A novel hepatitis C virus (HCV) subtype from Somalia and its classification into HCV clade 3

Karim Abid,[1] Rafael Quadri,[1] Anne-Lise Veuthey,[2] Antoine Hadengue[1] and Francesco Negro[1,3]

[1] Division of Gastroenterology and Hepatology, University Hospital, 24 rue Micheli-du-Crest, 1211 Geneva, Switzerland

[2,3] Swiss Institute of Bioinformatics[2] and Division of Clinical Pathology[3], CMU, 1 rue Michel-Servet, 1211 Geneva, Switzerland

Hepatitis C virus (HCV) sequences from throughout the world have been grouped into six clades, based on recently proposed criteria. Here, the partial sequences and clade assignment are reported for three HCV isolates from chronic hepatitis C patients from Somalia, for whom conventional assays failed to identify the genotype. Phylogenetic analysis of the sequences of the core, envelope 1 and part of the non-structural 5b regions suggests that all three isolates belong to a distinct HCV genetic group, tentatively classified as subtype 3h. This novel HCV subtype shows the highest sequence similarity with HCV isolates from Indonesia. Despite the fact that these patients were infected with HCV clade 3, none of them responded to standard interferon treatment.

Introduction

Hepatitis C virus (HCV) is a common human pathogen with a single-stranded RNA genome about 9600 nt in length (Choo *et al.*, 1991). The sequence diversity among different HCV isolates has been recognized since early studies and can reach as much as 30% nucleotide sequence divergence over the entire genome. Different HCV isolates cluster into a two-tiered classification, in that the major genetic groups, referred to as clades, are further divided into several subtypes (Robertson *et al.*, 1998). This HCV diversification is the result of the accumulation of mutations due to the lack of proof-reading activity of the RNA-dependent RNA polymerase and several selective pressures, among which epidemiological factors play a major role.

Specific criteria have been established to standardize and facilitate type assignment once new sequences are obtained from HCV isolates across the world. According to these criteria, six clades are now recognized, which correspond to the former genotypes 1–6; genotypes 7–9 and 11 (Tokita *et al.*, 1994a, 1995) have been reassigned to clade 6, and genotype 10 (Tokita *et al.*, 1996) has been reassigned to clade 3, based on phylogenetic analysis rather than sequence identity (Mizokami *et al.*, 1996; Simmonds *et al.*, 1996; de Lamballerie *et al.*,

1997; Robertson *et al.*, 1998). Although the classification of new HCV sequences should be preferably based on complete genome sequences, tentative clade/subtype assignments can be made based on phylogenetic analysis of nucleotide sequences of at least two coding regions (Robertson *et al.*, 1998). The precise subtype assignment of HCV isolates has not only taxonomic consequences, but may help in guiding the clinician in the decision-making process, especially before antiviral treatment (Poynard *et al.*, 1998; McHutchison *et al.*, 1998). Although the gold standard in type assignment is sequence analysis, the more widely used assays currently available in the clinical setting are a line-probe assay (INNO-LIPA; Innogenetics), RFLP analysis of sequences in the 5′ UTR of the HCV genome (Davidson *et al.*, 1995), and an RT–PCR assay amplifying the nucleocapsid-encoding region of HCV in a type-dependent manner (Okamoto *et al.*, 1992). These assays allow the correct identification of the HCV genotype (and in some cases the subtype) in more than 90% of cases. Controversies over the clade determination in some cases can be solved by direct sequencing.

We report here the sequence and phylogenetic analyses of three HCV isolates from Somalia, for whom genotype assignment gave conflicting results based on two of the above-mentioned conventional techniques.

Methods

■ **Patients.** Three patients from Somalia were found to be anti-HCV-positive by a second generation ELISA (Abbott) on their arrival in Switzerland. Patient 1 was a 43-year-old man with no risk factors for

Author for correspondence: Francesco Negro (at University Hospital). Fax +41 22 3729366.

e-mail Francesco.Negro @dim.hcuge.ch

HCV transmission. The liver biopsy showed mild chronic hepatitis with no fibrosis (A2F0 according to the METAVIR scoring system) (Bedossa & Poynard, 1996). Serum HCV RNA level (Amplicor Monitor HCV; Hoffmann-La Roche) was 8.6×10^6 genome equivalents/ml.

Patient 2 was a 57-year-old man without risk factors for HCV infection and had, at liver biopsy, moderate necroinflammatory activity with portal, focally extensive fibrosis (A2F2). Serum HCV RNA level was 1.6×10^6 genome equivalents/ml.

Patient 3, a 50-year-old woman, was a registered nurse in Mogadishu. In 1977, on occasion of her first delivery, she had received a blood transfusion. There were no other risk factors for HCV infection, but she reported that one of her sisters had died of an unspecified liver disease. She underwent a liver biopsy showing the presence of moderate necroinflammation with portal, focally extensive fibrosis (A2F2). Her serum HCV RNA level was 2.235×10^5 genome equivalents/ml.

All three patients were treated with α-IFN (3 million units, three times a week), but they all failed to respond (i.e. persisting HCV RNA in serum after 12 weeks of treatment).

■ **Genotype assignment by conventional assays.** Patients' genotype was tested by line-probe assay (INNO-LIPA; Innogenetics) and verified by RFLP analysis of sequences amplified in the 5′ UTR (Davidson et al., 1995).

