Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus

N. J. Adams,1 L. E. Prescott,1 L. M. Jarvis,1 J. C. M. Lewis,2 M. O. McClure,3 D. B. Smith1 and P. Simmonds1

1 Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK
2 International Zoo Veterinary Group, Keighley, Yorkshire BD21 1AG, UK
3 Department of Communicable Diseases, Jefferiss Research Trust, St Mary’s Hospital Medical School, Praed Street, London W2 1NY, UK

Infection with hepatitis G virus (HGV) or GB virus-C (GBV-C) is widely distributed in human populations. Viruses related to GBV-C/HGV have been recovered from several New World primate species, including tamarins, owl monkeys and marmosets. To understand more about the relationship between GB viruses and their hosts, we used primers from the 5′ non-coding (5′ NC), non-structural 3 (NS3) and NS5 regions in nested polymerase chain reactions to screen for related viruses infecting non-captive chimpanzees (Pan troglodytes, troglodytes and verus subspecies). Sequences from the 5′ NCR and NS5 regions were amplified from samples taken from 3 of 39 chimpanzees, and from one chimpanzee in the NS3 region. Sequence comparisons of each region revealed that the GB virus infecting chimpanzees was distinct from both GBV-C/HGV and from any of the known GBV-A sequences, but was more closely related to human viruses. GB viruses recovered from different chimpanzees were more diverse than variants of GBV-C/HGV found in humans, with 25% sequence divergence in the 5′ NCR and 20% (9.5 ± 5% amino acid) sequence divergence in NS5 between variants recovered from the troglodytes and verus subspecies, compared with 7.4% and 10.4% (1.9 ± 9% amino acid) divergence amongst GBV-C/HGV variants infecting humans. Finding GBV-C/HGV-related viruses in an Old World monkey species suggests that GB-like viruses may be widely distributed in simians, and suggests a close evolutionary relationship with their natural hosts.

Introduction

Considerable progress has recently been made in the search for agents implicated in post-transfusion and other forms of acute and chronic hepatitis (Linnen et al., 1996; Leary et al., 1996a, b; Nishizawa et al., 1997; Choo et al., 1991). Amongst these, hepatitis C virus (HCV) has been shown to be one of the principal causes of post-transfusion non-A, non-B hepatitis, and the development of screening assays for blood donors has led to a dramatic reduction in its incidence. More recently, two groups have independently identified a second virus that infects humans, but without a clear association with chronic hepatitis or other recognizable disease syndrome (Linnen et al., 1996; Leary et al., 1996a, b). The virus, described variously as hepatitis G virus (HGV) or GB virus-C (GBV-C), is most closely related to HCV, although it shows only distant nucleotide sequence similarity with HCV and other members of the Flaviviridae which is limited to specific enzymatic motifs in the NS3 and NS5B genes. There is little or no similarity in the number and arrangement of genes encoding the structural capsid and envelope genes.

Analysis of samples collected from tamarins and marmosets experimentally inoculated with serum from an individual (G.B.) with acute hepatitis of unknown aetiology led to the chance discovery of two novel flaviviruses that infect New World primates (Simons et al., 1995). One virus, GBV-B, showed a genome organization most similar to that of HCV, with an internal ribosome entry site (IRES) of similar structure to that of HCV and pestiviruses, and a homologue of the core gene of HCV. This virus causes acute hepatitis in tamarins...
Fig. 1. Nucleotide sequences from (a) the 5'NCR of the GBV-C/HGV-related virus recovered from three chimpanzees (Chimp 30 and Chimp 30/2 (subspecies *troglodytes*), Chimp 23 and Chimp 33 (subspecies *verus*) and inferred amino acid sequences from (b) NS3 and (c) NS5 regions. Sequences were compared with representative variants of GBV-C/HGV and GBV-A (sources listed in Methods). Sequences numbered from the 5' end with reference to nucleotide (a) or amino acid (b, c) positions in the sequence of PNF2161 (Linnen et al., 1996).

Symbols: '-': gap introduced to preserve alignment; '.': nucleotide or amino acid residue same as PNF2161; ' ': nucleotide or codon not determined.
Schlauder et al., 1995), but has not to date been recovered from wild-caught primates in the New World, and its natural reservoir remains unknown.

