Journal of Virology

Serology-Enabled Discovery of Genetically Diverse Hepaciviruses in a New Host

Peter D. Burbelo, Edward J. Dubovi, Peter Simmonds, Jan L. Medina, Jose A. Henriquez, Nischay Mishra, Jason Wagner, Rafal Tokarz, John M. Cullen, Michael J. Iadarola, Charles M. Rice, W. Ian Lipkin and Amit Kapoor

	Updated information and services can be found at: http://jvi.asm.org/content/86/11/6171
	These include:
REFERENCES	This article cites 41 articles, 23 of which can be accessed free at: http://jvi.asm.org/content/86/11/6171#ref-list-1
CONTENT ALERTS	Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Journals.ASM.org

JVI

Journals.ASM.org

Serology-Enabled Discovery of Genetically Diverse Hepaciviruses in a New Host

Peter D. Burbelo,[a] Edward J. Dubovi,[b] Peter Simmonds,[c] Jan L. Medina,[d] Jose A. Henriquez,[d] Nischay Mishra,[d] Jason Wagner,[a] Rafal Tokarz,[d] John M. Cullen,[e] Michael J. Iadarola,[a] Charles M. Rice,[f] W. Ian Lipkin,[d] and Amit Kapoor[d]

Neurobiology and Pain Therapeutics Section, Laboratory of Sensory Biology, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA[a]; Animal Health Diagnostic Center, College of Veterinary Medicine at Cornell, Ithaca, New York, USA[b]; Centre for Immunity, Infection and Evolution, Ashworth Laboratories, University of Edinburgh, Edinburgh, United Kingdom[c]; Center for Infection and Immunity, Columbia University, New York, New York, USA[d]; College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, USA[e]; and Center for the Study of Hepatitis C, Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, New York, USA[f]

Genetic and biological characterization of new hepaciviruses infecting animals contributes to our understanding of the ultimate origins of hepatitis C virus (HCV) infection in humans and dramatically enhances our ability to study its pathogenesis using tractable animal models. Animal homologs of HCV include a recently discovered canine hepacivirus (CHV) and GB virus B (GBV-B), both viruses with largely undetermined natural host ranges. Here we used a versatile serology-based approach to determine the natural host of the only known nonprimate hepacivirus (NPHV), CHV, which is also the closest phylogenetic relative of HCV. Recombinant protein expressed from the helicase domain of CHV NS3 was used as antigen in the luciferase immunoprecipitation system (LIPS) assay to screen several nonprimate animal species. Thirty-six samples from 103 horses were immunoreactive, and viral genomic RNA was present in 8 of the 36 seropositive animals and none of the seronegative animals. Complete genome sequences of these 8 genetically diverse NPHVs showed 14% (range, 6.4% to 17.2%) nucleotide sequence divergence, with most changes occurring at synonymous sites. RNA secondary structure prediction of the 383-base 5′ untranslated region of NPHV was refined and extended through mapping of polymorphic sites to unpaired regions or (semi)covariant pairings. Similar approaches were adopted to delineate extensive RNA secondary structures in the coding region of the genome, predicted to form 27 regularly spaced, thermodynamically stable stem-loops. Together, these findings suggest a promising new nonprimate animal model and provide a database that will aid creation of functional NPHV cDNA clones and other novel tools for hepacivirus studies.

The identification and characterization of animal virus homologs provide insights into the pathogenesis of human-pathogenic viruses and, in some instances, *in vivo* models for investigating prevention and treatment of human disease (41). Well-characterized animal viruses include simian immunodeficiency virus, animal poxviruses, herpesviruses, murine norovirus, and woodchuck hepatitis virus (28). Hepatitis C virus (HCV), in contrast, has few known animal relatives (3, 21). Moreover, HCV naturally infects only humans and chimpanzees, resulting in a paucity of animal models for studies of its pathogenesis, immunity, and treatment (14, 26, 32, 40). An estimated 2% of the world's population is chronically infected with HCV. The ability to study hepacivirus pathogenesis in more tractable animal models would dramatically enhance HCV research (14, 30).

The genus *Hepacivirus*, one of four genera in the family *Flaviviridae*, comprises HCV and GB virus B (GBV-B) (38). GBV-B was isolated during laboratory passage of plasma from an individual with unexplained hepatitis through tamarins and other New World monkey species but was never again recovered from a human sample (38). The natural host of GBV-B has thus remained elusive (4, 5, 31, 38). We recently identified canine hepacivirus (CHV) in respiratory samples of domestic dogs (21). CHV is the first nonprimate hepacivirus (NPHV) discovered, and comparative phylogenetic analysis confirmed it as the closest genetic relative of HCV described to date (3, 21). The envelope protein E2 of HCV, for example, is among the most variable portions of its genome, yet it has significant sequence similarity to CHV (21). Furthermore, CHV was detected in canine hepatocytes, although

its link with hepatitis and the persistence of infection was not studied (3).

Recent advances in sequencing technologies have helped identification of many highly divergent human and animal viruses, including CHV (2, 18–24). However, detection of viral nucleic acids alone, particularly in feces or respiratory samples where they may simply represent ingested contaminants, is insufficient to establish infection, let alone disease association (7, 27). Similarly, although the demonstration of specific adaptive immune responses against viral structural and nonstructural proteins cannot in itself prove a causal relationship to disease, it does provide definitive evidence of host infection (2, 7). Here, we describe the usefulness of a sensitive serological assay (7–11) that can be quickly established for new pathogens to screen different animal species for evidence of virus infection and thereby for identification of a novel virus's natural reservoir and host range. Our serology data indicated the presence of CHV-like viruses in horses and led to the genetic characterization of eight novel and genetically diverse nonprimate hepaciviruses (NPHVs).

Burbelo et al.

MATERIALS AND METHODS

Serum samples. Serum samples from different animal species included sera of 80 dogs, 14 rabbits, 81 deer, 84 cows, and 103 horses. All were residual samples collected for diagnostic or commercial use, and investigators had no other sample identifiers, except that all animals were living in the area of New York State. All serum samples were stored at $-80°C$, thawed, and then left at $4°C$ prior to processing for luciferase immunoprecipitation system (LIPS) analysis.

