Hepatitis C Virus Genotype 7, a New Genotype Originating from Central Africa

Donald G. Murphy,[a] Erwin Sablon,[b*] Jasmine Chamberland,[a] Eric Fournier,[a] Raymond Dandavino,[c] Cécile L. Tremblay[a]

Laboratoire de Santé Publique du Québec, Institut National de Santé Publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada[a]; Innogenetics NV, Ghent, Belgium[b]; Département de Néphrologie, Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada[c]

We report a new hepatitis C virus (HCV) genotype identified in patients originating from the Democratic Republic of Congo. The prototype QC69 virus is shown to be a new lineage distinct from genotypes 1 to 6. Three additional patients were also found to be infected by a virus from this lineage, confirming its circulation in humans. We propose that these viruses be classified into HCV genotype 7.

Hepatitis C virus (HCV) has emerged as a major cause of chronic liver disease throughout the world (1). Globally, 3% of the world's population is estimated to be infected with this blood-borne pathogen, and a vaccine is not presently available. The introduction of antibody and nucleic acid screening has more or less eliminated the risk of HCV infections by the transfusion of blood components. In industrialized countries, transmission occurs mainly between intravenous drug users. Other known modes of HCV transmission include perinatal transmission and invasive medical procedures. The long-term presence of HCV in certain areas of Africa and Asia, however, suggests alternative modes of transmission, presumably through exposure to blood.

HCV is in the genus *Hepacivirus* of the family *Flaviviridae*. The HCV genome consists of a single strand of positive-sense RNA that encodes a single long polyprotein arranged in the following gene order: 5′-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3′ (2). Since the identification of HCV in the late 1980s, six major genotypes have been recognized. Genotypes 1, 2, 3, 4, and 6 are further subdivided into a series of subtypes. The complete genomes of genotypes differ from each other by ≥30% at the nucleotide level, while those of subtypes within a given genotype differ typically by 15% to 25% (3, 4). HCV genotypes have a worldwide distribution and, except for genotype 5, display high genetic diversity.

Here, we report the finding of a new genotype of HCV identified in four subjects originating from the Democratic Republic of Congo (DRC), which confirms the circulation of this new lineage in the human population. The complete genome sequence of one isolate (QC69) as well as the partial genome sequences of three other isolates (QC272, IG93306, and IG93305) were obtained. We

show that these four isolates cluster together and are genetically distinct from genotypes 1 to 6. These findings fulfill current criteria for the designation of a new genotype of HCV (4). We propose that these viruses be classified as HCV genotype 7 and that the prototype sequence QC69 be classified as subtype 7a (4). This is the first published report of a new major genotype of HCV in 20 years.

