Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups

Hajime Tokita,1 Hiroaki Okamoto,1 Hisao Iizuka,2 Junichi Kishimoto,3 Fumio Tsuda,4 Laurentius A. Lesmana,5 Yuzo Miyakawa6 and Makoto Mayumi1*

1 Immunology Division, Jichi Medical School, Minamikawachi-Machi, Tochigi-Ken 329-04, Japan, 2 Japanese Red Cross Saitama Blood Center, Saitama-Ken 338, Japan, 3 Institute of Immunology, Tokyo 112, Japan, 4 Department of Medical Sciences, Toshiba General Hospital, Tokyo 140, Japan, 5 Department of Internal Medicine, Faculty of Medicine, University of Indonesia, Jakarta 10430, Indonesia and 6 Mita Institute, Tokyo 108, Japan

Hepatitis C virus (HCV) isolates from 126 hepatitis patients in Jakarta, Indonesia were genotyped by PCR with genotype-specific primers deduced from the HCV core gene. Fifty-five isolates (44%) were classified as genotype II/1b, 15 (12%) as lc, 33 (26%) as III/2a, and 1 (1%) as V/3a, while the remaining 22 (17%) were not classifiable into any of the five common genotypes (I/la, II/1b, III/2a, IV/2b and V/3a) or lc. Sequences of a part of the NSSb region [1093 bp (nucleotides 8279-9371)] of the 22 isolates of unclassifiable genotype were subjected to pair-wise comparison and phylogenetic analysis along with those of 62 isolates of 25 genotypes in nine genetic groups. Seven of the isolates were classified into 2e and two into 2f, representing novel genotypes in genetic group 2, while ten and three were classified into two new genetic groups, 10 and 11, respectively, and their genotypes were provisionally designated 10a and 11a. The isolates of genotype 10a (JK049) and 11a (JK046) were sequenced in full. Comparison of 24 HCV genomes including those of JK049 and JK046, over the entire genome and subgenomic regions, supported the classification of HCV into 11 genetic groups.

Introduction

Hepatitis C virus (HCV) is a positive-stranded linear RNA virus of approximately 9500 nucleotides (nt), and is the major causative agent of blood-borne non-A, non-B hepatitis worldwide (for a review, see Houghton et al., 1991). It has a single, long open reading frame flanked by 5' and 3' untranslated regions (UTR). The polyprotein precursor encoded by the open reading frame is cleaved by viral and host proteases into core (C) protein, envelope 1 (E1) and envelope 2/nonstructural 1 (E2/NS1) glycoproteins, as well as six nonstructural proteins designated NS2, NS3, NS4a, NS4b, NS5a and NS5b (Houghton et al., 1991; Grakoui et al., 1993).

HCV is one of the few viruses pathogenic for humans for which detection of viral RNA remains the sole means of establishing current infection, and sequence divergence in various areas of the genome is used for classifying the virus. Since the divergence in the entire genome and parts of the NS5b region distributes in two distinct tiers, HCV is classified into genetic groups or types named 1, 2, 3 etc. which divide further into genotypes (Okamoto et al., 1994; Tokita et al., 1995) or subtypes called a, b, c etc. (Simmonds et al., 1993; Simmonds, 1995). Up to the present, at least 34 genotypes in nine genetic groups have been reported (Cha et al., 1992; Mori et al., 1992; Bukh et al., 1993, 1994; Simmonds et al., 1993; Stuyver et al., 1993, 1994; Okamoto et al., 1994; Tokita et al., 1994a, b, 1995). HCV genotypes have distinct geographical distributions. It is not clear, however, how HCV genotypes should be classified in a purely virological context or for any clinical purposes.

From the standpoint that HCV genotypes should not be characterized fully except by determining the entire genomic sequence, we have proposed five common genotypes designated I, II, III, IV and V (Sakamoto et al., 1994). These five represent essentially all HCV genotypes in Europe, North America, Oceania and most areas in Asia (Bukh et al., 1995; Miyakawa et al., 1995; Simmonds, 1995). Genotype I corresponds to 1a, II to 1b, III to 2a, IV to 2b and V to 3a.

