The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae

Jack T. Stapleton,1 Steven Foung,2 A. Scott Muerhoff,3 Jens Bukh4,5 and Peter Simmonds5

1Department of Internal Medicine, Veterans Administration Medical Center and the University of Iowa, Iowa City, IA, USA
2Department of Pathology, Stanford University, Stanford, CA, USA
3Abbott Diagnostics Division, Abbott Laboratories, Abbott Park, IL, USA
4Copenhagen 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 Sciences, University of Copenhagen, Denmark
5Centre for Infectious Diseases, University of Edinburgh, Edinburgh, UK

In 1967, it was reported that experimental inoculation of serum from a surgeon (G.B.) with acute hepatitis into tamarins resulted in hepatitis. In 1995, two new members of the family Flaviviridae, named GBV-A and GBV-B, were identified in tamarins that developed hepatitis following inoculation with the 11th GB passage. Neither virus infects humans, and a number of GBV-A variants were identified in wild New World monkeys that were captured. Subsequently, a related human virus was identified [named GBV-C or hepatitis G virus (HGV)], and recently a more distantly related virus (named GBV-D) was discovered in bats. Only GBV-B, a second species within the genus Hepacivirus (type species hepatitis C virus), has been shown to cause hepatitis; it causes acute hepatitis in experimentally infected tamarins. The other GB viruses have however not been assigned to a genus within the family Flaviviridae. Based on phylogenetic relationships, genome organization and pathogenic features of the GB viruses, we propose to classify GBV-A-like viruses, GBV-C and GBV-D as members of a fourth genus in the family Flaviviridae, named Pegivirus (pe, persistent; g, GB or G). We also propose renaming ‘GB’ viruses within the tentative genus Pegivirus to reflect their host origin.

Introduction

The International Committee on Taxonomy of Viruses (ICTV) provides guidelines for virus nomenclature and classification based on orders (-virales), families (-viridae), subfamilies (-virinae), genera (-virus) and species (The Universal Virus Database of the International Committee on Taxonomy of Viruses; http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm). A species is a ‘polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche’. Several properties must be present to differentiate individual species, including differences in genome sequences, host range, cell and tissue tropism, pathogenicity or cytopathology, physical properties and antigenic properties. Within the family Flaviviridae, the ICTV has classified hepatitis C virus (HCV) as the type species within the genus Hepacivirus, and GB virus B (GBV-B) has tentatively been assigned as a second species within this genus (http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm). The related GBV-A and GBV-A-like agents, and GBV-C (or hepatitis G virus; HGV) viruses have also been assigned to the family Flaviviridae, but not to a genus (http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm). A related virus was recently discovered in Old World frugivorous bats (Pteropus giganteus) and was termed GBV-D (Epstein et al., 2010). In this review, we propose to assign GBV-A, GBV-C/HGV and GBV-D as species within a new genus, Pegivirus. In addition, we suggest that the GB viruses within this fourth genus of the family Flaviviridae be renamed to reflect better their biological and pathogenic properties.

History

Following the discovery of hepatitis A virus (HAV) and hepatitis B virus (HBV) in the 1960s and 1970s, it became
clear that neither virus was associated with a mild form of chronic hepatitis frequently observed in recipients of blood transfusions (Feinestone et al., 1975). Considerable research was directed towards identifying the causative agent of post-transfusion or ‘non-A, non-B’ hepatitis (Feinestone et al., 1975; Feinestone & Purcell, 1978; Prince et al., 1974), and well characterized human sera were shown to contain an infectious agent that caused chronic, relapsing and mild hepatitis in experimentally infected chimpanzees (Bradley et al., 1983). Using a molecular cloning approach, Choo et al. (1989) discovered an RNA virus in the serum and tissues of a chimpanzee experimentally inoculated with serum from an individual with chronic, non-A, non-B hepatitis (Choo et al., 1989). This virus was shown to be epidemiologically associated with non-A, non-B hepatitis (Kuo et al., 1989), and chimpanzee studies confirmed that the virus induced hepatitis (Bradley, 2000). The virus was called hepatitis C virus (HCV) and it is classified as the type species member of the genus Hepacivirus within the family Flaviviridae (Choo et al., 1989).

In the process of studying non-A, non-B hepatitis, Dienhardt and colleagues obtained serum from a surgeon on day 3 of acute hepatitis (Deinhardt et al., 1967). This serum apparently induced hepatitis when inoculated into tamarins, a type of New World monkey (Saguinus labiatus). Passage of serum obtained from the inoculated animals at the time of hepatitis into new tamarins produced similar hepatitis both in newly inoculated tamarins and in other New World monkey species (Deinhardt et al., 1967). Based on the initials of the surgeon, the transmissible agent was called the ‘GB agent’, and this agent was studied extensively as a putative cause of non-A, non-B hepatitis (Deinhardt et al., 1967; Deinhardt & Deinhardt, 1984; Gust & Feinestone, 1988). In 1995, a group of investigators from Abbott Laboratories identified two viruses in the serum and liver of tamarins inoculated with the 11th tamarin passage of the GB agent (Simons et al., 1995b). These viruses were named GB virus A and B (GBV-A and GBV-B) due to the pedigree of the infectious serum (Simons et al., 1995b). Using degenerate primers to amplify related viral sequences in human serum samples, a third virus was identified and termed GBV-C (Simons et al., 1995a). Simultaneously, a research group at Genelabs identified novel RNA virus sequences in the serum of humans with non-A, non-B hepatitis, and called this virus hepatitis G virus (HGV) (Linnen et al., 1996).

