Journal of General Virology (2003), **84**, 2333–2341 DOI 10.1099/vir.0.19240-0

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

Jean Ndjomou,[1]† Oliver G. Pybus[2] and Bertfried Matz[1]

[1]Institute of Medical Microbiology and Immunology, University of Bonn, Sigmund-Freud Strasse 25, 53105 Bonn, Germany

[2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Correspondence
Jean Ndjomou
ndjomou_j@hotmail.com

Received 19 March 2003
Accepted 9 May 2003

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 subSaharan 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 subSaharan Africa is, therefore, necessary to provide insights into the global epidemiology, epidemic history and origin of HCV. Furthermore, availability of more sequence data from subSaharan 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

†Present address: Department of Microbiology and Immunology, Indiana University School of Medicine, R2 302, 950 W. Walnut St. Indianapolis, IN 46202, USA.

The sequences reported in this study are deposited in GenBank under the following accession numbers: AY257069–AY257103 for the NS5B region and AY256784–AY256820 for the core region.

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 5 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 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 $MgCl_2$, 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 $MgCl_2$, 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 $MgCl_2$, 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 $MgCl_2$, 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 $MgCl_2$, 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 *Eco*RI 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

Table 1. Epidemiological data and genotyping results

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

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, Gabon, Benin and Ivory Coast; ‖Saudi Arabia; ¶Nigeria; #France, Belgium and Italy; **Gabon and Nigeria; ††Central African Republic.

