Published in final edited form as:

J Med Virol. 2011 April ; 83(4): 685–694. doi:10.1002/jmv.22029.

Evidence for extensive genotypic diversity and recombination of GB virus C (GBV-C) in Germany

Markus Neibecker^{a,*}, Carolynne Schwarze-Zander^{a,*}, Jürgen K. Rockstroh^a, Ulrich Spengler^a, and Jason T. Blackard^b

^a Department of Internal Medicine I, University Hospital Bonn, Bonn, Germany

^b Division of Digestive Diseases, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Abstract

Multiple genotypes of GB virus C (GBV-C) – a non-pathogenic flavivirus – have been identified to date, although they are not uniformly distributed worldwide. It has also been suggested that GBV-C genotype may play a role in modulating HIV disease; however, the prevalence and genotype distribution of GBV-C has not been adequately studied in most countries. Among 408 HIV positive subjects in Germany, 97 (23.8%) had detectable GBV-C RNA. Based on sequencing of the 5' untranslated region (5'UTR), the GBV-C genotypes were 1 (n = 8; 8.2%), 2 (n = 81; 83.5%), and 3 (n = 2; 2.1%), as well as a unique genotype not previously reported (n = 6; 6.2%). Among 17 samples also sequenced in the envelope 2 (E2) region, 14 had concordant genotype results when comparing the 5'UTR and E2, while evidence of intergenotypic recombination was observed among E2 sequences from 3 individuals. These results suggest that genotypic diversity and viral recombination contribute to the overall genetic variability of GB virus C.

Keywords

GB virus C (GBV-C) HIV; diversity; genotype; recombination

Introduction

GB virus C (GBV-C) is a member of the flaviviridae family of positive-sense, single-stranded RNA viruses. The GBV-C genome consists of a 5' untranslated region (5'UTR) that contains an internal ribosomal entry site and is followed by a single open reading frame that encodes for a polyprotein of approximately 3,000 amino acids (reviewed in [Mohr EL and Stapleton, 2009]). This polyprotein encodes for at least two structural proteins – Envelope protein 1 (E1) and Envelope protein 2 (E2) – as well as multiple non-structural proteins – NS2 to NS5. To date, six major GBV-C genotypes have been identified based on sequence analysis of the 5' untranslated region [Muerhoff AS et al., 2006; Muerhoff AS et al., 2005]. These genotypes are not evenly distributed worldwide and display significant geographic restriction. For instance, GBV-C genotype 1 is found mainly in West Africa, genotype 2 in Europe and the United States, genotype 3 in parts of Asia and South America, genotype 4 in Southeast Asia, genotype 5 in sub-Saharan Africa, and genotype 6 in Asia.

Address also requests for reprints to: Jason Blackard, PhD, Division of Digestive Diseases, University of Cincinnati College of Medicine, ML 0595, 231 Albert Sabin Way, Cincinnati, OH 45267, Phone: (513) 558-4389, Fax: (513) 558-1744, jason.blackard@uc.edu.

*These authors contributed equally

GBV-C was first identified in subjects with acute hepatitis; however, subsequent studies have found no association between GBV-C infection and liver disease [Linnen et al., 1996; Simons JN et al., 1995]. In contrast, multiple studies have demonstrated a beneficial effect of ongoing GBV-C replication on HIV disease resulting in lower HIV viral loads, slower CD4 cell decline [Williams CF et al., 2004], and prolonged AIDS-free survival compared to persons without GBV-C infection [Stapleton, 2003; Tillmann et al., 2001]. However, other studies have not shown such a survival advantage [Birk et al., 2002; Bjorkman et al., 2004]. While it has been suggested that GBV-C genotype may play a role in modulating HIV disease progression [Alcalde R et al., 2010; Muerhoff AS et al., 2003; Schwarze-Zander C et al., 2006] and could partially explain the divergent findings among these studies, additional investigation is required to fully address this hypothesis. Furthermore, the prevalence and genotype distribution of GBV-C has not been adequately studied in many geographic regions. Therefore, in the current study, we analyzed the prevalence of GBV-C co-infection and the GBV-C genotype distribution in a large cohort of HIV positive subjects in Germany.

