



## Hepatitis C virus infections in the Democratic Republic of Congo exhibit a cohort effect



James C. Iles<sup>a</sup>, G.L. Abby Harrison<sup>b</sup>, Sinead Lyons<sup>c</sup>, Cyrille F. Djoko<sup>d</sup>, Ubald Tamoufe<sup>e</sup>, Matthew Lebreton<sup>e</sup>, Bradley S. Schneider<sup>e</sup>, Joseph N. Fair<sup>e</sup>, Felix M. Tshala<sup>f</sup>, Patrick K. Kayembe<sup>g</sup>, Jean Jacques Muyembe<sup>h</sup>, Samuel Edidi-Basepeo<sup>i</sup>, Nathan D. Wolfe<sup>e,j</sup>, Paul Klennerman<sup>b</sup>, Peter Simmonds<sup>c</sup>, Oliver G. Pybus<sup>a,\*</sup>

<sup>a</sup> Department of Zoology, University of Oxford, South Parks Road, Oxford, UK

<sup>b</sup> Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK

<sup>c</sup> Centre for Immunology, Infection and Evolution, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK

<sup>d</sup> Global Viral, Yaoundé, Cameroon

<sup>e</sup> Metabiota, San Francisco, USA

<sup>f</sup> Division of Military Health, Ministry of Defense, Kinshasa, Democratic Republic of the Congo

<sup>g</sup> Kinshasa School of Public Health, Kinshasa, Democratic Republic of the Congo

<sup>h</sup> National Institute of Biomedical Research, Kinshasa, Democratic Republic of the Congo

<sup>i</sup> National AIDS Control Program, Reference Laboratory, Kinshasa, Democratic Republic of the Congo

<sup>j</sup> Stanford University Program in Human Biology, Stanford, USA

### ARTICLE INFO

#### Article history:

Available online 16 February 2013

#### Keywords:

Hepatitis C  
Phylogeny  
Democratic Republic of Congo  
Prevalence  
RNA  
Human pegivirus

### ABSTRACT

The prevalence and genetic diversity of hepatitis C virus (HCV) and human pegivirus (HPgV) in many regions of sub-Saharan Africa is poorly characterized, including in the Democratic Republic of Congo – the largest country in the region and one of the most populous. To address this situation we conducted a molecular epidemiological survey of HCV and HPgV (previously named GB Virus C or hepatitis G virus) in samples collected in 2007 from 299 males from the DRC, whose ages ranged from 21 to 71 years old. Samples were tested for the presence of HCV antibodies by ELISA and reactive samples were subsequently tested for HCV RNA using RT-PCR in which both the HCV Core and NS5B genome regions were amplified. Remaining samples were tested for HPgV RNA and the HPgV NS3 genome region of positive samples was amplified. For HCV, 13.7% of the samples were seropositive (41/299) but only 3.7% were viremic (11/299). HPgV RNA was found in 12.7% (33/259) of samples. HCV viremia was strongly associated with age; the percentage of samples that contained detectable HCV RNA was ~0.5% in those younger than 50 and 13% in those older than 50. Our study represents the first systematic survey of HCV genetic diversity in the DRC. HCV sequences obtained belonged to diverse lineages of genotype 4, including subtypes 4c, 4k, 4l and 4r, plus one unclassified lineage that may constitute a new subtype. These data suggest that HCV in the DRC exhibits an age ‘cohort effect’, as has been recently reported in neighbouring countries, and are consistent with the hypothesis that HCV transmission rates were higher in the mid-twentieth century, possibly as a result of parenteral, iatrogenic, or other unidentified factors. Different HCV subtypes were associated with individuals of different ages, implying that HCV infection in the DRC may have arisen through multiple separate HCV epidemics with different causes.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Hepatitis C virus (HCV) is a globally-distributed human pathogen, present in approximately 130–170 million people worldwide, and an estimated 3–4 million people are thought to be infected with HCV each year (Shepard et al., 2005). The introduction of screening strategies for HCV following its discovery in 1989 (Choo et al., 1989) has greatly reduced the transmission of the virus

through blood transfusion and blood products, and the main transmission route of HCV in developed countries is now the use of contaminated needles by injecting drug users (Pondé, 2011). In developing countries, non-sterile injections and other unsafe medical interventions are thought to contribute to continuing HCV incidence (Kane et al., 1999).

