Virology 476 (2015) 355–363

Contents lists available at ScienceDirect

Virology

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

An expanded taxonomy of hepatitis C virus genotype 6: Characterization of 22 new full-length viral genomes

Chunhua Li [a], Eleanor Barnes [b], Paul N. Newton [c,d], Yongshui Fu [e], Manivanh Vongsouvath [c], Paul Klenerman [b], Hiroaki Okamoto [f], Kenji Abe [g], Oliver G. Pybus [h], Ling Lu [a,*]

[a] The Viral Oncology Center, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS, USA
[b] The Peter Medawar Building for Pathogen Research, University of Oxford, South Parks Road and Oxford NIHR Biomedical Research Centre, OX1 3SY, UK
[c] Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao Democratic People's Republic
[d] Centre for Tropical Medicine and Global Health, Churchill Hospital, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK
[e] Guangzhou Blood Center, Guangzhou 510095, China
[f] Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, 3311-1 Yakushiji, Shimotsuke-shi Tochigi 329-0498, Japan
[g] Department of Pathology, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan
[h] Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, UK

ARTICLE INFO

Article history:
Received 17 October 2014
Returned to author for revisions
6 December 2014
Accepted 16 December 2014

Keywords:
HCV
Full-length genome
Genotype 6
Sequence
Laos

ABSTRACT

We characterized the full-length genomes of 22 hepatitis C virus genotype 6 (HCV-6) isolates: 10 from Vietnam (classified into subtypes 6e, 6h, 6p, 6r, 6s, and 6u), one from China (confirmed as a new subtype 6xd), and 11 from the Lao PDR (representing a new subtype 6xe plus eight novel variants). With these 22 new genomes, HCV-6 now has a diverse and extended taxonomic structure, comprised of 28 assigned subtypes (denoted 6a-6xe) and 27 unassigned lineages, all of which have been represented by full-length genomes. Our phylogenetic analyses also included many partially-sequenced novel variants of HCV-6 from Lao PDR. This revealed that Lao HCV isolates are genetically very diverse and are phylogenetically distributed in multiple lineages within genotype 6. Our results suggest that HCV-6 has been maintained in Laos, a landlocked country, since the common ancestor of genotype 6 and indicates historical dispersal of HCV-6 across Southeast Asia.

Introduction

Hepatitis C virus (HCV) is a blood-borne pathogen that has a global prevalence of about 3%, affecting approximately 170–220 million people worldwide (WHO, 1999). Infection with HCV results in chronic hepatitis in about 70–85% of the infected individuals, which causes a raised risk of developing cirrhosis, hepatocellular carcinoma, and liver failure (Jahan et al., 2012). The frequency of HCV infection varies considerably among geographic regions, with Egypt displaying the highest prevalence, at 27% (Saeed et al., 1991). In contrast, north-western Europe, North America, and Australia typically exhibit low prevalence of HCV infection, at about 1% (Touzet et al., 2000; Sy and Jamal, 2006). HCV prevalence in Asian countries is variable and in some instances is significantly higher than the global average (WHO, 2012) such that reported in China (3.2%) (Xia et al., 1996), Pakistan (4.95–5.31%) (Khokhar et al., 2004; Muhammad and Jan, 2005), Thailand (3.2–5.6%) (Apichartpiyakul et al., 1999; Songsivilai et al., 1997; Sunanchaikarn et al., 2007), Cambodia (2.3%) (Akkarathamrongsin et al., 2011), and Vietnam (2–9%) (Nakata et al., 1994; Tran et al., 2003). The seroprevalence of anti-HCV among first time blood donors in Vientiane City, Lao PDR (Laos), was detected to be at 1.1% (Jutavijittum et al., 2007), although that study was not a national survey.