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

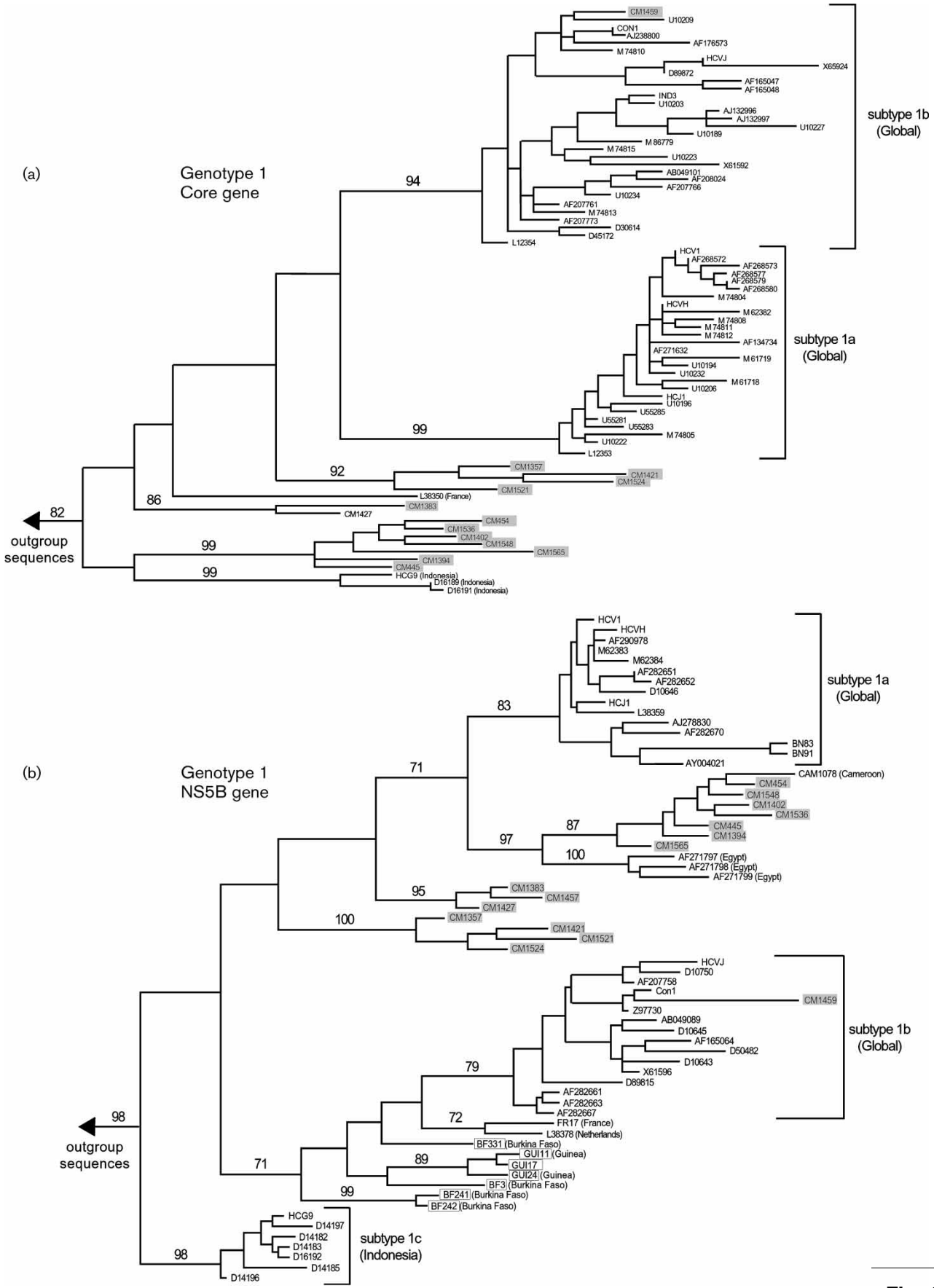
(a)
Genotype 1
Core gene
CM1459
U10209
CON1
AJ238800
AF176573
M74810
HCVJ
D89872
X65924
AF165047
AF165048
IND3
U10203
AJ132996
AJ132997
U10227
U10189
M86779
M74815
U10223
X61592
AB049101
AF208024
AF207766
U10234
AF207761
M74813
AF207773
D30614
D45172
L12354
subtype 1b
(Global)
94
HCV1
AF268572
AF268573
AF268577
AF268579
AF268580
M74804
HCVH
M62382
M74808
M74811
M74812
AF134734
AF271632
M61719
U10194
U10232
M61718
U10206
HCJ1
U10196
U55285
U55281
U55283
M74805
U10222
L12353
subtype 1a
(Global)
99
92
CM1357
CM1421
CM1524
CM1521
L38350 (France)
86
CM1383
82
CM1427
outgroup
sequences
CM454
CM1536
CM1402
CM1548
99
CM1565
CM1394
CM445
99
HCG9 (Indonesia)
D16189 (Indonesia)
D16191 (Indonesia)
(b)
Genotype 1
NS5B gene
HCV1
HCVH
AF290978
M62383
M62384
AF282651
AF282652
D10646
HCJ1
L38359
AJ278830
AF282670
BN83
BN91
AY004021
subtype 1a
(Global)
83
71
CAM1078 (Cameroon)
CM454
CM1548
CM1402
CM1536
87
CM445
CM1394
97
CM1565
100
AF271797 (Egypt)
AF271798 (Egypt)
AF271799 (Egypt)
95
CM1383
CM1457
CM1427
100
CM1357
CM1421
CM1521
CM1524
HCVJ
D10750
AF207758
Con1
CM1459
Z97730
AB049089
D10645
AF165064
D50482
D10643
X61596
D89815
subtype 1b
(Global)
79
AF282661
AF282663
AF282667
72
FR17 (France)
L38378 (Netherlands)
98
outgroup
sequences
BF331 (Burkina Faso)
GUI11 (Guinea)
89
GUI17
GUI24 (Guinea)
71
BF3 (Burkina Faso)
99
BF241 (Burkina Faso)
BF242 (Burkina Faso)
HCG9
D14197
D14182
D14183
D16192
D14185
98
subtype 1c
(Indonesia)
D14196