Methods

Patient population

408 consecutive subjects who received clinical care for chronic HIV infection at the Bonn HIV clinic from 2005 to 2008 were recruited for the current analysis. The study was approved by the local ethics committee.

RT-PCR detection of GBV-C RNA

Viral RNA was extracted from 140µl of serum using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. GBV-C RNA was detected by nested reverse transcriptase polymerase chain reaction (RT-PCR) using primers corresponding to the 5'UTR as previously described [Schwarze-Zander C et al., 2006]. The E2 region was amplified by nested RT-PCR in several overlapping fragments as previously described [Smith DB et al., 2000] among 17 randomly selected subjects with sufficient serum for additional analysis.

Phylogenetic analysis

The 5'UTR from full-length genome GenBank references corresponding to GBV-C genotypes 1–6 were initially used to determine the GBV-C genotype. These included genotype 1: AB003291 (Japan), U36380 (Ghana), AB013500 (Ghana); genotype 2: D87255 (Japan), U44402 (United States), AY196904 (United States), AB003289 (Japan), AF121950 (United States) AF104403 (France), U63715 (Egypt); genotype 3: U94695 (China), D87712 (Japan), AB003293 (Japan), D87708 (Japan), D87709 (Japan), D87711 (Japan), D87713 (Japan), D90601 (Japan), AB003290 (Japan), AB008335 (Japan), AB008342 (Japan); genotype 4: AB018667 (Vietnam); genotype 5: AY949771 (South Africa); and genotype 6: AB003292 (Japan). Previously published E2 sequences from sub-Saharan Africa were also included when available [Muerhoff AS et al., 2005; Sathar MA and York, 2001; Schleicher SB and Flehmig, 2003]. An internal gap of approximately 38 nucleotides was noted in several E2 sequences reflecting limited sequence coverage in this region. After removal of this gap, sequences were concatenated and analyzed as follows. Nucleotide alignments were initially performed using the neighbor-joining method of CLUSTAL X 2.0.11 [Larkin MA et al., 2007]. Phylogenetic trees were constructed with gaps excluded and a correction for multiple substitutions. The statistical robustness and reliability of the branching order within trees was evaluated by bootstrap analysis with 1000 replicates [Felsenstein, 1985].

Additional phylogenetic inference was performed using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in the Bayesian Evolutionary Analysis by

Sampling Trees (BEAST) v1.5.0 program [Drummond and Rambaut, 2007] under an uncorrelated log-normal relaxed molecular clock and the generalized time reversible (GTR) model with nucleotide site heterogeneity estimated using a gamma distribution. The BEAST MCMC analysis was run for a chain length of 50,000,000. Three independent runs were performed for each dataset. Combined results were visualized in Tracer v1.4 to confirm adequate chain convergence, and the effective sample size (ESS) was calculated for each parameter. All ESS values were >200 indicating sufficient sampling. The maximum clade credibility tree was selected from the posterior tree distribution after a 10% burn-in using TreeAnnotator v1.5.0.

Recombination analysis

To identify possible recombination events, bootscanning analysis of E2 sequences was performed as implemented in SimPlot version 3.5.1 using the Kimura 2-parameter with a 200 base pair (bp) window, a 20 bp step increment, and 1000 bootstrap replicates [Lole KS et al., 1999]. Each E2 sequence was compared to consensus sequences generated for genotypes 1, 2, 3, and 5 using the references described above. For GBV-C genotypes 4 and 6, only limited sequence data are available; therefore, references AB018667 and AB003292 were used. If >70% of the permuted trees showed similarity to more than one genotype across the E2 region analyzed, these “parental” reference/consensus sequences were retained in a second bootscanning analysis along with an outlier and the query sequence. Recombination was further evaluated using the Recombinant Identification Program (RIP) version 3.0 (available at <http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>) with a 200 bp window size and a confidence threshold of 90%. Finally, principal coordinate analysis [Higgins, 1992] was performed to assess patterns in the sequence data using the PCOORD program accessible at <http://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html>.