■ **Amplification and sequencing of liver-derived HCV RNA.** Total liver RNA was extracted (Chomczynski & Sacchi, 1987) and reverse-transcribed using random hexamers and AMV reverse transcriptase (Promega Biotech). The cDNA was then amplified in three separate reactions to obtain amplicons corresponding to the core, envelope 1 (E1) and part of the non-structural 5b (NS5b) regions. Primer selection was based on different approaches. For the core-encoding region, we used a universal genomic primer 5UTR43S (5′ GGGTG-CTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTG 3′), which spans positions −43 to −5 of most known HCV isolates (numbering according to Choo et al., 1991). The sequence of the antigenomic primer 3aASE1 (5′ ACAGTCGTTGGTAACGGCGTAGAGGCCGGAGGTG-TTCCG 3′), spanning positions 621–583, was chosen by alignment of different sequences of HCV clade 3, based on the provisional assignment of our isolates to this clade upon sequencing of the 5′ UTR obtained for RFLP analysis (see Results). PCR conditions were two 35-cycle rounds of PCR using 1.25 U Pfu polymerase (Promega) and the above primers. After a 5 min hot start at 95 °C, each cycle consisted of a denaturation step at 94 °C for 30 s, an annealing step at 55 °C for 45 s and an elongation step at 72 °C for 2 min. For the E1 region, primer sequences were deduced from the alignment of the core sequences of patients belonging to clade 3. The outer genomic primer 3aCORE461 had the sequence 5′ GCGTGAGGGCCCTTGAAGACGGGATAAATTTCGC-AACAGG 3′, spanning positions 461–500, whereas the outer anti-genomic primer 3aE2110 had the sequence 5′ ATGGAVTCATTGC-AGTTCAGGGCAGTACTGTTGATGTGCC 3′, spanning positions 1298–1259. For the second round of amplification, we chose the genomic primer 3aCORE476 (5′ AAGACGGGATAAATTTCGCAACAGGGA-ACTTGCCCGGTTG 3′), covering positions 476–515, and the anti-genomic primer 3aE297 (5′ AGTTCAGGGCAGTACTGTTGATGTG-CCACGAGCCATTGGT 3′), covering positions 1285–1246. PCR conditions were the same as those used for amplification of the core-encoding region. For amplification of the NS5b region, we used the primers and conditions reported by Mellor et al. (1995), with some exceptions: among the inner four genomic primers tested, we chose primer 554 since it gave the most consistent results; the annealing temperature was 55 °C instead of 45 °C; and the number of cycles was 35 instead of 30 for each round.

PCR products were purified with the Wizard DNA PCR Preps purification system (Promega) and directly sequenced using standard protocols for the ABI 377 automated sequencer. Primers used for the sequencing reactions (both directions) were the same as those used for the PCR amplifications.

■ **Sequence and phylogenetic analyses.** Sequences were first aligned by multiple sequence alignment with hierarchical clustering, using the MultAlin algorithm (Corpet, 1988). All subsequent phylogenetic analyses were carried out using the PHYLIP package (Felsenstein, 1993). Evolutionary distances were estimated with the DNADIST program. Unrooted trees with all branch lengths drawn to scale were constructed using the NEIGHBOR program. The robustness of the grouping was tested by the bootstrap resampling procedure; the values obtained for the tree drawn with programs SEQBOOT, DNADIST, NEIGHBOR and CONSENS were reported on the tree showing evolutionary distances. Neighbour-joining trees with 1000 replicates were used in the bootstrap analysis and values above 700 (70%) were considered to support the grouping (Robertson et al., 1998). Representative sequences from the six different HCV clades were chosen among those reported in the GenBank or EMBL databases and compared with the corresponding sequences of the three Somali isolates. We analysed the complete core-encoding region (573 nt), the complete E1-encoding region (576 nt) and the 222 nt portion of NS5b region (position 7975–8196; numbering according to Choo et al., 1991). Phylogenetic analysis was carried out using 51, 48 and 49 (not including the Somali isolates) different sequences for the core, E1 and NS5b regions, respectively. Trees representing clade 3 alone (for each region) were constructed with the same sequences used for the main trees, plus two, three and six additional sequences for the core, E1 and NS5b regions, respectively. Names of the isolates used in the present analyses are as follows (accession numbers are given in parentheses).

Core region. Clade 1: US11 (U10232); DR4 (U10196); HC-G9 (D14853); DK1 (U10193); HK3 (U10199); HK4 (U10200); YS117 (D16189). Clade 2: NE92 (L29631); S83 (U10211); T4 (U10228); T9 (U10230); T2 (U10226); SW3 (U10224); DK11 (U10190); T8 (U10229). Clade 3: S2 (U10208); NE125 (D14203); DK12 (U10191); HK10 (U10197); Th85 (D14307); S52 (U10210); US114 (D14309); NE048 (D16612); NE274 (D16620); NE145 (D16618); HCV-TR (D49374); NE137 (D16616); JK030 (D49747); JK049 (D49749); JK055 (D49750); JK070 (D49750); QC29 (U33437). Clade 4: Z7 (U10239); Z6 (U10238); CAM600 (L29589); Z4 (U10236); Z5 (U10237); Z1 (U10235). Clade 5: SA4 (U10218); SA7 (U10221); SA13 (U10215); BE95 (L29577). Clade 6: VN530 (D88471); VN507 (D88470); VN004 (D88465); VN085 (D88466); VN571 (D88476); VN569 (D88475); VN538 (D88473); VN540 (D88474); VN843 (D88478); VN998 (D31971); VN235 (D88467); VN405 (D17472). Somali isolates: SOM1 (AF216792); SOM2 (AF216793); SOM3 (AF216794).

E1 region. Clade 1: HK3 (L16660); HCV-JT (D11168); HN3 (X76414); YS117 (D16189); HC-G9 (D14853); US11 (L16676); DR4 (L16633); SW1 (L16671); DK1 (L16629). Clade 2: T4 (L16674); T9 (L16675); T2 (L16673); S83 (L16641); HN4 (X76415); SW3 (L16672); DK11 (L16655); T8 (L16649). Clade 3: US114 (D14309); S2 (L16639); Th85 (D14307); DK12 (L16630); S52 (L16640); NE274 (D16620); NE048 (D16612); NE145 (D16618); HCV-TR (D49374); NE137 (D166616); NE125 (D14203); JK030 (D49747); JK049 (D49749); JK055 (D49750); JK070 (D49752). Clade 4: Z1 (L16677); Z4 (L11652); Z6 (L16678); Z7 (L16653); DK13 (L16656). Clade 5: SA1 (L16642); SA4 (L16669); SA5 (L16644); SA6 (L16670). Clade 6: VN507 (D88470); VN531 (D88472); VN405 (D88468); VN085 (D88466); VN004 (D88465); VN569 (D88475); VN538 (D88473); VN540 (D88474); VN998 (D31971);

VN235 (D88467). Somali isolates: SOM1 (AF216786); SOM2 (AF216787); SOM3 (AF216788).

