Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.

Elisa Martro, Ana Valero, Elena Jordana-Lluch, Verónica Saludes, Ramón Planas, Fernando Gonzalez, Vicente Ausina, Maria Alma Bracho

► To cite this version:

Elisa Martro, Ana Valero, Elena Jordana-Lluch, Verónica Saludes, Ramón Planas, et al.. Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.. Journal of Medical Virology, 2011, 83 (5), pp.820. 10.1002/jmv.22054 . hal-00624172

HAL Id: hal-00624172

<https://hal.science/hal-00624172>

Submitted on 16 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.

Journal:	<i>Journal of Medical Virology</i>
Manuscript ID:	JMV-10-2119.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	24-Dec-2010
Complete List of Authors:	<p>Martro, Elisa; Hospital Universitari Germans Trias i Pujol, Microbiology Service</p> <p>Valero, Ana; Centre for Infectious Diseases. University of Edinburgh., Microbial Pathogenesis Unit.</p> <p>Jordana-Lluch, Elena; Hospital Universitari Germans Trias i Pujol, Microbiology Service</p> <p>Saludes, Verónica; Hospital Universitari Germans Trias i Pujol, Microbiology Service</p> <p>Planas, Ramón; Hospital Universitari Germans Trias i Pujol, Liver Unit</p> <p>Gonzalez, Fernando; University of Valencia, Instituto Cavanilles Biodiversidad y Biología Evolutiva</p> <p>Ausina, Vicente; Hospital Universitari Germans Trias i Pujol, Microbiology Service</p> <p>Bracho, Maria Alma; CSISP-UVEG/Institut Cavanilles, Joint Unit of Research in Genomics and Health</p>
Keywords:	HCV, genotype 2, subtype 2q, complete genome

SCHOLARONE™
Manuscripts

Figure 1. Maximum likelihood phylogenetic trees. (a) Unrooted phylogenetic tree of the coding region (positions 342 to 9336), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.301, gamma shape parameter: 0.841); (b) genotype 2 subtree of the Core region (positions 342 to 914), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (c) genotype 2 subtree of the E1 region (positions 927 to 1262), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (d) genotype 2 subtree of the NS5B region (positions 8283 to 8591), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.279, gamma shape parameter: 0.833). Positions are given according to the H77 reference sequence. Sequences are identified by genotype, subtype and GenBank accession number. Sequences that have not been assigned to any subtype are denoted as "2_". Sequences newly described in this study are underlined. Nodes supported by a bootstrap support value >70% are indicated. The scale bar represents number of substitutions per nucleotide position.

47x35mm (600 x 600 DPI)

Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.

Elisa Martró,^{1,2*} Ana Valero,³ Elena Jordana-Lluch,¹ Verónica Saludes,^{1,2} Ramón Planas,^{4,5} Fernando González-Candelas^{2,6}, Vicente Ausina^{1,7}, and Maria Alma Bracho.^{2,6}

¹ Microbiology Service, Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain.

² CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain.

³ Microbial Pathogenesis Unit. Centre for Infectious Diseases. University of Edinburgh. UK.

⁴ Liver Unit, Hospital Universitari Germans Trias i Pujol, Badalona, Spain.

⁵ CIBER Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, Spain.

⁶ Joint Unit of Research in Genomics and Health CSISP-UVEG/Institut Cavanilles, Valencia, Spain.

⁷ CIBER Enfermedades Respiratorias (CIBERES), Bunyola, Spain.

Shortened title: Confirmation of HCV subtype 2q

*Corresponding author: Elisa Martró. Mailing address: Microbiology Service, Hospital Universitari Germans Trias i Pujol, Ctra. del Canyet s/n, 08916 Badalona, Spain. Phone: +34 934978 894. Fax: +34 934978 895. E-mail: emartro.igtp.germanstrias@gencat.cat

