Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses

Amit Kapoor,[a] Peter Simmonds,[b] Troels K. H. Scheel,[c,d] Brian Hjelle,[e] John M. Cullen,[f] Peter D. Burbelo,[g] Lokendra V. Chauhan,[a] Raja Duraisamy,[a] Maria Sanchez Leon,[a] Komal Jain,[a] Kurt Jason Vandegrift,[h] Charles H. Calisher,[a] Charles M. Rice,[c] 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]; Center for the Study of Hepatitis C, Laboratory of Virology and Infectious Disease, the Rockefeller University, New York, New York, USA[c]; Copenhagen Hepatitis C Program (CO-HEP), Department of Infectious Diseases and Clinical Research Centre, Copenhagen University Hospital, Hvidovre, Denmark, and Department of International Health, Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark[d]; Department of Pathology, University of New Mexico, Albuquerque, New Mexico, USA[e]; North Carolina College of Veterinary Medicine, Raleigh, North Carolina, USA[f]; Clinical Dental Research Core, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA[g]; Center for Infectious Disease Dynamics, the Pennsylvania State University, University Park, Pennsylvania, USA[h]

ABSTRACT Hepatitis C virus (HCV) and human pegivirus (HPgV or GB virus C) are globally distributed and infect 2 to 5% of the human population. The lack of tractable-animal models for these viruses, in particular for HCV, has hampered the study of infection, transmission, virulence, immunity, and pathogenesis. To address this challenge, we searched for homologous viruses in small mammals, including wild rodents. Here we report the discovery of several new hepaciviruses (HCV-like viruses) and pegiviruses (GB virus-like viruses) that infect wild rodents. Complete genome sequences were acquired for a rodent hepacivirus (RHV) found in *Peromyscus maniculatus* and a rodent pegivirus (RPgV) found in *Neotoma albigula*. Unique genomic features and phylogenetic analyses confirmed that these RHV and RPgV variants represent several novel virus species in the *Hepacivirus* and *Pegivirus* genera within the family *Flaviviridae*. The genetic diversity of the rodent hepaciviruses exceeded that observed for hepaciviruses infecting either humans or non-primates, leading to new insights into the origin, evolution, and host range of hepaciviruses. The presence of genes, encoded proteins, and translation elements homologous to those found in human hepaciviruses and pegiviruses suggests the potential for the development of new animal systems with which to model HCV pathogenesis, vaccine design, and treatment.

IMPORTANCE The genetic and biological characterization of animal homologs of human viruses provides insights into the origins of human infections and enhances our ability to study their pathogenesis and explore preventive and therapeutic interventions. Horses are the only reported host of nonprimate homologs of hepatitis C virus (HCV). Here, we report the discovery of HCV-like viruses in wild rodents. The majority of HCV-like viruses were found in deer mice (*Peromyscus maniculatus*), a small rodent used in laboratories to study viruses, including hantaviruses. We also identified pegiviruses in rodents that are distinct from the pegiviruses found in primates, bats, and horses. These novel viruses may enable the development of small-animal models for HCV, the most common infectious cause of liver failure and hepatocellular carcinoma after hepatitis B virus, and help to explore the health relevance of the highly prevalent human pegiviruses.

Address correspondence to Amit Kapoor, ak3117@columbia.edu.

Hepatitis C virus (HCV) and human pegiviruses (HPgVs) infect an estimated 2% and 5% of the world's population, respectively. HCV, HPgV (formerly referred to as GB virus C [GBV-C] or hepatitis G virus), and other genetically related viruses belong to two genera in the *Flaviviridae* family, *Hepacivirus* and *Pegivirus*, respectively (1). HCV is hepatotropic and can trigger liver damage characterized by fibrosis, cirrhosis, and hepatocellular carcinoma (2). HPgV is lymphotropic (3), but its pathogenicity for humans, if any, is unknown. HPgV is more prevalent in people with blood-borne or sexually transmitted infections than in the general population. Up to 40% of HIV-infected individuals have HPgV viremia (1, 4, 5). Viruses genetically most similar to human HCV include GBV-B and the recently discovered nonprimate hepaciviruses (NPHVs) (6). These viruses show extensive gene homology and conserved genomic elements with HCV, as well as genus-specific features (1) that include a core protein, a type IV internal ribosomal entry site (IRES), and hepatotropism. Horses are the natural host for NPHVs (6, 7). The origin and natural host of GBV-B are unknown. Pegiviruses infect a wide range of mammals, including chimpanzees, New World primates (GBV-A or simian PgV [SPgV]), horses (equine PgV [EPgV] [A. Kapoor, et al., submitted for publication]), and bats (GBV-D or bat PgV [BPgV]) (1, 8, 9).

Despite differences in their pathogenic potentials, HCV and

HPgVs share many biological properties, including similar structural and genomic organizations, high inter- and intrahost genetic diversity, and, frequently, persistent infection of their natural hosts. Studies of HCV and HPgV in nonhuman primates have led to important advances in our understanding of their biology and pathogenesis (1, 10–13). However, basic questions regarding virulence determinants, organ tropism, systemic host responses, viral dynamics, and mechanisms of disease induction remain unanswered (10, 14). The currently characterized hepaciviruses and pegiviruses have narrow host ranges. HCV infects only humans and chimpanzees. Pegiviruses infecting humans, Old World primates, and New World primates are species-specific. Small-animal models often supply the extensive genetic and immunologic tools required to evaluate viral pathogenesis and immunity. The lack of such a model has impeded research on this important group of viruses (10). To address this challenge, we initiated a methodical search for such viruses in several species of wild rodents.

RESULTS

Identification and genetic characterization of new rodent viruses. Plasma samples of >400 wild-caught rodents, predominantly deer mice, were screened using two degenerate sets of primers targeting conserved virus helicase motifs of hepaciviruses and PgVs. Sequencing of PCR products confirmed the presence of hepaciviruses and PgVs in 18 samples of rodents belonging to four species: hispid pocket mice (*Chaetodipus hispidus*), deer mice (*Peromyscus maniculatus*), desert wood rats (*Neotoma lepida*), and white-throated wood rats (*Neotoma albigula*). The majority of samples were from these rodent species, which may have induced sampling bias. Although the sequencing of PCR products provided only 300-nucleotide (nt)-long viral fragments, the highly conserved nature of the sequenced helicase region allowed accurate phylogenetic classification of all new viruses (Fig. 1). Appropriate classification of well-characterized viruses using the corresponding sequence region validated our analysis (1, 6, 7). Moreover, the phylogenetic tree constructed with the complete protein sequences of new rodent viruses shows clustering concordant with that obtained using partial helicase sequences (Fig. 1 and 2). Following the International Committee for Taxonomy of Viruses guidelines of using host names to describe hepaciviruses and pegiviruses (1, 6, 7), we tentatively named these new viruses rodent hepacivirus (RHV) and rodent pegivirus (RPgV). We refrained from naming viruses on the basis of host species, given that their natural host and species tropism requires further investigation (7).

Despite their high genetic diversity, all new rodent virus sequences fell into the hepacivirus or pegivirus clade, supporting their provisional assignment to these genera of the family *Flaviviridae* (1). Based on the phylogenetic analyses of partial helicase sequences and intraspecies genetic distances of known viruses (HCV and NPHV), the RHV sequences identified in our study can be tentatively classified as five new virus species. Of these, three new RHV species were found in a single host species, deer mice. Two new RHV species were found in desert wood rats and hispid pocket mice. Despite our limited sampling of host animals, one of the new RHV species showed intraspecies genetic diversity (RHV-pm4144, RHV-pm4062, RHV-pm3243, RHV-pm3252, RHV-pm5198, RHV-pm4109, and RHV-pm5263) that was equivalent to that reported among different HCV subtypes (Fig. 1). Our anal-ysis also showed that the observed genetic diversity of RHV species exceeded that reported for all known hepaciviruses. The natural host of GBV-B remains obscure, and noticeably, GBV-B fell within the genetic diversity of RHV species.

