Infection, Genetics and Evolution 63 (2018) 135–143

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Genetic diversity of bat orthohepadnaviruses in China and a proposed new nomenclature

Ling'en Yang[a,b,1], Jianmin Wu[c,1], Tingsong Hu[d], Shaomin Qin[c], Bo Deng[d], Jinfeng Liu[c], Fuqiang Zhang[d,**], Biao He[a,e,*], Changchun Tu[a,b,e,*]

[a] Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Institute of Military Veterinary Medicine, Academy of Military Medical Sciences, Changchun, Jilin Province, China
[b] Fujian A&F University, College of Animal Science, Fuzhou, Fujian Province, China
[c] Guangxi Key Laboratory of Veterinary Biotechnology, Guangxi Veterinary Research Institute, Nanning, Guangxi Province, China
[d] Centers for Disease Control and Prevention of Chengdu Military Command, Kunming, Yunnan Province, China
[e] Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou University, Yangzhou, Jiangsu Province, China

ARTICLE INFO

Keywords:
Bats
Orthohepadnavirus
Genetic diversity
Nomenclature

ABSTRACT

The orthohepadnaviruses, which include the major human pathogen hepatitis B virus, exist in a wide range of hosts. Since 2013, a large group of orthohepadnaviruses has been identified in bats worldwide and classified as 4 species within the genus *Orthohepadnavirus*. To further investigate orthohepadnaviruses in the Chinese bat population, 554 archived bat samples from 20 colonies covering 3 southern provinces were screened with results showing that 9 (1.6%) were positive. A systematic phylogenetic analysis has indicated the need for a new nomenclature for bat hepatitis B virus-like viruses: BtHBV, with the addition of 3 new species, one being divided into 6 genotypes. Viruses identified here shared 9.0–19.2% full genome divergence and classified into 3 different genotypes. This study illustrates the genetic diversity of orthohepadnaviruses in the Chinese bat population, and emphasizes need for further investigation of their public health significance.

1. Introduction

Hepatitis B virus (HBV), the prototype member of the genus *Orthohepadnavirus* within the family *Hepadnaviridae*, is one of the most infectious human pathogens, responsible for > 250 million chronic infections worldwide and around 887,000 human deaths annually (ICTV, 2017; Mason et al., 2011; Ott et al., 2012; Seeger et al., 2013; WHO, 2017). Orthohepadnaviruses have also been characterized in other primates and rodents (Seeger et al., 2013). Additionally, 14 species of bats from the Old World (Myanmar, China and Gabon) and New World (Panama) have recently been found to harbor a variety of orthohepadnaviruses (Drexler et al., 2013; He et al., 2013, 2015; Nie et al., 2018; Wang et al., 2017), of which 4 have been approved in the latest report of the International Committee on Taxonomy of Viruses (ICTV) as new species within the genus *Orthohepadnavirus*: *Long-fingered bat hepatitis B virus*, *Pomona bat hepatitis B virus*, *Roundleaf bat hepatitis B virus*, and *Tent-making bat hepatitis B virus* (ICTV, 2017). Molecular studies have shown that the Panamanian tent-making bat hepatitis B virus (TBHBV) is capable of infecting primary human hepatocytes through binding to the same receptor as human HBV, thereby suggesting that bat orthohepadnaviruses are ancestors of primate HBVs (Drexler et al., 2013; Rasche et al., 2016).

Orthohepadnaviruses are much more diverse in bats than in other host species. A single bat species can harbor diverse strains of these viruses, and a single orthohepadnavirus species can infect different bat species: e.g., *Miniopterus schreibersi* (*M. schreibersi*) bats have been found carrying both Neixiang-Ms69 and Anlong-Ms258 viruses, two viruses showing enough genetic diversity to be considered as two distinct species (Nie et al., 2018), whereas Neixiang-Ms69 and Neixiang-Rpu92, sharing almost 100% full genome nucleotide (nt) similarity, have been found in two distinct bat species of two separate families, *M. schreibersi* and *Rhinolophus pusillus* (*R. pusillus*) (Nie et al., 2018). Currently nomenclature of a bat orthohepadnavirus species uses HBV after the common name of the host in which they have been first detected; e.g., "*Long-fingered bat hepatitis B virus*" representing a virus species identified in *Miniopterus* bats, but this could refer to either strain 776 from

