Expanded Classification of Hepatitis C Virus Into 7 Genotypes and 67 Subtypes: Updated Criteria and Genotype Assignment Web Resource

Donald B. Smith,1 Jens Bukh,2 Carla Kuiken,3 A. Scott Muerhoff,4 Charles M. Rice,5 Jack T. Stapleton,6 and Peter Simmonds1

The 2005 consensus proposal for the classification of hepatitis C virus (HCV) presented an agreed and uniform nomenclature for HCV variants and the criteria for their assignment into genotypes and subtypes. Since its publication, the available dataset of HCV sequences has vastly expanded through advancement in nucleotide sequencing technologies and an increasing focus on the role of HCV genetic variation in disease and treatment outcomes. The current study represents a major update to the previous consensus HCV classification, incorporating additional sequence information derived from over 1,300 (near-)complete genome sequences of HCV available on public databases in May 2013. Analysis resolved several nomenclature conflicts between genotype designations and using consensus criteria created a classification of HCV into seven confirmed genotypes and 67 subtypes. There are 21 additional complete coding region sequences of unassigned subtype. The study additionally describes the development of a Web resource hosted by the International Committee for Taxonomy of Viruses (ICTV) that maintains and regularly updates tables of reference isolates, accession numbers, and annotated alignments (http://talk.ictvonline.org/links/hcv/hcv-classification.htm). The Flaviviridae Study Group urges those who need to check or propose new genotypes or subtypes of HCV to contact the Study Group in advance of publication to avoid nomenclature conflicts appearing in the literature. While the criteria for assigning genotypes and subtypes remain unchanged from previous consensus proposals, changes are proposed in the assignment of provisional subtypes, subtype numbering beyond “w,” and the nomenclature of intergenotypic recombinant. Conclusion: This study represents an important reference point for the consensus classification of HCV variants that will be of value to researchers working in clinical and basic science fields. (HEPATOLOGY 2014;59:318-327)

Soon after the publication of the first nearly complete genome sequence of hepatitis C virus (HCV) in 1989,1 it became apparent that isolates from different individuals or countries showed substantial genetic diversity. After much research and surveying by groups worldwide, this variation was summarized and variants assigned as genotypes and subtypes in a consensus classification and nomenclature system and formal rules were agreed for the assignment and naming of future variants.2 Genotype and subtype assignments required: (1) one or more complete coding region sequence(s); (2) at least three epidemiologically unrelated isolates; (3) a phylogenetic group distinct from previously described sequences; (4) exclusion of intergenotypic or intersubtypic recombination.

Abbreviations: HCV, hepatitis C virus; ICTV, International Committee for Taxonomy of Viruses.

From the 1Centre for Immunity, Infection and Evolution, University of Edinburgh, Scotland, UK; 2Copenhagen Hepatitis C Program (CO-HEP), Department of Infectious Diseases and Clinical Research Centre, Copenhagen University Hospital, Herlev, and Department of International Health, Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; 3Theoretical Biology and Biophysics group, Los Alamos National Laboratory, Los Alamos, NM, USA; 4Abbott Diagnostics Research and Development, Abbott Park, IL, USA; 5Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, NY, USA; 6Medical Service, Iowa City Veterans Affairs Medical Center, Departments of Internal Medicine and Microbiology, University of Iowa, Iowa City, IA, USA.

Received July 25, 2013; accepted August 29, 2013.

Supported by a grant from the Wellcome Trust to the Centre for Immunity, Infection and Evolution at the University of Edinburgh, Scotland, UK.
The application of these criteria confirmed the assignment of six distinct genotypes, comprising 18 subtypes. In addition, 58 subtypes were provisionally assigned pending the availability of a complete coding region sequence or additional isolates. This agreement on nomenclature was mirrored by the establishment of several curated databases that organized HCV sequences as they became available and indicated which genotypes and subtypes were confirmed or provisionally assigned (Los Alamos HCV Sequence Database, euhCVdb, Hepatitis Virus Database: http://s2as02.genes.nig.ac.jp/). Concurrently, a proposal was made to unify the numbering of HCV with reference to the genotype 1a isolate H77 (AF009606).