Subsequent analysis of the genome sequences of HGV and GBV-C revealed that they were minor variants of the same virus species, while GBV-A and GBV-B were distinct (Kim & Fry, 1997; Leary et al., 1996b; Muerhoff et al., 1995). All of the ‘GB’ viruses are distantly related to HCV (Linnen et al., 1996; Simons et al., 1995a, b) and based on their predicted genome structure and nucleotide sequence relationships, the three ‘GB’ viruses were classified as members of the family Flaviviridae (http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm). Phylogenetic analysis of conserved regions of translated sequences from the helicase and polymerase domains revealed a closer relationship of HCV to GBV-B, while GBV-A and GBV-C/HGV form a separate cluster (Muerhoff et al., 1995; Simons et al., 1995a). GBV-B represented the true GB-agent. Although it apparently did not originate from the surgeon GB and does not infect humans or chimpanzees, it caused acute hepatitis in experimentally infected tamarins, including animals transfected intrahepatically with RNA transcripts of recombinant GBV-B (Bukh & Apgar, 1999; Lanford et al., 2003; Martin et al., 2003; Nam et al., 2004). In contrast, GBV-A represented indigenous tamarin viruses not associated with hepatitis (Simons et al., 1995b), and a number of GBV-A-like agents were subsequently identified from New World monkeys (Bukh & Apgar, 1997; Leary et al., 1996a; Simons et al., 1995b). GBV-C was found to be a frequent human virus that was not associated with viral hepatitis (Alter, 1997; Alter et al., 1997a; Simons et al., 1995a).

Recently, using unbiased, high-throughput pyrosequencing methods, a virus more distantly related to GBV-A, GBV-B and GBV-C/HGV was identified in serum samples obtained from Old World frugivorous bats (P. giganteus) in Bangladesh (Epstein et al., 2010). Two full-length sequences were generated that share approximately 50% amino acid sequence identity with GBV-A and GBV-C/HGV (Epstein et al., 2010). This novel virus was named GBV-D.

For three reasons, we suggest that the current nomenclature assigned to the ‘GB’ viruses should be changed. Firstly, there is no evidence that the surgeon (GB) for whom these viruses are named was infected with GBV-A, GBV-B, GBV-C/HGV or GBV-D. Secondly, infections and susceptibility to GBV-A and GBV-B have subsequently been shown to be restricted to New World primates, and to date, GBV-D appears to be restricted to bats. Thus, it is highly unlikely that they originated from the surgeon ‘GB’. Finally, GBV-C/HGV does not cause hepatitis in humans. In this manuscript, we review the clinical and virological aspects of these viruses and re-examine their genetic relationships. Based on these observations we propose a change in nomenclature of GBV-A, GBV-C/HGV and GBV-D to better and more clearly describe these viruses. Since GBV-B represents what for years has been referred to as the GB-agent, and since the natural host(s) of this virus remain unknown, we at present propose to remove the ‘B’ designation and call the virus ‘GB’, as it will be the only GB virus under the proposed classification system.

**Virology**

**Epidemiology and transmission**

HCV and GBV-C/HGV infections occur worldwide (reviewed by Lauer & Walker, 2001; Stapleton, 2003). It has been estimated that perhaps as many as 3% of the world’s population has been infected with HCV (The Global Burden of Disease Working Group, 2004). Frequencies of GBV-C/HGV infection are difficult to determine, but prevalence studies suggest that 1–4% of
healthy blood donors in most developed countries are viraemic at the time of blood donation, and another 5–13% have anti-E2 antibodies, indicating prior infection (Blair et al., 1998; Gutierrez et al., 1997; Pilot-Matias et al., 1996a; Tacke et al., 1997). In developing countries, blood donor viraemia prevalence is higher, approaching 20% in some regions of the world (reviewed by Mohr & Stapleton, 2009; Polgreen et al., 2003).

Among people with blood-borne or sexually transmitted infections, GBV-C/HGV is more prevalent (Scallan et al., 1994), and in one study of human immunodeficiency virus (HIV)-infected homosexual men, 39.6% had viraemia and 46% had E2 antibody detected for a total exposure rate of 85.6% (Williams et al., 2004). These data, in combination with the blood donor studies, suggest that at least one quarter of the world’s population has been infected with GBV-C/HGV. In contrast, sexual transmission of HCV is inefficient, and most transmission occurs through exposure to blood or from mother to child during birth (Lauer & Walker, 2001).

