Recombination among GB virus C (GBV-C) isolates in the United States

Jason T. Blackard,1 Gang Ma,1 Clarissa Polen,1 Juwen C. DuBois,1 Jonathon Gast,1 Caleb M. Radens,1 Richard K. Sterling2 and Kenneth E. Sherman1

1Division of Digestive Diseases, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
2Division of Gastroenterology, Hepatology, and Nutrition, Virginia Commonwealth University, Richmond, Virginia, USA

GB virus C (GBV-C) is a non-pathogenic flavivirus that may play a role in modulating HIV disease. Multiple genotypes of GBV-C that have been identified to date that may differentially regulate HIV; however, the number of complete GBV-C sequences published to date is very limited. We sequenced full-length GBV-C genomes from four individuals with HIV/HCV co-infection in the United States. Intergenotypic recombination was evident in two of these individuals. Evaluation of additional full-length GBV-C genomes would facilitate the creation of full-length, replication-competent molecular clones of GBV-C to evaluate the phenotypic diversity of GBV-C genotypes and provide important molecular data on this understudied virus.

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 (ORF) that encodes for a polyprotein of approximately 3000 amino acids (reviewed by Mohr & Stapleton, 2009). This polyprotein encodes for at least two structural proteins – envelope protein 1 (E1) and envelope protein 2 (E2) – as well as for multiple non-structural proteins – NS2 to NS5. To date, at least six GBV-C genotypes have been identified (Feng et al., 2005, 2006). These are typically classified by phylogenetic comparison of the 5′UTR. These genotypes are not evenly distributed worldwide and display significant geographic restriction with genotype 1 in West Africa, genotypes 2 and 2 in Europe and the United States, genotype 3 in parts of Asia, genotype 4 in Southeast Asia and genotype 5 in South Africa (Katayama et al., 1998; Muerhoff et al., 1996, 1997, 2005; Mukaide et al., 1997; Okamoto et al., 1997; Tucker & Smuts, 2000; Tucker et al., 1999).

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 and prolonged AIDS-free survival compared to persons without GBV-C infection (Stapleton, 2003; Tillmann et al., 2001; Williams et al., 2004). However, some other studies have not shown such a survival advantage (Birk et al., 2002; Björkman et al., 2004). While it has been suggested that GBV-C genotype may play a role in modulating HIV disease progression (Alcalde et al., 2010; Muerhoff et al., 2003; Schwarze-Zander et al., 2006), which could explain the divergent findings among these studies, additional investigation is required to evaluate this hypothesis adequately. As a necessary first step, we characterized complete ORF sequences from individuals with GBV-C infection in the United States.

Eighteen subjects were considered as part of a prospective cohort of HIV/HCV co-infected patients designed to characterize changes in HCV viral load and alanine aminotransferase (ALT) levels following combination anti-retroviral therapy (ART) initiation as described in detail previously (Sherman et al., 2014). All patients provided informed consent and the study protocols were approved by the Institutional Review Boards at the enrolling sites (University of Cincinnati, Virginia Commonwealth University and New York University).

For the current study, the first 12 individuals from the prospective cohort were screened for GBV-C infection at week 8 post-ART initiation. Viral RNA was extracted from 140 μl 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 et al., 2006). Two supplementary tables are available in the online Supplementary Material.
PCR products were analysed by agarose gel electrophoresis for the presence of a 256-nucleotide band corresponding to nucleotides 107–362 of GenBank accession number AY196904. Complete ORF sequences were amplified as overlapping fragments using multiple primers as outlined in Table S1a, b (available in the online Supplementary Material).

Sequences were aligned with a database reference using Clustal X 2.1 (Larkin et al., 2007). The GenBank reference sequences used to confirm GBV-C genotype included AB013500, U36380 and KC618399 (genotype 1); U63715, AF121950, AB003289, AF031828, AF081782, D87255, AY196904, AF104403, D90600 and U45966 (genotype 2); AB003288, AB003293, D87263, D87262, D87712, D87708, D90601, D87711, D87709, AB008342, AB008335, AB003290, D87710, D87713, AF06500 and U94695 (genotype 3); HQ331233, HQ331234, HQ331235, AB003292, AB21287 and AB18667.

