Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon

Jean Ndjomou,1 Oliver G. Pybus2 and Bertfried Matz1

1Institute of Medical Microbiology and Immunology, University of Bonn, Sigmund-Freud Strasse 25, 53105 Bonn, Germany
2Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Hepatitis C virus (HCV) is a blood-borne pathogen that poses a significant threat to public health worldwide. The genetic diversity and distribution of HCV genotypes in non-Western countries, particularly sub-Saharan Africa, is poorly documented. This study reports a phylogenetic analysis of core and NS5B gene sequences of 37 HCV strains sampled in Cameroon. A high level of genetic diversity of both genotypes 1 and 4 was found, indicating a unique pattern of long-term HCV infection that has not been observed elsewhere. These results lead to the hypothesis that these HCV genotypes originated and diversified in west Central Africa before spreading to other regions.

INTRODUCTION

Hepatitis C virus (HCV) is a leading global cause of liver disease. It is estimated that more than 170 million people worldwide are chronically infected with HCV and are at risk of developing liver cirrhosis and liver cancer (WHO, 1997).

HCV is a single-stranded, positive-sense RNA virus, with a genome of 9-7 kb long containing a single ORF. The ORF encodes three structural proteins (including the capsid protein, core) and six non-structural proteins (including the RNA polymerase, NS5B). The ORF is flanked by two NCRs located at the 5’ and 3’ ends (Choo et al., 1991).

HCV is genetically diverse and is classified into six major genotypes, each of which is divided further into several subtypes (Robertson et al., 1998). HCV genotypes differ at more than 30% of nucleotides across the entire genome, while HCV subtypes vary at more than 20% of sites (Simmonds, 1995). HCV genotypes show different geographical distributions and levels of diversity, reflecting their different epidemic histories and routes of transmission. HCV isolates from Western countries typically have limited sequence diversity, resulting from the recent introduction of a few strains (such as subtypes 1a, 1b and 3a) from endemic areas. These strains have spread rapidly through infected blood products and intravenous drug use (Pybus et al., 2001). The distribution of HCV genotypes in non-Western countries is less well documented, although small-scale surveys have found considerable sequence diversity of genotypes 1, 2 and 4 in African countries (Xu et al., 1994; Fretz et al., 1995; Ruggieri et al., 1996; Wansbrough-Jones et al., 1998; Jeannel et al., 1998), suggesting that these strains have been endemic to the continent for several hundred years (Smith et al., 1997; Pybus et al., 2001). The investigation of HCV genetic diversity in sub-Saharan Africa is, therefore, necessary to provide insights into the global epidemiology, epidemic history and origin of HCV. Furthermore, availability of more sequence data from sub-Saharan Africa may help to refine the HCV classification system.

In this work we present a phylogenetic analysis of core and NS5B gene sequences obtained from 37 HCV strains isolated in Cameroon. We show that HCV genotypes 1 and 4 are both highly divergent in Cameroon, indicating the long-term presence of both strains in west Central Africa. This pattern of HCV endemicity is not observed in neighbouring regions: divergent genotype 1 strains are found in West Africa, whereas genotype 4 appears endemic to Central Africa. We postulate that the common ancestor of genotypes 1 and 4 originated in the area of modern Cameroon, spread subsequently to other regions of Africa and then to the rest of the world.

METHODS

Patients and samples. A total of 482 blood samples was obtained in 1998 from sexually transmitted disease clinic attendees, infectious disease clinic attendees and patients clinically suspected for having
human immunodeficiency virus (HIV) infection, from a general clinic in the Littoral Province of Cameroon. Of these patients, 52% were male (age range, 11–90 years; median age, 33 years) and 48% were female (age range, 9–64 years; median age, 28 years). After giving informed consent, each participant was interviewed and socio-demographic data were recorded. Samples of venous blood (10 ml) were collected in EDTA tubes and plasma samples were obtained by centrifugation. Samples were stored in two aliquots at −20°C.