The HCV genome exhibits considerable sequence heterogeneity and is classified using phylogenetic methods into six confirmed genotypes, each of which is further subdivided into numerous subtypes (Simmonds, 2004). A seventh provisional genotype was isolated in Canada in 2007 from an individual originally from the

\* Corresponding author.

E-mail address: oliver.pybus@zoo.ox.ac.uk (O.G. Pybus).

**Table 1**

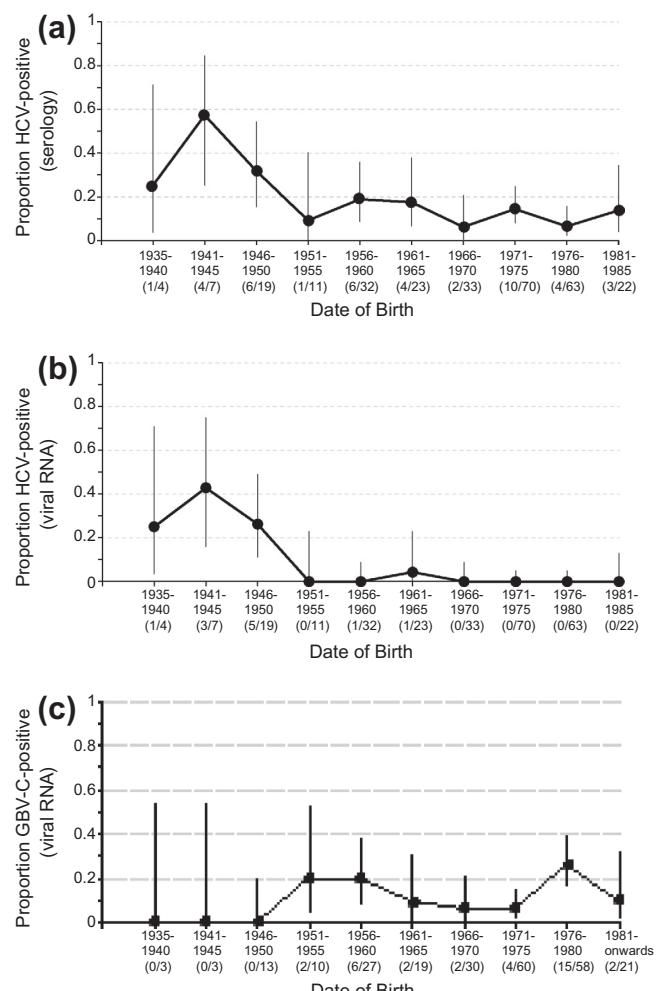
Primers used in the RT-PCR, PCR and sequencing reactions of the 5'UTR, Core and NS5B regions of HCV, and of the NS3 region of HPgV.

Primer	Primer sequences	Genotype
5'UTREx400F	CCTTGTGGTACTGCCGTAG	All
5'UTRIn405F	CTGATAGGGTGCCTGGAGTG	All
CoreIn980R	AGTGC CARRAGGAAGATAGA	All
E'In1420R	CCAGTTCATCATCATGTC	All
E'Ex1423R	GGRCTCCAGTCATCATGTC	All
5'UTRExF1	CCCTGTGAGGAACCTWCTGCTTACGC	All
5'UTRInF3	TCTAGCCATGGCGTTAGTRYGAG	All
CoreExR2	GGTGCACGGTCTACGAGACT	All
CoreInR4	CACTCGAAGCACCTATCAGGAGT	All
NS5BExF	TGGGGATCCGTATGATACCCGCTGCTTGA	1–5 and 7
NS5BExR	CGGAATTCTGGTCATAGCCTCCGAA	1–5 and 7
NS5BExFG6	CCHATGGGGTTCTAYGACAC	6
NS5BExRG6	GNGCYGAGTAYCTGGTCATGGC	6
NS5BInF	GACACCCGCTGCTTGA	All
NS5BInR	GAGTCTCACGGAGGCTATGACNAGGTA	All
GBVNS3_4126a	GSGCNATGGNCNTAYATGGA	HPgV
GBVNS3_4735as	GTNACYTCVACNACCTCTCYACCA	HPgV
GBVNS3_4275s	GTGGTNATHGYGAGTYCA	HPgV
GBVNS3_4543as	TCRCACTCMRCCTKGARTGRCARAA	HPgV

