Arch Virol 143/12 (1998)

VDN Virology Division News

Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization

B. Robertson[1], **G. Myers**[2], **C. Howard**[3], **T. Brettin**[2], **J. Bukh**[4], **B. Gaschen**[2], **T. Gojobori**[5], **G. Maertens**[6], **M. Mizokami**[7], **O. Nainan**[1], **S. Netesov**[8], **K. Nishioka**[9], **T. Shin-i**[5], **P. Simmonds**[10], **D. Smith**[10], **L. Stuyver**[6], and **A. Weiner**[11]

[1] Hepatitis Branch, Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.
[2] Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A.
[3] Royal Veterinary College, Royal College Street, London, U.K.
[4] Hepatitis Viruses Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, U.S.A.
[5] Center of Informatics Biology, National Institute of Genetics, Mishima, Japan
[6] Innogenetics, Hepatitis Program, Gent, Belgium
[7] Second Department of Medicine, Nagoya City University Medical School, Nagoya, Japan
[8] State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk, Russia
[9] Viral Hepatitis Research Foundation of Japan, Tokyo, Japan
[10] Department of Medical Microbiology, University of Edinburgh, Edinburgh, Scotland, U.K.
[11] Chiron Corporation, Emeryville, California, U.S.A.

Summary. This paper presents a summary of the recommendations that were formulated for the purposes of unifying the nomenclature for hepatitis C virus (HCV), based upon guidelines of the International Committee on Virus Taxonomy (ICTV), and provides guidelines for the incorporation of sequence data into an HCV database that will be available to researchers through the internet. Based upon the available data, the genus *Hepacivirus* should be regarded as comprising a single species with HCV-1 as the prototype. All currently known isolates of HCV can be divided into six phylogenetically distinct groups, and we recommend that these groups are described as clades 1 to 6. Whether or not these should be regarded as different species within the *Hepacivirus* genus requires additional clinical, virological, and immunological information. Clades 1, 2, 4, and 5 would correspond to genotype 1, 2, 4, and 5 while clade 3 would comprise genotype 3 and genotype 10, and clade 6 comprise genotypes 6, 7, 8, 9, and 11. We propose that existing subtype designations are reassigned within these clades based upon publication priority, the existence of a complete genome sequence and prevalence. The assignment of isolates to new clades and subtypes should be confined to isolates characterized from epidemiologically unlinked individuals. Comparisons should be based on nucleotide sequences of at least two coding regions and preferably of complete genome sequences, and should be based on phylogenetic analysis rather than percent identity. A forum for discussion and contributions to these recommendations will be made available at the international HCV database at http://s2as02.genes.nig.ac.jp.

Background

The long-awaited discovery of hepatitis C virus (HCV), the virus responsible for over 90% of non-A, non-B hepatitis, by Chiron Corporation in 1989 [11] and the simultaneous development of the technology of polymerase chain reaction led to a surge of sequence information for this virus (reviewed in [5, 27, 29, 47]). The first indication of the genetic diversity of HCV came from investigators in Japan [12, 19] and subsequent studies have identified at least six major genetic groups [1–4, 7–10, 12, 20, 22, 36, 38–40, 42, 44, 45, 48, 53]. Due to the rapid growth of sequence information, nomenclature describing the genetic variability has been inconsistent, resulting in considerable confusion for researchers and students alike who enter the field. In 1993, two research groups defined the existence of 6 major genotypes of HCV [3, 44]. Using regions within the core, E1, and NS5B, the proportion of identical nucleotides was proposed to distinguish types, subtypes, and isolates [44, 53]. However, the wide range of HCV genetic variability that has been discovered upon sequencing of subsequently identified strains revealed the existence of isolates that did not fit within these defined ranges [18, 28, 55–57, 58]. In light of these problems, a small but internationally representative group of HCV researchers gathered in Santa Fe, New Mexico, USA, in late August 1997 to generate a consensus regarding the taxonomic position of HCV and the nomenclature of HCV isolates.

Framework for discussions

The initial discussion focused upon the guidelines for virus nomenclature and classification provided by the International Committee on Taxonomy of Viruses (ICTV). The ICTV categorizes viruses as orders (-virales), families (-ae), subfamilies (virinae), genera (-virus) and species. A species is recognized as a "polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche". As such, more than one of the following properties have to be present to differentiate individual species: 1) sequence differences; 2) host range; 3) cell and tissue tropism; 4) pathogenicity or cytopathology; 5) physical properties; or 6) antigenic properties. In the case of a virus that cannot be propagated in cell culture (such as HCV), the complete genome sequence must be known and the virus must have further distinguishing characteristics such as virus neutralization [34]. At present, the ICTV recognizes the *Hepacivirus* genus as containing HCV [59]; the question of whether or not the 6 to 11 genotypes that have been reported should be categorized as individual species remains to be determined.