HCV antibodies and RNA detection. Detection of HCV antibodies and RNA has been described previously (Ndjomou et al., 2002). Briefly, antibodies were screened by a microparticle enzyme immunoassay (AxSYM HCV3.0; ABBOT Diagnostics) and reactive samples were confirmed by a strip immunoblot assay (RIBA HCV3.0; Chiron Corporation). All samples were screened for the presence of HCV RNA by nested RT-PCR with primers located in the highly conserved 5′ NCR.

RNA preparation, PCR amplification, cloning and sequencing. Viral RNA was isolated from 140 μl plasma using the QIAamp Viral RNA Mini kit (Qiagen), according to the manufacturer’s protocol, and eluted with 60 μl elution buffer in 0-2 ml tubes containing 3 mM DTT and 40 U RNasin (Promega). RNA was reverse-transcribed using the Expand Reverse Transcriptase Enzyme system (Roche) and PCR was carried out using the Expand High Fidelity system (Roche) and PCR was carried out using the Expand High Fidelity PCR system (Roche), as described below. All PCR products were checked on 2% agarose gels prior to cloning.

(a) Core region PCR. Some samples were amplified in a single round and others in a semi-nested PCR using the primers P874A, P417S and P439S, described previously by Viazov et al. (1997), generating an approximately 450 bp fragment. RNA was reverse-transcribed at 42°C for 30 min in a 20 μl reaction volume containing 10 pmol outer antisense primer P874A, 0-625 mM each dNTP, 5 mM MgCl2, 1× Expand RT buffer and 25 U of Expand RT enzyme. First-round PCR was carried out in a 50 μl reaction mixture containing 10 pmol each outer primer P874A and P417S, 0-25 mM each dNTP, 1-6 mM MgCl2, 1× Expand High Fidelity (HF) buffer and 1-75 U Expand HF enzyme. Second-round PCR used 5 μl of the first-round PCR product in a final volume of 50 μl containing 10 pmol each primer P874A and P439S, 0-2 mM each dNTP, 1 mM MgCl2, 1× Expand HF buffer and 1-75 U Expand HF enzyme. Cycling parameters for both rounds were as follows: denaturation at 94°C for 2 min, followed by 4 cycles at 94°C for 15 s, 75–55°C for 15 s, with an auto-decrement temperature of 5°C for each cycle, and 72°C for 35 s; then 32 cycles at 94°C for 10 s, 58°C for 10 s and 72°C for 35 s, with a final extension for 7 min at 72°C.

(b) NS5B region PCR. cDNA was generated using 30 pmol random hexamer primer, 5 mM MgCl2, 1 mM each dNTP, 1× Expand RT buffer and 35 U Expand RT enzyme and the reaction was incubated for 10 min at 30°C, 42°C for 45 min and 95°C for 5 min. All samples were amplified in a single round of PCR with the primers NS1204 and NS1203, described previously by Mellor et al. (1995), producing an approximately 405 bp fragment. The reaction volume of 50 μl contained 0-25 mM each dNTP, 1-6 mM MgCl2, 5 pmol each primer NS1204 and NS1203, 1× Expand HF buffer and 1-75 U Expand HF enzyme. Cycling conditions were as follows: 94°C for 2 min, followed by 6 cycles at 94°C for 10 s, 80–55°C for 10 s, with an auto-decrement temperature of 5°C each cycle, and 72°C for 15 s; then 30 cycles at 94°C for 10 s, 55°C for 10 s, 72°C for 15 s, with a final extension at 72°C for 7 min.

PCR-generated amplicons were cloned directly into the pCR4-TOPO plasmid vector (Invitrogen) and used to transform chemically competent TOP10 Escherichia coli (Invitrogen). Positive clones were detected by EcoRI digestion or HCV-specific PCR. For each isolate, two to three clones were sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) and sequencing products were purified with the DyeEx Spin kit (Qiagen). Electrophoresis and data collection were done on an automated ABI PRISM 310 genetic analyser (Applied Biosystems). Consensus nucleotide sequences derived from the two or three sequenced clones of each isolate were used for phylogenetic analysis.