Fig. 1. Phylogenetic inference of complete open reading frame sequences for Cin03, Cin08, Cin11 and VCU02 (denoted by –) based on a Bayesian MCMC approach as implemented in the BEAST program. GenBank reference sequences are indicated by their accession numbers and genotype. Relevant posterior probabilities are shown. The scale bar indicates 0.9 nucleotide substitutions per site. Previously identified recombinant sequences are denoted by an asterisk.

J. T. Blackard and others
Fig. 2. Bootscanning analysis of recombination for complete ORF sequences using a 600 bp window, a 20 bp step increment and 1000 bootstrap replicates. Genotypes 1 and 2 are shown as the red and green lines, respectively, while the grey line represents the outlier genotype 4. The dashed line indicates the 80% threshold used to denote significance between genotypes. (a) Cin03 is a non-recombinant genotype 2 isolate. (b) Cin08 is a non-recombinant genotype 2 isolate. (c) Cin11 is a genotype 2→1→2 recombinant isolate. (d) VCU02 is a genotype 1→2→1→2→1 recombinant isolate.
(genotype 4); KC618398, KC618400, KC618401 and AY949771 (genotype 5); AB008336 and AB003291 (genotype 6). Putative recombinants included D87715, U75356, AB013501 and AB021287. Outlier sequences included AF070476 – isolated from a chimpanzee – and GB virus A references AF023425, NC001837, U22303 and AF023424. The statistical robustness and reliability of the branching order within the phylogenetic tree was confirmed by bootstrap analysis using 1000 replicates. 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) v1.8.0 program (Drummond et al., 2012) under an uncorrelated log-normal relaxed molecular clock and the
Fig. 3. The 30-amino acid segment of NS5A that inhibits HIV replication is relatively well conserved among the existing full-length GBV-C references (shown by their accession number and genotype) published to date and the additional sequences described in the current analysis (boxed sequences). Asterisks denote previously reported recombinant sequences, while arrows denote polymorphic sites.
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 200,000,000 to yield sample size values >500 indicating sufficient sampling. The maximum clade credibility tree was selected from the posterior tree distribution after a 10 % burn-in using TreeAnnotator v1.7.5. Pairwise distances were calculated in MEGA v7.0.

To identify possible recombination events, bootscanning analysis of complete ORF sequences was performed as implemented in SimPlot version 3.5.1 using the Kimura 2-parameter with a 600 base pair (bp) window, a 20 bp step increment and 1000 bootstrap replicates (Lole et al., 1999). Each complete ORF sequence was compared to consensus sequences generated using available GenBank references for genotypes 1 (n=3), 2 (n=13), 3 (n=16), 4 (n=5) and 5 (n=4). Because of the limited number of sequences available for genotype 6, reference AB003291 was included. If >80 % of the permuted trees showed similarity to more than one genotype across the ORF analysed, the ‘parental’ sequences were retained within a second bootscanning analysis along with the consensus genotype 4 sequence as an outlier, since no GBV-C isolates from this cohort were identified as belonging to genotype 4 and the query sequence. Sequences were submitted to GenBank under the accession numbers KU685420–KU685423.

GBV-C RNA was detected in 5 of 12 (41.7 %) HIV/HCV co-infected individuals screened for GBV-C infection. Overlapping genomic fragments of varying length were generated using multiple primer combinations, although fragments were not always positive for every sample (Table S1a, b). Full-length ORF sequences were obtained for four of five individuals; however, a small fragment of Cin06 did not provide high-quality sequence data and was removed from further analysis. Using a Bayesian inference approach, three samples – Cin03, Cin08 and Cin11 – were identified as belonging to GBV-C genotype 2 (Fig. 1). One sample – VCU02 – belonged to GBV-C genotype 1. Pairwise comparisons showed that VCU02 was 10.3 %–11.8 % different from genotype 1 references. Cin03, Cin08 and Cin11 were 7.8 %–11.6 % different from genotype 2 references and did not cluster amongst themselves. The primary purpose of the original cohort was not to evaluate the impact of GBV-C on HIV disease progression; thus, comparison of demographic or clinical variables based on GBV-C RNA status was not possible. As shown in Table S2, GBV-C RNA-positive individuals included two black males, two white females and one black female. At the time of GBV-C RNA amplification, CD4 cell counts for these individuals ranged from 376 to 744, while HIV viral loads ranged from undetectable to 2437.

