Journal of General Virology (2007), **88**, 1505–1518 **DOI** 10.1099/vir.0.82604-0

Complete genomes for hepatitis C virus subtypes 6f, 6i, 6j and 6m: viral genetic diversity among Thai blood donors and infected spouses

Ling Lu,[1] Chunhua Li,[1,2] Yongshui Fu,[3] Lakkana Thaikruea,[4] Satawat Thongswat,[5] Niwat Maneekarn,[6] Chatchawann Apichartpiyakul,[6] Hak Hotta,[7] Hiroaki Okamoto,[8] Dale Netski,[9,10] Oliver G. Pybus,[11] Donald Murphy,[12] Curt H. Hagedorn[1] and Kenrad E. Nelson[9,10]

Correspondence
Ling Lu
llu@kumc.edu

Kenrad E. Nelson
kenelson@jhsph.edu

[1]Division of Gastroenterology/Hepatology, Department of Medicine, University of Kansas Medical Center, 4035 Delp, MS 1023, Kansas City, KS 66160, USA

[2]The First People's Hospital of Yunnan Province, Kunming, China

[3]Gunagzhou Blood Center, Guangzhou City, Guangdong Province, China

[4]Department of Community Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[5]Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand

[6]Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[7]Department of Microbiology, Kobe University School of Medicine, Chuo-ku, Kobe, Hyogo 650, Japan

[8]Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, Tochigi-Ken, Japan

[9]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[10]Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

[11]Department of Zoology, Oxford University, Oxford, UK

[12]Institut National de Santé Publique du Quebec, Laboratoire de Santé Publique du Quebec, Sainte-Anne-de-Bellevue, Quebec, Canada

In this study, the first complete genome sequences for hepatitis C virus (HCV) subtypes 6f, 6i, 6j and 6m, obtained from infected blood donors in Chiang Mai, Thailand, are reported. Pairwise genome-wide nucleotide similarities between some of these isolates were higher than the 75–80 % value used previously to define different HCV subtypes. To investigate further, the entire genomes of four prototype isolates, Th602 (6i), Th553 (6j), B4/92 (6m) and D86/93 (6n), were sequenced. Pairwise comparison of these sequences gave a similar range of nucleotide similarities, thereby providing new information for HCV subtype classification. In order to study the hypothesis of interspousal HCV transmission, four additional complete HCV genome sequences were obtained from two infected Thai blood donors and their spouses, C-0044 and C-0046 (6f), and C-0192 and C-0185 (6m). Pairwise comparison of the sequences revealed that C-0044 and C-0046 share a nucleotide similarity of 98.1 %, whilst C-0185 and C-0192 have a similarity of 97.8 %. Several other studies of partial HCV sequences of different genomic regions from HCV-infected couples have shown nucleotide similarities ranging from 96.3 to 100 %. The similarities of the complete genome sequences from the two couples in the current study are consistent with HCV transmission between spouses.

Received 29 September 2006
Accepted 25 January 2007

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this study are DQ835760–DQ835770.

Supplementary tables showing PCR primers and amplification strategy, and origin of the retrieved sequences reanalysed are available in JGV Online.

INTRODUCTION

Hepatitis C virus (HCV) is a blood-borne pathogen that infects an estimated 170 million people worldwide (3 % of the global population). HCV infection can be asymptomatic for many years, but in many patients it leads eventually to adverse effects and is a major cause of liver cirrhosis and hepatocellular carcinoma. Alone or in combination with alcoholism, HCV-related end-stage liver disease has become the leading indication for liver transplantation in most USA transplant programmes (Moreno & Berenguer, 2002). HCV is transmitted efficiently through blood transfusion and injection drug use (Murphy *et al.*, 2000). Transmission is also associated with other parenteral exposures, such as tattooing, ear or body piercing, surgery, acupuncture, haemodialysis, occupational needle-stick injuries and other health-care procedures (Alter, 1994; Alter *et al.*, 1990). Although the efficiency of HCV transmission through sexual or household contact is not yet understood entirely (Luksamijarulkul *et al.*, 2000; Romanowski *et al.*, 2003), the observation of closely related HCV sequences obtained from infected couples or sexual partners could be interpreted as evidence of HCV sexual transmission (Capelli *et al.*, 1997; Chayama *et al.*, 1995; Halfon *et al.*, 2001; Healey *et al.*, 1995; Kao *et al.*, 1992, 2000; Komine *et al.*, 1999; Nakayama *et al.*, 2005; Quer *et al.*, 2003; Rice *et al.*, 1993; Romanowski *et al.*, 2003; Ross *et al.*, 1999; Thaikruea *et al.*, 2004; Yagura *et al.*, 2002).

Currently, in most industrialized countries, HCV antibody screening and mini-pool nucleic acid amplification testing (NAT) are mandatory for virtually all blood donations collected. However, in many developing countries, HCV transmission through blood transfusion remains a critical health problem. For example, in Thailand, resources preclude the general implementation of NAT and, in some rural areas, HCV antibody screening of blood donors is not performed consistently (Luksamijarulkul *et al.*, 2004; Wiwanitkit, 2005). The seroprevalence of HCV among the general population in Thailand has been estimated to be approximately 5 % (3.5 million of the total 63 million population) (Songsivilai *et al.*, 1997; Wiwanitkit & Suyaphan, 2002). The group with the highest risk of HCV infection in Thailand is injection drug users, 95 % of whom have been found to be HCV-infected (Luksamijarulkul & Plucktaweesak, 1996). Another important risk group in Thailand is female sex workers, who have been reported to have an infection rate of 9.5 % (Luksamijarulkul & Deangbubpha, 1997). Minority hill-tribe populations in northern Thailand have also been reported to have a high HCV prevalence of approximately 8 % (Wiwanitkit & Suyaphan, 2002). The seroprevalence of HCV among Thai blood donors has increased since 1991 (Nantachit *et al.*, 2003; Songsivilai *et al.*, 1997) and a number of unique HCV subtypes have been identified in Thailand (Thaikruea *et al.*, 2004).

HCV is classified in the genus *Hepacivirus* of the family *Flaviviridae*. It has a single-stranded, positive-sense RNA genome of about 9600 nt in length. The genome contains a single open reading frame (ORF) that encompasses nearly the entire genome range. Flanked by non-coding regions (NCRs) at both 5′ and 3′ ends, this ORF encodes three structural (core, E1, E2) and seven non-structural (P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) proteins. HCV has been classified by phylogenetic analysis into six major genotypes; variants within each genotype are grouped into a number of subtypes. Recently, HCV nomenclature has been revised and 71 subtypes in total were defined. Among these subtypes, 17 have been assigned to genotype 6 (subtypes 6a–6q). Geographically, genotype 6 infections are restricted exclusively to south-eastern Asia or immigrants from this region (Lu *et al.*, 2006).

In one of our previous studies, partial sequences representing six HCV subtypes (6a, 6f, 6i, 6j, 6m and 6n) were obtained from samples of blood donors in Chiang Mai, northern Thailand. Despite the enormous importance of genetic diversity for our understanding of HCV infection and evolution, only eight of the 17 subtypes of genotype 6 have had their whole genomes sequenced to date (subtypes 6a, 6b, 6d, 6e, 6g, 6h, 6k and 6n; Li *et al.*, 2006; Lu *et al.*, 2006; Simmonds *et al.*, 2005). As a step towards a comprehensive understanding of HCV diversity, we have sequenced the entire genomes of HCV isolates of subtypes 6f, 6i, 6j and 6m, and report the data in this paper. By using phylogenetic analysis, we found that subtype 6i was genetically similar to 6j and that subtype 6m was similar to 6n. In addition to the four prototypic isolates (Th602, Th553, B4/92 and D86/93) that represent subtypes 6i, 6j, 6m and 6n, we also sequenced the complete genomes of HCV isolates obtained from two HCV-infected couples (Thaikruea *et al.*, 2004), enabling us to study the possibility of interspousal HCV transmission.

METHODS

Subjects and samples. All of the subjects studied were residents of northern Thailand. Among them, C-0046, C-0159, C-0192, C-0208 and C-0667 were male replacement blood donors. Subject C-0044 was the spouse of donor C-0046; they had lived together since 1992. Subject C-0185 was the spouse of donor C-0192; they had been married since 1978. Donor C-0192 reported a sexual history of 10 female partners and one occasion of sex with a female sex worker. Several features were shared by the other three male blood donors, C-0159, C-0208 and C-0667. None were married, but each had initiated sexual activities when 14–19 years old. They each had a history of sex with female sex workers on seven or eight occasions and had had five to ten female partners during their lifetime. These subjects were also associated with other risk factors, as listed in Table 1. From the remaining four individuals, partial sequences had been determined previously (Apichartpiyakul *et al.*, 1994; Doi *et al.*, 1996; Tokita *et al.*, 1995), which have been used for classifying HCV subtypes 6i, 6j, 6m and 6n, respectively (Simmonds *et al.*, 2005). These individuals included a healthy voluntary blood donor (B4/92), an injection drug user (D86/93) and two patients receiving kidney transplantation (Th553 and Th602).