NS5b region. Clade 1: HC-G9 (D14197); YS117 (D14196); Td34/92 (D26389); T1801 (L23439); US17 (L23438); 4TY4 (L23447). Clade 2: HC-J5 (D10076); SAC640 (L23448); T104 (L23450); NE92 (L29634); ARG8 (L23458); T983 (L23460); HCV-K2B (D10649). Clade 3: NE048 (D16613); NE145 (D16619); NE274 (D16621); US114 (D14310); T-7 (D10079); HEM26 (D14312); T-1 (D10078); Eb-1 (D10130); T1787 (L23467); T-10 (D10081); NE137 (D16617); HCV-TR (D49374); NE125 (D16615); Td-35/93 (D37898); IND1751 (X91417); IND1452 (X91418); S. T. (D10585); JK030 (D49762); JK049 (D49764); JK055 (D49765); JK070 (D49767). Clade 4: CAM600 (L29590); GB358 (L29607); EG-13 (L23469); EG-19 (L23470); G22 (L29596); GB438 (L29611). Clade 5: BE96 (L29584); 34REV (L23474). Clade 6: VN787 (D87364); VN540 (D87361); VN998 (D30797); VN235 (D87354); VN506 (D87356); VN569 (D87362); BB9 (D28542); B4/92 (D28543); VN531 (D87359); VN530 (D87358); VN405 (D88468); VN085 (D87353); VN004 (D87352). Somali isolates: SOM1 (AF216789); SOM2 (AF216790); SOM3 (AF216791).

Results

Genotype assignment by conventional assays

The genotype of two of the three patients (numbers 1 and 3), as determined by the line-probe assay, was found to be 5a, whereas patient number 2 was infected with genotype 1b. Since the HCV genotype 5a is mostly present in southern Africa (Cha *et al.*, 1992), we verified the type assignment by RFLP analysis (Davidson *et al.*, 1995). Fig. 1 shows the RFLP pattern in the 5′ UTR (positions −279 to −29) for all three patients. *Hae*III/*Rsa*I digestion produced a fragment profile

compatible with genotypes 2 (Fig. 1*a*, lane 5) or 3 (see for example Davidson & Simmonds, 1998, Fig. 1 A, lane *g*). On the other hand, *Mva*I/*Hin*fI digestion yielded a pattern similar, but not identical, to that of genotype 1 (Fig. 1*b*, lane 4). Thus, since a definite genotype assignment could not be made for any of our isolates, we directly sequenced the PCR products used for the RFLP analysis. Fig. 2(*a*) shows the alignment of the sequences of the three Somali isolates (SOM1 from patient 1, SOM2 from patient 2 and SOM3 from patient 3). The two isolates SOM1 and SOM3 had identical sequences, whereas a single substitution was found in SOM2 at position −138. Analysis of the restriction map of the three Somali isolates is reported in Fig. 2(*b*, *c*), together with the *Rsa*I/*Hae*III and *Mva*I/*Hin*fI digestion patterns of other known HCV genotypes. As shown with the Somali isolates, *Rsa*I/*Hae*III digestion produced restriction fragments of 9, 23, 26, 33, 46 and 114 bp (Fig. 2*b*), whereas *Mva*I/*Hin*fI digestion produced fragments of 7, 41, 53, 56 and 94 bp (Fig. 2*c*).

According the widely used, RFLP-based classification of HCV types (Davidson & Simmonds, 1998), and considering the restriction maps of HCV sequences reported so far in the GenBank database (Fig. 2*b*, *c*), the above patterns do not allow a clear HCV genotype assignment. Recently, however, well-defined recommendations have been issued by an *ad hoc*, international HCV collaborative study group (Robertson *et al.*, 1998) to correctly classify within the known HCV genetic groups new sequences that may not fit type assignment as determined by conventional assays. Based on those recommendations, we determined the sequences of the core-, E1- and
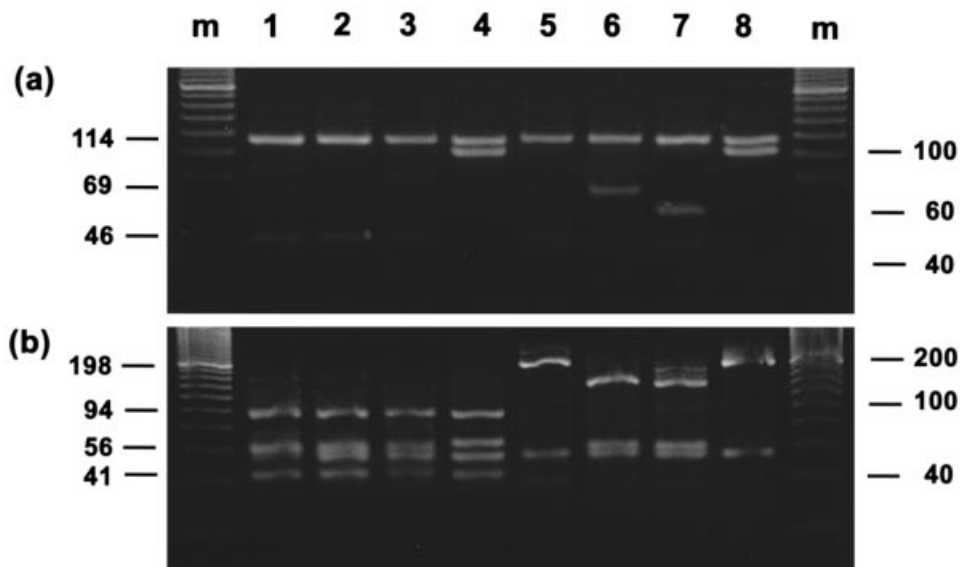

Fig. 1. Restriction patterns of three Somali HCV isolates compared with controls. RFLP patterns, as described by Davidson *et al.* (1995), were obtained by digestion of the amplicons with *Hae*III/*Rsa*I (*a*) and with *Mva*I/*Hin*fI (*b*). Samples were electrophoresed in a 4% MetaPhor (FMC Bioproducts) agarose gel. Numbers designate patterns relating to different HCV isolates: 1–3, Somali isolates; 4–8, representative HCV genotypes 1–5, respectively; m, molecular size markers (expressed in bp).