Recently, this remarkable agreement and cooperation in HCV nomenclature has been complicated by several developments. None of the HCV sequence databases are now actively curated and responsibility for naming new genotypes and subtypes has reverted de facto to individual researchers. This, combined with publication delays, has created new contradictions in which isolates assigned to the same subtype (4b: FJ462435, FJ025855, FJ025856, and FJ025854; 6k: DQ278891 and DQ278893; 6u: EU408330, EU408331, and EU408332) belong to different subtypes according to the consensus criteria. Another challenge is that the number of complete coding region sequences has increased from 238 in 2005 to more than 1,300. Similarly, the number of variants matching the criteria for assignment as confirmed genotypes/subtypes has expanded from 18 to 67; several recent publications contain figures that are illegible with regard to isolate name and/or accession number, complicating subsequent comparisons.

Finally, advances in sequencing technology have accelerated the rate at which HCV sequences are generated. Recent articles have reported the partial sequences of 282 isolates from Vietnam and 393 isolates from China, in each case identifying additional subtypes of genotype 6. Technological advances have also made it easier to obtain HCV complete coding region sequences through both dideoxysequencing and pyrosequencing. The latter technique was recently used to obtain 31 complete coding region sequences belonging to 13 different subtypes. More than 225,000 HCV sequences are now available on GenBank and about 30,000 added every year. This volume of sequence information and the diversity of known HCV variants make it increasingly important for researchers to have a single curated resource to refer to for accurate subtype designations, reference genomes and alignments.

This article updates the genotype and subtype assignments and the nomenclature rules, and describes the establishment of a reference Website hosted by the International Committee for the Taxonomy of Viruses (ICTV) to validate new genotype and subtype assignments, and provide updated reference alignments.

Revision of Confirmed Genotypes and Subtypes

Unique HCV complete or nearly complete coding region sequences available on NCBI Genome (969 sequences, http://www.ncbi.nlm.nih.gov/genome) and the Los Alamos HCV sequence database (1,364 sequences >8,000 nt from http://hcv.lanl.gov/content/index) were aligned within SSEv1.12 using Muscle v3.8.313 and refined manually. Phylogenetic analysis of sequences containing >95% of the coding region reveals seven major phylogenetic groupings corresponding to genotypes 1-7 (Fig. 1). Within these genotypes, grouping of the constituent subtypes is supported by 100% of bootstrap replications.

Based on the consensus criteria, confirmed subtypes (indicated by a letter following the genotype) require a complete or nearly complete coding region sequence differing from other sequences by at least 15% of nucleotide positions and sequence information from at least two other isolates in core/E1 (>90% of the sequence corresponding to positions 869 to 1,292 of the H77 reference sequence [accession number AF009606 numbered according to reference5] and NS5B (>90% of positions 8,276 to 8,615) (Table 1). The use of a 15% threshold over the complete coding region is supported by analysis of the large number of potential subtypes now sequenced (Fig. 2). This reveals major and consistently placed gaps in the distribution of pairwise distances between and within subtypes of
each genotype as follows: genotype 1: 12.9%-17.0%, genotype 2: 13.1%-17.6%, genotype 3: 12.5%-19.6%, genotype 4: 12.7%-15.3% (except distances of 14% and 14.2% between JX227963 and two subtype 4g sequences), and genotype 6: 9.9%-14.9% (except distances of 13.1%-13.7% between EU246931 and three subtype 6e sequences). Hence, for all genotypes and with remarkably few exceptions, a clear division can be made between isolates that differ by <13% over their complete coding region sequences (members of the same subtype) and those that differ by >15% (different genotypes or subtypes). This analysis includes sequences distinct from any of the confirmed HCV subtypes but not currently represented by three or more independent isolates that remain unclassified subtypes (Table 2). Whether the exceptions noted are due to technical problems or to differing epidemiological histories is unknown.

