Identification of GBV-D, a Novel GB-like Flavivirus from Old World Frugivorous Bats (Pteropus giganteus) in Bangladesh

Jonathan H. Epstein1, Phenix-Lan Quan2, Thomas Briese2, Craig Street2, Omar Jabado2, Sean Conlan2, Shahneaz Ali Khan1,3, Dawn Verdugo2, M. Jahangir Hossain4, Stephen K. Hutchison5, Michael Egholm5, Stephen P. Luby4, Peter Daszak1*, W. Ian Lipkin2

1 Conservation Medicine Program, Wildlife Trust, New York, New York, United States of America, 2 Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York, United States of America, 3 Chittagong Veterinary & Animal Sciences University, Chittagong, Bangladesh, 4 Programme on Infectious Disease and Vaccine Sciences, International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B), Dhaka, Bangladesh, 5 454 Life Sciences, Branford, Connecticut, United States of America

Abstract

Bats are reservoirs for a wide range of zoonotic agents including lyssa-, henipah-, SARS-like corona-, Marburg-, Ebola-, and astroviruses. In an effort to survey for the presence of other infectious agents, known and unknown, we screened sera from 16 Pteropus giganteus bats from Faridpur, Bangladesh, using high-throughput pyrosequencing. Sequence analyses indicated the presence of a previously undescribed virus that has approximately 50% identity at the amino acid level to GB virus A and C (GBV-A and -C). Viral nucleic acid was present in 5 of 98 sera (5%) from a single colony of free-ranging bats. Infection was not associated with evidence of hepatitis or hepatic dysfunction. Phylogenetic analysis indicates that this first GB-like flavivirus reported in bats constitutes a distinct species within the Flaviviridae family and is ancestral to the GBV-A and -C virus clades.

Introduction

Bats (order Chiroptera), after rodents, comprise the most diverse group of mammals with more than 1,100 species. They are present on six continents, often have substantial habitat overlap with humans [1] and harbor several zoonotic viruses causing significant human morbidity and mortality, including Ebola- and Marburg-virus, Nipah virus (NiV), and SARS-like coronaviruses [2–5]. Proximity of bats to human populations may facilitate the zoonotic transmission of viruses either through direct contact, via amplifying domestic animal hosts, or through food-borne routes [6–8].

In close association with humans and feeding on cultivated fruit [14]. NiV transmission from bats to humans has been linked with the harvest and consumption of raw date palm sap, which becomes contaminated with bat feces, urine or saliva overnight when bats such as P. giganteus come to feed from the collecting pots [14,15]. Date palm sap or other foods eaten by both bats and people, may also serve as a vehicle for transmission of other bat-borne agents.

Several zoonotic flaviviruses, including Japanese encephalitis virus, West Nile virus, and Kyasanur forest virus have been identified in bats; however, to date, GB viruses have not [1]. GB viruses A and C (GBV-A and -C) represent two recently identified species that are currently unassigned members of the family Flaviviridae [16]. GBV-A viruses have been described in New World primates and are not known to infect humans [17–19], while GBV-C (also known as Hepatitis G virus (HGV)) have frequently been isolated from humans in many regions of the World, including India and Bangladesh [19–23], and from wild chimpanzees (Pan troglodytes) in Africa [24,25]. Here we describe discovery of a virus in the serum of healthy bats in Bangladesh, tentatively named GB virus D (GBV-D), that is distantly related to GBV-A and -C and represents a new member of the family Flaviviridae.

