Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8234–8238, September 1993
Biochemistry

At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide

(non-A, non-B hepatitis/genetic heterogeneity/polymerase chain reaction/dendrogram/taxonomy)

JENS BUKH, ROBERT H. PURCELL, AND ROGER H. MILLER

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Robert H. Purcell, June 4, 1993

ABSTRACT **In a previous study we sequenced the 5′ noncoding (NC) region of 44 isolates of hepatitis C virus (HCV) and identified heterogeneous domains that provided evidence for additional genetic groups of HCV not previously recognized. In this study we have determined the complete nucleotide sequence of the putative envelope 1 (E1) gene in 51 HCV isolates from around the world and found that they could be grouped into at least 12 distinct genotypes. The E1 gene sequence of 8 of these genotypes has not been reported previously. Although the genetic relatedness of HCV isolates determined by the previous analysis of the 5′ NC region predicted the relationships observed in the E1 gene, analysis of the 5′ NC sequence alone did not accurately predict all HCV genotypes. The nucleotide and amino acid sequence identities of the E1 gene among HCV isolates of the same genotype were in the range of 88.0–99.1% and 89.1–98.4%, respectively, whereas those of HCV isolates of different genotypes were in the range of 53.5–78.6% and 49.0–82.8%, respectively. The latter differences are similar to those found when comparing the envelope gene sequences of the various serotypes of the related flaviviruses as well as other RNA viruses. We found that some genotypes of HCV were widely distributed around the world, whereas others were identified only in discreet geographical regions. Four genotypes were identified exclusively in Africa and comprised the majority of HCV isolates on that continent. The E1 gene was exactly 576 nucleotides in length in all 51 HCV isolates with no in-frame stop codons. Analysis of the predicted E1 protein identified several conserved domains that may be important for maintaining its biological function: (*i*) eight invariant cysteine residues, (*ii*) three potential N-linked glycosylation sites, (*iii*) a domain of nine amino acids (GHRMAWDMM), and (*iv*) an amino acid doublet (GV) near the putative cleavage site at the C terminus of the protein. In conclusion, the discovery of at least 12 genotypes of HCV has important implications for HCV diagnosis and vaccine development.**

Hepatitis C virus (HCV), the etiological agent of parenterally transmitted non-A, non-B hepatitis, contains a positive-stranded linear RNA genome ≈9.5 kb in length (1). The genomic structure and organization of HCV are similar to that of pesti- and flaviviruses with 5′ and 3′ noncoding (NC) regions enclosing a single long open reading frame. The gene order is core (C), envelope 1 (E1) glycoprotein, envelope 2 or nonstructural 1 (E2/NS1) glycoprotein, and nonstructural 2–5 (NS2–NS5) proteins (reviewed in ref. 2). Recently, Okamoto and coworkers (3) found that all HCV isolates available could be assigned to four genotypes based on analysis of the complete genome sequence of seven isolates: genotypes I (1, 4), II (5–7), III (8), and IV (3). The distinct differences among the four genotypes were found to be maintained throughout the protein coding region of the HCV genome and were most significant in the E1 and NS2 genes.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

HCV appears to be most closely related to pesti- and flaviviruses based on colinear nucleotide and protein sequence similarities (1, 2, 5, 9). We previously demonstrated that two nucleotide domains within the 5′ NC region of HCV share significant similarity with 5′ NC sequences of pestiviruses and found that these domains are highly conserved among different HCV isolates (10, 11). Interestingly, we identified several HCV isolates with unique 5′ NC sequences not previously reported that were significantly different from those of genotypes I–IV. In this study, we have determined the nucleotide sequence of the entire E1 gene of the virus genome of 51 HCV isolates* including those possessing the distinct 5′ NC sequences observed previously. Based on comparison of the nucleotide and predicted amino acid sequences of the E1 gene, we find that these HCV isolates can be assigned to at least 12 different genotypes.