* Author for correspondence. Fax +81 285 44 1557.

The sequences reported in this paper have been deposited in the DDBJ, GenBank and EMBL nucleotide sequence databases (accession nos D49745-D49780, D63821 and D63822).
Sequence divergence of the entire HCV genome is reflected, to a greater or lesser extent, within the various genomic regions. Hence, new genotypes keep being recorded based on divergence in partial genomic sequences, typified by 222 bp in the NS5b region representing comparison, such as 576 bp of the E1 gene (Bukh et al., 1993; Tokita et al., 1994b), 573 bp of the C gene (Bukh et al., 1994), and parts of the NS5b region spanning 329 bp (nt 8279–8607) (Tokita et al., 1994b), 340 bp (nt 8276–8615) (Enomoto et al., 1990; Styuver et al., 1994) and 1093 bp (nt 8279–9371) (Tokita et al., 1994b).

Of 126 HCV isolates from hepatitis patients in Jakarta, Indonesia, 22 (17%) were not classifiable into any of the five common genotypes (I/1a, II/1b, III/2a, IV/2b and V/3a) or into genotype lc (for which the entire genomic sequences have been characterized) by PCR with type-specific primers deduced from the HCV core gene. Pair-wise comparison and phylogenetic analysis of the 1093 bp (nt 8279–9371) NS5b sequence indicated that these new isolates belong to novel genotypes 2e and 2f in genetic group 2, as well as 10a and 11a in new genetic groups. Further, full-length sequences of two HCV genomes of genotype 10a or 11a were determined and were compared with the complete genomes of 22 HCV isolates of various genotypes.

Methods

Sera containing HCV. Sera from 126 hepatitis patients in Jakarta, Indonesia, testing positive for HCV RNA, were genotyped by PCR with type-specific primers deduced from the HCV core gene (Okamoto et al., 1992a, 1992b, 1993, 1994). Genotype lc was determined by PCR with primers specific for this genotype deduced from the core gene (Okamoto et al., 1994).

Amplification by PCR and sequence analysis. Nucleic acids were extracted from 50–100 μl of serum, and reverse-transcribed to cDNA with HCV-specific 20-mer antisense primers [#299 (nt 250–269), #122 (nt 828–847), a mixture of #127, #204 and #337 (nt 1587–1606), and #316 (nt 8608–8627)] or a non-specific 43-mer primer (#165) with (A)7 as described previously (Okamoto et al., 1992a, 1993; Tokita et al., 1994a, b). Nucleotides were numbered from the putative 5' end of the HCV genome of genotype II/1b [HC-14/83 (accession no. D01217)]. Amplification and sequencing of the 5'-terminal 1.6 kb (containing the 5'UTR, C gene, E1 gene and part of the E2/NS1 gene) and the 3'-terminal 1.1 kb (containing parts of the NS5b region and the 3'UTR) were performed by methods described elsewhere (Okamoto et al., 1993; Tokita et al., 1994a, b). The consensus sequence was determined from three independent clones.

We also sequenced HCV isolates in group 4 (HEMA51) and group 5 (FR741). HEMA51 was 89–91% similar in the NS5b 222 bp sequence to the genotype 4a isolates (EG-13 and EG-19) reported by Simmonds et al. (1993) but only 75% similar in the entire E1 gene sequence to the genotype 4a isolate (Z4) documented by Bukh et al. (1993); therefore, it was deduced to be of genotype 4a (Simmonds et al., 1993). FR741 was 95–97% similar in the NS5b 222 bp sequence when compared with the genotype 5a isolates (SA30, SA156, SA183 and 34REV) reported by Simmonds et al. (1993) and 89–91% similar in the entire E1 gene sequence to the genotype 5a isolates (SA1, SA4-7 and SA13) reported by Bukh et al. (1993). Hence, it was of genotype 5a.

Construction of the phylogenetic tree. A phylogenetic tree for HCV was constructed by the unweighted pair-group method with the arithmetic mean of Nei (1987) using a molecular evolutionary analysis system for DNA and amino acid sequences (treeupg of the ODE programs version 1.1.1, National Institute of Genetics, Mishima, Japan).

