

Unique NS5b Hepatitis C Virus Gene Sequence Consensus Database Is Essential for Standardization of Genotype Determinations in Multicenter Epidemiological Studies

Syria Laperche,¹ Karine Saune,² Paul Dény,³ Gilles Duverlie,⁴ Sophie Alain,⁵ Marie-Laure Chaix,⁶ Catherine Gaudy,⁷ Françoise Lunel,⁸ Jean-Michel Pawlotsky,⁹ Christopher Payan,⁸ Bruno Pozzetto,¹⁰ Catherine Tamalet,¹¹ Vincent Thibault,¹² Sophie Vallet,¹³ Françoise Bouchardeau,¹ Jacques Izopet,² and Jean-Jacques Lefrère^{14,15*}

*Centre National de Référence pour les hépatites B et C en transfusion, Département des Agents Transmissibles par le Sang, Institut National de la Transfusion Sanguine, Paris, France*¹; *Laboratoire de Virologie, Centre Hospitalo-Universitaire Purpan, Toulouse, France*²; *Laboratoire de Virologie-Bactériologie associé au CNR des hépatites B et C, Centre Hospitalo-Universitaire Avicenne, Bobigny, France*³; *Laboratoire de Virologie, Centre Hospitalo-Universitaire, Amiens, France*⁴; *Laboratoire de Virologie-Bactériologie, Centre Hospitalo-Universitaire, Limoges, France*⁵; *Laboratoire de Virologie, Centre Hospitalo-Universitaire Necker, Paris, France*⁶; *Université François Rabelais, EA 3856 et laboratoire de Bactériologie-Virologie, Centre Hospitalo-Universitaire Bretonneau, Tours, France*⁷; *Laboratoire de Virologie-Bactériologie, Centre Hospitalo-Universitaire, Angers, France*⁸; *Laboratoire de Virologie-Bactériologie, Centre Hospitalo-Universitaire Henri-Mondor, Créteil, France*⁹; *Laboratoire de Virologie-Bactériologie, Centre Hospitalo-Universitaire, Saint Etienne, France*¹⁰; *Laboratoire de Virologie-Bactériologie, Centre Hospitalo-Universitaire La Timone, Marseille, France*¹¹; *Laboratoire de Virologie, Centre Hospitalo-Universitaire Pitié-Salpêtrière, Paris, France*¹²; *Laboratoire de Microbiologie, Centre Hospitalo-Universitaire, Brest, France*¹³; *Laboratoire d'Hématologie, Centre Hospitalo-Universitaire, Amiens, France*¹⁴; and *Département des Agents Transmissibles par le Sang, Institut National de la Transfusion Sanguine, Paris, France*¹⁵

Received 13 September 2005/Returned for modification 1 November 2005/Accepted 29 November 2005

A multicenter study of NS5b hepatitis C virus (HCV) genotype determination involving 12 laboratories demonstrates that any laboratory with expertise in sequencing techniques would be able to provide a reliable HCV genotype for clinical and epidemiological purposes as long as they are provided a consensus reference sequence database.

In a previous multicenter study designed to assess the applicability of hepatitis C virus (HCV) genotyping methods (5), we found a wide heterogeneity of results depending on both the laboratories and the genotyping methods. We conducted a new multicenter study of genotype determination based on HCV NS5b sequencing in order to further evaluate the performances of the laboratories involved in the previous study and to define a consensus method for further epidemiological studies.

The panel included 12 samples collected from HCV-infected blood donors and selected as a subset of subtypes currently found in Europe. Each HCV-positive sample was characterized by its viral load and by its HCV genotype determined with a method based on the NS5b region sequence analysis (7). The genotype of each sample was determined by comparison (with a neighbor-joining phylogenetic tree [3]) of its sequence with the same region of a selection of 92 genomes obtained from GenBank (17 genomes of genotype 1, 30 of genotype 2, 10 of genotype 3, 25 of genotype 4, 4 of genotype 5a, and 6 of genotype 6) and selected for representing the main genotypes encountered in France. Genotypes were classified according to the nomenclature proposed by Simmonds et al. (9). The nu-

cleic acid sequences obtained in the laboratory which characterized the panel were considered the reference sequences. The characteristics of these samples are given in Table 1.

Each laboratory could use the NS5b genotyping method of its choice. However, the database of the 92 selected reference sequences was proposed to the laboratories. Twelve laboratories (A to L) participated in this study. Ten used the same primers (7). Two laboratories (J and L) used different sets of primers (2, 6). Four interpreted the results with the recommended database, while eight used their own sequence database. The result was considered exact when both the genotype and the subtype were correctly identified. An incomplete result was defined as an exact genotype result with an unidentified subtype. A correct genotype associated with an incorrect subtype was interpreted as a misclassification. The accuracy was defined as the percentage of correct results (exact genotype and subtype) among the 12 samples. The protocol stated that nucleotide sequences of all misclassified samples obtained in any laboratory should be sent to the organizing committee and compared to all sequences obtained by the other laboratories in order to define the cause of the misclassification.