Democratic Republic of Congo (DRC; Murphy et al., 2007a,b). Some regions, such as sub-Saharan Africa and South-East Asia, harbor unusually diverse HCV strains, likely reflecting the long-term endemic transmission of HCV in these locations (Simmonds, 2004; Pybus et al., 2007). For example, highly-diverse lineages of HCV genotypes 1 and 2 are present in West Africa (Jeannel et al., 1998; Candotti et al., 2003; Markov et al., 2009). Further east, HCV genotypes 1 and 4 are commonly found in central African countries such as Cameroon, Gabon and the Central African Republic (CAR), and strains of HCV genotype 4 are also found in Egypt and the Middle East (Ray et al., 2000; Pybus et al., 2003; Shepard et al., 2005; Njouom et al., 2009). Overall, considerable sequence diversity of HCV genotypes 1, 2 and 4 in sub-Saharan Africa has been observed, and it is likely these strains have been present there for at least several centuries (Smith et al., 1997; Pybus et al., 2001; Pybus et al., 2007; Ndjomou et al., 2003).

Previous studies have reported high HCV seroprevalence in many central and West African countries and the WHO estimates that the region contains 18.8% of HCV infections worldwide (Kane et al., 1999; Berkes and Cotler, 2005). For example, 5.2% of blood donors in Ghana were seropositive (Candotti et al., 2003), as were 11.6% of blood donors in Cameroon (Nerrienet et al., 2005) and 6.5% of subjects in Gabon (Delaporte et al., 1993). In addition to high overall prevalence in the general population, several studies of central African populations have reported a significant increase in HCV prevalence with age, for example in Cameroon (Nerrienet et al., 2005; Pépin et al., 2010b), Gabon (Ndong-Atome et al., 2008), Equatorial Guinea (Basaras et al., 1999) and the Republic of the Congo (Cantaloube et al., 2010). The observation of high HCV prevalences in those aged 50 or older could be explained by a defined period of increased transmission in the past, possibly as a result of non-sterile medical interventions, as has been proposed for Cameroon between the 1920s and 1960s (Njouom et al., 2007). Support for this interpretation of the age distribution of HCV comes from Egypt, where variation in HCV prevalence among age groups and among locations closely matches the level of exposure of those groups to parenteral anti-schistosomiasis therapy, which was widely administered in the first half of the twentieth century (Frank et al., 2000).