Materials and Methods

Ethics statement

Every effort was made to minimize bat stress and avoid injury during capture, restraint, and sampling procedures. This study was...
Bats are important reservoirs for emerging zoonotic viruses with significant impact on human health including lyssaviruses, filoviruses, henipaviruses and coronaviruses. Opportunities for transmission to humans are particularly prominent in countries like Bangladesh, where people live in close association with bats. Whereas previous studies of bats have employed assays that test for known pathogens, we present the first application of an unbiased molecular approach to pathogen discovery in this reservoir for emerging zoonotic disease. Unbiased pyrosequencing of serum from Pteropus giganteus bats enabled identification of a novel flavivirus related to Hepatitis C and GB viruses. Viral nucleic acid was present in 5 of 98 (5%) sera, and in the saliva of one animal. Sequence identification of two strains of the virus, tentatively named GBV-D, suggests P. giganteus as a natural reservoir. Detection of viral nucleic acid in saliva provides a plausible route for zoonotic transmission. Phylogenetic analysis indicates that GBV-D is ancestral to GBV-A and -C, and separate from the recently classified genus Hepacivirus. Our findings provide new insight into the range of known hosts for GB-like viruses and demonstrate the power of unbiased sequencing to characterize the diversity of potentially zoonotic pathogens carried by bats and other reservoirs.

Conducted following Wildlife Trust institutional guidelines under IACUC approval G2907 issued by Tufts New England Medical Center, Boston, Massachusetts.

Bat sample collection
As part of a longitudinal surveillance study of Nipah virus in bats, 98 free-ranging P. giganteus bats were caught from a colony in the Faridpur district of Bangladesh in December 2007 (Figure 1). Each bat was anesthetized using isoflurane gas; morphometric measurements (weight, forearm length, head length, and body condition) were taken and bats were aged [10]. Each bat was marked for future identification using an RFID microchip (AVID corp, www.avidid.com) implanted subcutaneously between the scapulae. Three mL of blood were collected and placed into serum separator tubes (vacutainer; Becton Dickinson, Franklin Lakes, NJ, USA). Serum was allowed to separate overnight at 4 °C then drawn off without centrifugation and immediately frozen using a liquid nitrogen dry shipper. To inactivate potentially infectious agents, serum samples were heat-treated at 65 °C for 30 min and then stored at −70 °C. For RNA extraction, 250 μL of serum was added to 750 μL TRI-Reagent LS (Molecular Research Center, Cincinnati, OH, USA). Saliva was collected from the bat’s throat using a sterile cotton swab. Urine was collected either by catching urine in a 1.0 mL sterile cryovial while the bat was urinating, or by urethral swab. Urine and saliva swabs were immediately placed into 1 mL TRI-Reagent LS and frozen in liquid nitrogen.

Unbiased high-throughput pyrosequencing (UHTS)
Total RNA from serum was extracted for UHTS analysis to screen for the presence of microorganisms. Five microliters of total RNA from each bat were combined into 4 pools: 4 pregnant bats; 4 non-pregnant female bats, and 2 pools of 4 adult male bats, respectively. Reverse transcription (RT) was performed on DNase I-treated (DNA-free, Ambion Inc., Austin, TX, USA) RNA pools to generate cDNA using Superscript II RT (Invitrogen, Carlsbad, CA, USA) and random octamers linked to a defined arbitrary, 17-mer primer sequence tail (MWG, Huntsville, AL, USA) [26]. After RNase H treatment cDNA was amplified by the polymerase chain reaction (PCR), applying a 9:1 mixture of the defined 17-mer primer sequence and the random octamer-linked 17-mer primer sequence, respectively [27]. Products of >70 base pairs (bp) were selected by column purification (MinElute, Qiagen, Hilden, Germany) and ligated to specific linkers for sequencing on the 4 34 Base Sequencer FLX (454 Life Sciences, Branford, CT, USA) without DNA fragmentation [28,29]. Sequences were analyzed using software applications implemented at the Greene Portal (http://tako.cpmc.columbia.edu/Tools/).