5'UTR and E2 sequences were submitted to GenBank under the accession numbers GU440639-GU440735 (5'UTR) and GU440736-GU440753 (E2).

Results

GBV-C infection status and phylogenetic analysis of 5'UTR sequences

Among 408 HIV positive subjects, GBV-C RNA was detected in 97 (23.8%) subjects. As shown in Figure 1a, GBV-C genotypes based on the 5'UTR included genotype 1 in eight (8.2%) subjects, genotype 2 in 81 (83.5%) subjects, and genotype 3 in two (2.1%) subjects. An additional six (6.2%) individuals were infected with viruses that were only moderately related to GBV-C genotypes 1–6. Specifically, sequences from subjects 039, 131, 168, 269, and 324 formed an “outlier” group that was separate from all other GBV genotypes and was supported by a bootstrap value of 781. Subject 193 was also included in this grouping albeit with a lower bootstrap value of 533. When these 5'UTR sequences were analyzed using a Bayesian inference approach, a distinct grouping for these six sequences was again highly supported by a posterior probability of 0.99 further suggesting a novel grouping of GBV-C sequences (Figure 1b). Interestingly, each of these six GBV sequences was derived from an individual born in a sub-Saharan African country, including Cameroon (n = 1), Ethiopia (n = 1), Kenya (n = 3), or Rwanda (n = 1) who then immigrated to Germany.

Phylogenetic analysis of E2 sequences and identification of recombination

We and others have previously observed that GBV-C genotyping based on the 5'UTR may not efficiently discriminate GBV-C genotypes/subtypes in all instances [Smith DB et al., 2000]. Therefore, in a subset of 17 individuals, a 1058 bp segment of the E2 gene was also amplified using the overlapping PCR fragment strategy outlined in Figure 2a. As shown in Figure 2b, E2 genotypes were 1 (n = 7), 2 (n = 5), 3 (n = 2), and 5 (n = 1). E2 sequences for

subjects 168 and 269 formed a distinct group that was separate from GBV-C genotypes 1–6 and supported by a bootstrap value of 952. When E2 sequences were analyzed using a Bayesian inference approach, a distinct grouping for subjects 168 and 269 was once again observed and highly supported by a posterior probability of 1.00 (Figure 2c). A principal coordinate analysis was also employed to explore genotypic diversity as reported recently for GBV-C [Branco C et al., 2010]. As shown in Figure 3, GBV-C genotypes 1, 2, 3, and 5 formed distinct clusters that were well separated from one another, although the limited number of sequences available for genotype 4 (n = 2) and 6 (n = 1) resulted in their clustering with genotype 1 references (Figure 3). Importantly, E2 sequences for subjects 168 and 269 again appeared as outliers that were distinct from genotypes 1–6. Additionally, E2 sequences from subjects 133 and 324 were outliers in this analysis.

For 14 of 17 (82.4%) individuals, the GBV-C genotypes were concordant when comparing 5'UTR and E2 results. However, for patient 133, the 5'UTR was genotype 3, while the E2 was genotype 1. For patient 193, the 5'UTR belonged to the unique outlier group, while the E2 belonged to genotype 5. For patient 324, the 5'UTR clustered with the unique outlier group, while the E2 belonged to genotype 3. Because these discordant genotype results suggest possible recombinant viruses, all E2 sequences were further analyzed for recombination events using SimPlot [Lole KS et al., 1999]. As summarized in Table 1, E2 sequences from 3 individuals – 133, 168, and 324 – showed evidence of intergenotype recombination. The E2 sequence for subject 133 was classified as genotype 1 at the 5' end but genotype 3 at the 3' end (i.e., a 1 → 3 recombinant). The E2 sequence for subject 168 had two breakpoints and was classified as a 2 → 4 → 2 recombinant. The E2 sequence for subject 324 was classified as a 5 → 3 recombinant. Similar evidence of intergenotypic recombination within E2 was detected via RIP analysis for these three subjects (data not shown).