(a)

```
                                          RsaI                    MvaI  HaeIII
             -279                          ▽                       ▼    ▽
SOM1     TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTACGAGTGTCGAACAGCCTCCAGGCCCCCCC
SOM2     TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTACGAGTGTCGAACAGCCTCCAGGCCCCCCC
SOM3     TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTACGAGTGTCGAACAGCCTCCAGGCCCCCCC
HCV-1a   TTCACGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCC

                                      RsaI    HinfI   MvaI
             -216                      ▽       ▼      ▼
SOM1     CTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATCCCCAGGACGACCG
SOM2     CTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATCCCCAGGACGACCG
SOM3     CTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATCCCCAGGACGACCG
HCV-1a   CTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCG

                                    MvaI
             -153                    ▼
SOM1     GGTCCTTTCTTGGATTAACCCGCTCCATGCCTGGGTATTTGGGCGTGCCCCCGCGAGACTGCT
SOM2     GGTCCTTTCTTGGATCAACCCGCTCCATGCCTGGGTATTTGGGCGTGCCCCCGCGAGACTGCT
SOM3     GGTCCTTTCTTGGATTAACCCGCTCCATGCCTGGGTATTTGGGCGTGCCCCCGCGAGACTGCT
HCV-1a   GGTCCTTTCTTGGATAAAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGACTGCT

                                 HaeIII      RsaI
             -90                   ▽          ▽                      -29
SOM1     AGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTCGAGAGTG
SOM2     AGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTCGAGAGTG
SOM3     AGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTCGAGAGTG
HCV-1a   AGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTG
```

(b)

	RsaI ▽	RsaI ▽	HaeIII ▽	RsaI ▽	HaeIII ▽	HaeIII ▽	RsaI ▽
a	33	23	46	114		9	26
b	102			114/5		9	26
c	44	58		114/5		9	26
d	44	12	46	58	56	9	26
e	44	12	46	114		9	26
f	56		46	114		9	26
g	56		46	58	56	9	26
h	33	69		114		9	26
i	83		18	114		9	26
j	101			114		9	26
k	44	172				9	26
l	217					9	26
l'	216					9	26
m	44	12	46	117		9	26

(c)

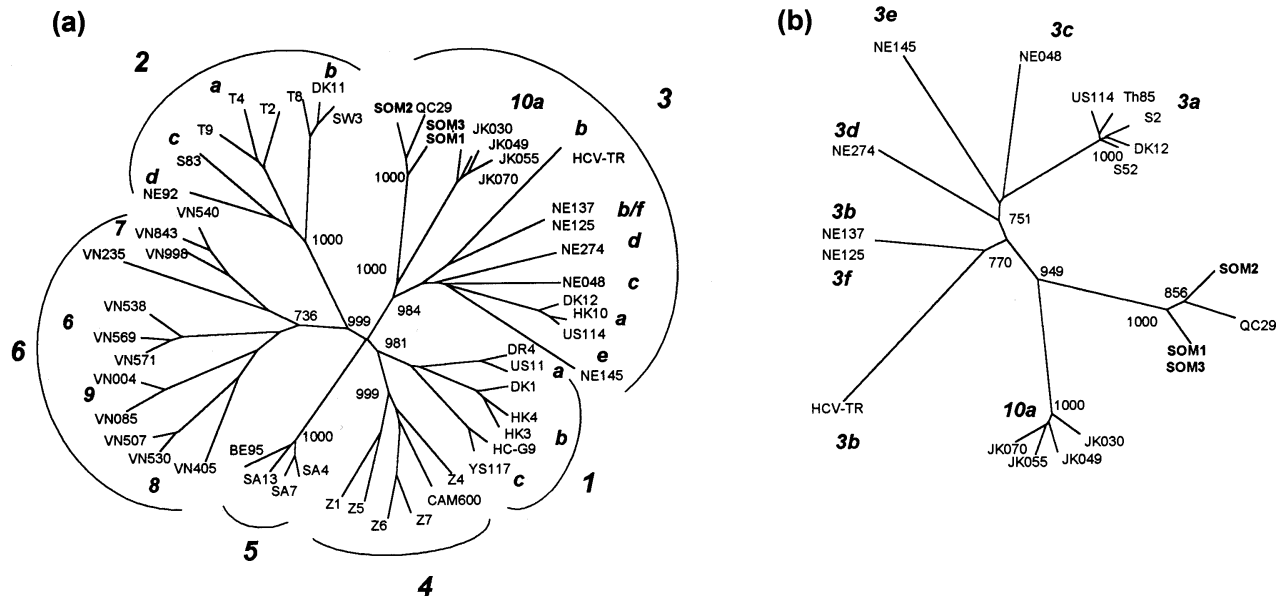|	MvaI ▼	HinfI ▼	MvaI ▼	MvaI ▼	
a	53	56	7	41	94
b	53	63		41	94
c	53	104			94
d	53	198			
e	109		142		
f	53	56	142/3		
g	53	63	43/4		94

Fig. 2. For legend see facing page.

Fig. 3. (*a*) Phylogenetic analysis of the HCV core-encoding region based on comparison of representative nucleotide sequences from all six clades. Bootstrap scores over 1000 replicates are shown for the clades and the Somali isolates branching. (*b*) Phylogenetic analysis as in (*a*) but limited to HCV clade 3. Bootstrap scores over 700 (70% of 1000 replicates) are shown. Note that isolates NE137 and NE125, assigned to the two different subtypes 3f and 3b, respectively, based on the NS5b sequence comparison, show here the same sequence.

NS5b-encoding regions of the three Somali isolates and compared them with the corresponding regions of representative subtypes chosen within the six different clades.