Of the non-human GB viruses, GBV-A and GBV-B can be experimentally transmitted to different species of New World monkeys via the blood-borne route. It is not clear if sexual, vertical or other modes of transmission occur for these two viruses. Following identification, GBV-D RNA was detected using real-time PCR methods in 5 of 98 (5%) P. giganteus (bat) serum samples in a population of wild bats in Bangladesh. No further epidemiological studies have been reported. Although there are no data to indicate the mode of GBV-D transmission, viral RNA was identified in the salivary gland of one of the five bats with viraemia, suggesting horizontal and potentially zoonotic transmission, and none of the viremic bats had GBV-D RNA detected in urine (Epstein et al., 2010).

Pathogenesis and cellular tropism

HCV and GBV-B viruses are primarily detected in the liver of naturally infected humans and experimentally infected New World monkeys, respectively, although viral genomes can be found in peripheral blood mononuclear cells (PBMCs) in some infected hosts (Beames et al., 2000; Bright et al., 2004; Bukh et al., 2001a; Fong et al., 1991; Ishii et al., 2007; Jacob et al., 2004; Laskus et al., 1997b; Simons et al., 1995b). GBV-B has not been identified in New World primates except in experimentally infected animals. High levels of virus are present in the blood, typically between 1–10 million genome equivalents ml⁻¹ for HCV and 10–100 million genome equivalents ml⁻¹ for GBV-B.

GBV-A and GBV-A-like agents, and GBV-C/HGV viruses are present in low or non-detectable levels in the liver of infected hosts, and the viruses are more readily detected in circulating lymphocytes, suggesting that GBV-A and GBV-C could be lymphotropic, and not hepatotropic (Kobayashi et al., 1999; Laskus et al., 1997a, 1998; Pessoa et al., 1998; Radkowski et al., 1999, 2000; Simons et al., 2000; Tucker et al., 2000). Although replication of HCV and GBV-C/HGV in hepatocyte and lymphocyte cell culture has been described, HCV and GBV-B replication is optimal in cultured cells of hepatocyte origin (Beames et al., 2000; Lanford et al., 1994; Lindenbach et al., 2005a; Waki et al., 2005; Zhong et al., 2005). In contrast, GBV-C/HGV replication is most frequently accomplished in PBMCs, including the CD4 and CD8 T lymphocyte subsets, and B lymphocytes (Fogeda et al., 1999; George et al., 2003, 2006; Xiang et al., 2000). Cell culture replication of GBV-A or GBV-D has not been described. Serum concentrations of GBV-D RNA ranged from 350 to 70 000 genome copies ml⁻¹ (Epstein et al., 2010), but bat liver, PBMCs or other cell types have not been assessed for evidence of viral replication.

Viral RNA transcripts and/or serum are infectious in animal models for HCV and GBV-B (Beames et al., 2001; Bukh et al., 1999, 2001a; Jacob et al., 2004; Kolykhalov et al., 1997; Lanford et al., 2001, 2003; Martin et al., 2003; Nam et al., 2004; Shabderali et al., 2001; Yanagi et al., 1997), and for cell culture models for HCV, GBV-B and GBV-C/HGV (Beames et al., 2000; Lindenbach et al., 2005a, 2006; Waki et al., 2005; Xiang et al., 2000; Zhong et al., 2005). Consistent with the tropism of these viruses, both HCV and GBV-B are associated with hepatitis (Bukh et al., 2001a; Hoofnagle, 1997; Simons et al., 1995b), but GBV-A and GBV-C/HGV are not associated with hepatitis in clinical and experimental studies (Alter et al., 1997a, b; Schlauder et al., 1995; Simons et al., 1995b). No experimental transmission studies of GBV-D have been reported; however, no difference in liver enzyme values in serum was identified in bats with GBV-D viraemia (Epstein et al., 2010).

Persistence and humoral immunity

HCV and GBV-A infection frequently leads to persistent viraemia, with approximately 80% of HCV infections and all GBV-A infections studied longitudinally, resulting in life-long infection (Hoofnagle, 1997; Lauer & Walker, 2001; Simons et al., 1995b). Neutralizing antibodies are detectable against HCV in at least 95% of persistently vireamic individuals (Lauer & Walker, 2001; Owsianka et al., 2008). Their role in chronic HCV hepatitis remains poorly defined. Although antibodies to GBV-A were not identified during acute or chronic infection, this may relate to the paucity of reagents available for GBV-A (Simons et al., 1995b).

GBV-B is usually cleared by the host 1–6 months after experimental infection of tamarins (Bukh et al., 2001a; Jacob et al., 2004; Simons et al., 1995b), and no persistent infection has been observed in animals infected with virus particles. However, infection was present at the time of sacrifice in two animals (90 weeks and 2 years, respectively) that were injected intrahepatically with full-length, synthetically transcribed GBV-B RNA (Martin et al., 2003; Nam et al., 2004). Further passage of GBV-B derived from
tamarins infected by intrahepatic injection resulted in self-limited infection, suggesting that the observed persistence was related to host genetic factors rather than a property of the specific GBV-B isolate (Jacob et al., 2004).