Phylogenetic analysis. The core and NS5B sequences obtained (35 NS5B sequences and 37 core sequences) were added to all other HCV sequences available for the same gene regions (428 NS5B sequences and 391 core sequences; Hepatitis Virus Database at http://s2as02.genes.nig.ac.jp/). Sequences were aligned using Se-Al v2.0 (http://evolve.zoo.ox.ac.uk) and edited by hand. Six alignments were produced, one for each combination of gene (core and NS5B) and genotype (1, 2 and 4). In each alignment, a set of about 35 sequences, representing all six genotypes and many subtypes, was retained for use as an outgroup. Many sequences from the common subtypes 1a and 1b were removed to improve computational efficiency. Molecular phylogenies were estimated for each alignment using a maximum-likelihood (ML) approach implemented in PAUP* v4.0b (Swofford, 2002). A general reversible model of nucleotide substitution was used, with gamma-distributed rate heterogeneity amongst sites (model GTR+G). Substitution model parameters were estimated on an initial neighbour-joining tree and were fixed throughout heuristic optimization, during which all three branch-swapping algorithms (NNI, TBR and SPR) were used. Likelihood and parameter values are available on request. Phylogenies were rooted using the outgroup sequences and measures of statistical support for each node in the ML tree were obtained using 500 bootstrap replicates of the original nucleotide alignment (Felsenstein, 1985).

RESULTS

Previously, Ndjomou et al. (2002) reported that HCV antibodies were detected in 54 (11%) of 482 blood samples from individuals at risk for HIV infection in Cameroon. A total of 37 samples were identified as viraemic based on RT-PCR amplification of the 5′ NCR, and sequence analysis of this region indicated that the samples belonged to HCV genotypes 1, 2 and 4 (Ndjomou et al., 2002). Here, the core and NS5B genes of these samples were sequenced and analysed, as they are more phylogenetically informative than the 5′ NCR. Furthermore, it is recommended that at least two coding regions are used to investigate the phylogenetic relationship between HCV strains (Robertson et al., 1998).

Table 1 summarizes the epidemiological data for the isolates and shows the results of HCV genotyping based on the core and NS5B regions. All 37 strains were amplified and analysed in the core region, but in spite of repeated attempts, two strains (CM9897 and CM1387) could not be amplified in the NS5B region. Phylogenetic analysis shows that the Cameroonian isolates all belong to HCV genotypes 1, 2 and 4. One strain (CM1457) was typed as genotype 1 in the core region but as genotype 2 in the NS5B region. Comparison of these results with those presented by Ndjomou et al. (2002) reveals that eight strains were mistyped previously when only the 5′ NCR sequences were analysed (seven isolates were typed as genotype 1 in the 5′ NCR but as genotype 4 in
core and NS5B; one isolate was typed as genotype 5 in the 5' NCR but as genotype 4 in core and NS5B). This mistyping probably results from poor phylogenetic resolution in the 5' NCR or from a lack of representative reference sequences that are close enough to the query sequences (Ohno et al., 1996).

Figs 1–3 show the estimated core and NS5B phylogenies for genotypes 1, 2 and 4, respectively. As illustrated in Fig. 1, Cameroonian genotype 1 strains are genetically very diverse. In the core phylogeny (Fig. 1a), they cluster mostly at the base of the tree and are distinct from the global subtypes 1a and 1b. However, one isolate (CM1459) clustered with subtype 1b probably represents an infection that was acquired abroad (this individual had travelled to France and Italy) or an infection acquired from imported blood products contaminated with HCV. One reference strain (L38350) from France grouped together with Cameroonian strains and may represent a French immigrant who was infected in Africa. The high diversity of Cameroonian type 1