The seven confirmed genotypes (discussed below) comprise 67 confirmed subtypes, 20 provisionally

---

**Fig. 1.** Phylogenetic tree of 129 representative complete coding region sequences. Up to two representatives of each confirmed genotype/subtype were aligned (together with a third extreme variant of subtypes 4g and 6e) and a neighbor joining tree constructed using maximum composite likelihood nucleotide distances between coding regions using MEGA5. Sequences were chosen to illustrate the maximum diversity within a subtype. Tips are labeled by accession number and subtype (*unassigned subtype). For genotypes 1, 2, 3, 4, and 6, the lowest common branch shared by all subtypes and supported by 100% of bootstrap replicates (n = 1,000) is indicated by ●.
assigned subtypes, and 21 unassigned subtypes. These tables have been posted on the ICTV Website at http://talk.ictvonline.org/links/hcv/hcv-classification.htm and will be updated regularly by the authors with information shared across existing HCV databases (http://hcv.lanl.gov/; http://euhcvdb.ibcp.fr/euHCVdb/), typing tools, and other resources (e.g., http://www.bioafrica.net/rega-genotype/html/subtypinghcv.html; http://comet.retrovirology.lu/; http://hcv.lanl.gov/content/sequence/phyloplace/; http://s2as02.genes.nig.ac.jp/; http://www.viprbrc.org/). Alignments including representatives of these subtypes are available on the