Determination of the full-length sequences of two HCV isolates in novel genetic groups. The entire genomic sequences of isolate JK049 of genotype 10a and isolate JK046 of genotype 11a were determined. The nucleotide sequences of the 5'-terminal 1.6 kb and the 3'-terminal 1.1 kb were determined as described above. The central portion was divided into four domains, nt 1469–1871 (403 bp), nt 1811–1958 (348 bp), nt 4900–5489 (590 bp) and nt 5439–8333 (2895 bp) for JK049, and nt 1447–1867 (421 bp), nt 1803–5232 (3431 bp), nt 5160–5485 (126 bp) and nt 5427–8346 (2920 bp) for JK046; nt were numbered in accordance with sequences specific to genotype 10a or 11a. These sequences were amplified by PCR with universal primers deduced from the conserved areas among HCV genomes of known full-length sequence or primers specific for JK049 or JK046. PCR was performed by the method described above or by means of long-distance PCR with Taq Extender PCR additive (Stratagene) and Takara Ex Taq DNA polymerase (Takara Biochemicals). Long-distance PCR was performed for 30 cycles with each cycle consisting of denaturation at 94 °C for 30 s, primer annealing at 55 °C for 30 s, and primer extension at 72 °C for 6 min.

Results

Classification of HCV isolates of unknown genotype from Jakarta, Indonesia by pair-wise comparison of a 1093 bp NS5b sequence

Of 126 sera from hepatitis patients in Jakarta, HCV genotype II/1b was detected in 55 (44%), 1c in 15 (12%), III/2a in 33 (26%) and V/3a in 1 (1%). HCV isolates of unclassifiable genotype in the remaining 22 (17%) sera were sequenced partially. They were subjected to pair-wise comparison, along with 62 HCV isolates of 25 genotypes in nine genetic groups, within a 1093 bp (nt 8279–9371) sequence in the NS5b region (Table 1). We have reported that HCV isolates can be classified by the criteria of isolate similarity of 91.7–99.2%, genotype similarity of 80.2–88.6% and group similarity of 66.5–80.0% (Tokita et al., 1994b, 1995). The 22 Jakarta isolates of unknown genotype showed a sequence similarity of ≤ 87.6% to the reported isolates of 25 genotypes, indicating that they would belong to distinct genotypes. They were classified into two novel genotypes in group 2 (2e and 2f) and two novel genotypes in new genetic groups (10a and 11a) for the reasons detailed below.

Seven of the Jakarta isolates (JK020, JK025, JK109, JK117, JK128, JK143 and JK151) differed from one another by < 7-7% and were classified as genotype 2e, while two of the isolates (JK081 and JK139), differing...
Table 1. Comparison of genotypes 2e, 2f, 10a and 11a from Jakarta, Indonesia with reported genotypes within a 1093 bp (nt 8279–9371) NS5b sequence