The other virus, the prototype variant of GBV-A [referred in this study as GBV-A(PT)], shows greatest similarity to GBV-C/HGV, with overall amino acid sequence similarity of 48% over the length of the genome. It contains an IRES similar in structure to GBV-C/HGV (Simons et al., 1996) (and distinct from HCV and GBV-B) and the same puzzling absence of an obvious coding sequence for the core protein. GBV-A is found in a wide range of New World primate species in which it causes a chronic infection without obvious disease associations (Schlauder et al., 1995). Variants of GBV-A, showing approximately 60–70% sequence similarity to each other, show specific associations with different host species; to date six different variants have been recovered from four tamarin species (Saguinus labiatus, S. mystax, S. nigricolor and S. oedips), and the more distantly related Aotus trivirgatus (owl monkey) and Callithrix jacchus (marmoset) (Erker et al., 1998; Leary et al., 1996a, b; Buhk et al., 1997).

As a strategy to detect GB-like viruses in other primate species, we have developed a PCR using primers corresponding to highly conserved regions (5’ NCR, NS3 and NS5) of currently available sequences of GBV-A and GBV-C/HGV. These primers were used to screen non-captive chimpanzees originating from different geographical areas in Central and West Africa to investigate whether members of this Old World primate species were also infected with viruses related to GBV-A and GBV-C/HGV.

Methods

Chimpanzee samples. Sera were collected from 39 chimpanzees originating in West Africa (Pan troglodytes, subspecies verus), and from Pan troglodytes, subspecies troglodytes originating from Cameroon and Nigeria. Their precise geographical origin was not accurately ascertained because samples from the chimpanzees were obtained after capture by local hunters. Samples were shipped frozen to St Mary’s Hospital Medical School where they were aliquoted, and dispatched for PCR testing in Edinburgh. Two samples were available from one chimpanzee (Chimp 30 and 30/2, verus subspecies), collected approximately 24 months apart.

Sample extraction and PCR. RNA was extracted from 50–100 µl volumes of plasma using proteinase K–SDS lysis buffer as previously described (Jarvis et al., 1994) and resuspended in a 20 µl final volume of RNase-free water; 5 µl of RNA was reverse transcribed using random hexamers in a total volume of 20 µl at 66 min at 42 °C, and 5 µl of the cDNA used in a nested PCR using previously described buffers and volumes (Jarvis et al., 1994). Primers for amplification of the 5’ NCR used in the first round of the PCR were 5’ AGGGTTTGAGGTGTGAAATCCCA 3’ [sense, 5’ base position 129 in the PNF2161 genome (Linnen et al., 1996), position 161 in GBV-A(PT); Simons et al., 1995] and 5’ TGCCACCGGACCTACCGGAGG 3’ [antisense, position 531 in PNF2161, 551 in GBV-A(PT)]. A second PCR was carried out using 1 µl of amplified product from the first round using 5’ TGGGCTTACCC-GGTTGAAATGA 3’ [sense, position 242 in PNF2161, 274 in GBV-A(PT)] and 5’ CCCCCACTGGGATTTGACAACCCACG 3’ [antisense, position 402 in PNF2161, 419 in GBV-A(PT)]. Temperatures and times for both first and second rounds of the PCR were as follows: 40 cycles of 60 s at 94 °C, 120 s at 54 °C and 180 s at 72 °C.

