Bats are a major natural reservoir for hepaciviruses and pegiviruses


*Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY 10032; †EcologyHealth Alliance, New York, NY 10001; ‡Poxvirus and Rabies Branch, Division of High-Consequence Pathogens and Pathology, National Center for Emerging Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333; §Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa; ¶Centers for Disease Control and Prevention Guatemala, 01015, Guatemala City, Guatemala; #Centro de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Mexico D. F., Mexico; †Centers for Disease Control and Prevention Nigeria, Abuja, Nigeria; ‡Infectious Diseases Pathology Branch, Division of High-Consequence Pathogens and Pathology, National Center for Emerging Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333; §Department of Veterinary Medicine, Ahmadu Bello University, Samaru, Zaria, Kaduna State, Nigeria; ¶Center for Health Studies, Universidad del Valle de Guatemala, 01015, Guatemala City, Guatemala; #Laboratoire National Vétérinaire, B.P. 503 Garoua, Cameroon; *Centre de Surveillance de la Biodiversité, Universidad de Kigali, B.P. 125 Kisangani, Democratic Republic of the Congo; ‡Mammalogy Section, National Museums of Kenya, 00100 Nairobi, Kenya; #Sydney Emerging Infections and Biosecurity Institute, School of Biological Sciences and Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia

Although there are over 1,150 bat species worldwide, the diversity of viruses harbored by bats has only recently come into focus as a result of expanded wildlife surveillance. Such surveys are of importance in determining the potential for novel viruses to emerge in humans, and for optimal management of bats and their habitats. To enhance our knowledge of the viral diversity present in bats, we initially surveyed 415 sera from African and Central American bats. Unbiased high-throughput sequencing revealed the presence of a highly diverse group of bat-derived viruses related to hepaciviruses and pegiviruses within the family Flaviviridae. Subsequent PCR screening of 1,258 bat specimens collected worldwide indicated the presence of these viruses also in North America and Asia. A total of 83 bat-derived viruses were identified, representing an infection rate of nearly 5%. Evolutionary analyses revealed that all known hepaciviruses and pegiviruses, including those previously documented in humans and other primates, fall within the phylogenetic diversity of the bat-derived viruses described here. The prevalence, unprecedented viral biodiversity, phylogenetic divergence, and worldwide distribution of the bat-derived viruses suggest that bats are a major and ancient natural reservoir for both hepaciviruses and pegiviruses and provide insights into the evolutionary history of hepatitis C virus and the human GB viruses.

Determining the natural reservoirs for emerging pathogens is critical for disease prediction and prevention. Nearly 60% of emerging infectious diseases in humans is zoonotic, with up to 70% originating from wildlife (1). Bats [order Chiroptera, suborders Yinpterochiroptera and Yangochiroptera (2, 3)] are a natural reservoir for many important zoonotic viruses that cause severe disease in humans, including lyssaviruses, severe acute respiratory syndrome (SARS)-related coronaviruses (SARS-related CoV), flaviviruses, henipaviruses, and other paramyxoviruses (4–13). In addition, several viruses establish persistent viral infections in bats, in which apparent clinical signs appear only rarely (5). Bats possess unique characteristics that may contribute to their capacity to function as a major reservoir host for viruses, including long lifespan, high species diversity, unique immune systems, gregarious roosting behaviors, and high spatial mobility and population densities (5).

Although hepatitis C virus (HCV) was discovered more than 20 years ago, its origin is unknown, and, until recently, humans appeared to be its only natural host (14, 15). HCV was initially isolated from the serum of a person with non-A, non-B hepatitis and is now a leading cause of chronic liver disease, cirrhosis, and hepatocellular carcinoma, with ~3% of the global population infected (16, 17). HCV and its distant relative, GB virus B (GBV-B), belong to the genus Hepacivirus within the family Flaviviridae of single-stranded, positive-sense RNA viruses (18, 19). New World monkeys can be experimentally infected with GBV-B, resulting in clinical hepatitis (20). Recently, nonprimate hepaciviruses (NPHV) were discovered in domestic animals (dogs, horses), expanding the natural host range of these agents (21–23). A new flavivirus genus, Pegivirus, has been proposed for GB viruses GBV-A, GBV-C, and GBV-D (19). GBV-A viruses have been identified in nonhuman primates and are not known to infect humans, whereas GBV-C is frequently isolated from humans and chimpanzees (19). GBV-D was recently identified in Old World frugivorous bats (24). No pegiviruses have yet been identified to cause disease in their hosts (19).