ABSTRACT

The hepatitis C virus (HCV) has been classified into 6 genotypes and more than 70 subtypes with distinct geographical and epidemiological distributions. While 18 genotype 2 subtypes have been proposed, only five have had their complete sequence determined. The aim of this study was to characterize HCV isolates from three patients from the Barcelona metropolitan area of Spain for whom commercial genotyping methods provided discordant results. Full-length genome sequencing was carried out for two of the three patients; for the third patient only partial NS5B sequences could be obtained. The generated sequences were subjected to phylogenetic, recombination and identity analyses. Sequences covering most of the HCV genome (9398 and 9566 nt in length) were obtained and showed a 90.3% identity to each other at the nucleotide level, while both sequences differed by 17.5-22.6% from the other fully sequenced genotype 2 subtypes. No evidence of recombination was found. The NS5B phylogenetic tree showed that sequences from the three patients cluster together with the only representative sequence of the provisionally designed 2q subtype, which also corresponds to a patient from Barcelona. Phylogenetic analysis of the full coding sequence showed that subtype 2q was more closely related to subtype 2k. The results obtained in this study suggest that subtype 2q now meets the requirements for confirmed designation status according to consensus criteria for HCV classification and nomenclature, and its epidemiological value is ensured as it has spread among several patients in the Barcelona metropolitan area.

KEY WORDS: HCV, genotype 2, subtype 2q, complete genome

INTRODUCTION

Hepatitis C virus (HCV) is estimated to infect 3% (170 million people) of the world's population, and is the most important cause of chronic liver disease and liver transplantation [World Health Organization 2010]. HCV is an enveloped virus belonging to the *Hepacivirus* genus within the *Flaviviridae* family. Its genome is a positive-sense single-stranded RNA of approximately 9.6 Kb. The HCV genome is flanked by noncoding (NC) regions, and includes a single open reading-frame encoding for both structural and nonstructural (NS) proteins [Choo et al. 1991]. HCV is characterized by a high level of genetic variability, which has allowed to classify the virus into 6 major genotypes (differing from each other by 31-33% at the nucleotide level), and a number of subtypes within each genotype (differing from each other by 20-25% at the nucleotide level) [Kuiken and Simmonds 2009]. On the basis of a single genomic sequence from a Canadian emigrant from the Democratic Republic of Congo, genotype 7a has been assigned provisionally [Murphy 2007]. Different genotypes show specific geographical locations, whilst certain subtypes (1a, 1b, 2a and 3a) have become widely distributed as a result of transmission through intravenous drug use, blood products, and health-care-related procedures [Kuiken and Simmonds 2009].

HCV genotype is a major predictor of a sustained virological response to antiviral therapy (pegylated alpha-interferon plus ribavirin); HCV genotyping is required to tailor dose and duration of treatment [Ghany et al. 2009]. The target of choice for commercialized molecular diagnostics and most genotyping assays is the 5'NC region, which is highly conserved. However, subtype identification based on this region is not always reliable. Although subtyping is not clinically relevant in the context of the current therapy, accurate subtype identification is useful for understanding the epidemiology and virological features of HCV. Besides, it might be clinically important

in the coming era of specifically targeted antiviral therapy for HCV, since subtype-specific resistance profiles have been described for subtypes 1a and 1b both in infected patients [Kieffer et al. 2007, Sarrazin et al. 2007] and *in vitro* [McCown et al. 2009]. Consensus criteria have been established for HCV classification and nomenclature, and confirmed designation status of new subtypes with epidemiological value requires rigorous phylogenetic analysis of one or more complete genome sequences of the candidate subtype with no epidemiological connection [Kuiken and Simmonds 2009]. Eighteen subtypes within genotype 2 have been assigned so far (2a to 2r), but only four have been confirmed (2a, 2b, 2c, and 2k) and a fifth one proposed (2i) on the basis of their complete genome sequences.

In a previous study comparing three commercial kits for HCV-genotyping based on the 5'NC region [Martró et al. 2008], three specimens that showed discordant results were subjected to NS5B sequencing. Upon phylogenetic analysis, they did not cluster with any of the previously confirmed genotype 2 subtypes, but formed a monophyletic group with the only representative sequence of the provisionally assigned 2q subtype [Tokita et al. 1998]. This study reports the HCV genomic sequence for two of those isolates and a partial NS5B sequence for the third one, and suggests changing the status of subtype 2q from proposed to confirmed designation.