The first partial genome sequences of QC69 were obtained in 2002 from a 48-year-old male patient (5). QC272 was identified in 2005 in a 15-year-old female patient. Patients infected with QC69 and QC272 were previously found to be anti-HCV positive by AxSYM HCV version 3.0 (Abbott Laboratories, Abbott Park, IL), Monolisa Anti-HCV Plus version 2 (Bio-Rad, Marnes-la-Coquette, France), and Ortho HCV 3.0 ELISA Test System with Enhanced SAVe (Ortho Clinical Diagnostics, Inc., Raritan, NJ) and HCV RNA positive by Amplicor Hepatitis C Virus (HCV) Test,
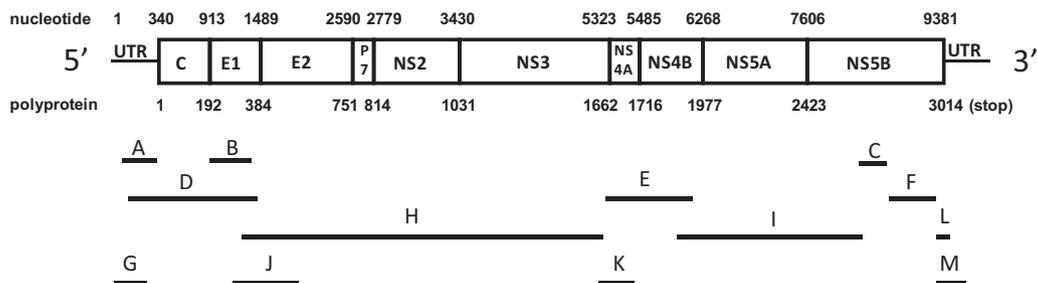
FIG 1 Genomic representation and organization of isolate QC69. The positions of 13 PCR fragments used to determine the full-length sequence are indicated.
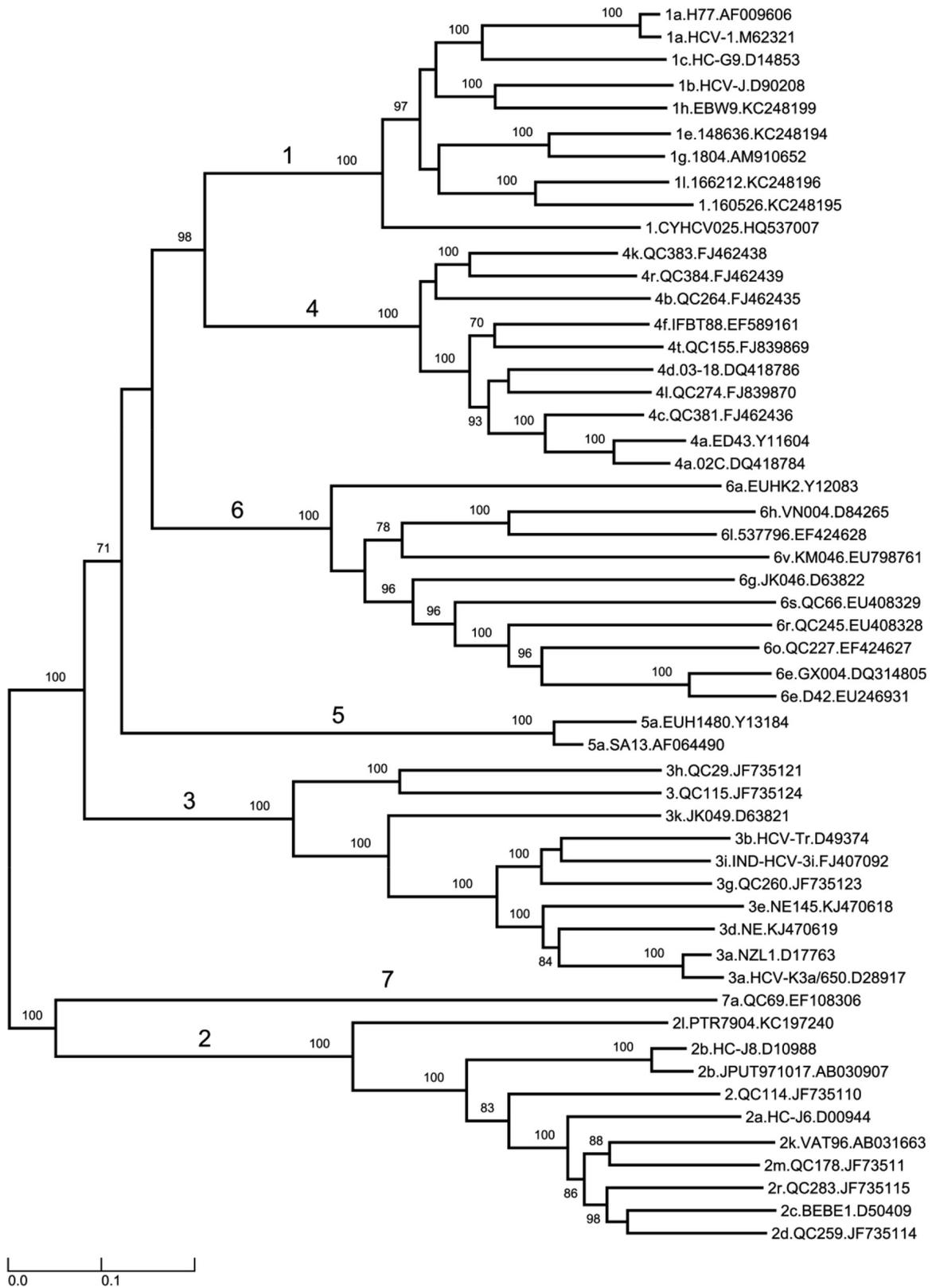
FIG 2 Maximum-likelihood phylogeny of nucleotide sequences for the complete coding region of representatives of the seven genotypes of HCV. Genotypes 1 to 4 and 6 are each represented by 10 genomes, including two of the same subtype. The major branches corresponding to each genotype are indicated within the tree. The numbers adjacent to the branch nodes represent the percent bootstrap robustness for 100 replicates; values of ≥70% are shown. Sequences are labeled to the right of each branch in order as follows: genotype/subtype, isolate identification, and GenBank accession number. Bar, nucleotide substitutions per site.