We found two new species of pegiviruses in rodents, one in white-throated wood rats (RPgV-cc61) and the other in deer mice (RPgV-pm5226, RPgV-pm6197, RPgV-pm6073, RPgV-pm6041, and RPgV-pm6087). All of the 5 RPgV variants found in deer mice clustered together, indicating a single genetically diverse RPgV species (Fig. 1). The latter variants from deer mice clustered together, with divergence between different variants comparable to the diversity observed between genotypes of human pegiviruses (Fig. 1).

Genetic analysis of rodent RHVs. We acquired the complete genome of RHV-339 (8,879 nt) and the nearly complete genome of RHV-089 (8,252 nt), which were found in plasma samples of two different deer mice. As with HCV, the RHV genome is predicted to encode a long polyprotein flanked by 5′ and 3′ untranslated regions (UTRs). In the open reading frame (ORF), 8% synonymous (dS) and <1% nonsynonymous (dN) mutations existed between the genomes of the two variants, indicative of strong purifying selection. The first in-frame polyprotein initiation codon was found at nucleotide position 403 of the RHV-339 genome and at the corresponding position in RHV-089. Although the proposed start codon did not include a classical Kozak consensus sequence (cca**C**ttATG**G**), an even-less-favorable Kozak context is found in GBV-B (tag**C**aaATG**C**) (where lowercase bases indicate variable positions), consistent with ribosomal positioning and translation initiation by a type IV internal ribosome entry site (IRES). The 5′ UTR region of RHV-089 and the corresponding region of RHV-339 showed homology to other hepacivirus sequences in the 200-nt region adjacent to the start of the polyprotein initiation codon (Fig. 3). The remaining 5′ UTR returned no matches from BLAST searching against the full GenBank sequence database, although a miR-122 seed site (UACACUCC) was found in the RHV-339 5′ UTR at nt 7 to 14, which may indicate the hepatotropic potential of these rodent hepaciviruses. Although convincing alignments on which to base structure predictions could not be achieved, the positioning and occurrence of covariant changes in sequences homologous to those of other hepaciviruses allowed us to establish a partial structural model of the IRES in the region corresponding to domain III of the type IV HCV IRES, which included each of the IIIa to IIIf stem-loops and the pseudoknot (IIIf) region sequences (Fig. 3). The region 5′ to the start of stem-loop III was shorter than corresponding sequences of HCV, GBV-B, and NPHV. Downstream of the polyprotein stop codon, we identified a 3′ UTR of 230 nt consisting of a structured region, potentially equivalent to the HCV 3′ variable region, a poly(C) tract of around 10 nt, and a 3′ X region of 158 nt, which is predicted to fold into 4 stem-loop structures (Fig. 3B). The structural elements of the 3′ UTR resembled those of HCV, with a short poly(C) tract replacing the long HCV poly(U/C) tract, and a longer 3′ X region. The RHV 3′ UTR shared no sequence homology with HCV or other database sequences.

The RHV open reading frame is predicted to encode a polyprotein of 2,748 amino acids (aa), shorter than those of HCV-1a (3,011 aa), NPHV (2,942 aa), and GBV-B (2,854 aa). Comparative genetic analysis predicts that the RHV-339 polyprotein contains three structural (core, E1, E2) and seven nonstructural (NS) (p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) proteins (Fig. 4).
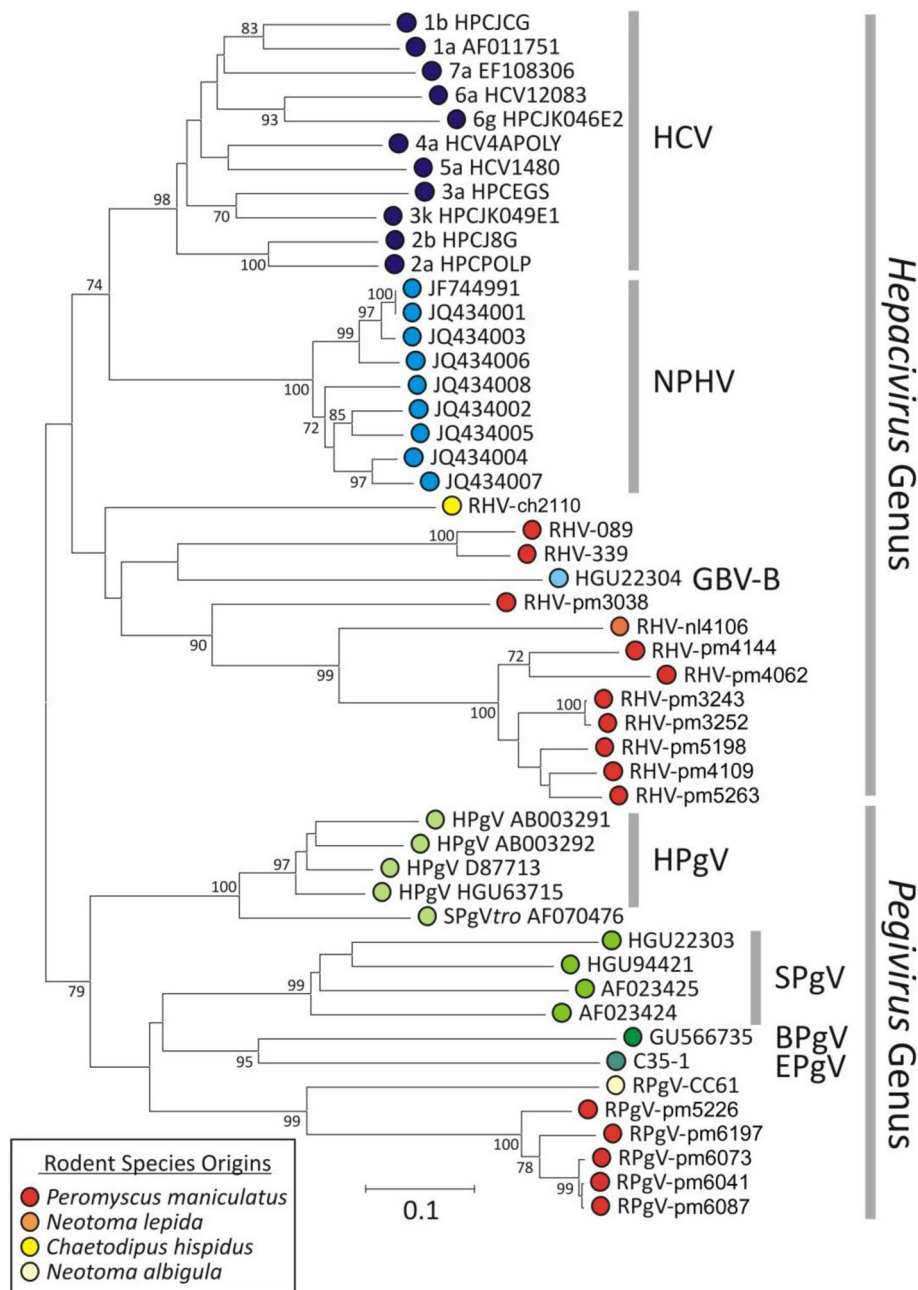
FIG 1 Phylogenetic analysis of new rodent viruses using the partial helicase sequences generated by generic PCR. Bootstrap resampling was used to determine the robustness of branches; values of ≥70% (from 1,000 replicates) are shown. Host origins of newly reported viral sequences are indicated by red and yellow circles (see the key) and for previously described hepaciviruses (blue circles) and pegiviruses (green circles).