MATERIALS AND METHODS

Serum used in this study was obtained from 84 individuals positive for antibodies to HCV (anti-HCV) that we previously found were positive for HCV RNA in a cDNA PCR assay with primer set *a* from the 5′ NC region of the HCV genome (10). These samples were from 12 countries: Denmark (DK), Dominican Republic (DR), Germany (D), Hong Kong (HK), India (IND), Sardinia, Italy (S), Peru (P), South Africa (SA), Sweden (SW), Taiwan (T), United States (US), and Zaire (Z). Viral RNA was extracted from 100 μl of serum and the final RNA solution was divided into 10 equal aliquots and stored at −80°C as described (10). The sequences of the synthetic oligonucleotides used in the cDNA PCR assay, deduced from the sequence of HCV strain H-77 (12), were as follows: 5′-GCGTCCGGGTTCTGGAAGACGGCGTGAACTATG-CAACAGG-3′ from nt 461–500 (external sense primer, e_1); 5′-AGGCTTTCATTGCAGTTCAAGGCCGTGCTATT-GATGTGCC-3′) from nt 1298–1259 (external antisense primer, e_2); 5′-AAGACGGCGTGAACTATGCAACAGG-GAACCTTCCTGGTTG-3′ from nt 476–515 (internal sense primer, e_3); 5′-AGTTCAAGGCCGTGCTATTGATGTGC-CAACTGCCGTTGGT-3′ from nt 1285–1246 (internal antisense primer, e_4); 5′-AAGACGGCGTGAATTCTGCAA-CAGGGAACCTTCCTGGTTG-3′ from nt 476–515 (internal sense primer with *Eco*RI site, $e_{3\text{EcoRI}}$); 5′-AGTTCAAGGC-CGTGGAATTCATGTGCCAACTGCCGTTGGT-3′ from nt 1285–1246 (internal antisense primer with *Eco*RI site, $e_{4\text{EcoRI}}$). One aliquot of the RNA solution was used for cDNA synthesis and the resulting cDNA was amplified in a "nested" PCR assay by *Taq* DNA polymerase (Amplitaq, Perkin–Elmer/Cetus) as described (10) with primer set *e* (primers e_1, e_2, e_3, and e_4). In most instances, amplified DNA (first or second PCR products) was reamplified with primers $e_{3\text{EcoRI}}$

Abbreviations: HCV, hepatitis C virus; NC, noncoding; E1, envelope 1.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L16628–L16678).

and e_{4EcoRI} prior to sequencing in order to facilitate future cloning of the E1 gene. Amplified DNA was purified by gel electrophoresis followed by glass-milk extraction (Geneclean, Bio 101) and both strands were sequenced directly by the dideoxynucleotide chain-termination method (13) with phage T7 DNA polymerase (Sequenase, United States Biochemical), dATP[α-^{35}S] (Amersham) or [α-^{33}P]dATP (Amersham or DuPont), and specific sequencing primers. RNA extracted from serum containing HCV strain H-77, previously sequenced in our laboratory (12), was amplified with primer set *e* and sequenced in parallel as a control. Data were analyzed with the computer programs GENALIGN (9) for multiple sequence alignment and CLUSTAL (14) for dendrogram construction.

RESULTS AND DISCUSSION

In a previous study we sequenced the 5′ NC region of 44 HCV isolates and reported that considerable genetic heterogeneity existed even within this highly conserved region of the HCV genome (11). We found that this heterogeneity was interspersed between regions of universal sequence conservation and postulated that the conserved regions represented important control elements and that the heterogeneous regions (which included insertions in some cases) provided strong evidence for additional genetic groups of HCV not previously recognized (Fig. 1). One such group of genetically distinct sequences was also described at approximately the same time by others (15–18). Prior to our previous study, representative isolates of only three genetic groups had been sequenced in their entirety (1, 4–6, 8). Subsequently, the fourth genotype was entirely sequenced (3) and genotype designations I–IV were assigned to these completely sequenced genomes. Thus, Okamoto *et al.* (3) set a precedent of assigning Roman numeral designations to completely sequenced genotypes. The fifth distinct genotype mentioned above was provisionally designated genotype V based upon only partial sequence analysis. Chan *et al.* (15) proposed a new nomenclature for the classification of HCV in which genotypes I and II were grouped together into genotype 1 and genotypes III and IV were similarly grouped into genotype 2. The provisional

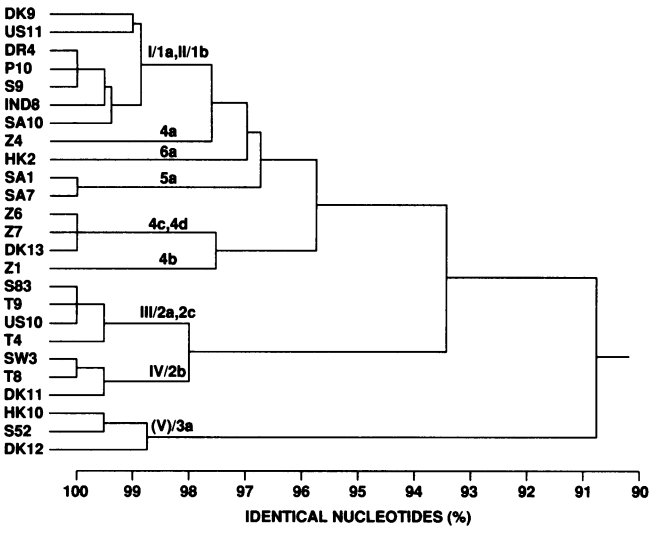

FIG. 1. Dendrogram of the genetic relationship of HCV isolates based on nucleotide identity of 5′ NC sequences (nt −246 to −51) from a study of 44 HCV isolates (11). This dendrogram was constructed by the program CLUSTAL (14), with a limit of 25 sequences and a gap penalty of 0. It should be noted that isolate HK2 had two separate nucleotide insertions. The scale showing percent identity was based upon manual calculation. Genotype designations of these HCV isolates, based upon analysis of the entire E1 gene region of the HCV genome, are indicated.

