Journal of Medical Virology 80:1740–1746 (2008)

Complete Genome Sequencing and Phylogenetic Analysis of HCV Isolates From China Reveals a New Subtype, Designated 6u

Xueshan Xia,[1] Wenhua Zhao,[2,3] Kok Keng Tee,[4] Yue Feng,[1] Yutaka Takebe,[4] Qihan Li,[2] Oliver G. Pybus,[5] and Ling Lu[6]*

[1]*Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunan, China*
[2]*Department of Viral Immunology, Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming, China*
[3]*The Key Laboratory of Tropical and Subtropical Animal Disease in Yunnan Province, Kunming, Yunnan, China*
[4]*Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Disease, Shinjuku-ku, Tokyo, Japan*
[5]*Department of Zoology, University of Oxford, South Parks Road, United Kingdom*
[6]*Division of Gastroenterology-Hepatology, and Nutrition, Department of Medicine, University of Utah, Salt Lake City, Utah*

Full length genome sequences were characterized for three novel hepatitis C virus (HCV) isolates (here named DH012, DH014, and DH028). The complete genomes were all isolated from injecting drug users (IDUs) who were co-infected with HIV-1 and lived in Dehong prefecture, Yunnan Province, China, which neighbors Myanmar. The three genomes are 9,443–9,470 nt in length and each contains a single open reading frame (ORF) of 9,069 nt long. Pairwise comparisons indicated nucleotide similarities of 97.9–98.6% among the three isolates, and similarities of 72.4–75.0% between the three isolates and 20 reference strains (representing HCV subtypes 6a–6q and 6t plus two unassigned genotype 6 isolates km41 and gz52557). Phylogenetic analyses demostrated that the three isolates formed a tight and well-supported monophyletic cluster in the genotype 6 clade. No evidence of viral recombination was found using similarity plots and bootscanning analyses. Based on the current HCV classification criteria, we have assigned the three isolates to a new subtype, 6u. Although another ''6u'' isolate D83 has been reported very recently, it is subtypically distinct from the three isolates we described here. Because its designation does not meet the criteria described in the updated HCV nomenclature proposal, this ''6u'' isolate should be reclassified. ***J. Med. Virol. 80:1740–1746, 2008.*** © 2008 Wiley-Liss, Inc.

KEY WORDS: HCV; genotype; sequence

INTRODUCTION

The hepatitis C virus (HCV) is a major causative agent of viral hepatitis. It infects an estimated 170 million people worldwide or about 3% of the global population [World Health Organization, 1999]. After infection, approximately 70–85% of the patients become chronically infected, 10–40% progress to liver cirrhosis, and a number goes on to develop hepatocellular carcinoma [Hoofnagle, 2002; Zoulim et al., 2003]. The main transmission route of HCV infection is through exposure to contaminated blood, via routes such as injecting drug use, unsafe medical practices, and blood transfusion. Transmission of HCV also occurs through sexual contacts and via other, as yet unknown, routes [Alter et al., 1999].

HCV is classified in the *Hepacivirus* genus of the Flaviridae family, whose members possess a single stranded positive RNA genome. The genome is about 9.6 kb in length and contains a single open reading frame (ORF) of nearly full genome size. The ORF

Sequences reported in this study have been submitted in Genbank. Their accession numbers are: EU408330–32.

Grant sponsor: Natural Science Foundation of China (NSFC; partial support); Grant number: 30460119; Grant sponsor: International Collaboration Project; Grant number: 2006GH07; Grant sponsor: Natural Science and Application Project (Yunnan Provincial Government, China); Grant number: 2007C038M.

*Correspondence to: Dr. Ling Lu, Division of Gastroenterolog, the University of Utah, 2000 Circle of Hope Drive, Rm3065, Salt Lake City, UT 84112. E-mail: ling.lu@hsc.utah.edu

Accepted 1 July 2008

DOI 10.1002/jmv.21287

Published online in Wiley InterScience (www.interscience.wiley.com)

© 2008 WILEY-LISS, INC.

encodes three structural (Core, E1, and E2) and seven non-structural (P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) proteins [Major and Feinstone, 1997]. Among these proteins NS5B serves as the viral RNA dependent RNA polymerase (RdRp). The RdRp lacks proof reading activity during viral RNA genome replication and drives the rapid generation of HCV genetic variation [Bartenschlager and Lohmann, 2000].