Here, we report the discovery of a highly diverse group of bat-derived viruses related to hepaciviruses and pegiviruses. The prevalence, unprecedented viral biodiversity, and broad geographic distribution of these viruses, identified throughout the order Chiroptera, suggest that bats are a key natural reservoir for both hepaciviruses and pegiviruses.

Results

Discovery of Bat-Derived Hepaciviruses and Pegiviruses. We surveyed serum specimens from 415 apparently healthy bats captured from five different countries (Guatemala, Cameroon, Nigeria, 
Democratic Republic of the Congo, and Kenya) that represent 33 species, 26 genera, and seven families (*SI Appendix, Tables S1 and S2*). Aliquots of sera were combined into 43 pools for nucleic acid extraction; total RNA was randomly amplified and subjected to unbiased high-throughput sequencing (UHTS) (25). After removal of host sequences, each pool yielded between 87 and 3,098 unique reads and 1–128 assembled contiguous sequences. The analysis of these sequence data at the nucleotide (nt) and amino acid (aa) levels revealed the presence of sequences with distant similarity to flaviviruses in 29 sera pools. Between 1 and 716 flavivirus sequences, ranging from 50 to 10,803 nt in length, were identified in the positive pools. The sequences showed aa sequence similarities ranging from 24% to 100% to members of the genus *Hepacivirus* or *Pegivirus*. Specific and degenerate primers targeting a conserved motif in the RNA-dependent RNA polymerase gene (RdRp or NS5B) were designed from the UHTS data and used to confirm the initial findings, and further for screening all individual serum specimens in the positive pools. Using this approach, genetic material from a total of 58 bat-derived viruses was detected (*SI Appendix, Tables S3 and S4*). To further explore the distribution of these viruses in other countries and bat species, an additional 1,166 sera/plasma specimens and 83 tissue specimens from 34 bats (kidney, liver, and lung), representing 40 species, 31 genera, and seven families of bats from Nigeria, Bangladesh, and Mexico were screened (*SI Appendix, Table S2*). Twenty-three sera/plasma and one lung specimen were positive. Additional specimens from 9 of the bats from Mexico with positive sera specimens, consisting of seven oral and two rectal swabs were also tested. One positive rectal swab was detected (*SI Appendix, Tables S3 and S4*).

In total, 1,673 specimens, collected between 2007 and 2011 from 1,615 bats representing 58 species, 44 genera, and eight families in seven countries worldwide were screened (*SI Appendix, Table S2*). The genetic material of bat-derived viruses was detected in six of the eight families of bats tested, representing all major phylogenetic lineages of the chiroptera. A total of 78 positive sera/plasma, one lung specimen, and one rectal swab were tested positive,
Genetic Diversity of Bat-Derived Hepaciviruses and Pegiviruses. Preliminary phylogenetic analysis of a conserved 300-nt region of the RdRp gene revealed the presence of a highly diverse group of viruses that clustered within the Hepacivirus and Pegivirus genera (Fig. 1). Phylogenetic trees based on the complete helicase and RdRp coding sequences of a representative subset of these viruses, along with previously known hepac- and pegivirus sequences, supported the clustering of these two genera into 11 strongly supported clades [Bayesian posterior probability (BPP) values in all cases ≥0.9] (SI Appendix, Fig. S1). Importantly, these analyses show that, in both hepaciviruses and pegiviruses, all previously described viruses fall within the phylogenetic diversity of the bat-derived viruses described here, indicating that bats are a major and ancient reservoir for genetic diversity in both genera. Within the Hepacivirus genus, the viruses fell into three highly divergent clades composed solely of viruses from two species of African bats (Hipposideros vittatus, Otomops martiensseni) (BPP = 1.0). Clade A viruses were distinct from, but most closely related to GBV-B whereas clade C and D viruses occupied a basal position relative to the clades containing NPHV (clade E) and HCV (clade F). The bat-derived pegiviruses identified in this study formed three distinct lineages (BPP = 1.0). Viruses from the clades G and K formed phylogenetically distinct clusters within the genus whereas clade H viruses clustered with the previously identified GBV-D viruses (24). Notably, the African, North American, and Asian bat pegivirus lineages were all paraphyletic, suggesting a long history of diversification and cross-species transmission. Although we were unable to detect a history of recombination in our dataset, coinfections were identified in individuals from three species of African bats (Mops condylurus, O. martiensseni and Taphozous sp.) and one species from Bangladesh (Pteropus giganteus), suggesting the potential for recombination to occur. Three coinfections involved two distinct pegiviruses from clades G or K whereas, in the fourth case, a bat was infected with both a hepacivirus (clade C) and a pegivirus (clade K) (SI Appendix, Table S4).

