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

Published Ahead of Print 4 April 2012.

Updated information and services can be found at: http://jvi.asm.org/content/86/11/6171

**REFERENCES**

These include:
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»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
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

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

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 undefined 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).
MATERIALS AND METHODS

Serum samples. Serum samples from different animal species included sera of 80 dogs, 14 rabbits, 81 deer, 44 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 immuno-precipitation 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′-GAGGGATCCAG = GTTACGAGCACGAGATAG-3′ and for CHV core, 5′-GAGGGATCCAG TAAATAAACCTAAAAAC-3′ and 5′-GAGCTGAGTCAGGCTCTTCCG AAGATAC-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 MgCl2, 1% Triton X-100) using a 96-well polystyrene microtiter plate. Antibody titers were measured by adding 40 μl of buffer A, 10 μl of diluted sera (1:1 equivalent), and 1 × 10−6 light units (LU) of each of the Ruc-CHV and HCV helicase antigen fragments containing crude Cos1 cell extract to wells of a polystyrene 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 Wildbad, 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/ak37OR1 and ak70F2/ak37OR2 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, GAGGGAGCTGTTAGGTGAA, and Qanti-5UF1, GAGACATCTCT ATCAGACCGT-3′) and probe (6-carboxyfluorescein [FAM]–CCAGAA GGAAGGGGCGG–(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 blast 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). Therefore, sequence validity was tested in 4′ genome coverage by classical denoisy 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 SSE 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 (MFEΔs), 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 hepacviruses (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 bind 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-
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...
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 hepacviruses, 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-
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 structure 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 semicovariances pairing 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 hepacviruses 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.

ACKNOWLEDGMENTS

We thank Natasha Qaisar for excellent technical assistance.

This work was supported by National Institutes of Health grants (AI090196, AI081132, AI079231, AI57158, AI070411, and EY017404), by an award from the Defense Threat Reduction Agency (Department of Defense) and USDA 58-1275-7-370, and by the Division of Intramural Research, National Institute of Dental and Craniofacial Research.

REFERENCES