Phylogenetically, HCV is classified into six major genotypes, and variants within the genotypes 1, 2, 3, 4, and 6 are further assigned into approximately 80 subtypes (http://hcv.lanl.gov). Geographically, subtypes 1a, 1b, 2a, 2b, 2c, and 3a represent a worldwide epidemic, whilst other subtypes are typically found in restricted geographic areas and represent long-term endemic infection in those regions. Although different HCV genotypes appear to have originated in different regions (genotypes 1, 2, and 4 in Africa, and genotype 3 and 6 in Asia), the global patterns of genotype distribution are changing as a result of modern transmission routes and human migration.

Genetically, the genotype 6 viruses show the highest diversity of HCV. To date, genotype 6 contains 20 (6a–6t) assigned subtypes and many novel variants [Simmonds et al., 2005; Murphy et al., 2007]. These novel variants may represent additional subtypes that are as yet unclassified. Previously, we have characterized the complete genomes of 12 assigned subtypes and three novel variants (km41, km45, and gz52557) within HCV genotype 6, using samples from patients in Southeast Asia or immigrants from this region. These studies constituted what was thought to be a complete panel of entire HCV genomes representing all subtypes of genotype 6 [Lu et al., 2008]. Recently, we obtained partial HCV genomic sequences from injecting drug users (IDUs) in China who were co-infected with HCV and HIV-1. Among these patients, a group of unique HCV genotype 6 variants were discovered, possibly representing a new and previously uncharacterized HCV subtype. Here, the full length genomes of these three variants were completely sequenced and classified as a new HCV genotype 6 subtype designated 6u.

When the current HCV nomenclature was updated in 2005, only 18 assigned subtypes of the six genotypes were confirmed with entire genome sequences. Although genotype 4 contains 19 subtypes, full length genomic data was determined only for the 4a subtype [Simmonds et al., 2005]. Recently, the complete sequences of subtypes 4d and 4f have been described [Hmaied et al., 2007; Timm et al., 2007]. Therefore, they were included in the analysis performed here.

MATERIALS AND METHODS

Serum samples were remained from our previous study [Xia et al., 2008], which were obtained from three male IDUs co-infected with HIV-1. The three IDUs were designated DH012, DH014, and DH028, and at the time of sampling they were 28, 32, and 30 years old, respectively. They lived in Deheng prefecture, Yunnan Province, China, bordering Myanmar. Guidelines set by the local ethical review committees were strictly followed and IRB approval was obtained. From these samples complete HCV genomes were amplified using previously described methods [Li et al., 2006]. Briefly, RNA was extracted using Tripure (Roche, Indianapolis, IN); cDNA was synthesized using random primers (Promega, Madison, WI) and AMV reverse transcriptase (Roche). HCV fragments were amplified using conventional PCR (Roche) with primers listed in Supplementary Table I. The amplicons were directly sequenced, and sequence information was analyzed using the following software: GCG (Wisconsin Sequence Analysis Package; Genetic Computer Group, Madison, Wisconsin, version 10.0), PHYML [Guindon and Gascuel, 2003], MEGA4 [Kumar et al., 2004], RDP2 [Martin et al., 2005], and the Oxford HCV Subtyping Tool (http://www.bioafrica.net/virus-genotype/html/subtypinghcv.html). In more detail, the obtained sequences were aligned using CLUSTAL_X program. Further adjustments to the alignments were manually made using visual correction. Phylogenetic trees were estimated using the maximum-likelihood method under the $HKY+I+\Gamma$ substitution model (gamma distribution approximated by using six rate categories; Hasegawa et al., 1985). The transition/transversion ratio, proportion of invariable sites and gamma distribution shape parameter were estimated from the data. The base frequencies were adjusted to maximize the likelihood. Bootstrap resampling was performed by using 500 neighbor-joining replicates. For a precise classification of the three novel HCV variants, 32 genotype 6 complete genome sequences were included in the phylogenetic analyses. In addition, 16 full length sequences representing all available subtypes for the other five HCV genotypes were included.

RESULTS

The three complete genomes, DH012, DH014, and DH028, were each amplified with eight overlapping fragments, which were 1,243, 1,325, 2,033, 2,012, 2,066, 1,254, 1,032, and 412–439 nt in length. These fragments spanned the three complete genomes that were 9,451, 9,470, and 9,443 nucleotides in length, respectively. The genomes each comprises a 5′UTR of 307 nt, a single ORF of 9,069 nt, and a 3′UTR of variable length (75, 94, 67 nt long, respectively) including a poly (U) tract of 26–57 nt in length. The three genomes shared 10 protein encoding regions with no length variation: Core (573 nt/191 aa), E1 (576 nt/192 aa), E2 (1,110 nt/370 aa), P7 (189 nt/63 aa), NS2 (651 nt/217 aa), NS3 (1,893 nt/631 aa), NS4A (162 nt/54 aa), NS4B (783 nt/261 aa), NS5A (1,356 nt/452 aa), and NS5B (1,776 nt/591 aa). In comparison with the 474 complete HCV genome sequences available in the Los Alamos HCV database (http://hcv.lanl.gov), our three isolates had the longest E2 regions observed to date, only equaled in length by isolate HCV-Tr (D49374/3b).