Genomic Organization of Bat-Derived Hepaciviruses and Pegiviruses. Twenty near full-length genome sequences were generated, representing all six clades in which the bat hepaciviruses (BHVs) and bat pegiviruses (BPgVs) were identified (SI Appendix, Table S7). Similar to all other members of the Flaviviridae, the BHVs and BPgVs have a positive-sense, single-stranded RNA genome that contains a single large ORF encoding a polyprotein precursor (26). The ORF is flanked at the 5′ and 3′ end by nontranslated regions (NTRs). The BHV and BPgV genome sequences comprise at least 8,916 nt and encode precursor polyproteins of 2,842–3,469 aa in length (Fig. 2, SI Appendix, Table S7). Conserved flavivirus protein domains consistent with structural and nonstructural proteins were recognized in the BHV and BPgV predicted polyproteins, enabling accurate prediction of the location and function of the encoded proteins (SI Appendix, Tables S8 and S9). A comparison of predicted protein products indicated that the envelope (E) and the nonstructural proteins NS2 and NS5A were most variable, with less than 34.5%, 33.5%, and 20.4% (hepaciviruses) and 48.3%, 59.2%, and 37.7% (pegiviruses) aa sequence identities, respectively. In contrast, NS3 and NS5B were the most conserved with aa sequence identities in the range of 39.2–53.3% (NS3) and 35.3–46% (NS5B) for hepaciviruses and 43.8–78.7% (NS3) and 39.5–71.1% (NS5B) for pegiviruses (Datasets S1 and S2).

Among the 20 BHVs and BPgVs with complete polyprotein-coding sequences, 14 presented a genomic organization similar to that of hepaciviruses or pegiviruses (Fig. 2, SI Appendix, Table S7). Resembling HCV (27–30), the genome of the BHV PDB-829 in clade D contains a potential alternate reading frame (ARF) that overlaps the core protein-coding gene (SI Appendix, Fig. S2). The HCV ARF proteins (ARFPs) are derived either through ribosomal frameshift or through internal translation initiation (27, 30). The putative BHV PDB-829 ARF begins at the alternative GUG start codon located 41 nt upstream of the main C ORF and has a coding capacity of 156 aa (31). In contrast to HCV, there is no in-frame stop codon that terminates the BHV PDB-829 ARF at the start of the main C ORF (27, 31). Interestingly, the ARF identified in BHV PDB-829 is absent in the BHVs from clade C. Genomes of viruses from clade D presented a distinct genomic feature, a region of variable length referred to as the variable region (VR, 273–843 nt in length) located upstream of the predicted E1 (Fig. 3). No sequence homology to this region is observed in members of the Flaviviridae or in any other viral sequence present in the GenBank/European Molecular Biology Laboratory (EMBL) databases. The predicted aa coding sequences deduced from the VRs are basic (pI = 8.49–9.22) with two to four transmembrane regions. Predicted signal peptidase cleavage sites are identified 18–21 aa downstream of the putative initiation codon (Fig. 3). Further studies are necessary to determine whether the VR is part of E1 or encodes for a separate protein. The translated envelope protein sequences (E1, E2) of the BHVs and BPgVs contain 5–9 potential N-linked glycosylation sites, compared with 10–16 and 3–7 sites in hepaciviruses and pegiviruses, respectively (SI Appendix, Tables S8, S9 and Fig. S3A). The BHVs contain 8–11 of the 18 E2 cysteine residues strictly conserved among HCV genotypes and shown to form nine
disulfide bridges essential for proper folding of the envelope glycoprotein (SI Appendix, Fig. S3A and B) (32).