genotype V was reclassified as genotype 3. Since the Okamoto classification is based upon sequencing of entire genomes, whereas the Chan classification is based upon limited sequencing data and an alternative grouping of types, we have used both nomenclatures in the present study but wish to point out the provisional nature of these classification schemes and the need for more data and international agreement before an official classification scheme is selected.

The purpose of the present investigation was to analyze the nucleotide and deduced amino acid sequences of an important structural gene region of the HCV genome, the E1 gene, of the genetically heterogeneous isolates that we identified previously by analysis of the 5′ NC region (11). By careful analysis of available HCV genomic sequences (1, 3–6, 8, 12, 19–23), we were able to design a primer set (*e*) spanning the entire E1 gene from the most highly conserved nucleotide domain identified in the 3′ end of the C gene to the 5′ end of the E2/NS1 gene. With primer set *e*, we were able to reverse-transcribe and, by PCR, amplify HCV sequences from 74 (88.1%) of the 84 samples that were found to be positive previously with the highly conserved 5′ NC primers among samples from 114 anti-HCV positive individuals (10). All 84 negative control samples interspersed among the test samples were negative for HCV RNA. This, plus the fact that no two E1 sequences were identical, ruled out contamination as a source of the positive results. Thus, we successfully amplified the entire E1 gene from isolates representing all of the genetic groups identified in Fig. 1.

We report here the nucleotide and deduced amino acid sequences of the putative E1 (nt 574–1149, aa 192–383; see ref. 2) of 51 HCV isolates. A multiple sequence alignment of the nucleotide (not shown) and the deduced amino acid sequences (Fig. 2) of the E1 gene of these isolates was performed. In all 51 HCV isolates, the E1 gene was exactly 576 nt in length and did not have any in-frame stop codons. The 51 HCV isolates comprised 12 genotypes, based upon the degree of variation of the E1 gene sequence. A dendrogram of the genetic relatedness of the E1 protein of selected HCV isolates representing the 12 genotypes is shown in Fig. 3. The nucleotide and amino acid sequence identities of HCV isolates of the same genotype were in the range of 88.0–99.1% and 89.1–98.4%, respectively, whereas those of HCV isolates of different genotypes were in the range of 53.5–78.6% and 49.0–82.8%, respectively (Table 1). Thus, analysis of the E1 gene sequence of 51 HCV isolates demonstrated extensive heterogeneity of this important gene. The remaining 23 of the 74 isolates that could be amplified by primer set *e* were partially sequenced and found to be genotype I/1a or II/1b. The distribution of the 12 genotypes among the 74 HCV isolates from around the world is depicted in Fig. 4. Genotypes I/1a and II/1b were the most common genotypes found, accounting for 48 (64.9%) of the 74 isolates. Analysis of the E1 gene sequences available in the GenBank data base at the time of this study (1, 3–8, 12, 18–27) revealed that all 44 such sequences were of genotypes I–IV (1a, 1b, 2a, and 2b). Thus, based upon E1 gene analysis, we have identified 8 previously unreported genotypes of HCV.

Different HCV genotypes were frequently found in the same country. The highest number of genotypes (five) was detected in Denmark. Genotypes I/1a, II/1b, III/2a, IV/2b, and V/3a were widely distributed. Most impressive, genotype II/1b was identified in 11 of 12 countries (Zaire was the only exception). Although genotypes I/1a or II/1b were predominant in the Americas, Europe, and Asia, several previously unidentified genotypes were predominant in Africa.

The genotype analysis using the entire nucleotide sequence of the E1 gene of 51 HCV isolates was in complete agreement with that using the predicted amino acid sequence of the E1 protein. Further analysis using segment lengths of 100 nt (i.e.,