Cleavage sites in the RHV polyprotein and in other hepaciviruses were predicted by alignment and homology to sites previously identified in HCV (15). Because of sequence variability around cleavage sites in structural proteins, the motifs for RHV and NPHV were independently predicted using the SignalP 4.1 server (16) and inferred from the cleavage motifs that aligned with those in HCV. Structural and nonstructural RHV proteins were similar in predicted size to those of HCV and other hepaciviruses (Fig. 4C), including a predicted 63-aa p7 transmembrane protein (Fig. 4C). However, the core protein was shorter than for HCV (168 versus 191 aa), as was E2 (279 versus 383 aa). The E1 and E2

proteins contained 2 and 4 predicted N-linked glycosylation sites, respectively, fewer than those found in the homologous glycoproteins of HCV (6 and 11, respectively), NPHV (4 and 10, respectively), and GBV-B (3 and 6, respectively). The region encoding the putative core protein lacked an alternative open reading frame equivalent to the proposed HCV alternative reading frame protein (17).

Despite the similarity in organization and structural features in the hepacivirus genomes, sequences were extraordinarily divergent from each other at the nucleotide and amino acid levels. Coding sequences were aligned, and pairwise distances between
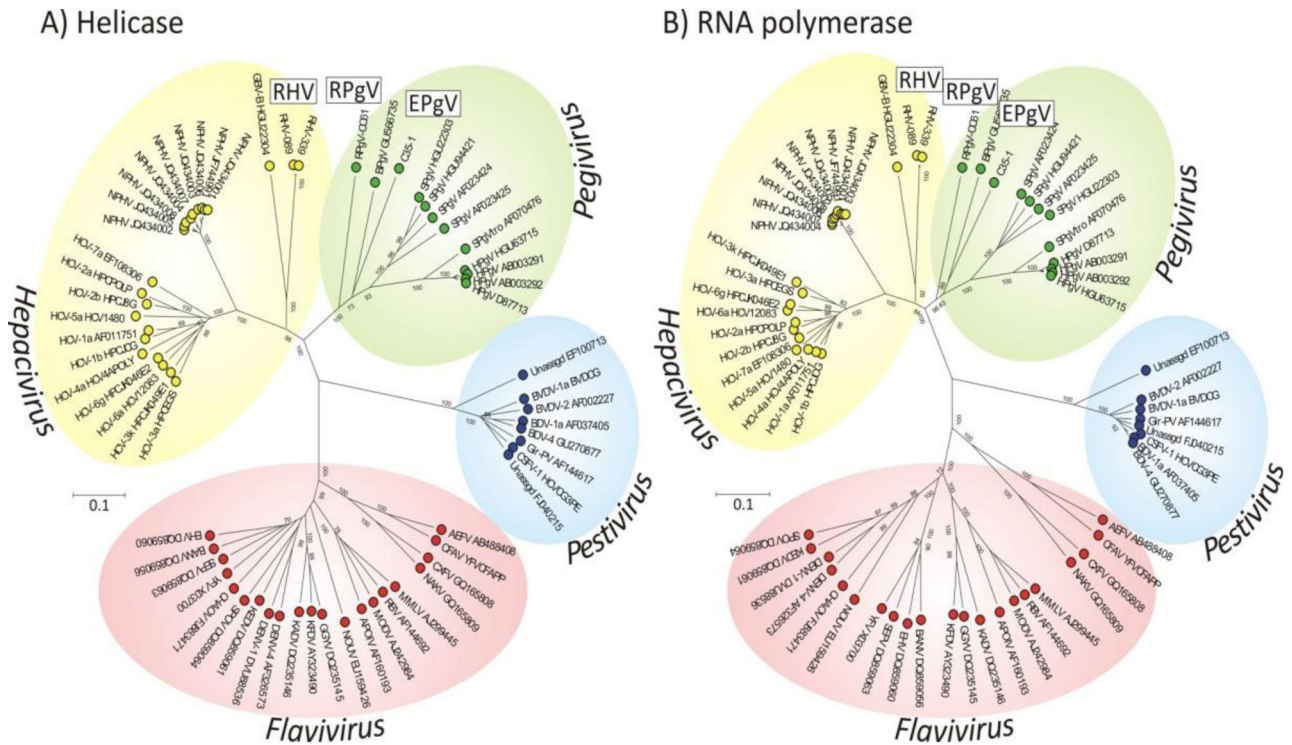
FIG 2 Phylogenetic analysis of conserved regions in the helicase (motifs I to VI) (A) and RdRp (B) genes of rodent hepaciviruses and pegiviruses aligned with representative members of the *Hepacivirus*, *Pegivirus*, *Pestivirus*, and *Flavivirus* genera. Trees were constructed by neighbor joining of pairwise amino acid distances with the program MEGA5 (according to the distance scale provided). Bootstrap resampling was used to determine the robustness of branches; values of ≥70% (from 1,000 replicates) are shown. Regions compared corresponded to positions 3667 to 4470 (helicase domain of NS3) and 7711 to 8550 (RdRp in NS5B; numbered according to the AF011751 HCV genotype 1a reference sequence).

structural and nonstructural genome regions were computed (Table 1). RHV-339 showed mean amino acid divergences ranging from 67% to 77% in the structural region and from 65% to 70% in the NS regions, similar to the divergences between HCV and GBV-B and substantially greater than those between HCV and NPHV. Genetic divergence between the entire genome of RHV-339 and the known hepaciviruses (HCV, GBV-B, and NPHV) was analyzed using a scanning window of 300 nt in 15-nt increments (Fig. 4). As observed for NPHV (6, 7), the most-conserved regions within viruses of the *Hepacivirus* genus were the NS3 and NS5B genes. High sequence divergence was observed in the E1 and E2 glycoproteins and NS4B; there was no apparent homology to other sequences in GenBank in extended regions of NS4A and NS5A. Sequence comparisons were extended to include homologous sequences from the other members of the *Flaviviridae* located in the NS3 and NS5B regions, and these sequences could be aligned (Fig. 2). Phylogenetic trees from the two genomic regions were topologically equivalent but different in relative branch lengths. The analysis confirmed the separate grouping of RHV-339 and RHV-89 from all other hepaciviruses (Fig. 2A and B), although hepaciviruses collectively formed a separate bootstrap that differentiated them from pegiviruses. RNA folding analysis of the RHV genome revealed that the minimum folding energy difference (MFED) value of RHV was 4.6%, which is well below the 8 to 9% determined for HCV and 9% for GBV-B. This suggested a less structured RHV genome.

Genetic analysis of rodent pegivirus. The complete genome (11,279 nt) of the rodent pegivirus (RPgV-cc61) found in a white-

throated wood rat included a 5′ UTR (349 nt), a polyprotein coding region (10,452 nt; 3,484 aa) and a 3′ UTR (475 nt). The RPgV-cc61 5′ UTR showed no significant similarity with any known pegivirus sequence; therefore, prediction of its secondary RNA structure was problematic in the absence of a structural alignment or covariance data. The first initiating codon in frame with the predicted encoded polyprotein was located at position 350. The 3′ UTR sequence contained two internal poly(C) tracts of around 9 nt, and stable stem-loop structures could be predicted immediately downstream of the stop codon and at the very 3′ end of the genome (Fig. 3C). Surprisingly, two repeat sequence elements (RSEs), potentially folding into similar stem-loop structures, were exact copies of a 24-nt region from the 5′ UTRs of human enterovirus, coxsackievirus, echovirus, and swine vesicular disease virus. No other homology to known viral sequences was found in the RPgV 3′ UTR.