Sequence approaches for classification

Genetic classification can be broadly approached from two perspectives generally identified as either distance-based or character-based [25]. A distance-based approach is based upon numerical values describing the similarity, irrespective of the evolutionary relatedness, with the values usually expressed as measures of distance. These numerical values can be graphically depicted as a frequency distribution or as a tree-like structure; the generation of a "tree" from these data does not necessarily imply an evolutionary correspondence, especially when relatively short sequences are under investigation. Unweighted pair group method analysis (UPGMA) and neighbor-joining (NJ) are commonly used distance-based

methods used to generate a tree; both methods generate a single tree, which can be a starting point for evolutionary analysis.

The other general approach is character-based and evaluates the phylogenetic relatedness of sequences based upon a subset of positions called "informative sites". Parsimony methods (weighted and unweighted) and maximum likelihood are two frequently used approaches. Typically multiple trees (cladograms) are generated and evaluated for accuracy, and so these methods are computationally more demanding.

Bootstrapping is a test that can be applied to trees generated by either character-based or distance-based methods in order to estimate the frequency with which a particular branch (node) occurs. Sequences are repeatedly reshuffled and a tree analysis under the appropriate model (N-J, parsimony, etc.) is performed. The percentage of bootstrap replicates in which a particular set of sequences cluster together gives an indication of the strength of the grouping. Values of greater than 70% out of 1000 replicates are considered as supporting the grouping [16].

Methods used for genetic analysis of HCV

Three coding regions, core, E1, and NS5B were identified initially as being useful for identifying genotypes by distance-based methods [3, 4, 36, 44, 45, 53]. Sequence distances (dissimilarities) have been calculated in several HCV laboratories using the PHYLIP program DNADIST [13, 14] with the Kimura 2-parameter metric for transitions and transversions. In other laboratories, 1-parameter (simple percent dissimilarities) and 6-parameter [15] measures have been employed. Plotting the values obtained from any of these approaches as a frequency distribution (the number of times each pair-wise value is observed in a large data set) usually results in three peaks, each with the characteristics of a random distribution. The troughs between these peaks have been used to demarcate types, subtypes, and isolates [30, 31, 36, 44, 53]. Although these criteria were satisfactory for distinguishing between isolates from Europe, Japan, the Middle East and North America, analysis of strains from Southeast Asia and Africa produced ambiguous results with many distances falling in the troughs. Whether a sequence was considered a type, subtype or isolate might depend upon which particular sequences it was compared with [28, 48]. In contrast, when distances between sequences are used to construct neighbor joining trees, six different clades are observed and are supported by bootstrap analysis [3, 4, 28, 31, 36, 48, 52, 53]. Similar conclusions arise from analysis of complete virus genomes by the character based methods of maximum likelihood [28, 52] or parsimony (Fig. 1). A virus previously described as genotype 10 clusters on the same branch as viruses within genotype 3, together forming clade 3, while a virus described as genotype 11 clusters with a genotype 6a virus, together forming clade 6. This concordance of distance-based and character-based methods is the basis of our proposal that HCV variants be grouped into six clades.

The GB viruses

The issues which relate to the taxonomic status of the GB and HGV viruses, GBV-A [32, 49, 50], GBV-B [32, 49, 50], and GBV-C/HGV [23, 26, 37, 49], are also whether or not these viruses are monophyletic or polyphyletic in relationship to HCV and also to each other. Figure 2 is a reconstruction of the relationships based upon inversely weighted