version 2.0 (Roche Diagnostics, Branchburg, NJ). IG93306 and IG93305 were obtained from patients residing in Belgium in the early 2000s. These patients had been found anti-HCV positive by the Innotest HCV Ab IV (Innogenetics, Ghent, Belgium). No close epidemiological link was established between the two isolates identified in Canada (QC69 and QC272) or Belgium (IG93306 and IG93305). The route of infection has not been confirmed for any of the four patients, but injection drug use was not suspected.

The nearly complete viral genome sequence of QC69 was derived from a serum sample drawn in 2003 that displayed an HCV RNA level of 7,210,000 IU/ml in the Cobas Amplicor HCV Monitor test version 2.0 (Roche Diagnostics Systems, Inc.). The sequence was derived from 13 overlapping DNA fragments obtained by reverse transcription (RT)-PCR (Fig. 1). The genuine 5′ terminus (fragment G; Fig. 1) was determined by use of a 2nd-generation 5′/3′ rapid amplification of cDNA ends (RACE) kit (Roche Applied Science, Mannheim, Germany). The remaining DNA fragments were obtained using a single-round PCR, except for fragment H, which was obtained using a nested-PCR approach. Viral RNA was extracted from either 140 µl of serum using the QIAamp viral RNA kit or 263 µl of serum using the QIAamp Virus BioRobot MDx kit (Qiagen, Inc., Mississauga, Ontario, Canada). Reverse transcription and DNA amplification by PCR for fragments A, B, C, D, E, F, J, K, and L were performed as previously described (5). For fragments H, I, and M, RT-PCR was performed using the SuperScript III One-Step RT-PCR system with Platinum DNA polymerase (Invitrogen, Burlington, Ontario, Canada). For fragment H, a second-round PCR was performed using the Expand high-fidelity PCR system (Roche Diagnostics). PCR products were purified and sequenced bidirectionally. Sequences were analyzed on an ABI Prism 3100xl genetic analyzer (Applied Biosystems, Foster City, CA).

The assembled genome sequence of QC69 consisted of 9,418 nucleotides excluding its poly(U) stretch and 3′ X tail. The 5′ untranslated region (UTR) consisted of 339 nucleotides, while the 3′ UTR variable region consisted of 40 nucleotides. QC69 displayed an open reading frame (ORF) of 3,013 amino acids. The amino acid lengths of the predicted cleavage proteins are 191 (Core), 192 (E1), 367 (E2), 63 (P7), 217 (NS2), 631 (NS3), 54 (NS4A), 261 (NS4B), 446 (NS5A), and 591 (NS5B). The genome length and organization of QC69 were similar to those of other HCV isolates.

The coding region sequence of QC69 was aligned with selected nonrecombinant full-length coding region sequences representative of genotypes 1 to 6. The evolutionary history was inferred using the maximum-likelihood method based on the general time-reversible model (6). The initial tree(s) for the heuristic search was obtained by applying the neighbor-joining and BIONJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood approach (7). A discrete gamma distribution (four categories) was used to model evolutionary rate differences among the sites. Evolutionary analyses and calculation of p distances using the pairwise deletion option were conducted in MEGA5 (8). The phylogenetic tree showed that QC69 clustered separately from sequences representative of genotypes 1 to 6 (Fig. 2). In the complete coding region sequence, QC69 showed 33.4% to 34.9% mean p-nucleotide distances and 28.3% to 29.8% mean p-amino acid distances, with the representative isolates of genotypes 1 to 6 appearing in Fig. 2 (Table 1). This level of divergence