Discussion

Genotypic variation among RNA viruses can have profound biological consequences. For instance, despite similar genome structures and replication strategies, HCV genotype is an important determinant of the virologic response to HCV therapy [Hnatyszyn, 2005; Pawlotsky, 2003], while differences in disease pathogenesis among genotypes may also exist [Adinolfi et al., 2001; Rubbia-Brandt et al., 2000]. To date, only a limited number of reports have addressed the impact of GBV-C genotype/subtype on HIV disease. Muerhoff *et al.* found that CD4 cell counts tended to be lower in subjects infected with subtype 2a compared to those with subtype 2b; however, other distinct genotypes were not circulating in the study population for further comparison [Muerhoff AS et al., 2003]. We subsequently observed a significant difference in CD4 counts in HIV positive persons co-infected with GBV-C genotype 2 compared to GBV-C genotype 1 even after controlling for race, HIV viral load, and antiretroviral therapy use [Schwarze-Zander C et al., 2006]. Similarly, another study found that the lowest CD4 cells counts were associated with genotype 1 and the highest with genotype 2b [Alcalde R et al., 2010]. However, in a recent study, no statistical difference in CD4 counts was found among HIV/HCV co-infected persons based on GBV-C genotype, although the predominance of a single genotype did not permit a robust comparison among multiple genotypes [Berzsenyi MD et al., 2009].

It has been suggested that the distinct geographical distribution of GBV-C genotypes is consistent with their co-evolution within humans during pre-historic migrations [Smith DB et al., 2000] and that current phylogenetic relationships among GBV-C isolates may reflect ancient human population movements [Worobey M and Holmes, 2001]. Importantly, the present study represents the largest analysis of GBV-C genotypes conducted to date. The phylogenetic analyses clearly identified the presence of four distinct GBV-C genotypes in

Germany, including an outlier group that could represent a previously unrecognized GBV-C genotype. Interestingly, each of these novel viruses was isolated from individuals born in sub-Saharan African, suggesting that the genetic diversity of GBV-C could be much wider than previously appreciated. To date, classification of GBV-C has relied mainly on sequencing of the 5' UTR. For example, the original identification of genotypes 4, 5, and 6 was based on examination of the 5' UTR [Muerhoff AS et al., 2006; Muerhoff AS et al., 2005; Naito H et al., 1999; Sathar MA et al., 1999; Tucker TJ et al., 1999], although a limited number of full-length genomic sequences are now available for these genotypes [Muerhoff AS et al., 2005; Naito H et al., 2000; Takahashi K et al., 1997]. Similarly, data on the genetic diversity of GBV-C from sub-Saharan Africa is also extremely limited. Nonetheless, the identification of multiple infections with a unique variant of GBV-C in people from sub-Saharan African that subsequently immigrated to Germany suggests that immigration and travel could act as a potential source for introducing and disseminating novel GBV-C strains into a population.

Several recombinant GBV-C isolates were also identified in the current study. While recombination appears to be rare during infection with the distantly related hepatitis C virus, as multiple sub-genomic regions reproduce the same phylogenetic relationships of full-length HCV genomes, Worobey *et al.* convincingly demonstrated that recombination occurs within and between GBV-C genotypes, thus highlighting the important role played by recombination in shaping GBV-C diversity [Worobey M and Holmes, 2001]. While the contribution of recombination to GBV-C pathogenesis itself has not been examined, studies of other highly recombinogenic viruses such as HIV (reviewed in [Blackard JT et al., 2002]) would suggest that GBV-C recombination may have important biological implications. For instance, GBV-C recombination may result in altered cell tropism, virulence, and drug resistance/sensitivity, and may also influence the impact of GBV-C on HIV disease progression.

Clearly, population movement and viral recombination contribute substantially to the overall genetic diversity of GB virus C. However, data on the extent of GBV-C global diversity are limited by restricted geographical sampling that has focused largely on developed countries. Thus, future studies to comprehensively characterize GBV-C genotypes in distinct geographical regions, including sub-Saharan Africa, would permit a more complete assessment of GBV-C diversity. Additional analysis of full-length genomes from the outlier group identified in African subjects is currently underway to explore whether this finding truly represents a novel GBV-C genotype.

Acknowledgments

These data were previously presented at the 11th European AIDS Conference held from October 24–27, 2007. The H.W. & J. Hector Foundation and NIH award R21 AI081564 to JTB funded this work. The authors would like to thank M. Tarek Shata for his critical review of this manuscript.