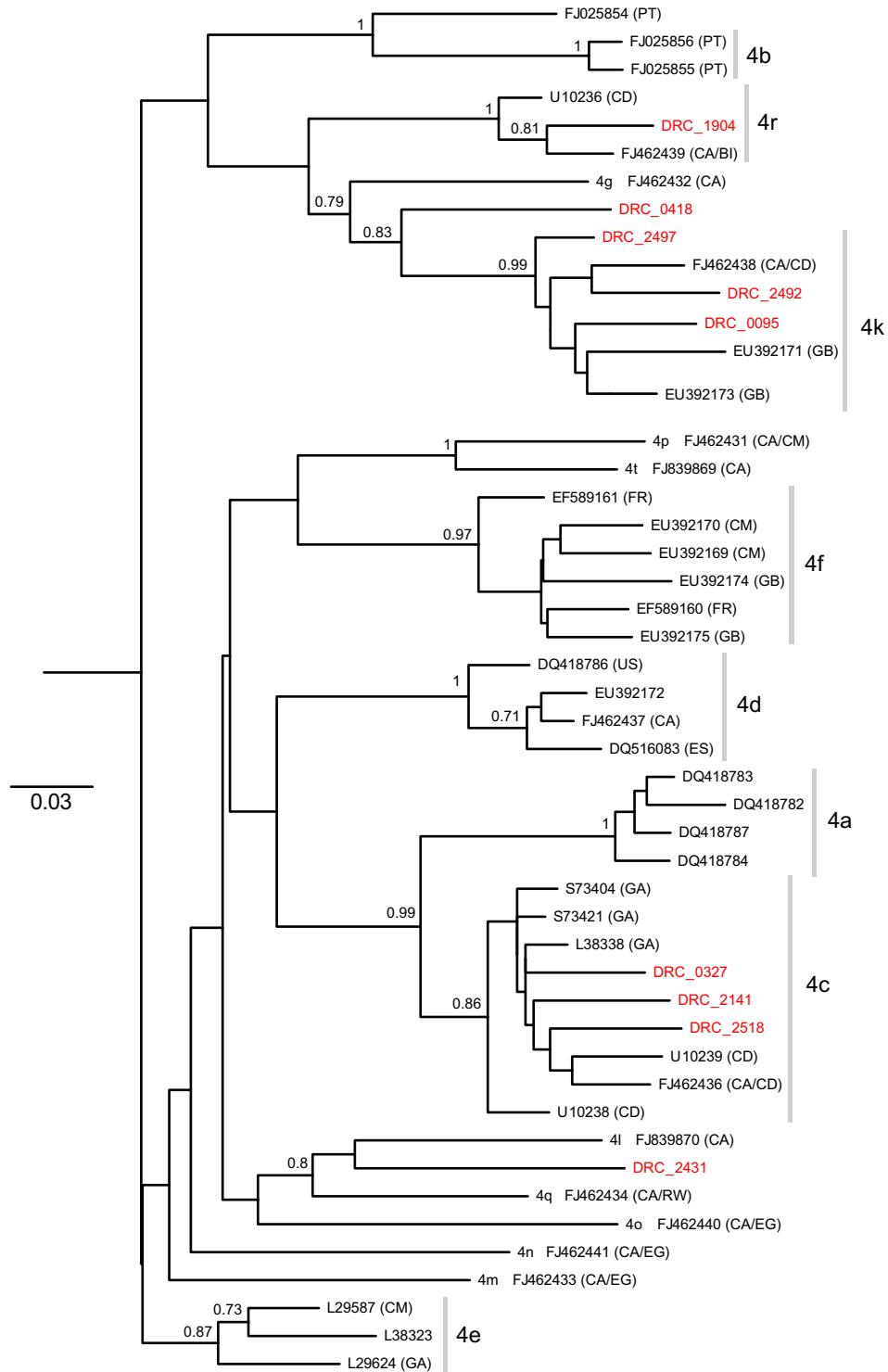
Despite its large size and geographically central position, there have been few studies to date of HCV in the DRC. Tibbs et al. (1991) screened 173 samples from rural populations in local hospitals for anti-HCV antibodies and estimated HCV seroprevalence to be 6.4%. Laurent et al. (2001) surveyed pregnant women and commercial



**Fig. 1.** The age distribution among 299 blood samples from the DRC of (a) HCV seroprevalence (all reactive samples) and (b) HCV RNA prevalence (Core and/or NS5b sequence). (c) The age distribution of HPgV RNA prevalence among HCV seronegative samples. Samples were assigned to one of ten age categories by date of birth. Numbers in parenthesis below each category are the number of positive samples/total number of samples. Fifteen samples (all negative) did not have date of birth information and are therefore not included in these numbers. The y-axis shows the proportion of samples in each age category that were positive. The errors bars represent 95% confidence limits of this proportion, estimated using the Adjusted Wald method (Agresti and Coull 1998). Individuals born before 1946 were significantly more likely to be positive for HCV RNA and anti-HCV antibodies than those born later.

sex workers (CSW) in Kinshasa and reported seroprevalences of 6.6% among CSW and 4.3% among pregnant women. Liu et al. (1999) also investigated pregnant women from Kinshasa ( $n = 97$ ) and found that while 10.3% carried human pegivirus (HPgV) RNA, only 1% carried HCV RNA. Most recently, Batina Agasa et al. (2010) screened 140 patients with sickle-cell disease in Kisangani for anti-HCV antibodies and found 7.9% were seropositive, all of whom had received blood transfusions before the introduction of HCV screening in 2004. Most of these prevalence studies have been limited in size and study population.