The majority of immune competent individuals infected with GBV-C/HGV clear viraemia within 2 years of infection (Berg et al., 1999; Tanaka et al., 1998). Unlike HCV, which elicits antibodies to several viral proteins during viraemia that usually persist throughout infection (Bauert et al., 2000), GBV-C/HGV antibodies are not generally detected during viraemia, although some studies have reported the detection of anti-GBV-C/HGV peptide reactivity (Fernandez-Vidal et al., 2007; Gomara et al., 2010; Pilot-Matias et al., 1996b; Schwarze-Zander et al., 2006; Tan et al., 1999; Van der Bij et al., 2005; Xiang et al., 1998). Following clearance of GBV-C/HGV viraemia, most individuals develop conformation-dependent antibodies to the envelope glycoprotein E2, and thus E2 antibody serves as a marker of prior infection (Barnes et al., 2007; Gutierrez et al., 1997; McLinden et al., 2006; Nakatsuji et al., 1992; Pilot-Matias et al., 1996a; Tacke et al., 1997; Tanaka et al., 1998). Detection of anti-GBV-C/ HGV antibodies occurs coincidently with clearance of viraemia and appears to be restricted to E2, suggesting that this E2 antigenic site is immunodominant in humans (McLinden et al., 2006). In addition, HCV and GBV-C/ HGV particles contain high concentrations of lipids, and as a result have very low buoyant densities (<1.10 g cm⁻³) (Agnello et al., 1999; Hijiikata et al., 1993; Melvin et al., 1998; Monazahian et al., 1999, 2000; Thomssen et al., 1992, 1993; Wunschmann et al., 2000, 2006; Xiang et al., 1998). It is possible that these virus-associated lipids mask the HCV and GBV-C/HGV neutralization epitopes and contribute to viral persistence. However, this does not explain the failure of humans to develop antibodies to non-structural (NS) proteins in GBV-C/HGV infections, and suggests that the virus interacts with the humoral immune response.

By comparing the number of healthy blood donors with either GBV-C/HGV viraemia or E2 antibody, it appears that approximately 80% of healthy people spontaneously clear viraemia (Gutierrez et al., 1997; Nakatsuji et al., 1992; Pilot-Matias et al., 1996a; Tacke et al., 1997; Tanaka et al., 1998). Among HIV-infected subjects, the frequency of GBV-C/HGV clearance appears to be reduced, as the prevalence of viraemia is increased, while E2 antibody prevalence is generally the same or higher than in HIV-uninfected individuals (Dorrucci et al., 1995; Heringlake et al., 1998; Williams et al., 2004). Thus, the proportion of individuals with viraemia compared with those with E2 antibody is increased. GBV-C/HGV viraemia has been documented to persist for decades (Alter, 1997; Barnes et al., 2007), and by analogy with other viruses such as HBV, it is possible that higher frequencies of persistence may occur in individuals exposed very early in life. GBV-C/ HGV infection of chimpanzees (GBV-C/ap or GBV-C/trm) may also persist in infected animals throughout 19 years of follow-up. However, like GBV-C/HGV in humans, it appears that the majority of infections in chimpanzees are self limited (Mohr et al., 2010).

Although HCV antibodies frequently persist in those who clear viraemia, HCV antibody titres wane over time, and longitudinal studies have revealed a substantial proportion of HCV-infected individuals without residual serological markers of infection (Takaki et al., 2000). Antibodies may prevent challenge with autologous virus (Farcì et al., 1994; Tabor et al., 1980), but they do not prevent superinfection (Farcì et al., 1992a). Nevertheless, neutralizing antibodies directed against conserved HCV epitopes have been identified, and are actively being studied for immunotherapy or as potential vaccine immunogens (Keck et al., 2004).

To date, no antibodies to GBV-A have been detected, suggesting that GBV-A also somehow evades host recognition. However, there is a paucity of studies and reagents available for GBV-A, so humoral immunity in GBV-A cannot be conclusively described (Schaluder et al., 1995). In contrast, GBV-B antibodies are present following viraemia clearance and were thought to prevent or attenuate experimental infection in marmosets (Schaluder et al., 1995). However, although a subsequent study found evidence of protection following GBV-B infection, this did not appear to involve humoral immunity (Bukh et al., 2008). Antibodies to GBV-B, HCV and GBV-C/HGV frequently decline following clearance of viraemia, sometimes below the limit of detection. Thus, antibody detection may underestimate the prevalence of prior infection. Information regarding GBV-D persistence or serological responses has not yet been described.

**Host range**

Despite the similarities in genome organization and existence of homologous proteins, specific host range differences exist among HCV, GBV-A, GBV-B, GBV-C/ HGV and GBV-D. HCV and GBV-C/HGV infect Old World primates, while GBV-A and GBV-B infect New World primates. Specifically, natural HCV infection is limited to humans, although experimental infection of chimpanzees is well documented (Farcí et al., 1992b; Shimizu et al., 1990; Tabor et al., 1979). Natural infection of humans and chimpanzees with GBV-C/HGV is well documented (Adams et al., 1998; Birkenmeyer et al., 1998; Linnen et al., 1996; Simons et al., 1995a), and sequences of isolates obtained from chimpanzee (GBV-C/ap or GBV- C/trm) form a separate phylogenetic group from human GBV-C/HGV (Adams et al., 1998; Birkenmeyer et al., 1998). The host range for experimental infection with HCV and human GBV-C/HGV infection appears to be restricted to humans and chimpanzees (Bukh et al., 1998, 1999, 2001b, 2008), although small studies suggest that HCV and GBV-C/ HGV may infect some Old World monkeys (Macaca) (Cheng et al., 2000; Krawczynski, 1997; Majerowicz et al., 2004; Ren et al., 2005; Vitral et al., 1997). This is controversial, as neither HCV nor GBV-C/HGV infection of macaques was reproduced by other laboratories (Bukh et al., 2008), although limited infection has been reported (Adams et al., 1998; Birkenmeyer et al., 1998).
2001b and personal communication; Steven Feinstone, David Thomas, Jian-Qiu Han, Binhua Ling, Jinhua Xiang and Jack Stapleton). Thus, it is not clear if organisms other than humans and chimpanzees are susceptible to HCV and GBV-C/HGV infection.