Genome sequencing
Multiple forward and reverse primers for RT-PCR (available upon request) were designed using the sequences obtained by UHTS in order to fill gaps between fragments. Amplifications were performed with Bio-X-act (Bioline, London, UK) according to manufacturer’s protocols. Products were size fractionated by electrophoresis and directly sequenced in both directions with ABI PRISM Big Dye Terminator 1.1 Cycle Sequencing kits (Perkin-Elmer Applied Biosystems, Foster City, CA, USA) at a commercial facility (Geneewiz, South Plainfield, NJ, USA). Additional methods applied to obtain the genome sequence included touch-down PCR [30], 2-step walking PCR [31], and 3′- and 5′- RACE (Invitrogen).

Quantitative real-time PCR
A real time Taqman PCR assay was developed to screen bat samples for GBV-D. Reactions were performed in a 25 μL volume by using commercial Taqman Universal Master Mix (Applied Biosystems, Foster City, CA, USA). Primers and probe were designed to target a 60 nt region in the NS4A gene region: Fadipro, 5′- gcAgCTGgGgTGgCCa; Fadi-reverse, 5′- aCACC-CATgATgTTACCACgAC; Fadi-probe, 5′- FAM-AggACCgTCgCTCCAgCA-T-BQX (TIB Molbiol, Adelphia, NJ, USA). Cycling conditions were: 50 °C for 2 min, and 95 °C for 10 min, followed by 45 cycles at 95 °C for 15 sec and 60 °C for 1 min. Thermal cycling was performed in an ABI 7300 real-time PCR system (Applied Biosystems).

Serum chemistry
A liver function panel was conducted at the International Center for Diarrheal Disease Research (Dhaka, Bangladesh) using non heat-treated bat sera (Automated Chemistry Analyzer AU 640, Olympus Corporation, Tokyo, Japan). The following parameters were analyzed: total protein, albumin, globulin, albumin:globulin ratio, total cholesterol, total bilirubin, alkaline phosphatase, alanine transference, aspartate aminotransferase, gamma glutamyltransferase, and lactate dehydrogenase.

Phylogenetic and sequence analyses
Sequence alignments were generated with ClustalW software [32] and phylogenetic relationships deduced using Geneious software [33]. Statistical significance was assessed by bootstrap re-sampling of 1000 pseudoreplicate data sets. Sequence relations were determined from p-distance matrices calculated with pairwise deletion for missing data and homogeneous patterns among lineages based on ClustalW alignments as implemented in MEGA software [34]. Sliding window similarity analysis was performed using SimPlot [35]. Potential signalase cleavage sites, glycosylation sites, and phosphorylation sites were analyzed using the respective prediction servers available at the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/).
Results

Identification of a GB-like agent from bats

Total RNA from the serum of healthy bats captured at a roost in the Faridpur district of Bangladesh was extracted for UHTS analysis. Extracts of 16 individual bats were combined into 4 pools consisting of 4 pregnant adult bats, 4 non-pregnant adult female bats, or 2 adult male bats. Each pool yielded between 1,400 and 2,000 assembled contigs or singleton reads (representing 50,000–75,000 reads ranging in size from 31–328 nt). Two reads of 238 and 215 nucleotides (nt) derived from the pregnant bat pool had distant homology to GBV-A sequences at the deduced amino acid (aa) level in the E2 and NS4A gene regions respectively (BLASTX); no homology was detected by searches at the nt level (BLASTN; local copy of the executables with standard settings except that the reward for a nucleotide match was set to 2 instead of 1). No viral sequences were detected in other pools at the nt or aa levels. Screening of the individual RNA preparations from the pregnant bat pool using primers derived from the UHTS reads confirmed the presence of the GBV-like sequence in the serum of bat 93. A quantitative real time PCR assay indicated a load of approximately 30,000 RNA copies in bat-93 serum extract, and identified an additional 4 positive bat sera from the original 98 samples (5/98; 5%), indicating serum loads ranging from 350 to 70,000 RNA copies per assay. These positive samples came from male bats that were not included in the initial UHTS pools. Extracts of saliva from the five positive bats indicated a load of approximately 200 RNA copies in bat 93; no signal was obtained with urine extracts from the five positive bats.