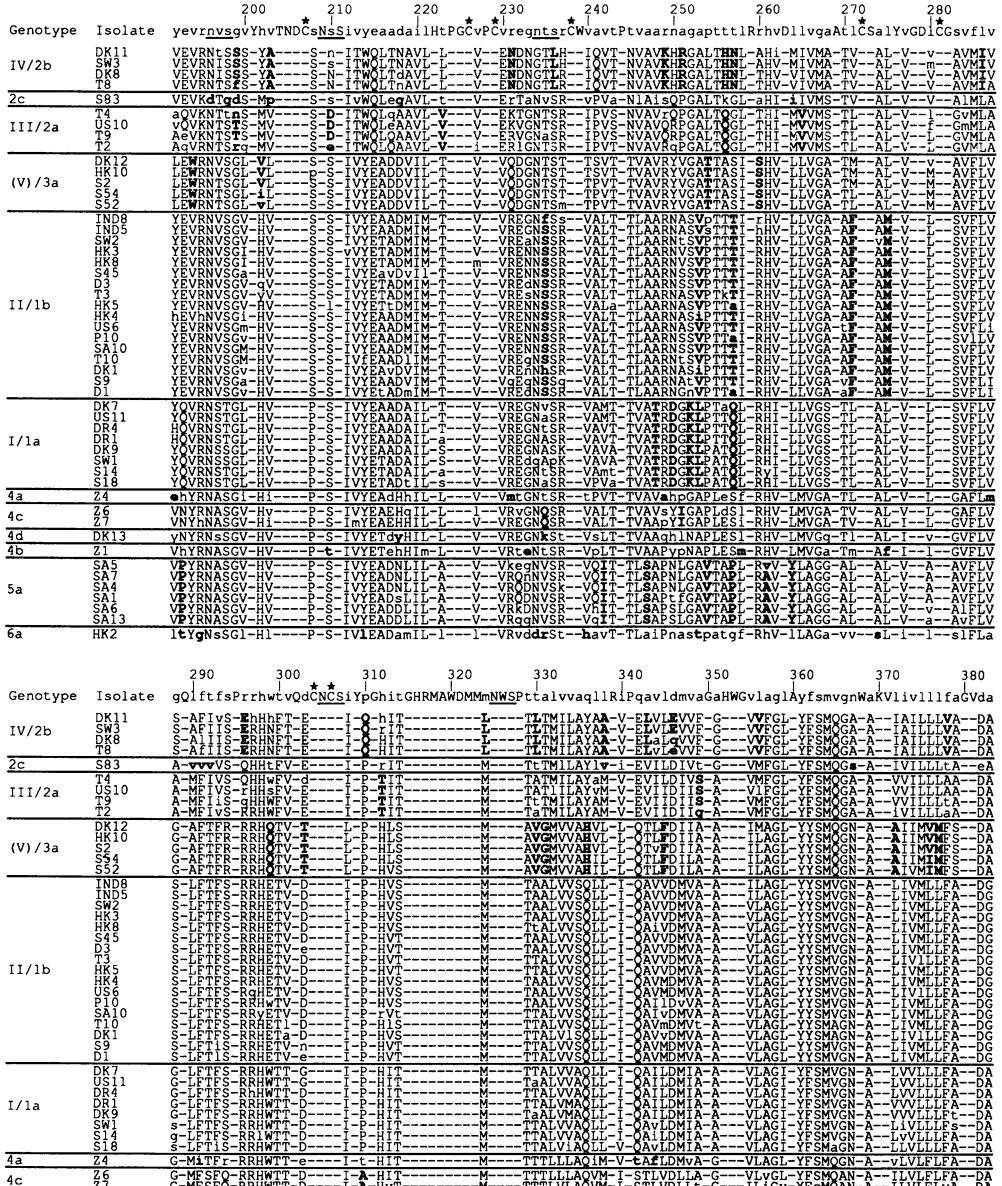

FIG. 2. Multiple sequence alignment of the deduced amino acid sequence of the E1 gene of 51 HCV isolates collected worldwide. The consensus sequence of the E1 protein is shown at the top with cysteine residues highlighted with stars, potential N-linked glycosylation sites underlined, and invariant amino acids capitalized. Variable amino acids are shown in lowercase letters. In the alignment, amino acids are shown in lowercase letters if they differed from the amino acids of both adjacent isolates. Amino acids that were invariant among all HCV isolates are shown as dashes (–) in the alignment. Map positions correspond to those of the HCV prototype sequence (HCV-1; ref. 1) with the first amino acid of the E1 protein at position 192. The sites at which all amino acids are unique for an individual genotype are highlighted by printing the amino acid at that site in boldface type. The grouping of isolates into 12 genotypes is indicated. When microheterogeneity in a sequence was observed, defined as more than one prominent nucleotide at a specific position, we reported the nucleotide that was identical to that of the HCV prototype (HCV1, ref. 1) if possible. Alternatively, we reported the nucleotide that was identical to the most closely related isolate. Microheterogeneity was observed in only 20 of the 51 HCV isolates at 0.2–2.6% of the 576 nucleotide positions of E1 and resulted in changes in 0.5–2.6% of the 192 amino acids in only 10 of these isolates. Changes resulting from microheterogeneity did not result in an appreciable change in the results of the analysis.

nt 574–673, 624–723, 674–773, 724–823, 774–873, 824–923, 874–973, 924–1023, 974–1073, and 1024–1123) revealed that characteristic genotypic changes were seen throughout the gene. Although analysis of each of these segments could predict the basic grouping of HCV isolates observed by analysis of the complete E1 gene sequence, it did not predict the exact genetic relationships observed by analysis of the complete gene. Furthermore, we identified a domain of only 31 nt (i.e., nt 1023–1053) within the E1 gene that was predictive of the division of all of our 51 HCV isolates into 12 genotypes. It is noteworthy that analysis of nt 987–1022, proposed by Cha *et al.* (18) to contain a motif predictive of genotype, could not accurately predict the division of all 51 isolates into the 12 genotypes.