Generation of Ruc-CHV helicase antigen fusion constructs. The template for NS3 serine protease/helicase coding sequence of CHV was generated by reverse transcription-PCR (RT-PCR) amplification from a respiratory sample of a dog (21). Due to the possibility of antibody cross-reactivity with the HCV helicase gene, a previously described fragment (11) encompassing the corresponding helicase region of human HCV was also tested as an antigen control. The primer adapter sequences used to clone each CHV protein fragment are as follows: for CHV NS3, 5′-GAG GGATCCATACACTTCGCAGATATG-3′ and 5′-GAGCTCGAGTCAG GTGTTACAGTCAGTAAC-3′, and for CHV core, 5′-GAGGGATCCAG TAATAAATCTAAAAAC-3′ and 5′-GAGCTCGAGTCAGGCCTCTCCG AAAGATAC-3′. Both protein fragments were subcloned downstream of *Renilla* luciferase (Ruc) using the pREN2 vector (10). DNA sequencing was used to confirm the integrity of the DNA constructs. The helicase protein fragment of CHV used in LIPS assay (amino acid positions 1173 to 1436 of AEC45560) was >32% and >28% different from HCV genotypes in nucleotide and protein sequences, respectively. Plasmid DNA was then prepared from these two different pREN2 expression vectors using a Qiagen Midi preparation kit. Following transfection of mammalian expression vectors, crude protein extracts were obtained as described for use as antigen (8).

LIPS assays. Briefly, animal sera were processed in a 96-well format at room temperature as previously described (6, 8, 9). Serum samples were first diluted 1:10 in assay buffer A (50 mM Tris, pH 7.5, 100 mM NaCl, 5 mM MgCl$_2$, 1% Triton X-100) using a 96-well polypropylene microtiter plate. Antibody titers were measured by adding 40 μl of buffer A, 10 μl of diluted sera (1-μl equivalent), and 1×10^7 light units (LU) of each of the Ruc-CHV and HCV helicase antigen fragments containing crude Cos1 cell extract to wells of a polypropylene plate and incubated for 60 min at room temperature on a rotary shaker. Next, 5 μl of a 30% suspension of Ultralink protein A/G beads (Pierce Biotechnology, Rockford, IL) in phosphate-buffered saline (PBS) was added to the bottom of each well of a 96-well filter HTS plate (Millipore, Bedford, MA). To this filter plate, the 100-μl antigen-antibody reaction mixture was transferred, and the plate was incubated for 60 min at room temperature on a rotary shaker. The washing steps of the retained protein A/G beads were performed on a Biomek Workstation or Tecan plate washer with a vacuum manifold. After the final wash, LU were measured in a Berthold LB 960 Centro microplate luminometer (Berthold Technologies, Bad Wilbad, Germany) using coelenterazine substrate mix (Promega, Madison, WI). All LU data were obtained from the averages of at least two separate experiments. GraphPad Prism software (San Diego, CA) was used for statistical analysis of LIPS data. For the calculation of sensitivity and specificity, a cutoff limit was used, which was derived from the combined value of the mean value plus 3 standard deviations (SD) of the replica samples containing only buffer, Ruc extract, and protein A/G beads. Horse serum samples highly positive for anti-CHV helicase antibodies were used as internal positive controls to standardize the LIPS parameters for testing of all serum samples.

Screening for CHV-like viruses and quantitative PCR. All serum samples were extracted using a Qiagen viral RNA extraction kit. Total RNA was converted to cDNA using random primers and used in nested PCR. The PCR assay for the CHV 5′ untranslated region (UTR) conserved motifs used primers ak70F1/ak370R1 and ak70F2/ak370R2 in the first and second rounds, respectively. The PCR assay for the CHV helicase gene conserved motifs used primers ak4360F1/ak4640R1 and ak4360F2/ ak4640R2 in the first and second rounds, respectively. Details of primers

and PCR conditions used are available on request. All PCR products were sequenced to confirm the presence of CHV-like viruses in the samples. Quantitative PCR to determine the NPHV genome copy number in serum samples was done using a TaqMan assay using primers (Qanti-5UF1, GAGGGAGCTGRAATTCGTGAA, and Qanti-5UR1, GCAAGCATCCT ATCAGACCGT) and probe (6-carboxyfluorescein [FAM])-CCACGAA GGAAGGCGGGGGC-(black hole quencher 1a [BHQ1a]-FAM) targeting 5′ UTR sequences. Plasmids containing the 5′ UTRs of all eight variants were used as controls to optimize the assay and determine its sensitivity.

Genome sequencing and phylogenetic analysis. Sequences with similarity to flaviviruses (sequences that showed expect [E] values of <10^{-10} in NCBI blastp using default parameters) were assembled against the prototype HCV strains. Gaps were filled by primer walking, using specific and degenerate flavivirus primers (21). Both termini of the genome were acquired using rapid amplification of cDNA ends (RACE) (23). Thereafter, sequence validity was tested in 4× genome coverage by classical dideoxy Sanger sequencing. Nucleotide compositions of different flaviviruses and NPHV were determined using EMBOSS compseq (http://emboss .bioinformatics.nl/cgi-bin/emboss/compseq). Nucleotide (5′ UTR) and translated protein sequences (coding regions) were aligned using the program MUSCLE as implemented in the SSE package (34). Sequence divergence scan and summary values for different genome regions were generated by the program Sequence Distance in the SSE package (34). All complete genome sequences were checked for recombination by using the program Genetic Algorithm Recombination Detection (GARD) in the Datamonkey package, which provides an interface with the HyPhy program (25, 33). Default parameters were used with an HKY substitution model and a gamma distribution of 6 discrete rate steps.

RNA structure prediction. Independent of phylogenetic information, the secondary structure of the NPHV 5′ UTR RNA was modeled with MFOLD. Labeling of the predicted structures in the 5′ UTR followed numbering used for reported homologous structures in HCV and GBV-B. The NPHV genome sequence was analyzed for evidence of genome-scale ordered RNA structure (GORS) by comparing folding energies of consecutive fragments of nucleotide sequence with random sequence order controls using the program Folding Energy Scan in the SEE package (34). Minimum folding energies (MFEs) of NPHVs were calculated by using the default setting in the program UNAFOLD. MFE results were expressed as MFE differences (MFEDs), i.e., the percent difference between the MFE of the native sequence and that of the mean value of the 50 sequence order-randomized controls.