TABLE 1 Mean nucleotide and amino acid pairwise distances between genotypes for the complete coding region

Genotype	Mean pairwise distance[a] (%) for genotype:						
	1	2	3	4	5	6	7
1		33.0[b]	32.1[b]	29.1[b]	30.7[b]	31.2[b]	33.4[b]
2	27.9[c]		34.9[b]	33.4[b]	33.8[b]	34.1[b]	34.8[b]
3	25.4[c]	29.4[c]		32.0[b]	32.5[b]	32.8[b]	34.9[b]
4	21.5[c]	28.3[c]	25.2[c]		30.7[b]	31.1[b]	34.1[b]
5	24.2[c]	29.1[c]	26.3[c]	24.6[c]		31.8[b]	34.7[b]
6	24.1[c]	28.7[c]	25.9[c]	23.8[c]	25.1[c]		34.7[b]
7	28.3[c]	29.1[c]	29.2[c]	28.8	29.8[c]	29.1[c]	

[a] Mean distances between genotypes were calculated using the sequences described in Fig. 1.
[b] Nucleotide.
[c] Amino acid.

was similar to that observed between the other genotypes. QC69 was assessed for the presence of recombination by similarity and bootscanning plots analyses using SimPlot 3.5.1 software (9). No evidence of recombination was found. The classification of QC69 as genotype 7a has been recognized by the HCV databases and supported by others (see http://hcv.lanl.gov and http://euhcvdb.ibcp.fr) (4, 10–12).

The coding region sequence of QC69 was made publically available on GenBank in 2007. Gottwein et al. (13) used this sequence to construct a QC69/JFH1(7a/2a) intergenotypic recombinant-containing core, E1, E2, P7, and NS2 gene sequences of QC69. It was found to successfully replicate in Huh7.5 cells, and its viability did not depend on adaptive mutations. More recently, this group of researchers developed J6/JFH1 recombinants expressing either the NS4A or NS5A proteins in Huh7.5 cells (14, 15). When exposed to several protease inhibitors, interferon alfa-2b, and a putative NS4A inhibitor, the genotype 7a-NS4A expressing recombinant responded in a manner similar to that of the genotype 1 to 6 NS4A-expressing recombinants (14). For the NS5A-expressing recombinants, mutations in the low-complexity sequence II region led to a high reduction in viral production for genotype 4a and 7a recombinants (15). These studies confirm the biological properties of genotype 7a virus proteins.