In contrast, the natural hosts of GBV-A and GBV-A-like variants include at least six species of New World monkeys including *Saguinus* species (*Saguinus labiatus, Saguinus mystax, Saguinus nigricollis and Saguinus oedipus*), *Callithrix* species (*Callithrix jacchus*) and *Aotus* species (*Aotus trivirgatus*) (Bukh & Apgar, 1997; Leary et al., 1996a; Muerhoff et al., 1995; Simons et al., 1995b). The 5’ non-translated region (NTR) and NS3 helicase sequences of GBV-A isolates from different host species were found to segregate into distinct groups, suggesting co-speciation of GBV-A isolates from different host species were found to segregate into distinct groups, suggesting co-speciation of GBV-A isolates from different host species. GBV-B infection of tamarins and aotus monkeys have been documented, but experimental inoculation into chimpanzees did not provide evidence of viral replication (Bukh et al., 2001a). Identification of the natural host of GBV-B has been identified. Experimental GBV-B infection of tamarins and aotus monkeys have been documented, but experimental inoculation into chimpanzees did not provide evidence of viral replication (Bukh et al., 2001a). Identification of the natural host of GBV-B has been identified. Experimental GBV-B infection of tamarins and aotus monkeys have been documented, but experimental inoculation into chimpanzees did not provide evidence of viral replication (Bukh et al., 2001a). Identification of the natural host of GBV-B may be complicated by the short duration of viremia and the lack of a reliable serological method to detect prior infection (Pilot-Matias et al., 1996b; Schaluder et al., 1995). Alternatively, tamarins may not be the natural host of GBV-B, and the virus may indeed be capable of establishing persistent infections in an alternative, natural host. GBV-D has only been reported in one bat species (Epstein et al., 2010). A summary of the host range, tropism and pathogenesis is presented in Table 1.

### Genome organization

Like other members of the family *Flaviviridae*, HCV, GBV-A, GBV-B, GBV-C/HGV and GBV-D have a positive-sense, ssRNA genome that contains a single long ORF encoding a multifunctional polyprotein (Fig. 1) (Birkenmeyer et al., 1998; Choo et al., 1991; Kim & Fry, 1997; Leary et al., 1996b; Muerhoff et al., 1995). The 5’ NTR contains an internal ribosomal entry site (IRES) element that directs translation of the polyprotein directly from viral genomic RNA (Simons et al., 1996; Yoo et al., 1992), and translation and release of the viral RNA-dependent RNA polymerase (RdRp) initiates RNA replication (reviewed by Major & Feinstone, 1997; Mohr & Stapleton, 2009; Moradpour et al., 2007; Robertson, 2001). The viral structural proteins are processed from the amino-terminal portion of the polyprotein by cellular signal peptidases or signal peptide peptidase, while the NS proteins are processed by two viral encoded proteases. An NS2–NS3 autoprotease catalyses the cleavage of NS2–NS3 for HCV, and based on the conservation of two essential amino acids for HCV NS2 function (His-952 and Cys-993), NS2-mediated cleavage of NS2–NS3 is thought to occur in GBV-A, GBV-B, GBV-C/HGV and GBV-D viruses as well. The remaining NS proteins are cleaved by the NS3 protease with its co-factor NS4A in HCV, GBV-B, GBV-C/HGV and presumably GBV-A and GBV-D (reviewed by Epstein et al., 2010; Major & Feinstone, 1997; Mohr & Stapleton, 2009; Moradpour et al., 2007; Robertson, 2001). Like NS2, NS3 amino acids essential for the processing of HCV NS proteins (His-1083, Asp-1107 and Ser-1165) are present in all four GB viruses and HCV. Their NS3 proteases are classified as chymotrypsin-like serine proteases (reviewed by Major & Feinstone, 1997; Mohr & Stapleton, 2009; Moradpour et al., 2007; Robertson, 2001).