We calculated pairwise sequence similarities among the three complete genomes, and between these isolates

TABLE I. Ranges of Nucleotide Similarities (%) in Comparing the Three 6u Isolates With 20 Reference Sequences in Different Genomic Regions

Compared[a]	Genome[b]	ORF	5′UTR	Core	E1	E2	P7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	3′UTR[c]
6a/EUHK2 (Y12083)	72.4–72.6	71.6–71.9	96.8–97.1	84.1–84.5	64.6–65.1	67.5–68.9	58.7–59.3	64.4–65.0	74.2–74.6	77.2–77.8	71.3–71.6	65.4–65.7	77.3–77.9	—
6b/Th580 (D84262)	73.0–73.1	72.1–72.3	98.0	83.4–83.6	63.5–63.9	69.3–69.6	59.8–61.4	63.4–63.7	75.2–75.7	75.9	72.2–72.4	67.6–67.7	77.0–77.3	72.3–76.6
6c/Th846 (EF424629)	74.4–74.6	73.6–73.8	98.0–98.3	85.8–86.0	68.4–69.3	69.7–70.5	69.3–70.9	65.9–66.4	76.5–76.9	71.5–72.2	73.7–74.1	67.0–67.1	78.2–78.5	79.7
6d/VN235 (D84263)	74.6–74.8	73.8–74.0	99.0–99.3	84.8–85.2	68.1–68.2	69.9–71.0	70.4–71.4	64.1–64.5	76.9–77.3	71.0	73.8–74.7	67.9–68.9	78.9	78.3
6e/GX004 (DQ314805)	74.6–75.0	73.8–74.2	99.0–99.3	85.2–85.5	67.6–68.1	70.5–71.5	74.6–75.1	62.7–64.2	75.2–75.8	74.1	73.5–73.8	70.0–70.8	78.9–79.0	79.7–81.4
6f/C–0044 (DQ835760)	74.6–74.8	73.8–73.9	97.7–98.0	84.8–85.1	71.5–72.2	69.5–70.3	68.6–70.2	66.2–66.3	75.6–75.9	75.9	73.0–73.5	68.4–69.0	78.6–78.8	84.5–86.2
6g/JK046 (D63822)	73.2	72.2	99.0–99.3	83.4–83.9	69.7–74.0	66.7–67.4	69.3–69.8	63.4–63.9	73.5–73.7	71.0	73.3–73.4	67.8–68.1	77.1–77.2	94.4
6h/VN004 (D84265)	73.3–73.5	72.5–72.7	98.3–98.7	84.5–84.8	68.9–69.1	69.4–70.1	61.9–64.0	63.6–64.5	73.7–74.0	72.2–72.8	73.4	68.5–68.9	76.7–76.9	73.3–75.6
6i/C–0159 (DQ835762)	73.3–73.4	72.5–72.6	98.7–99.0	83.8–84.3	69.3–69.4	68.5–69.5	63.5–65.1	63.2–64.0	74.3–74.4	70.4–71.0	71.5–71.9	68.3–68.5	77.8–77.9	61.0–62.7
6j/C–0667 (DQ835761)	73.9–74.0	73.1–73.2	99.0–99.3	85.1–85.5	70.2–70.8	69.5–70.0	61.9–63.0	64.3–64.7	75.2–75.3	71.4–72.0	73.6–73.8	68.7–69.1	77.2–77.6	75.5–77.4
6k/VN405 (D84264)	74.4–74.5	73.7–73.8	97.3	85.3–85.7	68.7–69.4	69.6–70.1	65.1–65.6	65.0–65.7	76.2–76.6	75.9	72.5–73.2	70.4–70.8	77.3–77.6	66.7–68.5
6l/537796 (EF424628)	74.5–74.6	73.7–73.9	98.7–99.0	85.4–85.6	67.3–68.0	72.1–72.8	60.8–62.4	64.3–64.7	75.1–75.4	72.5	73.2–73.9	69.6–69.8	79.0–79.4	67.8–69.5
6m/C–0208 (DQ835763)	74.2–74.4	73.4–73.6	97.7–98.0	85.6–85.8	68.7–69.1	69.4–69.8	67.7–68.3	66.5–66.6	76.3–76.4	73.5–74.1	72.8–73.2	68.0–68.5	77.6–77.8	76.3
6n/km42 (DQ278894)	74.2–74.3	73.5–73.6	96.7	83.9–84.0	69.0–69.4	69.1–69.4	64.6	63.3–63.4	76.5–76.7	74.7–75.3	72.2–72.5	69.5–69.8	79.1–79.2	68.6–70.6
6o/QC227 (EF424627)	74.7–74.9	73.8–74.0	98.7	83.2–83.5	67.4–68.0	70.2–70.9	69.8–70.9	66.6–67.3	76.2–76.8	69.1	75.2–75.6	68.1–68.4	79.1–79.3	84.7–88.1
6p/QC216 (EF424626)	74.3–74.4	73.4–73.6	98.3–98.7	83.0–83.1	69.7–70.1	69.8–71.7	67.2–68.3	63.5–63.6	75.3–75.7	71.9	74.2–74.6	68.5–69.6	78.4–78.6	78.0–79.7
6q/QC99 (EF424625)	74.3–74.4	73.5–73.6	96.7	85.7–85.8	68.6–69.5	69.5–69.9	73.0–74.7	67.4–67.6	75.3–75.8	69.1–69.8	72.6–73.2	69.0–69.1	77.5–77.6	78.0
6t/VT21 (EF632071)	74.1–74.2	73.3–73.5	98.7	84.6–85.2	67.4–67.8	71.7–72.1	69.8–70.4	64.1–64.6	74.8–75.3	70.4–71.0	72.7–72.8	69.5–70.2	77.2–77.6	59.4–62.5
km41 (DQ278893)	73.9–74.1	73.1–73.3	98.3–98.7	85.3–85.5	68.1–68.6	70.2–70.9	57.7–58.2	65.4–66.8	75.4–75.7	76.4	73.0–73.1	69.3–69.8	76.7–77.0	56.1–58.5
gz52557 (DQ278892)	72.7–72.8	71.8–71.9	98.3–98.7	81.8–82.2	66.1–67.0	66.8–67.6	59.8–60.3	64.4–65.0	73.6–73.8	69.1–69.8	72.0–72.7	69.5–69.6	76.5–76.8	86.5–88.5
DH012 v.s DH014	98.6	98.6	99.7	99.7	99.1	96.8	97.4	98.3	97.9	98.8	99.1	99.0	99.7	96.6
DH012 vs. DH028	97.9	97.8	99.7	99.0	98.4	95.9	98.4	96.9	97.4	99.4	98.3	97.6	99.1	100.0
DH014 vs. DH028	98.1	98.0	100.0	99.3	98.3	95.4	96.8	97.1	98.5	99.4	98.2	98.1	99.9	96.6