Table 1. Epidemiological data and genotyping results

<table>
<thead>
<tr>
<th>Virus sample</th>
<th>Age</th>
<th>Sex</th>
<th>Group</th>
<th>Trip abroad?</th>
<th>Core Genotype</th>
<th>Accession no.</th>
<th>NS5B Genotype</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>98CM445</td>
<td>28</td>
<td>F</td>
<td>CSP</td>
<td>No</td>
<td>1</td>
<td>AY257796</td>
<td>1</td>
<td>AY257096</td>
</tr>
<tr>
<td>98CM454</td>
<td>49</td>
<td>M</td>
<td>CSP</td>
<td>No</td>
<td>1</td>
<td>AY257788</td>
<td>1</td>
<td>AY257077</td>
</tr>
<tr>
<td>98CM481</td>
<td>38</td>
<td>M</td>
<td>CSP</td>
<td>No</td>
<td>4</td>
<td>AY257818</td>
<td>4</td>
<td>AY257078</td>
</tr>
<tr>
<td>98CM1357</td>
<td>50</td>
<td>F</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257784</td>
<td>1</td>
<td>AY257069</td>
</tr>
<tr>
<td>98CM1363</td>
<td>55</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257811</td>
<td>4</td>
<td>AY257085</td>
</tr>
<tr>
<td>98CM1383</td>
<td>62</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257789</td>
<td>1</td>
<td>AY257083</td>
</tr>
<tr>
<td>98CM1387</td>
<td>72</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257819</td>
<td>NA</td>
<td>–</td>
</tr>
<tr>
<td>98CM1394</td>
<td>55</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257791</td>
<td>1</td>
<td>AY257090</td>
</tr>
<tr>
<td>98CM1402</td>
<td>44</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257793</td>
<td>1</td>
<td>AY257092</td>
</tr>
<tr>
<td>98CM1414</td>
<td>47</td>
<td>M</td>
<td>TBC</td>
<td>Yes*</td>
<td>2</td>
<td>AY257804</td>
<td>2</td>
<td>AY257088</td>
</tr>
<tr>
<td>98CM1418</td>
<td>80</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257807</td>
<td>4</td>
<td>AY257103</td>
</tr>
<tr>
<td>98CM1421</td>
<td>59</td>
<td>F</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257785</td>
<td>1</td>
<td>AY257071</td>
</tr>
<tr>
<td>98CM1423</td>
<td>32</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257814</td>
<td>4</td>
<td>AY257097</td>
</tr>
<tr>
<td>98CM1427</td>
<td>49</td>
<td>F</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257792</td>
<td>1</td>
<td>AY257091</td>
</tr>
<tr>
<td>98CM1431</td>
<td>55</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257815</td>
<td>4</td>
<td>AY257098</td>
</tr>
<tr>
<td>98CM1453</td>
<td>50</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>2</td>
<td>AY257802</td>
<td>2</td>
<td>AY257084</td>
</tr>
<tr>
<td>98CM1457</td>
<td>56</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>2</td>
<td>AY257803</td>
<td>1</td>
<td>AY257086</td>
</tr>
<tr>
<td>98CM1458</td>
<td>32</td>
<td>F</td>
<td>TBC</td>
<td>Yes†</td>
<td>4</td>
<td>AY257808</td>
<td>4</td>
<td>AY257072</td>
</tr>
<tr>
<td>98CM1459</td>
<td>29</td>
<td>M</td>
<td>TBC</td>
<td>Yes‡</td>
<td>1</td>
<td>AY257786</td>
<td>1</td>
<td>AY257073</td>
</tr>
<tr>
<td>98CM1464</td>
<td>53</td>
<td>M</td>
<td>TBC</td>
<td>Yes§</td>
<td>2</td>
<td>AY257805</td>
<td>2</td>
<td>AY257099</td>
</tr>
<tr>
<td>98CM1496</td>
<td>64</td>
<td>F</td>
<td>TBC</td>
<td>Yes††</td>
<td>2</td>
<td>AY257806</td>
<td>2</td>
<td>AY257100</td>
</tr>
<tr>
<td>98CM1512</td>
<td>42</td>
<td>M</td>
<td>TBC</td>
<td>Yes‡‡</td>
<td>4</td>
<td>AY257816</td>
<td>4</td>
<td>AY257101</td>
</tr>
<tr>
<td>98CM1521</td>
<td>44</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257790</td>
<td>1</td>
<td>AY257087</td>
</tr>
<tr>
<td>98CM1524</td>
<td>51</td>
<td>M</td>
<td>TBC</td>
<td>Yes#</td>
<td>1</td>
<td>AY257787</td>
<td>1</td>
<td>AY257074</td>
</tr>
<tr>
<td>98CM1536</td>
<td>76</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257797</td>
<td>1</td>
<td>AY257102</td>
</tr>
<tr>
<td>98CM1548</td>
<td>29</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257795</td>
<td>1</td>
<td>AY257095</td>
</tr>
<tr>
<td>98CM1565</td>
<td>40</td>
<td>F</td>
<td>TBC</td>
<td>No</td>
<td>1</td>
<td>AY257794</td>
<td>1</td>
<td>AY257094</td>
</tr>
<tr>
<td>98CM1573</td>
<td>52</td>
<td>M</td>
<td>TBC</td>
<td>Yes**</td>
<td>4</td>
<td>AY257817</td>
<td>4</td>
<td>AY257103</td>
</tr>
<tr>
<td>98CM1578</td>
<td>50</td>
<td>F</td>
<td>TBC</td>
<td>No</td>
<td>4</td>
<td>AY257809</td>
<td>4</td>
<td>AY257075</td>
</tr>
<tr>
<td>98CM1581</td>
<td>50</td>
<td>M</td>
<td>TBC</td>
<td>No</td>
<td>2</td>
<td>AY257798</td>
<td>2</td>
<td>AY257076</td>
</tr>
<tr>
<td>98CM9774</td>
<td>63</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>4</td>
<td>AY257812</td>
<td>4</td>
<td>AY257089</td>
</tr>
<tr>
<td>98CM9854</td>
<td>41</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>2</td>
<td>AY257799</td>
<td>2</td>
<td>AY257079</td>
</tr>
<tr>
<td>98CM9897</td>
<td>53</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>4</td>
<td>AY257820</td>
<td>NA</td>
<td>–</td>
</tr>
<tr>
<td>98CM9905</td>
<td>35</td>
<td>M</td>
<td>STD</td>
<td>Yes††</td>
<td>4</td>
<td>AY257813</td>
<td>4</td>
<td>AY257093</td>
</tr>
<tr>
<td>98CM9919</td>
<td>30</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>2</td>
<td>AY257800</td>
<td>2</td>
<td>AY257080</td>
</tr>
<tr>
<td>98CM9921</td>
<td>46</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>2</td>
<td>AY257801</td>
<td>2</td>
<td>AY257081</td>
</tr>
<tr>
<td>98CM9928</td>
<td>36</td>
<td>M</td>
<td>STD</td>
<td>No</td>
<td>4</td>
<td>AY257810</td>
<td>4</td>
<td>AY257082</td>
</tr>
</tbody>
</table>