Sequence analysis of the core and E1 regions

As far as the core region is concerned, FASTA analysis showed that isolate QC29 (Bernier *et al.*, 1996) had the highest sequence similarity with the three Somali isolates, 95% similarity with SOM1 and SOM3, and 96% similarity with SOM2. The second highest level of similarity was with isolate JK049 (Tokita *et al.*, 1996), 87% with SOM1 and SOM3, and 86% with SOM2. As quoted above, isolate QC29 was found in another Somali patient in Montreal, Canada, but its genotype assignment was withheld pending acquisition of further sequence data and the identification of additional isolates (D. Murphy, personal communication). JK049 is an isolate from

Jakarta, Indonesia, initially referred to as belonging to a new genetic group named 10a (Tokita *et al.*, 1996) and later assigned to a distinct subtype within clade 3 (Mizokami *et al.*, 1996; Simmonds *et al.*, 1996; de Lamballerie *et al.*, 1997; Robertson *et al.*, 1998). Phylogenetic analysis showed that our three Somali isolates, together with isolate QC29, form a distinct cluster within clade 3 (Fig. 3*a*), and bootstrap values of 100% over 1000 replicates, assigned to this cluster, support this grouping. The tree indicates a close relatedness with the group formerly referred to as type 10a mentioned above. Phylogenetic analysis performed within clade 3 only (Fig. 3*b*) confirms the degree of relatedness between different subtypes and the separate branching of the Somali isolates, with bootstrap values of 100% over 1000 replicates.

The approach used for the analysis of the core-encoding region was applied to the E1-encoding region. The highest similarity scores were found with isolate JK070 (72%, 74% and

Fig. 2. (*a*) Aligned genomic sequences of the HCV 5′ UTR from the three Somali isolates SOM1, SOM2 and SOM3 and the HCV prototype sequence 1a (Inchauspe *et al.*, 1991). Restriction sites used to type HCV by RFLP (Davidson *et al.*, 1995) are shown for both the first *Rsa*I/*Hae*III reaction (▽) and the second *Mva*I/*Hin*fI reaction (▼), respectively. (*b*) Restriction patterns for the *Rsa*I/*Hae*III reaction for all HCV 5′ UTR sequences reported in the GenBank database. Restriction sites are indicated as in (*a*). Pattern **a** is found in all three Somali isolates SOM1, SOM2 and SOM3 and in some isolates belonging to HCV type 3. Patterns **b** and **c** are common to types 1, 4 and 5. Patterns **d**–**g** are all exclusively found in type 2. Type 3 may present with patterns **h**–**j** (more rarely with pattern **a**). Type 4 may show patterns **h**, **k** and **l**. Type 5 (as well as patterns **b** and **c**) may also present pattern **l′**, which differs from pattern **l** only by a 1 nt deletion within the major restriction fragment. Finally, pattern **m** is the only one described for type 6. (*c*) Same as (*b*) except restriction patterns for the *Mva*I/*Hin*fI reaction are shown. The three Somali isolates have the pattern shown as **a**, common also to some isolates belonging to type 4. Type 4 may also present pattern **f**. Type 1 has patterns **b** and **c**, type 2 has patterns **c** and **d**, type 3 has patterns **e** and **f**, type 5 has patterns **d** and **f**, whereas pattern **g** is the only one reported for type 6.
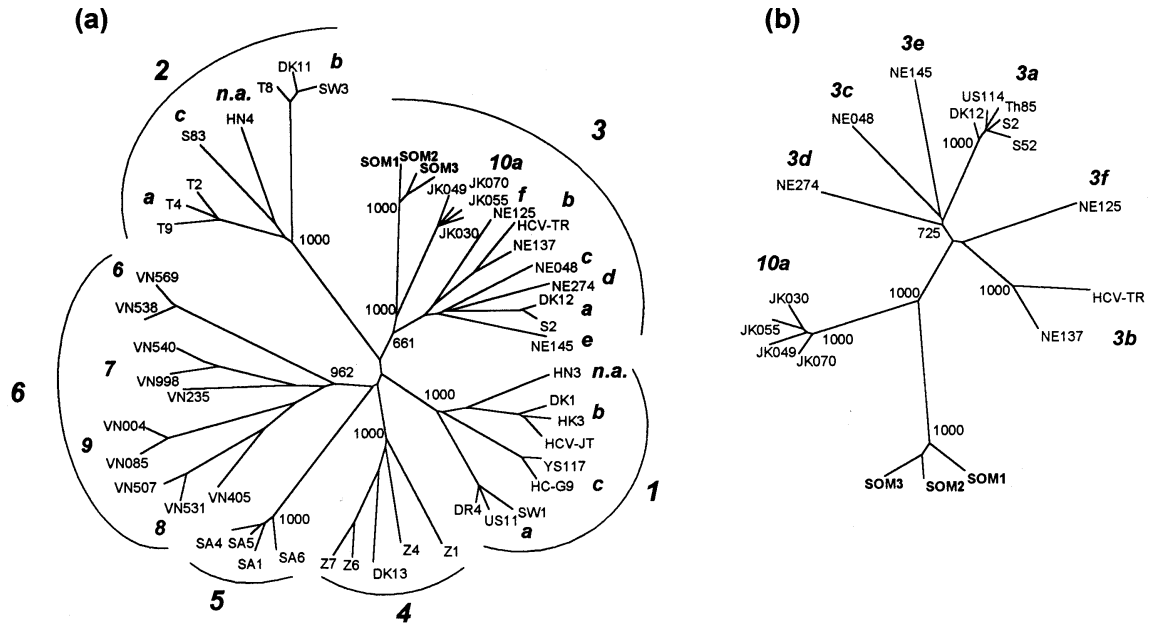
Fig. 4. (a) Phylogenetic analysis of the HCV E1 region based on comparison of representative nucleotide sequences from all six clades. Isolates HN3 and HN4 represent not assigned (*n.a.*) subtypes of clades 1 and 2, respectively. Bootstrap scores are shown for the clades and the Somali isolate branching. (b) Phylogenetic analysis as in (a) but limited to HCV clade 3. Bootstrap values over 700 (70 % of 1000 replicates) are shown.
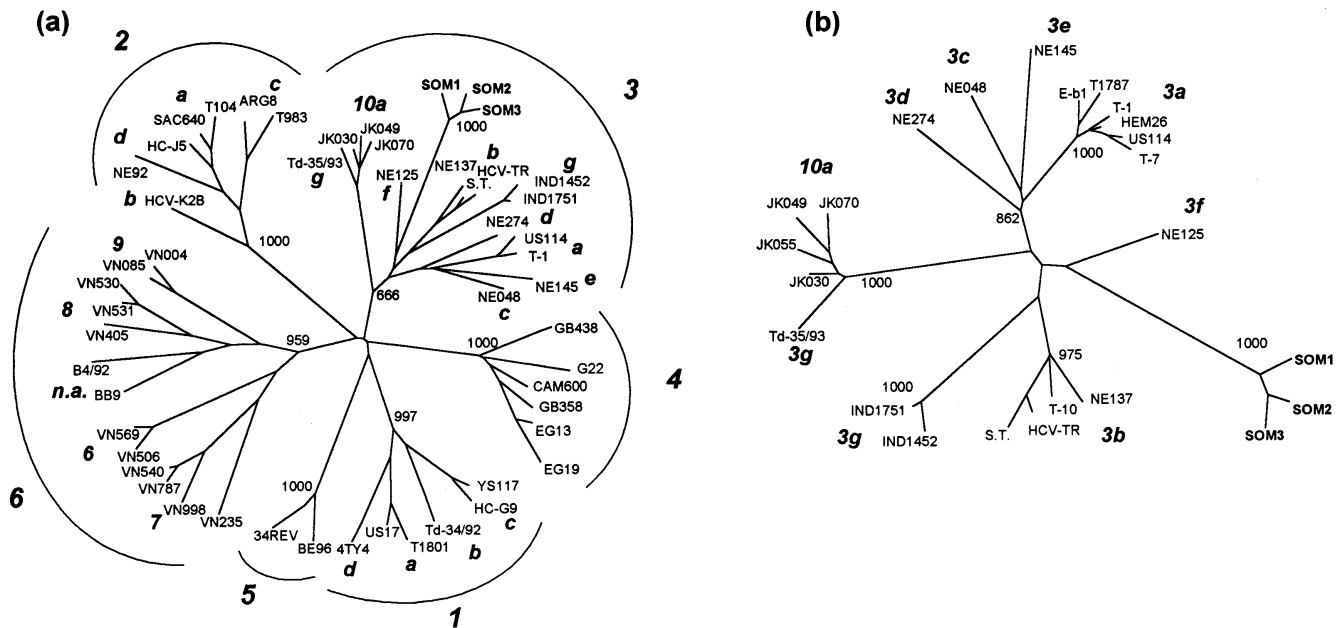