* Corresponding authors at: Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Institute of Military Veterinary Medicine, Academy of Military Medical Sciences, Changchun, Jilin Province, China.
** Correspondence to: Fuqiang Zhang, Centers for Disease Control and Prevention of Chengdu Military Command, Kunming, Yunnan Province, China.
E-mail addresses: zfq1968@aliyun.com (F. Zhang), heb-001001@163.com (B. He), changchun_tu@hotmail.com (C. Tu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.meegid.2018.05.024
Received 5 December 2017; Received in revised form 23 May 2018; Accepted 25 May 2018
Available online 26 May 2018
1567-1348/

Myanmar or Neixiang-Ms-69 from China — two different viruses (He et al., 2013; Nie et al., 2018). Such a mismatch can cause confusion. Accordingly, a new species nomenclature for bat HBV-like viruses (BtHBV) is necessary.

With the exception of humans and possibly rodents, bats are the most abundant as well as the most widely distributed mammals (Wilson and Reeder, 2005). They provide a huge natural virus bank from which > 200 viruses have been isolated or identified, including such lethal agents as lyssaviruses, henipaviruses, ebolaviruses, and SARS-related coronaviruses (Moratelli and Calisher, 2015). Apart from orthohepadnaviruses in bats in America, Asia and Africa, homologues of hepatitis A, C and E virus have also been sporadically identified in these animals (Drexler et al., 2015; Drexler et al., 2012; Quan et al., 2013). To further understand the genetic diversity of HBV-like virus, we screened hepatitis viruses from bats collected in south China, and found another three variants. Systematic and phylogenetic analyses of the BtHBVs identified 7 species, for which we propose a new nomenclature that will be more helpful for understanding host species specificity and evolution of these viruses.

2. Materials and methods

2.1. Sample collection

The procedures for sampling of bats in this study were reviewed and approved by the Administrative Committee on Animal Welfare of the Institute of Military Veterinary, Academy of Military Medical Sciences, China (Laboratory Animal Care and Use Committee Authorization, permit numbers: JSY-DW-2010-02 and JSY-DW-2015-01). All live bats were maintained and handled according to the Principles and Guidelines for Laboratory Animal Medicine (2006), Ministry of Science and Technology, China. To investigate orthohepadnaviral ecology in Chinese bats, animals were collected between 2005 and 2015 with nets near or in human-inhabited communities in Yunnan, Guangxi and Guangdong provinces. All bats were apparently healthy at capture, and were initially identified by morphological examination by a trained field biologist and confirmed by sequence analysis of the mt-*cyt b* gene (Wang et al., 2003). Following euthanization their livers and other organs were immediately collected and stored at −80 °C (Fig. 1 and Table 1).

2.2. Virus detection and full genome amplification

To screen for orthohepadnaviruses, viral DNA was extracted from liver tissue of each of the 554 bats using the QIAamp DNA Mini Kit (Qiagen) in a QIAcube (Qiagen).Virus detection was conducted using the TaKaRa PCR Kit (TaKaRa) as previously described (He et al., 2013). Double-distilled water was used as a negative control. Two samples of each lineage were chosen for amplification of their complete genomes using LA Taq polymerase (TaKaRa) and previously described primers (He et al., 2013). Positive PCR amplicons were ligated into pMD18T vector (TaKaRa) and used to transfect competent *Escherichia coli* DH5α cells (Tiangen). > 3 clones of each amplicon were randomly picked for sequencing by the Sanger method in an ABI 3730 sequencer (ComateBio). Overlapping amplicons were assembled with SeqMan v7.0 into full genomic sequences.