TBC, Tuberculosis clinic attendees; STD, sexually transmitted disease attendees; CSP, HIV clinical suspect patient; NA, not available.

*France; †France, Germany, Italy and Switzerland; ‡France and Italy; §France, Togo, Benin and Ivory Coast; ‡‡Saudi Arabia; ††Nigeria; #France, Belgium and Italy; **Gabon and Nigeria; ‡‡Central African Republic.
Fig. 1

(a) Genotype 1
Core gene

(b) Genotype 1
NS5B gene

subtype 1b
(Global)

subtype 1a
(Global)

subtype 1c
(Indonesia)
Endemic HCV infection in Cameroon

Fig. 2
strains is shown clearly in the NS5B phylogeny also (Fig. 1b). Most of these strains cluster at the base of the global subtype 1a, whereas West African strains occupy the equivalent position for the global subtype 1b.

On the genotype 2 phylogenies (Fig. 2), the Cameroonian strains formed a cluster distinct from the global subtypes 2a, 2b and 2c. One strain (CM1581) fell outside this cluster in the core phylogeny (Fig. 2a), probably because of low bootstrap support arising from limited phylogenetic information. As Fig. 2(b) shows, the Cameroonian isolates fall within a larger group of West African strains that are genetically very diverse and tend to be found at the base of the global subtypes 2a, 2b and 2c, which are distributed throughout the tree. Therefore, it is reasonable to suppose that genotype 2 originated in West Africa and spread subsequently to other parts of Africa and the rest of the world. This observation is not clear from the core phylogeny (Fig. 2a), probably because of the paucity of core gene sequences from the region.