Analysis of the consensus sequence of the E1 protein of the 51 HCV isolates from this study demonstrated that a total of 60 (31.3%) of the 192 amino acids of the E1 protein was invariant among these isolates (Fig. 2). Most impressive, all eight cysteine residues, possibly used to maintain the conformation of the E1 protein through disulfide linkages, were invariant among all HCV isolates, as were six of eight proline residues. It is noteworthy that the cysteine residues of the structural proteins of the related flaviviruses are also highly conserved (28). The most abundant amino acids (e.g., alanine, valine, and leucine) showed a very low degree of conservation. The consensus sequence of the E1 protein contained five potential N-linked glycosylation sites (N-X-T/S). Three sites, at positions 209, 305, and 325, were maintained in all 51 HCV isolates. Further analysis revealed a highly conserved amino acid domain (aa 302–328) in the E1 protein with 20 (74.1%) of 27 amino acids invariant among all 51 HCV isolates. It is possible that the 5′ and 3′ ends of this domain are conserved because of important cysteine residues and N-linked glycosylation sites. The central sequence, 5′-GHRMAWDMM-3′ (aa 315–323), may be conserved because of additional functional constraints on the protein structure. Finally, although the amino acid sequence surrounding the putative E1 protein cleavage site was variable, an amino acid

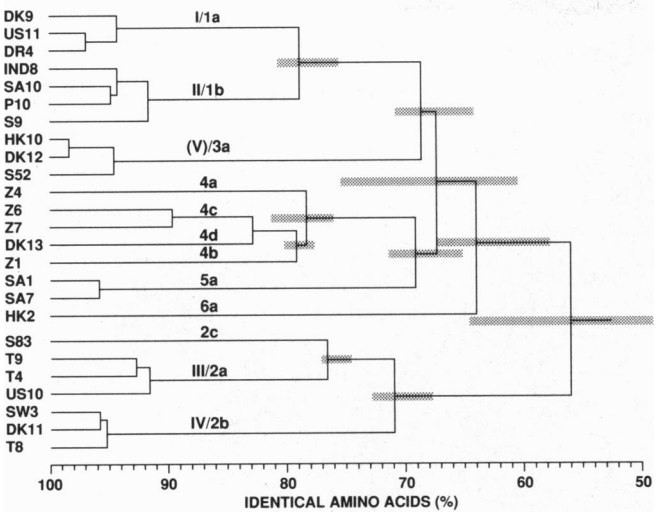

FIG. 3. Dendrogram of the genetic relatedness of selected HCV isolates based on the percent amino acid identity of the putative E1 protein. The selected HCV isolates, the same isolates analyzed in Fig. 1, represent the 12 different genotypes. The scale showing percent identity was based upon manual calculation. The range of amino acid sequence identity of the putative E1 protein among all HCV isolates of different genotypes is depicted as a shaded bar at major branch points (see Table 1).

doublet (GV) at position 380 was invariant among all HCV isolates. Thus, despite the high degree of sequence heterogeneity among the E1 proteins of the different genotypes of HCV, there are conserved domains that may be important for the function of the protein.

In this study we have grouped 74 HCV isolates into 12 genotypes based on the nucleotide and deduced amino acid sequence of the putative E1 gene. These data confirm our previous finding of additional genotypes based upon analysis of the 5′ NC region (11). As seen in Fig. 1, nine genetically distinct groups could be identified among the 5′ NC sequences. Three of these could be further subdivided into two genotypes each (I/1a, II/1b; III/2a, 2c; 4c, 4d) based upon the sequence of the E1 gene. A single isolate from Sardinia, Italy, most closely related to genotype III/2a, was designated genotype 2c. This genotype has not been reported previously. Based upon 5′ NC sequences of HCV isolates, our genotype V/3a appears to be similar to that described by Chan *et al.* (15) as genotype 3. We identified genotypes 4a–4c only in samples from Zaire; a closely related genotype, 4d, was a single isolate from Denmark. Based upon their 5′ NC sequence, some of these isolates resemble isolates from Egypt designated as genotype 4 by Simmonds *et al.* (29).