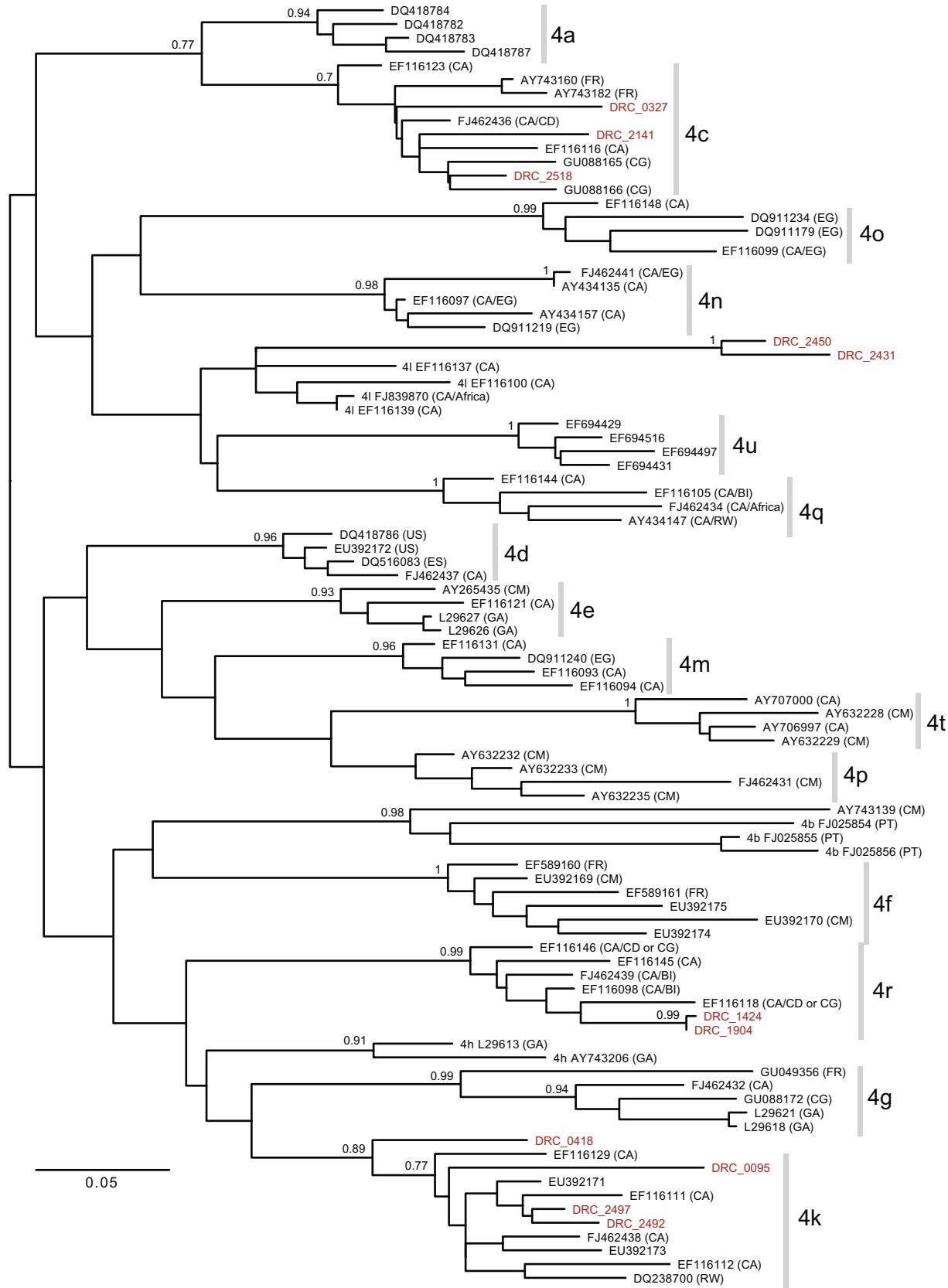
In addition to our poor understanding of HCV epidemiology in the DRC, there is scant information about the genetic diversity of HCV in the country. No country-specific surveys have been published and the HCV sequence database (Kuiken et al., 2005) contains only 21 sub-genomic sequences from 9 isolates labeled as originating from the DRC. These sequences belong to various subtypes of genotype 4, although classification is uncertain, as many of