RESULTS

Use of the LIPS to search CHV-related viruses in other animal species. CHV, the only known nonprimate hepacivirus and the virus phylogenetically most related to HCV, has the potential to become a valuable model system to study infection by and pathogenesis of hepaciviruses (3). A limited number of canine samples were identified as CHV RNA positive in our previous study (21). To conduct a more thorough survey to identify a repository of CHV-infected samples, we employed the recently developed luciferase-based luciferase immunoprecipitation system (LIPS) seroscreening approach (7). Given the high genetic diversity observed in other RNA viruses, including HCV, we used the evolutionarily conserved CHV helicase protein as the target antigen. The CHV serine protease/helicase (NS3) coding region corresponding to the highly immunogenic region of HCV helicase protein (11) was cloned into pREN2 eukaryotic expression vector, for recombinant expression of an NS3-*Renilla* luciferase fusion protein in COS1 cells (6–10). CHV-luciferase fusion proteins specifically bound to antibodies immobilized on protein A/G beads and were measured by a standard luciferase assay (Fig. 1A).

We tested serum samples of 80 dogs; surprisingly, all were neg-
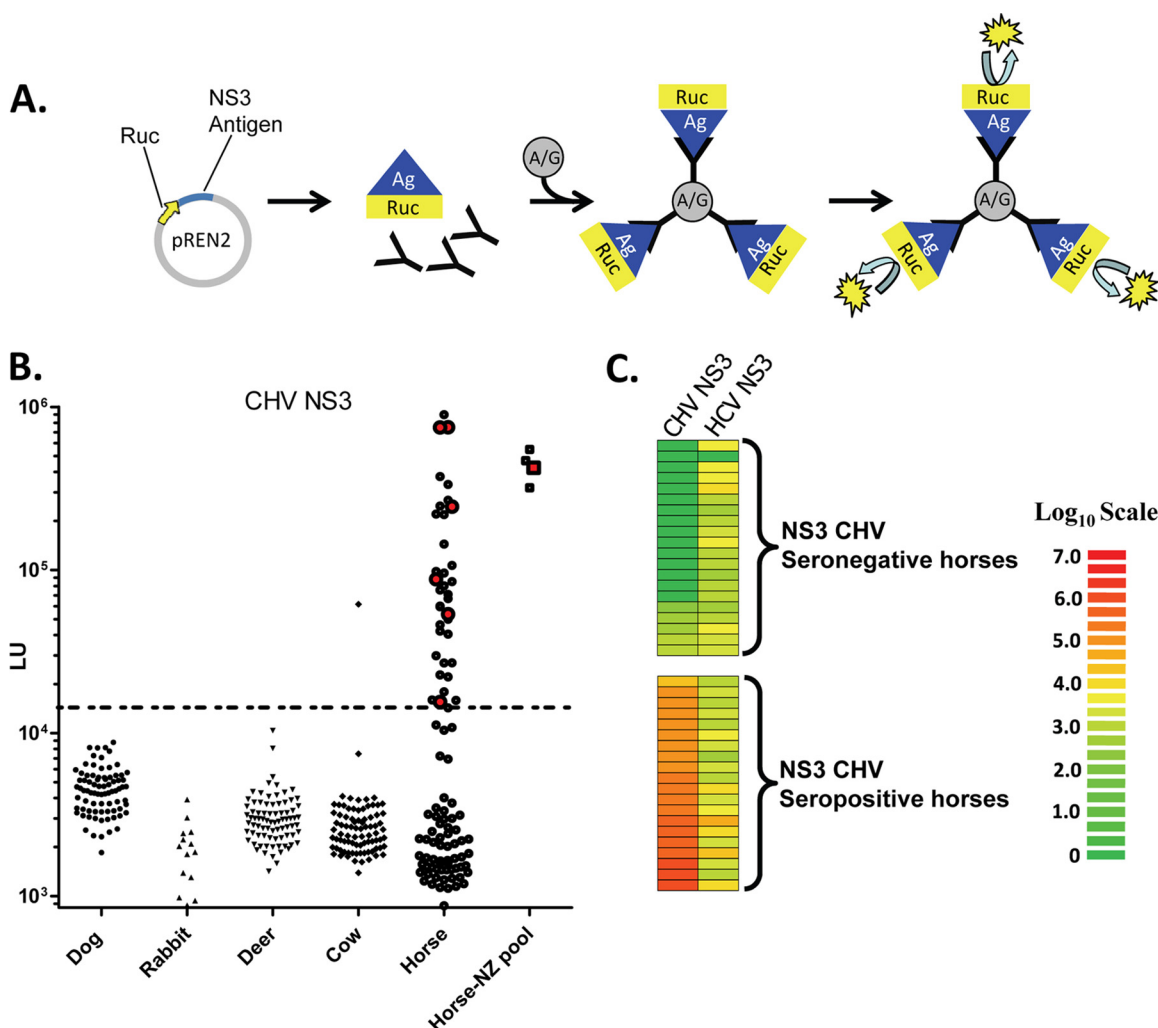
FIG 1 LIPS detection of robust anti-CHV NS3 antibody titers in horses. (A) Schematic of LIPS serological screening. Recombinant NS3 of CHV was genetically fused to the C terminus of *Renilla* luciferase (Ruc) and produced in Cos1 cells. The Ruc-NS3 protein extract was incubated with serum samples, Ruc-NS3 antibody complexes were then captured by protein A/G beads, and light units were measured. (B) LIPS detection of anti-NS3 CHV antibodies in different samples, including 80 dogs, 14 rabbits, 81 deer, 84 cows, 99 U.S. horses, and 4 pooled horse serum samples from New Zealand. The antibody titers from each sample are plotted in light units (LU) on a $\log_{10}$ scale on the y axis, and equine samples positive by PCR are colored red. (C) Heat map analysis of equine and human samples for anti-CHV antibodies. Anti-CHV NS3-seronegative horse ($n = 20$) and -seropositive horse ($n = 20$) samples were analyzed for anti-CHV and anti-HCV NS3 antibodies. Antibody titers against the antigens were $\log_{10}$ transformed, and the levels were then color coded as indicated by the scale on the right, where signal intensities range from high (red) to low (green). Each individual row represents the antibody titers in a single serum sample.

ative. We then tested 81 deer, 84 cows, 103 horses, and 14 rabbits for the presence of anti-CHV helicase IgG. Using a conservative cutoff, high-titered IgG antibodies were detected in 35% of the serum samples from horses, while serum from one cow also showed intermediate reactivity (Fig. 1B). To examine possible antigenic cross-reactivity, we tested the equine seropositive and seronegative samples against helicase protein of HCV (11) and found all these samples to be nonreactive (Fig. 1C). These serological results suggested infection of horses by a hepacivirus(es) more closely related to CHV than HCV in the helicase protein.