Fig. 5. (a) Phylogenetic analysis of the 222 nt sequence of the HCV NS5b region based on comparison of representative nucleotide sequences from all six clades. Isolates B4/92 and BB9 represent not assigned (*n.a.*) subtypes of HCV clade 6. Bootstrap scores are shown for the clades and the Somali isolate branching. (b) Phylogenetic analysis as in (a) but limited to HCV clade 3. Isolate Td-35/93, described as belonging to subtype 3g, was found to group with type 10a, and was only distantly related to isolates IND1751 and IND1452, i.e. subtype 3g proper. Bootstrap scores over 700 (70 % of 1000 replicates) are shown.

71 % with SOM1, SOM2 and SOM3, respectively) and with isolates JK055 (71 % with SOM1 and 70 % with SOM3) and JK072 (72 % with SOM2). The phylogenetic trees of the E1-encoding region displayed the same topology encountered in the trees obtained with the core-encoding region, namely a cluster of the three Somali isolates distinct from other existing subtypes, as confirmed by bootstrap values of 100 % on 1000 replicates for both trees (Fig. 4 *a, b*).

Analysis of the NS5b region

As far as the NS5b region is concerned, the closest sequence to the Somali isolates, as determined by FASTA, did not match that found with the analyses of the core and E1 regions. In fact, the highest similarity scores for SOM1 and SOM2 were found with an isolate from Nepal (NE125) classified as subtype 3f (Tokita *et al.*, 1994 *b*). In both cases, this similarity was 77 %. The highest similarity found for SOM3 was observed with isolate S. T. from Japan (Chayama *et al.*, 1994), belonging to subtype 3b. The second highest similarity score for the three Somali isolates was observed with isolate HCV-TR, which belonged to subtype 3b (Chayama *et al.*, 1994), with 75 % similarity for SOM1 and SOM2, and 76 % for SOM3. The phylogenetic trees of the NS5b region, for all six HCV clades considered together and for clade 3 alone (Fig. 5 *a*, *b*), confirmed the results obtained by FASTA analysis, i.e. the Somali isolates were more closely related to subtype 3f than to isolates from Jakarta. Bootstrap values of 100 % over 1000 replicates in both trees support the branching of the Somali isolates into a separate cluster, like those observed for the core- and E1-encoding regions.

Discussion

Based on the recently proposed criteria, which follow the guidelines of the International Committee on Virus Taxonomy (ICTV) and allow the subdivision of HCV into six clades, we have described a novel HCV subtype, which is tentatively classified as subtype 3h within clade 3. All patients infected with this new HCV subtype come from Somalia, as did one patient formerly described by Bernier *et al.* (1996) and who had emigrated to Montreal, Canada. In that case, the isolate QC29 was not further assigned to a precise virus type, pending a more thorough sequence characterization although, based on phylogenetic analysis, it is very likely that it belongs to the same subtype as the present HCV isolates. Apart from the same geographical origin, there were no other evident epidemiological links among the three patients, although the source of infection was not documented with certainty for any of them.