Fig. 1

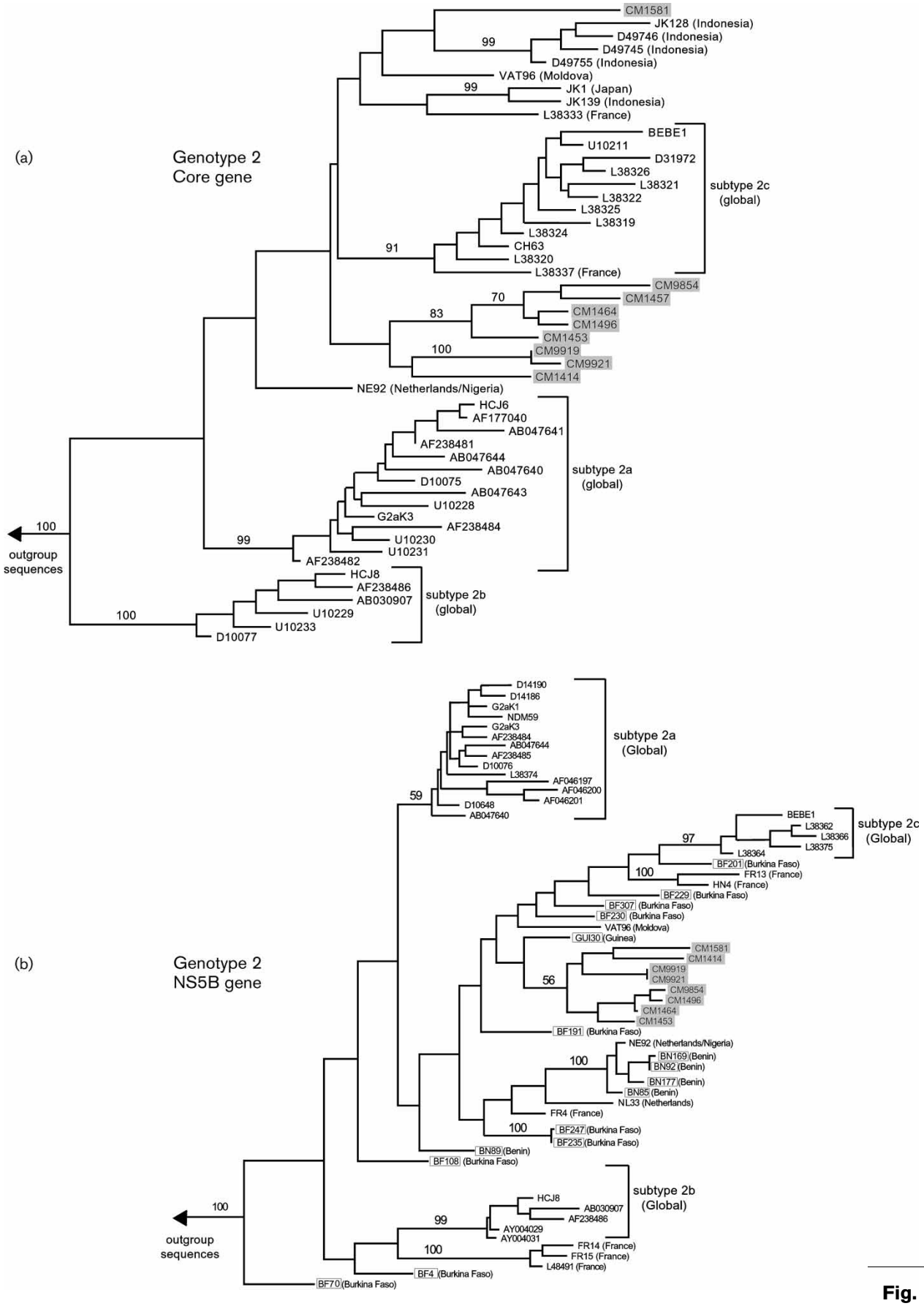
(a)
Genotype 2
Core gene
CM1581
JK128 (Indonesia)
D49746 (Indonesia)
D49745 (Indonesia)
D49755 (Indonesia)
VAT96 (Moldova)
JK1 (Japan)
JK139 (Indonesia)
L38333 (France)
BEBE1
U10211
D31972
L38326
L38321
L38322
L38325
L38319
L38324
CH63
L38320
L38337 (France)
subtype 2c
(global)
CM9854
CM1457
CM1464
CM1496
CM1453
CM9919
CM9921
CM1414
NE92 (Netherlands/Nigeria)
HCJ6
AF177040
AB047641
AF238481
AB047644
AB047640
D10075
AB047643
U10228
G2aK3
AF238484
U10230
U10231
AF238482
subtype 2a
(global)
HCJ8
AF238486
AB030907
U10229
U10233
D10077
subtype 2b
(global)
outgroup
sequences
99
91
83
70
100
99
100
100
(b)
Genotype 2
NS5B gene
D14190
D14186
G2aK1
NDM59
G2aK3
AF238484
AB047644
AF238485
D10076
L38374
AF046197
AF046200
AF046201
D10648
AB047640
subtype 2a
(Global)
BEBE1
L38362
L38366
L38375
L38364
subtype 2c
(Global)
BF201 (Burkina Faso)
FR13 (France)
HN4 (France)
BF229 (Burkina Faso)
BF307 (Burkina Faso)
BF230 (Burkina Faso)
VAT96 (Moldova)
GUI30 (Guinea)
CM1581
CM1414
CM9919
CM9921
CM9854
CM1496
CM1464
CM1453
BF191 (Burkina Faso)
NE92 (Netherlands/Nigeria)
BN169 (Benin)
BN92 (Benin)
BN177 (Benin)
BN85 (Benin)
NL33 (Netherlands)
FR4 (France)
BF247 (Burkina Faso)
BF235 (Burkina Faso)
BN89 (Benin)
BF108 (Burkina Faso)
HCJ8
AB030907
AF238486
AY004029
AY004031
FR14 (France)
FR15 (France)
L48491 (France)
BF4 (Burkina Faso)
BF70 (Burkina Faso)
subtype 2b
(Global)
outgroup
sequences
97
100
56
100
59
100
99
100
100