Genomic characterization of GBV-D

Near full-length genome sequence was generated from bat-93 and a second positive serum (bat 68), applying primers crossing

| Table 1. Percent sequence similarity between GBV-D (bat-68), -A, -C, and hepaciviruses. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| nt*             | aa              | GBV-D           | GBV-C           | GBV-A           | GBV-B           | HCV-1           |
| GBV-D           | 48              | 46              | 39              | 41              |                 |
| GBV-C           | 55              | 41              |                 |                 |                 |
| GBV-A           | 47              | 37              | 38              |                 |                 |
| GBV-B           | 24              | 24              | 24              | 41              |                 |
| HCV-1           | 22              | 22              | 22              | 27              |                 |

*Sequence similarity at nt level in upper right, and at aa level in lower left portion of the table. 
doi:10.1371/journal.ppat.1000972.t001

Figure 1. Map showing the location of the bat colony in Faridpur district, Bangladesh from which GBV-D was identified. 
doi:10.1371/journal.ppat.1000972.g001
gaps between UHTS reads as well as touch-down PCR [30], 2-step walking PCR [31], and 3'- and 5'-RACE (Invitrogen) protocols. The two genome sequences were 96% identical at the nt level (GenBank Accession nos. GU566734 and GU566735), indicating two strains of the same virus. Comparison of deduced polyprotein sequence to other GBV and hepaciviruses indicated highest nt and aa sequence identities to GBV-A and -C (Table 1, Figure 2). The genomic sequence of the GBV-like virus identified in *P. giganteus* bats, tentatively named GBV-D, comprises 9,633 nt with 52 nt of potentially 5'-untranslated region (UTR), one continuous open reading frame (ORF) of 9318 nt (3106 aa) and 265 nt of 3'-UTR (Figure 3).

Mature structural proteins in GB viruses, as well as other flaviviruses, are the product of cleavage by host signal peptidase [36]. In GBV-D the first potential signal sequence cleavage site is present after a stretch of 57, largely basic aa (6 kDa, p1 = 12), followed by sequence homologous to E1 (pfam 01539, http://pfam.sanger.ac.uk/) (Figure 3). The single glycosylation site N177IT present in that sequence is located in a position comparable to GBV-C, -A, -B and HCV glycosylation sites. Identification of the downstream E2 termini is less apparent as the next 580 aa contain multiple potential signal sequences and 10 potential glycosylation sites that indicate no homology to hepaciviral E2/NS1 (pfam 01560), until the sequence aligns with N-terminal NS2 motifs (pfam 01538) (Figure 2, Figure 3). However, despite similarity to pfam 01538 no signal sequence compatible with cleavage at A279G/F was found; cleavage may occur at G126/R, which combined with potential signalase cleavage at A759/F may indicate the existence of a heavily glycosylated potential 26 kDa product instead of the p7 transmembrane protein identified in HCV [37–39] or the 13 kDa variant described in GBV-B [40,41]. Conserved C-terminal motifs of the autocatalytic NS2/NS3 endoprotease domain are compatible with NS2/NS3 cleavage at S1067/A and comparable to other GBV and HCV [42]. Figure 3 indicates potential cleavage sites for NS3 (peptidase S29, pfam 02907; DEAD box helicase, pfam 07652; helicase C, pfam 00271), NS4A (pfam 01006), NS4B (pfam 01001), NS5A (domain-1a zinc finger, pfam 08300; domain-1b, pfam 08301), and NS5B (pfam 00998).