The NS5B partial genome sequences of QC272, IG93306, and IG93305 were obtained as previously described (5, 16). Phylogenetic analyses showed that QC69, QC272, IG93306, and IG93305 clustered together but separately from representatives of genotypes 1 to 6 (Fig. 3). QC272 was found to be more distantly related to the other genotype 7 isolates, but this evolutionary distance was less than that observed between various subtypes of genotype 6. For example, the length of the branches between genotype 7 isolates QC272 and IG93305 was 0.753 substitutions/site, whereas the branch length between genotype 6a EUHK2 and 6e D42 viruses was 0.775 substitutions/site (Fig. 3). The inclusion of QC272 within genotype 7 was also supported by phylogenetic analysis of C/E1 sequences (our unpublished data). The NS5B sequence data indicate that QC272, IG93306, and IG93305 would each belong to a distinct subtype different from subtype 7a. This considerable level of genetic diversity suggests that genotype 7 has been propagated and maintained for a long time in the DRC. Interestingly, in a recent survey of HCV genotypes in the DRC, Iles et al. (17) did not detect isolates belonging to genotype 7. However, genotype 7 may have remained undetected as the NS5B sequence data used for accessing the genotypes were not obtained for 20% of the sam-
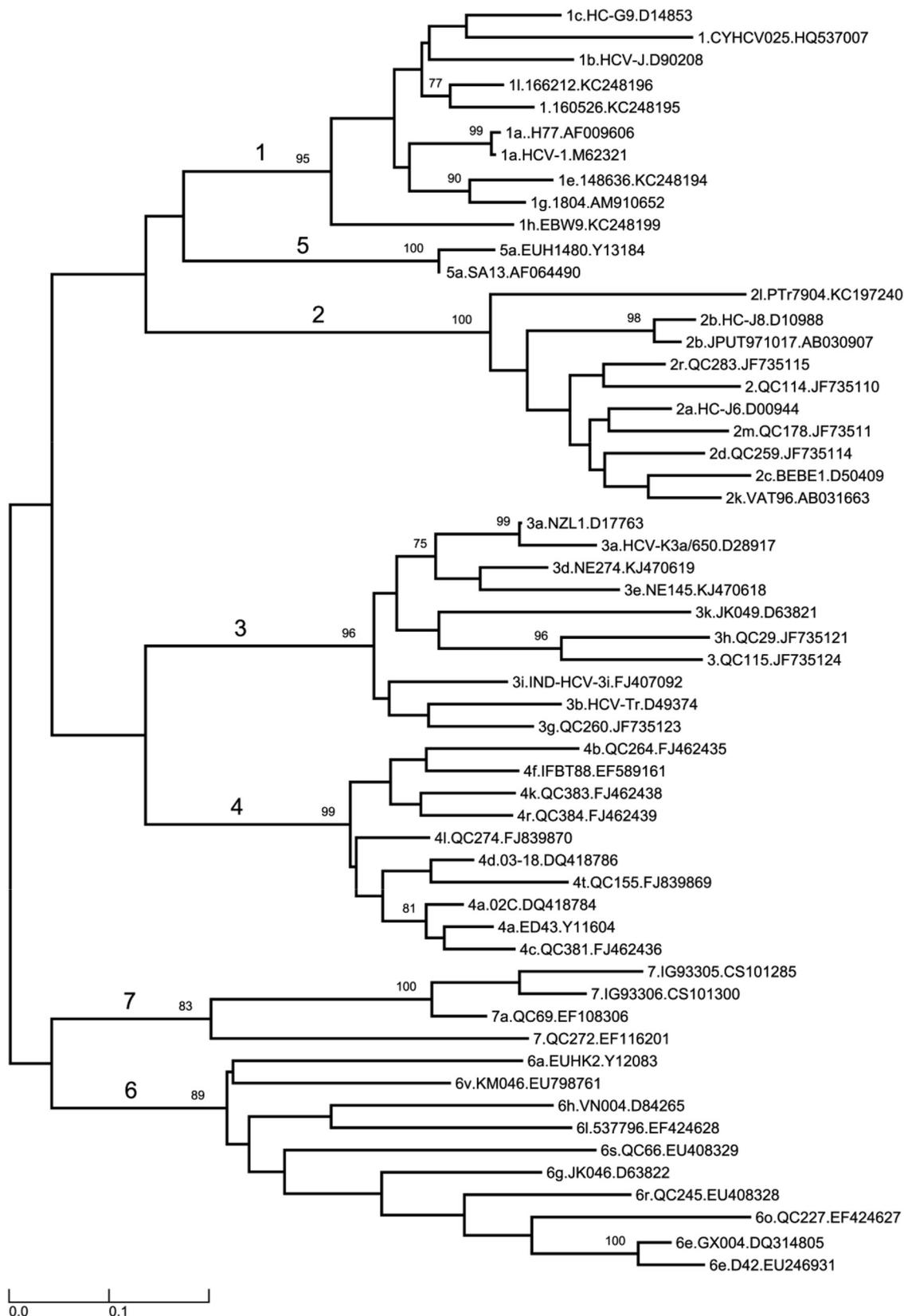
FIG 3 Maximum-likelihood phylogeny of NS5B sequences corresponding to positions 8276 to 8615 (340 bp) of representatives of the seven genotypes of HCV. The phylogeny was reconstructed using NS5B sequences of genomes included in Fig. 1 and of isolates QC272, IG93306, and IG93305. The major branches corresponding to each genotype are indicated within the tree. The numbers adjacent to the branch nodes represent the percent bootstrap robustness for 1,000 replicates; values of ≥70% are shown. Sequences are labeled to the right of each branch in order as follows: genotype/subtype, isolate identification, and GenBank accession number. Bar, nucleotide substitutions per site.