Sequence amplification and analysis. Complete HCV genomic sequences were each amplified from 100 μl serum by using modified, previously described approaches (Li *et al.*, 2006). Briefly, RNA was extracted by using Tripure (Roche), cDNA was synthesized by using

Table 1. Epidemiological data for the five blood donors and two spouses

	Isolate						
	C-0159	**C-0208**	**C-0667**	**C-0046**	**C-0044**	**C-0192**	**C-0185**
Age (years)	24	34	23	40	39	47	42
Gender	M	M	M	M	F	M	F
Occupation	Military	Employee	Employee	Labourer	Employee	Merchant	Civil servant
Donations (*n*)	4	7	1	8	Spouse	1	Spouse
Reason for donation	Relative	Relative	Relative	Relative	−	Relative	−
Relative with hepatitis	+	−	−	+	−	−	−
History of:							
Jaundice	+	−	−	−	−	+	−
Surgery	−	+	+	−	+	−	+
Suture	+	+	+	−	−	−	+
Transfusion	−	−	+	−	+	−	+
Tattooing	+	−	+	−	−	−	−
Body piercing	+	−	+	−	+	−	+
Injection drug use	+	−	−	−	−	−	−
Shared instrument when snorting drugs	−	−	+	−	−	−	−
No. sexual partners of opposite gender	5	10	5	1	2	10	1
No. times sex with sex workers	8	7	8	0	0	1	0
Age (years) at first time of sex	14	15	19	25	25	15	20

avian myeloblastosis virus reverse transcriptase (Roche) and random primers (Promega), and overlapping fragments were amplified by using conventional PCR (Roche) with the strategies illustrated in Fig. 1 and primers listed in Supplementary Table S1 (available in JGV Online). Standard procedures were adopted to avoid nested RT-PCR false positives. These included at least a negative control, a positive control and a water blank, which were tested during each batch of RNA extraction, reverse transcription and cDNA amplification. All of these steps and the final product-resolving step were completed in different spaces with considerable distance and floor separation. After amplification, the fragments were sequenced directly by using methods described previously (Lu *et al.*, 2006). All sequence information was analysed by using GCG version 10.0 (Wisconsin Sequence Analysis Package; Genetic Computer Group), PHYML (Guindon & Gascuel, 2003) and MEGA3 (Kumar *et al.*, 2004) software. Initially, multiple alignments were performed with PILEUP and adjustments to the alignments were performed with PRETTY. Phylogenetic trees were reconstructed by using the maximum-likelihood method under the HKY+I+Γ 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 neighbour-joining replicates. For comparison between isolates, pairwise nucleotide similarities were calculated by using the MEGA3 software. For the detection and analysis of potential genetic recombination events among all available genotype 6 sequences, the RDP2 software (Recombination Detection Program, version 2) was used (Martin *et al.*, 2005). The program was run by using default settings with the following adjustments: (i) window size was 40 nt, (ii) the option of linear sequences was chosen, (iii) six different methods (RDP, GENECONV, MaxChi, Bootscan, Chimaera and SiScan) were run simultaneously against the multiple sequence alignment, and (iv) listing events detected by more than two methods. In addition, the complete genomes were analysed by using the bootscanning approach implemented in the hepatitis C virus subtyping tool available from http://www.bioafrica.net/virus-genotype/.

RESULTS

Characterization of the complete sequences for C-0159, C-0208 and C-0667

The complete genomic sequences of samples C-0159, C-0208 and C-0667 were amplified and sequenced by using strategies shown in Fig. 1. Their genome lengths and genomic organizations are described in Table 2. Pairwise comparisons of nucleotide sequences over the entire genome and within the ten protein-encoding regions showed that C-0159 was related closely to C-0667, whilst C-0208 was related closely to the reference strain km42. The genome-wide nucleotide similarity between C-0159 and C-0667 was 82.7 %; it was 81.3 % between C-0208 and km42 (Table 3). It has been proposed that HCV genotypes differ by 31–35 % and HCV subtypes by 20–25 % of nucleotides over the entire genome length (Simmonds *et al.*, 1994, 2005). This is consistent with our analysis, which included 189 retrievable complete HCV genomic sequences from GenBank (data not shown). We found that the highest inter-subtype nucleotide similarities were 79.7–80.0 % between subtypes 1a and 1c and 79.5 % between subtypes 6a and 6b. Based on the above criteria, C-0159 and C-0667 qualify as belonging to one subtype, whereas C-0208 and km42 qualify as belonging to another. Phylogenetic analysis was performed with complete genome sequences from 30 reference isolates that

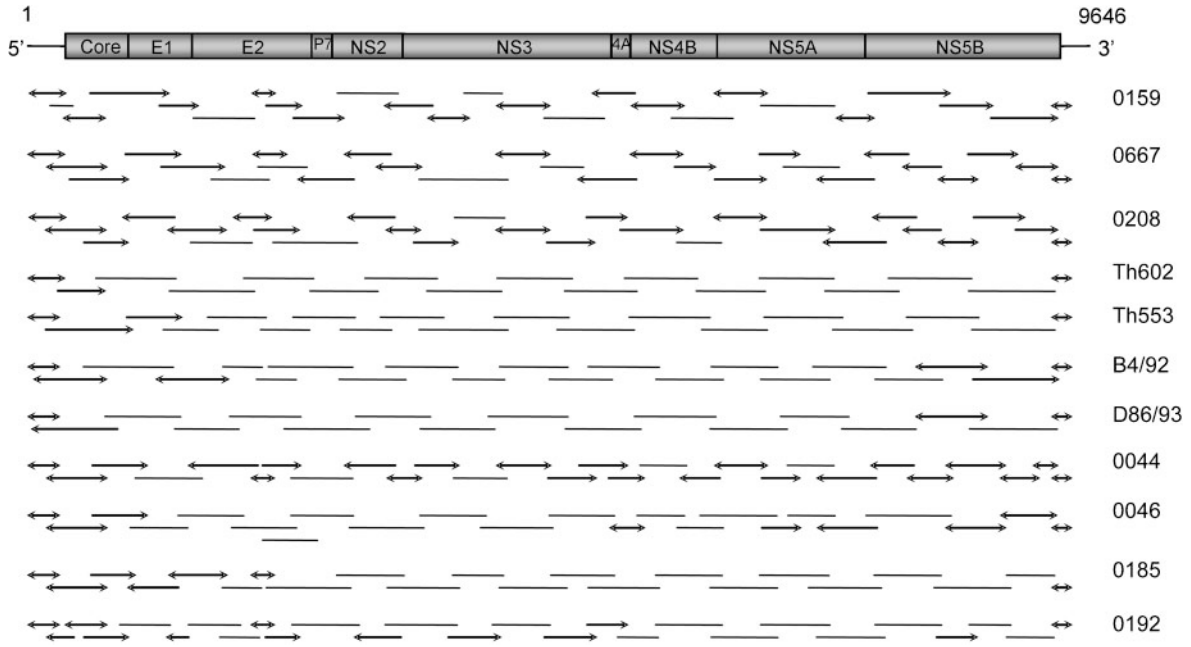

Fig. 1. Strategies used to amplify the 11 complete HCV genomic sequences. The bar at the top represents the genomic organization of HCV and shows the 10 protein-encoding regions of various lengths. Two lines attached to the bar at both sides indicate the 5′ and 3′ NCRs. The nucleotides start at 1 and end at 9646, according to the numbering of the H77 genome (GenBank accession no. NC_004102). Lines and arrows under the bar represent the overlapping fragments amplified for the HCV isolates, with their designations shown on the right. Among them, double-ended arrows identify fragments amplified by using conserved or degenerate primers, single-ended arrows indicate fragments amplified by using conserved or degenerate primers at the arrowed end and strain-specific primers at the other end, and lines without arrowed ends designate fragments amplified by using strain-specific primers.

represented various HCV genotypes and subtypes; the analysis showed that C-0159 clustered with C-0667 and that C-0208 clustered with km42, with each cluster having a bootstrap score of 100 % (Fig. 2).

Characterization of the complete sequences for Th602 and Th553

To test whether C-0159 and C-0667 represent a single HCV subtype or two different ones, complete genomic sequences were subsequently determined for Th602 and Th553 (Fig. 1). These isolates have the same ORF and encoded protein sizes as isolates C-0159 and C-0667 (Table 2). Pairwise comparisons revealed that C-0159 and Th602 had a genome-wide nucleotide similarity of 95.3 %, whilst C-0667 and Th553 had a genome-wide similarity of 96.8 % (Table 3). Phylogenetic analysis demonstrated that Th602 clustered most closely with C-0159, and Th553 clustered most closely with C-0667, with each cluster having a bootstrap value of 100 % (Fig. 2). In previous studies, partial sequences of Th602 had been classified into subtype 9b and partial sequences of Th553 classified into subtype 9c (Tokita *et al.*, 1995). In the recent consensus HCV nomenclature proposal, 9b and 9c correspond to subtypes 6i and 6j, respectively (Simmonds *et al.*, 2005). With the addition of the complete genomic sequences of Th602 and Th553, it is possible to classify C-0159 definitively as subtype 6i, and C-0667 as subtype 6j. Subtypes 6i and 6j have genome-wide nucleotide similarities of 82.7–83.4 %. Although this range includes the typical range of values for a single HCV subtype, 6i and 6j still qualify as two distinct subtypes. This is because the two clusters (6i and 6j) are maintained consistently and significantly in phylogenetic analyses of different genomic regions (trees not shown), complying with the criteria for HCV subtype classification (Simmonds *et al.*, 2005). Furthermore, when 6i and 6j sequences were tested against the complete genome sequences of all available genotype 6 subtypes by using a variety of different statistical approaches, no recombination events were confirmed within the ORFs or between the isolates whose sequences were determined in this study (data not shown). This provides strong evidence that subtypes 6i and 6j are not recombinants of other genotype 6 subtypes.

Characterization of the complete sequences for B4/92 and D86/93

Similarly, C-0208 grouped with km42 and the two isolates appeared to represent another single subtype (Fig. 2). For verification, we determined the entire genomic sequences for B4/92 and D86/93, which were the prototype isolates of

Table 2. Genomic organization of the 11 HCV isolates

Bold text highlights lengths that are slightly different.