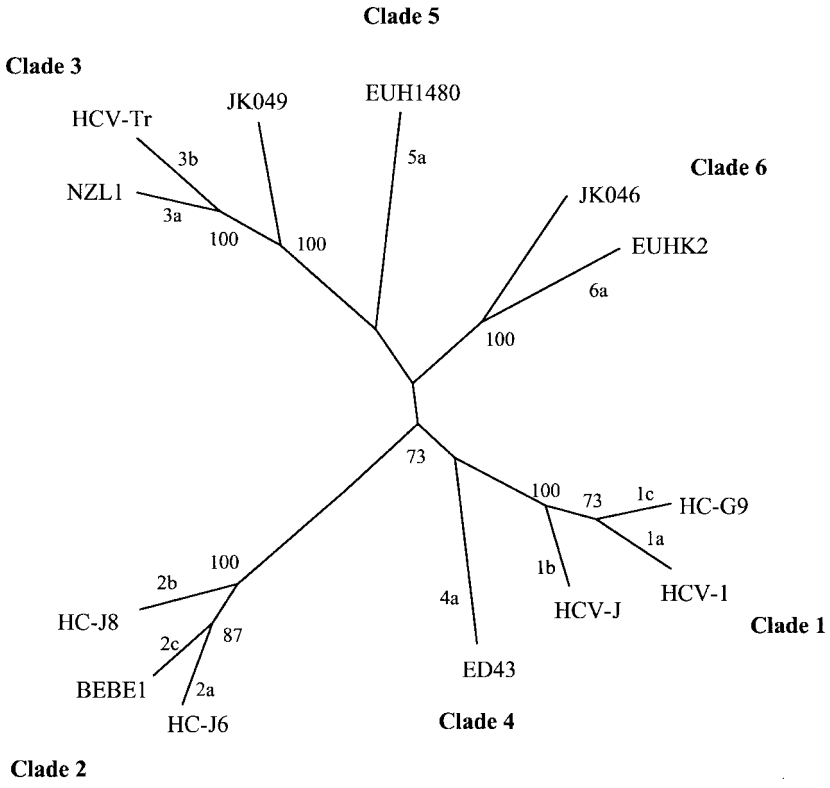

Fig. 1. Parsimony analysis of representative HCV sequences. Inversely weighted parsimony was performed using PAUP 3.1.1 as previously described [17, 21]. To minimize noise, only homologous second base positions from complete genomic sequences were studied: of the 3183 total sites that were aligned, 1216 varied sites were analyzed. A virtually identical topology was obtained using maximum likelihood analysis at all nucleotide positions [52]. Bootstrap analysis was subsequently performed (DNABOOT) and values of 70% or more are indicated at the appropriate nodes; in general, bootstrap values of 70% may correspond to confidence values of 95% [16]. Accession numbers (not identified in the figure) are: HC-G9 (D14853), HCV-1 (M62321), HCV-J (D90208), ED43(Y11604), HC-J6 (D00944), HC-J8 (D10988), NZL1(D17763), HCV-Tr (D49347), JK049 (D63821), EUH1480 (Y13184), JK046 (D63822), and EUHK2 (Y12083)

parsimony using only second base positions within the NS3 helicase region. The inclusion of other flavivirus sequences in this analysis is important as it tends to argue against a monophyletic relationship.

The GBV-A and GBV-C/HGV viruses are closely related to each other phylogenetically, have similar genome structures (with a short or absent potential coding sequence for a homologue of the core gene) and are non-hepatotrophic. Based upon such criteria, an argument could be made for these viruses being considered as members of a new genus of non-hepatotropic *Flaviviridae*. The known sequences of the GBV-C/HGV viruses can be grouped into three geographic variants based upon sequence of the whole genome or sequence of the 5'UTR [33, 41, 51]. However, other subgenomic regions smaller than one-quarter of the complete genome, with the exception of the first 1200 5' terminal nucleotides or a fragment spanning nucleotides 3600–4800, do not reflect the pattern seen within the whole genome [51]. Given these observations, and the recent findings by Takahashi et al. [54], investigators are cautioned against assigning names until these

relationships are clarified. The GBV-A viruses have sequences that reflect the species from which they were derived [6, 24] although experimental studies have demonstrated that cross-species transmission can occur [6, 24].

The GBV-B virus causes hepatitis in New World monkeys [6, 43] and is more closely related to HCV than to GBV-A and GBV-C (Fig. 2); furthermore, it has a genome structure that resembles HCV, specifically with the presence of a reading frame encoding a core polyprotein. Based upon these characteristics, it may be considered as a member of the genus *Hepacivirus* or alternatively as a member distinct from, but sharing some similarities with the hepaciviruses.

Summary of recommendations

Hepacivirus *genus*

At this time, the genus *Hepacivirus* should be regarded as consisting of a single species, with HCV-1 as the prototype of the species. The future designation of species corresponding to the 6 clades would be appropriate if additional information such as cross-neutralization or cross-protection studies clearly distinguish these viruses. In the absence of such information, however, we conclude that it is difficult to argue for more than a single species.

Type nomenclature

To avoid confusion with any previous terminology and to reflect the inclusive nature of the groupings, we recommend that the term clade be used to identify the 6 genetic groups. Thus, genotypes 1, 2, 4, and 5 would be identified as clades 1, 2, 4, and 5, respectively; genotypes 3 and 10 would become members of clade 3, while genotypes 6, 7, 8, 9, and 11 would be members of clade 6 (Fig. 1).