<table>
<thead>
<tr>
<th>Groups/genotypes</th>
<th>n</th>
<th>2e (n = 7)</th>
<th>2f (n = 2)</th>
<th>10a (n = 10)</th>
<th>11a (n = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>18</td>
<td>66.2–68.9</td>
<td>67.4–69.5</td>
<td>68.0–72.0</td>
<td>70.4–73.2</td>
</tr>
<tr>
<td>(I/1a, II/1b, 1c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>2</td>
<td>83.5–85.9</td>
<td>84.7–85.6</td>
<td>69.3–70.4</td>
<td>68.2–69.3</td>
</tr>
<tr>
<td>III/2a</td>
<td>2</td>
<td>82.2–83.3</td>
<td>82.9–83.8</td>
<td>69.1–71.5</td>
<td>69.3–70.4</td>
</tr>
<tr>
<td>IV/2b</td>
<td>2</td>
<td>86.2–87.5</td>
<td>86.9–87.6</td>
<td>69.5–70.5</td>
<td>68.3–69.0</td>
</tr>
<tr>
<td>2c</td>
<td>1</td>
<td>66.2–67.5</td>
<td>66.8–68.2</td>
<td>78.0–81.0</td>
<td>71.4–72.4</td>
</tr>
<tr>
<td>Group 3</td>
<td>5</td>
<td>67.7–69.2</td>
<td>67.0–68.4</td>
<td>78.4–81.3</td>
<td>71.3–72.9</td>
</tr>
<tr>
<td>V/3a</td>
<td>4</td>
<td>66.4–67.7</td>
<td>66.5–67.2</td>
<td>80.1–82.2</td>
<td>71.7–73.1</td>
</tr>
<tr>
<td>3b</td>
<td>1</td>
<td>66.8–68.6</td>
<td>68.1–68.6</td>
<td>80.6–82.9</td>
<td>72.6–73.0</td>
</tr>
<tr>
<td>3c</td>
<td>1</td>
<td>67.2–68.3</td>
<td>67.8–68.0</td>
<td>74.7–79.9</td>
<td>71.3–72.1</td>
</tr>
<tr>
<td>3d</td>
<td>1</td>
<td>68.6–69.2</td>
<td>69.2–69.3</td>
<td>80.6–82.0</td>
<td>71.8–72.1</td>
</tr>
<tr>
<td>3e</td>
<td>1</td>
<td>68.4–69.7</td>
<td>69.0–69.6</td>
<td>73.9–75.0</td>
<td>73.4–73.6</td>
</tr>
<tr>
<td>3f</td>
<td>1</td>
<td>68.4–69.7</td>
<td>69.0–69.6</td>
<td>73.9–75.0</td>
<td>73.4–73.6</td>
</tr>
<tr>
<td>Group 4</td>
<td>5</td>
<td>69.2–70.0</td>
<td>70.9–71.4</td>
<td>71.3–72.1</td>
<td>70.1–70.8</td>
</tr>
<tr>
<td>Group 5</td>
<td>6</td>
<td>67.8–69.7</td>
<td>68.3–71.0</td>
<td>71.0–73.5</td>
<td>75.9–79.1</td>
</tr>
<tr>
<td>(6a, 6b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 7</td>
<td>4</td>
<td>66.4–68.9</td>
<td>67.2–68.8</td>
<td>69.4–71.5</td>
<td>78.8–80.1</td>
</tr>
<tr>
<td>7a</td>
<td>1</td>
<td>66.5–68.4</td>
<td>66.7–68.1</td>
<td>70.4–71.8</td>
<td>78.7–79.2</td>
</tr>
<tr>
<td>7b</td>
<td>1</td>
<td>65.9–68.6</td>
<td>66.7–67.6</td>
<td>68.9–71.2</td>
<td>78.4–80.7</td>
</tr>
<tr>
<td>7c</td>
<td>1</td>
<td>67.9–69.6</td>
<td>67.7–68.5</td>
<td>70.5–71.5</td>
<td>80.2–81.0</td>
</tr>
<tr>
<td>7d</td>
<td>1</td>
<td>67.9–69.6</td>
<td>67.7–68.5</td>
<td>70.5–71.5</td>
<td>80.2–81.0</td>
</tr>
<tr>
<td>Group 8</td>
<td>4</td>
<td>67.2–69.0</td>
<td>68.2–68.8</td>
<td>69.8–71.6</td>
<td>76.3–78.6</td>
</tr>
<tr>
<td>(8a, 8b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 9</td>
<td>5</td>
<td>66.3–69.2</td>
<td>67.4–69.4</td>
<td>68.9–71.6</td>
<td>74.6–76.9</td>
</tr>
</tbody>
</table>

from each other by 6.1%, were classified as genotype 2f. These two new genotypes had a similarity of only 83.4–86.1%, and were distinct from genotypes in genetic group 2, such as III/2a, IV/2b and 2c, based on the above criteria. Genotypes 2e and 2f also differed from genotype 2d (Stuyver et al., 1994), for which the 1093 bp sequence is not available, based on sequence similarity of only 73.2–78.6% in a 384 bp stretch of the E1 gene and 82.0–85.6% in a 334 bp stretch of NS5b.