METHODS

Patients and specimens. Three patients with a discordant genotyping result based on the 5'NC region were selected retrospectively from a previous study [Martró et al. 2008]. Isolates from patients 963 and 852 were genotyped as 2a/2c by reverse hybridization (Inno-LiPA HCV II Assay, Innogenetics, Gent, Belgium) and were indeterminate by real-time PCR (HCV genotyping analyte specific reagent [ASR] assay; Abbott

Molecular Inc., Des Plaines, IL), while isolate 985 was genotyped as 2a/2c by reverse hybridization, 2c by sequencing (TruGene HCV 5'NC genotyping kit, Bayer HealthCare, Berkeley, CA), and 2a by the real-time PCR assay. Only by means of NS5B sequencing and phylogenetic analysis evidence could be obtained showing that the three isolates belonged probably to the previously proposed subtype 2q.

The clinical and epidemiological characteristics of the three patients are summarized in Table I. Sera obtained at the date of HCV diagnosis had been conserved at -80°C. Patient 985 had insufficient serum volume and only HCV isolates from patients 963 and 852 could be subjected to full-length genome sequencing.

This study was approved by the Clinical Research Ethics Committee ("Comité Ético de Investigación Clínica", CEIC) at Hospital Universitari Germans Trias i Pujol (HUGTIP).

RNA extraction, RT-PCR and sequencing. HCV RNA was extracted from 200 µl of serum using the High Pure Viral RNA Kit (Roche) or QIAamp[®] viral RNA Mini kit (QIAGEN[®]) following the manufacturer's protocol. Reverse transcription (RT) was performed with 10 µl of extracted RNA in a total reaction volume of 20 µl containing 4 µl of 5× First-Strand buffer, 5 mM DTT, 200 U of SuperScript[®] III Reverse Transcriptase (Invitrogen), 40 U of rRNasin[®] Ribonuclease Inhibitor (Promega), 0.5 mM of each dNTP, and either 0.75 µg of random hexadeoxynucleotides (Amersham-Pharmacia Biotech) or 0.5 µM of antigenomic primer. The mix containing dNTP, random hexadeoxynucleotides or primer, and the RNA was incubated at 65°C for 5 min. Once the rest of components were added, the reaction was incubated at 50°C for 60 min (for random hexadeoxynucleotides) or 55°C for 60 min (for primer), followed by 15 min at 70°C.

Table I in supplementary material lists the primers used to obtain overlapping RT-PCR products covering most of the viral genome. Primer design was initially based on primers used in the sequencing of the 2i/6p recombinant (GenBank accession number DQ155560) [Noppornpanth et al. 2006]. First round and hemi-nested PCR amplifications were carried out in a final reaction volume of 50 µl containing either 5 µl of the RT product (in the case of first round PCR) or 1 µl of the first round PCR product (in the case of hemi-nested PCR), 5 µl of 10× PCR buffer with 15 mM MgCl₂, 200 µM of each dNTP, 400 nM each of both genomic sense and antigenomic sense primers, and 1 U of TaKaRa Taq™ (Takara Bio Europe SAS). Amplification was performed on 2700 or 9700 GeneAmp® PCR systems (Applied Biosystems) under the following conditions: 94°C for 2 min; 5 cycles, 94°C for 30 s, 50–65°C (depending on the primers used) for 30 s, 72°C for 3 min; 35 cycles, 94°C for 30 s, 52°C for 30 s and 72°C for 3 min; 72°C for 10 min. For the 3'NC region, the following conditions were used: 94°C for 2 min; 5 cycles, 94°C for 30 s, 55°C for 1 min, 72°C for 2 min; 35 cycles, 94°C for 30 s, 52°C for 1 min, 72°C for 2 min; 72°C for 10 min. Amplified products were checked by electrophoresis on 1.8% agarose gels stained with ethidium bromide, and purified with either High Pure PCR Products Purification Kit (Roche) or QIAquick PCR purification kit (QIAGEN®).

Direct sequencing of PCR products was performed for both strands using the BigDye™ Terminator v3.1 Ready Reaction Cycle Sequencing Kit on ABI Prism 3730 or 3100-Avant Genetic Analyzers (Applied Biosystems) and the primers listed in Table II in supplementary material. The obtained chromatogram files were assembled, verified, and edited using the Staden Package [Staden et al. 2000].

The obtained sequences were deposited in the EMBL Nucleotide Sequence Database with accession numbers FN666428 (patient 963), FN666429 (patient 852), and FN666430 (patient 985).