References

- Adinolfi L, Gambardella M, Andreana A, Tripodi M, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology*. 2001; 33(6):1358–1364. [PubMed: 11391523]
- Alcalde R, Nishiya A, Casseb J, Inocêncio L, Fonseca LA, Duarte A. Prevalence and distribution of the GBV-C/HGV among HIV-1-infected patients under anti-retroviral therapy. *Virus Research*. 2010; 151(2):148–152. [PubMed: 20420864]
- Berzsenyi MD, Bowden DS, Roberts SK, Revill P. GB virus C genotype 2 predominance in a hepatitis C virus/HIV infected population associated with reduced liver disease. *Journal of Gastroenterology and Hepatology*. 2009; 24(8):1407–1410. [PubMed: 19702909]

- Birk M, Lindback S, Lidman C. No influence of GB virus C replication on the prognosis in a cohort of HIV-1-infected patients. *AIDS*. 2002; 16(18):2482–2485. [PubMed: 12461426]
- Bjorkman P, Flamholz L, Naucler A, Molnegren V, Wallmark E, Widell A. GB virus C during the natural course of HIV-1 infection: viremia at diagnosis does not predict mortality. *AIDS*. 2004; 18(6):877–886. [PubMed: 15060435]
- Blackard JT, Cohen DE, Mayer K. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clinical Infectious Diseases*. 2002; 34(8):1108–1114. [PubMed: 11915000]
- Branco C, Esteves A, Piedade J, Parreira R. A new genotype 2 subcluster identified among GBV-C strains circulating in the Lisbon metropolitan area of Portugal. *Journal of Medical Virology*. 2010; 82(3):452–459. [PubMed: 20087931]
- Drummond A, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 2007; 7(1):214. [PubMed: 17996036]
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985; 39:783–791.
- Higgins D. Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets. *Computer applications in the biosciences : CABIOS*. 1992; 8(1):15–22. [PubMed: 1568121]
- Hnatyszyn H. Chronic hepatitis C and genotyping: the clinical significance of determining HCV genotypes. *Antiviral Therapy*. 2005; 10(1):1–11. [PubMed: 15751759]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins D. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 2007(23):21.
- Linnen J, Wages J Jr, Zhang-Keck ZY, Fry KE, Krawczynski KZ, Alter H, Koonin E, Gallagher M, Alter M, Hadziyannis S, Karayiannis P, Fung K, Nakatsuji Y, Shih JW, Young L, Piatak M Jr, Hoover C, Fernandez J, Chen S, Zou JC, Morris T, Hyams KC, Ismay S, Lifson JD, Hess G, Fong SK, Thomas H, Bradley D, Margolis H, Kim JP. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science*. 1996; 271(5248):505–508. [PubMed: 8560265]
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray S. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*. 1999; 73(1):152–160. [PubMed: 9847317]
- Mohr EL, Stapleton J. GB virus type C interactions with HIV: the role of envelope glycoproteins. *Journal of Viral Hepatitis*. 2009; 16(11):757–768. [PubMed: 19758271]
- Muerhoff AS, Dawson GJ, Desai S. A previously unrecognized sixth genotype of GB virus C revealed by analysis of 5'-untranslated region sequences. *Journal of Medical Virology*. 2006; 78(1):105–111. [PubMed: 16299729]
- Muerhoff AS, Leary TP, Sathar MA, Dawson GJ, Desai S. African origin of GB virus C determined by phylogenetic analysis of a complete genotype 5 genome from South Africa. *Journal of General Virology*. 2005; 86(6):1729–1735. [PubMed: 15914851]
- Muerhoff AS, Tillmann HL, Manns MP, Dawson GJ, Desai S. GB virus C genotype determination in GB virus-C/HIV co-infected individuals. *Journal of Medical Virology*. 2003; 70(1):141–149. [PubMed: 12629656]
- Naito H, Hayashi S, Abe K. The entire nucleotide sequence of two hepatitis G virus isolates belonging to a novel genotype: isolation in Myanmar and Vietnam. *Journal of General Virology*. 2000; 81(1):189–194. [PubMed: 10640557]
- Naito H, Win KM, Abe K. Identification of a novel genotype of hepatitis G virus in Southeast Asia. *Journal of Clinical Microbiology*. 1999; 37(4):1217–1220. [PubMed: 10074558]
- Pawlotsky J. Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Research*. 2003; 59(1):1–11. [PubMed: 12834855]
- Rubbia-Brandt L, Quadri R, Abid K, Giostra E, Male PJ, Mentha G, Spahr L, Zarski JP, Borisch B, Hadengue A, Negro F. Hepatocyte steatosis is a cytopathic effect of hepatitis C virus genotype 3. *Journal of Hepatology*. 2000; 33(1):106–115.