### Table 1. Confirmed HCV Genotypes/Subtypes

<table>
<thead>
<tr>
<th>Genotype*</th>
<th>Locus/Isolate(s)</th>
<th>Accession number(s)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>HPCPU/PRE, HPCGAA</td>
<td>M62321, M67463</td>
<td>29,30</td>
</tr>
<tr>
<td>1b</td>
<td>HPCID, HPCUMR</td>
<td>D90208, MS8335</td>
<td>31,32</td>
</tr>
<tr>
<td>1c</td>
<td>HPCG3S, AV051292</td>
<td>D14863, AV051292</td>
<td>33</td>
</tr>
<tr>
<td>1a</td>
<td>148636</td>
<td>KC248194</td>
<td>9</td>
</tr>
<tr>
<td>1b</td>
<td>1804</td>
<td>AM91065</td>
<td>34</td>
</tr>
<tr>
<td>1h</td>
<td>EBW443, EBW9</td>
<td>KC248198, KC248199</td>
<td>9</td>
</tr>
<tr>
<td>1i</td>
<td>136142, EBW42</td>
<td>KC248193, KC248197</td>
<td>9</td>
</tr>
<tr>
<td>Genotype 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>HPCPOLP, JFH-1</td>
<td>D00944, AB047639</td>
<td>35,36</td>
</tr>
<tr>
<td>2b</td>
<td>HPCJ8G, JPUT971017</td>
<td>D10988, AB030097</td>
<td>37,38</td>
</tr>
<tr>
<td>2c</td>
<td>BEBE1</td>
<td>D50409</td>
<td>39</td>
</tr>
<tr>
<td>2d</td>
<td>QC559</td>
<td>JF735114</td>
<td>40</td>
</tr>
<tr>
<td>2e</td>
<td>QC64</td>
<td>JF735120</td>
<td>40</td>
</tr>
<tr>
<td>2f</td>
<td>D54</td>
<td>D10988</td>
<td>41</td>
</tr>
<tr>
<td>2g</td>
<td>1779, QC232</td>
<td>M777358, JF735113</td>
<td>4,40</td>
</tr>
<tr>
<td>2h</td>
<td>VA91</td>
<td>AB031663</td>
<td>42</td>
</tr>
<tr>
<td>2m</td>
<td>QC178, BID-G1314</td>
<td>JF735111, JX227967</td>
<td>40,8</td>
</tr>
<tr>
<td>2q</td>
<td>936, 852</td>
<td>FN666428, FN666429</td>
<td>43</td>
</tr>
<tr>
<td>2r</td>
<td>QC283</td>
<td>JF735115</td>
<td>40</td>
</tr>
<tr>
<td>Genotype 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>HPCEGS, HPCK3A</td>
<td>D17763, D28917</td>
<td>44,45</td>
</tr>
<tr>
<td>3b</td>
<td>HPCFG</td>
<td>D49374</td>
<td>46</td>
</tr>
<tr>
<td>3g</td>
<td>BID-G1243, QC260</td>
<td>JX227954, JF735123</td>
<td>8,21</td>
</tr>
<tr>
<td>3h</td>
<td>QC29</td>
<td>JF735121</td>
<td>21</td>
</tr>
<tr>
<td>3i</td>
<td>IND-HCV, BID-G1244</td>
<td>FJ407092, JX227955</td>
<td>8</td>
</tr>
<tr>
<td>3k</td>
<td>HPCKGK22, QC105</td>
<td>D63821, JF735122</td>
<td>47,21</td>
</tr>
<tr>
<td>Genotype 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a</td>
<td>ED43</td>
<td>Y11604</td>
<td>48</td>
</tr>
<tr>
<td>4b</td>
<td>QC264</td>
<td>FJ462435</td>
<td>16</td>
</tr>
<tr>
<td>4c</td>
<td>QC381</td>
<td>FJ462436</td>
<td>16</td>
</tr>
<tr>
<td>4d</td>
<td>03-18, QC382</td>
<td>D4108786, FJ462437</td>
<td>49,16</td>
</tr>
<tr>
<td>4f</td>
<td>IFR3, PS6</td>
<td>EFS391161, EU392175</td>
<td>50,51</td>
</tr>
<tr>
<td>4g</td>
<td>QC193</td>
<td>FJ462432</td>
<td>16</td>
</tr>
<tr>
<td>4k</td>
<td>PS3, QC383</td>
<td>EU392173, FJ462438</td>
<td>51,16</td>
</tr>
<tr>
<td>4l</td>
<td>QC274</td>
<td>FJ462433</td>
<td>16</td>
</tr>
<tr>
<td>4m</td>
<td>QC249</td>
<td>FJ462441</td>
<td>16</td>
</tr>
<tr>
<td>4o</td>
<td>QC313</td>
<td>FJ462431</td>
<td>16</td>
</tr>
<tr>
<td>4p</td>
<td>QC262</td>
<td>FJ462434</td>
<td>16</td>
</tr>
<tr>
<td>4r</td>
<td>QC384</td>
<td>FJ462439</td>
<td>16</td>
</tr>
<tr>
<td>4t</td>
<td>QC155</td>
<td>FJ389869</td>
<td>16</td>
</tr>
<tr>
<td>4v</td>
<td>CYHCV0713, BID-G1248</td>
<td>HQ537009, JX227959</td>
<td>52,8</td>
</tr>
<tr>
<td>4w*</td>
<td>P212, P245</td>
<td>FJ025855, FJ025856</td>
<td>14</td>
</tr>
<tr>
<td>Genotype 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>EUH1480, SA13§</td>
<td>Y13184, AFO64490</td>
<td>53,54</td>
</tr>
<tr>
<td>Genotype 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a</td>
<td>EUH2.6a33</td>
<td>Y12083, AY859526</td>
<td>55,56</td>
</tr>
<tr>
<td>6b</td>
<td>Ths580</td>
<td>D64262</td>
<td>57</td>
</tr>
<tr>
<td>6c</td>
<td>Ths580</td>
<td>D64262</td>
<td>57</td>
</tr>
<tr>
<td>6d</td>
<td>Ths580</td>
<td>D64262</td>
<td>57</td>
</tr>
<tr>
<td>6e</td>
<td>QCX004</td>
<td>DQ314805</td>
<td>60</td>
</tr>
<tr>
<td>6f</td>
<td>C-0044</td>
<td>DQ385760</td>
<td>60</td>
</tr>
<tr>
<td>6g</td>
<td>HPC2464E2</td>
<td>D63822</td>
<td>47</td>
</tr>
<tr>
<td>6h</td>
<td>VN004</td>
<td>D64265</td>
<td>57</td>
</tr>
<tr>
<td>6i</td>
<td>Ths53</td>
<td>DQ385770</td>
<td>60</td>
</tr>
<tr>
<td>6k</td>
<td>VN004</td>
<td>D64265</td>
<td>57</td>
</tr>
<tr>
<td>6l</td>
<td>S37796</td>
<td>EF424628</td>
<td>58</td>
</tr>
<tr>
<td>6m</td>
<td>B4/92</td>
<td>DQ385767</td>
<td>60</td>
</tr>
<tr>
<td>6n</td>
<td>KM42, D66/93</td>
<td>DQ278894, DQ385768</td>
<td>17,60</td>
</tr>
<tr>
<td>6o</td>
<td>QC227</td>
<td>EF424627</td>
<td>58</td>
</tr>
</tbody>
</table>

Additions and changes from assignments proposed in 2 shown in bold.