Phylogenetic reconstructions and identity analysis. The nucleotide sequences corresponding to the complete polyprotein from patients 963 and 852 were aligned with 27 homologous sequences representative of the main HCV genotypes and subtypes (see GenBank accession numbers, genotypes and subtypes in Figure 1) with ClustalW [Thompson et al. 1994] implemented in MEGA v4.0 [Tamura et al. 2007]. Alignments corresponding to partial genomic regions (Core, E1 and NS5B) were also obtained. The alignments included the most similar sequences found in a BLAST search against HCV sequences, including all 2k deposited sequences and sequences for all genotype 2 subtypes available, as well as reference sequences for the other genotypes. The sequence from patient 985 was also considered for the analysis of the NS5B region.

Maximum likelihood phylogenetic trees were constructed from the corresponding alignments with the evolutionary model implemented in PHYML that best fitted the data according to jModeltest [Posada 2008]. The robustness of the tree topologies was assessed by bootstrap analysis implemented in PHYML [Guindon and Gascuel 2003] with 1000 pseudorandom resamplings of the sequences. GenBank accession numbers of the reference sequences used in each phylogenetic reconstruction are shown in Figure 1. Pairwise nucleotide identities between genotype 2 subtypes including the newly described 2q sequences were calculated using MEGA v4.0, using an alignment encompassing nucleotide positions 31 (5'NC region) to 9336 (end of the 2i sequence DQ155561) of the H77 reference sequence (GenBank accession number AF009606). Reference sequences for the following HCV subtypes were used in this analysis:

Subtype 2a (GenBank accession number D00944), 2b (D10988), 2c (D50409), 2k (AB031663) and 2i (DQ155561).

Recombination analysis. Recombination analysis was performed using the recombination detection program RDP v3.44 [Martin and Rybicki 2000] with an alignment that included the two genomic sequences generated in this study along with the 27 representative complete sequences used in the phylogenetic analyses [Heath et al. 2006]. This program implements several methods for the identification of recombinant sequences and recombination breakpoints. Recombination analysis was performed using six of them: two phylogenetic methods that detect recombination when different regions of the genome lead to discordant topologies (RDP, and Bootscan/Recscan [Martin et al. 2005]), and four nucleotide substitution methods that search either for a significant clustering of substitutions or for a fit to an expected statistical distribution (GeneConv [Padidam et al. 1999], Chimaera [Posada and Crandall 2001], MaxChi [Maynard Smith 1992], and SiScan [Gibbs et al. 2000]). Default RDP v3.44 settings were used throughout, except that sequences were considered to be linear. Only putative recombination events inferred by more than two methods were taken into account, as performance evaluation of these recombination detection methods indicated that the results of a single method are not completely reliable [Posada 2002].

RESULTS

The HCV genomic sequences obtained were 9398 and 9566 nt in length for patients 852 and 963, respectively. Both sequences were partial at the 5' end of the 5'NC region; the first one ended 18 nt before the end of the polyprotein coding region (the NS5B region is partial at the 3' end), and the later included part of the 3'NC region.

Figure 1a shows the unrooted ML phylogenetic tree corresponding to the polyprotein-encoding nucleotide sequence. For genotype 2, only the five fully sequenced subtypes could be included in the analysis (2a, 2b, 2c, 2i and 2k), since complete genome sequences are still lacking for the other 13 assigned subtypes. This tree shows a highly supported grouping of sequences from patients 852 and 963, which would constitute a highly supported clade along with subtype 2k. Similar results were obtained in the phylogenetic analysis of the full Core region (573 nt) in which a few additional subtype 2 sequences were included (Figure 1b).

Phylogenetic analysis of partial E1 and NS5B regions provided additional information. The phylogenetic tree of a NS5B subregion (309 nt) included the three patients considered in this study (852, 963 and 985, for which only this partial NS5B sequence could be obtained). As shown in Figure 1d, these three sequences cluster together with the only representative sequence of the provisionally assigned 2q subtype (GenBank accession number D86529) with a high bootstrap support. Interestingly, the latter corresponds to isolate BA045, which was obtained in the 90's from a patient residing in Barcelona, Spain, who had not travelled abroad, and had acquired HCV infection through transfusion (M. A. Barrera, personal communication). Sequence AF037240 from a patient from Burkina Faso (West Africa) and described as a putative new genotype 2 subtype [Jeannel et al. 1998], was the sequence related most closely to the 2q cluster. Similarly, the phylogenetic tree of the E1 region (336 nt) shows that the two 2q sequences clustered with maximum bootstrap support (Figure 1c), and non-typable sequences EF115812 and EF115840 from patients in Quebec, Canada [Murphy et al. 2007], were the most closely related sequences.