[a]Isolates to which the three 6u sequences were pairwise compared. Each subtype was represented by one isolate as indicated with its Genbank accession number shown in parenthesis.
[b]Over the entire genome length.
[c]Y12084 lacks the 3′UTR sequence.

and a panel of 20 genotype 6 reference strains. The latter represent subtypes 6a–6q and 6t, plus two unassigned isolates km41 and gz52557. Compared to each other, DH012, DH014, and DH028 had nucleotide similarities of 97.9–98.6% over the entire genome, indicating these isolates represent a single subtype. When compared to the genotype 6 reference sequences, the similarities were 72.4–75.0% over the entire genome; these similarities fall into a range that indicate different subtypes of the same genotype. The similarities also indicate that the three isolates are approximately equidistant from the 20 reference sequences. Among the 10 protein encoding regions, the core (mean 84.6%) and NS5B (mean 77.9%) regions have the highest similarities, whilst the P7 (mean 66.1%) and NS2 (mean 64.9%) regions have the lowest (Table I). Based on the previously described HCV classification criteria [Simmonds et al., 1996], these similarities indicate that the three isolates (DH012, DH014, and DH028) represent a new genotype 6 subtype.

A phylogenetic tree was estimated using the complete genome sequences (Fig. 1a). DH012, DH014, and DH028 formed a close cluster in the tree with a full (100%) bootstrap support. There is weak bootstrap support (37%) suggesting that the new strains group with a major cluster comprising subtypes 6h, 6j, 6i, 6m, 6n, 6l, 6k, and two unassigned isolates km41 and km45. A second major cluster, containing subtypes 6t, 6q, 6e, 6o, 6p, 6c, 6d, 6f, 6g plus two unassigned isolates gz52557 and D83, has a bootstrap support of 67%. Within this cluster the D83 isolate was recently classified as "6u" subtype [Noppornpanth et al., 2008]. However, it closely groups with the 6e isolate GX004 and this group has a full bootstrap support of 100%. Becasue the designation does not meet the criteria described in the updated HCV nomenclature proposal [Simmonds

