5°C each cycle to allow mutation tolerance during primer hybridization. For the second round, PCR conditions included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 64°C for 1 min, and 72°C for 45 s, 30 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 40 s, and a final extension at 72°C for 5 min. The GBVc-NS3 PCR assay used the primer pair GBVc-ak-F1 (5′-CCTTGGACCCAGGTICCNACIGA-3′) and GBVc-ak-R1 (5′-CCTGGTGGGGTRGCIGTNGC-3′) in the first round of PCR and the pair GBVc-ak-F2 (5′-GGACCCAGGTGCCIACGGAYGC-3′) and GBVc-ak-R2 (5′-CCTGGTGGGGTRGCGGTNGCRTA-3′) in the second round of PCR. For the first round, the PCR cycle included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 63°C for 1 min, and 72°C for 30 s, 30 cycles of 95°C for 30 s, 60°C for 1 min, and 72°C for 40 s, and a final extension at 72°C for 5 min. In first 10 cycles, the annealing temperature was ramped down by 0.5°C each cycle to allow mutation tolerance during primer hybridization. For the second round, PCR conditions included 8 min of denaturation at 95°C, 10 cycles of 95°C for 40 s, 71°C for 1 min, and 72°C for 30 s, 30 cycles of 95°C for 30 s, 64°C for 30 s, and 72°C for 40 s, and a final extension at 72°C for 5 min.

Phylogenetic and RNA secondary structure analysis. Nucleotide sequences (5′ UTRs) and translated protein sequences (coding regions) were aligned using the program MUSCLE as implemented in the SSE package (37). Sequence divergence scans were performed, and summary values for different genome regions were generated by the program Sequence Distance in the SSE package. Bootstrapped maximum-likelihood trees for the NS3 helicase region and NS5B polymerase regions were constructed using the maximum-likelihood algorithm as implemented in MEGA, using the optimal model selected by Model Test. For both regions, this was the Le-Gascuel model with a gamma distribution (5 rates) and invariant sites (LG + γ + I). Phylogenetic analysis of NS3 nucleotide sequences was performed similarly using the Jukes-Cantor substitution model (selected by Model Test for this data set).

RNA structures were predicted by Mfold and by homology searching and structural alignment with bases conserved in other hepaciviruses. Structure prediction for the pseudoknot region in HCV (IIIf) and homologous pairings in other hepaciviruses could not be predicted by Mfold or other conventional RNA secondary-structure prediction algorithms.

Structure predictions upstream of stem-loop III were performed by Mfold.

Cleavage sites between HHpgV-1 structural proteins were predicted using the SignalP version 4.1 program (16); these were concordant with positions predicted from the sequence alignment. Cleavage between nonstructural genes was predicted by alignment of the NS2/NS3, NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B sites previously proposed for simian pegiviruses (18).

HHpgV-1 and other pegivirus sequences were analyzed for the presence of GORS by comparing folding energies of consecutive fragments of nucleotide sequence with random sequence order controls using the program's MFED scan in the SEE package (37). Minimum folding energies (MFEs) of rodent virus genomes were calculated by using the default setting in the program Zipfold. MFE results were expressed as MFEDs, i.e., the percentage difference between the MFE of the native sequence from that of the mean value of the 50 sequence order-randomized controls (32).

Nucleotide sequence accession number. The complete genome of HHpgV-1 has been submitted to GenBank under accession number KT439329.

REFERENCES

1. **Dwyre DM, Fernando LP, Holland PV**. 2011. Hepatitis B, hepatitis C and HIV transfusion-transmitted infections in the 21st century. Vox Sang **100**: 92–98. http://dx.doi.org/10.1111/j.1423-0410.2010.01426.x.

2. **Kuhns MC, Busch MP**. 2006. New strategies for blood donor screening for hepatitis B virus: nucleic acid testing versus immunoassay methods. Mol Diagn Ther **10**:77–91. http://dx.doi.org/10.1007/BF03256447.

3. **Goedert JJ, Brown DL, Hoots K, Sherman KE**. 2004. Human immunodeficiency and hepatitis virus infections and their associated conditions and treatments among people with haemophilia. Haemophilia **10**(Suppl 4): 205–210. http://dx.doi.org/10.1111/j.1365-2516.2004.00997.x.