Discovery of genetically diverse NPHVs in horses. Expecting that, like HCV infection, hepacivirus infection in animals might persist (30), we developed two broadly reactive PCR assays targeting highly conserved sequence motifs in CHV 5′ UTR and helicase to detect the genetically related hepacivirus genomic RNA in serum samples of horses and cows. Of the 84 cow and 103 horse

serum samples tested, 8 horse samples were positive for hepacivirus RNA. Initial sequencing identified a series of genetically diverse viruses. Comparison of the serological data with the PCR results revealed that the 8 samples positive for hepacivirus RNA were those highly reactive in the LIPS CHV-helicase assay (red circles in Fig. 1B). The primer walking approach was then used to acquire additional genomic sequences of all 8 hepacivirus variants. Since these viruses were found in a different natural host and had substantial genetic diversity compared to CHV, we tentatively named them nonprimate hepaciviruses (NPHVs 1 to 8).

Complete genomes and genetic diversity of NPHVs. We acquired complete genomic sequences of all 8 NPHV variants (GenBank sequence accession numbers JQ434001 to JQ434008) directly from horse serum samples for the purpose of phylogenetic classification and estimation of their genome-wide diversity. Complete genome sequences of NPHV were almost completely
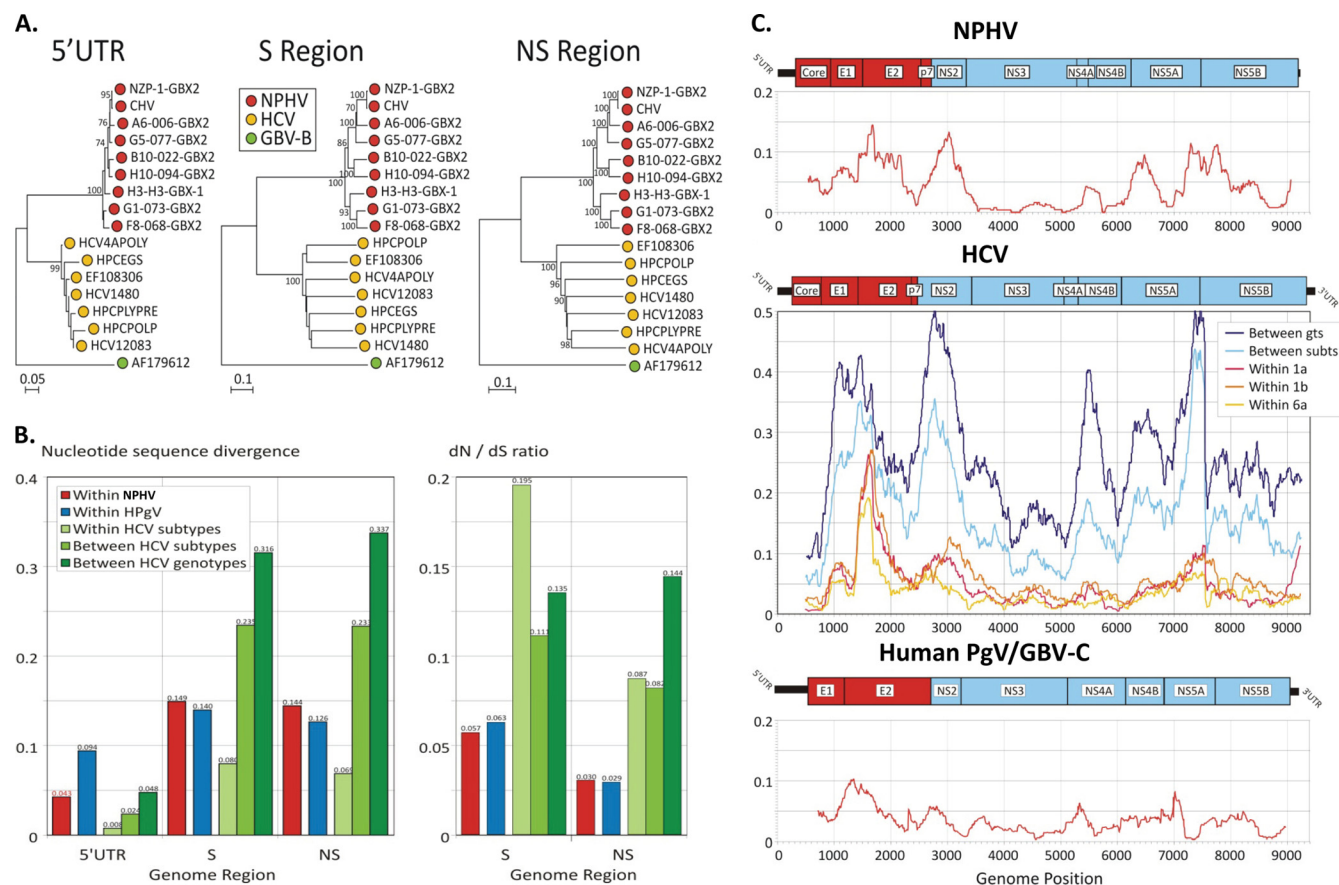
FIG 2 Phylogenetic analysis, genetic composition, and genome-wide divergence scanning of the eight NPHV genomes (A) Neighbor-joining trees of nucleotide sequences from different genome regions of NPHV and corresponding regions of HCV (genotypes 1a to 7a) and GBV-B. Trees were constructed from Jukes-Cantor-corrected pairwise distances calculated using the program MEGA version 5 (34); data sets were bootstrap resampled 500 times to indicate robustness of branching (values of ≥70% shown on branches). (B) Mean nucleotide pairwise distances (uncorrected, y axis) and ratios of synonymous to nonsynonymous Jukes-Cantor distances (dN/dS) between horse-derived hepaciviruses in different genome regions (red bars). These values were compared with equivalent calculations for GBV-C/HPgV (blue bars) and HCV (green bars). (C) Amino acid sequence divergence across the genome of horse-derived NPHV sequences (top plot) and comparison with HCV and GBV-C/HPgV (middle and bottom plots, respectively) using 300-base fragments increasing by 9 bases across each virus alignment (midpoint plotted on y axis, positions numbered using NXP-1-GBX2 as a reference sequence). Genome diagrams above each graph show gene boundaries using the same x axis scale as that on the divergence graph.

colinear, with four sites of 1- to 3-base insertions among variants in the 5′ UTR and three regions of 1- to 4-amino-acid insertions in the coding region. Compared to the original CHV 5′ UTR sequence, all NPHV variants were 17 bases longer in the 5′ UTR, indicating that the originally reported CHV genomic sequence (JF744991) was likely incomplete at the 5′ end. Our many attempts to find the 3′ UTR (X-tail) in all new NPHV genomes remained unsuccessful. Moreover, unlike other hepaciviruses, the 3′ termini of all eight NPHV genomes showed the presence of poly(A) tails of variable lengths.