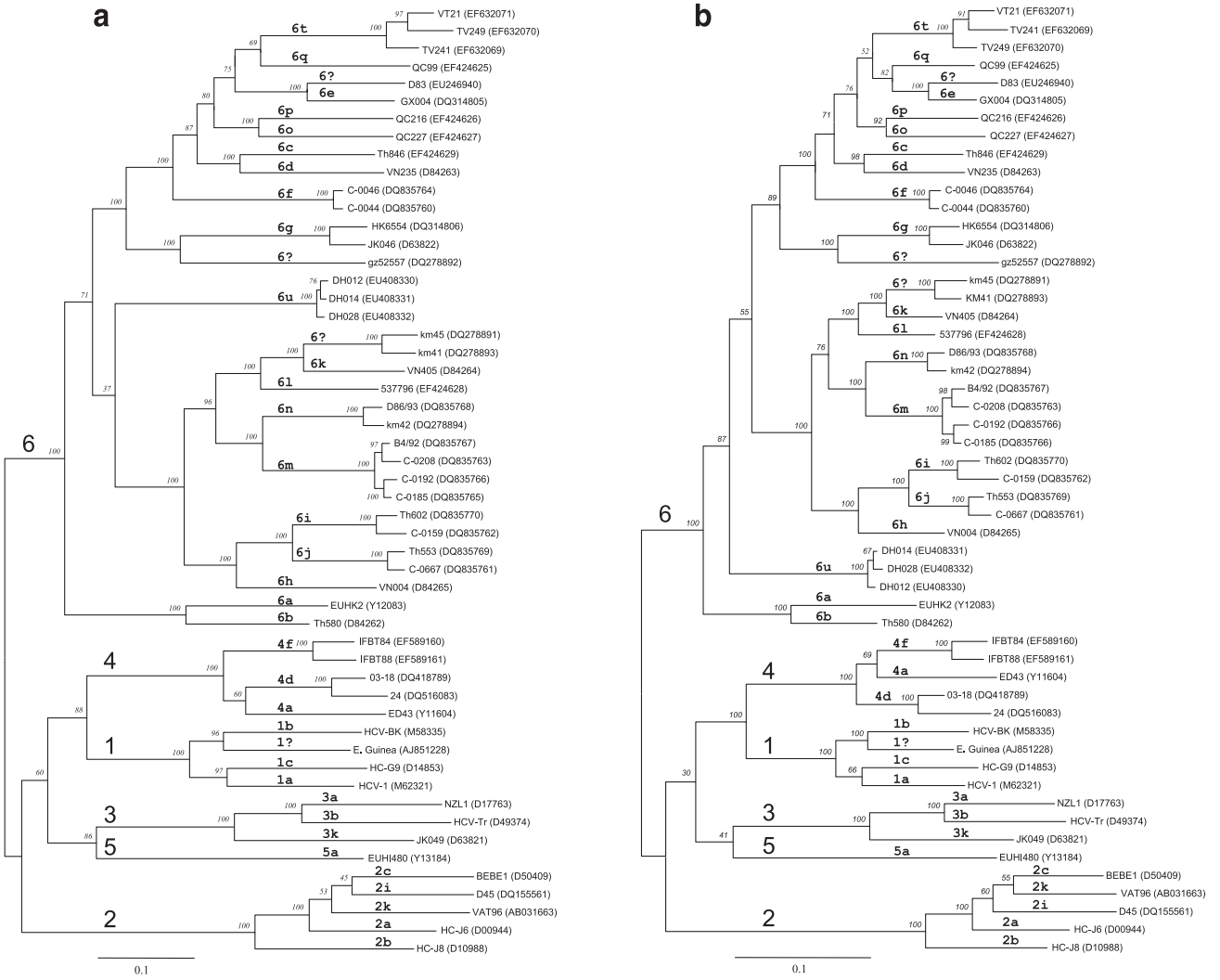

Fig. 1. Phylogenetic tree based on (**a**) complete nucleotide sequences and (**b**) deduced amino acid sequences. The six HCV genotypes are indicated by numbers 1–6; subtypes are designated 1a–6t, 1? and 6? (subtype unassigned). Reference HCV sequences are indicated by an isolate name followed by their Genbank accession number in parenthesis. Bootstrap analysis values are shown in italics. Scale bar represets 0.1 nucleotide or amino acid substitutions per site.

et al., 2005], this D83 isolate was further discussed. A third and distinct cluster contains subtypes 6a and 6b, which finally joined to complete the genotype 6 clade. The internal branch leading to the isolates DH012, DH014, and DH028 is long and these strains do not group closely with any other subtype of lineage. As a result of this, we designate these three isolates as subtype 6u, following from the recently proposed subtypes 6r, 6s, and 6t [Murphy et al., 2007; Lu et al., 2008]. Since three complete genomes have been characterized, this designation can be confirmed, in accordance with the updated HCV nomenclature proposal [Simmonds et al., 2005].