- Sathar MA, Soni PN, Pegoraro R, Simmonds P, Smith DB, Dhillon AP, Dusheiko G. A new variant of GB virus C/hepatitis G virus (GBV-C/HGV) from South Africa. *Virus Research*. 1999; 64(2):151–160. [PubMed: 10518711]
- Sathar MA, York D. Group 5: GBV-C/HGV isolates from South Africa. *Journal of Medical Virology*. 2001; 65(1):121–122. [PubMed: 11505453]
- Schleicher SB, Flehmig B. Genotyping of GB virus C by restriction pattern analysis of the 5' untranslated region. *Journal of Medical Virology*. 2003; 71(2):226–232. [PubMed: 12938197]
- Schwarze-Zander C, Blackard JT, Zheng H, Addo MM, Lin W, Robbins GK, Sherman KE, Zdunek D, Hess G, Chung R. for the AIDS Clinical Trial Group A5071 Study Team. GB virus-C (GBV-C) infection in hepatitis C virus (HCV)/HIV co-infected patients receiving HCV treatment: importance of the GBV-C genotype. *Journal of Infectious Diseases*. 2006; 194(4):410–419. [PubMed: 16845623]
- Simons JN, Desai SM, Schultz DE, Lemon SM, Mushahwar I. Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *Journal of Virology*. 1996; 70(9):6126–6135. [PubMed: 8709237]
- Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar I. Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine*. 1995; 1(6):564–569.
- Smith DB, Basaras M, Frost S, Haydon D, Cuceanu N, Prescott L, Kamenka C, Millband D, Sathar MA, Simmonds P. Phylogenetic analysis of GBV-C/hepatitis G virus. *Journal of General Virology*. 2000; 81(3):769–780. [PubMed: 10675415]
- Stapleton J. GB virus type C/Hepatitis G virus. *Seminars in Liver Disease*. 2003; 23(2):137–148. [PubMed: 12800067]
- Takahashi K, Hijikata M, Hino K, Mishiho S. Entire polyprotein-ORF sequences of Japanese GBV-C/HGV isolates: implications for new genotypes. *Hepatology Research*. 1997; 8(2):139–148.
- Tillmann H, Heiken H, Knapik-Botor A, Heringlake S, Ockenga J, Wilber J, Goergen B, Detmer J, McMorro M, Stoll M, Schmidt R, Manns M. Infection with GB virus C and reduced mortality among HIV-infected patients. *New England Journal of Medicine*. 2001; 345(10):715–724. [PubMed: 11547740]
- Tucker TJ, Smuts H, Eickhaus P, Robson SC, Kirsch R. Molecular characterization of the 5' non-coding region of South African GBV-C/HGV isolates: major deletion and evidence for a fourth genotype. *Journal of Medical Virology*. 1999; 59(1):52–59. [PubMed: 10440808]
- Williams CF, Klinzman D, Yamashita TE, Xiang J, Polgreen PM, Rinaldo C, Liu C, Phair J, Margolick JB, Zdunek D, Hess G, Stapleton J. Persistent GB virus C infection and survival in HIV-infected men. *New England Journal of Medicine*. 2004; 350(10):981–990. [PubMed: 14999110]
- Worobey M, Holmes E. Homologous recombination in GB virus C/hepatitis G virus. *Molecular Biology & Evolution*. 2001; 18(2):254–261. [PubMed: 11158384]

Figure 1.