*Consensus proposed genotype/subtype names. Where multiple sequences of a HCV genotype are available, two sequences have been listed, prioritized by (a) publication date or (b) submission date when unpublished.

†Locus (or isolate name if locus is the same as the accession number).

§Sequence obtained from acute phase plasma of a chimpanzee experimentally infected with (human-derived) isolate SA13.

kPreviously described as 6u.18
The process of producing these tables has detected a small number of variants with conflicting assignments. Isolates P026, P212, P245, (FJ025854-6) are described as subtype 4b, but these complete coding region sequences show <85% identity to the core/E1 of isolate Z1 (U10235, L16677), provisionally assigned as 4b15 that is more closely related to core/E1 of the complete coding region sequence of isolate QC264 (FJ46243516). P212 and P245 belong to the same, novel subtype for which NS5B sequence is available from a third isolate (P213, GU049362), so this becomes confirmed subtype 4w. Isolate P026 differs from all other genotype 4 sequences by >17.5% but being represented by a single sequence remains currently unassigned (Table 2).

Similarly, isolates KM45 and KM41 (DQ278891,3) have been assigned to subtype 6k,17 but differ by >17% in complete coding region sequence from the subtype 6k isolate VN405 (D84264) and 6.7% from each other, and so remain an unclassified subtype of genotype 6. Two distinct groups of isolates have been assigned to subtype 6u; EU408330-218 and EU246940.19 The latter was submitted first to GenBank and is represented by NS5B sequences from two additional isolates and so is assigned subtype 6u, while EU408330, EU408331, and EU408332 are designated subtype 6xa (see below).

Finally, our analysis of both phylogenetic groupings and sequence distances suggests that a number of isolates20 described in their GenBank accessions as “subtype k-related” (QC273, TV257, TV476, KM35), “subtype l-related” (TV533, L349), “intermediate between subtypes 6m and 6n” (DH027), or “intermediate between subtypes 6j and 6i” (QC271) should be considered as unassigned novel subtypes.

### Additional Taxonomic Levels

In making this taxonomic distinction into virus genotypes and subtypes we are aware of the difficulties of imposing a discrete classification scheme on a complex taxonomy. In particular, for genotypes 3 and 6 there are undoubtedly several hierarchies of taxonomic relationships. For example, subtypes 6k and 6l form a clade along with several unassigned genotype 6 isolates.20 A higher-level clade includes these sequences and subtypes 6m and 6n, while further grouping consists of these subtypes and subtypes 6i and 6j (Fig. 1). These phylogenetic hierarchies are reflected in the discontinuous distribution of p-distances between complete coding region sequences (Fig. 2), which comprises three almost merging distributions (roughly 15% to 20%, 20% to 25%, and 25% to 30%). Three distributions of intersubtype distances were also observed for genotype 3 (20% to 25%, 25% to 27%, and 27% to 30%), two distributions for genotype 2 (18% to 22.5%, 23% to 26.5%), and uniform distributions for genotype 1 (17.7% to 25.4%) and genotype 4 (15.3% to 23.1%). However, the internal divisions defined by the multiple distributions of distances within genotypes 2, 3, and 6 have not been shown to correspond with geographical or epidemiological differences. The higher-level grouping of subtypes 3b, 3g, and 3i does not reflect a common geographical origin distinct from that of 3h and 3k.21 There is also no geographical correlation with the groupings of subtypes 6k, 6l, and various unassigned isolates; for 6m, 6n, and an unassigned isolate; for 6h, 6i, 6j, and an unassigned isolate; for 6a and 6b; for 6f and 6r; or for 6r and 6e.22 Similarly, there are currently no known virological or clinical reasons to recognize these higher-level groupings. Without practical