Complete genomes for the other five HCV genotypes were also analyzed. These include four subtypes (1a, 1b, 1c, and 1?) from genotype 1, five subtypes (2a, 2b, 2c, 2i, and 2k) from genotype 2, three subtypes from genotype 3 (3a, 3b, and 3k) and from genptype 4 (4a, 4d, and 4f), and a single subtype (5a) of genotype 5. Subtypes 4d and 4f have only recently been described at the complete genome level [Hmaied et al., 2007; Timm et al., 2007]. The sequence AJ851228 was recently isolated in Equatorial Guinea; it belongs to genotype 1 but no subtype has been assigned to this strain yet [Bracho et al., 2006]. Both genotype 1 and 4 were sister lineages and together they comprise a cluster with a significant bootstrap support of 88%, as previously noted by Salemi and Vandamme [2002]. The cluster containing both genotypes 1 and 4 appears to contain less diversity than the genotype 6 clade. Pairwise comparison revealed nucleotide similarities of 70.7–72.3% between genotype 1 and 4 sequences, whilst comparison of subtype 6a (EUHK2) with 6d (VN235), 6e (GX004), 6j (C–0667), 6t (VT21 and TV249), and gz52557 showed similarities of 71.8–72.0%. These values indicate that the genotype 6 clade and the joint genotype 1 + 4 cluster contain similar levels of diversity and are therefore expected to be of similar ages. Genotypes 1 and 4 represent a major HCV lineage, as perhaps do genotypes 3 and 5. Genotypes 2 and 6 represent two other major lineages. The lineages constitute the diversity of HCV, which includes six major genotypes, over 80 assigned subtypes, many unassigned variants, and, very likely, additional missing genotypes and subtypes (http://hcv.lanl.gov).

A second phylogeny was reconstructed from the translated amino acid sequences (Fig. 1b). It showed a similar topology to Figure. 1a, and the 6u subtype remained a distinct lineage that does not cluster with any other branches. Again, genotypes 1 and 4 can be seen to cluster together with strong bootstrap support (100%). Pairwise amino acid similarities between these two genotypes ranged from 76.7% to 80.3%, while the amino acide similarities of 6a (EUHK2) with C–0667, gz52557, JK046, C–0159 range from 78.9% to 79.3%, again demonstrating that genotype 6 alone is as diverse as the combined genotype 1 + 4 clade.

The DH012, DH014, and DH028 genomes were further analyzed to identify related variants that have only been described with partial sequences. An additional three subtype 6u variants were identified in the Los Alamos HCV database (http://hcv.lanl.gov): MYAN-3E-3 (AB103143 and AB254861), MYAN-TC148 (AB269346 and AB269353), and MYAN-C184 (AB269339). For isolates MYAN-3E-3 and MYAN-TC148, partial sequences were available in both the core and NS5B regions. For MYAN-C184, only a core region sequence was found (tree not shown). The partially-sequenced strains were all isolated from blood donors in Myanmar [Shinji et al., 2004; Lwin et al., 2007] that neighbors the area from which the DH012, DH014, and DH028 samples were collected.

In order to exclude the potential virus recombination, pairwise nucleotide comparisons within short windows (diversity plots) were performed. Comparison of DH012, DH014, and DH028 with 20 reference genotype 6 sequences generated similarity curves with no evidence of viral recombination. In addition, a bootscanning approach was used to search for recombination, as implemented in the Oxford HCV Subtyping Tool, available from http://www.bioafrica.net/virus-genotype/html/subtypinghcv.html. As with the similarity plots no evidence of recombination was found.

DISCUSSION

Isolates of HCV genotype 6 have three unique features. First, they are endemic in Southeast Asia or its vicinity. Second, they represent the most genetically diverse and complex HCV lineage. Third, they are recognized as having the longest evolutionary history [Salemi and Vandamme, 2002]. In the updated HCV nomenclature, HCV genotype 6 is reported to contain 17 assigned subtypes, 6a–6q, all of which have been entirely sequenced [Simmonds et al., 2005; Lu et al., 2007]. Recently, subtypes 6r–6t have been proposed, but only the 6t subtype was validated with complete genomes [Murphy et al., 2007; Lu et al., 2008]. In this report, a new subtype designated 6u was further identified and confirmed with full length genome sequences determined for three different variants. This study has therefore increased the number of genotype 6 subtypes to 21, the genetic diversity of which exceeds any of the other five HCV genotypes.