With the exception of the original CHV variants, which were highly similar to NZP-1-GBX2 (maximum of 0.35% divergence), the 8 horse-derived NPHV sequences showed moderate sequence divergence from each other (6.4% to 17.2%) over the length of the genome (mean, 14.0%). At the nucleotide level, sequences were more divergent in the structural (S) and nonstructural (NS) gene regions (encoding core, E1 and E2 proteins, and NS2-NS5B, respectively) than in the 5′ UTR (Fig. 2), although the S region showed greater amino acid sequence divergence than did NS genes (6.7% compared to 4.0%, respectively). However, most sequence

diversity between NPHV variants occurred at synonymous sites with extremely low dN/dS ratios (ratios of nonsynonymous to synonymous substitutions) in both coding regions (0.057 and 0.030) (Fig. 2). These figures are conservative estimates because calculated Jukes-Cantor-corrected synonymous distances of between 1 and 2 likely substantially underestimate the frequency of multiple substitutions. Despite the differences in sequence variability between sequence regions, phylogenetic relationships between the 8 equine sequences and CHV were consistent across the genome (Fig. 2) with no bootstrap-supported changes in branching order indicative of recombination between NPHV variants (39). Formal testing for the occurrence of recombination was carried out using GARD in the Datamonkey package. One possible recombination breakpoint was detected at position 3525 ($P = 0.019$). However, phylogenetic trees constructed by neighbor joining using maximum composite likelihood distances of either side of the proposed breakpoint (positions 3126 to 3525 and 3526 to 3925) were topologically identical with similar bootstrap support for branches and branch lengths.

Sequence diversity within NPHV was greater than subtype di-

A.

B. Folding energy scan

C. Alifold structure prediction
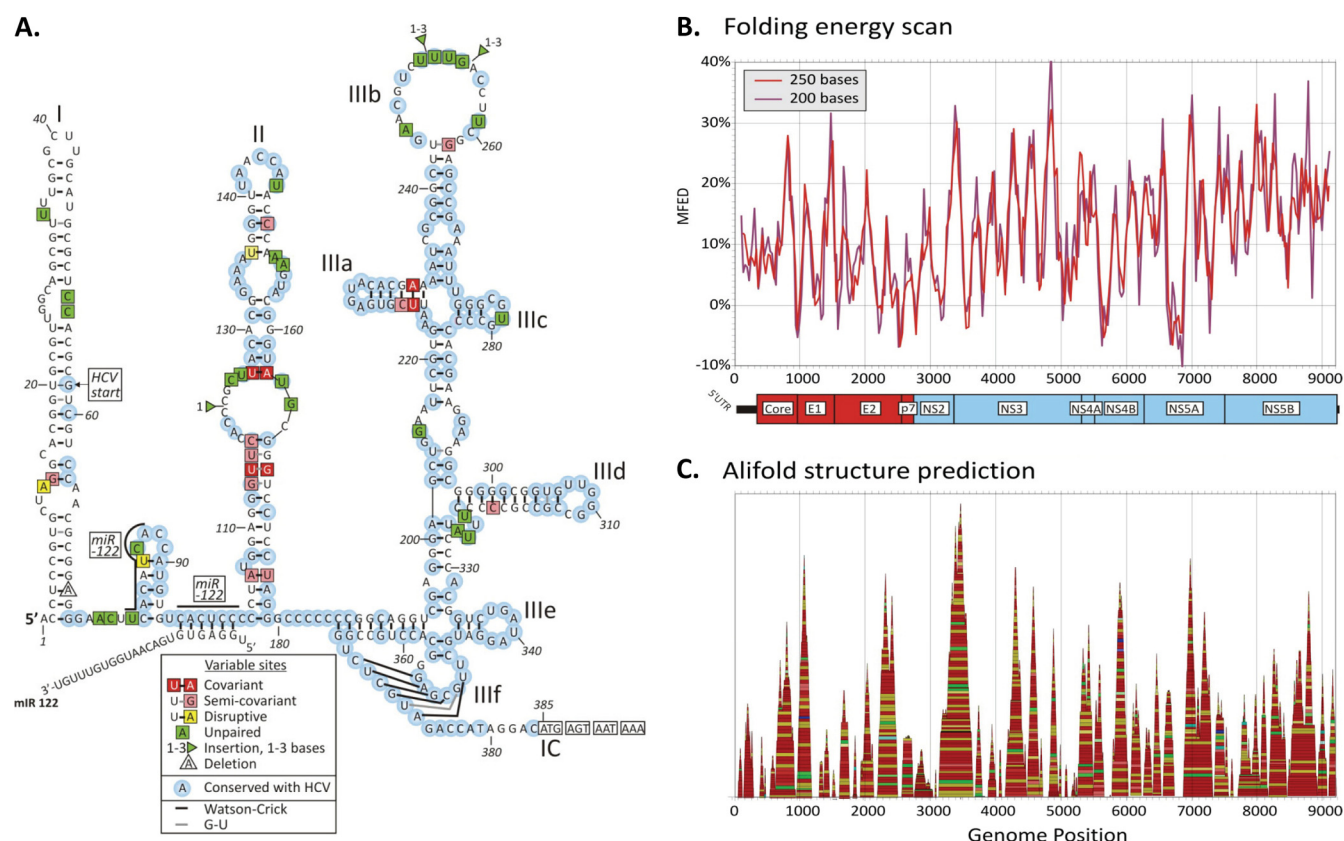
FIG 3 RNA structure analysis of NPHV 5′ UTR and complete genomes. (A) Predicted RNA structure for the 5′ UTR of NPHV based on minimum free energy predictions and comparison with homologous sequences of HCV and GBV-B (21). Bases were numbered using NXP-1-GBX2 as a reference sequence; stem-loops were numbered as in reference 16. Sequences homologous to targets of miRNA-122 (17) are indicated by heavy lines. (B and C) Secondary structure prediction for NPHV genome sequences using mean MFED differences (y axis) of 200- and 250-base fragments (30-base increment; midpoint plotted on x axis) for CHV and the 8 horse-derived hepacivirus sequences (B) and analysis of the 8 NPHV sequences by ALIFOLD using default parameters (C) (see reference 31 for explanation of color coding). Due to restriction in the server, this figure was built as a composite of 6 separate overlapping 2,000-base fragments increasing by 1,500 bases (C).