Conserved aa motifs were recognized in NS proteins. RNA-dependent RNA polymerase (RdRp) motifs in RdRp block III that are conserved with respect to other GBV and hepaciviruses were identified in NS5B (Figure 3) [43–46]. Potential phosphorylation sites are present at multiple serine (9), threonine (14) and tyrosine (4) residues in NS5A, compatible with its possible function as a phosphorylation-regulated mediator of viral replication [47]. However, significant conservation of primary sequence is not obvious for phosphorylation sites, proline-rich, or interferon-sensitivity
Figure 4. Phylogenetic relationship of GBV-D to other GBV and hepacviruses. GBV-D amino acid sequences for A: NS5B, B: NS3, and C: the polyprotein (PPT) were analyzed in comparison to representative sequences of GBV-A, -B, -C and hepatitis C viruses. GenBank accession numbers for
determining region motifs [48–50]. The C-terminal portion of NS3 has homology to conserved NTPase/helicase motifs [51]; the N-terminal portion includes conserved active triad residues H1123, D1147, S1204 of serine protease [52], the viral protease responsible for cleavage of mature non-structural proteins [53]. Likewise, the active triad H991, E1011, C1032 of the cis-acting protease activity in the C-terminal portion of NS2 is conserved with respect to other GBV and HCV [42]. The only other discernable motif identified was a well-conserved N75 C/D C motif at the N-terminus of E1 [42].

Phylogenetic analysis

Phylogenetic analysis of GBV-D was performed in comparison to selected representatives of GBV-A, GBV-B, GBV-C and HCV. Analysis of NS5B aa sequence (Figure 4A) confirmed a closer relationship of GBV-D to GBV-A and -C than to GBV-B or HCV as also indicated by pairwise sequence comparisons (Table 1). The same relationships were also apparent when NS3, or the complete polyprotein sequence were analyzed (Figure 4B and C, respectively). All three trees show GBV-D consistently at the root of the GBV-A/C viruses, indicating an independent phylogenetic clade compatible with a separate species distinct from the recently created genus Hepacivirus [16].

Table 2. Liver function values from Pteropus giganteus bats.