4. **Ehmann WC, Eyster ME, Wilson SE, Andes WA, Goedert JJ**. 1994. Relationship of CD4 lymphocyte counts to survival in a cohort of hemophiliacs infected with HIV. Multicenter Hemophilia Cohort Study. J Acquir Immune Defic Syndr **7**:1095–1098.

5. **Eyster ME, Fried MW, Di Bisceglie AM, Goedert JJ**. 1994. Increasing hepatitis C virus RNA levels in hemophiliacs: relationship to human immunodeficiency virus infection and liver disease. Multicenter Hemophilia Cohort Study. Blood **84**:1020–1023.

6. **Goedert JJ, Cohen AR, Kessler CM, Eichinger S, Seremetis SV, Rabkin CS, Yellin FJ, Rosenberg PS, Aledort LM**. 1994. Risks of immunodeficiency, AIDS, and death related to purity of factor VIII concentrate. Multicenter Hemophilia Cohort Study. Lancet **344**:791–792. http://dx.doi.org/10.1016/S0140-6736(94)92345-0.

7. **Kroner BL, Rosenberg PS, Aledort LM, Alvord WG, Goedert JJ**. 1994. HIV-1 infection incidence among persons with hemophilia in the United States and Western Europe, 1978–1990. Multicenter Hemophilia Cohort Study. J Acquir Immune Defic Syndr **7**:279–286.

8. **Rosenberg PS, Goedert JJ, Biggar RJ**. 1994. Effect of age at seroconversion on the natural AIDS incubation distribution. Multicenter Hemophilia Cohort Study and the International Registry of Seroconverters. AIDS **8**:803–810. http://dx.doi.org/10.1097/00002030-199406000-00013.

9. **Kapoor A, Lipkin WI**. 2014. Virus discovery in the 21st century. In eLS. John Wiley & Sons, New York, NY.

10. **Kapoor A, Mehta N, Esper F, Poljsak-Prijatelj M, Quan PL, Qaisar N, Delwart E, Lipkin WI**. 2010. Identification and characterization of a new bocavirus species in gorillas. PLoS ONE **5**:e11948. http://dx.doi.org/10.1371/journal.pone.0011948.

11. **Kapoor A, Simmonds P, Cullen JM, Scheel TK, Medina JL, Giannitti F, Nishiuchi E, Brock KV, Burbelo PD, Rice CM, Lipkin WI**. 2013. Identification of a pegivirus (GB virus-like virus) that infects horses. J Virol **87**:7185–7190. http://dx.doi.org/10.1128/JVI.00324-13.

12. **Kapoor A, Simmonds P, Scheel TK, Hjelle B, Cullen JM, Burbelo PD, Chauhan LV, Duraisamy R, Sanchez Leon M, Jain K, Vandegrift KJ, Calisher CH, Rice CM, Lipkin WI**. 2013. Identification of rodent homologs of hepatitis C virus and pegiviruses. mBio **4**:e00216-13. http://dx.doi.org/10.1128/mBio.00216-13.

13. **Kapoor A, Tesh RB, Duraisamy R, Popov VL, Travassos da Rosa APA, Lipkin WI**. 2013. A novel mosquito-borne orbivirus species found in South-East Asia. J Gen Virol **94**:1051–1057. http://dx.doi.org/10.1099/vir.0.046748-0.

14. **Delwart E**. 2013. A roadmap to the human virome. PLoS Pathog **9**:e1003146. http://dx.doi.org/10.1371/journal.ppat.1003146.

15. **Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E**. 2005. New DNA viruses identified in patients with acute viral infection syndrome. J Virol **79**:8230–8236. http://dx.doi.org/10.1128/JVI.79.13.8230-8236.2005.

16. **Osburn WO, Fisher BE, Dowd KA, Urban G, Liu L, Ray SC, Thomas DL, Cox AL**. 2010. Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. Gastroenterology **138**:315–324. http://dx.doi.org/10.1053/j.gastro.2009.09.017.

17. **Grebely J, Thomas DL, Dore GJ**. 2009. HCV reinfection studies and the door to vaccine development. J Hepatol **51**:628–631. http://dx.doi.org/10.1016/j.jhep.2009.07.002.

18. **Walker CM, Grakoui A**. 2015. Hepatitis C virus: why do we need a vaccine to prevent a curable persistent infection? Curr Opin Immunol **35**: 137–143. http://dx.doi.org/10.1016/j.coi.2015.06.010.