versity within HCV (mean pairwise distances in S and NS regions ranged from 6 to 10% and 5 to 12% in representative subtypes 1a, 1b, and 6a, compared to 15% and 14% in NPHV, respectively). NPHV diversity in the two regions was, however, substantially less than the mean divergence between HCV subtypes and genotypes (24%/23% and 32%/34%, respectively). HCV additionally differed from NPHV in its greater frequency of nonsynonymous substitutions; although less divergent overall, the mean divergence within subtype amino acid sequences of 1a, 1b, and 6a in the two regions (7.2% and 6.5%) was greater than that within NPHV (6.7% and 4.0%). The pattern of diversity was indeed more similar to that of GB virus C (GBV-C)/hepatitis G virus (29, 36, 37) or the proposed revised name human pegivirus (HPgV) (38). Diversity among human variants occurred at a similar level (14% and 12.5% nucleotide sequence divergence in S and NS regions, respectively) and similarly low dN/dS ratios (0.063 and 0.029) (Fig. 2).

RNA secondary structures in the NPHV 5′ UTR and coding region. The availability of multiple sequences from NPHVs enabled verification and refinement of the 5′ UTR secondary structure prediction as well as an exploration of the nature of large-scale RNA structure in the coding part of the genome. The additional 17 bases at the 5′ end of the 5′ UTR extend the terminal loop, creating a conserved, thermodynamically supported struc-

ture among a clear majority of predicted energetically most favored sequences and the first four suboptimal folding sequences of the 6 sequences complete to the 5′ end. This stem-loop is both larger and more conserved than the structure found within the homologous region in HCV. 5′ UTR sequences showed a mean divergence of approximately 4% between horse-derived NPHV variants, and the distribution of this variability was investigated to determine whether substitutions could be accommodated within the previously proposed secondary structure (Fig. 3A) (21). Most of the 44 polymorphic sites occurred in regions of no predicted base-pairing ($n = 26$; 59% [green boxes]). All but two of the remainder were covariant (i.e., substitutions occurred in pairs to maintain base-pairing; $n = 6$) or semicovariant (G-C ↔ G-U or A-U ↔ G-U; $n = 10$). All insertions/deletions (green triangles) occurred in unpaired loop regions (in stem-loops II and IIIb) or could be accommodated through changing pairing partners (stem-loop I) (Fig. 3A).

Variability in the 5′ UTR sequences was concentrated in stem-loops II and IIIb and the region homologous to the microRNA 122 (miRNA-122) binding site 1 in HCV (17, 21). In general, regions that were conserved between NPHV and HCV (blue circles) were invariant between NPHV variants, while other regions such as the IIIb terminal region were variable in both sequence and length in

both viruses (Fig. 3A). Regions of the internal ribosome entry site (IRES) with known or suspected functional roles in ribosome binding/translation initiation were invariant in NPHV and mostly identical in sequence to homologous regions in HCV (16). The exception was the base-paired region between position 5′-185-193 and 5′-357-365, which was nonhomologous to paired base regions in HCV.

Sequence variability and the use of phylogenetic information (such as the occurrence of covariant sites) were used to explore RNA secondary structure in the coding region of the genome (Fig. 3B). Previous thermodynamic folding analysis of the CHV genome revealed a 14% free energy difference between its minimum folding energy (MFE) and that of sequence order-randomized controls, observations consistent with the presence of genome-scale ordered RNA structure in the CHV genome (21). MFE differences (MFEDs) in the coding region of the 8 horse-derived hepacivirus sequences ranged from 12.5% to 13.9% (mean, 13.0%). MFEDs of successive fragments with lengths between 250 and 400 bases revealed the presence of 27 regularly spaced stem-loops running through the coding part of the genome (Fig. 3B). Mean stem-loop separations (between peak MFED values) were 295 (standard deviation, ±80) and 306 (±71) bases in separate analyses using fragment lengths of 250 and 200 bases, respectively, for scanning. Positions and spacing of structures predicted by MFED scanning were consistent with ALIFOLD (15), which computes pairing likelihoods based on phylogenetic conservation and covariance weighted structure prediction on an underlying thermodynamic model (Fig. 3C). Through analysis of the predicted pairings, the substantial sequence diversity between NPHV sequences in coding regions (14%) was compensated by semicovariant and fully covariant sites and concentration of polymorphisms in predicted unpaired terminal loop regions analogous to the pattern observed in the 5′ UTR. A full analysis of the coding region of NPHV and other viruses with large-scale RNA secondary is in preparation.

DISCUSSION

HCV and its genetically related viruses were considered to be restricted to primates until the recent discovery of CHV indicated a wider host range (3, 21). Initially, CHV was found in dogs, but our subsequent efforts to find similar viruses in canids remained largely unsuccessful. To explore the hypothesis that the natural host of CHV was indeed another mammalian species (and that the outbreaks of infection in dog kennels were due to cross-species virus transmission), a wide range of other nonprimate animal species were tested. We chose to use serology for its advantages of tolerance for sequence divergence, a capacity to detect resolved as well as acute or active infections, and high sample throughput. We expected that like other RNA viruses, including HCV, different CHV variants would be genetically diverse, and we therefore synthesized antigen from the helicase domain of NS3, the most conserved viral protein in the genome and one that shows documented antigenicity in HCV (11).