Genotype 10a, represented by ten Jakarta isolates, resembled genotypes in group 3 with a sequence similarity of 77.4–82.9%, which is on the boundary of group and genotype similarity. The 10 isolates differed from the other groups, however, with a similarity of < 75%. Likewise genotype 11a, represented by three Jakarta isolates, resembled group 7 with a similarity of 78.4–81.0%; it differed from all the others with a similarity of < 79.1%.

In order to classify the 22 Jakarta isolates of novel genotype by another method, they were subjected to phylogenetic analysis, within 1093 bp (nt 8279–9371) in the NS5b region, along with 62 HCV isolates of 25 genotypes in nine genetic groups (Fig. 1). The phylogenetic tree supported the classification of nine Jakarta isolates into genotypes 2e and 2f. Ten Jakarta isolates (JK012, JK030, JK049, JK055, JK070, JK072, JK092, JK093, JK111 and JK114) made a group which bifurcated from the same branch as the genetic group 3. They were separated from group 3 at an evolutionary distance of 0.236, and therefore were classified into a putative new genetic group designated group 10. The ten Jakarta isolates, with a distance of < 0.068 from each other, were of the same genotype which was provisionally named 10a.

The remaining three Jakarta isolates (JK046, JK065 and JK148) were considered to form another new genetic group tentatively designated 11. It was on the same branch as group 7, but separated by a distance of 0.239. Since this distance was much closer to the value of 0.246, which separates group 8 from group 9, than the range 0.191–0.214 which separates genotypes of group 7 (7a–7d), it would constitute a new genetic group.

Determination of the entire genomic sequences of Jakarta isolates of genotypes 10a and 11a

In order to confirm the new genetic groups 10 and 11, isolate JK049 of genotype 10a and isolate JK046 of genotype 11a were sequenced in full (accession nos D63821 and D63822, respectively). JK049 had a genomic length of 9433 bp [poly(T) tail excluded] with a 5'UTR and a 3'UTR of 339 and 37 bp, respectively. Its open
reading frame coded for a 3019 amino acid (aa) precursor which could be C, E1, NS2, NS3, NS4a, NS4b and NS5b proteins of sizes identical to those in other genotypes, except for E2/NS1 with 431 aa and NS5a with 451 aa. The genomic length of JK046 was 9439 bp [poly(T) tail excluded] with a 5'UTR and a 3'UTR of 338 and 35 bp, respectively. The open reading frame coded for a 3022 aa precursor with E2/NS1 of 430 aa and NS5a of 455 aa, which, as in JK049, differed in size from those previously reported.

The entire genomic sequences of JK049 (genotype 10a) and JK046 (genotype 11a), as well as their various subgenomic areas, were subjected to pair-wise comparison with the 22 HCV genomes for which the full-length sequences are known (Fig. 2). Within the entire genomic sequence, group similarity spanned 65.0–75.0%, genotype similarity 76.9–80.0% and isolate similarity 90.1–99.1%, which were clearly separated without overlapping. Such a clear separation was seen only in the comparison of NS3 and NS5b regions among various subgenomic areas.

JK049 showed a similarity of 74.4–75.0% to genomes of HCV isolates of genotype V/3a (NZL1 and K3a/650) or 3b (HCV-Tr) within the entire genomic sequence, which justified its classification into a new genetic group separate from group 3, as deduced from phylogenetic analysis of the 1093 bp NS5b region. Likewise, JK046 was only 66.2–70.0% similar to reported HCV genomes in the entire genomic sequence, and therefore was in a novel genetic group. JK049 was 67.5% similar to JK046, thereby indicating that these two isolates belong to distinct genetic groups.

Comparison of 5'-terminal 1.6 kb and 3'-terminal 1.1 kb stretches among Jakarta isolates of four novel genotypes

Four Jakarta isolates (JK020, JK025, JK109 and JK128) were selected randomly from the seven of genotype 2e, as were four (JK030, JK055, JK070 and JK072) from the ten of genotype 10a. Along with both Jakarta isolates of genotypes 2f (JK081 and JK139) and 11a (JK065 and JK114), they were sequenced within 5'-terminal 1.6 kb and 3'-terminal 1.1 kb stretches which covered approximately 30% of the entire genome. They were compared along with JK049 and JK046 for which the entire genomic sequences had been determined.