The accuracy ranged from 66.7% to 100% (Table 2). Four laboratories correctly identified all samples; four gave an incorrect result for a unique sample (three provided incomplete genotypes, and one misclassified one sample); two gave an

* Corresponding author. Mailing address: Institut National de la Transfusion Sanguine, 6 rue Alexandre-Cabanel, 75015 Paris, France. Phone: 01 44 49 30 51. Fax: 01 44 49 30 59. E-mail: lefrere@ints.fr.

TABLE 1. HCV genotype determination results for the laboratories (A to L)

Sample no.	Genotype	HCV RNA ^a (log IU/ml)	Genotype result for laboratory ^b											
			A	B	C	D	E*	F	G*	H*	I	J	K*	L
1	1a	4.93	1a	1a	1a	1a	Negative†	1a	1a	1a	1a	1a	1a	1a
2	1a	6.75	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	Negative†
3	1b	5.68	1b	1b	1b	1b	1b	1b	1b	1b	1b	1b	1b	1b
4	2a	5.38	2a	Negative†	2a	2a	2a	2a	2a	2a	Negative†	2a	2a	2a
5	2b	5.64	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b
6	2i	>5.93	2i	2†	2†	2i	2†	2i	2†	2†	2i	2c†	2†	2c†
7	3a	5.86	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
8	3a	5.11	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
9	4a	5.68	4a	4a	4a	4a	4a	4a	4a	4a	4a	4a	4a	4c†
10	4d	5.76	4d	4d	4d	4d	4d	4d	4d	4d	4d	4d	4d	4d
11	4h	5.56	4h	4h	4h	4h	4h	4h	4h	4a†	4h	4h	4h	4c†
12	5a	5.45	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a

^a Amplicor HCV Monitor version 2.0; Roche, Meylan, France.

^b The genotype 2i isolates included in the recommended database were initially classified as merely genotype 2. *, laboratories whose genotyping was based only on the recommended sequence database. For all other laboratories, genotyping was not based only on the recommended database. †, incorrect results.

incomplete result for one sample and missed one other sample; one laboratory had one incomplete result, one misclassification, and one false-negative result; and one participant misclassified three samples and had a false-negative result. Of the 144 correct expected results, 129 (89.6%) were accurate. Among the 15 incorrect results involving six samples, 4 were false-negative results, 5 were misclassifications, and 6 were incomplete genotypes.

PCR amplification of HCV RNA failed for three samples (four laboratories), once for samples no. 1 (genotype 1a, 4.9 log IU/ml) and no. 2 (genotype 1a, 6.8 log IU/ml) and twice for sample no. 4 (genotype 2a, 5.4 log IU/ml). Six samples (genotypes 1b, 2b, 3a [two samples], 4d, and 5a, with viral loads ranging from 5.1 to 5.8 log IU/ml) were correctly identified by all laboratories. Sample no. 6 was correctly identified as genotype 2i by four laboratories, while six laboratories classified it as genotype 2 and two misclassified it as genotype 2c. The percentage of sequence similarities with the reference sequence ranged from 92.5% to 100% (mean, 98%), and the majority of divergences were due not to nucleotide mutations but rather to uninvestigated ambiguities. The phylogenetic analysis performed with sequences provided retrospectively by each participant confirmed that all laboratories would have classified this sample as 2i by using an appropriate sequence database. Sample no. 9 was classified as genotype 4a by all participants but one (laboratory L), which concluded it was genotype 4c. The percentage of sequence similarities with the reference sequence ranged from 94.9% to 100% (mean, 97.9%) and was estimated at 98% for laboratory L. According to the analysis of the provided sequence by this laboratory, this

sample would have been classified as 4a. Finally, sample no. 11 (genotype 4h) was wrongly classified as genotypes 4a and 4c by laboratories H and L, respectively. The percentage of sequence similarities with the 4h reference sequence ranged from 83.8% to 100% (mean, 98%). The phylogenetic comparison of the reference sequences and the one obtained by the participating laboratories indicates that the laboratory H sequence was of genotype 4h but that the sequence from laboratory L was rather of genotype 4a.