ples that had detectable HCV RNA in the 5′ UTR (17). This indicates that although genotype 7 may well have its origin in the DRC, it is not highly prevalent in this country or is delimited to certain regions. In Quebec, Canada, genotypes have now been determined in more than 17,000 persons by sequencing analysis of the NS5B region (5). All seven HCV genotypes, numerous subtypes, and various unclassified variants have been found in either the local or immigrant population. QC69 and QC272 were the only genotype 7 isolates identified, suggesting that this genotype has not spread significantly.

Phylogenetic analysis performed on full-length coding region sequences showed that QC69 branched more closely to genotype 2 variants (Fig. 2). The greatest diversity of genotype 2 is observed in western Africa. It has been proposed that HCV genotype 2 originated from West Africa with subsequent movement to the east (18, 19). Since genotype 7 isolates have been found in patients originating from central Africa, the phylogeographic relationship between genotypes 7 and 2 remains to be elucidated. Further epidemiological studies and sequencing of coding regions are required to determine the prevalence and spread of genotype 7 in Africa. The finding of a new genotype further indicates the endemic nature of HCV in certain parts of Africa and shows that HCV infection may have evolved from a common ancestor originating from this region of the world.

Genotypes are not only crucial for our understanding of the epidemiology and evolution of HCV; they are clinically significant as current treatment regimens are tailored to the infecting genotype (20). Thus, inaccurate identification of the infecting genotype may expose patients to suboptimal treatment regimens. Commercial genotyping assays that target the 5′ UTR are widely used by clinical laboratories for assessing the HCV genotype (21). These assays have been developed for the identification of genotypes 1 to 6. Consequently, genotype 7 isolates would be either not attributed to a genotype or inaccurately classified by these assays. We tested QC69 and QC272 using the Versant HCV amplification 2.0 kit and the Versant HCV Genotype 2.0 assay (LiPA) (Siemens Healthcare Diagnostics, Erlangen, Germany). QC69 reacted with lines 6, 11, and 20, while QC272 reacted with lines 5, 9, 11, and 20, as was expected from their derived 5′ UTR nucleotide sequences (22). Thus, QC272 would have been incorrectly identified as genotype 2 and QC69 would not have been classified since the observed line pattern was not attributed to a genotype. IG93305 and IG93306 also displayed atypical line patterns on LiPA strips that were not assigned to a genotype (16). In clinical practice and for epidemiological studies, correct identification of all seven HCV genotypes is relevant and can be achieved only by the sequencing of a coding region able to discriminate genotypes and subtypes, such as C/E1 or NS5B.

Nucleotide sequence accession numbers. The newly described sequences in this study were deposited in GenBank under the accession numbers EF108306 (QC69), CS101285 (IG93305), CS101300 (IG93306), CS101293 (IG93305), and CS101294 (IG93306).