Recombination signals were investigated by means of phylogenetic and nucleotide substitution methods. These analyses based on the complete genome alignment showed

no evidence that the two newly described 2q genomes had been involved in recombination events involving other genotypes or genotype 2 subtypes.

Finally, when considering the coding region, the two subtype 2q genomes sequenced showed a 90% identity to each other at the nucleotide level, while both sequences differed by 17.5 to 22.6% from the other confirmed genotype 2 subtypes (Table II). The highest identity was observed between subtypes 2q and 2c (82.23%) while the identity between 2q and 2k was slightly lower (80.54%). However, genetic distances estimated by maximum likelihood (which attempt to estimate the true amount of change occurred between two subtypes) were 0.40 and 0.28 respectively, which reveals the closest relationship between 2q and 2k subtypes. When the length of the full coding region as well as the ten protein regions were compared between genotype 2 subtypes, all subtype 2q protein regions were identical in length to those of subtypes 2a, 2b and 2k (Table III).

DISCUSSION

Genetic diversity is greatest within HCV genotypes 6, 3 and 2, which contains more divergent sequences than other genotypes [Kuiken and Simmonds 2009]. Genotype 2 is highly prevalent and genetically diverse in sub-Saharan Africa and in Martinique [Candotti et al. 2003, Jeannel et al. 1998, Martial et al. 2004, Ruggieri et al. 1996]. In Europe, previous studies found a prevalence of 11% in France, with a considerable diversity [Thomas et al. 2007], and a 3% in Spain [Echevarria et al. 2006]. Genotype 2 represents 4.3% of the patients treated in the Hospital Universitari Germans Trias i Pujol (period 2003-2009, N=493), where a variety of subtypes have been identified, including 2a, 2b, 2c, 2i, 2j and 2q [Martró et al. 2008].

New genotype 2 subtypes with limited distribution have been proposed on the basis of a few partial genome sequences in certain geographical regions not only in Africa [Jeannel et al. 1998, Ruggieri et al. 1996], but also in Europe [Thomas et al. 2007, van Doorn et al. 1995]. Among HCV-2 subtypes, 2a, 2b, 2c and 2k are worldwide distributed while subtypes 2d to 2p, 2r and five other tentative subtypes [Cantaloube et al. 2008] have been detected in only one or a few countries. HCV 2q, which had been proposed previously on the basis of a single partial NS5B sequence from a patient from Barcelona [Tokita et al. 1998] is another example of a geographically restricted subtype. In this study, another three patients in the Barcelona metropolitan area were identified as bearing this subtype, and most of its genome has been sequenced in two of the three cases. Patients' ages and clinical records agree with the hypothesis that HCV-2q could have been emerged in Spain as a result of blood transfusion, probably before HCV screening was implemented in blood banks in 1990. In fact, the most common transmission route among patients infected by minority genotype 2 subtypes in the close area of Midi-Pyrénées in France was through blood transfusion, while subtype 2a was transmitted via intravenous drug use [Thomas et al. 2007].

Three commercial genotyping assays based on the 5'NC region either failed to assign a genotype or to subtype correctly the three 2q isolates. Nucleotide sequencing and phylogenetic analysis of more variable genomic regions has been recommended for HCV genotyping in consensus proposals [Simmonds et al. 2005]. Accordingly, commercialized assays based on the Core (Versant HCV genotype 2.0 assay, Siemens Medical Solutions Diagnostics) or the NS5B regions (Abbott RealTime HCV Genotype II, Abbott Molecular Diagnostics) have been developed over the last years for a better discrimination of genotype 1 subtypes. However, subtype assignment for the other genotypes still relies on the 5'NC region and often leads to errors [Chen and Weck

2002, Martró et al. 2008, Tamalet et al. 2003, Zeuzem et al. 1995]. Furthermore, only sequencing and phylogenetic analysis of coding regions, such as NS5B or Core, provides the chance to identify variants not yet classified.