Among the seven genotypes of HCV, genotype 6 (HCV-6) exhibits the highest genetic diversity (Salemi and Vandamme, 2002; Wang et al., 2013). Currently, a total of 26 subtypes, denoted 6a-6xc, are formally assigned to HCV-6, and each has at least one full-length genome determined. In addition, 19 lineages have also been completely sequenced, and many novel variants have been detected through the sequencing of partial genomic regions (Smith et al., 2014; Li et al., 2014). Geographically, HCV-6 isolates are typically found in Southeast Asia or among expatriates from this region, which may suggest that this region represents the ancestral and endemic region of HCV-6 (Bernier et al., 1996; Mellor et al., 1996; Noppornpanth et al., 2006; Pham et al., 2011; Pybus et al., 2009; Shinji et al., 2004; Stuyver et al., 1995; Thaikruea et al., 2004; Theamboonlers et al., 2002). In several

* Corresponding author. Fax: +1 913 945 6836.
E-mail address: llu@kumc.edu (L. Lu).

http://dx.doi.org/10.1016/j.virol.2014.12.025
0042-6822/

Southeast Asian countries HCV-6 is highly prevalent. For example, HCV-6 accounts for about 21–49% of HCV infections in Myanmar (Akkarathamrongsin et al., 2011; Lwin et al., 2007; Shinji et al., 2004), 47% in Vietnam (Nguyen et al., 2010; Pham et al., 2009), 56% in Cambodia (Akkarathamrongsin et al., 2011), and a notably high percentage of 95.6% in Laos (Hübschen et al., 2012).

It is not clear why HCV-6 is so dominant in Laos. It has been hypothesized that the globally-prevalent subtypes of 1a, 1b, 2a, and 3a may be rarer in Laos due to the country's landlocked environment and historical lack of medical infrastructure, resulting in a lower likelihood of parenteral exposure (Pybus et al., 2009). There have only been two studies of HCV diversity in Laos. The first study investigated samples from 31 patients with anti-HCV antibodies admitted to Mahosot Hospital in Vientiane City. Fifteen patients were found to be HCV RNA-positive and all isolates belonged to HCV-6. Of these, one was classified as subtype 6h, one as subtype 6o, and the remaining 13 isolates represented unclassified variants (Pybus et al., 2009). The second, larger, study investigated samples obtained from 105 first-time blood donors in Vientiane City that were positive for anti-HCV. This resulted in 45 HCV isolates being sequenced in Core-E1 and/or NS5B regions. Among these 45, two represented the globally prevalent subtype 1b, while 43 were classified as HCV-6. Of the 43 HCV-6 isolates, 11 belonged to subtypes 6b, 6h, 6k, 6l, 6n, and 6q, whilst 32 represented novel variants (Hübschen et al., 2011). A common finding from both studies is that Laos contains a very high proportion of novel and unclassified HCV-6 variants that have not been reported in surrounding countries. The high genetic diversity of HCV-6 in Laos suggests the virus has circulated endemically in the country for a very long time.

To better understand the genetic diversity of HCV-6 and to investigate the phylogeny of HCV strains from Laos, we obtained complete genomic sequences representing 11 unclassified HCV-6 variants (Pybus et al., 2009). These new sequences substantially extend the current taxonomic structure of HCV-6 and will form the basis of future studies on the molecular epidemiology and evolution of HCV in Southeast Asia.

Results

Genome sequences and organization

Nearly full-length genomes of HCV were determined for 22 HCV-6 isolates (L23, L132, L176, L250, L310, L344, L347, L350, L390, L373, L394, TV280, TV395, TV396, TV407, TV412, TV406, TV462, TV503, TV546, KM98, and VN085), each with 16–20 overlapping fragments (see "Materials and methods" for full sample details). These genomes were 9393–9529 nucleotides (nt) in length, measured from the start of the 5′-UTR through to the variable region or X-tail of the 3′-UTR (Table S1). Each genome contained a single ORF of 9045–9063 nt. The 5′-UTRs were 324–339 nt long, while the 3′-UTRs varied from 1 to 139 nt in length. The sizes of 10 protein encoding regions were the same as those in the H77 reference genome (Table S1) except for the E2 (1089–1107 nt) and NS5A (1053 to 1056 aa) genes, whose lengths varied among the 22 isolates, and for the E1 gene of isolate L350. The latter exhibited a single codon insertion between nucleotide positions corresponding to 1042–1043 in the H77 reference genome.