Primers for amplifications of NS3 were 5’ ATGGGCCCNTACAT-GGA 3’ [sense, position 3933 in PNF2161, 4178 in GBV-A(PT) and GGTCSACYTNCNAACCTCCTC [antisense, position 4544 in PNF2161, 4790 in GBV-A(PT)]; temperatures and times were 40 cycles of 60 s at 94 °C, 120 s at 50 °C and 180 s at 72 °C. The second PCR used primers 5’ ACNTAYGGNAGRTTYNGGC 3’ [sense, position 4038 in PNF2161, 4284 in GBV-A(PT)] and 5’ ACCCTCTCMACCAAYAANNCCACA 3’ [antisense, position 4531 in PNF2161, 4777 in GBV-A(PT)]; temperatures and times were 40 cycles of 60 s at 94 °C, 120 s at 54 °C and 180 s at 72 °C. Primers for amplifications of NS5 were 5’ TGGACTCTCGGATAGCTGA 3’ [sense, position 7867 in PNF2161, 8242 in GBV-A(PT)] and 5’ ARGNCGCTNW5ACYTRATGTARCA 3’ [antisense, position 8265 in PNF2161, 8640 in GBV-A(PT)]; temperatures and times were 40 cycles of 60 s at 94 °C, 120 s at 55 °C.

<table>
<thead>
<tr>
<th>Virus group</th>
<th>HGBV-C/HGV</th>
<th>Chimpanzee</th>
<th>Human</th>
<th>GBV-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) 5’ NCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;CPZ&lt;/sub&gt;</td>
<td>3</td>
<td>24±7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;HUM&lt;/sub&gt;</td>
<td>9</td>
<td>35±7</td>
<td>7±4</td>
<td></td>
</tr>
<tr>
<td>GBV-A</td>
<td>7</td>
<td>41±4</td>
<td>43±0</td>
<td>25±9</td>
</tr>
<tr>
<td>(b) NS3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;CPZ&lt;/sub&gt;</td>
<td>1</td>
<td>-- (--)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;HUM&lt;/sub&gt;</td>
<td>8</td>
<td>26±7(13–6)</td>
<td>15±6(1–6)</td>
<td></td>
</tr>
<tr>
<td>GBV-A</td>
<td>4</td>
<td>39±2(39–7)</td>
<td>40±9(38–3)</td>
<td>34±1(25–1)</td>
</tr>
<tr>
<td>(c) NS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;CPZ&lt;/sub&gt;</td>
<td>3</td>
<td>19±3(9–5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C/HGV&lt;sub&gt;HUM&lt;/sub&gt;</td>
<td>8</td>
<td>28±8(19–7)</td>
<td>10±4(1–9)</td>
<td></td>
</tr>
<tr>
<td>GBV-A</td>
<td>4</td>
<td>43±3(39–6)</td>
<td>40±5(37–3)</td>
<td>32±0(24–0)</td>
</tr>
</tbody>
</table>
Fig. 2. For legend see facing page.
Fig. 2. Phylogenetic analysis of sequences from (a) the 5'NCR using the alignment shown in Fig. 1, (b) NS3 and (c) NS5, using uncorrected nucleotide distances and neighbour-joining. Numbers on branches indicates number of bootstrap resamplings from 100 supporting observed phylogeny, restricted to values of 75% or greater. Sequences compared with representative sequences from GBV-C/HGV, and available sequences from GBV-A (sources listed in Methods). p distances indicated on scale bar.

and 180 s at 72 °C. The second PCR used primers 5' AAGMTNAT-YYTGGGNGACC 3' [sense, position 7887 in PNF2161, 8262 in GBV-A(PT)] and 5' ACCCCNTGNGWNACCAT 3' [antisense, position 8173 in PNF2161, 8548 in GBV-A(PT)]; temperatures and times were 40 cycles of 60 s at 94 °C, 120 s at 55 °C and 180 s at 72 °C.

Following amplification, PCR product was analysed by agarose gel electrophoresis and ethidium bromide staining, and the sizes of the bands estimated by standard size markers. Amplified DNA was directly sequenced by cycle sequencing using a thermostable DNA polymerase (Amersham) from both ends using the same primers as used for the PCR.