HCV clade 3, together with clade 6, includes a great number of subtypes characterized by extensive sequence variability. Most recently discovered HCV (sub)types have been in fact assigned to those two clades, such as genotypes 7, 8, 9 and 11 (assigned to clade 6) and genotype 10 (assigned to clade 3). The close phylogenetic relatedness of the three present Somali isolates with HCV sequences found in geographically distant areas, such as the Indian subcontinent (Tokita *et al.*, 1994 *b*; Valliammai *et al.*, 1995; Panigrahi *et al.*, 1996) or Indonesia (Tokita *et al.*, 1996; Hotta *et al.*, 1994) is puzzling. However, an Indian minority is present in major coastal cities of East African countries, including Somalia, and HCV could have originated from this source and spread to the indigenous Somali population, possibly via nosocomial or traditional medical procedures. Due to the sequence divergence, we postulate that branching of subtype 3h from an ancestor common to other Indian/Indonesian subtypes may have occurred a long time ago, but the exact modalities (i.e. where, when and how) which led to the differentiation of the Somali subtype can only be the object of speculation, until more sequences from these areas are available for study. However, we cannot exclude the possibility that the close relatedness of subtype 3h with type 10a is only apparent, since it was based on sequence analysis of the core and E1 regions, but not the NS5b region. Subtype 3h may in fact form an intermediate genetic group (like 10a) between another clade 3 subtype and another HCV clade – further sequence data may confirm or dismiss this hypothesis. Pending the availability of more complete sequence data, the evolution of virus subtypes within clade 3 may follow unexpected routes, making every speculation premature at the present time.

We are fully aware that a definitive subtype assignment of novel HCV sequences may only be based on full genomic sequencing. In fact, the minimal length of the subgenomic fragments sufficient to correctly classify new HCV (sub)types is controversial. Different studies have been performed on part or complete nucleotide sequences of the core-encoding region (Mellor *et al.*, 1995; Panigrahi *et al.*, 1996; Bukh *et al.*, 1994), the E1 region (Bukh *et al.*, 1993; Simmonds *et al.*, 1994), the NS4 region (Bhattacherjee *et al.*, 1995) and various lengths of nucleotide fragments within the NS5b-encoding region (Simmonds *et al.*, 1993, 1994; Apichartpiyakul *et al.*, 1994; Valliammai *et al.*, 1995; Mellor *et al.*, 1995). However, other authors have suggested that a precise classification may require analysis of longer fragments (Stuyver *et al.*, 1994), if not analysis of the complete genome sequence (Tokita *et al.*, 1994 *a*). Sequence similarity or evolutionary distance analyses have yielded identical results, namely the existence of six different groupings; these are irrespective of the region being analysed. The major determinant is the number and the representativity of the sequences considered, rather than the choice of more accurate subgenomic regions for the classification of HCV genotypes (Mellor *et al.*, 1995). In keeping with this assumption, and following the aforementioned guidelines (Robertson *et al.*, 1998), we restrained our present analysis to the three subgenomic regions encoding the core, E1 and (partially) NS5b proteins, pending further sequence data. Incidentally, the drawing of the HCV clade 3 phylogenetic tree underscores a certain degree of confusion in the nomenclature of the subtypes reported so far, and underlines the need for a proper reclassification based on stringent and homogeneous criteria.

Finally, we are concerned about the failure to correctly identify the virus genotype in all three cases by conventional typing assays. A single mutation at position 204 within the 5′ UTR shifted the type assignment (by the line-probe assay) from genotype 5a (SOM1 and SOM3) to 1b (SOM2). When

we performed the sequence analysis, it turned out that none of isolates had been correctly typed. Moreover, RFLP analysis gave a unique restriction pattern, common to all three isolates, which was incompatible with the HCV sequences reported in the literature. Correct virus genotyping has proven to be critical for determining optimal treatment duration (Poynard *et al.*, 1998) and may also be important when interpreting anatomo-clinical correlations (Rubbia-Brandt *et al.*, 2000). We did not test the Somali sera with other assays, such as the serotyping assay (commercially unavailable in Switzerland), although it may be worth trying. For the sake of cost and simplicity, we favoured RFLP analysis, since, in case of dubious results, it leads directly towards a direct sequencing analysis, which is the gold standard procedure for HCV genotyping (Gretch, 1997). In this respect, the impossibility of assigning a virus type (as with RFLP) is obviously better than a wrong assignment (as was the case with the line-probe assay). For future reference, the unique restriction pattern of subtype 3h may help to identify this subtype in other patients. Further work is however needed to confirm and extend our above findings, as well as to clarify the phylogenetic history of this novel HCV subtype from Somalia.

References

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. *Journal of Clinical Microbiology* **32**, 2276–2279.

Bedossa, P. & Poynard, T. (1996). An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* **24**, 289–293.

Bernier, L., Willems, B., Delage, G. & Murphy, D. G. (1996). Identification of numerous hepatitis C virus genotypes in Montreal, Canada. *Journal of Clinical Microbiology* **34**, 2815–2818.

Bhattacherjee, V., Prescott, L. E., Pike, I., Rodgers, B., Bell, H., El-Zayadi, A. R., Kew, M. C., Conradie, J., Lin, C. K., Marsden, H., Saeed, A. A., Parker, D., Yap, P.-L. & Simmonds, P. (1995). Use of NS-4 peptides to identify type-specific antibody to hepatitis C virus genotypes 1, 2, 3, 4, 5 and 6. *Journal of General Virology* **76**, 1737–1748.

Bukh, J., Purcell, R. H. & Miller, R. H. (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. *Proceedings of the National Academy of Sciences, USA* **90**, 8234–8238.

Bukh, J., Purcell, R. H. & Miller, R. H. (1994). Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proceedings of the National Academy of Sciences, USA* **91**, 8239–8243.

Cha, T. A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G. & Urdea, M. S. (1992). At least five related, but distinct, hepatitis C viral genotypes exist. *Proceedings of the National Academy of Sciences, USA* **89**, 7144–7148.