Although HCV and the four GB viruses have somewhat similar genome organization and predicted protein structure, there are distinct differences. All of the GB viruses studied have an IRES element in the 5’ NTR, although their structures differ, with GBV-B having a type 3 IRES, like HCV, while the IRES element in GBV-A and GBV-C/HGV conform better with type 4 IRES elements (Kieft, 2008). However, others state that the GBV-A and GBV-C/HGV IRES do not conform to any recognized IRES class (Bakhshesh et al., 2008). IRES activity has not been examined in GBV-D. Direct comparative data do not exist

### Table 1. Comparison of tropism, host range and pathogenesis of GB viruses and Hepaciviruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>Proposal*</th>
<th>Tropism</th>
<th>Host range</th>
<th>Pathogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegivirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C</td>
<td>HPgV</td>
<td>Lymphocytes†</td>
<td>Humans, chimpanzee</td>
<td>None identified</td>
</tr>
<tr>
<td>GBV-C&lt;sub&gt;cpz&lt;/sub&gt;</td>
<td>SPgV&lt;sub&gt;cpz&lt;/sub&gt;</td>
<td>IDA</td>
<td>Chimpanzee</td>
<td>None identified</td>
</tr>
<tr>
<td>GBV-A</td>
<td>SPgV</td>
<td>Lymphocytes†</td>
<td>NWP</td>
<td>None identified</td>
</tr>
<tr>
<td>GBV-D</td>
<td>BPgV</td>
<td>IDA</td>
<td>Bats</td>
<td>IDA</td>
</tr>
<tr>
<td>Hepivirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCV</td>
<td>HCV</td>
<td>Hepatocytes</td>
<td>Humans, chimpanzee</td>
<td>Hepatitis</td>
</tr>
<tr>
<td>GBV-B</td>
<td>GBV</td>
<td>Hepatocytes</td>
<td>NWP</td>
<td>Hepatitis</td>
</tr>
</tbody>
</table>

*Genus (italics) and virus nomenclature proposed in this review.
†Data support lymphotropism, but other sites of replication are possible.
on the relative translational efficiency of the IRES elements for GB viruses. The 5' NTRs of HCV isolates are approximately 340 nt long, whereas the 5' NTR of GBV-B is 445 nt, and the contained IRES directs the translation of a core protein. In contrast, the 5' NTRs for GBV-A and GBV-C/HGV are predicted to be longer based on in vitro translation studies that demonstrated that the AUG at position 556 of GBV-C/HGV was the codon that initiated translation (Simons et al., 1996). Sequence numbering is based on the infectious clone isolate (Xiang et al., 2000) (GenBank accession no. AF121950). Thus, GBV-A and some GBV-C/HGV isolates do not appear to encode a core protein (Kim & Fry, 1997; Leary et al., 1996b; Xiang et al., 1998). A signal peptidase cleavage site is predicted to occur 17 or 21 aa downstream of the putative initiation codon in GBV-C/HGV (Mohr & Stapleton, 2009), although it is doubtful that this small peptide could serve as the core (nucleocapsid) protein. Biophysical characterization of GBV-C/HGV particles, however, found that they appear to have a nucleocapsid (Xiang et al., 1998). Although there is limited experimental evidence, several potential hypotheses have been put forward to explain the potential source of the nucleocapsid protein. These include the possibility that the capsid forms from the very small cleaved peptide at the N terminus of the polyprotein (Xiang et al., 1998), or that a longer core protein is translated off an alternative reading frame on the genomic or negative strands of the GBV-C/HGV genome. Alternatively, the hypothesis that the virus utilizes a cellular protein to serve as the nucleocapsid protein has been raised (Theodore & Lemon, 1997).

It is unclear which AUG codon initiates translation in GBV-D. Like GBV-C, there are multiple potential initiation codons in-frame with the GBV-D coding sequence. Specifically, there are five AUG codons between the 5' end of the GBV-D genome and nt 744 that are in-frame with the long ORF (GenBank accession nos GU566734 and GU566735). Of note, the predicted amino acid sequence of GBV-D starting at nt 744 is MAVLLLLSTGLAEG. The GBV-C predicted amino acid sequence starting at the AUG shown to initiate translation in vitro (Simons et al., 1996) is MAVLLLLLVVEAGA, thus sharing complete identity with the first seven amino acids of GBV-D. GBV-A and

---

**Fig. 1.** Genome organization of the GB viruses and HCV. All four viruses contain a positive-polarity ssRNA genome with a 5' NTR and 3' NTR. The genome encodes a polyprotein that is co- and post-translationally cleaved into individual viral proteins. The structural proteins include core (C) and envelope glycoproteins (E1 and E2), and the NS proteins include NS2–NS5B. The presence of a genomic coding region for a C protein has not been identified for GBV-A or GBV-C. Structural proteins are cleaved by cellular signal peptidases (open arrows), and the NS2–NS3 cleavage is accomplished by the NS2–NS3 autoprotease (shaded arrows). The remaining NS proteins are cleaved by the NS3–NS4A protease complex (solid block arrows). The predicted genome organization of GBV-D was based on a polyprotein starting nt 18 of GU566735. *, The predicted sizes of the proteins analogous to the HCV p7 are 21 kDa for GBV-A and 6 kDa for GBV-C. The existence of a GBV-D p7-like protein is not clear from sequence analysis. **, The size of the protein corresponding to the HCV p7 in GBV-B is 13 kDa; this protein could be cleaved into p7 and p6 proteins, of which the p7 protein, but not the p6 protein is critical for viability in vivo. Proposed names for GB viruses are in parentheses.
GBV-A-like viruses share sequence homology in this potential polyprotein initiation region as well. The GBV-A<sub>trn</sub> sequence is MEVLLVLLKTAHALGA, GBV-A<sub>lab</sub> sequence is MELLLLLVLAPAGA and the GBV-A sequence is MASL-WVFVHLLPLGGGG. Until the GBV-D translation initiation codon is identified, it will be difficult to assign precise predicted structural protein sizes. Analysis of the GBV-D polyprotein starting at the first AUG demonstrates four predicted signal peptidase sites in the polyprotein using the AUG at nt 57 as the translation start (Epstein et al., 2010). These are located at amino acid numbers 57–58, 247–248, 584–585 and 826–827. A 57 aa residue long protein (6 kDa) that is highly basic (pI 12) was proposed as the nucleocapsid or core protein for GBV-D (Epstein et al., 2010), the genome would then have to have an extremely short 5′ NTR (57 nt), unless the true 5′ end sequence was not identified. Further experimental work is required to determine the genome organization of this newly described virus.