Cleavage sites in the RPgV polyprotein were predicted by alignment and homology to sites previously predicted between nonstructural proteins in simian and bat pegiviruses (GBV-A [18]; GBV-D [8]; sites NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B) and by comparison to predicted signalase sites between structural proteins of RPgV and other pegiviruses. This analysis identified homologous cleavage sites that aligned with those in BPgV, including the boundaries of the proposed novel X protein (Fig. 5). Our analysis indicated that the RPgV-cc61 genome might harbor an additional signalase site after nt 1015 and before a coding sequence that was clearly homologous to E1 of other pegiviruses. The presence of a hydrophobic sequence of
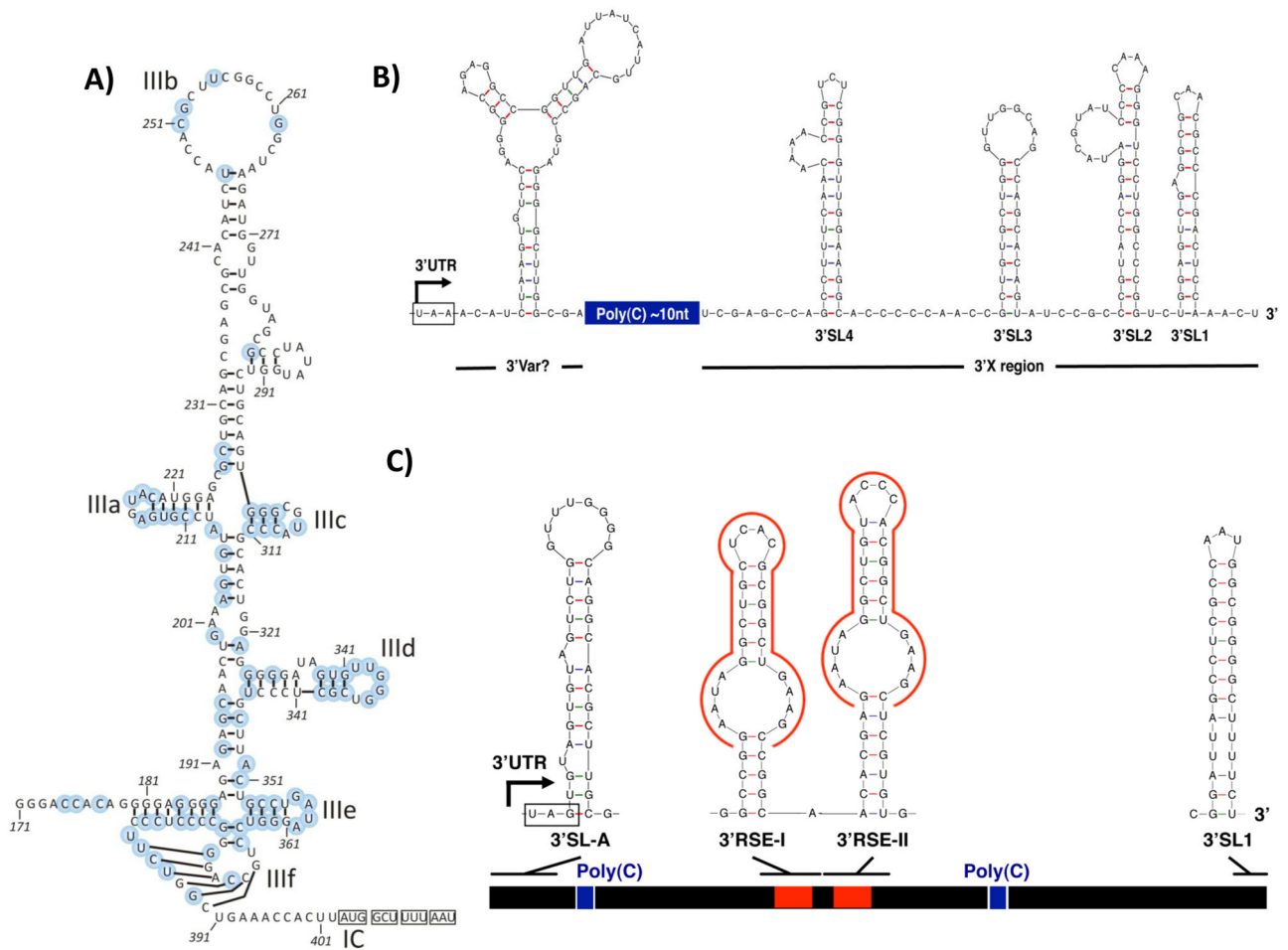
FIG 3 Predicted structures in the RHV-339 and RPgV-cc61 UTRs. (A) RHV-339 5′ UTR structure predictions were made by homology searching and structural alignment within domain III. Bases conserved in all hepaciviruses (HCV, NPHV, and GBV-B) are highlighted with blue circles. (B) Structure prediction of the RHV 3′ UTR using Mfold. The ORF stop codon is boxed, and a putative variable region equivalent to that of HCV is indicated. A poly(C) tract of around 10 nt separates this from the downstream 3′ X region, which was predicted to fold into four stem-loops (3′SL1 to -4). (C) Predicted structure of the RPgV 3′ UTR. The black bar cartoon of the 3′ UTR (475 nt) was drawn to scale, with poly(C) tracts indicated in blue. Strong stem-loop structures immediately downstream of the ORF stop codon (boxed) and at the very end of the genome are indicated. RSEs, potentially folding into similar stem-loop structures, are outlined in red. The structure prediction of the RPgV 3′ UTR was done using Mfold.

31 aa at the beginning of the polyprotein possibly functions as a signal peptide that leads to translocation of a predicted 223-aa protein (labeled "Y" in Fig. 5) into the endoplasmic reticulum (ER) (which is analogous to E1 processing for other pegiviruses). The presence of two N-linked glycosylation sites in this coding sequence suggests a possible fourth glycoprotein (in addition to E1, E2, and X) in the RPgV envelope. Alignment of the RPgV sequence with simian and bat pegiviruses allowed prediction of cleavage sites of the nonstructural proteins NS2, NS3, NS4A, NS4B, NS5A, and NS5B. Each was comparable in size to homologs from other pegiviruses.

The genetic relatedness of RPgV-cc61 to other pegiviruses was assessed by alignment of coding sequences and calculations of pairwise distances between structural and nonstructural genome regions (Table 2). RPgV-cc61 was substantially divergent from HPgV, SPgV, BPgV, and EPgV sequences, with amino acid divergence ranging from 78% to 81% in the structural proteins and 54% to 56% in the NS region (Table 2). The degree of genetic divergence across the genome of RPgV-cc61 from those of other

pegiviruses was analyzed as described for RHV (Fig. 5). Consistently with previous analyses (6, 7), the most-conserved regions within viruses of the genus *Pegivirus* were the NS3 and NS5B genes, with high sequence divergence in the E1 and E2 glycoproteins and NS4B and no apparent homology to other sequences in GenBank in extended regions of X, NS4A, and NS5A (Fig. 5). Phylogenetic analysis of the NS3 and NS5B regions confirmed the separate grouping of RPgV-cc61 from all other pegiviruses (Fig. 2A and B). RNA folding analysis of the RPgV genome revealed that the MFED value of the RPgV genome was 9.7%, an observation consistent with the presence of a genome-scale, ordered RNA structure (7). This MFED value was similar to those of human (mean, 12.8% [11.7% to 13.3%]), simian (mean, 13.3% [12.7% to 13.8%]), bat (9.7% and 10.7%), and equine (10.7%) pegiviruses (6).