**Sequence analysis.** Nucleotide sequences were aligned using the ClustalV program (5'NCR) or manually (NS3, NS5). For comparison of GBV-C/HGV sequences, the following complete genomic sequences were used: GBV-C: U36380, PNF2161: U44402; R10291: U45966; GBV-C(EA): U63715; HGV-Iw: U73536; GT110: D90600; GT230: D90601; HGV-Iw: D87255; GBV-A(PT): U22303; GBV-A_LAB: U194241 (T1059); GBV-A_mys: AF023424 (70047) and GBV-A_tr: AF023425 (1122). Partial sequences from the 5'NCR compared were U59533, U59538, U59542, U93251 (owl769), U93252 (owl896), U93254 (T1044), U84941 (SN98) and U84942 (SO14). Sequences were labelled by their species of origin in full and/or by the abbreviations: SL, SM, AT and CJ (Fig. 1). The sequences mx70047 (U93245) and mx80149 (U93246) were originally described as originating from S. mystax, but have been relabelled as originating from S. mystax in the current study as their actual origin from marmosets was recently reported. Nucleotide sequences were compared using the MEGA package (Kumar et al., 1993) using p distances for the 5'NCR, NS3 and NS5 regions, the latter two including a separate comparison of inferred amino acid sequences. Sequence groupings were found in the phylogenetic trees by bootstrap re-sampling. Sequences obtained in this investigation have been submitted to GenBank and bear the accession numbers AF068905–AF068913.

**Results**

Primers for amplification of the 5'NCR were chosen to match as closely as possible sequences from known variants of both GBV-C/HGV and GBV-A from different New World primate species. DNA amplicons corresponding approximately to the expected lengths of 146–161 bp for GBV-A(PT) and GBV-C/HGV respectively were obtained upon amplification of 3 of the 39 samples collected from wild-caught chimpanzees from West Africa. Chimp 23 and Chimp 33 were the *troglodytes* subspecies, and originated from Cameroon and Nigeria respectively. Chimp 30 was the *verus* subspecies originating from West Africa. Nucleotide sequences of the amplified DNA fragments were determined by direct sequencing of the PCR product (Fig. 1a). The sequences obtained differed considerably from those reported previously for GBV-A and GBV-C/HGV, containing a large number of unique nucleotide substitutions and insertions/deletions. Using the alignment shown in Fig. 1, the chimpanzee sequences (referred to provisionally as GBV-C/HGV_CPZ) showed a mean sequence divergence of 36% with representative sequences of GBV-C/HGV from humans (GBV-C/HGV_HUM) and 41% with those of GBV-A (Table 1a).
These degrees of sequence divergence were greater than found amongst variants of GBV-C/HGV<sub>HUM</sub> (7-4%), or amongst variants of GBV-A recovered from the same species (maximum 7%).

Phylogenetic analysis of the sequences showed the chimpanzee sequences formed a separate branch in the tree (Fig. 2a) whose position was consistent with their greater degree of sequence similarity to GBV-C/HGV<sub>HUM</sub> than to GBV-A. Sequences from GB viruses infecting New and Old World primate species (including humans) clustered separately with 98% bootstrap support. Within the Old World group of GB viruses, sequences recovered from the chimpanzees were monophyletic (78% bootstrap support), but were considerably more heterogeneous than variants of GBV-C/HGV<sub>HUM</sub> recovered from the current human population, and between GBV-A variants recovered from pairs of infected New World primate species (Table 1a). Amongst the chimpanzee variants, the sequences from the <i>troglodytes</i> subspecies (Chimp 23 and Chimp 33) showed 16% divergence from each other, but 28% divergence from variants recovered from the <i>vulvus</i> subspecies (Chimp 30).

The three samples from which sequences from the 5’NCR were amplified were retested using the NS3 and NS5 primers. One of the three samples (Chimp 33) produced amplicons of approximate size 500 bp using the NS3 primers, close to the expected sizes of NS3 sequences amplified from GBV-C/HGV and GBV-A(PT). Its nucleotide sequence showed dissimilarity to the homologous sequences of both GBV-C/HGV<sub>HUM</sub> and GBV-A at both the nucleotide and amino acid level (Fig. 1b, 2b; Table 1b), although it was more closely related to GBV-C/HGV<sub>HUM</sub> consistent with relationships observed in the 5’NCR. Confirming the close relatedness of GB viruses from human and chimpanzees, the mean amino acid sequence divergence between variants of GBV-C/HGV<sub>CPZ</sub> and GBV-C/HGV<sub>HUM</sub> was 13.6%, substantially less than the 39.7% mean divergence between GBV-C/HGV<sub>CPZ</sub> and GB viruses from New World primate species (Table 1b). Their closer relatedness was also indicated by phylogenetic analysis of NS3 sequences which showed that the chimpanzee sequence shared a common ancestor with GBV-C/HGV<sub>HUM</sub> distinct from the ancestor of GBV-A (Fig. 2b).