Isolate	Amplicons (*n*)*	Length (nt)													
		Genome	ORF	5′ NCR	Core	E1	E2	P7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	3′ NCR	Poly(U)
C-0159	25	9458	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	69	41
C-0208	27	9450	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	61	32
C-0667	27	9442	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	53	28
Th602	17	9447	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	58	27
Th553	19	9454	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	65	32
B4/92	21	9444	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	55	26
D86/93	17	9447	**9048**	338	573	576	**1092**	189	651	1893	162	783	1353	1776	61	32
C-0044	28	9454	**9057**	338	573	576	**1089**	189	651	1893	162	783	**1365**	1776	59	35
C-0046	23	9454	**9057**	338	573	576	**1089**	189	651	1893	162	783	**1365**	1776	59	35
C-0185	22	9449	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	60	30
C-0192	27	9449	9051	338	573	576	1095	189	651	1893	162	783	1353	1776	60	30

*Number of overlapping fragments to cover the complete genomic sequence for the specified HCV isolate.

subtypes 6m and 6n (Simmonds *et al.*, 1996, 2005). The ORF and encoded protein sizes of the two isolates were similar to those of C-0208 and km42, respectively (Table 2). Pairwise comparison revealed that C-0208 and B4/92 had a genome-wide nucleotide similarity of 97.4 %, whilst that between D86/93 and km42 was 93.1 % (Table 3). Phylogenetic analysis showed that B4/92 clustered with C-0208, and D86/93 clustered with km42, with each cluster having a bootstrap value of 100 % (Fig. 2). The pairwise nucleotide similarities of the two clusters were 81.3–81.7 %. Therefore, subtypes 6m and 6n are related closely to each other, but are distinct, a pattern similar to that observed between subtypes 6i and 6j. As before, the distinction of the two subtypes was further confirmed by phylogenetic analyses of different genomic regions and by the lack of evidence for recombination between the complete genomic sequences from all available genotype 6 subtypes (data not shown).

Characterization of the complete sequences for C-0044, C-0046, C-0185 and C-0192

In order to study the possible transmission of HCV between spouses and because no complete genome sequences have been reported for subtype 6f, four isolates from two infected couples were sequenced completely (C-0044, C-0046, C-0185 and C-0192). The sequences of C-0044 and C-0046 were 9454 nt long; those of C-0185 and C-0192 were 9449 nt long. C-0185, C-0192 and C-0208 had the same ORF and encoded protein lengths. However, the C-0044 and C-0046 genomes had smaller E2 regions (1089 nt or 363 aa) and larger NS5A regions (1365 nt or 455 aa) (Table 2). Phylogenetic analysis demonstrated that C-0044 resembled C-0046 closely; both were found on a branch designated subtype 6f. C-0185 resembled C-0192 closely, and they both grouped closely with another subset containing C-0208 and B4/92. The two pairs form a clade of closely related sequences, designated subtype 6m (Fig. 2). In phylogenetic analyses of various genomic regions and in recombination tests against all available complete genotype 6 genome sequences, both C-0044 and C-0046 remained closely related and distinct from all other subtypes. Pairwise comparison showed that C-0044 and C-0046 had a genome-wide nucleotide similarity of 98.1 %, whilst C-0185 and C-0192 had a similarity of 97.8 %. When the complete C-0208 sequence was compared, the nucleotide similarity to C-0185 was 96 % and that to C-0192 was 95.5 % (Table 3).

Based on paired sequences from a single patient with a separation time of 13 years, the rate of HCV evolutionary change was estimated to be 1.92×10^{-3} substitutions per site per year for the Hutchinson strain (Ogata *et al.*, 1991). By using this molecular clock, we estimated the age of the common ancestor of the pairs of sequences sampled from the two infected couples. A nucleotide dissimilarity of 1.8 % between C-0044 and C-0046 corresponds to a common ancestor existing about 4.9 years ago. Likewise, a nucleotide divergence of 2.2 % between C-0185 and C-0192 corresponds to a common ancestor about 5.7 years ago. These estimates are a reasonable reflection of the known epidemiological data: couple C-0044 and C-0046 had been married for 10 years and couple C-0185 and C-0192 had been married for 24 years.

Co-analyses with retrieved partial HCV sequences

The 11 complete genomic sequences of HCV characterized in this study were analysed phylogenetically in four different genomic regions together with many other sequences retrieved from the Los Alamos HCV database (Kuiken *et al.*,

Table 3. Pairwise nucleotide similarities (%) among completely sequenced genotype 6 HCV isolates

Similarities derived from the four prototype sequences are shown in bold. Similarities between isolates from the two infected couples are shown in italics.

Variant	Compared with	Genome	5′ NCR	Core	E1	E2	P7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	3′ NCR
C-0159/6i	EUHK2/6a	72.3	96.4	85.0	62.5	68.4	62.4	69.4	73.8	74.1	70.4	66.5	75.7	–
	Th580/6b	72.6	97.3	87.1	64.9	68.8	67.7	68.2	73.3	77.2	72.5	66.9	73.9	48.9
	VN235/6d	73.5	99.7	84.3	67.9	68.5	65.1	65.7	75.1	68.5	73.8	68.1	76.4	65.2
	GX004/6e	73.8	100.0	85.2	67.6	69.5	64.0	67.4	75.8	71.0	72.3	68.5	76.0	67.2
	HK6554/6g	72.8	99.7	85.3	68.1	67.6	66.7	65.7	73.8	72.8	71.6	67.3	75.5	56.4
	VN004/6h	78.9	99.7	90.2	71.7	74.2	72.5	74.5	79.9	77.2	76.8	74.1	82.4	75.6
	Th602/6i	**95.3**	**100.0**	**97.2**	**93.6**	**92.5**	**91.5**	**93.4**	**95.5**	**95.0**	**96.4**	**94.4**	**97.0**	**89.7**
	VN405/6k	77.1	97.6	87.8	70.3	71.4	69.8	70.9	79.2	79.6	77.4	73.2	78.7	77.8
	km42/6n	77.4	96.7	88.2	70.4	74.2	66.7	71.7	79.1	74.7	76.4	72.3	80.3	80.4
	km41/6?	76.4	99.1	86.2	68.7	71.5	63.0	71.0	77.9	78.3	76.2	72.2	79.2	70.7
	gz52557/6?	73.0	99.1	85.0	67.9	67.0	67.2	66.5	74.6	72.8	71.5	67.4	76.2	61.5
C-0667/6j	EUHK2/6a	71.9	96.8	84.9	65.4	68.4	55.6	68.6	73.4	77.6	69.7	65.5	74.9	–
	Th580/6b	72.8	97.6	86.7	68.0	69.0	57.1	67.5	73.8	79.5	71.4	67.3	74.3	63.8
	VN235/6d	73.6	100.0	84.8	69.4	69.2	67.7	65.6	74.7	73.3	73.6	67.7	76.0	69.6
	GX004/6e	73.6	100.0	86.0	67.8	68.0	65.1	67.0	74.8	71.4	73.7	68.1	76.6	69.8
	HK6554/6g	73.1	99.7	86.7	68.2	70.1	63.5	64.6	73.5	73.3	70.1	68.9	75.5	66.0
	VN004/6h	79.5	99.7	89.0	72.0	74.6	73.0	75.2	81.0	78.9	77.9	74.6	82.9	86.7
	C-0159/6i	82.7	100.0	91.1	77.9	77.1	71.4	79.8	81.4	83.2	81.5	80.8	87.2	81.1
	Th553/6j	**96.8**	**100.0**	**98.8**	**96.8**	**93.5**	**96.3**	**96.6**	**96.4**	**98.1**	**96.5**	**97.8**	**97.4**	**98.1**
	VN405/6k	76.8	97.9	86.7	71.4	71.4	67.2	71.2	78.6	82.0	75.7	72.8	78.6	74.0
	km42/6n	77.6	97.0	87.3	73.2	73.3	70.9	71.6	78.4	77.6	77.7	73.0	80.3	73.7
	km41/6?	76.6	99.1	86.5	69.9	72.4	66.1	71.8	77.8	78.1	76.3	72.1	79.0	70.7
	gz52557/6?	72.9	99.4	83.7	66.5	69.5	61.9	65.3	74.0	73.3	70.3	68.2	75.8	75.0
C-0208/6m	EUHK2/6a	72.5	94.3	83.9	63.2	69.1	57.1	69.2	75.0	74.7	71.5	65.2	76.4	–
	Th580/6b	73.0	95.9	85.1	63.0	68.6	56.1	68.3	75.3	73.5	73.2	67.1	76.2	68.1
	VN235/6d	74.3	97.9	84.2	67.2	71.1	64.0	67.4	75.5	70.4	74.6	68.5	77.6	78.3
	GX004/6e	73.6	98.5	85.8	68.8	69.4	60.3	65.1	74.7	69.8	73.5	68.4	77.0	75.4
	HK6554/6g	73.4	97.9	84.9	67.2	70.5	62.4	66.3	73.4	76.5	73.0	68.2	76.7	70.9
	VN004/6h	77.3	97.9	88.1	69.6	73.7	68.3	73.5	77.1	79.0	77.8	73.1	80.1	75.6
	C-0159/6i	77.3	98.2	86.3	71.5	73.0	64.0	71.2	78.5	76.5	76.8	73.4	80.8	75.4
	C-0667/6j	77.2	98.2	85.1	72.9	73.5	62.4	71.5	78.0	79.5	74.2	74.8	80.3	73.6
	VN405/6k	78.6	96.4	86.9	75.1	73.5	64.6	74.6	80.7	79.6	78.9	74.4	80.0	90.7
	B4/92/6m	**97.4**	**99.4**	**98.9**	**95.1**	**94.7**	**96.3**	**96.0**	**97.9**	**97.5**	**98.6**	**97.8**	**98.1**	**100.0**
	C-0185/6m	96.0	98.8	98.9	93.2	93.5	94.7	93.9	96.8	95.7	95.6	95.9	97.0	98.3
	C-0192/6m	95.5	98.8	98.2	94.1	91.5	93.1	93.3	96.3	96.3	95.4	95.5	96.5	100.0
	km42/6n	81.3	97.0	90.5	75.8	75.8	74.6	76.6	82.7	85.8	79.8	77.5	84.4	88.2
	km41/6?	78.3	98.5	88.3	72.4	72.5	62.4	74.4	80.1	75.8	78.8	73.3	81.3	85.4
	gz52557/6?	73.1	97.6	82.3	69.3	68.3	58.7	64.8	74.6	74.1	71.9	68.1	76.5	75.0
km42/6n	**D86/93/6n**	**96.1**	**98.5**	**97.9**	**95.0**	**93.1**	**93.7**	**94.0**	**96.9**	**98.1**	**96.4**	**96.3**	**96.9**	**98.0**
C-0044/6f	EUHK2/6a	72.2	96.1	86.4	63.3	69.2	61.4	64.2	74.1	69.1	72.8	66.1	75.3	–
	Th580/6b	73.0	96.4	86.2	64.8	68.4	61.9	63.2	75.5	72.8	73.9	67.8	75.3	69.6
	VN235/6d	76.2	98.8	88.5	73.2	68.9	76.2	69.2	78.5	77.2	78.7	71.7	80.9	82.2
	GX004/6e	77.2	99.1	88.7	73.3	69.3	74.6	70.3	77.2	77.2	77.7	72.8	81.4	81.7
	HK6554/6g	74.8	98.8	87.8	71.3	71.3	67.2	65.7	75.2	74.7	72.7	68.3	78.8	83.3
	VN004/6h	73.6	99.4	85.0	69.0	71.9	64.6	65.8	74.8	69.1	73.1	68.0	75.2	75.0
	C-0159/6i	73.4	99.1	84.6	70.6	69.4	65.1	67.2	73.8	72.6	72.9	67.5	76.2	67.2
	C-0667/6j	73.9	99.1	86.5	69.6	69.6	66.1	67.0	74.6	72.0	71.8	69.1	76.4	71.2
	VN405/6k	73.8	97.3	86.4	70.1	69.1	66.0	65.8	75.1	71.0	73.8	68.3	75.8	75.5
	C-0208/6m	74.2	97.3	86.3	69.2	70.1	65.1	65.9	74.9	71.0	75.0	68.8	77.1	76.7
	km42/6n	74.4	95.9	86.5	68.2	70.4	60.8	66.9	75.8	72.2	72.7	69.7	78.1	74.0
	km41/6?	74.2	98.2	85.5	66.2	72.9	66.1	65.3	75.6	69.6	74.9	68.4	76.7	67.5
	gz52557/6?	74.3	98.8	83.2	68.6	67.5	65.6	70.8	76.8	74.7	72.4	68.1	77.4	84.3
C-0044	*C-0046*	*98.1*	*100.0*	*98.1*	*98.3*	*95.8*	*97.3*	*97.5*	*98.1*	*98.8*	*98.2*	*98.7*	*98.8*	*93.2*
C-0185	*C-0192*	*97.8*	*99.1*	*98.9*	*96.9*	*95.5*	*98.4*	*97.5*	*98.2*	*98.1*	*97.6*	*97.8*	*98.5*	*98.3*