Serology provided evidence of CHV-like virus infection of horses that was confirmed by the detection of diverse viral genomes in 8 of the 36 samples with anti-NS3 reactivity. Horses are also known to support replication of several other flaviviruses, including the vector-borne West Nile virus, Japanese encephalitis virus, dengue virus, and St. Louis encephalitis virus, which are transmitted between several mammalian species, including hu-

mans (1, 12). Although most of the NPHV variants detected in horses were genetically distinct from CHV infecting dogs, one found in a commercial horse serum pool (from New Zealand) was almost identical to CHV, providing direct evidence for an ability of NPHV to jump species. We tested all the dog serum samples with PCR and found no positives; however, the absence of NPHV infection in dogs from a specific geographical location might not represent the global ecology of NPHV and should be studied further. NPHV genomes were not detected in any of the 83 cow serum samples despite rare seropositivity (Fig. 1B), suggesting either further cross-species transmission events or widespread distribution of further, potentially equally diverse NPHV variants. The frequent detection of viremia in seropositive horses (22%) provides evidence that infections were persistent (copresence of IgG antibodies and viral genomes). In this respect, NHPV is unlike HCV, which persistently infects over 50% of humans exposed, and more similar to GBV-C/HPgV (approximately 25% persistence). Although our failure to detect viral sequences in more than half of the seropositive horses may reflect sequence divergence that confounds consensus PCR, we have an alternative hypothesis whereby NPHV may be cleared in the majority of equine hosts. If confirmed, investigation of the mechanism by which this occurs could yield insights that will lead to new strategies for management of human HCV infection.

Comparative sequence analysis of related viruses can be used to identify evolutionarily conserved and therefore functionally important genomic regions. Particularly striking was the predicted secondary and pseudoknot (tertiary) structure of the 5′ UTR of NPHV and its structural conservation with homologous regions in HCV and GBV-B (Fig. 3A). Although the validity of structure predictions in this region requires verification by nuclease mapping or by the more recently developed SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) analysis methods, the pattern of sequence changes (the occurrence of multiple covariant and semicovariant changes in predicted paired regions and the restriction of noncompensated substitutions to unpaired regions) provides strong evolutionary support for the proposed structure model. Most evident was the structural similarity of the stem-loop III region of NPHV to homologous regions forming the IRES in HCV and pestiviruses and type IV IRESs in picornaviruses. An unexpected structural difference existed, however, in the extreme 5′ end of the 5′ UTR. Sequence similarity to the equivalent, much shorter region in HCV (16 bases) diminished in the first 74 bases of the NPHV 5′ UTR sequence. NPHV forms a predicted thermodynamically stable stem-loop encompassing positions 2 to 74 that is much larger than that described in the equivalent region of HCV. While the distribution of covariant and noncompensated substitutions supports the NPHV structure model, the existence of potential alternative pairing in a proportion of suboptimal folds of this region and the striking structural difference from HCV require that the region be physically mapped to verify the structure predictions.

The HCV 5′ UTR contains two miRNA-122 binding sites that are highly conserved among all genotypes and required for replication in liver cells. While the original predicted secondary structure of the CHV 5′ UTR showed occlusion of both miRNA-122 binding sites (21), the availability of likely more complete sequences from NPHV allowed some modification and refinement of the 5′ UTR secondary structure prediction. In the revised structure of the NHPV 5′ UTR, the second miRNA-122 seed site was

both open and completely conserved (Fig. 3A). Given the tissue-specific expression of miRNA-122 in the liver and the potential for an equivalent cooperative interaction between NPHV and the horse homolog of miRNA-122, these findings suggest hepatotropism comparable to that of HCV. At this stage, however, we have no experimental data to support the specific targeting of the liver by NPHV *in vivo*, and this represents a major priority in ongoing studies.

MFED scanning implemented in the SSE package has been used to document the presence of genome-scale ordered RNA structure (GORS) in RNA virus genomes (35). MFEDs averaged over the NPHV genomes fell into a relatively narrow range from 12.5% to 13.9% (mean, 13.0%), similar to that described for CHV and consistently higher than those for HCV, human and simian GB viruses/pegiviruses, and most other RNA viruses (21, 35). Detection of high MFED values, restricted generally to viruses that establish persistence in their natural hosts (35), is consistent with the observation that over 20% of exposed horses were viremic (see discussion above). Scanning sequences using fragments comparable to the lengths of stem-loops in the NPHV genome produced a plot comprising regularly spaced peaks of high MFED values (corresponding to fragments containing the 5′ and 3′ sides of the stem-loop) alternating with fragments of low folding energy, representing folding energies of naturally nonpairing sequences from the 3′ side of one stem-loop and the 5′ side of the next. This revealed the presence of a series of quite regularly spaced stem-loops running through the whole coding sequence of NPHV, an RNA structure that is likely to substantially modify the structural configuration of the genomic RNA, as demonstrated for other viruses with GORS (13), and potentially modulates its interaction with host cell defenses in a way that promotes virus persistence, as previously hypothesized (15). The substantial diversity at third codon positions and the likely conservation of RNA structure between NPHV variants enabled use of structure prediction programs that exploit phylogenetic information such as pairing partners and the occurrence of covariant sites to validate thermodynamic predictions. The mountain plot produced by ALIFOLD (15) closely reproduced the peak and trough prediction of the MFED scan. The concordance between predictions from two different RNA structure prediction methods recapitulates the results of previous detailed analysis of predicted unstructured and structured viral genomes (13). This showed substantial, algorithm-independent concordance between MFED, ALIFOLD, PFOLD, and RNAz methods in their prediction of both the presence and the intensity of RNA pairing in an extended viral sequence data set. Furthermore, viral genomes predicted to be structured showed a range of biophysical differences from unstructured virus, including hybridization accessibility and scanning electron microscopy appearance (13). From these previous data, we can expect the NPHV genome to be extensively internally base-paired and to share a closed genomic configuration that may modulate its interaction with host cell defenses (13, 35). For future analysis, the comparative sequence data from the 8 complete genomes of NPHV indeed provide a wealth of sequence data to validate RNA pairing predictions, analyze the evolutionary conservation of paired and unpaired sites, and explore the relationship between sequence change and evolution of RNA structures that preserve secondary structure and closed genomic RNA configurations that seem central to its role in persistence.