Jakarta isolates of the same genotype (2e, 2f, 10a or 11a) showed a similarity in nucleotide sequence of >91.8% and in amino acid sequence of >91.9%. As in the other genotypes, the 5'UTR sequence was most conserved with a similarity >99.1%, while the 3'UTR sequence was most divergent. The hypervariable region 1 in E2/NS1 (Hijikata et al., 1991; Okamoto et al., 1992a) spanned 72–84 bp; it differed by 27–54% in nucleotide
Hepatitis C virus genotypes

Fig. 2. Pair-wise comparison of the 24 HCV genomes of various genotypes over the nucleotide sequence of the entire genome and subgenomic regions. Ranges of group, genotype and isolate similarity are shown for the entire genome and for various subgenomic regions. The comparison was made among the 22 HCV genomes for which the full-length sequences have been reported, and for isolates JK049 of genotype 10a and JK046 of genotype 11a. The 22 genomes with known full-length sequences are classified in three genetic groups and seven different genotypes and are deposited under the following designations and accession numbers: M62321 (HCV-1), M67463 (HCV-H) and D10749 (HCV-JI) of genotype 1/1a; D90208 (HCV-J2), M58335 (HCV-BK), M84754 (HCV-T), D01217 (HC-J4/83), D10750 (HC-J4/91), D01171 (HCV-JT), D01172 (HCV-JT), X61596 (HCV-JK1), D10934 (HC-C2), S62220 (HCV-N), L02836 (HCV-HB), M96362 (HCV-L1) and U01214 (HCV-L2) of II/1b; D14853 (HC-G9) of 1c; D00944 (HC-J6) of III/2a; D01221 (HC-JS) of IV/2b; D17763 (NZL1) and D28917 (K3a/650) of V/3a; and D26556 (HCV-Tr) of 3b.

Fig. 3 compares the sequence of the 3'UTR [poly(T) tail excluded] among HCV isolates of known genotype including the four novel Jakarta genotypes. All 22 Jakarta isolates had a stop codon at nucleotide positions 9372–9374: TAG for genotypes 2e, 2f and 10a and TAA for genotype 11a. A second in-phase stop and 40–79% in amino acid sequences even among Jakarta isolates of the same genotype. JK148 of genotype 11a had an in-phase insertion of 9 bp, while the other two isolates of genotype 11a (JK046 and JK046) and three of the four isolates of genotype 2e (JK020, JK109 and JK128) had an in-phase deletion of 3 bp at the start of the E2/NS1 protein. The four novel genotypes of Jakarta isolates were characterized further in distinct regions of the genome in comparison with reported genotypes.

5'UTR. Of four Jakarta isolates of genotype 2e sequenced, three (JK025, JK109 and JK128) had an insertion of A after nt 14 of the 5'UTR, resulting in a length of 342 bp. By contrast, both isolates of genotype 2f lacked nt 6, resulting in a 5'UTR of 340 bp. All five isolates of genotype 10a had a deletion of 2 bp (nt 13 and 14) which shortened the length of the 5'UTR to 339 bp, and all three of genotype 11a were 338 bp long due to the absence of nt 6, 15 and 35 as seen in isolates of genotypes 7a–7c, 8a, 8b and 9a–9c. All six Jakarta isolates of genotype 2e or 2f had A as nt 210 and T as nt 214, which were shared by every reported isolate in group 2 but not by any genotypes in the other groups.

Amino acid sequence of the E1 protein. Fig. 3 compares the entire amino acid sequence of the E1 protein of HCV isolates of known genotypes and the 14 Jakarta isolates of four novel genotypes. All Jakarta isolates of genotypes 2e, 2f and 10a possessed the five potential N-glycosylation sites, Asn at amino acid positions 196, 209, 234, 305 and 325, as reported for isolates of genotypes 1/1a, 1c, III/2a, V/3a, 3b, 3d–3f, 4a–4d and 5a. All three Jakarta isolates of genotype 11a had another potential N-glycosylation site, Asn-250, as reported for isolates of genotypes II/1b, 6a, 7a–7d, 8a, 8b and 9a–9c. All 14 Jakarta isolates of novel genotype conserved the eight Cys residues at positions 207, 226, 229, 238, 272, 281, 304 and 306 as did the other isolates of all other genotypes. In addition, all five Jakarta isolates of genotype 10a had Cys-376, which is not found in any isolates of group 3.