DISCUSSION

The identification and characterization of animal virus homologs can provide insights into the pathogenesis of human viruses and,
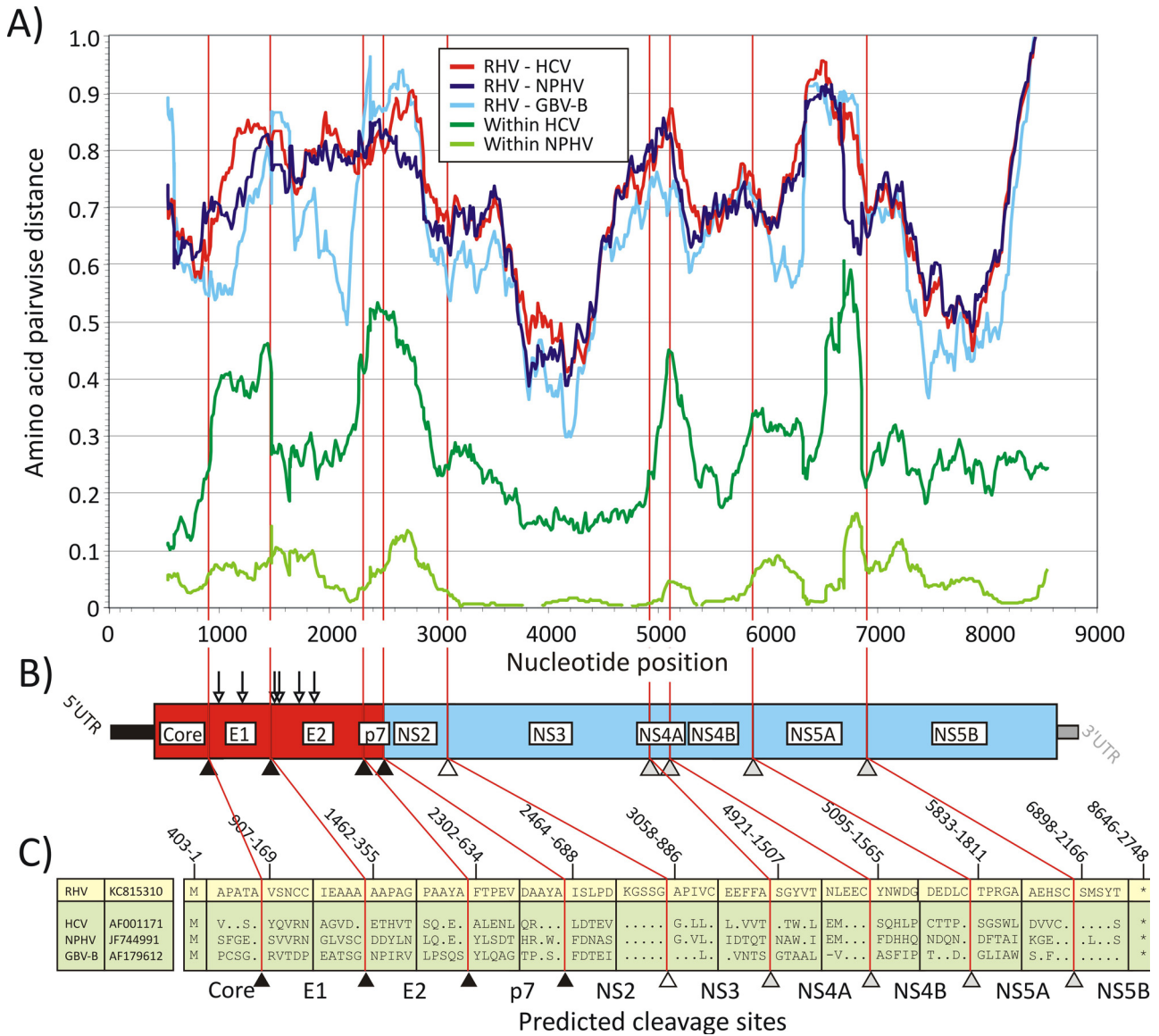
FIG 4 (A) Amino acid sequence divergence between RHV and human virus (HCV), equine virus (NPHV), and GBV-B using 300-nt fragments by 15-nt increments across the sequence alignment (the midpoint is plotted on the *y* axis). Within-species distances for human and equine hepaciviruses are included for comparison. (B, C) Genome diagram of RHV showing predicted N-linked glycosylation sites (downward-pointing arrows) and proposed cleavage sites of cellular peptidase (black triangles), NS2-3 protease (white triangle), and NS3-4A protease (gray triangles) (sequence positions were numbered using the RHV sequence).

in some instances, *in vivo* models for investigating methods for the prevention and treatment of human disease (19). Examples where well-characterized animal viruses have provided such insights include simian immunodeficiency virus, animal poxviruses, herpesviruses, murine norovirus, and woodchuck hepatitis virus (20). HCV, in contrast, has no satisfactory homolog (6, 21), and only chimpanzees can be experimentally infected with HCV (22–25). Even before the recent U.S. Institute of Medicine recommendations to restrict the use of chimpanzees for biomedical research, limited access to these animals was a challenge for HCV research. NPHV and GBV-B are the most genetically similar to HCV (6) and could therefore be used as surrogate models of HCV infection. The natural host of NPHV is the horse (6, 7, 17), in which high frequencies of viremia (from 3 to 8%) have been reported in separate studies (17). GBV-B was initially detected in a laboratory

tamarin (New World monkeys of the family *Callitrichidae*). However, subsequent attempts to identify its natural host that concentrated primarily on the screening of New World primates have been unsuccessful. Nonetheless, GBV-B-infected tamarins and marmosets have been used as surrogate models for HCV pathogenesis. The identification of rodent hepaciviruses may finally provide a promising small-animal model for the study of hepaciviruses, with possible relevance to HCV.

Here we identified several lineages of RHV in deer mice that are as highly divergent from each other as are HCV and GBV-B. In light of the recent finding of hepaciviruses infecting horses and dogs (6, 7, 17), which are considerably more similar to HCV than GBV-B or RHV, it is unlikely that hepaciviruses coevolved with their hosts. The basal radiation of three different lineages of hepaciviruses infecting deer mice (Fig. 1) means either that the variants

TABLE 1 Amino acid and nucleotide (shaded) sequence divergence of structural genes (core, E1, and E2) and nonstructural genes (NS2 to NS5B) between RHV and other hepaciviruses

Genes	Virus	% sequence divergence from indicated virus			
		RHV	HCV	NPHV	GBV-B
Structural	RHV		60.8	60.9	57.7
(core, E1, E2)	HCV	75.7		52.7	61.8
	NPHV	73.9	59.6		61.7
	GBV-B	67.2	76.2	75.7	
Nonstructural	RHV		57.9	57.1	55.9
(NS2 to NS5B)	HCV	70.4		47.7	58.6
	NPHV	69.0	49.6		56.7
	GBV-B	65.2	68.1	67.0	

diversified within this host species and subsequently infected another rodent species (*Neotoma lepida*) and ultimately a tamarin (GBV-B) or that deer mice became infected with highly genetically distinct hepaciviruses from other host species. Either explanation requires the occurrence of multiple cross-species events that cannot be dated *a priori*. Thus, without a chronological anchor, we did not attempt to estimate the evolutionary rates of hepacivirus lineages and their divergence times. The hepaciviruses identified by us in this and previous studies (7) may have cross-species transmission potential. The high genetic diversity observed among RHV species raises the possibility that hepaciviruses (HCV, NPHV, and GBV-B) may have actually originated in rodents. Serology-enabled approaches, such as the one we recently used to study the host tropism of NPHV (7), will be very useful in determining the host range and cross-species transmission potential of these novel rodent viruses and in identifying related viruses that infect other animal species.