A level of nucleotide differences of at least 15% is expected between subtypes within each genotype, including newly described variants [Kuiken and Simmonds 2009]. Accordingly, subtype 2q shows a nucleotide difference of 17.5-22.6% from the other confirmed genotype 2 subtypes. Phylogenetic analysis of the coding region showed that, among confirmed genotype 2 subtypes, subtype 2q was most closely related to subtype 2k. Furthermore, in the phylogenetic analysis of the NS5B and E1 regions, non-typable genotype 2 sequences in patients from Burkina Faso and Canada were identified as the most closely related sequences to subtype 2q. Several recombinants of different genotypes including genotype 2 have been described in several countries: 2k/1b [Kalinina et al. 2004], 2i/6p [Noppornpanth et al. 2006], and 2b/1b [Kageyama et al. 2006], and 2k/5 [Legrand-Abravanel et al. 2007]. Nevertheless, the two newly described 2q sequences showed no evidence of recombination.

In conclusion, phylogenetic, identity, and recombination analyses indicate that subtype 2q now meets the requirements for confirmed designation status according to consensus criteria for HCV classification and nomenclature [Kuiken and Simmonds 2009, Simmonds et al. 2005]. Furthermore, subtype 2q is of epidemiological value, since it has spread among several patients in the Barcelona metropolitan area.

ACKNOWLEDGEMENTS

This work was supported partially by CIBER Epidemiología y Salud Pública (CIBERESP), projects PI05/1131 (EM, RP, VA), CP09/0004 (EM), CP07/00078 (MAB), and BFU2008-03000 (FG-C, MAB) from “Instituto de Salud Carlos III - Fondo

de Investigaciones Sanitarias” (ISCH-FIS), and grants CD05/00258 (“Contratos Postdoctorales de Perfeccionamiento”) and “Miguel Servet” from “Ministerio de Ciencia e Innovacion” (MICINN), within the “Plan Nacional de Investigación científica, Desarrollo e Innovación Tecnológica (I+D+I)” (EM).

For Peer Review

REFERENCES

Candotti D, Temple J, Sarkodie F, Allain JP. 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol* 77:7914-7923.

Cantaloube JF, Gallian P, Laperche S, Elghouzzi MH, Piquet Y, Bouchardeau F, Jordier F, Biagini P, Attoui H, de Micco P. 2008. Molecular characterization of genotype 2 and 4 hepatitis C virus isolates in French blood donors. *J Med Virol* 80:1732-1739.

Chen Z, Weck KE. 2002. Hepatitis C virus genotyping: interrogation of the 5' untranslated region cannot accurately distinguish genotypes 1a and 1b. *J Clin Microbiol* 40:3127-3134.

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos D, Coit A, Medina-Selby P, Barr J, Weiner AJ, Bradleyt DW, Kuo G, and Houghton M. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 88:2451-2455.

Echevarria JM, Leon P, Pozo F, Avellon A. Follow-up of the prevalence of hepatitis C virus genotypes in Spain during a nine-year period (1996-2004). 2006. *Enferm Infecc Microbiol Clin* 24:20-25.

Ghany MG, Strader DB, Thomas DL, Seeff LB. 2009. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 49:1335-1374.

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573-582.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704.

Heath L, van der WE, Varsani A, Martin DP. 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 80:11827-11832.