We are provisionally assigning genotype designations of 5 and 6 to additional genotypes not previously described. Genotypes designated 5a were all from South African specimens and formed a unique, relatively closely related group of isolates. A single isolate, designated 6a, was from Hong Kong and contained two separate nucleotide insertions in the 5′ NC region. Two other isolates containing single nucleotide insertions in the 5′ NC region, from Zaire, could not be amplified with primer set *e* and may represent additional new genotypes not previously described.

The plethora of new classification schemes for HCV has led to considerable confusion. The Okamoto system, based upon full-length genome sequences, separated HCV isolates into groups that differed by about the same degree that the structural proteins of different serotypes of other RNA viruses differ from each other (28, 30). In contrast, the classification system of Chan *et al.* (15) was based upon a hierarchical scheme with genotype designations for genetic differences of the order of magnitude of those found between subgenera of other RNA viruses and, within these types, subtypes with genetic differences equivalent to those between serotypes of other RNA viruses. Individual isolates were then grouped within these subtype designations. This classification scheme is based primarily upon limited sequence data from the C and/or NS5 genes of HCV isolates. However, we wish to point out that the classification is not entirely confirmed by analysis of the 5′ NC and E1 regions. Thus, as seen in Fig. 1, subtypes 1a and 1b, 4c and 4d, and 2a and 2c cannot be differentiated on the basis of 5′ NC sequences. However, IV/2b can be differentiated from III/2a and 2c in this region. Similarly, in the E1 gene region, the genetic difference between genotype III/2a or 2c and IV/2b is equivalent to the difference between, for example, I/1a and 4a–4d (Table 1 and Fig. 3). It is also noteworthy that genotype 4a (but not 4b, 4c, and 4d) is most closely related to genotypes I/1a and II/1b in the 5′ NC region (Fig. 1). Thus, a formal classification of HCV genotypes should be postponed until more sequence data are available.

In summary, we have confirmed and extended our previous report of multiple genotypes of HCV and have identified at least 12 such genotypes, based upon an analysis of the sequence of the E1 gene. These were identified among 84 PCR-positive serum samples drawn from a collection of only 114 anti-HCV positive sera. Thus, it is highly likely that a number of additional genotypes will be identified as more samples from diverse geographic regions are examined.

Table 1. Percent nucleotide (nt) and amino acid (aa) sequence identity of the E1 gene among the 12 HCV genotypes

	I/1a	II/1b	III/2a	IV/2b	2c	(V)/3a	4a	4b	4c	4d	5a	6a	nt
	89.9–97.6	72.0–76.2	59.2–63.7	56.1–58.3	60.8–62.8	63.0–66.3	63.9–67.2	64.9–66.8	62.7–64.4	67.7–69.4	62.3–67.2	62.2–63.9	I/1a
aa		88.9–97.9	58.3–62.2	53.8–57.5	60.1–61.5	63.9–67.2	60.9–63.7	63.4–65.8	61.6–65.1	63.0–65.5	62.2–66.5	61.6–63.0	II/1b
I/1a	91.1–98.4		88.0–91.3	69.1–71.0	72.7–73.6	58.0–60.8	61.5–62.7	58.9–60.4	59.7–63.4	58.7–61.3	56.6–60.8	55.0–56.8	III/2a
II/1b	75.5–80.7	90.1–97.9		92.7–95.0	67.5–68.9	56.3–58.3	58.9–60.8	56.4–57.6	57.1–59.9	57.5–59.0	53.5–56.6	53.6–55.2	IV/2b
III/2a	58.3–64.6	52.6–56.8	89.1–92.7		—	57.5–58.2	59.2	58.5	58.0–58.3	58.9	56.9–57.1	57.6	2c
IV/2b	54.2–56.8	51.0–54.2	69.3–72.9	93.8–96.4		93.8–99.1	64.4–65.3	62.7–64.1	60.9–62.5	62.3–63.9	61.8–64.4	58.0–58.9	(V)/3a
2c	56.3–60.4	52.6–55.7	74.5–77.1	67.7–69.8	—		—	74.8	75.5–78.0	74.8	62.8–64.6	62.0	4a
(V)/3a	64.1–68.8	66.7–70.8	54.7–58.9	54.2–56.8	52.1–53.6	94.3–98.4		—	74.0–74.8	72.0	63.9–64.6	62.7	4b
4a	69.3–73.4	64.6–67.2	62.0–63.0	58.9–60.4	58.3	66.1–68.8	—		90.1	77.6–78.6	62.7–64.8	63.0–64.4	4c
4b	66.7–69.3	66.1–70.3	53.6–56.3	52.1–53.1	53.6	62.0–64.6	76.0	—		—	64.4–66.1	64.1	4d
4c	66.1–72.9	64.6–69.3	55.2–61.5	54.2–58.3	54.7–58.3	63.0–65.6	77.1–81.3	79.2–80.2	89.6		90.1–95.7	60.6–63.2	5a
4d	73.4–75.5	66.7–70.3	56.3–58.9	55.2–55.7	54.2	63.5–64.6	78.1	77.6	82.8	—		—	6a
5a	66.1–73.4	64.1–70.3	52.6–57.3	50.5–53.1	54.2–56.3	60.4–64.1	67.2–68.2	65.1–67.2	67.7–71.4	69.3–71.4	92.7–97.4		
6a	64.6–65.6	62.5–65.6	49.0–51.0	49.0–50.5	50.5	57.8–58.9	66.1	62.5	66.1–67.2	66.7	62.0–63.5	—	