Viruses genetically related to HPgV include its primate homologs (SPgV), an uncharacterized virus from bats (BPgV) (8), and a recently identified distinct variant infecting horses (EPgV [Kapoor et al., submitted]). Studies thus far indicate a narrow host range for these viruses, with HPgV being found only in humans and chimpanzees, SPgV being found in New World monkeys, and BPgV and EPgV being found only in bats and horses, respectively (1). These findings are consistent with the phylogenetic relationships between pegiviruses infecting rodents and other mammalian species (Fig. 1). Indeed, the two lineages of RPgVs infecting deer mice (*Peromyscus maniculatus*) and white-throated wood rats (*Neotoma albigula*) are more similar to each other than to pegiviruses found in other mammalian species, an observation that is consistent with virus-host cospeciation. However, further investigation of pegiviruses infecting other rodents and mammalian species will be required to solidify or refute the hypothesis that pegiviruses are species specific and have codiverged with the evolution of mammals.

The deduced genome organizations of rodent hepaciviruses and RPgV were similar to those of other members of these genera (1). The 5′ UTRs of RHV and RPgV are long, consistent with the presence of IRES elements found in other hepaci- and pegiviruses. In the case of RHV, we were able to model an RNA structure based on the structurally conserved domains III found in other hepaciviruses, providing support for this structure's function as a type IV IRES. Interestingly, the RHV 3′ UTR elements, but not the pri-

mary sequence, resembled that of HCV, with a putative variable region immediately downstream of the ORF, followed by a polypyrimidine tract and a 3′ X region. However, a short poly(C) tract replaced the longer poly(U·C) tracts found in HCV isolates. The RPgV 3′ UTR did not have homology to other pegiviruses but, surprisingly, contained repeat sequence elements (RSEs) identical to 5′ UTR sequences from human enterovirus, coxsackievirus, echovirus, and swine vesicular disease virus. It is as yet unclear how RPgV acquired these sequence elements and what function they might have.

Analysis of the RPgV polyprotein sequences revealed both similarities and differences from previously identified pegivirus isolates. Unlike hepaciviruses, pegiviruses typically do not encode a core (nucleocapsid) protein (26, 27). Nonetheless, biophysical characterization of HPgV particles suggests the presence of a nucleocapsid, although its origin and composition remain a mystery (27). The RPgV sequence also lacks a convincing capsid protein sequence in either the polyprotein-coding or alternative open reading frames. Rather, the pegivirus polyprotein typically initiates with a signal peptide immediately downstream of the initiation codon that translocates E1 into the ER (position 17 or 21 in human pegiviruses) (28). For RPgV, this is also the case, but RPgV-cc61 also possessed an additional 223-residue Y protein preceding E1, which may be targeted to the ER and glycosylated. Following the E2 homolog, the RPgV sequence encoded a predicted, 249-residue-long acidic X protein (Fig. 5), potentially homologous to, although highly divergent from, those predicted in EPgV and BPgV (8). RPgV also possesses an additional predicted signalase site between E2 and NS2 (position 736) that could give rise to yet another glycosylated membrane protein.

Much of our current knowledge of the replication, host interactions, immune responses, and pathogenesis of HCV and pegiviruses comes from experimental infection of primates or cell culture systems. *In vitro* models have proven valuable for investigating virus replication (13), yet these systems fail to mimic the endogenous milieu of the target organ (liver) and may not accurately recapitulate life cycle events, such as polarized cell entry. Finally, cell culture systems cannot reproduce the interaction between virus and immune system, nor do they allow for studies of pathogenesis (12). The identification and genetic characterization of RHV and RPgV reported here provide a unique opportunity to develop tractable-animal models to study the infection, transmission, immunity, and pathogenesis of hepaciviruses and pegiviruses. Although the current study design precluded direct examination of tissues of infected rodents, it is interesting that the 5′ UTR of RHV contains an miR-122 binding site. These have been previously described in the HCV 5′ UTR (two miR-122 seed sites) as highly conserved among all genotypes and functionally required for replication in hepatocytes (29). Similarly, we recently reported the presence of one miR-122 site in NPHV (7), while the GBV-B 5′ UTR contains sites at positions 8 and 23. Tissue-specific expression of miR-122 in the liver of vertebrates (including rodents) is consistent with potential hepatotropism of all hepaciviruses identified to date, including RHV. It will be interesting to define the sites of RHV replication in rodents and NPHV in horses in future investigations. If RHV does indeed resemble HCV in its tissue tropism and pathogenesis, rodents could prove to be a very useful small-animal model. A rodent model for pegivirus infections is also important for studies focused on the viral and host factors underlying virus persistence. Estimates suggest that >20%
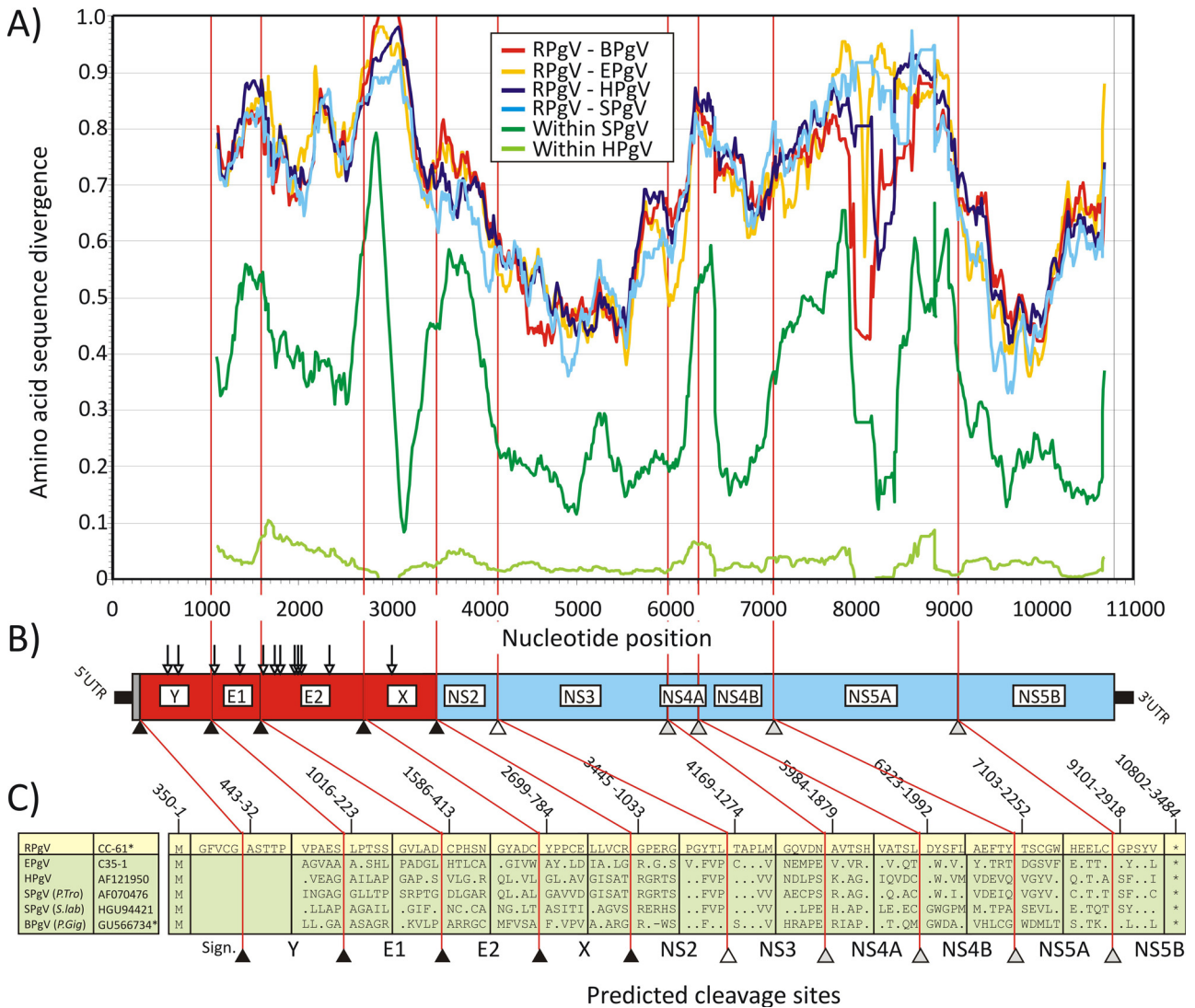
FIG 5 (A) Amino acid sequence divergence between RPgV and HPgV, New World primate PgV (SPgV), equine PgV (EPgV), and bat PgV (BPgV) using 300-base fragments by 15-base increments across the sequence alignment (the midpoint is plotted on the *y* axis). Within-species distances for human and simian pegiviruses (from New World primates [GBV-A]) are included for comparison. The divergence scan commenced at the predicted signalase cleavage site at the start of the E1-encoding genes (position 1016). (B, C) Genome diagram of RPgV showing predicted N-linked glycosylation sites (downward-pointing arrows) and proposed cleavage sites of cellular peptidase (black triangles), NS2-3 protease (white triangle), and NS3-4A protease (gray triangles) (sequence positions were numbered using the RPgV sequence).