Using the NS5 primers, sequences were amplified from all four chimpanzee samples positive in the 5’NCR. Sequence comparisons with GBV-C/HGV<sub>HUM</sub> and GBV-A revealed similar relationships to those found in NS3 and the 5’NCR (Fig. 1c, 2c, Table 1c). As found from analysis of the 5’NCR, the degree of sequence divergence between different chimpanzees (mean 19.3% nucleotide, 9.5% amino acid) was much greater than between variants of HGV/GBV-C recovered from humans (mean 10.4% nucleotide, 1.9% amino acid). The degree of divergence of HGV/GBV-C<sub>CPZ</sub> from different chimpanzee subspecies was, however, less than found between GBV-A variants recovered from different New World primate species (mean 32% nucleotide, 24% amino acid).

### Discussion

GBV-A shows a highly species-specific pattern of sequence divergence, in which virus variants recovered from different species (Erker et al., 1998; Bukh et al., 1997; Leary et al., 1996a, b, 1997) reproduce the underlying degree of relatedness of their natural hosts (Jones et al., 1992). If it can be demonstrated that viruses infecting different New World primate species had diverged in concert with their hosts, the current divergence between variants of GBV-A of 30–45% must have originated by a process of divergent evolution over the past 5–10 million years. Over this timescale, the long-term rate of sequence change in both the 5’NCR and in coding regions such as NS3 would have been considerably lower than the rate recently calculated from comparisons of nucleotide sequences of GBV-C/HGV<sub>HUM</sub> recovered from an infected individual over a period of 8-4 years (Nakao et al., 1997).

In the current study, the closer relationship of the chimpanzee sequences to those found in humans compared with those of New World primates is consistent with the proposed species-specific association of GB viruses with their hosts. The time of divergence of humans from chimpanzees is more recent (around 7 million years) than the minimum estimates from the divergence of Old and New World primates (35 million years). However, there was also a high degree of sequence divergence between 5’NCR and NS5 sequences amplified from different subspecies of the chimpanzees (Table 1a, c), greater than found between variants of GBV-C/HGV<sub>HUM</sub> recovered from geographically dispersed human populations, but less than between-species variation of GBV-A recovered from different New World primates (Bukh & Apgar, 1997; Leary et al., 1996a, b). One explanation is that the variants recovered from the different chimpanzees represent GB agents transmitted from other primate species, although none of the infected chimpanzees tested in the current study had been housed with other primates before sampling. The potential for cross-species transmission is illustrated by the experimental infection of a chimpanzee with GBV-C/HGV<sub>HUM</sub> (J. Bukh and others, presented at the 4th International Meeting on Hepatitis C and Related Viruses, Kyoto, March, 1996), and by the passaging of the prototype GBV-A strain through both <i>S. mystax</i> and <i>S. labiatus</i> (Simons et al., 1995).

However, the grouping together of the GBV-C/HGV<sub>CPZ</sub> sequences (78% and 100% bootstrap support in 5’NCR and NS5, Fig. 2a, c) is inconsistent with the phylogeny expected if the variants originated from different primate species. Instead, the great variability of GBV-C/HGV<sub>CPZ</sub> is more likely to have resulted from the relative age and diversity of the surviving wild chimpanzee populations in Central and West Africa. Different subspecies of <i>Pan troglodytes</i> are likely to have diverged at least 1 million years ago, considerably earlier than the age of the human population, which is proposed to have originated from a common ancestor as recently as
100000–150000 years ago. Using the detection methods developed in the current paper, a more extensive sampling of other Old World primates, such as gorillas, orang-utans and more distantly related species is clearly possible, and comparison of the sequences obtained may provide more evidence for the species-specificity of GB agents, and the unusual relationship they have with their hosts.

References


Received 11 February 1998; Accepted 29 April 1998