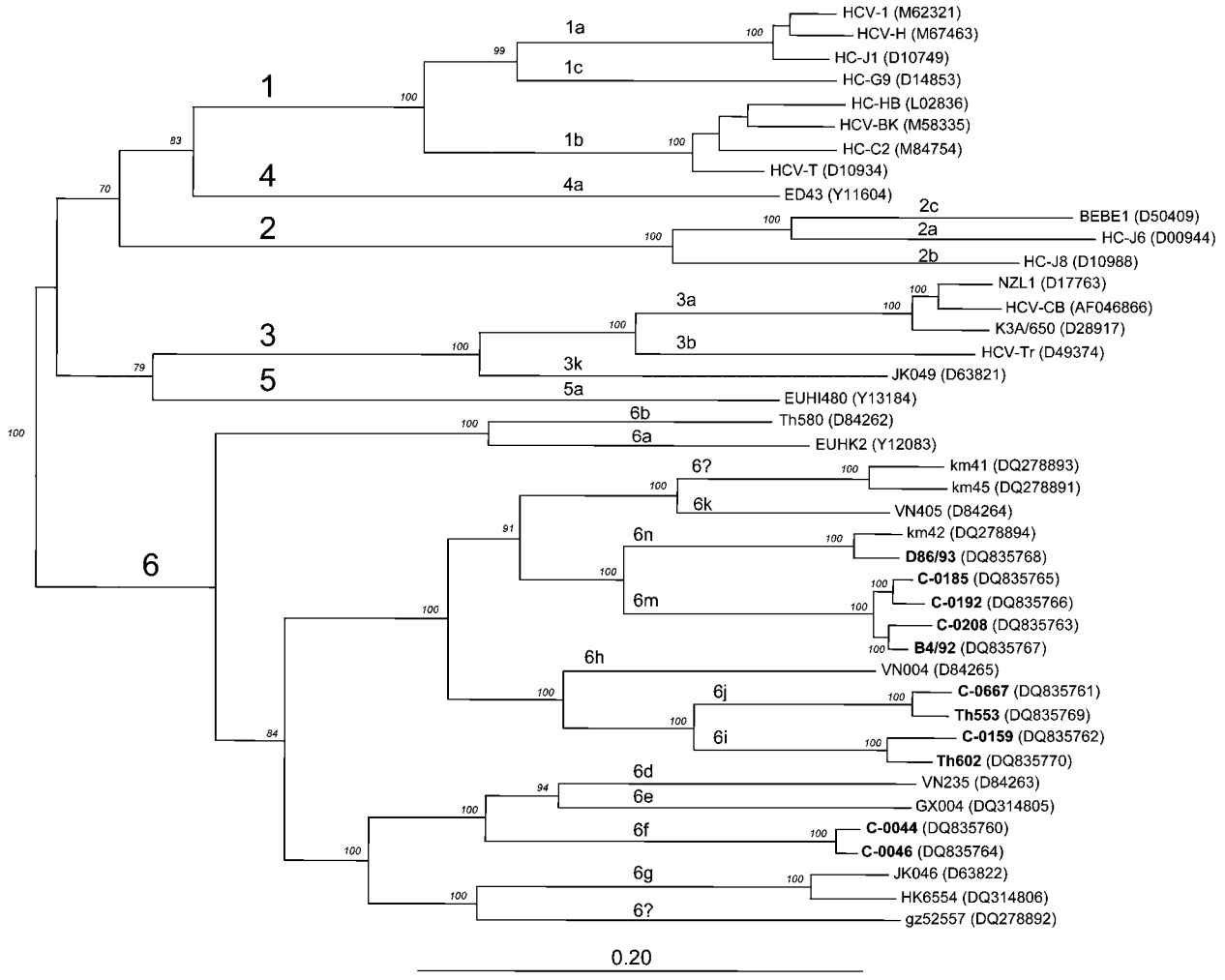

Fig. 2. Phylogenetic tree based on complete HCV genome sequences. The six HCV genotypes are indicated by numbers 1–6; subtypes are designated 1a–6n and 6? (subtype unassigned). Reference HCV sequences are indicated by an isolate name followed by a GenBank accession number in parentheses. The 11 HCV variants sequenced completely in this study are shown in bold. Bootstrap-analysis values are shown in italics. Bar, 0.20 nucleotide substitutions per site.

2005). The four regions included a 302 nt partial core region, a 453 nt partial E1 region and two partial NS5B regions. One of the latter spanned 305 nt in the middle of the NS5B region and the other spanned 366 nt close to the 3′ end. Based on the new consensus numbering system proposed for HCV sequences (Kuiken *et al.*, 2006), these four regions corresponded to nt 376–676, 843–1295, 8282–8586 and 8808–9173, respectively, of the complete H77 genome (GenBank accession no. NC_004102). Phylogenetic trees based on the core (Fig. 3a), E1 (Fig. 3b) and 366 nt NS5B (Fig. 3d) regions demonstrated that the four subtypes 6f, 6i, 6j and 6m each clustered with between two and 22 retrieved sequences, all from Thailand (Mellor *et al.*, 1995, 1996; Simmonds *et al.*, 1996; Sugiyama *et al.*, 1995; Thaikruea *et al.*, 2004; Theamboonlers *et al.*, 2002; Tokita *et al.*, 1995). The tree of the 305 nt NS5B region (Fig. 3c) identified 10 sequences that were from Myanmar, which borders Thailand (Shinji *et al.*, 2004). A similar pattern of geographical distribution is therefore strongly suggested. In order to give a better interpretation of the phylogenies, detailed information about the origin of the reference HCV sequences is summarized in Supplementary Table S2 (available in JGV Online). Although no recombination was detected by using the RDP2 software, clusters representing the four HCV subtypes 6f, 6i, 6j and 6m were positioned in different parts of the phylogenies. This probably reflects a lack of phylogenetic resolution when subgenomic regions are used to estimate trees; although the sequences within each subtype group with high bootstrap support, the bootstrap values for the grouping together of different subtypes are typically much lower (Figs 3, 4).