of the world population has been exposed to GBV-C, with chronic infection established in 1 to 5% of healthy adults (4, 5, 30). Studying RPgV infections in a natural host amenable to genetic manipulation should provide a powerful approach for unraveling mechanisms favoring resolved infection versus persistence.

Rodent models also provide an opportunity to investigate routes of transmission for RPgV and RHV and how this might relate to HCV transmission, which is due largely to blood-borne routes of exposure. Such studies performed in rodents, including deer mice, have been extremely valuable for understanding hantavirus transmission (31, 32). Comparative genetic analysis and functional characterization of viral entry may help to unravel the determinants of host specificity and tissue tropism and provide insight into possible routes of cross-species transmission (29, 33). In addition, defining the natural history of RHV infection, the rate

of chronicity, the immune determinants of clearance and protection, and possible disease association holds promise for establishing a highly relevant preclinical model for the development of HCV vaccine strategies and interventions to prevent or reverse virus-associated liver disease.

MATERIALS AND METHODS

Rodent samples. Plasma samples from rodents of eight species were collected for a program in hantavirus ecology from sites in the southwestern United States during the period of 2007 to 2009. Samples were stored at −80°C until nucleic acid (NA) extraction. Residual plasma samples were used in this study and included 43 hispid pocket mice (*Chaetodipus hispidus*), 9 black-tailed prairie dogs (*Cynomys ludovicianus*), 9 prairie voles (*Microtus ochrogaster*), 9 wood rats (*Neotoma cinerea*), 4 desert wood rats (*Neotoma lepida*), 342 deer mice (*Peromyscus maniculatus*), 58 white-

TABLE 2 Amino acid and nucleotide (shaded) sequence divergence of structural genes (E1, E2, and X) and nonstructural genes (NS2 to NS5B) between RPgV and other pegiviruses

Genes	Virus	% sequence divergence from indicated virus				
		RPgV	EPgV	HPgV	SPgV	BPgV
Structural	RPgV		64.1	60.0	62.8	63.4
(E1, E2, X)	EPgV	81.3		60.7	62.7	64.1
	HPgV	80.1	76.1		53.6	55.2
	SPgV	80.7	77.8	61.4		57.9
	BPgV	78.5	77.1	72.9	71.2	
Nonstructural	RPgV		64.4	63.5	64.7	62.2
(NS2 to NS5B)	EPgV	55.6		58.1	59.1	56.7
	HPgV	54.3	52.2		48.7	56.2
	SPgV	54.9	52.3	46.4		57.5
	BPgV	53.8	50.1	51.1	51.5	

throated wood rats (*Neotoma albigula*), 10 western harvest mice (*Reithrodontomys megalotis*), and 9 yellow-pine chipmunks (*Tamias amoenus*).

High-throughput sequencing (HTS) and generic PCR assays for hepaciviruses and pegiviruses. Plasma samples were treated with nucleases to digest free NAs for enrichment of viral NA (34–36) and then extracted in NucliSens buffer using the automated easyMAG system (bioMérieux, United States). NAs were reverse transcribed using Superscript II reverse transcriptase and converted to double-stranded DNA using the Klenow fragment (NEB; catalog no. M0212S). Double-stranded DNA (dsDNA) was fragmented using an Ion Shear Plus reagent kit (catalog no. 4471248). Fragmented dsDNA products were ligated to Ion Xpress adapters and unique Ion Xpress bar codes (catalog no. 4471250). Bar-coded libraries were amplified using the Ion Plus fragment library kit (catalog no. 4471252) and the Ion OneTouch system using the Ion OneTouch 200 template kit (v2, catalog no. 4478316). Sequencing was done with the Ion Personal Genome Machine (PGM) system by using the Ion PGM 200 sequencing kit (catalog no. 4474004). Two highly degenerate nested-PCR assays were designed to amplify genetically diverse viruses related to HCV and HPgV. All PCR mixtures used AmpliTaq gold 360 master mix (Applied Biosystems; catalog no. 4398881) and 3 μl of cDNA. The first degenerate PCR assay used primer pair HGLV-ak1 (5′-TACGCIACNGCIACNCCICC-3′) and HGLV-ak2 (5′-TCGAAGTTCCCIGTRTANCCIGT-3′) in the first round of PCR and HGLV-ak3 (5′-GACIGCGACICCICCIGG-3′) and HGLV-ak4 (5′-TCGAAGTTCCCIGTRTAICCIGT-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, 60°C for 1 min, and 72°C for 40 s, 30 cycles of 95°C for 30 s, 55°C for 45 s, and 72°C for 40 s, and a final extension at 72°C for 5 min. 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 40 s, 30 cycles of 95°C for 30 s, 57°C for 45 s, and 72°C for 40 s, and a final extension at 72°C for 5 min. The second degenerate PCR assay used primer pair AK4340F1 (5′-GTACTTGCTACTGCNACNCC-3′) and AK4630R1 (5′-TACCCTGTCATAAGGGCRTC-3′) for the first round of PCR. Primers AK4340F2 (5′-CTTGCTACTGCNACNCCWCC-3′) and AK4630R2 (5′-TACCCTGTCATAAGGGCRTCNGT-3′) were used in second round. For the first round, the PCR cycle 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 40 s, 30 cycles of 95°C for 30 s, 56°C for 45 s, and 72°C for 40 s, and a final extension at 72°C for 10 min. 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 40 s, 30 cycles of 95°C for 30 s, 58°C for 45 s, and 72°C for 40 s, and a final extension at 72°C for 10 min [6]). 5′ UTRs were determined using rapid identification of 5′ cDNA ends (5′ RACE) (36). 3′ UTRs were determined by poly(A), -(G), or -(U) tailing of viral RNA

using poly(A) polymerase (USB Affymetrix), followed by reverse transcription using adaptor-containing primers and subsequent PCR amplification. Thereafter, sequence validity was tested with 4-fold genome coverage by classical dideoxy Sanger sequencing.