**Fig. 2.** Maximum likelihood phylogeny of HCV genotype 4, estimated from the Core region sequences. New DRC isolates from this study are in red; other sequences are reference isolates from Genbank or the HCV Sequence Database. Numbers next to nodes represent maximum likelihood bootstrap support values; only values >0.7 are shown. Grey bars indicate the location of HCV subtypes. Reference isolates are labeled with accession number and, where available, the location of sampling (ISO 3166 two-letter country codes, in parenthesis). For some isolates two locations are noted: the first corresponds to the country of sampling and the second corresponds to the country of origin (information obtained from the primary literature). The phylogeny is mid-point rooted and the scale bar is in units of nucleotide substitutions per site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

these sequences are <500 nt in length. However, both patients in which the provisional genotype 7 was discovered were originally from the DRC (Murphy et al., 2007a,b) suggesting that the country may harbor further undetected diversity. The presence of high viral diversity in a region can aid the search for the geographic origin

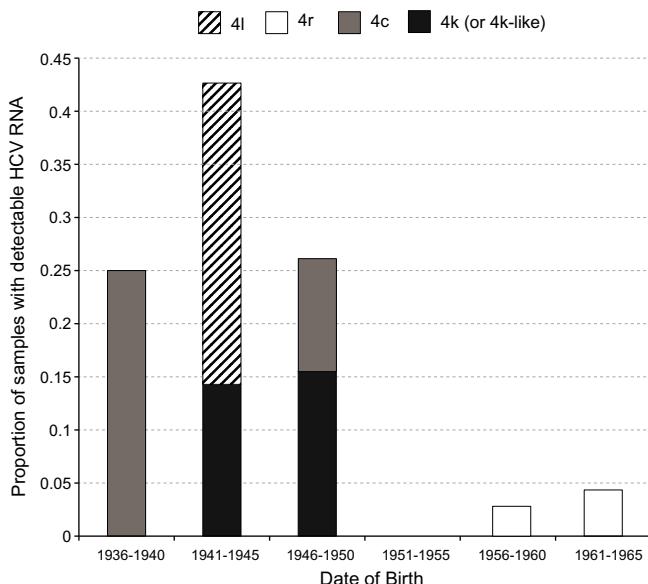
and source population of a virus, neither of which is known for HCV. The only other known viruses in the genus Hepacivirus are GB Virus B (GBV-B) and the Canine Hepacivirus/Non-Primate Hepacivirus (CHV/NPHV), both of which are highly divergent from HCV and not known to infect humans: GBV-B has not been found



**Fig. 3.** Maximum-likelihood phylogeny of HCV genotype 4, estimated from NS5B region sequences. For further details see legend to Fig. 2.

in a natural host, and to date CHV/NPHV has been found only in dogs and horses (Kapoor et al., 2011; Stapleton et al., 2011; Burbelo et al., 2012; Lyons et al., 2012).

In order to improve our knowledge of the molecular epidemiology and genetic diversity of HCV in the DRC 299 blood samples from the country were screened for HCV using both serological as-



**Fig. 4.** The age distribution of HCV subtypes among samples with detectable HCV RNA in the Core and/or NS5B region. Samples from the oldest HCV-positive subjects belonged to subtype 4c. Subtypes 4k, 4I and 4r were most common in samples from subjects born between 1941 and 1950.

says and PCR. From those samples that contained HCV RNA we attempted to obtain viral sequences from the Core and NS5B genes. In addition, the same samples were also screened for HPgV RNA using PCR and part of the NS3 gene was sequenced from HPgV positive samples. Note that HPgV was previously termed GB Virus C or hepatitis G virus and its new name has been provisionally approved by the ICTV Flavivirus Study Group. Using available demographic information the age distribution of HCV infections in our sample set were reconstructed and compared.