Analysis of full-length genomes

A maximum likelihood (ML) tree was obtained based on 99 full-length HCV-6 genome sequences, together with 11 further reference genomes to represent the other six genotypes of HCV (Fig. 1), in addition to 15 sequences of animal hepaciviruses as an outlier group. The latter included three sequences from horses (AB863589, NC024889, and KJ472766), four sequences from bats (KC796074, KC796078, KC796090, and KC796091), and eight sequences from other nonprimate animals (JQ434001–JQ434008). For simplicity, in Fig. 1 all non-HCV-6 genomes and the 15 sequences from animal hepaciviruses have been collapsed into a single branch. Based on phylogenetic analysis and distance-based methods (Simmonds et al., 2005), the 99 HCV-6 sequences were divided into 28 subtypes, 6a-6xe, and 27 unassigned lineages. Each of these exhibited full bootstrap support of 100% in the phylogeny, with two exceptions: the divergence between L310 and subtype 6c (bootstrap score was 89%) and the divergence between L250 and subtype 6j (bootstrap score was 87%). Among the 22 newly-generated sequences from this study, TV280, TV395, and TV503 were classified into subtype 6e, TV407, TV412, and VN085 into 6h, TV462 into 6p, TV406 into 6r, TV396 into 6s, and TV546 into 6u. These 10 genomes, which were all sampled in Vietnam, showed pairwise nucleotide similarities to their nearest references of 87.9–95.0%. Based on the criteria of the current HCV classification system (Simmonds et al., 2005; Smith et al., 2014), these genetic similarities are sufficiently high to classify these 10 isolates into the six subtypes outlined above. With the completion in this study of the genomes for isolates TV280, TV395, and TV503, subtype 6e is now represented by a total of seven full-length sequences. These sequences formed two well-supported clusters in Fig. 1, with TV280, TV503, and D42 belonging to one group and D88, TV395, 537798 and GX004 to the other.

In contrast, the other 12 genomes obtained in this study could not be classified into any of the currently known subtypes (6a-6xc). Among these, 11 were generated from samples from Laos, and one (KM98) was sampled in China. For convenience of description, we here divided the genetic diversity of HCV-6 into four subsets (Fig. 1). Subset 1 contained three (KM98, L176, and L250) of the 12 unclassified genomes, subset 2 included four (L373, L394, L23, and L347), and subset 3 included five (L310, L132, L350, L390, and L344). In subset 1, KM98 was most closely related to the reference sequence DH027 from China, and the nucleotide similarity between the two was 95%. Further, in the phylogenies obtained for the Core-E1 and NS5B region sequences (Figs. 2 and 3), KM98 and DH027 were joined by another isolate, KM05 (accession numbers KF586079 and KF585672) (Lu et al., 2014). Nucleotide similarities among these three isolates were 95.3–98.4% in the Core-E1 region and 95.3–98.2% in NS5B. These three isolates, which were all sampled in China, represent a new subtype that is denoted 6xd. In subset 2, L23, L347 and L394 formed a cluster that was located close to subtype 6b. Nucleotide similarities among these three genomes were 86.5–88.1% and they differed from the nearest subtype 6b isolate (Th580) in 15.5–15.9%, hence they meet the criteria to confirm a second new subtype, which is denoted 6xe.

The eight sequences from Laos (L132, L176, L250, L310, L344, L350, L390, and L373) each appeared to represent a new subtype but unassigned. Among them, L250 was placed between isolate QC271 and subtype 6j (as represented by isolates C-0667 and Th553) and showed *p*-distances of 14.3%, 14.0%, and 14.1% to the QC271, C-0667 and Th553 sequences, respectively. Excluding isolate L250, the remaining seven Lao sequences all exhibited $>15\%$ *p*-distances to their most closely related relatives. In subset 1, L176 was placed close to subtype 6h and differed from 6h by *p*-distances of 21.2–21.8%. In subset 2, L373 was placed near to subtype 6a and differed from 6a by *p*-distances of 21.4–22.1%. In subset 3, L310 was located between subtypes 6c and 6d and showed *p*-distances of 23.5–24.4% from 6c and 6d. L132 was more similar to subtype 6d than to subtype 6c and showed a *p*-distance of 16.8% from 6d. L350 was placed close to TV453 with a *p*-distance of 21%. Both L344 and L390 formed a cluster together with subtype 6q, but they differed from 6q by *p*-distances of 16.6–21.9%.

A range of nucleotide differences, 13–15%, has been recently defined as the genetic distance threshold required to separate