Chayama, K., Tsubota, A., Koida, I., Arase, Y., Saitoh, S., Ikeda, K. & Kumada, H. (1994). Nucleotide sequence of hepatitis C virus (type 3b) isolated from a Japanese patient with chronic hepatitis C. *Journal of General Virology* **75**, 3623–3628.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P. J., Weiner, A., Bradley, D. W., Kuo, G. & Houghton, M. (1991). Genetic organization and diversity of the hepatitis C virus. *Proceedings of the National Academy of Sciences, USA* **15**, 2451–2455.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881–10890.

Davidson, F. & Simmonds, P. (1998). Determination of HCV genotypes by RFLP. In *Hepatitis C Protocols*, pp. 175–181. Edited by J. Y.-N. Lau. Totowa, NJ: Humana Press.

Davidson, F., Simmonds, P., Ferguson, J. C., Jarvis, L. M., Dow, B. C., Follett, E. A. C., Seed, C. R. G., Krusius, T., Lin, C., Medgyesi, G. A., Kiyokawa, H., Olim, G., Duraisamy, G., Cuypers, T., Saeed, A. A., Teo, D., Conradie, J., Kew, M. C., Lin, M., Nuchaprayoon, C., Ndimbie, O. K. & Yap, P. L. (1995). Survey of major genotypes and subtypes of hepatitis C virus using RFLP of sequences amplified from the 5′ non-coding region. *Journal of General Virology* **76**, 1197–1204.

de Lamballerie, X., Charrel, R. N., Attoui, H. & De Micco, P. (1997). Classification of hepatitis C virus variants in six major types based on analysis of the envelope 1 and nonstructural 5B genome regions and complete polyprotein sequences. *Journal of General Virology* **78**, 45–51.

Felsenstein, J. (1993). PHYLIP inference package version 3.5. University of Washington, Seattle: Department of Genetics.

Gretch, D. R. (1997). Diagnostic tests for hepatitis C. *Hepatology* **26**, 43S–47S.

Hotta, H., Handajani, R., Lusida, M. I., Soemarto, W., Doi, H., Miyajima, H. & Homma, M. (1994). Subtype analysis of hepatitis C virus in Indonesia on the basis of NS5b region sequences. *Journal of Clinical Microbiology* **32**, 3049–3051.

Inchauspe, G., Zebedee, S., Lee, D. H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991). Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proceedings of the National Academy of Sciences, USA* **88**, 10292–10296.

McHutchison, J. G., Gordon, S. C., Schiff, E. R., Shiffman, M. L., Lee, W. M., Rustgi, V. K., Goodman, Z. D., Ling, M. H., Cort, S. & Albrecht, J. K. (1998). Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *New England Journal of Medicine* **339**, 1485–1492.

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

Mizokami, M., Gojobori, T., Ohba, K., Ikeo, K., Ge, X. M., Ohno, T., Orito, E. & Lau, J. Y. (1996). Hepatitis C virus types 7, 8 and 9 should be classified as type 6 subtypes. *Journal of Hepatology* **24**, 622–624.

Okamoto, H., Sugiyama, Y., Okada, S., Kurai, K., Akahane, Y., Sugai, Y., Tanaka, T., Sato, K., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1992).

Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *Journal of General Virology* **73**, 673–679.

Panigrahi, A. K., Roca, J., Acharya, S. K., Jameel, S. & Panda, S. K. (1996). Genotype determination of hepatitis C virus from Northern India: identification of a new subtype. *Journal of Medical Virology* **48**, 191–198.

Poynard, T., Marcellin, P., Lee, S. S., Niederau, C., Minuk, G. S., Ideo, G., Bain, V., Heathcote, J., Zeuzem, S., Trepo, C. & Albrecht, J. (1998). Randomised trial of interferon alpha2b plus ribavirin for 48 weeks or for 24 weeks versus interferon alpha2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. International Hepatitis Interventional Therapy Group. *Lancet* **352**, 1426–1432.

Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin-I, T., Simmonds, P., Smith, D., Stuyver, L. & Weiner, A. (1998). Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. *Archives of Virology* **143**, 2493–2503.

Rubbia-Brandt, L., Quadri, R., Abid, K., Giostra, E., Male, P.-J., Mentha, G., Spahr, L., Zarski, J.-P., Borisch, B., Hadengue, A. & Negro, F. (2000). Hepatocyte steatosis is a cytopathic effect of hepatitis C virus genotype 3. *Journal of Hepatology* (in press).

Simmonds, P., Holmes, E. C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *Journal of General Virology* **74**, 2391–2399.

Simmonds, P., Smith, D. B., McOmish, F., Yap, P. L., Kolberg, J., Urdea, M. S. & Holmes, E. C. (1994). Identification of genotypes of hepatitis C virus by sequence comparison in the core, E1 and NS-5 region. *Journal of General Virology* **75**, 1053–1061.

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. *Journal of General Virology* **77**, 3013–3024.

Stuyver, I., Vanarherm, W., Wyseur, A., Hernandez, F., Delaporte, E. & Maertens, G. (1994). Classification of hepatitis C virus based on phylogenetic analysis of the envelope 1 and nonstructural 5b regions and identification of five additional subtypes. *Proceedings of the National Academy of Sciences, USA* **91**, 10134–10138.

Tokita, H., Okamoto, H., Tsuda, F., Song, P., Nakata, S., Chosa, T., Iizuka, H., Mishiro, S., Miyakawa, Y. & Mayumi, M. (1994 *a*). Hepatitis C virus variants from Vietnam are classifiable into the seventh, eight and ninth major groups. *Proceedings of the National Academy of Sciences, USA* **91**, 11022–11026.

Tokita, H., Shrestha, S. M., Okamoto, H., Sakamoto, M., Horikita, M., Iizuka, H., Shrestha, S., Miyakawa, Y. & Mayumi, M. (1994 *b*). Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. *Journal of General Virology* **75**, 931–936.

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. *Journal of General Virology* **76**, 2329–2335.

Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F., Lesmana, L. A., Miyakawa, Y. & Mayumi, M. (1996). Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. *Journal of General Virology* **77**, 293–301.

Valliammai, T., Thyagarajan, S. P., Zuckerman, A. J. & Harrison, T. J. (1995). Diversity of genotypes of hepatitis C virus in southern India. *Journal of General Virology* **76**, 711–716.