ACKNOWLEDGMENTS

REFERENCES

1. **Shepard CW, Finelli L, Alter MJ.** 2005. Global epidemiology of hepatitis C infection. Lancet Infect Dis **5:**558–567. http://dx.doi.org/10.1016/S1473-3099(05)70216-4.
2. **Lindenbach BD, Rice CM.** 2005. Unravelling hepatitis C virus replication from genome to function. Nature **436:**933–938. http://dx.doi.org/10.1038/nature04077.
3. **Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe 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. http://dx.doi.org/10.1002/hep.20819.
4. **Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapelton JT, Simmonds P.** 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology **59:**318–327. http://dx.doi.org/10.1002/hep.26744.
5. **Murphy DG, Willems B, Deschênes M, Hilzenrat N, Mousseau R, Sabbah S.** 2007. Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5′ untranslated region sequences. J Clin Microbiol **45:**1102–1112. http://dx.doi.org/10.1128/JCM.02366-06.
6. **Tavare S.** 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci **17:**57–86.
7. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406–425.
8. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
9. **Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol **73:**152–160.
10. **Kuiken C, Yusim K, Boykin L, Richardson R.** 2005. The Los Alamos hepatitis C sequence database. Bioinformatics **21:**379–384. http://dx.doi.org/10.1093/bioinformatics/bth485.
11. **Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Le Mercier P, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky JM, Rice CM, Trépo C, Penin F, Deléage G.** 2007. euHCVdb: the European hepatitis C virus database. Nucleic Acids Res **35:**D363–D366. http://dx.doi.org/10.1093/nar/gkl970.
12. **Nakano T, Lau GMG, Lau GML, Sugiyama M, Mizokami M.** 2012. An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. Liver Int **32:**339–345. http://dx.doi.org/10.1111/j.1478-3231.2011.02684.x.
13. **Gottwein JM, Scheel TK, Jensen TB, Lademann JB, Prentoe JC, Knudsen ML, Hoegh AM, Bukh J.** 2009. Development and characterization of hepatitis C virus genotype 1-7 cell culture systems: role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. Hepatology **49:**364–377. http://dx.doi.org/10.1002/hep.22673.
14. **Gottwein JM, Jensen SB, Serre SB, Ghanem L, Scheel TK, Jensen TB, Krarup H, Uzcategui N, Mikkelsen LS, Bukh J.** 2013. Adapted J6/JFH1-based hepatitis C virus recombinants with genotype-specific NS4A show similar efficacies against lead protease inhibitors, alpha interferon, and a putative NS4A inhibitor. Antimicrob Agents Chemother **57:**6034–6049. http://dx.doi.org/10.1128/AAC.01176-13.
15. **Scheel TK, Prentoe J, Carlsen TH, Mikkelsen LS, Gottwein JM, Bukh J.** 2012. Analysis of functional differences between hepatitis C virus NS5A of genotypes 1-7 in infectious cell culture systems. PLoS Pathog **8:**e1002696. http://dx.doi.org/10.1371/journal.ppat.1002696.
16. **Sablon E, Quint W, van Doorn LJ.** June 2005. New hepatitis C virus clade and phototype sequences thereof. European patent 1536024 A1.
17. **Iles JC, Raghwani J, Harrison GL, Pepin J, Djoko CF, Tamoufe U, LeBreton M, Schneider BS, Fair JN, Tshala FM, Kayembe PK, Muyembe JJ, Edidi-Basepeo S, Wolfe ND, Simmonds P, Klenerman P, Pybus OG.** 2014. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology **464–465:**233–243. http://dx.doi.org/10.1016/j.virol.2014.07.006.
18. **Markov PV, Pepin J, Frost E, Deslandes S, Labbe AC, Pybus OG.** 2009. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. J Gen Virol **90:**2086–2096. http://dx.doi.org/10.1099/vir.0.011569-0.

19. **Njouom R, Nerrienet E, Dubois M, Lachenal G, Rousset D, Vessière A, Ayouba A, Pasquier C, Pouillot R.** 2007. The hepatitis C virus epidemic in Cameroon: genetic evidence for rapid transmission between 1920 and 1960. Infect Genet Evol **7:**361–367. http://dx.doi.org/10.1016/j.meegid.2006.10.003.

20. **American Association for the Study of Liver Diseases, Infectious Diseases Society of America.** 2014. Recommendations for testing, managing, and treating hepatitis C. http://www.hcvguidelines.org.

21. **Bouchardeau F, Cantaloube JF, Chevaliez S, Portal C, Razer A, Lefrere JJ, Pawlotsky JM, De Micco P, Laperche S.** 2007. Improvement of hepatitis C virus (HCV) genotype determination with the new version of the Inno-LiPA HCV assay. J Clin Microbiol **45:**1140–1145. http://dx.doi.org/10.1128/JCM.01982-06.

22. **Stuyver L, Wyseur A, van Arnhem W, Hernandez F, Maertens G.** 1996. Second-generation line probe assay for hepatitis C virus genotyping. J Clin Microbiol **34:**2259–2266.