<p>| Table 2. Unassigned Complete Coding Region Sequences |
|-------------------|-------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th>Genotype*</th>
<th>Locus/Isolate(s)</th>
<th>Accession no(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype 1</td>
<td>1_A1851228</td>
<td>A1851228</td>
<td>A1851228</td>
</tr>
<tr>
<td></td>
<td>1_NC248195</td>
<td>160526</td>
<td>NC248195</td>
</tr>
<tr>
<td></td>
<td>1_HQ537007</td>
<td>CYNHC025</td>
<td>HQ537007</td>
</tr>
<tr>
<td>Genotype 2</td>
<td>2_JF735119</td>
<td>QC331</td>
<td>JF735119</td>
</tr>
<tr>
<td></td>
<td>2_JF735112</td>
<td>QC182</td>
<td>JF735112</td>
</tr>
<tr>
<td></td>
<td>2_JF735110</td>
<td>QC114</td>
<td>JF735110</td>
</tr>
<tr>
<td></td>
<td>2_JF735117</td>
<td>QC297</td>
<td>JF735117</td>
</tr>
<tr>
<td></td>
<td>2_JF735116</td>
<td>QC289</td>
<td>JF735116</td>
</tr>
<tr>
<td></td>
<td>2_JF735118</td>
<td>QC302</td>
<td>JF735118</td>
</tr>
<tr>
<td>Genotype 3</td>
<td>3_JF735124</td>
<td>QC115</td>
<td>JF735124</td>
</tr>
<tr>
<td>Genotype 4</td>
<td>4_JX227964</td>
<td>BID-G1253</td>
<td>JX227964</td>
</tr>
<tr>
<td></td>
<td>4_FJ025854</td>
<td>P026</td>
<td>FJ025854</td>
</tr>
<tr>
<td>Genotype 6</td>
<td>6_DQ278891</td>
<td>KM45,KM41</td>
<td>DQ278891,DQ278893</td>
</tr>
<tr>
<td></td>
<td>6_XI83550</td>
<td>QC273</td>
<td>XI83550</td>
</tr>
<tr>
<td></td>
<td>6_XI83552</td>
<td>TV476</td>
<td>XI83552</td>
</tr>
<tr>
<td></td>
<td>6_XI83549</td>
<td>KM35</td>
<td>XI83549</td>
</tr>
<tr>
<td></td>
<td>6_XI83551</td>
<td>TV257</td>
<td>XI83551</td>
</tr>
<tr>
<td></td>
<td>6_XI83553</td>
<td>TV533</td>
<td>XI83553</td>
</tr>
<tr>
<td></td>
<td>6_XI83554</td>
<td>L349</td>
<td>XI83554</td>
</tr>
<tr>
<td></td>
<td>6_XI83557</td>
<td>DH027</td>
<td>XI83557</td>
</tr>
<tr>
<td></td>
<td>6_XI83558</td>
<td>QC271</td>
<td>XI83558</td>
</tr>
</tbody>
</table>
*Classification of sequences into genotypes but without subtype assignments using the format “genotype_Accession number.”
†Locus (or isolate name if locus is the same as the accession number).
‡Previously described as 4b.14
§Previously described as 6k.17
utility, we therefore propose that the observed within-
genotype hierarchies are not given any formal recogni-
tion in their nomenclature.

Proposed Updates and Changes to Rules
for Genotype/Subtype Assignments

Subtype Names. By definition, subtype name
assignments would be limited to a maximum of 26 if
designated by a single letter suffix (e.g., 2a-2z). We
therefore suggest that subtypes are assigned up to the
letter “w” and subsequent designations follow the
eXtended form xa, xb, … xz, in turn followed by ya, …
yz, za, … zz, potentially giving a total of 101 subtypes
of each genotype. This avoids potentially ambiguous
terms such as “subtype 6x,” which could be interpreted
as “genotype 6 of unknown subtype,” or designations
such as “subtype 3aa,” which might suggest a relation-
ship with 3a.

Provisional Genotypes. According to the 2005
consensus classification protocol2 new genotypes could
be provisionally assigned from a single complete cod-
ing region sequence, but partial or complete coding
region sequences from additional isolates would be
required to confirm these assignments. Since then only
one provisional genotype has been identified (7a) re-
presented by a single isolate (QC69, EF108306). Thus,
in contrast to subtype assignments, the number of
genotypes appears relatively limited and the require-
ment to sequence multiple isolates now seems over-
onerous. We propose that only a single complete
coding region sequence is needed to confirm a new
genotype assignment; QC69 is therefore confirmed as
genotype 7a.