## 2. Materials and methods

### 2.1. Study population

EDTA blood samples were collected from informed consenting members of the uniformed services as part of a screening program for HIV and other infectious diseases. These samples have been studied previously for HIV-1 (Djoko et al., 2011, wherein full details of sample collection can be found) and human parvovirus 4 (Sharp et al., 2010). Collection took place in Kinshasa, capital of the DRC, between June and September 2007. The samples were anonymised although patient date of birth and date of collection were available for most. All samples were from male individuals. Mean age was  $38.7 \pm 1.28$ , range 21–71.

### 2.2. HCV serology, RT-PCR and sequencing

Serological tests for HCV were conducted using Ortho 3.0 Enhanced SAVe (Ortho Clinical Diagnostics) as per the manufacturer's instructions. All HCV samples that were reactive for anti-HCV antibodies were tested for HCV-RNA. Viral RNA was extracted from sera using the Qiagen miniprep kit (QIAGEN) as per the manufacturer's instructions, with one modification: 500  $\mu$ l of sera was centrifuged at 4000 rpm for 1 h to concentrate viral particles, from which 360  $\mu$ l of sera was removed. RNA was subsequently extracted from the remaining 140  $\mu$ l. Three subgenomic regions were amplified (5'UTR, Core, and NS5B) using Superscript III (Invitrogen,

Life Technologies) followed by nested PCR with the Fast Start High Fidelity PCR System (Roche Applied Science) using standard protocols. Controls were run in parallel at each step. Primers were designed from an alignment of representative genomes from each of the 7 HCV genotypes; primers are listed in Table 1. The internal primers were used for sequencing with BigDye Terminator v3.1 (Applied Biosystems). Traces were examined using Sequencher 5.0 (Gene Codes). The HCV sequences have accession numbers KC012607-KC12616 (Core sequences) and KC506766-KC506776 (NS5B sequences).

### 2.3. Phylogenetic analysis

Nucleotide sequences were aligned by hand using Se-Al v2.0 together with HCV and HPgV reference sequences, obtained from the HCV sequence database (Kuiken et al., 2005) and from GenBank. Phylogenies were estimated for each alignment using maximum likelihood (ML), as implemented in GARLI v0.951 (Zwickl, 2006) under a GTR nucleotide substitution model with gamma-distributed among-site rate variation. Statistical support for phylogenetic clustering was calculated using a ML bootstrap approach with 500 bootstrap replicates; bootstrap scores were summarized using the Consense package in PHYLIP (Felsenstein, 1989). Phylogenies were visualized and annotated using FigTree (<http://tree.bio.ed.ac.uk/software/figtree>).

### 2.4. HPgV RT-PCR and sequencing

All samples that were negative for HCV antibodies were subsequently screened for HPgV through the amplification and sequencing of a 268 nt amplicon located at positions 4126–4275 in the NS3 region. Primers are listed in Table 1. RNA was extracted from 200  $\mu$ l plasma using QIAamp MinElute Virus Spin Kit (Qiagen) and converted to cDNA with random primers using Superscript III. PCR amplification was performed using Fast Start High Fidelity (Roche Applied Science) and the products were sequenced using Big Dye Terminator v3.1. Traces were examined using Sequencher 5.0. The HPgV sequences have accession numbers KC506736-KC506765.