Genotype/isolate(s): I/1a/DK7, DK9, DR1, DR4, S14, S18, SW1, and US11; II/1b/D1, D3, DK1, HK3, HK4, HK5, HK8, IND5, IND8, P10, S9, S45, SA10, SW2, T3, T10, and US6; III/2a/T2, T4, T9, and US10; IV/2b/DK8, DK11, SW3, and T8; 2c/S83; (V)/3a/DK12, HK10, S2, S52, and S54; 4a/Z4; 4b/Z1; 4c/Z6 and Z7; 4d/DK13; 5a/SA1, SA4, SA5, SA6, SA7, and SA13; 6a/HK2.

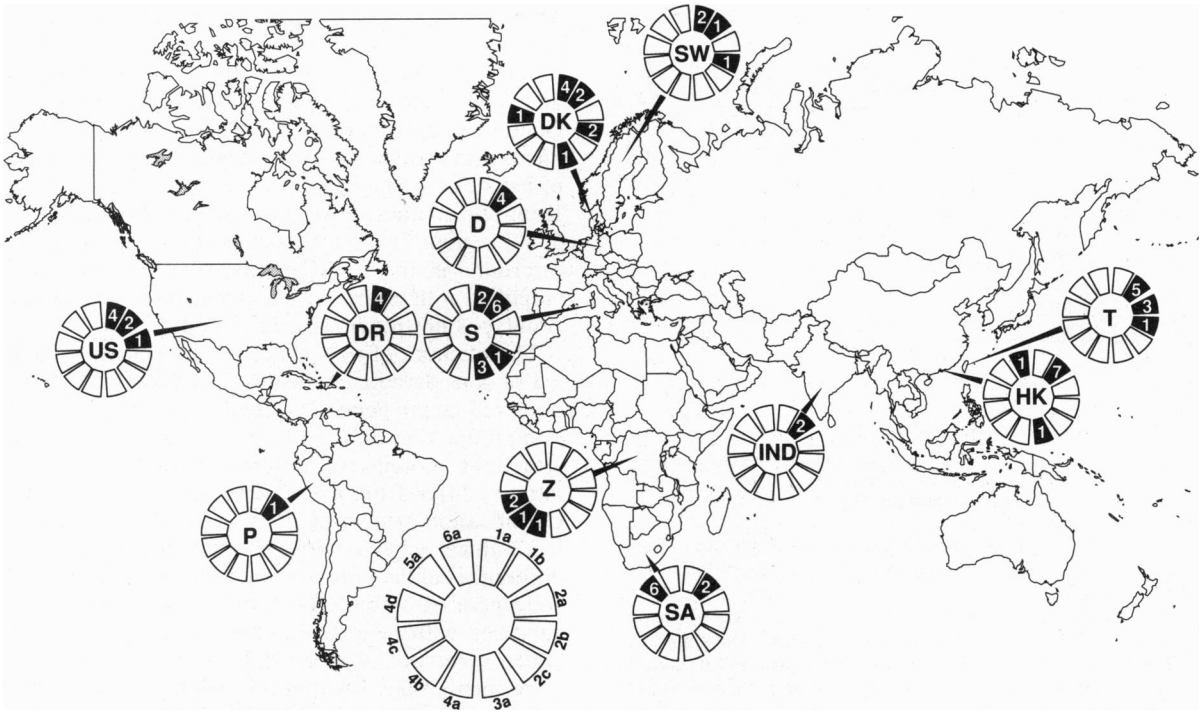

FIG. 4. Distribution of different HCV genotypes in the 12 countries studied. We used the genotype nomenclatures proposed by Chan *et al.* (15) and by Okamoto *et al.* (3), where genotype 1a corresponds to genotype I, 1b to II, 2a to III, and 2b to IV, respectively. The complete E1 gene sequence was determined in 51 of these HCV isolates, including 8 isolates of genotype I/1a, 17 isolates of genotype II/1b, and 26 isolates comprising genotypes III/2a, IV/2b, 2c, 3a, 4a–4d, 5a, and 6a. In the remaining 23 isolates, all of genotypes I/1a and II/1b, the genotype assignment was based on a partial E1 gene sequence since they did not represent additional genotypes in any of the 12 countries. The number of isolates of a particular genotype is given in each of the 12 countries studied. Although we successfully determined the genotype of 74 of 84 HCV isolates studied, we could not amplify the E1 gene in 10 isolates: one isolate from each of five countries (D, DR, IND, P, and US); 2 isolates from Zaire (Z); and 3 isolates from South Africa (SA). National borders depicted represent those existing at the time of sampling.