Subtype nomenclature

The designation of subtype names may be important for the study of clinical differences within different clades of HCV, for the study of new therapies and vaccines, and for epidemiological investigations. A proposal for renaming currently identified subtypes based upon publication priority, number of isolates, and the presence of a complete genome sequence is being developed (Maertens et al., manuscript in prep.). In the future, we recommend that at least two epidemiologically unlinked isolates be identified which cluster together phylogenetically and are clearly distinguished from other subtypes.

Approach to classify new sequences

In the interest of defining a common approach to classification of clades, subtypes, and isolates the following steps (illustrated in Fig. 3) are recommended : 1) Sequence the C, E1, or NS5B region. 2) Align sequences with representatives of each of the six clades. 3) Determine % similarity; the highest % similarity gives a tentative clade assignment. 4) Calculate DNADIST (Kimura 2 parameter option) using PHYLIP followed by the generation of a neighbor-joining tree using representatives of all known clades; bootstrap to confirm. These steps will result in identifying the clade assignment, or if the sequence appears to be a new clade, indicate that additional (complete genome) sequence information is needed.

To identify a subtype, steps 2–4 should be repeated with representatives of subtypes within the clade. Subtype identification can result in three possible outcomes: 1) The isolate is identified as belonging to an existing subtype. 2) The sequence appears to be a new subtype. However, prior to assigning a new subtype name, the sequence of at least one additional sample which is epidemiologically unrelated is needed. Alternatively, a full length sequence of the proposed new subtype would provide adequate data to support the designation of a new subtype. 3) The analyses are ambiguous. The sequences can be analysed by other phylogenetic methods or sent to one of the HCV research labs capable of performing in-depth analyses. As has been proposed previously [46], the sequence of two sub-genomic regions should be obtained and compared to verify the phylogenetic relationship of new HCV subtypes.

Database policies and development

For clarity in communication, investigators should note that subtypes are referred to with small letters (e.g., "5a"), whereas the protein components of HCV and HGV/GBV-C are characterized by capital letters (e.g., "NS5A"). It was suggested that, for consistency,

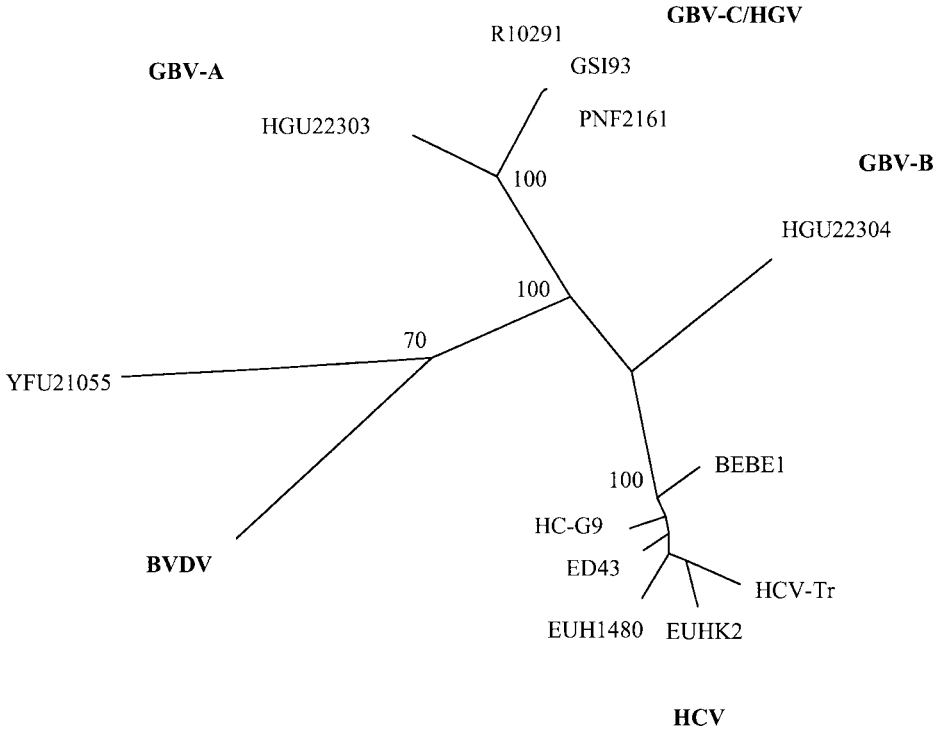

Fig. 2. Parsimony analysis illustrating the relationship of representative hepaciviruses, BVDV (pestivirus), yellow fever virus (flavivirus), GBV-A, GBV-B, and GBV-C. Inversely weighted parsimony was performed using PAUP 3.1.1 as previously described [17, 21]. To minimize noise, only homologous second base positions from NS3 coding sequences were studied: of the 601 total alignable sites, 526 varied sites were analyzed. Bootstrap analysis was subsequently performed (PAUP 3.1.1) and values of 70% or more are indicated at the appropriate nodes; in general, bootstrap values of 70% may correspond to confidence values of 95% [16]. Accession numbers (not identified in the figure) are: HCV-Tr (D49347), EUHK2 (Y12083), EUH1480 (Y13184), ED43 (Y11604), HC-J9 (D14853), BVDV (M31182),YFU21055 (U21055), HGU22303 (U22303), HGU44402 (U44402), R10291(U45966), GSI93 (D87263), and PNF2161 (U22304)