19. **Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M**. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244**:359–362. http://dx.doi.org/10.1126/science.2523562.

20. **Scheel TKH, Simmonds P, Kapoor A**. 2015. Surveying the global virome: identification and characterization of HCV-related animal hepaciviruses. Antiviral Res **115**:83–93. http://dx.doi.org/10.1016/j.antiviral.2014.12.014.

21. **Pfaender S, Brown RJ, Pietschmann T, Steinmann E**. 2014. Natural reservoirs for homologs of hepatitis C virus. Emerg Microbes Infect **3**:e21. http://dx.doi.org/10.1038/emi.2014.19.

22. **Stapleton JT, Foung S, Muerhoff AS, Bukh J, Simmonds P**. 2011. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. J Gen Virol **92**:233–246. http://dx.doi.org/10.1099/vir.0.027490-0.

23. **Burbelo PD, Dubovi EJ, Simmonds P, Medina JL, Henriquez JA, Mishra N, Wagner J, Tokarz R, Cullen JM, Iadarola MJ, Rice CM, Lipkin WI, Kapoor A**. 2012. Serology-enabled discovery of genetically diverse hepaciviruses in a new host. J Virol **86**:6171–6178. http://dx.doi.org/10.1128/JVI.00250-12.

24. **Lyons S, Kapoor A, Sharp C, Schneider BS, Wolfe ND, Culshaw G, Corcoran B, McGorum BC, Simmonds P**. 2012. Nonprimate hepaciviruses in domestic horses, United Kingdom. Emerg Infect Dis **18**: 1976–1982. http://dx.doi.org/10.3201/eid1812.120498.

25. **Kapoor A, Simmonds P, Gerold G, Qaisar N, Jain K, Henriquez JA, Firth C, Hirschberg DL, Rice CM, Shields S, Lipkin WI**. 2011. Characterization of a canine homolog of hepatitis C virus. Proc Natl Acad Sci U S A **108**:11608–11613. http://dx.doi.org/10.1073/pnas.1101794108.

26. **Drexler JF, Corman VM, Müller MA, Lukashev AN, Gmyl A, Coutard B, Adam A, Ritz D, Leijten LM, van Riel D, Kallies R, Klose SM, Gloza-Rausch F, Binger T, Annan A, Adu-Sarkodie Y, Oppong S, Bourgarel M, Rupp D, Hoffmann B, Schlegel M, Kummerer BM, Kruger DH, Schmidt-Chanasit J, Setien AA, Cottontail VM, Hemachudha T, Wacharapluesadee S, Osterrieder K, Bartenschlager R, Matthee S, Beer M, Kuiken T, Reusken C, Leroy EM, Ulrich RG, Drosten C**. 2013. Evidence for novel hepaciviruses in rodents. PLoS Pathog **9**:e1003438. http://dx.doi.org/10.1371/journal.ppat.1003438.

27. **Quan P-, Firth C, Conte JM, Williams SH, Zambrana-Torrelio CM, Anthony SJ, Ellison JA, Gilbert AT, Kuzmin IV, Niezgoda M, Osinubi MOV, Recuenco S, Markotter W, Breiman RF, Kalemba L, Malekani J, Lindblade KA, Rostal MK, Ojeda-Flores R, Suzan G, Davis LB, Blau DM, Ogunkoya AB, Alvarez Castillo DA, Moran D, Ngam S, Akaibe D,**

Agwanda B, Briese T, Epstein JH, Daszak P, Rupprecht CE, Holmes EC, Lipkin WI. 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proc Natl Acad Sci U S A 110:8194–8199. http://dx.doi.org/10.1073/pnas.1303037110.

28. Lauck M, Sibley SD, Lara J, Purdy MA, Khudyakov Y, Hyeroba D, Tumukunde A, Weny G, Switzer WM, Chapman CA, Hughes AL, Friedrich TC, O'Connor DH, Goldberg TL. 2013. A novel hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. J Virol 87:8971–8981. http://dx.doi.org/10.1128/JVI.00888-13.