We thank Ms. T. Tsareva for synthesis of oligonucleotides. We gratefully acknowledge the generosity of the following investigators for providing sera for this study: Drs. H. Alter, V. Arankalle, D.-S. Chen, P. Farci, M. C. Kew, K. Krogsgaard, A. Lok, T. Quinn, F. Renger, M. Sjogren, and A. Widell. Computer-assisted nucleic acid analysis was through the GenBank Online Service and PC/GENE.

1. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
2. Houghton, M., Weiner, A., Han, J., Kuo, G. & Choo, Q.-L. (1991) *Hepatology* **14**, 381–388.
3. Okamoto, H., Kurai, K., Okada, S.-I., Yamamoto, K., Lizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. & Mishiro, S. (1992) *Virology* **188**, 331–341.
4. Inchauspe, G., Zebedee, S., Lee, D.-H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10292–10296.
5. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. & Okayama, H. (1991) *J. Virol.* **65**, 1105–1113.
6. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
7. Okamoto, H., Kojima, M., Okada, S.-I., Yoshizawa, H., Iizuka, H., Tanaka, T., Muchmore, E. E., Peterson, D. A., Ito, Y. & Mishiro, S. (1992) *Virology* **190**, 894–899.
8. Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. & Mayumi, M. (1991) *J. Gen. Virol.* **72**, 2697–2704.
9. Miller, R. H. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
10. Bukh, J., Purcell, R. H. & Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 187–191.
11. Bukh, J., Purcell, R. H. & Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4942–4946.
12. Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3392–3396.
13. Bachmann, B., Luke, W. & Hunsmann, G. (1990) *Nucleic Acids Res.* **18**, 1309.
14. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237–244.
15. Chan, S.-W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. F., Follett, E., Yap, P. L. & Simmonds, P. (1992) *J. Gen. Virol.* **73**, 1131–1141.
16. Mori, S., Kato, N., Yagyu, A., Tanaka, T., Ikeda, Y., Petchclai, B., Chiewsilp, P., Kurimura, T. & Shimotohno, K. (1992) *Biochem. Biophys. Res. Commun.* **183**, 334–342.
17. Lee, C.-H., Cheng, C., Wang, J. & Lumeng, L. (1992) *J. Clin. Microbiol.* **30**, 1602–1604.
18. Cha, T.-A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G. & Urdea, M. S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7144–7148.
19. Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.-L. & Houghton, M. (1991) *Virology* **180**, 842–848.
20. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. & Shimotohno, K. (1991) *Biochem. Biophys. Res. Commun.* **175**, 220–228.
21. Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1990) *Jpn. J. Exp. Med.* **60**, 167–177.
22. Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *J. Gen. Virol.* **71**, 3027–3033.
23. Chen, P.-J., Lin, M.-H., Tai, K.-F., Liu, P.-C., Lin, C.-J. & Chen, D.-S. (1992) *Virology* **188**, 102–113.
24. Liu, K., Hu, Z., Li, H., Prince, A. M. & Inchauspe, G. (1992) *Gene* **114**, 245–250.
25. Tanaka, T., Kato, N., Nakagawa, M., Ootsuyama, Y., Cho, M.-J., Nakazawa, T., Hijikata, M., Ishimura, Y. & Shimotohno, K. (1992) *Virus Res.* **23**, 39–53.
26. Abe, K., Inchauspe, G. & Fujisawa, K. (1992) *J. Gen. Virol.* **73**, 2725–2729.
27. Honda, M., Kaneko, S., Unoura, M., Kobayashi, K. & Murakami, S. (1993) *Arch. Virol.* **128**, 163–169.
28. Hahn, Y. S., Galler, R., Hunkapiller, T., Dalrymple, J. M., Strauss, J. H. & Strauss, E. G. (1988) *Virology* **162**, 167–180.
29. Simmonds, P., McOmish, F., Yap, P. L., Chan, S.-W., Lin, C. K., Dusheiko, G., Saeed, A. A. & Holmes, E. C. (1993) *J. Gen. Virol.* **74**, 661–668.
30. Palmenberg, A. C. (1989) in *Molecular Aspects of Picornavirus Infection and Detection*, eds. Semler, B. L. & Ehrenfeld, E. (Am. Soc. Microbiol., Washington, DC), pp. 211–241.