Discussion

Our study suggests that bats are a major natural reservoir for both hepaciviruses and pegiviruses. Hepaciviruses, formerly detected only in primates, were recently identified in horses and dogs (21–23). Pegiviruses have been found only in primates and in one species of Old World frugivorous bat (P. giganteus) (24). The viruses characterized in this study showed an unprecedented viral biodiversity and represent the most diverse repertoire of hepaciviruses and pegiviruses described to date (Fig. 1 and SI Appendix, Fig. S1). In addition, our phylogenetic analyses suggest that the bat-derived viruses are likely to occupy a basal position relative to the previously described hepaciviruses and pegiviruses although a wider sampling of various mammalian species is required to determine whether bats represent the ultimate reservoir host for these viruses. The BHVs and BPgVs are widely distributed, are present on several continents of the New and Old World, and are identified in both suborders of bats (Yinpterochiroptera and Yangochiroptera) from 21 species, 20 genera, and six families, indicating a widespread circulation (Fig. 4). Nearly five percent of the bats studied here were infected with BHVs or BPgVs (0.6% and 4%, respectively; 0.4% were undetermined). All bats collected were apparently healthy despite the high levels of viremia detected, suggesting that BHVs and BPVs may not be pathogenic in their hosts.

A long evolutionary history of these viruses in bats is further supported by the distribution across the different taxonomic groups of bats and geographic areas in which the viruses were found. The three distinct clades of BHVs that were all identified in Kenyan specimens were each associated with common insectivorous African bats (O. martiensseni and H. vittatus) of the distinctly related families Molossidae and Hipposideridae (Fig. 1). An additional BHV of undetermined clade was also identified in an insectivorous Scotocerus sp. (Family Vespertilionidae) bat from Kenya (SI Appendix, Table S4). The BPgVs clustered in three distinct clades, all of which were paraphyletic, indicating a long history of diversification and cross-species transmission. Whereas the viruses from clades G and H were separated by deep branches associated with different bat species (with the exception of PDB-303 and PDB-1479), the clusters in clade K were associated with multiple bat species from different genera and families, representing Yinpterochiroptera and Yangochiroptera bats. Moreover, within each cluster, the closely related BPgVs originated from different countries or even continents as well as from taxonomically distinct bat hosts. Transmission of viruses between bats may occur through shedding in excreta, via mating, or during gestation or parturition. However, the mechanism by which these viruses (or their ancestors) moved between continents remains to be elucidated. Further studies, especially in Asian bats, are likely to
provide additional information on the evolutionary history and geographical distribution of these viruses.

If bats are currently or historically implicated in transmission of hepatic- and pegivirus to other species, potential routes for transmission include consumption of fresh bat bodies, contamination of food with bat excreta, or direct exposure to bat blood or excreta or infection via intermediate hosts. Such mechanisms were identified for Nipah virus (NiV), where humans become exposed through contact with infected pigs, consumption of date palm sap contaminated with infected bat excreta, or through direct exposure to NiV-infected blood/excreta from the Indian flying fox, P. giganteus (33). In the case of SARS, the emergence of SARS-CoV in the human population probably occurred from bats to humans via palm civets present in Chinese wet markets (34–36). Whatever the mechanism of interspecies transmission, the distinct bat-derived pegivirus lineages identified in six bat families from Africa, Asia, and Central America and their widespread distribution indicate a much older association of pegiviruses with bats than with any human or nonhuman primate host. Similarly, whereas earlier studies have suggested that the evolution of both GBV-A and GBV-C is compatible with virus–host coevolution within primates (37–39), our phylogenetic analysis reveals multiple cross-species transmissions—both within bats and likely among other mammalian species—over an unknown time-scale. Therefore, the evolutionary history of the hepaticviruses and pegiviruses is clearly more complex than previously thought.