29. Epstein JH, Quan PL, Briese T, Street C, Jabado O, Conlan S, Ali Khan S, Verdugo D, Hossain MJ, Hutchison SK, Egholm M, Luby SP, Daszak P, Lipkin WI. 2010. Identification of GBV-D, a novel GB-like flavivirus from Old World frugivorous bats (Pteropus giganteus) in Bangladesh. PLoS Pathog 6:e1000972. http://dx.doi.org/10.1371/journal.ppat.1000972.

30. Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, Van Blaricum AJ, Tennant BC, Kistler AL. 2013. Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. Proc Natl Acad Sci U S A 110:E1407–E1415. http://dx.doi.org/10.1073/pnas.1219217110.

31. Bernardin F, Operskalski E, Busch M, Delwart E. 2010. Transfusion transmission of highly prevalent commensal human viruses. Transfusion 50:2474–2483. http://dx.doi.org/10.1111/j.1537-2995.2010.02699.x.

32. Hsu H, Ni Y, Chen H, Kao J, Chang M. 2003. TT virus infection in healthy children, children after blood transfusion, and children with non-A to E hepatitis or other liver diseases in Taiwan. J Med Virol 69:66–71. http://dx.doi.org/10.1002/jmv.10249.

33. Burbelo PD, Ching KH, Esper F, Iadarola MJ, Delwart E, Lipkin WI, Kapoor A. 2011. Serological studies confirm the novel astrovirus HMOAstV-C as a highly prevalent human infectious agent. PLoS One 6:e22576. http://dx.doi.org/10.1371/journal.pone.0022576.

34. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Jr., Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol 87:11966–11977. http://dx.doi.org/10.1128/JVI.02323-13.

35. Weiner AJ, Kuo G, Bradley DW, Bonino F, Saracco G, Lee C, Rosenblatt J, Choo QL, Houghton M. 1990. Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet 335:1–3. http://dx.doi.org/10.1016/0140-6736(90)90134-Q.

36. Asnani M, Kumar P, Hellen CU. 2015. Widespread distribution and structural diversity of type IV IRESs in members of Picornaviridae. Virology 478:61–74. http://dx.doi.org/10.1016/j.virol.2015.02.016.

37. Firth AE, Brierley I. 2012. Non-canonical translation in RNA viruses. J Gen Virol 93:1385–1409. http://dx.doi.org/10.1099/vir.0.042499-0.

38. Reineke LC, Lloyd RE. 2011. Animal virus schemes for translation dominance. Curr Opin Virol 1:363–372. http://dx.doi.org/10.1016/j.coviro.2011.10.009.

39. Randall RE, Goodbourn S. 2008. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. J Gen Virol 89:1–47. http://dx.doi.org/10.1099/vir.0.83391-0.

40. Helle F, Duverlie G, Dubuisson J. 2011. The hepatitis C virus glycan shield and evasion of the humoral immune response. Viruses 3:1909–1932. http://dx.doi.org/10.3390/v3101909.

41. Phogat S, Wyatt RT, Karlsson Hedestam GB. 2007. Inhibition of HIV-1 entry by antibodies: potential viral and cellular targets. J Intern Med 262:26–43. http://dx.doi.org/10.1111/j.1365-2796.2007.01820.x.

42. Zeisel MB, Felmlee DJ, Baumert TF. 2013. Hepatitis C virus entry. Curr Top Microbiol Immunol 369:87–112. http://dx.doi.org/10.1007/978-3-642-27340-7_4.

43. Davis M, Sagan SM, Pezacki JP, Evans DJ, Simmonds P. 2008. Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. J Virol 82:11824–11836. http://dx.doi.org/10.1128/JVI.01078-08.

44. Simmonds P, Tuplin A, Evans DJ. 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. RNA 10:1337–1351. http://dx.doi.org/10.1261/rna.7640104.

45. McFadden N, Arias A, Dry I, Bailey D, Witteveldt J, Evans DJ, Goodfellow I, Simmonds P. 2013. Influence of genome-scale RNA structure disruption on the replication of murine norovirus—similar replication kinetics in cell culture but attenuation of viral fitness in vivo. Nucleic Acids Res 41:6316–6331. http://dx.doi.org/10.1093/nar/gkt334.

46. Koonin EV, Dolja VV. 2006. Evolution of complexity in the viral world: the dawn of a new vision. Virus Res 117:1–4. http://dx.doi.org/10.1016/j.virusres.2006.01.018.

47. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res 21:1552–1560. http://dx.doi.org/10.1101/gr.120618.111.