We and others have reported recombination among GBV-C genotypes (Neibecker et al., 2011; Worobey & Holmes, 2001). Thus, the four full-length ORF sequences were analysed for recombination events. As shown in Fig. 2(a, b), Cin03 and Cin08 were non-recombinant genotype 2 isolates. In contrast, Cin11 was identified as a genotype 2 → 1 → 2 recombinant isolate, while VCU02 was shown to be a genotype 1 → 2 → 1 → 2 → 1 recombinant isolate (Fig. 2c, d).

Genotypic variation among RNA viruses can have profound biological consequences. The existence of multiple GBV-C genotypes has led several authors to suggest that differences in GBV-C strains circulating within populations might impact HIV disease (Berzsenyi et al., 2005; Kaye et al., 2005; Muerhoff et al., 2003). For instance, Muerhoff et al. reported that CD4 cell counts were lower in subjects infected with GBV-C subtype 2a compared to those with subtype 2b (Muerhoff et al., 2003). We observed a significant difference in CD4 cell 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 ART use (Schwarzke-Zander et al., 2006). GBV-C genotype 2 was also marginally more sensitive to interferon-based HCV therapy than genotype 1. Similarly, another study reported lower CD4 cell counts associated with genotype 1 compared to genotype 2b (Alcalde et al., 2010). However, in a recent study, no statistical difference in CD4 cell 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 rigorous multi-genotype comparison (Berzsenyi et al., 2009).

Whether GBV-C genotypes differ in their biological phenotypes is unclear as in vitro comparison of viral genotypes is rare, and only a single GBV-C molecular clone is currently available (Xiang et al., 2000). Nevertheless, several studies have demonstrated that sensitivity to interferon and cell tropism differ among GBV-C variants (Fogeda et al., 2000; Kato et al., 1998; Shimizu et al., 1999; Xiang et al., 2005). For instance, Xiang et al. demonstrated that the NS5A protein from an interferon-resistant individual inhibited RNA-activated protein kinase (PKR) in vitro, while an interferon-sensitive NS5A did not inhibit PKR function (Xiang et al., 2005). Subsequent analysis identified a 30-amino acid segment of NS5A that was sufficient to inhibit HIV replication (Chang et al., 2007; Xiang et al., 2006). This region is relatively well conserved among the existing full-length GBV-C references, as well as among the additional sequences described in the current analysis (Fig. 3). However, several polymorphic sites are present; thus, the impact of viral variability and the roles of genotypically diverse GBV-C proteins in regulating other cellular pathways require additional evaluation. Importantly, clinical GBV-C isolates also vary in their ability to persist in culture, and sequence variability in key regulatory regions may affect growth in PBMC cultures (George et al., 2003).

Inter-genotypic recombination was evident in two of the four full-length ORF sequences analysed in the current study. Similarly, a previous study conducted by Worobey et al. demonstrated that recombination occurs within and between GBV-C genotypes, thus highlighting the important role played by recombination in shaping GBV-C diversity.
GBV-C recombination in the US

( Worobey & 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 by Blackard et al., 2002) suggest that GBV-C recombination may have important implications for cell tropism, virulence and drug resistance/sensitivity, and may influence the impact of GBV-C on HIV disease progression. However, the primary purpose of the cohort from which these sequences were derived was not to evaluate the impact of GBV-C on HIV disease progression; thus, comparison of demographic or clinical variables based on GBV-C RNA status is not possible, although this has been investigated in other studies as noted above.

Additional analysis of full-length genomes would facilitate the creation of full-length, replication-competent molecular clones of GBV-C to evaluate the phenotypic diversity of GBV-C genotypes and explore the consequences of viral recombination in more depth.

Acknowledgements
This work was supported by the National Institute of Allergy and Infectious Diseases (award AI081564 to JTB and award AI065256 to KES).

References