Jakarta isolates of genotypes 2e and 2f highlighted amino acids specific for group 2. These were Trp-Gln-Leu at positions 214–216, as well as Val-220, Gln-223, Asn-245, Leu-254, Met-285 and Phe-300. His-336, which is characteristic of group 3, was not found in isolates of any of the other groups; it was not shared by Jakarta isolates of genotype 10a, either. Gly-290 and Arg/His-295 and Cys-376 were specific for genotype 10a, as was Gln-260 for 11a. These amino acids reinforced the distinctness of 10a and 11a from genotypes in groups 3 and 7, respectively, although they were close in the phylogenetic tree (Fig. 1) and similar in sequence (Table 1).

3'UTR. Fig. 4 compares the sequence of the 3'UTR [poly(T) tail excluded] among HCV isolates of known genotype including the four novel Jakarta genotypes. All 22 Jakarta isolates had a stop codon at nucleotide positions 9372–9374: TAG for genotypes 2e, 2f and 10a and TAA for genotype 11a. A second in-phase stop
Table 1: Representative HC-V isolates of 29 genotypes are compared along with 14 Jakarta isolates of novel genotypes 2e, 2f, 10a and 11a. Potential N-glycosylation sites are boxed, and Cys residues are marked by dots. Amino acid sequences characteristic of groups 2, 3, 10 and 11 are shaded.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Amino Acid Sequence</th>
<th>Characteristic Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>298</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 3:** Amino acid sequences of the entire E1 protein, as represented by selected HC-V isolates, are compared along with 14 Jakarta isolates of novel genotypes 2e, 2f, 10a and 11a. Potential N-glycosylation sites are boxed, and Cys residues are marked by dots. Amino acid sequences characteristic of groups 2, 3, 10 and 11 are shaded.
Fig. 4. Nucleotide sequences of the 3'UTR. Sequences of representative HCV isolates of 25 genotypes are compared along with 22 Jakarta isolates of novel genotypes 2e, 2f, 10a and 11a. The CACTCC motif is shaded, and in-frame stop codons are underlined. Dashes represent nucleotides that are the same as in HC-J1 of genotype I/1a shown at the top.

codon was found in six of the seven Jakarta isolates of genotype 2e and in both isolates of genotype 2f; it was TAG at positions 9387–9389 as in isolates of genotype III/2a or IV/2b.

All ten Jakarta isolates of genotype 10a had a TAA stop codon at nt 9384–9386 as in isolates of genotypes I/1a, II/1b, 5a and 9a. Along with the presence of a second stop codon, this makes the genotype 10a isolates markedly different from isolates of genotypes in group 3.

All three Jakarta isolates of genotype 11a lacked a second stop codon as reported for isolates of genotypes in groups 3 and 6–9 (with the exception of 9a). The length of the 3'UTR before the poly(T) tail was 38–42 bp for genotype 2e, 44 bp for 2f, 22–33 bp for 10a and 34–35 bp for 11a. Thus, genotype 10a had a short 3'UTR comparable with isolates in groups 3–9.

Discussion

Classification of HCV into genotypes now faces a critical turning point. Genotypes defined by pair-wise comparison and phylogenetic analysis of nucleotide sequences, within the entire genome or over shorter stretches, now number at least 34. Enduring enthusiasm
to pursue new HCV variants is driven by the high prevalence of HCV throughout the world and its strong disease-inducing activity, resulting in chronic hepatitis in > 80% of those who become infected. An endless list of new genotypes is being logged, whenever HCV variants are obtained from previously unexamined areas of the world (Bukh et al., 1995). A long-standing dilemma of taxonomists is that they are inclined to be either ‘splitters’ or ‘lumpers’. Are we going to split HCV genotypes any further when there appears to be little, if any, definitive virological or clinical significance known for them? Alternatively, should we start to lump them together into fewer categories based on common features in various regions of the genome?