Additional differences between the various GB viruses and HCV occur in the extent of predicted glycosylation of the two envelope proteins (E1 and E2). HCV is the most heavily glycosylated, followed by GBV-B, GBV-A and GBV-C/HGV (reviewed by Mohr & Stapleton, 2009). GBV-D is predicted to have 13 glycosylation sites, which would place it similar to GBV-B and HCV. There may be an additional glycoprotein between E2 and NS2 of GBV-D. This region of the polyprotein was called the ‘X’ protein by the group that discovered the virus (Epstein et al., 2010). HCV and GBV-B has a p7 and a p13 protein, respectively, between E2 and NS2 that is essential for their viability (Sakai et al., 2003; Takikawa et al., 2006). The HCV p7 protein is believed to be important for virus assembly and release (Sakai et al., 2003; Lindenbach & Rice, 2005b). It is not known whether GBV-A and GBV-C have a corresponding protein. The 3′ NTR of HCV and GBV-B contain poly-U tracts, while GBV-A, GBV-C and GBV-D do not (Birkenmeyer et al., 1998; Kim & Fry, 1997; Leary et al., 1996b; Muerhoff et al., 1995). In addition, HCV and GBV-B have highly structured 3′ terminal sequences (Lindenbach & Rice, 2005b). Finally, a number of differences occur in the predicted size of the NS5 proteins; however, with the exception of HCV and GBV-B, the experimental evidence to demonstrate differences is lacking. A summary of genome organizational features that differ among the GB viruses and HCV is shown in Table 2.

### Phylogenetic relationships

Unweighted pair group method analysis (UPGMA) and neighbour-joining (NJ) analyses are common methods used to generate a single tree, which can be a starting point for evolutionary analysis. Alternatively, a ‘character-based’ approach evaluates the relatedness of sequences based on a subset of positions called ‘informative sites’. Weighted and unweighted Parsimony methods and maximum-likelihood are frequently used approaches that generate multiple trees (cladograms), which can be evaluated for accuracy (Smith et al., 2000). Repeat analyses of a group of sequences in which a proportion of sites are randomly resampled and used for phylogenetic analysis is called ‘bootstrapping’, which estimates the frequency that a particular branch (node) in the tree occurs for each sample. Support for a grouping is considered present when a branch occurs in more than 70% of 1000 replicates (Smith et al., 2000).

Analysis of conserved amino acid sequence motifs involved in the enzymic function of HCV, and GBV-A, GBV-B, GBV-C and GBV-D may elucidate their evolutionary relationship to each other and other members of the family Flaviviridae (Bukh & Apgar, 1997; Muerhoff et al., 1995; Robertson, 2001; Sathar et al., 1999; Schlauder et al., 1995; Smith et al., 2000). The helicase region within NS3 share amino acid sequence identity in six domains (Supplementary Fig. S1, available in JGV Online), while the RdRp protein contains eight conserved motifs (Supplementary Fig. S2, available in JGV Online) (Adams

### Table 2. Genome features of GB viruses and HCV

<table>
<thead>
<tr>
<th>Virus</th>
<th>Proposal*</th>
<th>5′ NTR</th>
<th>Core protein</th>
<th>Envelope protein glycosylation</th>
<th>3′ NTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBV-A</td>
<td>SPgV</td>
<td>~540 nt</td>
<td>Not identified</td>
<td>3–7</td>
<td>No poly U</td>
</tr>
<tr>
<td>GBV-B</td>
<td>GBV</td>
<td>445 nt, type 3 IRES</td>
<td>156 aa</td>
<td>10</td>
<td>Poly U</td>
</tr>
<tr>
<td>GBV-C</td>
<td>HPgV</td>
<td>555 nt, NC</td>
<td>Not identified</td>
<td>4†</td>
<td>No poly U</td>
</tr>
<tr>
<td>GBV-C&lt;sub&gt;cpz&lt;/sub&gt;</td>
<td>SPgV&lt;sub&gt;cpz&lt;/sub&gt;</td>
<td>~550 nt, NC</td>
<td>Not identified</td>
<td>3</td>
<td>No poly U</td>
</tr>
<tr>
<td>GBV-D</td>
<td>BPgV</td>
<td>Not identified</td>
<td>Not identified</td>
<td>13†</td>
<td>No poly U</td>
</tr>
<tr>
<td>HCV</td>
<td>HCV</td>
<td>~340 nt, type 3 IRES</td>
<td>191 aa</td>
<td>14–16</td>
<td>Poly U-UC</td>
</tr>
</tbody>
</table>