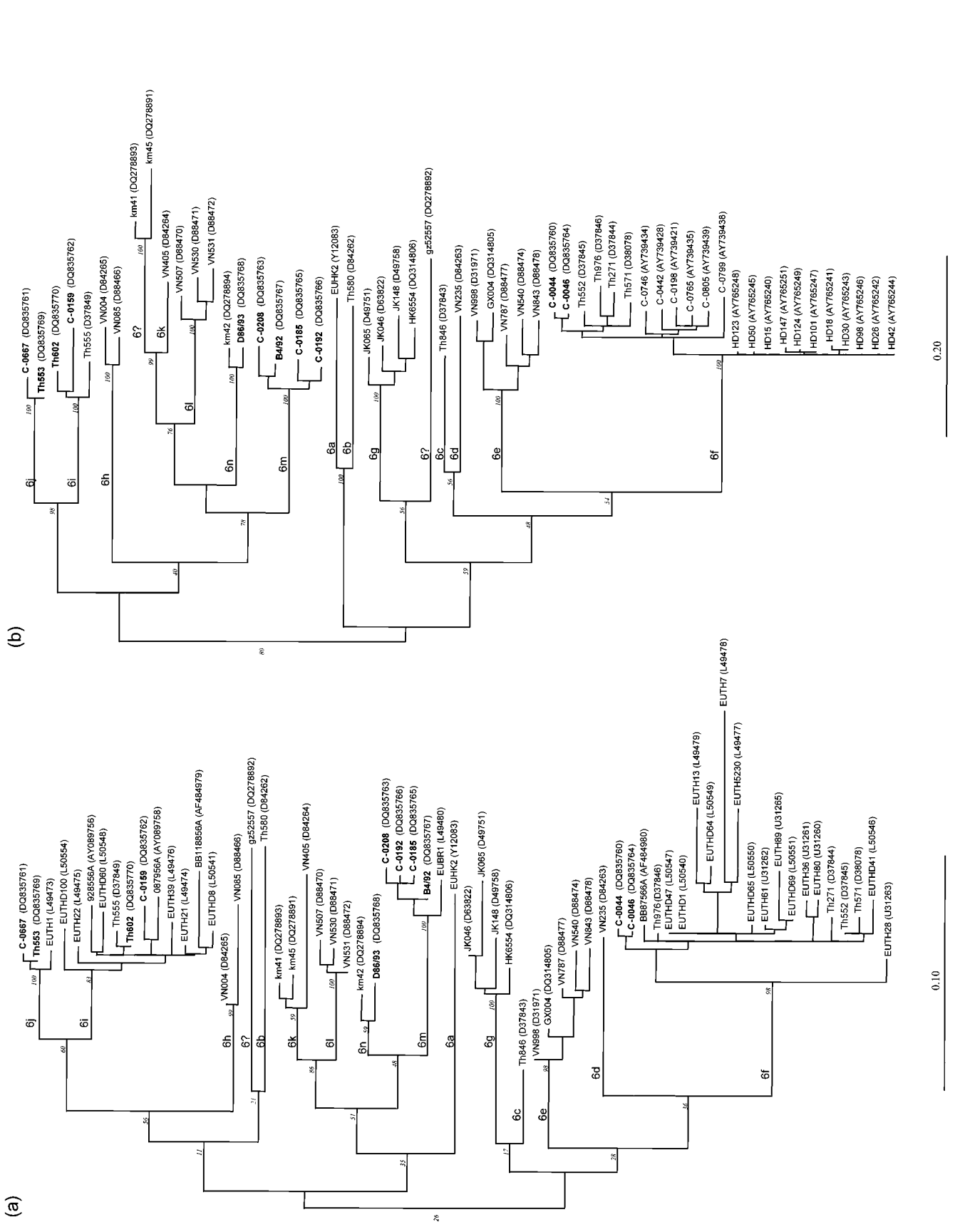
(a)
C-0667 (DQ835761)
Th553 (DQ835769)
EUTH1 (L49473)
EUTHD100 (L50554)
EUTH22 (L49475)
928556A (AY089756)
EUTHD60 (L50548)
Th555 (D37849)
Th602 (DQ835770)
C-0159 (DQ835762)
087956A (AY089758)
EUTH39 (L49476)
EUTH21 (L49474)
BB118856A (AF484979)
EUTHD8 (L50541)
VN004 (D84265)
VN085 (D88466)
6h
6?
g252557 (DQ278892)
Th580 (D84262)
6b
km41 (DQ278893)
km45 (DQ278891)
6k
VN405 (D84264)
VN507 (D88470)
VN530 (D88471)
6l
VN531 (D88472)
km42 (DQ278894)
D86/93 (DQ835768)
6n
C-0208 (DQ835763)
C-0192 (DQ835766)
C-0185 (DQ835765)
6m
B4/92 (DQ835767)
EUBR1 (L49480)
EUHK2 (Y12083)
6a
JK046 (D63822)
JK065 (D49751)
JK148 (D49758)
HK6554 (DQ314806)
6g
Th846 (D37843)
6c
VN998 (D31971)
GX004 (DQ314805)
6e
VN787 (D88477)
VN540 (D88474)
VN843 (D88478)
6d
VN235 (D84263)
C-0044 (DQ835760)
C-0046 (DQ835764)
BB87566A (AF484980)
Th976 (D37846)
EUTHD47 (L50547)
EUTHD1 (L50540)
EUTH13 (L49479)
EUTHD64 (L50549)
EUTH7 (L49478)
EUTH5230 (L49477)
EUTHD65 (L50550)
EUTH61 (U31262)
EUTH89 (U31265)
EUTHD69 (L50551)
EUTH36 (U31261)
EUTH80 (U31260)
Th271 (D37844)
Th552 (D37845)
Th571 (D38078)
EUTHD41 (L50546)
EUTH28 (U31263)
6f
0.10
(b)
C-0667 (DQ835761)
Th553 (DQ835769)
Th602 (DQ835770)
C-0159 (DQ835762)
Th555 (D37849)
VN004 (D84265)
VN085 (D88466)
6?
km41 (DQ278893)
km45 (DQ278891)
6k
VN405 (D84264)
VN507 (D88470)
VN530 (D88471)
6l
VN531 (D88472)
km42 (DQ278894)
D86/93 (DQ835768)
6n
C-0208 (DQ835763)
B4/92 (DQ835767)
C-0185 (DQ835765)
C-0192 (DQ835766)
6m
EUHK2 (Y12083)
Th580 (D84262)
6a
6b
JK065 (D49751)
JK046 (D63822)
JK148 (D49758)
HK6554 (DQ314806)
6g
g252557 (DQ278892)
Th846 (D37843)
6?
6c
VN235 (D84263)
6d
VN998 (D31971)
GX004 (DQ314805)
VN787 (D88477)
6e
VN540 (D88474)
VN843 (D88478)
C-0044 (DQ835760)
C-0046 (DQ835764)
Th552 (D37845)
Th976 (D37846)
Th271 (D37844)
Th571 (D38078)
C-0746 (AY739434)
C-0442 (AY739428)
C-0198 (AY739421)
C-0765 (AY739435)
C-0805 (AY739439)
C-0799 (AY739438)
6f
HD123 (AY765248)
HD50 (AY765245)
HD15 (AY765240)
HD147 (AY765251)
HD124 (AY765249)
HD101 (AY765247)
HD18 (AY765241)
HD30 (AY765243)
HD98 (AY765246)
HD26 (AY765242)
HD42 (AY765244)
0.20

Fig. 3. Phylogenetic trees based on (a) partial core region sequences corresponding to nt 376–676; (b) partial E1 region sequences corresponding to nt 843–1295 of the H77 genome (GenBank accession no. NC_004102). Designations are the same as those described in the legend to Fig. 2. Bootstrap-analysis values are shown in italics.

DISCUSSION

Identification and geographical distribution of subtypes 6f, 6i, 6j and 6m

In this study, we report, for the first time, the complete genomic sequences of four HCV subtypes: 6f, 6i, 6j and 6m. Previously, isolates classified into these four subtypes were only represented by partial sequences. These included 51 isolates of subtype 6f, 15 of 6i, six of 6j and 16 of 6m (see the Los Alamos HCV database; http://hcv.lanl.gov/components/hcv-db/), comprising 88 isolates in total. Of these 88 isolates, 70 were reanalysed selectively in the present study (Figs 3, 4; Supplementary Table S2, available in JGV Online). Historically, subtype 6f has been identified from an original HCV variant BB7 (Apichartpiyakul *et al.*, 1994) and then as a novel genotype VII (Sugiyama *et al.*, 1995), a new genotype 7c (Tokita *et al.*, 1995) and the genotype NG(II) (Mellor *et al.*, 1995, 1996). In a subsequent analysis, related variants were reclassified as subtype 6f and this designation has since been used widely (Simmonds *et al.*, 1996). In our previous studies, 13 isolates of this subtype were determined from haemodialysis patients (GenBank accession numbers AY765240–AY765252) and eight isolates were sequenced from blood donors and a spouse (Thaikruea *et al.*, 2004). The latter included C-0044 and C-0046, which were sequenced completely in this study. Initially, subtype 6i was identified from a new HCV variant, D10/93 (Apichartpiyakul *et al.*, 1994), and then as a novel genotype VIII (Sugiyama *et al.*, 1995), a novel genotype 9b (Tokita *et al.*, 1995) and variants having a restriction fragment-length polymorphism (RFLP) pattern identical to that of subtype 1b (Mellor *et al.*, 1996). In subsequent analysis, the related variants were reclassified as subtype 6i (Simmonds *et al.*, 1996). C-0159 was the sole isolate of this subtype from our previous report (Thaikruea *et al.*, 2004) and was sequenced completely here. For subtype 6j, TH33 was the first identified isolate (Sugiyama *et al.*, 1995). Th553 was initially characterized as belonging to genotype 9c (Tokita *et al.*, 1995). EUTH1 was found to have an RFLP pattern identical to that of subtype 1b (Mellor *et al.*, 1996) and then reclassified into subtype 6j (Simmonds *et al.*, 1996). C-0667 was classified into this subtype (Thaikruea *et al.*, 2004) and has been sequenced completely here. Of subtype 6m, B4/92 was the first isolate (Apichartpiyakul *et al.*, 1994), followed by EUBUR1 (Mellor *et al.*, 1996). They both represented the prototypic sequences for reclassifying subtype 6m (Simmonds *et al.*, 1996). Recently, 11 isolates were characterized among blood donors from Myanmar (Shinji *et al.*, 2004) and three isolates (C-0185, C-0208 and C-0229) were from our previous report (Thaikruea *et al.*, 2004); they all belong to subtype 6m.