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 of hepaciviruses and pegiviruses were generated using RAxML with the PROTGAMMA model (gamma distribution for rates over sites and Dayhoff amino acid similarity matrix with all model parameters estimated by RAxML) and 100 bootstraps (11). NS3 and NS5B trees for members of all four genera of flaviviruses (*Flavivirus*, *Pestivirus*, *Hepacivirus*, and *Pegivirus*) were generated by neighbor joining of Poisson-corrected pairwise distances.

RNA structures were predicted by Mfold and by homology searching and structural alignment with bases conserved in other hepaciviruses. Reliable structure prediction for the pseudoknot region in HCV (IIIf) and homologous pairings in other hepaciviruses cannot be predicted by Mfold or other conventional RNA secondary-structure prediction algorithms. Structure predictions were not attempted upstream of stem-loop III in the absence of detectable homology to other hepacivirus sequences or comparative sequence data from other RHV variants to support covariance or phylogenetic conservation analysis. Labeling of the predicted structures in the 5′ UTR followed the numbering used for reported homologous structures in HCV, GBV-B, and NPHV (7). We were unable to predict the structure the RPgV 5′ UTR due to insufficient data for structural alignment. Cleavage sites in the RHV polyprotein sequence were predicted for RHV at sites homologous to those of HCV that have been experimentally determined. Signalase sites between structural proteins were highly divergent between different hepaciviruses and could not be aligned, therefore those for RHV and NPHV were independently predicted using the SignalP version 4.1 program (16) and concordant with positions predicted from the sequence alignment. RPgV cleavage sites were similarly predicted by alignment of the NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B sites previously proposed for simian pegiviruses (18) and comparison of SignalP version 4.1 predictions for structural proteins of RPgV and the NS2/NS3 cleavage site of BPgV (8).

All complete genome sequences were examined for recombination using the programs Genetic Algorithm Recombination Detection (GARD) in the DataMonkey package, which provides an interface to the HyPhy program (38, 39). Default parameters were used with a Hasegawa, Kishino, and Yano (HKY) substitution model and a gamma distribution of 6 discrete rate steps. Rodent virus genome sequences were analyzed for evidence of genome-ordered RNA structures (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 numbers. The nucleotide compositions of viruses were determined using EMBOSS compseq (http://emboss.bioinformatics.nl/cgi-bin/emboss/compseq). All sequences generated in this study were submitted to GenBank under accession no. KC815310 to KC815327.

REFERENCES

1. **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.

2. **Beames B, Chavez D, Guerra B, Notvall L, Brasky KM, Lanford RE.** 2000. Development of a primary tamarin hepatocyte culture system for GB virus-B: a surrogate model for hepatitis C virus. J. Virol. **74**: 11764–11772.

3. **Berg T, Müller AR, Platz KP, Höhne M, Bechstein WO, Hopf U, Wiedenmann B, Neuhaus P, Schreier E.** 1999. Dynamics of GB virus C viremia early after orthotopic liver transplantation indicates extrahepatic tissues as the predominant site of GB virus C replication. Hepatology **29**:245–249.

4. **Blair CS, Davidson F, Lycett C, McDonald DM, Haydon GH, Yap PL, Hayes PC, Simmonds P, Gillon J.** 1998. Prevalence, incidence, and clinical characteristics of hepatitis G virus/GB virus C infection in Scottish blood donors. J. Infect. Dis. **178**:1779–1782.

5. **Williams CF, Klinzman D, Yamashita TE, Xiang J, Polgreen PM, Rinaldo C, Liu C, Phair J, Margolick JB, Zdunek D, Hess G, Stapleton JT.** 2004. Persistent GB virus C infection and survival in HIV-infected men. N. Engl. J. Med. **350**:981–990.

6. **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.

7. **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.

8. **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.

9. **Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML, Van Sant CL, Mushahwar IK.** 1995. Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc. Natl. Acad. Sci. U. S. A. **92**:3401–3405.

10. **Bukh J.** 2012. Animal models for the study of hepatitis C virus infection and related liver disease. Gastroenterology **142**:1279–1287.

11. **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.

12. **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.

13. **Morin TJ, Broering TJ, Leav BA, Blair BM, Rowley KJ, Boucher EN, Wang Y, Cheslock PS, Knauber M, Olsen DB, Ludmerer SW, Szabo G, Finberg RW, Purcell RH, Lanford RE, Ambrosino DM, Molrine DC, Babcock GJ.** 2012. Human monoclonal antibody HCV1 effectively prevents and treats HCV infection in chimpanzees. PLoS Pathog. **8**:e1002895. http://dx.doi.org/10.1371/journal.ppat.1002895.

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

15. **Kuiken C, Simmonds P.** 2009. Nomenclature and numbering of the hepatitis C virus. Methods Mol. Biol. **510**:33–53.

16. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.

17. **Branch AD, Stump DD, Gutierrez JA, Eng F, Walewski JL.** 2005. The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. Semin. Liver Dis. **25**:105–117.

18. **Simons JN, Desai SM, Schultz DE, Lemon SM, Mushahwar IK.** 1996. Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. J. Virol. **70**: 6126–6135.

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

20. **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.

21. **Bukh J.** 2011. Hepatitis C homolog in dogs with respiratory illness. Proc. Natl. Acad. Sci. U. S. A. **108**:12563–12564.

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

23. **Lindenbach BD, Evans MJ, Syder AJ, Wölk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, Rice CM.** 2005. Complete replication of hepatitis C virus in cell culture. Science **309**:623–626.

24. **Pietschmann T, Lohmann V, Kaul A, Krieger N, Rinck G, Rutter G, Strand D, Bartenschlager R.** 2002. Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. J. Virol. **76**: 4008–4021.

25. **Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Kräusslich HG, Mizokami M, Bartenschlager R, Liang TJ.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat. Med. **11**:791–796.

26. **Theodore D, Lemon SM.** 1997. GB virus C, hepatitis G virus, or human orphan flavivirus? Hepatology **25**:1285–1286.

27. **Xiang J, Klinzman D, McLinden J, Schmidt WN, LaBrecque DR, Gish R, Stapleton JT.** 1998. Characterization of hepatitis G virus (GB-C virus) particles: evidence for a nucleocapsid and expression of sequences upstream of the E1 protein. J. Virol. **72**:2738–2744.

28. **Mohr EL, Stapleton JT.** 2009. GB virus type C interactions with HIV: the role of envelope glycoproteins. J. Viral Hepat. **16**:757–768.

29. **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.

30. **Giret MT, Kallas EG.** 2012. GBV-C: state of the art and future prospects. Curr. HIV/AIDS Rep. **9**:26–33.

31. **Schountz T, Acuña-Retamar M, Feinstein S, Prescott J, Torres-Perez F, Podell B, Peters S, Ye C, Black WC, Hjelle B.** 2012. Kinetics of immune responses in deer mice experimentally infected with Sin Nombre virus. J. Virol. **86**:10015–10027.

32. **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.

33. **Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, Rice CM.** 2009. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature **457**:882–886.

34. **Kapoor A, Li L, Victoria J, Oderinde B, Mason C, Pandey P, Zaidi SZ, Delwart E.** 2009. Multiple novel astrovirus species in human stool. J. Gen. Virol. **90**:2965–2972.

35. **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.

36. **Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaukat S, Sharif S, Alam MM, Angez M, Wang C, Shafer RW, Zaidi S, Delwart E.** 2008. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. Proc. Natl. Acad. Sci. U. S. A. **105**: 20482–20487.

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

38. **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.

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