Fig. 2

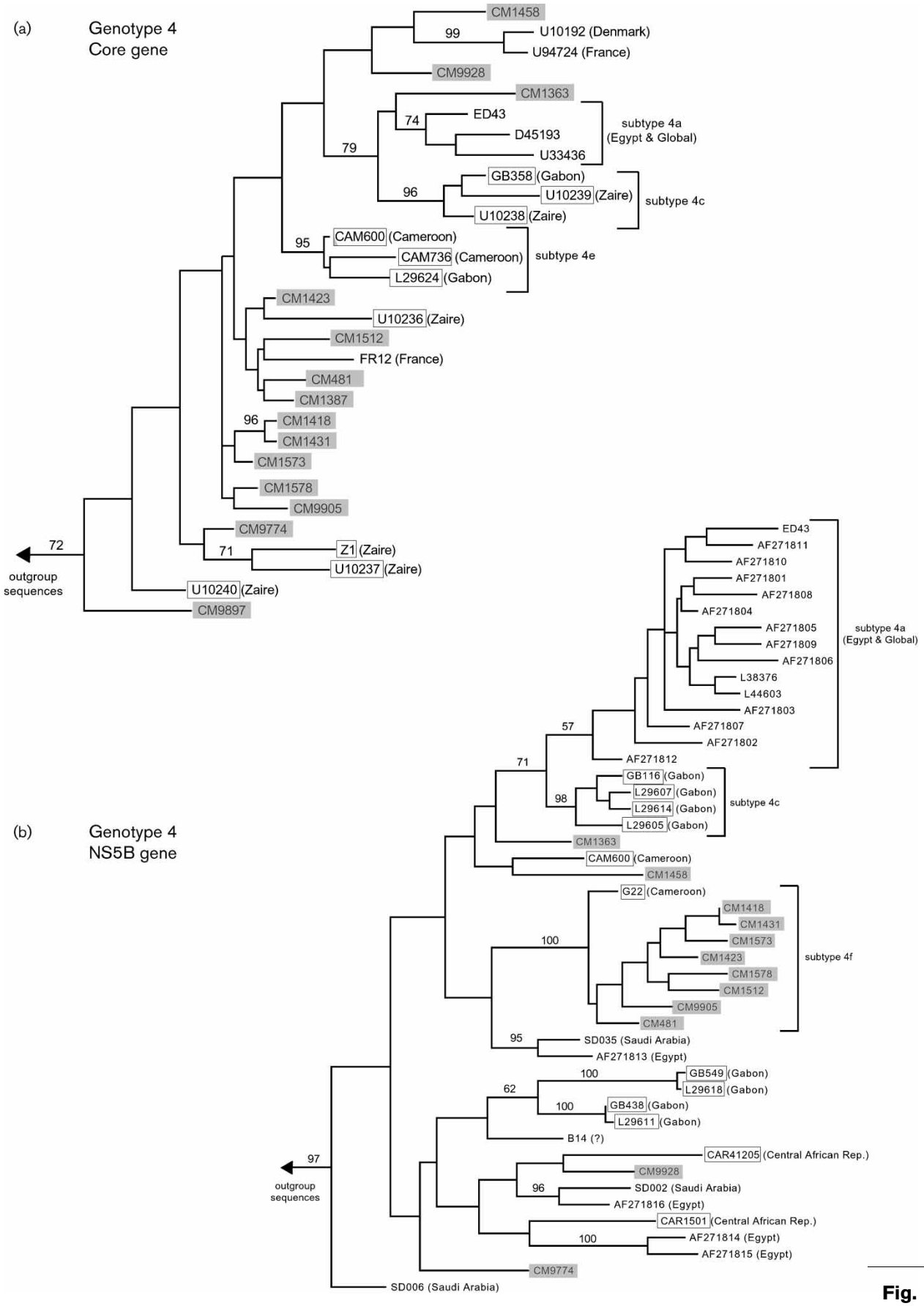
(a)
Genotype 4
Core gene
CM1458
99
U10192 (Denmark)
U94724 (France)
CM9928
CM1363
74
ED43
D45193
U33436
subtype 4a
(Egypt & Global)
79
GB358 (Gabon)
96
U10239 (Zaire)
U10238 (Zaire)
subtype 4c
95
CAM600 (Cameroon)
CAM736 (Cameroon)
L29624 (Gabon)
subtype 4e
CM1423
U10236 (Zaire)
CM1512
FR12 (France)
CM481
CM1387
96
CM1418
CM1431
CM1573
CM1578
CM9905
CM9774
72
71
Z1 (Zaire)
U10237 (Zaire)
outgroup
sequences
U10240 (Zaire)
CM9897
(b)
Genotype 4
NS5B gene
ED43
AF271811
AF271810
AF271801
AF271808
AF271804
AF271805
AF271809
AF271806
subtype 4a
(Egypt & Global)
L38376
L44603
AF271803
57
AF271807
AF271802
AF271812
71
GB116 (Gabon)
L29607 (Gabon)
98
L29614 (Gabon)
L29605 (Gabon)
subtype 4c
CM1363
CAM600 (Cameroon)
CM1458
G22 (Cameroon)
CM1418
CM1431
100
CM1573
CM1423
CM1578
CM1512
CM9905
CM481
subtype 4f
95
SD035 (Saudi Arabia)
AF271813 (Egypt)
100
GB549 (Gabon)
62
L29618 (Gabon)
100
GB438 (Gabon)
L29611 (Gabon)
B14 (?)
97
CAR41205 (Central African Rep.)
CM9928
outgroup
sequences
96
SD002 (Saudi Arabia)
AF271816 (Egypt)
CAR1501 (Central African Rep.)
100
AF271814 (Egypt)
AF271815 (Egypt)
CM9774
SD006 (Saudi Arabia)