<table>
<thead>
<tr>
<th>Bat ID</th>
<th>T.Prot (g/dL)</th>
<th>Alb (g/dL)</th>
<th>Glob A:G</th>
<th>T. Chol (mg/dL)</th>
<th>T. Bill (mg/dL)</th>
<th>Alk Phos (U/L)</th>
<th>ALT (U/L)</th>
<th>AST (U/L)</th>
<th>GGT (U/L)</th>
<th>LDH (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref Range (human)</td>
<td>64.0–82.0</td>
<td>34–50</td>
<td>23–35</td>
<td>1.1–1.8</td>
<td>3.6–6.3</td>
<td>0.1–19</td>
<td>30–120</td>
<td>0.1–41</td>
<td>0.1–38</td>
<td>9–40</td>
</tr>
<tr>
<td>Ref Range (P. giganteus)</td>
<td>65–84</td>
<td>30–41</td>
<td>0–2.3</td>
<td>1.2–8.6</td>
<td>30.2–141.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bat ID</td>
<td>Alb (g/dL)</td>
<td>Glob</td>
<td>T. Chol (mg/dL)</td>
<td>T. Bill (mg/dL)</td>
<td>Alk Phos</td>
<td>ALT (U/L)</td>
<td>AST (U/L)</td>
<td>GGT (U/L)</td>
<td>LDH (mg/dL)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>86</td>
<td>50.7</td>
<td>35.3</td>
<td>1.44</td>
<td>1.65</td>
<td>1.8</td>
<td>1572.4</td>
<td>13.9</td>
<td>67.6</td>
<td>87.90</td>
</tr>
<tr>
<td>3</td>
<td>79.7</td>
<td>49.4</td>
<td>30.3</td>
<td>1.63</td>
<td>2.74</td>
<td>2.1</td>
<td>605.7</td>
<td>20.2</td>
<td>67.4</td>
<td>49.20</td>
</tr>
<tr>
<td>4</td>
<td>70.4</td>
<td>42.9</td>
<td>27.5</td>
<td>1.56</td>
<td>0.5</td>
<td>2.3</td>
<td>532.1</td>
<td>12.3</td>
<td>66.1</td>
<td>75.40</td>
</tr>
<tr>
<td>21</td>
<td>76</td>
<td>48</td>
<td>28</td>
<td>1.71</td>
<td>1.06</td>
<td>1.2</td>
<td>794.9</td>
<td>23.8</td>
<td>83.7</td>
<td>35.70</td>
</tr>
<tr>
<td>23</td>
<td>71.5</td>
<td>46.2</td>
<td>25.3</td>
<td>1.83</td>
<td>1.49</td>
<td>1.7</td>
<td>1650</td>
<td>12</td>
<td>115.1</td>
<td>102.00</td>
</tr>
<tr>
<td>33</td>
<td>80.5</td>
<td>43.8</td>
<td>36.7</td>
<td>1.19</td>
<td>1.09</td>
<td>2.1</td>
<td>891.6</td>
<td>22.2</td>
<td>146</td>
<td>32.10</td>
</tr>
<tr>
<td>36</td>
<td>89.9</td>
<td>51</td>
<td>38.9</td>
<td>1.31</td>
<td>1.22</td>
<td>2.5</td>
<td>580.4</td>
<td>11.1</td>
<td>109.4</td>
<td>69.00</td>
</tr>
<tr>
<td>37</td>
<td>76.6</td>
<td>47</td>
<td>29.6</td>
<td>1.59</td>
<td>0.85</td>
<td>1.3</td>
<td>205.8</td>
<td>12.6</td>
<td>50.3</td>
<td>72.50</td>
</tr>
<tr>
<td>39</td>
<td>70.4</td>
<td>43.3</td>
<td>27.1</td>
<td>1.6</td>
<td>0.5</td>
<td>1.8</td>
<td>328.7</td>
<td>20.5</td>
<td>77.8</td>
<td>31.00</td>
</tr>
<tr>
<td>68</td>
<td>78.9</td>
<td>44.9</td>
<td>34</td>
<td>1.32</td>
<td>2.54</td>
<td>1.2</td>
<td>813.3</td>
<td>7.9</td>
<td>65.6</td>
<td>46.10</td>
</tr>
<tr>
<td>73</td>
<td>82.4</td>
<td>46.1</td>
<td>36.3</td>
<td>1.27</td>
<td>1.17</td>
<td>1.8</td>
<td>958.8</td>
<td>19.7</td>
<td>185.4</td>
<td>37.00</td>
</tr>
<tr>
<td>82</td>
<td>72.3</td>
<td>43.8</td>
<td>28.5</td>
<td>1.54</td>
<td>0.5</td>
<td>1.4</td>
<td>458.4</td>
<td>22.1</td>
<td>82.1</td>
<td>48.50</td>
</tr>
<tr>
<td>84</td>
<td>82.6</td>
<td>46.2</td>
<td>36.4</td>
<td>1.27</td>
<td>1.43</td>
<td>2.7</td>
<td>923.6</td>
<td>22.9</td>
<td>107.6</td>
<td>46.40</td>
</tr>
<tr>
<td>93</td>
<td>82.1</td>
<td>47.5</td>
<td>34.6</td>
<td>1.37</td>
<td>1.46</td>
<td>2.6</td>
<td>262.2</td>
<td>8.6</td>
<td>59.9</td>
<td>47.10</td>
</tr>
<tr>
<td>96</td>
<td>76.5</td>
<td>46.8</td>
<td>29.7</td>
<td>1.58</td>
<td>0.5</td>
<td>1.1</td>
<td>574.6</td>
<td>20.9</td>
<td>79.4</td>
<td>101.80</td>
</tr>
<tr>
<td>Mean uninfected</td>
<td>78.73</td>
<td>47.15</td>
<td>31.58</td>
<td>1.518</td>
<td>1.226</td>
<td>1.81</td>
<td>763.9</td>
<td>18.7</td>
<td>97.55</td>
<td>60.85</td>
</tr>
<tr>
<td>Mean infected</td>
<td>77.7</td>
<td>45.22</td>
<td>32.48</td>
<td>1.406</td>
<td>1.288</td>
<td>1.9</td>
<td>540.96</td>
<td>12.72</td>
<td>77.58</td>
<td>54.64</td>
</tr>
<tr>
<td>STDv uninfected</td>
<td>6.52</td>
<td>2.63</td>
<td>4.71</td>
<td>0.19</td>
<td>0.68</td>
<td>0.52</td>
<td>391.61</td>
<td>4.62</td>
<td>35.31</td>
<td>27.40</td>
</tr>
<tr>
<td>STDv infected</td>
<td>4.56</td>
<td>1.99</td>
<td>3.80</td>
<td>0.17</td>
<td>0.78</td>
<td>0.62</td>
<td>311.19</td>
<td>5.71</td>
<td>38.77</td>
<td>18.63</td>
</tr>
</tbody>
</table>