The Cameroonian type 4 isolates also display great genetic diversity (Fig. 3). The Cameroonian strains are interspersed with many strains sampled in Central Africa and the phylogenies show that HCV type 4a, which is highly prevalent in Egypt and is found also outside Africa, has its source in the Central African region. Isolate CM99897 from Cameroon formed an outgroup to the genotype 4 core tree (Fig. 2a), but this strain could not be amplified with the NS5B primers, possibly because it was too divergent.

In summary, our phylogenetic analysis of HCV isolates from Cameroon and reference strains from other parts of the world reveals great genetic diversity of genotypes 1 and 4 in Cameroon, but less diversity of genotype 2.

**DISCUSSION**

The most striking feature of the results presented above is the discovery of both highly divergent genotype 1 and 4 strains in Cameroon, indicating a pattern of long-term endemic infection that has not been observed in other regions of African. As discussed below, this result suggests new hypotheses regarding the origin and global epidemiology of HCV.

Our results demonstrate that HCV genotypes 1 and 2 are present in Cameroon – previous studies have typically found genotype 4 in Central Africa (Xu *et al.*, 1994; Fretz *et al.*, 1995; Stuyver *et al.*, 1994, 1995). The Cameroonian genotype 4 strains are more diverse than those sampled from other Central African countries or from Egypt; in the latter country, the HCV epidemic is quite homogeneous and is dominated by subtype 4a (Rapicetta *et al.*, 1998; Ray *et al.*, 2000). The presence of so many divergent and co-existing lineages of genotype 4 is strong evidence for the long-term presence of this strain in west Central Africa (modern Cameroon) and for its stability in transmission within the population. It is perhaps surprising that studies conducted in countries west of Cameroon have not revealed the presence of genotype 4 (Jeannele *et al.*, 1998). Only one study, concerning Nigeria, the closest West African neighbour of Cameroon, has documented the presence of this genotype (Oni & Harrison, 1996).

Substantial genetic diversity was also found among Cameroonian strains belonging to genotype 1. These strains are distinct from subtypes 1a and 1b prevalent in industrialized nations. No other study has reported such diversity of genotype 1 in the Central Africa region, although this is unsurprising, since the few studies performed so far have been restricted to small-scale samplings. Using the Inno-Lipa genotyping tool, Nkengasong *et al.* (1995b) reported the predominance of HCV genotypes 1 and 2 elsewhere in Cameroon. In contrast, there is clear evidence for considerable diversity of HCV type 1 in West Africa (Burkina Faso, Benin and Guinea; Jeannele *et al.*, 1998). The phylogeny presented in Fig. 1(b) suggests that subtype 1a (and one Egyptian strain) may have originated in Central Africa, whereas subtype 1b emerged possibly from West Africa. Although the phylogenetic position of the Indonesian subtype 1c is, as yet, unresolved, the limited diversity in this clade suggests that it represents a relatively recent introduction from Africa. Since the diversity of genotype 1 found in Cameroon is greater than that encountered in West Africa, it is reasonable to assume that this genotype was present earlier in the former region.

---

**Fig. 1. (on page 2336)** HCV genotype 1 phylogenies. Cameroonian isolates obtained in this study are highlighted in grey. The sampling location of some reference strains is shown in parentheses after the sequence name. Reference strains from West Africa are boxed. Many isolates from within the global clades were removed to improve clarity. Numerical values on branches indicate the percentage of bootstrap replicates that support that phylogenetic split. (a) Phylogeny estimated using core sequences. (b) Phylogeny estimated using NS5B sequences.

**Fig. 2. (on page 2337)** HCV genotype 2 phylogenies. Many isolates from within the global clades were removed to improve clarity. See the legend for Fig. 1 for other details. (a) Phylogeny estimated using core sequences. (b) Phylogeny estimated using NS5B sequences.

**Fig. 3. (on page 2338)** HCV genotype 4 phylogenies. Reference strains from Central Africa are boxed. See the legend for Fig. 1 for other details. (a) Phylogeny estimated using core sequences. (b) Phylogeny estimated using NS5B sequences.
Cameroonian genotype 2 sequences showed less genetic diversity and were restricted to a single cluster that seems to have originated in West Africa. Considerable diversity of this genotype has been reported in several West African countries (Jeannel et al., 1998; Wansbrough-Jones et al., 1998; Ruggieri et al., 1996), strongly suggesting a long-term maintenance of this genotype in West Africa.