- annel D, Fretz C, Traore Y, Kohdjo N, Bigot A, Pe GE, Jourdan G, Kourouma K, Maertens G, Fumoux F, Fournel JJ, Stuyver L. 1998. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol* 55:92-97.
- Kageyama S, Agdamag DM, Alesna ET, Leano PS, Heredia AM, Abellanos-Tac-An IP Jereza LD, Tanimoto T, Yamamura J, Ichimura H. 2006. A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. *J Med Virol* 78:1423-1428.
- Kalinina O, Norder H, Magnus LO. 2004. Full-length open reading frame of a recombinant hepatitis C virus strain from St. Petersburg: proposed mechanism for its formation. *J Gen Virol* 85:1853-1857.
- Kieffer TL, Sarrazin C, Miller JS, Welker MW, Forestier N, Reesink HW, Kwong AD, Zeuzem S. 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. *Hepatology* 46:631-639.
- Kuiken C, Simmonds P. 2009. Nomenclature and numbering of the hepatitis C virus. In: Tang H, editor. *Hepatitis C: Methods and Protocols*. Humana Press, p 33-53.
- Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, Sandres-Saune K, Pasquier C, Izopet J. 2007. New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* 81:4357-4362.
- Martial J, Morice Y, Abel S, Cabié A, Rat C, Lombard F, Edouard A, Pierre-Louis S, Garsaud P, Béra O, Chout R, Gordien E, Deny P, Césaire R. 2004. Hepatitis C virus (HCV) genotypes in the Caribbean island of Martinique: evidence for a large radiation of HCV-2 and for a recent introduction from Europe of HCV-4. *J Clin Microbiol* 42:784-791.
- Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562-563.

- 1
2
3 Martin DP, Posada D, Crandall KA, Williamson C. 2005. A modified bootscan
4 algorithm for automated identification of recombinant sequences and recombination
5 breakpoints. *AIDS Res Hum Retroviruses* 21:98-102.
6
7
8
9
10 Martró E, González V, Buckton AJ, Saludes V, Fernández G, Matas L, Planas R,
11 Ausina V. 2008. Evaluation of a new assay in comparison with reverse hybridization
12 and sequencing methods for hepatitis C virus genotyping targeting both 5' noncoding
13 and nonstructural 5b genomic regions. *J Clin Microbiol* 46:192-197.
14
15
16
17
18
19 Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *J Mol Evol* 34:126-
20 129.
21
22
23
24
25 McCown MF, Rajyaguru S, Kular S, Cammack N, Najera I. 2009. GT-1a or GT-1b
26 subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-
27 796. *Antimicrob Agents Chemother* 53:2129-2132.
28
29
30
31
32 Murphy DG, Willems B, Deschenes M, Hilzenrat N, Mousseau R, Sabbah S. 2007. Use
33 of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus
34 with reference to C/E1 and 5' untranslated region sequences
35 10. *J Clin Microbiol* 45:1102-1112.
36
37
38
39
40
41 Noppornpanth S, Lien TX, Poovorawan Y, Smits SL, Osterhaus AD, Haagmans BL.
42 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus.
43 *J Virol* 80:7569-7577.
44
45
46
47
48 Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses
49 by frequent recombination. *Virology* 265:218-225.
50
51
52
53 Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253-
54 1256.
55
56
57
58 Posada D. 2002. Evaluation of methods for detecting recombination from DNA
59 sequences: empirical data. *Mol Biol Evol* 19:708-717.
60

- Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 98(24):13757-13762.
- Ruggieri A, Argentini C, Kouruma F, Chionne P, D'Ugo E, Spada E, Dettori S, Sabbatani S, Rapicetta M. 1996. Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *J Gen Virol* 77:2073-2076.
- Sarrazin C, Kieffer TL, Bartels D, Hanzelka B, Muh U, Welker M, Wincheringer D, Zhou Y, Chu HM, Lin C, Weegink C, Reesink H, Zeuzem S, Kwong AD. 2007. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology* 132:1767-1777.
- Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspé G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42:962-973.
- Staden R, Beal KF, Bonfield JK. The Staden package, 1998. 2000. *Methods Mol Biol* 132:115-130.
- Tamalet C, Colson P, Tissot-Dupont H, Henry M, Tourres C, Tivoli N, Botta D, Ravaux I, Poizot-Martin I, Yahi N. 2003. Genomic and phylogenetic analysis of hepatitis C virus isolates: a survey of 535 strains circulating in southern France. *J Med Virol* 71:391-398.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
- Thomas F, Nicot F, Sandres-Saune K, Dubois M, Legrand-Abravanel F, Alric L, Peron JM, Pasquier C, Izopet J. 2007. Genetic diversity of HCV genotype 2 strains in south western France. *J Med Virol* 79:26-34.

Thomas F, Nicot F, Sandres-Saune K, Dubois M, Legrand-Abravanel F, Alric L Peron JM, Pasquier C, Izopet J. 2007. Genetic diversity of HCV genotype 2 strains in south western France. *J Med Virol* 79:26-34.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.