No indication of hepatitis or impaired liver function was observed; no significant differences between mean values for infected (bold) or non-infected bats were apparent. The following parameters were analyzed: Total Protein (T. Prot); Albumin (Alb), Globulin (Glob), Alb:Glob ratio, Total Cholesterol (T. Chol); Total Bilirubin (T. Bill); Alkaline Phosphatase (Alk. Phos); Alanine Transferase (ALT); Aspartate Aminotransferase (AST); Gamma Glutamyltransferase (GGT) and Lactate Dehydrogenase (LDH). doi:10.1371/journal.ppat.1000972.t002

Discussion

Molecular analyses of sera from Pteropus giganteus bats from Faridpur, Bangladesh led to the identification of a 9,633 nt sequence consistent in genomic organization with known GBV and other species within the family Flaviviridae [16]. Whereas previous studies of bats have employed assays that test for known pathogens, ours is the first to report a novel flavivirus in bats.
first report of an unbiased molecular approach to pathogen discovery in this important reservoir of emerging infectious diseases. The modest yield of novel microbial sequences may reflect the choice of sample (e.g., serum vs feces, tissue or another specimen), competition between host and microbial template during unbiased amplification, or both. Efforts to address template competition are under way that include subtraction of host nucleic acids or the use of semi-random primers that do not amplify host sequences. Such efforts will likely enhance the sensitivity and throughput of unbiased sequencing technologies for pathogen discovery.

The discovery of this chiropteran flavivirus broadens both the taxonomical and geographical distribution of GB-like viruses. Three types of GB viruses have been described: GBV-A, -B and -C [18,19,24,25,54,57]. GBV-B, which has never been found in humans and was only reported in captive tamarins after serial passage of the original human GB serum [58], is most closely related to HCV and was recently classified together with HCV into a new genus, *Hepacivirus*, within the family *Flaviviridae* [16]. GBV-A and -C remain unclassified members of the family. GBV-A have been isolated from several New World monkeys. Different genotypes appear to be associated with specific monkey species of the genera *Saguinus*, *Callithrix* (*Callithrichidae* family) and *Aotus* (*Aotidae* family), without any clinical signs associated with infection [24,54,57]. GBV-C have been isolated from humans with non-A-E hepatitis; however, its pathogenicity is unknown and the virus is widespread in the human population [21,59-61]. Population studies showed that GB viruses are enzoonotic and species-specific within both Old and New World nonhuman primates as well as humans, and have likely co-evolved with their hosts over long periods of time [62]. Previously, the only GBV found in the Old world was GBV-C from chimpanzees (in Africa) and humans. Although GBV-C were found in humans, GB viruses have not been previously reported in primates or other animals on the Indian subcontinent.