The sequences representing the four HCV subtypes 6f, 6i, 6j and 6m are of great interest, because all were identified exclusively in a contiguous geographical region: Thailand and its neighbouring country Myanmar. Common sources of infection were therefore suggested. Epidemiologically, HCV infection has been characterized as having epidemic and endemic patterns. The epidemic pattern is found with HCV genotypes 1a, 1b, 2a, 2b, 2c and 3a. They have spread worldwide over the past 50–70 years as a result of efficient transmission through blood transfusion, injection drug use and unsafe medical practices (Pybus *et al.*, 2001). However, some variants of other genotypes, for example the genotype 4 variants in west-central Africa and genotype 6 in south-eastern Asia, are genetically more diverse, but restricted to a few geographical regions (Ndjomou *et al.*, 2003; Simmonds, 2004). They have been suggested to be endemic with a long-term circulation, a low level of infection and possibly unique patterns of transmission. Eleven genotype 6 isolates that have been sequenced completely in the current study are representatives of such endemic HCV strains.

Subtype classification

A criterion has been proposed that HCV genotypes should differ by 31–33 % and subtypes by 20–25 % of nucleotides over the entire genome length. Moreover, it was found that the nucleotide differences between various subtypes of genotype 6 ranged between 21 and 29 %, with a mean difference of 27 % (Simmonds *et al.*, 1994, 2005). In this study, pairwise comparison of complete sequences revealed that subtypes 6i (C-0159) and 6j (C-0667) only differed by 17.3 % of nucleotides, whereas 6m (C-0208) and 6n (km42) differed by 18.7 %. When the four subtypes were proposed, only partial sequences were available (Simmonds *et al.*, 1996). However, after the representative (C-0159, C-0667, C-0208 and km42) and prototypic (Th602, Th553, B4/92 and D86/93) isolates were sequenced completely, smaller ranges of differences have been obtained. Optimally, precise classification of viral sequences should be based on entire genome sequences. However, complete HCV sequences have been always difficult to obtain, for many reasons. The use of modified procedures in this study has enabled the complete sequencing of 11 isolates, representing four HCV subtypes, each obtained from a serum sample of 100 µl. It has also been recommended recently that accurate HCV classification should be based on extensive phylogenetic analysis, preferably upon analysis of the complete coding region (Simmonds *et al.*, 2005). By following this standard, 6i, 6j, 6m and 6n stand firmly as four different subtypes. This is because they remain consistently and significantly distinct in phylogenetic analyses of entire genome sequences and of partial

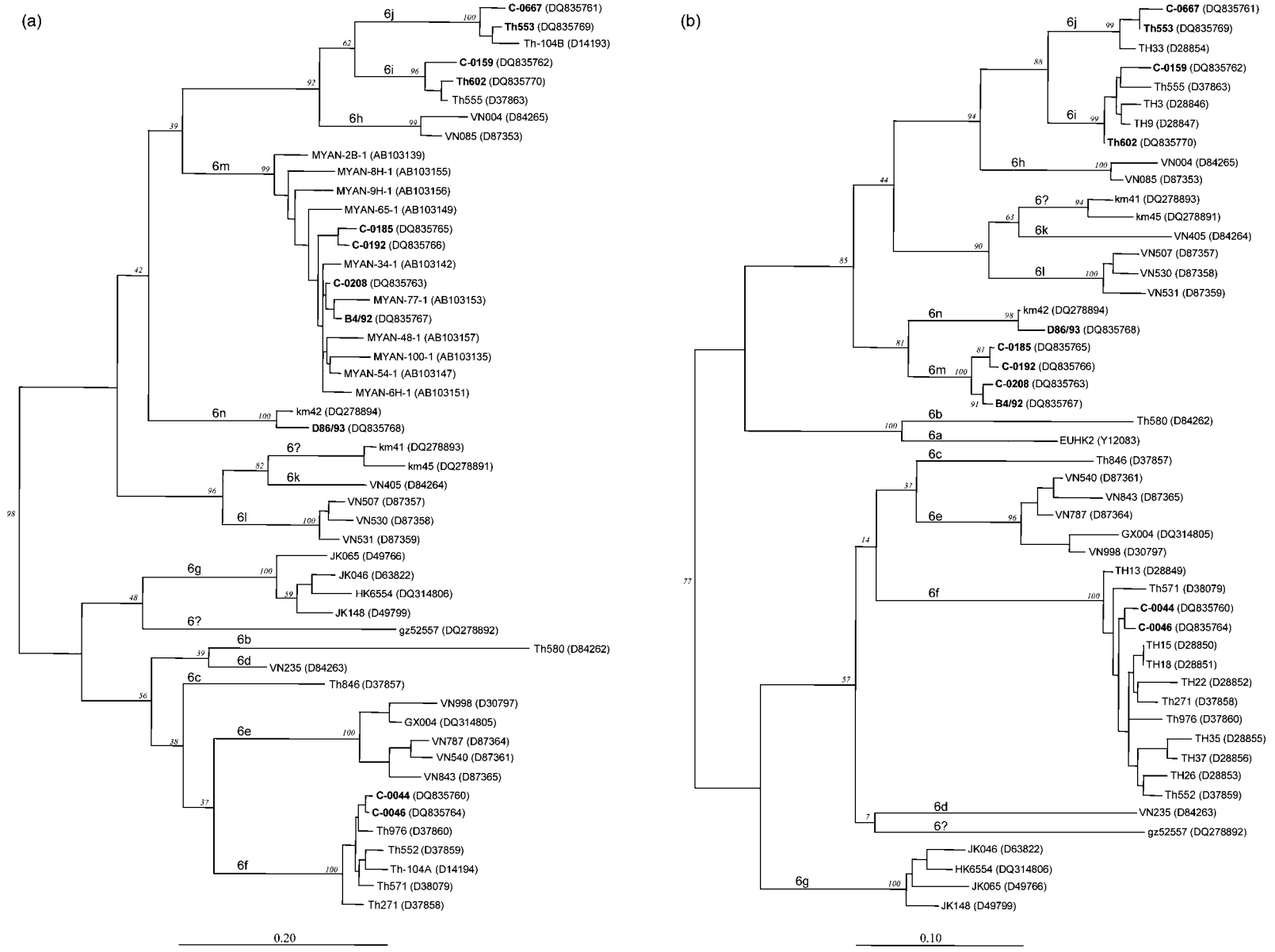

Fig. 4. Phylogenetic trees based on (a) partial NS5B region sequences corresponding to nt 8282–8586; (b) partial NS5B region sequences corresponding to nt 8808–9173 of the H77 genome (GenBank accession no. NC_004102). Designations are the same as those described in the legend to Fig. 2. Bootstrap-analysis values are shown in italics.

subgenomic sequences, and in recombination tests against all other genotype 6 subtypes.

Genetic heterogeneity among related viral variants derives from the accumulation of mutations during long-term virus evolution. Although the number of mutations varies among different genomic regions and according to viral and host selection forces, the process of accumulation of mutations driving virus evolution is continual. Although such differences are continual variables, they do tend to show clear discontinuities between genotypes, subtypes and isolates (Simmonds *et al.*, 1996). The classification of HCV into genotypes and subtypes is most useful when the classification scheme matches the epidemiological processes that generate discontinuities in genetic variation – when genetic diversity thresholds are used to define subtype boundaries, they should be flexible enough to accommodate important epidemiological information. The current HCV classification system has distinguished six HCV genotypes and a great number of subtypes. However, some HCV variants may be characterized poorly at the subtype level, and incomplete sampling could in part explain current discontinuities in HCV genetic diversity. As more such variants are identified and sequenced completely, the continual range of subtypes may become wider, with discontinuities being filled by the novel variants observed.

Sexual and interspousal HCV transmission

Although injection drug use and other parenteral exposures are the most efficient modes of HCV transmission, many studies have suggested that the virus can be transmitted by sexual contact. High prevalence of HCV antibody has been detected among commercial sex workers (Mesquita *et al.*, 1997), homosexual men (Osmond *et al.*, 1993), active heterosexuals (Daikos *et al.*, 1994) and patients attending clinics for sexually transmitted diseases (Thomas *et al.*, 1995). It has also been suggested that persons with multiple partners or those at risk of acquiring sexually transmitted diseases have a higher risk for HCV seroconversion than those in long-term, monogamous partnerships (Terrault, 2002). An explosive AIDS epidemic in northern Thailand was attributed primarily to heterosexual transmission of human immunodeficiency virus and to some unique traditions and cultural practices (Beyrer *et al.*, 2005; Morrison, 2004; Nantachit *et al.*, 2003; Weniger *et al.*, 1991). Transmission of HCV between sexual partners may be common in some populations in this region; this could partly explain the limited geographical distribution of subtypes 6f, 6i, 6j and 6m. Four blood donors, C-0159, C-0192, C-0208 and C-0667, each had a history of sex with female sex workers on one to eight occasions and each had had five to ten female partners. These qualify as risk factors associated significantly with HCV infection among blood donors in Thailand (Thaikruea *et al.*, 2004).