Tokita H, Okamoto H, Iizuka H, Kishimoto J, Tsuda F, Miyakawa Y, Mayumi M. 1998. The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups. *J Gen Virol* 79:1847-1857.

van Doorn LJ, Kleter GE, Stuyver L, Maertens G, Brouwer JT, Schalm SW, Heijntink RA, Quint WG. 1995. Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries. *J Gen Virol* 76:1871-1876.

World Health Organization. 2010. Viral Cancers: Hepatitis C virus. www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html#disease%20burden.

Zeuzem S, Ruster B, Lee JH, Stripf T, Roth WK. 1995. Evaluation of a reverse hybridization assay for genotyping of hepatitis C virus. *J Hepatol* 23:654-661.

FIGURE LEGENDS

Figure 1. Maximum likelihood phylogenetic trees. (a) Unrooted phylogenetic tree of the coding region (positions 342 to 9336), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.301, gamma shape parameter: 0.841); (b) genotype 2 subtree of the Core region (positions 342 to 914), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (c) genotype 2 subtree of the E1 region (positions 927 to 1262), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (d) genotype 2 subtree of the NS5B region (positions 8283 to 8591), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.279, gamma shape parameter: 0.833). Positions are given according to the H77 reference sequence. Sequences are identified by genotype, subtype and GenBank accession number. Sequences that have not been assigned to any subtype are denoted as “2_”. Sequences newly described in this study are underlined. Nodes supported by a bootstrap support value >70% are indicated. The scale bar represents number of substitutions per nucleotide position.

TABLES

Table I. Characteristics of the four patients identified so far as being infected with HCV subtype 2q.

Characteristic	Patient ID			
	852	963	985	BA045
Gender	Male	Female	Female	Male
HCV risk factor	Surgery, several transfusions (1959)	Multiparous, two abortions (<1990), curettage (1992)	Cervical carcinoma surgery (1990), possible transfusion	Transfusion
Date of HCV diagnosis	2001	2002	2001	-
Age (years)	60	51	69	-
HCV treatment	No	No	Not eligible for treatment	Yes
Clinical outcome	Persistently normal transaminase levels Discharged (2005)	Cirrhosis, portal hypertension (esophagic varices)	Pluripathology (advanced renal insufficiency) Exitus (2005)	Sustained virological response
City of residence (province)	Santa Coloma (Barcelona)	Badalona (Barcelona)	Badalona (Barcelona)	Barcelona (Barcelona)

Table II. Identity (%) among genotype 2 subtypes that have been characterized by complete genome sequencing. GenBank accession numbers of reference sequences: D00944 (2a), D10988 (2b), D50409 (2c), AB031663 (2k) and DQ155561 (2i). Nucleotide identity was computed between nucleotide positions 31 to 9336 according to the H77 reference sequence. Patient ID is indicated in parenthesis.

	2a	2b	2c	2i	2k	2q (852)
2a	-					
2b	76.75	-				
2c	79.44	76.09	-			
2i	79.93	77.33	80.29	-		
2k	80.16	76.64	80.09	81.10	-	
2q (852)	79.62	77.42	82.52	80.40	79.47	-
2q (963)	80.04	77.58	82.23	80.54	80.10	90.34

Table III. Size composition of HCV protein regions among the genotype 2 subtypes that have been characterized by complete genome sequencing. ^aBased on curated alignments available at the HCV LANL Database (<http://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html>). ^bPartial sequence at its 3' end. ^cBased only on the 963 isolate. GenBank accession numbers of reference sequences: D00944 (2a), D10988 (2b), D50409 (2c), AB031663 (2k) and DQ155561 (2i).

	Region length (nt) ^a										
	Core	E1	E2	p7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	Polypeptide
2a	573	576	1101	189	651	1893	162	783	1398	1776	9102
2b	573	576	1101	189	651	1893	162	783	1398	1776	9102
2c	573	576	1101	189	651	1893	162	783	1410	1776	9114
2i	573	576	1101	189	651	1893	162	783	1395	1734 ^b	9057 ^b
2k	573	576	1101	189	651	1893	162	783	1398	1776	9102
2q	573	576	1101	189	651	1893	162	783	1398	1776 ^c	9102 ^c