Thus, previous studies have documented a high diversity of HCV genotypes 1 and 2 in West Africa and a high diversity of genotype 4 in Central African countries (Fretz et al., 1995; Stuyver et al., 1994, 1995; Mellor et al., 1995; Ruggieri et al., 1996; Wansbrough-Jones et al., 1998; Jeannel et al., 1998). Here we have reported a high genetic diversity of both genotypes 1 and 4 in Cameroon, a state that lies geographically between the West and Central regions of Africa. Phylogenetic analysis has shown that of all six HCV genotypes, only genotypes 1 and 4 reliably group together in a phylogenetic tree (Salem & Vandamme, 2002); therefore, these two genotypes share a common evolutionary origin and ancestor. These observations lead us to postulate a new hypothesis regarding the epidemiological history of HCV in Africa: the common ancestor of HCV genotypes 1 and 4 originated in the area where modern Cameroon is located, diverged and then spread to other parts of Africa and the rest of the world. This hypothesis implies a westward spread of genotype 1 into West Africa and an eastward spread of genotype 4 into Central Africa and the Middle East. Further sampling in Central and West Africa and the sequencing of larger subgenomic regions is required to confirm this hypothesis. If true, it may help to identify possible zoonotic reservoirs of HCV. It is more probable that genotype 2 originated in West Africa and spread to other regions through continuous migration among populations.

Phylogenetic and population genetic methods can be used to estimate the timescale of HCV endemicity in Cameroon. Smith et al. (1997) estimated that the HCV genotypes diversified about 500–2000 years ago, and Pybus et al. (2001) estimated that the most recent common ancestor of genotype 4 existed about 350 years ago. In contrast, the most recent common ancestors of subtypes 1a and 1b existed about 100 years ago. It is also clear that several genotype 1 and 4 lineages were present in Egypt about 80 years ago, since these lineages were all spread within Egypt by parenteral antischistosomal treatment (PAT) campaigns initiated around 1920 (Frank et al., 2000; Ray et al., 2000; Pybus et al., 2003).

The higher prevalence of HCV infection in Cameroon compared to other Central and West African countries is consistent with long-term endemicity in Cameroon (Louis et al., 1994; Kowo et al., 1995; Nkengasong et al., 1995a; Ndjomou et al., 2002; Xu et al., 1994; Fretz et al., 1995; Ruggieri et al., 1996; Jeannel et al., 1998; Wansbrough-Jones et al., 1998). We are unaware of a scenario in Cameroon similar to the Egyptian PAT campaign, so it is possible that the high prevalence of HCV in Cameroon has been favoured by long-term stability in virus transmission, facilitated possibly by unsafe parenteral practices.

We have avoided naming new subtypes in this study, as the diversity observed in Cameroon and other endemic areas leads simply to an unmanageable profusion of names and illustrates the need to rationalise HCV subtype classification. Several studies have described a correlation between virus type and the response to interferon treatment (Kanai et al., 1992; Takada et al., 1992; Yoshioka et al., 1992; Chemello et al., 1994; Hino et al., 1994; Tisminetzky et al., 1994) and it is possible that disease outcome also depends on virus type (Simmonds et al., 1996). Furthermore, the wide genetic diversity of HCV represents a major challenge for future vaccine development. Therefore, efficient HCV genotyping and classification is of importance to clinical practice, and therapies that are effective against the greatest possible range of HCV variants should be sought.

ACKNOWLEDGEMENTS

We are grateful to clinicians and patients for their collaboration. This work was supported by a grant from the University of Bonn (grant number 0-151.0013 BONFOR). J.N. was supported by the Deutscher Akademischer Austauschdienst (DAAD). O.G.P. was funded by the Wellcome Trust. We thank M. Dauemer, R. Kaiser, A. Yassin and B. Kupfer for help with sequence determination.

REFERENCES