Provisional Subtypes. The 2005 consensus proto-
col also proposed that provisional subtypes could be
assigned on the basis of sequence comparisons in the
core/E1 and NS5B regions for at least three independ-
ent isolates, requiring in addition a complete coding
region sequence before being confirmed. Of the 58
subtypes provisionally assigned in the 2005 article, 38
have now been confirmed (Table 1). However, it is
now much easier to obtain complete coding region
sequences and very few additional provisional subtypes
have been proposed. Instead, some authors have inco-
sistently labeled unusual isolates with the suffix “?,”
“unassigned group I”11,23 or “subtype 1(I).”9 We pro-
pose that provisional subtype designations should no
longer be provided for variants where complete
genoome sequences are lacking. The 20 remaining
 provisionally assigned subtypes will be maintained (Table
3), since they already exist in the literature. Future
subtype assignments will only be made (as confirmed
assignments) when sequence data from three or more
isolates including at least one complete or nearly com-
plete coding region is provided. Where a complete
coding region sequence is available but there are fewer
than three isolates, we propose that these remain unas-
signed. In Table 2 these are labeled using the form
“Genotype_Accession number,” e.g., 1_AJ851228.

Recombinant and Other Forms. One issue that
was not addressed in the 2005 consensus protocol2
was the naming of the newly discovered recombinant
forms of HCV, their importance being unknown. Nine
different recombinant forms of HCV have now
been described (Table 4), of which only one (2k/1b) is
represented by multiple isolates; no multiple recombi-
nants have been reported (reviewed in reference24). In
this context it does not seem necessary to revise the
nomenclature generally used in the literature in which
“RF” (recombinant form) is followed by the contribu-
tory subtypes separated by “/” in the order in which
they appear in the complete genome sequence. We
suggest that recombinant forms with the same geno-
typic structure but with different breakpoints or where
the component genomic sections are unrelated are
numbered consecutively with a numerical suffix (for
example, RF2b/1b_1).

Proposals for New Genotype/Subtype Assign-
ments. The ICTV Flaviviridae Study Group is willing
to take a coordinating role in the assignment of newly
described variants of HCV. We urge researchers who
have characterized new HCV variants that potentially
qualify as new types or subtypes to contact Donald
Smith (D.B.Smith@ed.ac.uk) or any member of the
Study group (listed on http://ictvonline.org/subcomm
ittee.asp?committee=25&sec=5) in confidence before
publication so that naming conflicts can be avoided
and appropriate assignments made.

Future Developments

Despite the increasing number and diversity of
HCV sequences, the system of classification of variants
into genotypes and subtypes has proven surprisingly
robust. The seven confirmed genotypes have strong
bootstrap support (Fig. 1), and the partition of these
genotypes into subtypes that differ over a complete
coding region sequence by >15% reflects a natural
hiatus in the distribution of sequence distances
(Fig. 2). We welcome any comments or suggestions for
the proposed classification guidelines. Areas of uncer-
tainty remain with respect to the region of endemicity
of genotype 5, represented by a single subtype
isolated in Europe, Brazil, North Africa, and South Africa, and genotype 7, isolated from an emigrant from the Congo. We might also anticipate the further discovery of other HCV-like viruses in the genus Hepacivirus, and variants closer genetically to HCV than the nonprimate hepacivirus that appears to be an endemic infection of horses worldwide. As more is learned about the host-specificity and diversity of hepativiruses, the genotype classification of HCV may be logically incorporated within a unified classification of hepativiruses at the species and potentially subspecies and subgenus levels.

Acknowledgment: We thank Professor Ling Lu of the University of Kansas Medical Centre for providing several genotype 1 sequences prior to their release on GenBank.

Note added in proofs: A recent paper by Jordier et al. (J. Med Virol 2013;85:1754-1764) describes sequences that allow the confirmation of subtypes 2l (KC197235, KC197240) and 2t (KC197238) and identifies three unassigned subtypes of genotype 2 (KC197236, KC197237, KC197239).

References


16. Li C, Lu L, Wu X, Wang C, Bennett P, Lu T, et al. Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. J Gen Virol 2009;90:1820-1826.


confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain. J Med Virol 2011;83:820-826.


73. Morice Y, Roulot D, Grando V, Stirmann J, Gaulth E, Jeantils V, et al. 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 2001;82:1001-1012.