nucleotide numbering should initiate at the AUG codon, with 5' terminal nucleotides given negative values numbered from <–1 to –341>. The term 5'UTR should be used to refer to the 5' region of the RNA that is untranslated. To avoid confusion and the possibility that identical names are given to individual samples, it was suggested that samples be identified by 1) the ISO two character country name [35], 2) identification of the laboratory in which the isolation and sequencing was performed, and 3) a sample number, i.e., USCDC10 (meaning United States, Centers for Disease Control, sample number 10). No patient initials should be used for virus identification, and the identifier should be kept to under 10 characters in total. In looking to the future, other useful information which could usefully be incorporated into a database format would include information such as age, sex, year of isolation, year of infection, etiology (transfusion, IVDU, sexual, unknown), disease stage (asymptomatic, CAH, LC, HCC) and ethnic origin.

Currently there is a database of all available HCV sequences (downloaded regularly from DDBJ [DNA DataBase of Japan]) available through the world wide web at http://s2as02.genes.nig.ac.jp. The master database contains all HCV sequences with an alignment based upon the reference strain HCV HC-J4 (D10750). A reproduction of the database home page containing this graphic alignment is shown in Fig. 4; individual sequences are illustrated by lines under a schematic of the HCV genome indicating the various gene regions and products (adapted from [5]). Genome regions can be selected from this graphic map of the HCV genome and all sequences from that region can then be viewed as fully annotated entries – nucleic acid or amino acid sequences – and compared to each other by tree analysis. In a separate section of this database, there is FASTA similarity search capability.

The purpose of this communication is to solicit feedback on these proposed guidelines, and to generate ideas for future categories that would be useful for database development. A forum for discussion and contributions to these recommendations will be made available at the international HCV database at http://s2as02.genes.nig.ac.jp. One of the services that this database could provide is a source of curatorial information for researchers interested in evaluating sequence information from a wide range of perspectives – molecular, epidemio-

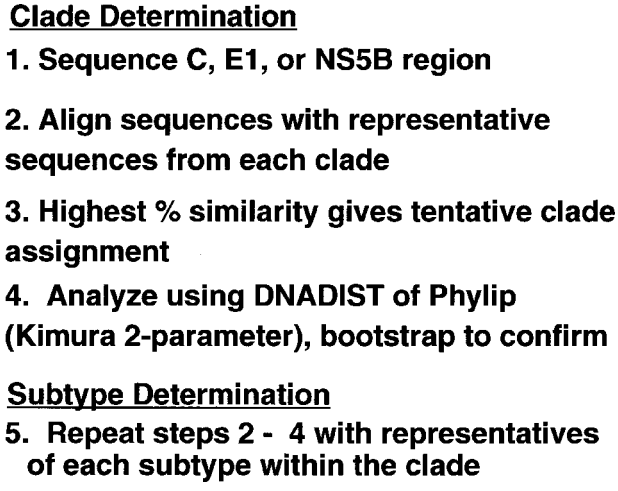

Fig. 3. Flow chart illustrating the steps recommended to identify HCV clades and subtypes