Analyses indicate that the genotype 6 clade and the joint genotype 1 and genotype 4 clade contain equivalent levels of diversity. This may be explained by the fact that both genotypes 1 and 4 have origins from Central and West Africa [Kamal and Nasser, 2008], and historically they belonged to a single ancient HCV "genotype" as proposed by Ndjomou et al. [2003]. Likewise, almost all genotype 6 isolates are identified in Southeast Asia or immigrants from that region, such that genotype 6 represents another ancient HCV lineage. Comparison of genotype 6 with other genotypes not only demonstrates the high diversity of genotype 6, but also provides an illustration of potential future trajectories of genotype 6 transmission. It is well understood that subtypes 1a and 1b have become distributed worldwide through modern transmission routes [Pybus et al.,

2001]. Recently, a similar pattern of rapid and intercontinental transmission has been repeated by genotype 4 variants, of which 12 subtypes have been found to be prevalent in southern Europe, mainly among immigrants and IDUs [Nicot et al., 2005; Cenci et al., 2007]. Furthermore, a similar trend is perhaps emerging in the US. [Timm et al., 2007]. Genotype 6 variants have become increasingly detected among Asian immigrants in North America [Murphy et al., 2007], and if these variants enter effective transmission networks, such as IDUs, then they may account for a larger percentage of emerging infections in the near future.

Isolates of subtype 6u were derived from IDUs who were co-infected with HIV-1 in Yunnan province, China. This finding indicates that the virus has already entered modern transmission networks. In recent years, Yunnan province has become a center for trafficking drugs from the "Golden Triangle" to the other parts of the world [Beyrer et al., 2000]. One concern is that genotype 6 may spread worldwide through drug trafficking routes, particularly if no effective measures are taken for surveillance. Longitudinally, endemic HCV genotype 6 infections have been maintained in Southeast Asia for more than 1,000 years [Smith et al., 1997], largely due to geographic separation, inefficient transmission, and possibly some degree of population immunity in the endemic region. When introduced to new areas and uninfected populations of high-risk individuals, such as IDUs and those with impaired immunity due to HIV-1 co-infection, significant changes in genotype 6 prevalence and associated disease may be induced.

Although a "6u" isolate D83 has been reported very recently [Noppornpanth et al., 2008], its designation does not meet the criteria described in the updated HCV nomenclature proposal [Simmonds et al., 2005]. The latter requires that at least three isolates should be obtained before a new HCV subtype is designated. In addition, evidence should be provided that the lineage is spreading in a particular transmission network or that the isolates represent separate infected individuals. Furthermore, the sequence data should be sent to one of the three HCV databases for evaluation: the Los Alamos HCV sequence database (http://hcv.lanl.gov), the European HCV Database (http://euhcvdb.ibcp.fr), or the Japan Hepatitis Virus DataBase (http://s2as02.genes.nig.ac.jp). All of the above steps have been undertaken for the three 6u isolates we reported here. Our sequences originated from a recent study of eight IDUs from a small geographic region in China neighboring Myanmar, who were found to be co-infected with HIV-1 and a putative new subtype of HCV [Xia et al., 2008]. These isolates were sent to the Los Alamos HCV sequence database and a confirmed designation of subtype 6u was announced in February 2008 (see http://hcv.lanl.gov/content/sequence/hcv/classfication/genotable.html). In contrast, both the European HCV Database and the Genbank sequence document (EU246940) suggest that the D83 isolate represents a single unassigned HCV genotype 6 variant. At present, no any other closely related variants to the D83 isolate have been identified. D83 is genetically distinct from the three 6u sequences reported here (Fig. 1a). Therefore we conclude that the D83 isolate reamins an unassigned HCV genotype 6 single variant, whilst the three isolates described here should be designated subtype 6u. Nevertheless, an absolute decision will be made by the International HCV Taxonomy Committee.

REFERENCES

Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. 1999. The prevalence of hepatitis C virus infection in the United States, 1988 through1994. N Engl J Med 341:556–562.

Bartenschlager R, Lohmann V. 2000. Replication of hepatitis C virus. J Gen Virol 81:1631–1648.

Beyrer C, Razak MH, Lisam K, Chen J, Lui W, Yu XF. 2000. Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. AIDS 14:75–83.

Bracho MA, Carrillo-Cruz FY, Ortega E, Moya A, González-Candelas F. 2006. A new subtype of hepatitis C virus genotype 1: Complete genome and phylogenetic relationships of an Equatorial Guinea isolate. J Gen Virol 87:1697–1702.