(A) Neighbor-joining tree based on consensus 5'UTR sequences for 97 HIV/GBV-C co-infected subjects (indicated by 3 digit identifiers). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences can be estimated using the 1% divergence bar shown in the upper left corner. Relevant bootstrap values out of 1000 are shown in italics. "Outlier" sequences are denoted by the shaded box. (B) Phylogenetic inference based on a Bayesian MCMC approach as implemented in the BEAST program. Relevant posterior probabilities are shown in italics. The scale bar indicates 0.02 nucleotide substitutions per site.

Figure 2.

(A) Location of primers used to amplify the GBV-C E2 gene relative to the AUG codon (position 524–526 of the reference sequence U36380 [Simons JN et al., 1996]). Overall, 2 single round amplicons (RT1 and RT2) and 3 nested amplicons (NE1–NE3) were generated per subject. (B) Neighbor-joining tree based on the consensus E2 sequences for 17 HIV/GBV-C co-infected subjects (shown in bold and indicated by a 3 digit identifier). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences can be estimated using the 2% divergence bar shown in the upper left corner. Relevant bootstrap values out of 1000 are shown in italics. “Outlier” sequences are denoted by the shaded box. Asterisks denote GBV-C sequences isolated from persons born outside of Germany. (C) Phylogenetic inference based on a Bayesian MCMC approach as implemented in the BEAST program. Relevant posterior probabilities are shown in italics. The scale bar indicates 0.03 nucleotide substitutions per site.

Figure 3.

Principal coordinate plots for E2 sequences from 17 HIV/GBV-C co-infected subjects, as well as representative GenBank reference sequences. Axes represent the first two dimensions that were extracted. Each black diamond represents a distinct E2 reference sequence, while open circles represent E2 sequences from the 17 subjects from Germany. Outlier and/or recombinant sequences are highlighted by arrows and the corresponding subject number.

Figure 4.

Bootscreening analysis of recombination within the E2 region using a 200 bp window, a 20 bp step increment, and 1000 bootstrap replicates. E2 sequences were initially compared to consensus sequences generated for GBV-C genotypes 1, 2, 3, and 5 and references

AB018667 and AB003292 for genotypes 4 and 6. If >70% of the permuted trees showed similarity to more than one genotype across the E2 region analyzed, these “parental” reference/consensus sequences were retained in a second bootscanning analysis along with an outlier and the query sequence as shown for subject 133 (genotype 1 → genotype 3) (**A**), subject 168 (genotype 2 → genotype 4 → genotype 2) (**B**), and subject 324 (genotype 5 → genotype 3) (**C**). The parental genotypes are shown as the top two colors within the legend, while the bottom color represents the outlier genotype/reference. The dashed line indicates the 70% threshold used to denote significance between genotypes.

Summary of genotypic and recombinant analysis of E2 sequences from 17 HIV/GBV-C co-infected subjects. *E2 recombination* refers to the findings of the bootscan analysis of the E2 sequence, while *Sample classification* includes genotyping data from the 5'UTR and the E2. Importantly, a patient's virus can be classified as *recombinant* if 1) the 5'UTR and the E2 genotypes are discordant, and/or 2) intergenotypic recombination is detected within E2.

Table 1

Sample ID	UTR genotype	E2 genotype	E2 recombination	Sample classification
007	2	2	2	Non-recombinant
119	1	1	1	Non-recombinant
133	3	outlier	1 → 3	Recombinant
134	1	1	1	Non-recombinant
139	2	2	2	Non-recombinant
146	1	1	1	Non-recombinant
151	2	2	2	Non-recombinant
168	Outlier group	Outlier group	2 → 4 → 2	Recombinant
193	Outlier group	5	5	Recombinant
252	1	1	1	Non-recombinant
258	2	2	2	Non-recombinant
269	Outlier group	Outlier group	5	Non-recombinant
308	2	2	2	Non-recombinant
320	3	3	3	Non-recombinant
324	Outlier group	3	5 → 3	Recombinant
338	1	1	1	Non-recombinant
371	1	1	1	Non-recombinant