Virus Research 238 (2017) 84–93

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Novel non-canonical genetic rearrangements termed "complex structural variations" in HBV genome

Kei Fujiwara*, Kayoko Matsunami, Etsuko Iio, Shunsuke Nojiri, Takashi Joh

Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

ARTICLE INFO

Keywords:
Hepatitis B virus
Non-canonical genetic rearrangement
Complex structural variations
Hepadnavirus variability

ABSTRACT

Backgrounds and aims: Chronic hepatitis B virus (HBV) infection is an important worldwide public health issue. Further knowledge on the characteristics of HBV will facilitate its eradication. Genome structural variations (SVs) are defined by its canonical form such as duplication, deletion, and insertion. However, recent studies have reported complex SVs that cannot be explained by those canonical SVs. A HBV strain (UK2) with an unusual genome structure rearrangement that was completely different from known mutations or rearrangements was previously reported. Thus, this study was conducted to confirm the rearrangement in UK2 as a novel complex SV, and to find additional HBV strains with complex SVs. Further, the contribution of complex SVs in hepadnavirus variability was investigated.

Methods: The genome rearrangement pattern in UK2 was analyzed. Further, a search of online databases retrieved additional HBV strains which were candidates to harbor complex SVs. The architecture of each rearrangement in the candidate strains was analyzed by bioinformatical tools. In addition, alignment of woolly monkey hepatitis virus (WMHV) and HBV from human and non-human primates was performed to investigate the contribution of complex SVs to variability of hepadnavirus.

Results: The rearrangement in UK2 was confirmed as a complex SV. An additional 15 HBV strains were retrieved from databases, and confirmed as harboring complex SVs. Complex combinations of deletion, insertion, and duplication characterized the novel rearrangements. The complex SVs in six strains (37.5%) were composed of deletion, insertion, and duplication. The complex SVs in another six strains (37.5%) consisted of deletion and insertion, followed by insertions and duplication in three strains (18.8%), and deletion and duplication in one strain (6.3%). In addition, unique preS1 promoter insertions, which contained the hepatocyte nuclear factor 1 binding site, were observed in seven (43.8%) of 16 strains. Further, analysis of the genetic sequences of WMHV and HBV from human and non-human primates showed that complex combinations of deletions and insertions accounted for their genetic differences.

Conclusions: Non-canonical genetic rearrangements termed complex SVs were observed in HBV. Further, complex SVs accounted for the genetic differences of WMHV and HBV from human and non-human primates.

1. Introduction

Hepatitis B virus (HBV) infection occurs through vertical, horizontal, sexual, or parenteral routes. Chronic hepatitis and liver cirrhosis, which are caused by chronic HBV infection, can lead to serious clinical problems such as liver failure and hepatocellular carcinoma (HCC). Hepatitis B surface antigen seroprevalence is estimated to be 3.61% worldwide with approximately 248 million individuals who are positive for it (Schweitzer et al., 2015). HBV infection results in 0.5–1.2 million deaths per year, caused by chronic hepatitis, cirrhosis, and HCC, the latter of which alone accounts for 320,000 deaths per year (Lavanchy, 2004). Therefore, chronic HBV infection is one of the important public health issues worldwide.

HBV belongs to the family *hepadnavirdae*, and its viral genome is composed of partially double-stranded DNA of approximately 3200 base pairs (bps). The virus has at least eight genotypes (A to H) defined by nucleotide differences of more than 8%, and each genotype has different geographical distribution (Okamoto et al., 1988; Norder et al., 1992; Arauz-Ruiz et al., 2002; Raimondi et al., 2010). Recently, two other tentative genotypes (I and J) have been reported (Tran et al.,

Abbreviations: HBV, hepatitis B virus; HCC, hepatocellular carcinoma; SVs, structural variations; HBeAg, hepatitis B e antigen; WMHV, woolly monkey hepatitis virus; CURS, core upstream regulatory region; BCP, basic core promoter; HNF1, hepatocyte nuclear factor 1

* Corresponding author at: Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, 1 Kawasumi, Mizuho, Mizuho, Nagoya, Aichi, 467-8601, Japan.

E-mail address: keifuji@med.nagoya-cu.ac.jp (K. Fujiwara).

http://dx.doi.org/10.1016/j.virusres.2017.06.009
Received 26 January 2017; Received in revised form 11 June 2017; Accepted 12 June 2017
Available online 13 June 2017
0168-1702/

2008; Yu et al., 2010; Tatematsu et al., 2009). HBV genotypes are related to the severity of liver disease and response to therapy such as interferon (Wai et al., 2002; Sugauchi et al., 2003a). Various HBV mutations and structural variations (SVs) have been reported in the past, and they are important because they affect HBV replication, immunological response of the host, hepatitis B e antigen (HBeAg) seroconversion, chronic liver disease progression, hepatocarcinogenesis, and drug resistance. Key genetic modifications in HBV include point mutations such as precore (Carman et al., 1989) or core promoter mutations (Okamoto et al., 1994), core deletions (Wakita et al., 1991), preS/S deletions (Sugauchi et al., 2003b; Weinberger et al., 1999), insertions (Pult et al., 1997; Gerolami et al., 2005), and intergenotypic recombinations (Bollyky et al., 1996; Morozov et al., 2000; Simmonds and Midgley, 2005; Suwannakarn et al., 2005; Araujo, 2015). In addition, previous data showed that the frequency of mutations was different among HBV genotypes (Orito et al., 2001).