Cenci M, Massi M, Alderisio M, De Soccio G, Recchia O. 2007. Prevalence of hepatitis C virus (HCV) genotypes and increase of type 4 in central Italy: An update and report of a new method of HCV genotyping. Anticancer Res 27:1219–1222.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

Hmaied F, Legrand-Abravanel F, Nicot F, Garrigues N, Chapuy-Regaud S, Dubois M, Njouom R, Izopet J, Pasquier C. 2007. Full-length genome sequences of hepatitis C virus subtype4f. J Gen Virol 88:2985–2990.

Hoofnagle JH. 2002. Course and outcome of hepatitis C. Hepatology 36:S21–S29.

Kamal SM, Nasser IA. 2008. Hepatitis C genotype 4: What we know and what we don't yet know. Hepatology 47:1371–1381.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163.

Li C, Fu Y, Lu L, Ji W, Yu J, Hagedorn CH, Zhang L. 2006. Complete genomic sequences for hepatitis C virus subtypes 6e and 6g isolated from Chinese patients with injection drug use and HIV-1 co-infection. J Med Virol 78:1061–1069.

Lu L, Li C, Fu Y, Gao F, Pybus OG, Abe K, Okamoto H, Hagedorn CH, Murphy D. 2007. Complete genomes of hepatitis C virus (HCV) subtypes 6c, 6l, 6o, 6p and 6q: Completion of a full panel of genomes for HCV genotype 6. J Gen Virol 88:1519–1525.

Lu L, Murphy D, Li C, Liu S, Xia X, Pham PH, Jin Y, Hagedorn CH, Abe K. 2008. Complete genomes of three subtype 6t isolates and analysis of many novel hepatitis C virus variants within genotype 6. J Gen Virol 89:444–452.

Lwin AA, Shinji T, Khin M, Win N, Obika M, Okada S, Koide N. 2007. Hepatitis C virus genotype distribution in Myanmar: Predominance of genotype 6 and existence of new genotype 6 subtype. Hepatol Res 37:337–345.

Major ME, Feinstone SM. 1997. The molecular virology of hepatitis C. Hepatology 25:1527–1538.

Martin DP, Williamson C, Posada D. 2005. RDP2: Recombination detection and analysis from sequence alignments. Bioinformatics 21:260–262.

Murphy DG, Willems B, Deschênes M, Hilzenrat N, Mousseau R, Sabbah S. 2007. Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5′ untranslated region sequences. J Clin Microbiol 45:1102–1112.

Ndjomou J, Pybus OG, Matz B. 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. J Gen Virol 84:2333–2341.

Nicot F, Legrand-Abravanel F, Sandres-Saune K, Boulestin A, Dubois M, Alric L, Vinel JP, Pasquier C, Izopet J. 2005. Heterogeneity of hepatitis C virus genotype 4 strains circulating in south-western France. J Gen Virol 86:107–114.

Noppornpanth S, Poovorawan Y, Lien TX, Smits SL, Osterhaus AD, Haagmans BL. 2008. Complete genome analysis of hepatitis C virus subtypes 6t and6u. J Gen Virol 89:1276–1281.

Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. 2001. The epidemic behavior of the hepatitis C virus. Science 292:2323–2325.

Salemi M, Vandamme AM. 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. J Mol Evol 54:62–70.

Shinji T, Kyaw YY, Gokan K, Tanaka Y, Ochi K, Kusano N, Mizushima T, Fujioka S, Shiraha H. 2004. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. Acta Med Okayama 58:135–142.

Simmonds P, Mellor J, Sakuldamrongpanich T, Nuchaprayoon C, Tanprasert S, Holmes EC, Smith DB. 1996. Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: Comparison with classifications based upon sequence similarity. J Gen Virol 77:3013–3014.

Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. 2005. Consensus proposal for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology 42:962–973.

Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, Yap PL, Simmonds P. 1997. The origin of hepatitis C virus genotypes. J Gen Virol 78:321–328.

Timm J, Neukamm M, Kuntzen T, Kim AY, Chung RT, Brander C, Lauer GM, Walker BD, Allen TM. 2007. Characterization of full-length hepatitis C virus genotype 4 sequences. J Viral Hepat 14: 330–337.

World Health Organization. 1999. Hepatitis C global prevalence. Wkly Epidemiol Rec 74:425–427.

Xia X, Lu L, Tee KK, Zhao W, Wu J, Yu J, Li X, Lin Y, Mukhtar MM, Hagedorn CH, Takebe Y. 2008. The unique HCV genotype distribution and the discovery of a novel subtype 6u among IDUs co-infected with HIV-1 in Yunnan, China. J Med Virol 80:1142–1152.

Zoulim F, Chevallier M, Maynard M, Trepo C. 2003. Clinical consequences of hepatitis C virus infection. Rev Med Virol 13:57–68.