Fig. 3

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 CM9897 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 (Jeannel *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; Jeannel *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 (Salemi & 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 anti-schistosomal 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 Deustcher 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

Chemello, L., Alberti, A., Rose, K. & Simmonds, P. (1994). Hepatitis C serotype and response to interferon therapy. *N Engl J Med* **330**, 143.

Choo, Q. L., Richman, K. H., Han, J. H. & 11 other authors (1991). Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A* **88**, 2451–2455.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Frank, C., Mohamed, M. K., Strickland, G. T. & 8 other authors (2000). The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**, 887–891.

Fretz, C., Jeannel, D., Stuyver, L., Hervé, V., Lunel, F., Boudifa, A., Mathiot, C., de Thé, G. & Fournel, J. J. (1995). HCV infection in a rural population of the Central African Republic (CAR): evidence for three additional subtypes of genotypes 4. *J Med Virol* **47**, 435–437.

Hino, K., Sainokami, S., Shimoda, K., Iino, S., Wang, Y., Okamoto, H., Miyakawa, Y. & Mayumi, M. (1994). Genotypes and titers of hepatitis C virus for predicting response to interferon in patients with chronic hepatitis C. *J Med Virol* **42**, 299–305.

Jeannel, D., Fretz, C., Traore, Y. & 9 other authors (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.

Kanai, K., Kako, M. & Okamoto, H. (1992). HCV genotypes in chronic hepatitis C and response to interferon. *Lancet* **339**, 1543.

Kowo, M. P., Goubau, P., Ndam, E.-C., Njoya, O., Sasaki, S., Seghers, V. & Kesteloot, H. (1995). Prevalence of hepatitis C virus and other blood-borne viruses in Pygmies and neighbouring Bantus in southern Cameroon. *Trans R Soc Trop Med Hyg* **89**, 484–486.

Louis, F. J., Maubert, B., Le Hesran, J. Y., Kemmegne, J., Delaporte, E. & Louis, J. P. (1994). High prevalence of anti-hepatitis C virus

antibodies in a Cameroon rural forest area. *Trans R Soc Trop Med Hyg* **88**, 53–54.

Mellor, J., Holmes, E. C., Jarvis, L. M., Yap, P. L. & Simmonds, P. (1995). Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J Gen Virol* **76**, 2493–2507.

Ndjomou, J., Kupfer, B., Kochan, B., Zekeng, L., Kaptue, L. & Matz, B. (2002). Hepatitis C virus infection and genotypes among human immunodeficiency virus high-risk groups in Cameroon. *J Med Virol* **66**, 179–186.

Nkengasong, J. N., De Beenhouwer, H., Claeys, H., Nyambi, P., Ayuk, J., van der Groen, G. & Ndumbe, P. (1995a). A pilot study of the prevalence of hepatitis C virus antibodies and hepatitis C virus RNA in southern Cameroon. *Am J Trop Med Hyg* **52**, 98–100.

Nkengasong, J. N., Nyambi, P., Claeys, H., De Beenhouwer, H., Collart, J.-P., Ayuk, J. & Ndumbe, P. (1995b). Predominantly hepatitis C virus genotypes 1 and 2 are found in Cameroon. *J Infect Dis* **171**, 1380–1381.

Ohno, T., Mizokami, M., Saleh, M. G. & 8 other authors (1996). Usefulness and limitation of phylogenetic analysis for hepatitis C virus core region: application to isolates from Egyptian and Yemeni patients. *Arch Virol* **141**, 1101–1113.