Taken together, our data indicate that bats are a major natural reservoir for hepaticviruses and pegiviruses. Although our current knowledge does not allow us to assess the true prevalence, host range, or geographical distributions of BHVs and BPgVs, it does reveal that these viruses are highly genetically diverse and are circulating in major lineages of Chiroptera in both the Old and New World. As such, our findings shed light on the possible evolutionary history of HCV and human GB viruses, which in turn opens avenues to calculating in major lineages of Chiroptera in both the Old and New World. Therefore, the evolutionary history of the hepaticviruses and pegiviruses is clearly more complex than previously thought.

Taken together, our data indicate that bats are a major natural reservoir for hepaticviruses and pegiviruses. Although our current knowledge does not allow us to assess the true prevalence, host range, or geographical distributions of BHVs and BPgVs, it does reveal that these viruses are highly genetically diverse and are circulating in major lineages of Chiroptera in both the Old and New World. As such, our findings shed light on the possible evolutionary history of HCV and human GB viruses, which in turn opens avenues to calculating in major lineages of Chiroptera in both the Old and New World. Therefore, the evolutionary history of the hepaticviruses and pegiviruses is clearly more complex than previously thought.

Taken together, our data indicate that bats are a major natural reservoir for hepaticviruses and pegiviruses. Although our current knowledge does not allow us to assess the true prevalence, host range, or geographical distributions of BHVs and BPgVs, it does reveal that these viruses are highly genetically diverse and are circulating in major lineages of Chiroptera in both the Old and New World. As such, our findings shed light on the possible evolutionary history of HCV and human GB viruses, which in turn opens avenues to calculating in major lineages of Chiroptera in both the Old and New World. Therefore, the evolutionary history of the hepaticviruses and pegiviruses is clearly more complex than previously thought.

Materials and Methods

**Bat Sample Collection.** Bats were captured at multiple sites in seven countries (SI Appendix, Table S1). Animals were collected with mist netting or hoop nets in caves and around human dwellings or from roost locations and foraging sites. All bats captured were apparently healthy. The species of each animal was identified by field biologists and recorded, as well as the sex and morphometric measurements (forearm, body lengths, and weight). Bats captured in Bangladesh and Mexico were anesthetized using isoflurane or manually restrained, respectively. Animals were safely released at the site of capture following sample collection. Oral and rectal swabs and blood were collected (when possible) from each animal. Bats sampled in Africa and Guatemala were anesthetized by intramuscular inoculation with ketamine hydrochloride (0.05–0.1 mg/kg of body weight) and euthanized under sedation by intracardiac exsanguination and cervical dislocation. Animals were necropsied by aseptic technique for tissue collection and immediately stored on dry ice or liquid nitrogen in the field, for final storage at –80 °C. One hundred microliters to 3 ml of blood were collected and placed immediately into serum separator tubes (BD vacutainer or microtainer; Becton Dickinson). Serum was kept on ice separated by centrifugation and immediately frozen on dry ice or liquid nitrogen. The capture and sample collection protocols were approved by the Institutional Animal Care and Use Committee of the Centers for Disease Control and Prevention (Atlanta, GA), the University of California (Davis, CA; protocol 16048), and the Tufts New England Medical Center (Boston, MA; protocol G2907) as well as by local authorities in the countries of sampling.

**Samples and High-Throughput Sequencing.** Aliquots of sera from bat specimens collected in Africa and Guatemala were combined into 43 pools for nucleic acid extraction using the NucliSens easyMAG (bioMérieux). Total nucleic acid was treated with DNase-I before random amplification and pyrosequencing (25). After assembly (Newbler v2.3; 454 Life Sciences), reads were clustered and assembled into contiguous fragments for comparison with the GenBank database of nucleic acids and proteins via the BLAST (40).