Sexual exposure has also been suggested to be crucial for interspousal transmission of HCV, with a longer duration of marriage being a more evident risk factor (Akahane *et al.*, 1994). However, studies on heterosexual monogamous spouses or partners of patients with hepatitis C have revealed only infrequent interspousal HCV transmission (Stroffolini *et al.*, 2001). Although confounded with other risk factors in some cases, a number of interspousal HCV transmissions have been indicated or confirmed by analyses of partial sequences from the core, E1, E2, NS3 and NS5B regions. These included the short span of hypervariable region 1 (HVR1) to best trace recent HCV transmission events. Nucleotide similarities obtained from the infected couples ranged between 96.3 and 100 %. Within the E1–E2 junction region, including HVR1, the similarities varied from 89.6 to 96.6 % (Capelli *et al.*, 1997; Chayama *et al.*, 1995; Halfon *et al.*, 2001; Healey *et al.*, 1995; Kao *et al.*, 1992, 2000; Komine *et al.*, 1999; Nakayama *et al.*, 2005; Quer *et al.*, 2003; Rice *et al.*, 1993; Romanowski *et al.*, 2003; Ross *et al.*, 1999; Thaikruea *et al.*, 2004; Yagura *et al.*, 2002). Notwithstanding, there is a paucity of complete viral sequences isolated from HCV-infected couples. HCV genotypes that have been analysed for interspousal transmission include 1a, 1b, 2a, 2b and 3a, but there are no reports on other genotypes (Table 4). In our previous study, closely related HCV sequences were identified from five HCV-infected couples (Thaikruea *et al.*, 2004) and HCV genomic sequences from two of these couples were determined completely in this study. By using the rate of accumulation of mutations, the level of HCV genetic divergence was in accordance with the timescale of the cohabitation of the couples studied. The present study provides the first report of complete viral sequences that is consistent with the hypothesis of sexual transmission of HCV. However, genetic data from cohabiting couples cannot by itself determine the actual transmission route definitively. An alternative common source of infection for the couples might be their close but less efficient household contacts or other unrecognized parenteral or blood-to-blood exposures (Alter *et al.*, 1990). Moreover, risk factors other than monogamous contact might be the actual cause, as both of the female spouses claimed histories of surgery, blood transfusion and body piercing. One of them had also had body suture (Table 1). Therefore, it is likely that the direction of HCV transmission was from the female spouses to their husbands. On the other hand, the couples may have acquired HCV infection individually from different sources, with the paired viral sequences appearing highly similar. Infections before marriage are also speculative, but the possibilities of these are small. Couple C-0044 and C-0046 had been married for 10 years, whereas the viruses were estimated to have evolved for approximately 5 years. Couple C-0185 and C-0192 had been married for 24 years, but the accumulation of mutations suggested a common virus ancestor existing 5.7 years ago. Both estimates are consistent with interspousal transmission.

Sequences C-0208 and B4/92 share a genome-wide nucleotide similarity of 97.4 %, slightly lower than that between C-0192 and C-0185 (97.8 %). Cross-contamination and carry-over were thought not to have occurred,

Table 4. Nucleotide similarities (%) of partial HCV sequences isolated from cases of suggested interspousal transmission

Reference	Index	Age (years)	Hepatitis	Suggested source of infection	Genotype	Region(s)	Similarity (mean)	Length (nucleotide position)*
Ross *et al.* (1999)	Female	24	Acute	Male sexual partner	1a	Core	100.0	216 (461–676)
Yagura *et al.* (2002)	Female	63	Acute	Spouse with chronic hepatitis C	1b	Core	98.1	?
						E1	96.3	?
Honda *et al.* (1993)	Spouse	?	?	Spouse?	1b	Core–E1	98.1	220 (840–1055)
Chayama *et al.* (1995)	Five couples	28–64	Chronic	Chronic	1b	E1	97.3–99.8	653 (717–1369)
Komine *et al.* (1999)	Female	21	Acute	Male sexual partner with chronic hepatitis C	2b	E1–E2 (HVR1)	98.0	324 (1293–1616)
Nakayama *et al.* (2005)	Female	65	Acute	Spouse, chronic hepatitis C and hepato-cellular carcinoma	1b	E1/E2 (HVR1)	96.7–99.6 (98.6)	320 (1287–1606)
						NS5B	99.9	1130 (8266–9395)
	Male	65	Acute	Spouse, chronic hepatitis C	1b	E1/E2 (HVR1)	89.6–98.9 (92.1)	320 (1287–1606)
						NS5B	99.1	1130 (8266–9395)
Quer *et al.* (2003)	Female	28	Acute	Partner, chronic hepatitis C	1a	E2	97.4	386 (1642–2085)
Capelli *et al.* (1997)	Female	27	Acute	Male partner, past injection drug use	3a	E2	93.4	176 (1428–1603)
Halfon *et al.* (2001)	Female	32	Acute	Male partner, chronic hepatitis C	3a	HVR1	97.0	79 (1496–1574)
Kao *et al.* (2000)	Female	36	Acute	Spouse, chronic hepatitis C	1b	HVR1	98.0	78 (1496–1573)
Kao *et al.* (1992)	Four spouses	>40		Spouses, chronic hepatitis C	1b	NS3	96.4–100	275 (4073–5047)
Rice *et al.* (1993)	Female	32	Acute	Male sexual partner who shared needles with drug users 10 years ago	2b	NS5B	98.8	166 (8350–8515)
Healey *et al.* (1995)	Female	29	Acute	Male sexual partner with chronic hepatitis C	1a	NS5B	⩽96.4	196 (8342–8538)

*Corresponding to the nucleotide numbering of the HCV genome (GenBank accession no. NC_004102).

because we have adopted strict procedures to avoid nested RT-PCR false positives. We consider it more likely that C-0208 and B4/92 are connected epidemiologically. Although isolate B4/92 was obtained in 1992 and C-0185 was collected in 2002, both were from blood donors recruited in the same hospital (Apichartpiyakul *et al.*, 1994; Doi *et al.*, 1996; Thaikruea *et al.*, 2004). However, as little information was available regarding the B4/92 sample, the true relationship between the two donors cannot be ascertained in this study.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grants 5P20 RR016443-04, 5U01-DA13032-5 and CA063640.

REFERENCES

Akahane, Y., Kojima, M., Sugai, Y., Sakamoto, M., Miyazaki, Y., Tanaka, T., Tsuda, F., Mishiro, S., Okamoto, H. & other authors (1994). Hepatitis C virus infection in spouses of patients with type C chronic liver disease. *Ann Intern Med* **120**, 748–752.

Alter, M. J. (1994). Transmission of hepatitis C virus – route, dose, and titer. *N Engl J Med* **330**, 784–786.

Alter, M. J., Hadler, S. C., Judson, F. N., Mares, A., Alexander, W. J., Hu, P. Y., Miller, J. K., Moyer, L. A., Fields, H. A. & other authors (1990). Risk factors for acute non-A, non-B hepatitis in the United States and association with hepatitis C virus infection. *JAMA* **264**, 2231–2235.

Apichartpiyakul, C., Chittivudikarn, C., Miyajima, H., Homma, M. & Hotta, H. (1994). Analysis of hepatitis C virus isolates among healthy blood donors and drug addicts in Chiang Mai, Thailand. *J Clin Microbiol* **32**, 2276–2279.

Beyrer, C., Sripaipan, T., Tovanabutra, S., Jittiwutikarn, J., Suriyanon, V., Vongchak, T., Srirak, N., Kawichai, S., Razak, M. H. & Celentano, D. D. (2005). High HIV, hepatitis C and sexual risks among drug-using men who have sex with men in northern Thailand. *AIDS* **19**, 1535–1540.

Capelli, C., Prati, D., Bosoni, P., Zanuso, F., Pappalettera, M., Mozzi, F., De Mattei, C., Zanella, A. & Sirchia, G. (1997). Sexual transmission of hepatitis C virus to a repeat blood donor. *Transfusion* **37**, 436–440.

Chayama, K., Kobayashi, M., Tsubota, A., Koida, I., Arase, Y., Saitoh, S., Ikeda, K. & Kumada, H. (1995). Molecular analysis of intraspousal transmission of hepatitis C virus. *J Hepatol* **22**, 431–439.

Daikos, G. L., Lai, S. & Fischl, M. A. (1994). Hepatitis C virus infection in a sexually active inner city population. The potential for heterosexual transmission. *Infection* **22**, 72–76.

Doi, H., Apichartpiyakul, C., Ohba, K. I., Mizokami, M. & Hotta, H. (1996). Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand, and identification of novel subtypes of HCV major type 6. *J Clin Microbiol* **34**, 569–574.

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

Halfon, P., Riflet, H., Renou, C., Quentin, Y. & Cacoub, P. (2001). Molecular evidence of male-to-female sexual transmission of hepatitis C virus after vaginal and anal intercourse. *J Clin Microbiol* **39**, 1204–1206.

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.

Healey, C. J., Smith, D. B., Walker, J. L., Holmes, E. C., Fleming, K. A., Chapman, R. W. & Simmonds, P. (1995). Acute hepatitis C infection after sexual exposure. *Gut* **36**, 148–150.

Honda, M., Kaneko, S., Unoura, M., Kobayashi, K. & Murakami, S. (1993). Risk of hepatitis C virus infections through household contact with chronic carriers: analysis of nucleotide sequences. *Hepatology* **17**, 971–976.

Kao, J. H., Chen, P. J., Yang, P. M., Lai, M. Y., Sheu, J. C., Wang, T. H. & Chen, D. S. (1992). Intrafamilial transmission of hepatitis C virus: the important role of infections between spouses. *J Infect Dis* **166**, 900–903.