Genome SV is defined by canonical forms such as duplication, deletion, insertion, and inversion. However, recent studies have reported complex structural variants that cannot be classified by those canonical forms (Quinlan et al., 2010; Quinlan and Hall, 2012; Yalcin et al., 2012). Quinlan and Hall (2012) demonstrated that complex structural variants comprise multiple breakpoints whose origin cannot be explained by a single end-joining or DNA exchange event. Yalcin et al. (2012) described complex SVs as two or more structural variants co-occurring at the same locus. Complex SVs may have more unpredictable and multi-faceted functional impacts compared with known canonical variants; for example, complex mutations can change regulatory regions such as promoters, enhancers, or repressors into a novel configuration (Quinlan and Hall, 2012).

Complex SV is different from intergenotypic recombination in HBV. Intergenotypic recombination in HBV is exchange of sequences between viruses belonging to different genotypes (Bollyky et al., 1996). Simmonds and Midgley (2005) described that both the existence of well-defined breakpoints and the exchange of large sections of sequences between genotypes indicate recombination. The difference between recombinant and non-recombinant reflects the genotypic differences in part of the HBV sequences. The genotypic differences are caused mainly by nucleotide substitutions. On the other hand, complex SVs are composed by combinations of multiple SVs. The nucleotide differences caused by complex SVs are generally higher than genotypic nucleotide substitutions in recombination due to combinations of SVs which do not show the sequence similarity to the original genetic sequences. In addition, sequence gaps may occur, although it depends on the types of SVs composing complex SVs.

In 2005, an unusual genomic rearrangement was reported in HBV (Fujiwara et al., 2005). The rearrangement consisted of an insertion, a deletion and a duplication co-occurring at the same locus. The rearrangement was novel and completely different from the mutations or SVs previously reported in the HBV genome. However, at that time, a defined characterization system for genomic SVs was not available. Currently, the system of complex SVs has been used to describe genome variations of several model species (Quinlan et al., 2010; Quinlan and Hall, 2012; Yalcin et al., 2012), and in this study, the characterization system used to describe the complex structural variants was adapted from the above-mentioned studies performed on eukaryotic genomes. The HBV strain with the novel genomic rearrangement was analyzed for complex SVs. Then, additional HBV strains with complex SVs were retrieved from available databases, and detailed analyses were performed in order to investigate rearrangement patterns and virological impact of complex SVs in HBV. Further, the relationship of complex SVs to hepadnavirus variability was investigated.

2. Materials and methods

2.1. Search strategy

A search of published articles in PubMed was performed. Research articles from 1980 to December 2016 were retrieved, using the keywords (“HBV” and “mutation”); (“HBV” and “recombination”); (“HBV” and “insertion”); (“HBV” and “deletion”); (“HBV” and “duplication”) or (“HBV” and “rearrangement”). Approximately 4800 abstracts or full articles were obtained. In addition; a search was also conducted in the Hepatitis Virus Database (http://s2as02.genes.nig.ac.jp/). Alignment of 6577 full genome nucleotide sequences was inspected. HBV genomes integrated into human genomes were not included in this study.

2.2. Reference sequences

Consensus reference sequences of HBV genotype A (HBV/A) to H and HBV from non-human primates, such as orangutan, chimpanzee, gorilla, and gibbon, were constructed from the full genome alignment using the CLUSTAL W software program (Thompson et al., 1994). The consensus reference sequences of HBV/A, HBV/B, HBV/C, HBV/D, HBV/E, HBV/F, HBV/G, HBV/H, orangutan HBV, chimpanzee HBV, gorilla HBV, and gibbon HBV were determined by the analysis of 150, 40, 168, 79, 38, 38, 13, 30, 8, 27, 6, and 27 full genome sequences, respectively. The list of HBV genome sequences from human and non-human primates are in the Supplementary Information.

2.3. Analysis of rearrangement patterns

Complex SVs are defined by the SVs with multiple breakpoints, and are composed of a complex mixture of deletions, insertions, inversions, and copy number gains such as duplications (Quinlan and Hall, 2012; Yalcin et al., 2012). Contrary to canonical SVs, complex SVs show great nucleotide differences when their genetic sequences are compared to reference sequences. The retrieved nucleotide sequences were aligned with the reference sequence (HBV/A, V00866) and the consensus reference sequences of HBV/A to HBV/E using the CLUSTAL W software program (Thompson et al., 1994), and the alignments were confirmed by visual inspection. When partial sequences with low homology to the reference sequence and the sequence gaps were detected by visual inspection, homology search for the unique sequence was carried out using NCBI BLAST 2.2.31 (Altschul et al., 1994), and then additional analysis with visual inspection was performed by considering the architecture of complex SVs in previous reports as references (Quinlan and Hall, 2012; Yalcin et al., 2012). Detailed data on each analyzed complex SV pattern are shown in the figures along with nucleotide alignments. The complex SV patterns are also represented as simplified schematic diagrams without the corresponding nucleotide sequences according to Gunther et al. (1996), Quinlan and Hall (2012) and Yalcin et al. (2012). According to the latter two reports, the sequence of each complex SV was compared with the reference HBV sequence to clarify the breakpoints. Further, representative consensus sequences from HBV/A to HBV/H and HBV from non-human primates (orangutan, chimpanzee, gorilla, and gibbon) were aligned with the woolly monkey hepatitis virus (WMHV) (AF046996) using the CLUSTAL W software program (Thompson et al., 1994), and the alignments were confirmed by visual inspection.

2.4. Phylogenetic analysis and recombination analysis

Phylogenetic analyses were performed by the MEGA software version 6 (Tamura et al., 2013), using the neighbor-joining method. Bootstrap resampling and reconstruction with 1000 replicates were carried out. Genetic distance calculation and pairwise distance comparisons were performed using the Kimura two-parameter model integrated into the MEGA software. Inter-genotype recombination of HBV