Oni, A. O. & Harrison, T. J. (1996). Genotypes of hepatitis C virus in Nigeria. *J Med Virol* **49**, 178–186.

Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C. & Harvey, P. H. (2001). The epidemic behaviour of the hepatitis C virus. *Science* **292**, 2323–2325.

Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. (2003). The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**, 381–387.

Rapicetta, M., Argentini, C., Dettori, S., Spada, E., Pellizzer, G. & Gandin, C. (1998). Molecular heterogeneity and new subtypes of HCV genotype 4. *Res Virol* **149**, 293–297.

Ray, S. C., Arthur, R. R., Carella, A., Bukh, J. & Thomas, D. L. (2000). Genetic epidemiology of hepatitis C virus throughout Egypt. *J Infect Dis* **182**, 698–707.

Robertson, B., Myers, G., Howard, C. & 14 other authors (1998). Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Taxonomy of Viruses. *Arch Virol* **143**, 2493–2503.

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.

Salemi, M. & Vandamme, A. M. (2002). Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* **54**, 62–70.

Simmonds, P. (1995). Variability of hepatitis C virus. *Hepatology* **21**, 570–583.

Simmonds, P., Mellor, J., Craxi, A. & 12 other authors (1996). Epidemiological, clinical and therapeutic associations of hepatitis C types in western European patients. *J Hepatol* **24**, 517–524.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L. & Simmonds, P. (1997). The origin of hepatitis C virus genotypes. *J Gen Virol* **78**, 321–328.

Stuyver, L., van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E. & Maertens, G. (1994). Classification of hepatitis C viruses based on phylogenetic analysis of the envelope 1 and nonstructural 5B regions and identification of five additional subtypes. *Proc Natl Acad Sci U S A* **91**, 10134–10138.

Stuyver, L., Wyseur, A., van Arnhem, W. & 7 other authors (1995). Hepatitis C virus genotyping by means of 5′-UR/core line probe assays and molecular analysis of untypeable samples. *Virus Res* **38**, 137–157.

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA, USA.

Takada, N., Takase, S., Enomoto, N., Takada, A. & Date, T. (1992). Clinical backgrounds of the patients having different types of hepatitis C virus genomes. *J Hepatol* **14**, 35–40.

Tisminetzky, S. G., Gerotto, M., Pontisso, P., Chemello, L., Ruvoletto, M. G., Baralle, F. & Alberti, A. (1994). Genotypes of hepatitis C virus in Italian patients with chronic hepatitis C. *Int Hepatol Commun* **2**, 105–112.

Viazov, S., Kuzin, S., Paladi, N., Tchernovetsky, M., Isaeva, E., Mazhul, L., Vasychova, F., Widell, A. & Roggendorf, M. (1997). Hepatitis C virus genotypes in different regions of the former Soviet Union (Russia, Belarus, Moldova, and Uzbekistan). *J Med Virol* **53**, 36–40.

Wansbrough-Jones, M. H., Frimpong, E., Cant, B., Harris, K., Evans, M. R. W. & Teo, C. G. (1998). Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana. *Trans R Soc Trop Med Hyg* **92**, 496–499.

WHO (1997). Hepatitis C: global prevalence. *Wkly Epidemiol Rec* **72**, 341–344.

Xu, L. Z., Larzul, D., Delaporte, E., Bréchot, C. & Kremsdorf, D. (1994). Hepatitis C virus genotype 4 is highly prevalent in central Africa (Gabon). *J Gen Virol* **75**, 2393–2398.

Yoshioka, K., Kakumu, S., Wakita, T., Ishikawa, T., Itoh, Y., Takayanagi, M., Higashi, Y., Shibata, M. & Morishima, T. (1992). Detection of hepatitis C virus by polymerase chain reaction and response to interferon-alpha therapy: relationship to genotypes of hepatitis C virus. *Hepatology* **16**, 293–299.