GBV-C and -A are remarkable for a truncated or missing capsid (C) protein [18,19]. Due to exhaustion of our samples we were unable to complete assessment of the 5'-terminal sequence; nonetheless, RACE experiments suggest that GBV-D likely codes for a short basic peptide, instead of a full-length C protein. The first methionine (M1) predicts a peptide of 37 aa (pI = 12); however, the more favorable Kozak context [63] of M1 indicates a 55 aa peptide. After signalase cleavage from the polyprotein precursor, this peptide may be functional, possibly influencing maturation of, or directly binding to, the E1 and/or E2 glycoproteins.

Phylogenetic analyses of NS5B, NS3 and complete polyprotein sequence place GBV-D at the root of the GBV-A and -C clades and are consistent with a model wherein GBV-D is ancestral to GBV-A and -C clades. Mixed relationships indicative of recombination events [64] were not evident (Figure 2, Figure 4). Both proropid bats and chimpanzees are restricted to the Old World. While the range of chimpanzees (Africa) and *P. giganteus* (the Indian subcontinent) do not overlap, it is possible that other primate species in Bangladesh or India, such as macaques, or other fruit bats in Africa such as *Epolallo spp.*, whose range overlaps that of chimpanzees, may carry related viruses. While GBV-A is only known from primates of the New World, an African origin has been suggested for GBV-C based on a 12-aa indel sequence in NS5A [65]. Although the NS5A sequence of GBV-D, similar to that of GBV-A, appears elongated in the indel region, compatible with their respective earlier phylogenetic branching compared to GBV-C, little sequence conservation is observed in that region.

The bats in this study, like primates infected with their associated GBV [66], all appeared to be healthy. The lack of chemical evidence of hepatic inflammation or dysfunction suggests that this virus may not target hepatic cells in bats. This is consistent with the behavior of GBV-A in its natural primate hosts [54]. In contrast, elevated alanine transferase levels and mild hepatitis are observed in experimental infections of macaques with GBV-C isolates from humans [67]. Five percent of the bats we studied were infected with one of at least two different strains of GBV-D, which suggests widespread viral circulation within this species. The observation that bats are asymptptomatically infected with diverse strains that constitute a distinct phylogenetic clade is compatible with a co-evolutionary relationship between GBV and their hosts [57,62], and supports the hypothesis that *P. giganteus* bats may be a natural reservoir for GBV-D. In one case we were able to detect GBV-D nucleic acid in saliva. This suggests a potential route for viral transmission via fighting or grooming behavior, or via food shared by bats.

*Pteropus giganteus* is a frugivorous bat species that carries NiV, a zoonotic paramyxovirus [10,11]. This species lives in close association with humans in Bangladesh and bats have been observed drinking from (and urinating into) date palm sap collecting pots [14]. Human consumption of contaminated palm juice is proposed to be a major route of NiV transmission [68]. Although it is unclear whether infectious virus was present in bat saliva, the observation that saliva can contain GBV-D nucleic acids provides a biologically plausible mechanism for transmission from infected bats to other hosts. While it is currently unknown whether GBV-D virus occurs in humans, up to 20% of non-A-E hepatitis cases remain unexplained [19].

**Acknowledgments**

We thank the Forestry Department of Bangladesh for permission to conduct this research; Md. Sheikh Gofur and Md. Pitu Biswas for help in sampling bats; A. Bennett, A. Tashmukhamedova, and R. Tokarz for technical support, and K. Olival for critical comments on the manuscript.

**Author Contributions**

Conceived and designed the experiments: JHE TB PD WIL. Performed the experiments: JHE PLQ DV SH. Analyzed the data: JHE PLQ TB CS OJ DV SKH ME PD. Contributed reagents/materials/analysis tools: JHE TB CS SC SAK MJH SKH ME PD. Contributed field and lab activities and logistics for work in Bangladesh; contributed to paper-writing: MJH. Coordinated permissions, field activities and logistics for work in Bangladesh; contributed to paper-writing: SPL.

**References**