The 22 HCV isolates from Jakarta, Indonesia of four novel genotypes, 2e, 2f, 10a and 11a, may provide some insight into the way forward for HCV genotyping. First, they allow us to argue against the two-tiered classification of HCV genotypes presently advocated. We have reported group similarities (66.5–80.0%) which did not overlap with the genotype similarities (80.2–88.6%) in a comparison of a 1093 bp NS5b sequence spanning nt 8279–9371 (Tokita et al., 1994b, 1995). The Jakarta HCV isolates of genotypes 10a and 11a, however, were 77.4–82.9% and 78.7–81.0% similar, respectively, to genotypes in groups 3 and 7 within this sequence. Therefore it is questionable whether groups 10 and 11 are independent or should be classified as genotypes in groups 3 and 7, respectively (Table 1).

For the purpose of establishing these two genetic groups, one Jakarta isolate of each of the suggested genotypes (10a or 11a) was sequenced in full, and compared with 22 HCV isolates for which the entire genomic sequences are known. The comparison substantiated the new genetic groups, which shared only < 75.0% and < 70.0% of the sequence, respectively, of genomes from the other genetic groups. In these comparisons, it has become evident that sequence divergence over the entire genome can distinguish genetic groups and genotypes most clearly. Two subgenomic regions, NS3 and NS5b, can be used to separate group from genotype similarity without overlapping, but with margins much less than those obtained by comparison of the entire genome. These results highlight the limitations of distinguishing genotypes by comparison of subgenomic sequences, and underscore the need to compare entire genomic sequence when new genotypes are proposed. Complete sequences for genetic groups 4–9 are needed urgently.

Although HCV isolates are classified into eleven genetic groups (or types) by us (Tokita et al., 1994a, b, 1995; present report) and six groups by P. Simmonds and co-workers (Simmonds et al., 1993; Simmonds, 1995), i.e. 1, 5, 3 (group 10 included), 4, 6 (groups 7, 11, 8 and 9 included), and 2, such groupings must be further evaluated for validity as sequence data for new HCV isolates are accumulated. Based on the phylogenetic analysis on the 1093 bp NS5b sequence, five groups (1; 5; 3, 10 and 4; 6, 7, 11, 8 and 9; 2), four groups (1 and 5; 3, 10 and 4; 6, 7, 11, 8 and 9; and 2) or even two groups (2 and all the others) can be distinguished by setting an arbitrary distance for grouping (Fig. 1).

From a practical point of view, only five of ≥ 34 genotypes prevail in most areas of the world. The five common genotypes, 1/1a, II/1b, III/2a, IV/2b and V/3a, account for essentially all HCV isolates in Europe and North America as well as North and Far East Asia and Oceania (Bukh et al., 1995; Miyakawa et al., 1995; Simmonds, 1995). Among all HCV genotypes presently recorded, only genotype II/1b (group 1) is associated with severe hepatitis and poor response to interferon (Yoshioka et al., 1992; Hino et al., 1994). It remains to be seen whether or not such clinically important features of genotype II/1b are shared by other genotypes in genetic group 1. Some groups of investigators report differences in the response to interferon between patients infected with HCV genotype I/1a and those infected with II/1b (Qu et al., 1994; Nousbaum et al., 1995), while others do not (Mahaney et al., 1994). Additional clinical and therapeutic trials are required to settle this.

The time may have come when we must start to ponder the future for logging new HCV genotypes by an intricate classification based on sequence divergence in partial sequences of the genome. There is concern that such an effort may overlook the true genomic divergence which may be defined only by comparing the entire genomic sequences. Furthermore, it can potentially distract from essential virological characters of HCV, and can isolate HCV genotyping from other fields of medicine. For exact classification of HCV further studies are required, in which virologists and clinicians proceed in close cooperation.

References


Hepatitis C virus genotypes


(Received 8 June 1995; Accepted 29 September 1995)