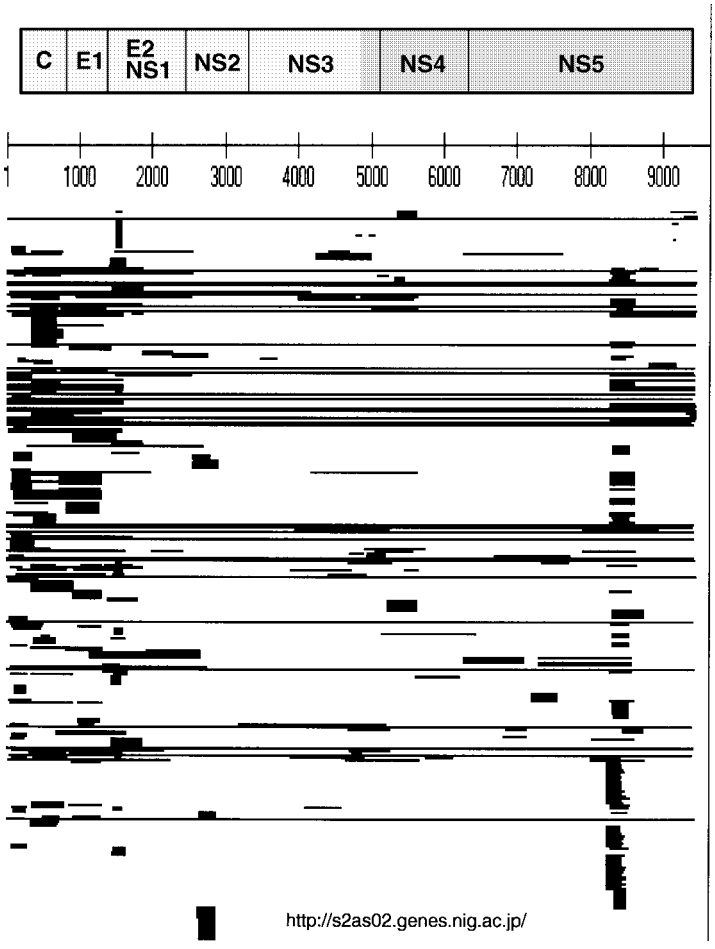

Fig. 4. Graphic of the HCV genome from the HCV Database (http://s2as02.genes.nig.ac.jp) illustrating alignment of complete and partial genome sequences. Each individual line represents an HCV sequence whose location has been aligned relative to the complete genome. The potential cleavage sites resulting in NS4A, NS4B, NS5A, and NS5B are not shown in this graphic

logic, and clinical. The types of information that are useful to these different groups may vary, and the ultimate usefulness will depend upon the inclusion of the maximum amount of appropriate information from each sample.

References

1. Adams NJ, Chamberlain RW, Taylor LA, Davidson F, Lin CK, Elliott RM, Simmonds P (1997) Complete coding sequence of hepatitis C virus genotype 6a. Biochem Biophys Res Commun 234: 393–396
2. Bukh J, Purcell RH, Miller RH (1992) Sequence analysis of the 5' noncoding region of hepatitis C virus. Proc Natl Acad Sci USA 89: 4942–4946
3. Bukh J, Purcell RH, Miller RH (1993) At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc Natl Acad Sci USA 90: 8234–8238
4. Bukh J, Purcell RH, Miller RH (1994) Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc Natl Acad Sci USA 91: 8239–8243
5. Bukh J, Miller RH, Purcell RH (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liver Dis 15: 41–63