Kao, J. H., Liu, C. J., Chen, P. J., Chen, W., Lai, M. Y. & Chen, D. S. (2000). Low incidence of hepatitis C virus transmission between spouses: a prospective study. *J Gastroenterol Hepatol* **15**, 391–395.

Komine, F., Yamaguchi, T., Moriyama, M., Segawa, T., Matsumura, H., Nakai, K., Shimizu, T., Hayashi, N., Arakawa, Y. & other authors (1999). A case of sexually transmitted acute hepatitis C: confirmation by analysis of viral genome. *Hepatol Res* **14**, 84–91.

Kuiken, C., Yusim, K., Boykin, L. & Richardson, R. (2005). The Los Alamos hepatitis C sequence database. *Bioinformatics* **21**, 379–384.

Kuiken, C., Combet, C., Bukh, J., Shin-I, T., Deleage, G., Mizokami, M., Richardson, R., Sablon, E., Yusim, K. & other authors (2006). A comprehensive system for consistent numbering of HCV sequences, proteins and epitopes. *Hepatology* **44**, 1355–1361.

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, C. H. & 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., Nakano, T., Li, C., Fu, Y., Miller, S., Kuiken, C., Robertson, B. H. & Hagedorn, C. H. (2006). Hepatitis C virus complete genome sequences identified from China representing subtypes 6k and 6n and a novel, as yet unassigned subtype within genotype 6. *J Gen Virol* **87**, 629–634.

Luksamijarulkul, P. & Deangbubpha, A. (1997). Hepatitis C antibody prevalence and risk factors of some female sex workers in Thailand. *Southeast Asian J Trop Med Public Health* **28**, 507–512.

Luksamijarulkul, P. & Plucktaweesak, S. (1996). High hepatitis C seroprevalence in Thai intravenous drug abusers and qualitative risk analysis. *Southeast Asian J Trop Med Public Health* **27**, 654–658.

Luksamijarulkul, P., Khemnak, P. & Pacheun, O. (2000). Human immunodeficiency virus and hepatitis C virus infections among patients attending sexually transmitted disease clinics, Regional 2, Thailand. *Asia Pac J Public Health* **12**, 41–45.

Luksamijarulkul, P., Thammata, N., Sujirarat, D. & Tiloklurs, M. (2004). Hepatitis C virus infection among Thai blood donors: antibody prevalence, risk factors and development of risk screening form. *Southeast Asian J Trop Med Public Health* **35**, 147–154.

Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.

Mellor, J., Holmes, E. C., Jarvis, L. M., Yap, P. L. & Simmonds, P. (1995). Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J Gen Virol* **76**, 2493–2507.

Mellor, J., Walsh, E. A., Prescott, L. E., Jarvis, L. M., Davidson, F., Yap, P. L. & Simmonds, P. (1996). Survey of type 6 group variants of hepatitis C virus in Southeast Asia by using a core-based genotyping assay. *J Clin Microbiol* **34**, 417–423.

Mesquita, P. E., Granato, C. F. & Castelo, A. (1997). Risk factors associated with hepatitis C virus (HCV) infection among prostitutes and their clients in the city of Santos, Sao Paulo State, Brazil. *J Med Virol* **51**, 338–343.

Moreno, R. & Berenguer, M. (2002). Hepatitis C and liver transplantation. *Ann Hepatol* **1**, 129–135.

Morrison, L. (2004). Traditions in transition: young people's risk for HIV in Chiang Mai, Thailand. *Qual Health Res* **14**, 328–344.

Murphy, E. L., Bryzman, S. M., Glynn, S. A., Ameti, D. I., Thomson, R. A., Williams, A. E., Nass, C. C., Ownby, H. E., Schreiber, G. B. & other authors (2000). Risk factors for hepatitis C virus infection in United States blood donors. NHLBI Retrovirus Epidemiology Donor Study (REDS). *Hepatology* **31**, 756–762.

Nakayama, H., Sugai, Y., Ikeya, S., Inoue, J., Nishizawa, T. & Okamoto, H. (2005). Molecular investigation of interspousal transmission of hepatitis C virus in two Japanese patients who acquired acute hepatitis C after 40 or 42 years of marriage. *J Med Virol* **75**, 258–266.

Nantachit, N., Robison, V., Wongthanee, A., Kamtorn, N., Suriyanon, V. & Nelson, K. E. (2003). Temporal trends in the prevalence of HIV and other transfusion-transmissible infections among blood donors in northern Thailand, 1990 through 2001. *Transfusion* **43**, 730–735.

Ndjomou, J., Pybus, O. G. & 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.

Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991). Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci U S A* **88**, 3392–3396.

Osmond, D. H., Charlebois, E., Sheppard, H. W., Page, K., Winkelstein, W., Moss, A. R. & Reingold, A. (1993). Comparison of risk factors for hepatitis C and hepatitis B virus infection in homosexual men. *J Infect Dis* **167**, 66–71.

Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C. & Harvey, P. H. (2001). The epidemic behavior of the hepatitis C virus. *Science* **292**, 2323–2325.

Quer, J., Murillo, P., Esteban, J. I., Martell, M., Esteban, R. & Guardia, J. (2003). Sexual transmission of hepatitis C virus from a patient with chronic disease to his sex partner after removal of an intrauterine device. *Sex Transm Dis* **30**, 470–471.

Rice, P. S., Smith, D. B., Simmonds, P. & Holmes, E. (1993). Heterosexual transmission of hepatitis C virus. *Lancet* **342**, 1052–1053.

Romanowski, B., Preiksaitis, J., Campbell, P. & Fenton, J. (2003). Hepatitis C seroprevalence and risk behaviors in patients attending sexually transmitted disease clinics. *Sex Transm Dis* **30**, 33–38.

Ross, R. S., Viazov, S., Varenholz, C. & Roggendorf, M. (1999). Inquiries on intraspousal transmission of hepatitis C virus: benefits and limitations of genome sequencing and phylogenetic analysis. *Forensic Sci Int* **100**, 69–76.

Shinji, T., Kyaw, Y. Y., Gokan, K., Tanaka, Y., Ochi, K., Kusano, N., Mizushima, T., Fujioka, S., Shiraha, H. & other authors (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. (2004). Genetic diversity and evolution of hepatitis C virus – 15 years on. *J Gen Virol* **85**, 3173–3188.

Simmonds, P., Alberti, A., Alter, H. J., Bonino, F., Bradley, D. W., Brechot, C., Brouwer, J. T., Chan, S. W., Chayama, K. & Chen, D. S. (1994). A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology* **19**, 1321–1324.

Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, S., Holmes, E. C. & Smith, D. B. (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. & other authors (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**, 962–973.

Songsivilai, S., Jinathongthai, S., Wongsena, W., Tiangpitayakorn, C. & Dharakul, T. (1997). High prevalence of hepatitis C infection among blood donors in northeastern Thailand. *Am J Trop Med Hyg* **57**, 66–69.

Stroffolini, T., Lorenzoni, U., Menniti-Ippolito, F., Infantolino, D. & Chiaramonte, M. (2001). Hepatitis C virus infection in spouses: sexual transmission or common exposure to the same risk factors? *Am J Gastroenterol* **96**, 3138–3141.

Sugiyama, K., Kato, N., Nakazawa, T., Yonemura, Y., Phornphutkul, K., Kunakorn, M., Petchclai, B. & Shimotohno, K. (1995). Novel genotypes of hepatitis C virus in Thailand. *J Gen Virol* **76**, 2323–2327.

Terrault, N. A. (2002). Sexual activity as a risk factor for hepatitis C. *Hepatology* **36**, S99–S105.

Thaikruea, L., Thongsawat, S., Maneekarn, N., Netski, D., Thomas, D. L. & Nelson, K. E. (2004). Risk factors for hepatitis C virus infection among blood donors in northern Thailand. *Transfusion* **44**, 1433–1440.

Theamboonlers, A., Chinchai, T., Bedi, K., Jantarasamee, P., Sripontong, M. & Poovorawan, Y. (2002). Molecular characterization of hepatitis C virus (HCV) core region in HCV-infected Thai blood donors. *Acta Virol* **46**, 169–173.

Thomas, D. L., Zenilman, J. M., Alter, H. J., Shih, J. W., Galai, N., Carella, A. V. & Quinn, T. C. (1995). Sexual transmission of hepatitis C virus among patients attending sexually transmitted diseases clinics in Baltimore – an analysis of 309 sex partnerships. *J Infect Dis* **171**, 768–775.

Tokita, H., Okamoto, H., Luengrojanakul, P., Vareesangthip, K., Chainuvati, T., Iizuka, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1995). Hepatitis C virus variants from Thailand classifiable into five novel genotypes in the sixth (6b), seventh (7c, 7d) and ninth (9b, 9c) major genetic groups. *J Gen Virol* **76**, 2329–2335.

Weniger, B. G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., Uneklabh, T., Thongcharoen, P. & Wasi, C. (1991). The epidemiology of HIV infection and AIDS in Thailand. *AIDS* **5** (*Suppl. 2*), S71–S85.

Wiwanitkit, V. (2005). Anti HCV seroprevalence among the voluntary blood donors in Thailand. *Hematology* **10**, 431–433.

Wiwanitkit, V. & Suyaphan, A. (2002). A note in the high prevalence of anti-HCV seropositivity among hilltribers in Mae Jam District, Northern Thailand. *Viral Immunol* **15**, 645–646.

Yagura, M., Fukui, H., Murai, S., Tokita, H., Kamitsukasa, H., Harada, H. & Sumida, H. (2002). Intraspousal transmission of hepatitis C virus occurring after 42 years of marriage: no relation to HCV RNA titers. *Am J Gastroenterol* **97**, 2925–2927.