By comparison with the previous study (5), the present work exhibits an improvement in the rate of correct genotyping results, which is essentially attributable to the decrease of false-negative results, which dropped from 18.3% to 3% in samples with comparable viral loads in the two panels. The false-negative results do not seem to be linked to the viral loads (which were relatively high) but rather to laboratory technical failures. The two laboratories which missed the same sample (no. 4) used the same consensus set of primers as the other eight laboratories. The lack of standardization for NS5b amplification and sequencing did not seem to affect the quality of the results, and a consensual HCV sequencing method would not be useful. However, we clearly identify as a pitfall the use of several uncoordinated reference sequences. Indeed, the minor divergences observed were linked solely to sequence databases used by each participant. It is noteworthy that more than 50% of the incorrect results were due to the inability to provide the subtype of one sample (no. 6, genotype 2i). This failure is explained by the fact that some participants used a genotype database that did not include any genotype 2i sequences or a database that misclassified 2i isolates as merely genotype 2 due

TABLE 2. Performance of the laboratories (A to L) in HCV genotype determination

Parameter	Result for laboratory											
	A	B	C	D	E	F	G	H	I	J	K	L
No. (%) of correct results	12 (100)	10 (83.3)	11 (91.7)	12 (100)	10 (83.3)	12 (100)	11 (91.7)	9 (75.0)	12 (100)	11 (91.7)	11 (91.7)	8 (66.7)
No. of incomplete results ^a	0	1	1	0	1	0	1	1	0	0	1	0
No. of misclassified samples ^b	0	0	0	0	0	0	0	1	0	1	0	3
No. of false-negative results	0	1	0	0	1	0	0	1	0	0	0	1

^a Correct genotype with a subtype not identified.

^b Correct genotype but incorrect subtype.

to low bootstrap value. The use of an incomplete database could also be incriminated for the five misclassifications of three samples (no. 6, 9, and 11). Indeed, the retrospective phylogenetic analysis of all laboratories' nucleic acid sequences of these three specimens led to a correct classification of the strains, except for one laboratory. In the latter case, the obtained sequence rooted as genotype 4a even though the sample was 4h. Had the laboratories used a unique sequence data set, the overall performance would have reached 96.5%. Consequently, the constitution of a consensual sequence database for multicenter studies appears crucial. Such a consensual bank should include a large amount of reference sequences corresponding to a relevant panel of published isolates. The constitution of this bank could first be based on international available sequences (Los Alamos [4], JSPs Kakehni, euHCVdb, and NIBSC [8] databases), where each sequence is annotated by its genotype, but should also take into account the geographical distribution and the epidemiological situation of the country. Moreover, due to the rapid evolution of HCV and consequently the continual characterization of new genotypes (especially in types 2 and 4 [1, 10]), this consensual bank should regularly include newly described subtypes.

This multicenter study demonstrates that any laboratory with expertise in sequencing techniques would be able to provide a reliable HCV genotype for clinical and epidemiological purposes as long as they are provided a consensus reference sequence database.

This work was supported by a grant from the Agence Nationale de Recherches sur le SIDA et les virus des hépatites.

We thank Joëlle Lerable, Annie Razer, and Christine Portal for their technical assistance.

REFERENCES

1. **Cantaloube, J. F., P. Gallian, H. Attoui, P. Biagini, P. De Micco, and X. de Lamballerie.** 2005. Genotype distribution and molecular epidemiology of hepatitis C virus in blood donors from southeast France. *J. Clin. Microbiol.* **43**:3624–3629.
2. **Enomoto, N., A. Takada, T. Nakao, and T. Date.** 1990. There are two major types of hepatitis C virus in Japan. *Biochem. Biophys. Res. Commun.* **170**:1021–1025.
3. **Felsenstein, J.** 1989. PHYLIP—phylogenetic inference package (version 3.2). *Cladistics* **5**:164–166.
4. **Kuiken, C., K. Yusim, L. Boykin, and R. Richardson.** 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics* **21**:379–384.
5. **Laperche, S., F. Lunel, J. Izopet, S. Alain, P. Deny, G. Duverlie, C. Gaudy, J. M. Pawlotsky, J. C. Plantier, B. Pozzetto, V. Thibault, F. Tosetti, and J. J. Lefrere.** 2005. Comparison of hepatitis C virus NS5b and 5' noncoding gene sequencing methods in a multicenter study. *J. Clin. Microbiol.* **43**:733–739.
6. **Mellor, J., E. C. Holmes, L. M. Jarvis, P. L. Yap, P. Simmonds, et al.** 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.
7. **Morice, Y., D. Roulot, V. Grando, J. Stirnemann, E. Gault, V. Jeantils, M. Bentata, B. Jarrousse, O. Lortholary, C. Pallier, and P. Deny.** 2001. Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. *J. Gen. Virol.* **82**:1001–1012.
8. **Rožanov, M., U. Plikat, C. Chappey, A. Kochergin, and T. Tatusova.** 2004. A web-based genotyping resource for viral sequences. *Nucleic Acids Res.* **32**:W654–W659.
9. **Simmonds, P., J. Mellor, T. Sakuldarnongpanich, C. Nuchaprayoon, S. Tanprasert, E. C. Holmes, and D. B. Smith.** 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.
10. **Zekri, A. R., H. M. Alam El-Din, A. A. Bahnassy, A. M. El-Shehabi, H. El-Leethy, A. Omar, and H. M. Khaled.** 2005. TRUGENE sequencing versus INNO-LiPA for sub-genotyping of HCV genotype-4. *J. Med. Virol.* **75**:412–420.