6. Bukh J, Apgar CL (1997) Five new or recently discovered (GBV-A) virus species are indigenous to New World Monkeys and may constitute a separate genus of *Flaviviridae*. Virology 229: 429–436
7. Cha T-A, Beall E, Irvine B, Kolberg J, Chien D, Kuo G, Urdea MS (1992) At least five related, but distinct, hepatitis C viral genotypes exist. Proc Natl Acad Sci USA 89: 77144–7148
8. Chamberlain RW, Adams N, Saeed AA, Simmonds P, Elliott RM (1997) Complete nucleotide sequence of a type 4 hepatitis C virus vairant, the predominant gnotype in the Middle East. J Gen Virol 78: 1341–1340
9. Chamberlain RW, Adams NJ, Taylor LA, Simmonds P, Elliott RM (1997) The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. Biochem Biophys Res Commun 236: 44–49
10. Chan SW, McOmish F, Holmes EC, Dow B, Peutherer JF, Follett E, Yap PL, Simmonds P (1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. J Gen Virol 73: 1131–1141
11. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a c DNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244: 359–362
12. Enomoto N, Takada A, Nakao T, Date T (1990) There are two major types of hepatitis C in Japan. Biochem Biophys Res Commun 170: 1021–1025
13. Felsenstien J (1991) PHYLIP Manual Version 3.4. University of California, Berkeley
14. Felsenstein, J (1993) PHYLIP, Phylogeny Inference Package, Version 3.5c. University of Washington, Seattle
15. Gojobori T, Ishii K, Nei M (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J Mol Evol 18: 414–423
16. Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method of assessing confidence in phylogenetic analysis. Syst Biol 42: 182–192
17. Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. Science 264: 671-677
18. Hotta H, Handajani R, Luida MI, Soemarto W, Doi H, Miyajima H, Homma M (1994) Subtype analysis of hepatitis C virus in Indonesia on the basis of NS5b region sequences. J Clin Microbiol 32: 3049–3051
19. Kato N, Ohkoshi S, Shimotohno K (1989) Japanese isolates of the non-A, non-B hepatitis viral genome show sequences variations from the original isolate in the USA. Proc Jpn Acad 665: 219–223
20. Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-a, non-B hepatitis. Proc Natl Acad Sci USA 87: 9524–9528
21. Korber BTM, Smith RF, MacInnes K, Myers G (1994) Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. J Virol 68: 6730–6744
22. Kubo Y, Takeuchi K, Boonmar S, Katayama T, Choo QL, Kuo G, Weiner AJ, Bradley DW, Houghton M, Saito I, Miyamura T (1989) A cDNA fragment of hepatitis C virus isolated from an implicated donor of post-transfusion non-A, non-B hepatitis in Japan. Nucleic Acids Res 17: 10367–10372
23. Leary TP, Muerhoff AS, Simons JN, Pilot-Matias TJ Erker JC, Chalmers MS, Schaluder GG, Dawson GJ, Desai SM, Mushahwar IK (1996) Sequence and genomic organization of GBV-C: a novel member of the Flaviviridae associated with human non-A-E hepatitis. J Med Virol 48: 60–67
24. Leary TP, Desai SM, Yamaguchi J Chalmers ML, Schlauder GG, Dawson GJ, Mushahwar IK (1996) Species-specific variants of GB virus A in captive monkeys. J Virol 70: 9028–9030
25. Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer Publishers, Sunderland
26. Linnen J, Wages J, Zhang-Keck Z-Y, Fry KE, Krawczynski KZ, Alter H, Koonin E, Gallagher M, Alter M, Hadziyannis S, Karayiannis P, Fung K, Nakatsuji Y, Shih JW, Young L, Piatak M Jr, Hoover C, Fernandes J, Chen S, Zou JC, Morris T, Hyams KC, Ismay S, Lifson JD, Kim JP (1996) Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Sicence 271: 505–508
27. Maertens G, Stuyver L (1997) Genotypes and genetic variation of hepatitis C virus. In: Harrison TJ, Zuckerman AJ (eds) The molecular medicine of viral hepatitis. Wiley, New York, pp 183–233
28. Mellor J, Holmes EC, Jarvis LM, Yap PL, Simmonds P (1995) Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. J Gen Virol 76: 2493–2507
29. Miyakawa Y, Okamoto H, Mayumi M (1995) Classifying hepatitis C virus genotypes. Mol Med Today 1: 20–25
30. Mizokami M, Gojobori T, Lau JYN (1994) Molecular evolutionary virology: its application to hepatitis C virus. Gastroenterology 107: 1181–1182