**Screening and Quantitative PCR.** All bat specimens were initially screened for viral genetic material in pools of 8–10 specimens. BHV and BPgV screening was performed using four nested PCR assays amplifying a 300-nt conserved fragment of the RdRp gene (N5SB) (corresponding to the nt positions 8127–8483 in the HCV-1a genome, GenBank accession number M62321). Consensus primers (SI Appendix, Table S11) were designed by multiple alignments of N5SB flavivirus nucleotide sequences identified in the unbiased high-throughput sequencing (UHTS) data. Thereafter, individual specimens from each positive pool were screened using the same set of primers. Additionally, specific primers were designed for all sera pools for which flavivirus sequences were identified in the UHTS data but were negative by N5SB consensus PCRs. Reverse transcription was performed using SuperScript II RT (Invitrogen). PCR primers were applied at 0.2 μM concentrations with 1.5 μL of cDNA (diluted 1:5) and HotStar polymerase (Qiagen). First round cycle conditions were: 95 °C for 7 min; 15 cycles at 95 °C for 30 s, 65 °C for 35 s (–1 °C per cycle), and 72 °C for 40 s; 35 cycles at 95 °C for 30 s, 48 °C for 30 s, and 72 °C for 40 s; and 1 cycle at 72 °C for 7 min. Second round cycling conditions were: 95 °C for 7 min; 35
cycles at 95 °C for 30 s, 48 °C for 30 s, and 72 °C for 40 s; and 1 cycle at 72 °C for 10 min. All PCR products were sequenced in both directions to confirm the presence of viral genetic material in samples. Quantitative SYBR green PCR assays were developed to determine the BHV or BPGV genome copy number. Standards were prepared for each assay by cloning the NS5B amplicon from representative viruses (see SI Appendix, Table S12 for primer sequences). Quantitative PCR (qPCR) reactions were cycled on the 7500 Fast Real-Time PCR System (Applied Biosystems) under the following conditions: 50 °C for 2 min, 95 °C for 10 min, then 45 cycles of 95 °C for 15 s and 60 °C for 60 s.

**Genome Sequencing and Phylogenetic and Evolutionary Analyses.** Details are provided in **SI Appendix, Table S6** and **S7**.

**ACKNOWLEDGMENTS.** We thank the King and Chairs of the Idanre community, Ondo State, Nigeria; the Federal Ministry of Health, Abuja, Nigeria; S. Wuyah, M. Lawal, M. Ari, A. Mohammed, I. Onaja, and G. Kia; the Faculty of Veterinary Medicine, Ahmadu Bello University (ABU), Zaria, Nigeria; and the Vice Chancellor and Management of ABU. We thank Luis Escobar, Alejandro Estevez, Maria Renée López, Ramon Medrano, Maria E. Morales, and Maria Luisa Muller (Center for Health Studies, Universidad del Valle de Guatemala) and Julio Martinez (Guatemala Ministry of Agriculture and Animal Health Department). We thank Sophonnya Yu, Edwin Dangwa, Evelyne Mulama, Solomon Gikundi, Leonardo Ndetiru, and Eric Ogola (Centers for Disease Control-Kenya), Rafael Ciraz (Guatemala Ministry of Public Health and Social Assistance), and Arif Islam (from Bangladesh) for excellent technical assistance and logistics. This work was supported by National Institutes of Health Grants AI051729 and AI051758 (Northeast Biodiversity Center) (to W.J.L.), National Institute of Allergy and Infectious Diseases Grant 5R01AI079321-02, US Agency for International Development PREDICT Grant GHNA 009 0001 000, and an award from the US Department of Defense. This study was also supported by the Emerging Pandemic Threats Program of the US Agency for International Development, the National Center for Emerging and Zoonotic Infectious Diseases, the Centers for Disease Control and Prevention (Atlanta, GA), and The Technical Support Corps funds from the Global Disease Detection Program of the Centers for Disease Control and Prevention (Atlanta, GA).