* Nomenclature proposed in this review.
† Predicted.
et al., 1998; Birkenmeyer et al., 1998; Bukh & Apparag, 1997; Gorbalenya & Koonin, 1989; Koonin, 1991; Koonin & Dolja, 1993; Leary et al., 1996b; Linnen et al., 1996; Muerhoff et al., 1995; Robertson, 2001; Simons et al., 1995a; Smith et al., 2000). Alignments were performed by hand, based on the alignments of Koonin et al. that were used as a guide (Gorbalenya & Koonin, 1989; Koonin & Dolja, 1993). Helicase sequences of members of the family Flaviviridae fall within the helicase supergroup II, and their RdRp sequences place them into the RdRp supergroup II. UPGMA analysis of the six conserved helicase domains and eight conserved RdRp motifs with their intervening sequences of HCV, GBV-A, GBV-B, GBV-C and GBV-D confirm that these viruses are related to the Pestivirus and Flavivirus genera within the family Flaviviridae (Fig. 2). GBV-A, GBV-C and GBV-D group together, and GBV-B groups with HCV, consistent with the proposed assignment as a second species within the genus Hepacivirus. NJ analysis of the helicase and RdRp sequences including the intervening sequences between the conserved domains identified similar genetic relationships (Fig. 3), as did tree methods using CLUSTAL alignments of these sequences (data not shown). Despite the agreement between trees constructed by the two methods used here, final assignment into specific genera should also consider genome structure, tissue tropism and pathogenesis.
Proposal to classify ‘GB’ viruses

We propose that GBV-A, GBV-A-like agents, GBV-C/HGV and GBV-D should be classified together within a new genus (Pegivirus) within the family Flaviviridae. This is based on their phylogenetic relationships, genome structure, ability to persist in vivo and apparent lack of pathogenicity. This designation indicates that these viruses cause persistent infection (Pe), and provides historical recognition of the relationship with the ‘GB’ agents and hepatitis G virus (g). We propose to rename GBV-A as simian pegivirus (SPgV). This indicates the primate host range (S) and through the genus designation provides historical reference to their relationship to the ‘GB’ serum inoculation (Pg). The host species for different GBV-A variants will be identified by subscript suffixes. For example, SPgV will include SPgVmys, SPgVtr, SPgVlab and SPgVjac. We propose to rename GBV-C/HGV as human pegivirus (HPgV) for similar reasons, and genotypes will be identified by subscript suffixes. The chimpanzee GBV-C variants would be called SPgVcpz to reflect its simian host. The recently reported GBV-D virus found in fruit bats would be described as Bat PgV (BPgV). Should related BPgV isolates be identified in different species, the host species of the current bat isolate will be designated by the subscript suffix (pgi) for the current virus identified in the species P. giganteus, and the new related viruses will be identified by their host species. This nomenclature avoids the suggestion that these viruses either cause hepatitis or were derived from the surgeon ‘GB’.

We support the proposal to classify GBV-B as a second species within the genus Hepacivirus, as this virus causes hepatitis in experimentally infected tamarins, thus it was the true GB-agent (Thiel et al., 2005). Although the surgeon GB would not have been infected with this virus, it is the agent responsible for the hepatitis observed in tamarins used in serial passage studies. Since, under the proposed nomenclature, there will not be any other ‘GB’ viruses, we propose to rename GBV-B as GB virus (GBV). Although the ‘GB’ serum was associated with hepatitis following inoculation into tamarins, the specific agent(s) responsible for the hepatitis remain unknown and the reported hepatitis may simply relate to the relatively common finding of non-specific enzyme elevation. A summary of the classification proposal is shown in Fig. 4.

Our proposal to rename GBV-A, GBV-C/HGV and GBV-D viruses and to classify them within a new genus (Pegivirus) will be submitted to the International Committee on...
Family Flaviridae
Genus Flavivirus
Representative species:
- Yellow fever virus
- West Nile virus
- Dengue virus
- Japanese encephalitis virus

Genus Pestivirus
Representative species:
- Bovine viral diarrhea virus

Genus Hepacivirus
Known species:
- HCV (seven genotypes)
- GBV (GBV-B)

Genus (new): Pegivirus
Known species:
- SPgV (GBV-A) (mys, lab, tri, jac)
- HPGV (GBV-C) (six genotypes)
- SPgVcpz (GBV-Ccpz)
- BPGV (GBV-D)

Acknowledgements
This work was supported in part by a Merit Review grant from the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development (J. T. S.), and a grant from the National Institutes of Health (RO1 AI-58740, J. T. S.). Jens Bukh is the recipient of a professorship at the University of Copenhagen with external funding from the Lundbeck Foundation.

References


The GB viruses


http://vir.sgmjournals.org