31. Mizokami M, Gojobori T, Ohba K, Ikeo K, Orito E, Lau JYN (1996) Hepatitis C virus type 7, 8, and 9 should be classified as type 6 subtypes. J Hepatol 24: 622–624
32. Muerhoff AS, Leary TP, Simons JN, Pilot-Matias TJ, Dawson GJ, Erker JC, Chalmers ML, Schlauder GG, Desai SM, Mushahwar IK (1995) Genomic organization of GB viruses A and B: two members of the Flaviviridae associated with GB agent hepatitis. J Virol 69: 5621–5630
33. Muerhoff AS, Simonds JN, Leary TP, Erker JC, Chalmers ML, Pilot-Matias TJ, Dawson GJ, Desai SM, Mushahwar IK (1996) Sequence heterogeneity within the 5'-terminal region of the hepatitis GB virus C genome and evidence for genotypes. Hepatology 25: 379–384
34. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (1995) Introduction to the universal system of virus taxonomy. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York, pp 1–13 (Arch Virol [Suppl] 10)
35. Myers G, Korber B, Foley B, Jeang K-T, Mellors JW, Wain-Hobson S (eds) (1996) Human retroviruses and AIDS 1996, III-108-III-112. Theoretical Division, Los Alamos National Laboratory, Los Alamos
36. Ohba K, Mizokami M, Ohno T, Suzuki K, Orito E, Ina Y, Lau JYN, Gojobori T (1995) Classification of hepatitis C virus into major types and subtypes based on molecular phylogenetic tree analysis. Virus Res 36: 201–214
37. Ohba K, Mizokami M, Lau JY, Orito E, Ikeo K, Gojobori T (1996) Evolutionary relationship of hepatitis C, pesti-, flavi-, plantviruses, and newly discovered GB hepatitis agents. FEBS Lett 378: 232–234
38. Okamoto H, Okada, S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M (1991) Nucleotide sequence of the gnomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J Gen Virol 72: 2697–2704
39. Okamoto H, Kurai K, Okada S, Yamamoto K, Lizuka H, Tanaka T, Fukuda S, Tsuda F, Mishiro S (1992) Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 188: 331–341
40. Okamoto H, Kojima M, Sakamoto M, Iizuka H, Hadiwandowo S, Suwignyo S, Miyakawa Y, Mayumi M (1994) The entire nucleotide sequence and classification of a hepatitis C virus iolates of a novel genotype from an Indonesian patient with chronic liver disease. J Gen Virol 75: 629–635
41. Okamoto H, Nakao H, Inoue T, Fukuda M, Kishimoto J, Iizuka H, Tsuda F, Miyakawa Y, Mayumi MJ (1997) The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. J Gen Virol 78: 737–745
42. Sakamoto M, Akahane Y, Tsuda F, Tanaka T, Woodfield DG, Okamoto H (1994) Entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a. J Gen Virol 75: 1761–1768.
43. Schlauder GG, Dawson GJ, Simons JN, Pilot-Matias TJ, Gutierrez RA, Heynen CA, Knigge MF, Kurpiewski GS, Buijk SL, Leary TP, Muerhoff AS, Desai SM, Mushahwar IK (1995) Molecular and serologic analysis in the transmission of the GB hepatitis agents. J Med Virol 46: 81–90
44. Simmonds P, Holmes EC, Cha T-A, Chan SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol 74: 2391–2399
45. Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, Urdea MS, Holmes EC (1994) Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1, and NS-5 regions. J Gen Virol 75: 1053–1061
46. Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, Brouwer JT, Chan S-W, Chayama K, Chen D-S, Choo Q-L, Colombo M, Cuypers HTM, Date T, Dusheiko GM, Esteban JI, Fay O, Hadziyannis SJ, Han H, Hatzakis A, Holmes EC, Hotta H, Houghton M, Irvine B, Khoara M, Kolberg JA, Kuo G, Lau JYN, Lelie PN, Maertens G, McOmish F, Miyamura T, Mizokami M, Nomoto A, Prince AM, Reesnik HW, Rice C, Roggendorf, M, Schalm SW, Shikata T, Shimotohno K, Stuyver L, Trepo C, Weiner A, Yap PL, Urdea MS (1994) A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology 19: 1321–1324
47. Simmonds P (1995) Variability of hepatitis C virus. Hepatology 21: 570–583
48. Simmonds P, Mellor J, Sakuldamrongpanich T, Nuchaprayoon C, Tanprasert S, Holmes EC, Smith DB (1996) Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. J Gen Virol 77: 3013–3024
49. Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar IK (1995) Isolation of novel virus-like sequences associated with human hepatitis. Nat Med 1: 564–569

50. Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML, Van Sant CL, Mushahwar IK (1995) Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc Natl Acad Sci USA 92: 3401–3405
51. Smith DB, Cuceanu N, Davidson F, Jarvis LM, Mikili JL, Hamid S, Ludlam CA, Simmonds P (1997) Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5' non-coding region. J Gen Virol 78: 1533–1542
52. Smith DB, Simmonds P (1988) Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. J Mol Evol (in press)
53. Stuyver L, Arnhem WV, Wyseur A, Hernandez F, Delaporte E, Maertens G (1994) Classification of hepatitis C viruses based on phylogenetic analysis of the envelope E1 and nonstructural 5B regions and identification of five additional subtypes. Proc Natl Acad Sci USA 91: 10134–10138
54. Takahashi K, Hijikata M, Hino K, Mishiro S (1997) Entire polyprotein-ORF sequences of Japanese GBV-C/HGV isolates: implications for new genotypes. Hepatology Res 8: 139–148
55. Tokita H, Shrestha SM, Okamoto H, Sakamoto M, Horikita M, Iizuka H, Shrestha S, Miyakawa Y, Mayumi M (1994a) Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. J Gen Virol 75: 931–936
56. Tokita H, Okamoto H, Tsuda F, Song P, Nakata S, Chosa T, Iizuka H, Mishiro S, Miyakawa Y, Mayumi M (1994b) Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth and ninth major genetic groups. Proc Natl Acad Sci USA 91: 11022–11026
57. Tokita H, Okamoto H, Luengrojanakul P, Vareesangthip K, Chainuvati T, Iizuka H, Tsuda F, Miyakawa Y, Mayumi M (1995) Hepatitis C virus variants from Thailand classifiable into five novel genotypes in the sixth (6b), seventh (7c, 7d) and ninth (9b, 9c) major genetic groups. J Gen Virol 76: 2329–2335
58. Tokita H, Okamoto H, Iizuka H, Kishimoto J, Tsuda F, Lesmana LA, 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. J Gen Virol 77: 293–301
59. Wengler G, Bradley DW, Collett MS, Heinz FX, Schlesinger RW, Strass JH. (1995) Flaviviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York, pp 415–427 (Arch Virol [Suppl] 10)

Authors' address: Dr. B. Robertson, Hepatitis Branch, Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, U.S.A.