Similarities and differences between HCV and NPHV will be equally informative with respect to hepacivirus biology. If NPHV resembles HCV in its pathogenesis, it could lead to a tractable *in vivo* model for the human virus. Where the species diverge, it will provide a unique opportunity to compare the molecular and cellular bases for those differences. The ability to compare closely related hepaciviruses *in vitro* will provide insights into the molecular biology of both viruses. Features such as entry factors, interactions of viral and host proteins, and the regulation of replication by genomic elements can be pursued. Moreover, an infectious clone for NPHV will pave the way for experimental animal infections. The data presented here will help in generating an NPHV consensus sequence from multiple isolated sequences. As for HCV, we expect that a consensus clone will be useful in recapitulating replication and potentially infectious virus production in cultured cells. Ultimately, these NPHV infectious clones may provide an ideal backbone for the development of recombinant HCV vaccines. Together, our data will assist in the design of studies to illuminate hepacivirus biology from a new angle. The availability of genetically distinct NPHV genomes and analysis of their evolutionary change and constraints and similarities and differences from the evolutionary processes reconstructed for HCV will therefore likely advance our understanding of the role that these genetic elements and proteins play in the viral life cycle, epidemiology, and pathogenesis.

REFERENCES

1. **Aguilar PV, et al.** 2011. Endemic Venezuelan equine encephalitis in the Americas: hidden under the dengue umbrella. Future Virol. **6**:721–740.
2. **Bruderer U, et al.** 2004. Differentiating infection from vaccination in foot-and-mouth-disease: evaluation of an ELISA based on recombinant 3ABC. Vet. Microbiol. **101**:187–197.
3. **Bukh J.** 2011. Hepatitis C homolog in dogs with respiratory illness. Proc. Natl. Acad. Sci. U. S. A. **108**:12563–12564.
4. **Bukh J, Apgar CL, Govindarajan S, Purcell RH.** 2001. Host range studies of GB virus-B hepatitis agent, the closest relative of hepatitis C virus, in New World monkeys and chimpanzees. J. Med. Virol. **65**:694–697.
5. **Bukh J, Apgar CL, Yanagi M.** 1999. Toward a surrogate model for hepatitis C virus: an infectious molecular clone of the GB virus-B hepatitis agent. Virology **262**:470–478.
6. **Burbelo PD, et al.** 2011. LIPS arrays for simultaneous detection of antibodies against partial and whole proteomes of HCV, HIV and EBV. Mol. Biosyst. **7**:1453–1462.
7. **Burbelo PD, et al.** 2011. Serological studies confirm the novel astrovirus HMOAstV-C as a highly prevalent human infectious agent. PLoS One **6**:e22576.
8. **Burbelo PD, Ching KH, Klimavicz CM, Iadarola MJ.** 2009. Antibody profiling by luciferase immunoprecipitation systems (LIPS). J. Vis Exp. **2009**(32):1549.
9. **Burbelo PD, et al.** 2007. Rapid antibody quantification and generation of whole proteome antibody response profiles using LIPS (luciferase immunoprecipitation systems). Biochem. Biophys. Res. Commun. **352**:889–895.
10. **Burbelo PD, Goldman R, Mattson TL.** 2005. A simplified immunoprecipitation method for quantitatively measuring antibody responses in clinical sera samples by using mammalian-produced Renilla luciferase-antigen fusion proteins. BMC Biotechnol. **5**:22.
11. **Burbelo PD, et al.** 2010. Proteome-wide anti-hepatitis C virus (HCV) and anti-HIV antibody profiling for predicting and monitoring the response to HCV therapy in HIV-coinfected patients. J. Infect. Dis. **202**:894–898.

12. **Centers for Disease Control and Prevention.** 2011. West Nile virus disease and other arboviral diseases—United States, 2010. MMWR Morb. Mortal. Wkly. Rep. **60**:1009–1013.

13. **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.

14. **Dolgin E.** 2011. Research technique: the murine candidate. Nature **474**:S14–S15.

15. **Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL.** 2008. The Vienna RNA websuite. Nucleic Acids Res. **36**:W70–W74.

16. **Honda M, Brown EA, Lemon SM.** 1996. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. RNA **2**:955–968.

17. **Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P.** 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. Science **309**:1577–1581.

18. **Kapoor A, et al.** 2009. Multiple novel astrovirus species in human stool. J. Gen. Virol. **90**:2965–2972.

19. **Kapoor A, et al.** 2011. Characterization of novel canine bocaviruses and their association with respiratory disease. J. Gen. Virol. **93**:341–346.

20. **Kapoor A, et al.** 2010. Identification and characterization of a new bocavirus species in gorillas. PLoS One **5**:e11948.

21. **Kapoor A, et al.** 2011. Characterization of a canine homolog of hepatitis C virus. Proc. Natl. Acad. Sci. U. S. A. **108**:11608–11613.

22. **Kapoor A, Simmonds P, Lipkin WI, Zaidi S, Delwart E.** 2010. Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. J. Virol. **84**:10322–10328.

23. **Kapoor A, et al.** 2008. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. Proc. Natl. Acad. Sci. U. S. A. **105**:20482–20487.

24. **Kapoor A, et al.** 2008. A highly divergent picornavirus in a marine mammal. J. Virol. **82**:311–320.

25. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23**:1891–1901.

26. **Lindenbach BD, et al.** 2005. Complete replication of hepatitis C virus in cell culture. Science **309**:623–626.

27. **Lipkin WI.** 2010. Microbe hunting. Microbiol. Mol. Biol. Rev. **74**:363–377.

28. **Menne S, Cote PJ.** 2007. The woodchuck as an animal model for pathogenesis and therapy of chronic hepatitis B virus infection. World J. Gastroenterol. **13**:104–124.

29. **Muerhoff AS, et al.** 1995. Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. J. Virol. **69**:5621–5630.

30. **Murray CL, Rice CM.** 2011. Turning hepatitis C into a real virus. Annu. Rev. Microbiol. **65**:307–327.

31. **Nam JH, et al.** 2004. In vivo analysis of the 3′ untranslated region of GB virus B after in vitro mutagenesis of an infectious cDNA clone: persistent infection in a transfected tamarin. J. Virol. **78**:9389–9399.

32. **Pietschmann T, et al.** 2002. Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. J. Virol. **76**:4008–4021.

33. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21**:676–679.

34. **Simmonds P.** 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res. Notes **5**:50.

35. **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.

36. **Simons JN, et al.** 1995. Isolation of novel virus-like sequences associated with human hepatitis. Nat. Med. **1**:564–569.

37. **Simons JN, et al.** 1995. Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc. Natl. Acad. Sci. U. S. A. **92**:3401–3405.

38. **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.

39. **Tamura K, et al.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

40. **Wakita T, et al.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat. Med. **11**:791–796.

41. **Wobus CE, Thackray LB, Virgin HW IV.** 2006. Murine norovirus: a model system to study norovirus biology and pathogenesis. J. Virol. **80**:5104–5112.