Classification of hepatitis C virus variants in six major types based on analysis of the envelope 1 and nonstructural 5B genome regions and complete polyprotein sequences

Xavier de Lamballerie, Remi N. Charrel, Houssam Attoui and Philippe De Micco

Laboratoire de Virologie, Faculté de Médecine, Hôpital de la Timone, Boulevard Jean Moulin 13385, Marseille cedex 05, France

The phylogenetic status of recently described isolates of hepatitis C virus (HCV) from Vietnam, Thailand and Indonesia (previously classified as types 7, 8, 9, 10 and 11) was re-analysed by the neighbour-joining method instead of the unweighted pair-group method with arithmetic mean (UPGMA) that was first used by the discoverers of these strains. The analysis of complete amino acid sequences and of nucleotide sequences of the envelope 1 (672 nt) and nonstructural 5B (1092 nt) genomic regions permitted the re-assignment of the type 7, 8, 9 and 11 isolates to type 6, and that of type 10 strains to type 3. Finally, this study made possible the classification of the previously described HCV strains (including these South-East Asian isolates) in six major types and at least 30 subtypes. It confirms that analysis of the E1 and NS5B genomic regions using the neighbour-joining method is a reliable tool for the assignment of most new isolates.

Genetic variants of hepatitis C virus (HCV) have been classified in three major types including many subtypes on the basis of phylogenetic studies done on complete genomes. A fourth, fifth and sixth type have been defined, even though complete sequences are not available at the present time (Simmonds et al., 1994). In a series of studies, Tokita et al. (1994, 1995, 1996) described new variants from Indonesia, Vietnam and Thailand and proposed an extension of the previous classification to types 7, 8, 9, 10 and 11. Recently, Mizokami et al. (1996) showed that types 7, 8 and 9 could, in fact, be classified as type 6 subtypes when using the neighbour-joining method instead of the less reliable unweighted pair-group method with arithmetic mean (UPGMA) that had been previously used for the classification of these South-East Asian strains. However, these findings were not supported by numerical or statistical data.

In this paper, we have studied, using the neighbour-joining method, the phylogenetic status of these new strains including isolates from Jakarta (Indonesia) first described as types 10a and 11a (Tokita et al., 1996). All sequences were collected from the GenBank database. Phylogenetic studies were done with the software program MEGA (Kumar et al., 1993) using the p-distance determination algorithm for amino acid sequences, the Jukes–Cantor algorithm for nucleic acid sequences and the neighbour-joining method for tree-drawing. The reliability of the different phylogenetic groupings was evaluated by using the bootstrap test from the MEGA program.

The polyproteins deduced from complete coding regions of isolates JK049 and JK046 (classified as types 10a and 11a respectively) were compared with those of completely sequenced strains (type 1a: HCV-I, HC-J1 and HCV-H; type 1b: HCV-J, HCV-BK, HCV-T, HCV-JK1, HCV-JT, HCV-JT', HC-J4/83, HC-J4/91, HC-C2, HCV-N, HCV-HB, HCV-K1-S2 and HPCUNKCDS; type 1c: HC-G9; type 2a: HC-J6; type 2b: HC-J8; type 2c: BEBE1; type 3a: NZL1 and K3a/650; type 3b: HCV-Tr). The phylogenetic tree deduced from polyprotein sequences of completely sequenced HCV strains is shown in Fig. 1. The distance between strain JK049 and the type 3 strains (0.1770–0.1879) was comparable to that between the different subtypes of other types (0.1326–0.1678) and much lower than that between strains of different types (0.2440–0.3116). This suggests that JK049 could be classified as a new subtype of type 3.

All other trees were constructed after alignment of nucleic acid sequences: 672 nt sequences (nt 574–1245) of the envelope 1 (E1) genomic region and 1092 nt sequences (nt 7852–8943) of the nonstructural 5B (NS5B) genomic region of type 10a and 11a strains were compared with those of previously described strains belonging to various HCV types and subtypes, including isolates from Vietnam (Tokita et al., 1994) and Thailand (Tokita et al., 1995) (see Fig. 2). In both the E1 and NS5B regions, the type 7, 8 and 9 isolates clustered with the type 6 isolates as reported by Mizokami et al. (1996). The type 11a isolates were found in the same group. This
Fig. 1. Phylogenetic tree deduced from polyprotein sequences of completely sequenced HCV strains; molecular evolutionary distances between these sequences for isolates belonging to the same subtype (Isolates), to different subtypes of the same type (Subtypes) and to different types (Types) are presented as minimum to maximum (average ± SD). The horizontal bar indicates the number of amino acids substitutions per site. Subtypes in parentheses, e.g. (11a), are those of Tokita et al.

The recent description of many South-East Asian variants of HCV highlights the difficulties in classification of this virus and the importance of the phylogenetic methods that are used. Although the results of genetic analysis of isolates of types 1–5 are similar when using various analysis procedures, this is not the case when these new isolates are studied, and their classification is therefore still controversial. As previously reported by Mizokami et al. (1996) the neighbour-joining method is the most accepted means of analysis for the classification of HCV isolates.

When using the Jukes–Cantor distance determination algorithm and the neighbour-joining method for the E1 and NS5B genomic regions all the isolates described at the present time clustered in six major types divided into at least 30 subtypes. However, it must be noted that the unavailability of complete sequences for the HCV strains of types 4, 5 and 6 (with the exception of the VK046 isolate) renders definitive classification difficult.
Fig. 2. For legend see pp. 48.
Fig. 2. Comparisons of 672 nt sequences of the envelope 1 HCV genomic region (a) and 1092 nt sequences of the nonstructural 5B genomic region (b); molecular evolutionary distances between these sequences for isolates belonging to the same subtype (Isolates), to different subtypes of the same type (Subtypes) and to different types (Types) are presented as minimum to maximum (average ± SD). The horizontal bar indicates the number of nucleotides substitutions per site. Subtypes in parentheses, e.g. (11a), are those of Tokita et al.
Classification of hepatitis C virus variants

Fig. 3. For legend see pp. 50.
Fig. 3. Results of the bootstrap test (500 replications). The condensed trees in the envelope 1 (a) and nonstructural 5B (b) HCV genomic regions show only the branches that are supported at a Bootstrap Confidence Level (BCL) over 95%. Subtypes in parentheses, e.g. (11a), are those of Tokita et al.

References


Tokita, H., Okamoto, H., Luengojanakul, P., Vareesanghip, K.,


Received 1 July 1996; Accepted 7 October 1996