2.3. Genomic characterization and phylogenetic characterization

Genomic structures were determined by Vector NTI v.8, followed by comparison with those of other orthohepadnaviruses. Full genomes and predicted open-reading frames (ORF) *Pol*, *preS1/S2/S*, *preC/C*, and *X* (only for orthohepadnaviruses) of representatives of hepadnaviruses, including unclassified ones, retrieved from Genbank were aligned with counterparts of viruses identified in this study using MAFFT v7.394 (Katoh and Standley, 2013), with nt identities being calculated using MegAlign v7.1 (DNASTAR, Inc., Madison, WI). Evolutionary models were determined using ModelGenerator v0.85 with Akaike Information Criterion 1(AIC1) (Keane et al., 2006). Phylogenetic and molecular evolutionary analyses were achieved using PhyML v3.3 by the maximum likelihood and best-fit substitution models under evaluation of 100 bootstraps (Guindon et al., 2010). The generated tree files were visualized with FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

2.4. Genbank accession numbers

Full genomic sequences obtained here were deposited in Genbank under the accession numbers KY905324-KY905329.

3. Results

3.1. Sample collection and viral detection

During 2005–2015, 554 frugivorous and insectivorous bats representing 13 species were collected from 1 to 3 adjacent colonies in 11 locations in Yunnan, Guangxi and Guangdong provinces (Fig. 1). Collection details are shown in Table 1. Pan-orthohepadnavirus PCR screening gave positive results in 9 bat samples from 3 colonies. Of these, 10.0% (4/40) and 9.4% (3/32) were from *Hipposideros pomona* (*H. pomona*) bats in Zhenyuan, Yunnan province, 2013, and Baise, Guangxi province, 2014 respectively. The remainder (2/3) came from *H. armiger* bats in Hechi, Guangxi province, 2015 (Table 1). The amplicons were sequenced, and phylogenetic analysis revealed that they formed three distinct lineages. Of note is that the three lineages were from three different bat colonies. All BtHBVs in this study were named as described previously (He et al., 2014), the first two letters representing the sampling location, with the remaining letters identifying the host species and numbers referring to the sampling order.

3.2. Full genomic characterization

Full genomes of six samples with two of each lineage were successfully sequenced. Analyses showed that the 6 full genomes ranged from 3254 to 3275 nt in length, consistent with previously published BtHBVs (3230–3287 nt) in Asia, including PEPRs in Pu'er and Rs3364 in Jinning, Yunnan, China, and 776 in Kachin state, Myanmar (He et al., 2013). As in all orthohepadnaviruses, the viruses identified here contained four ORFs: polymerase (*Pol*), surface (*S*), core (*C*), and *X*. Positions of all ORFs in these bat orthohepadnaviruses were similar to ORFs of currently reported members of the *Orthohepadnavirus* genus but were clearly distinct from the ORFs of *Avihepadnavirus*. The S protein genes encoded in the large ORF contained preS1, preS2, and S domains. The preS1 domain contained a signal of N-myristoylation at glycine-2. Additional to their ORF organization, human HBV and the BtHBVs shared a similar location for the direct repeat (DR) sequences DR1 and DR2 involved in genome replication. Secondary structure prediction highlighted the structural similarities between BtHBVs and HBVs of other animal origin in their ε-loops, which serve as templates for the priming of reverse transcription of pre-genomic RNA in all hepadnaviruses (Seeger et al., 2013).

3.3. Phylogenetic analysis and sequence comparison

All representatives of hepadnaviruses available in Genbank were retrieved and aligned with the sequences obtained here. The best-fit substitution models, under AIC1, for full genomes, ORFs *Pol*, *preS1/S2/S*, *preC1/C* and *X* were selected as GTR + G [Gamma distribution parameter alpha (a) = 0.61], TVM + G (a = 0.61), GTR + G (a = 0.62), GTR + I + G (*I* = 0.05, a = 0.63), and TVM + G (a = 0.45), respectively. Phylogenetic trees based on the full genomic and ORF nucleotide sequences are shown in Fig. 2–4, with the