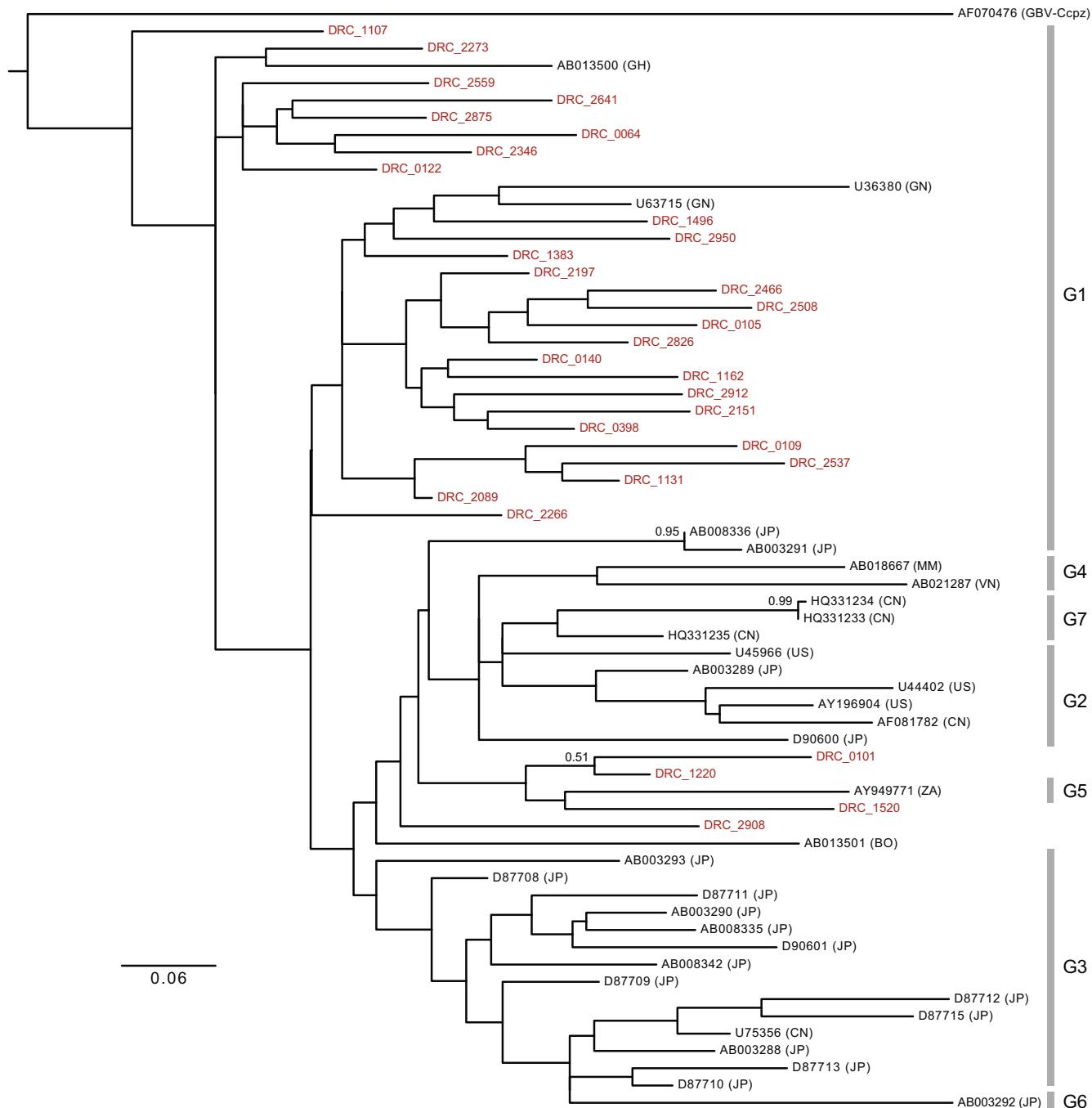
## 3. Results

### 3.1. HCV serology

Of the samples screened for anti-HCV antibodies, 13.7% were reactive (41/299). Fig. 1a shows the age distribution of the 41 HCV-seropositive samples. Samples from individuals born before 1950 were significantly more likely to be HCV seropositive than those born after that date ( $p < 0.005$ ; Fisher's exact test). The estimated seroprevalence in the three oldest age classes ranged between 25% and 57%, whereas in the remaining younger age classes the estimated seroprevalence was 6–19%. Of the 41 seropositive samples, only 24 had a signal to cut-off ratio  $>3.8$  and might be considered to represent active infections (Alter et al., 2003).

### 3.2. HCV RT-PCR and sequencing

Only about one quarter of the HCV samples that were reactive for anti-HCV antibodies contained detectable HCV RNA: among all samples 3.7% (11 out of 299) were PCR-positive. Possible reasons for this substantial difference between seropositivity and RNA-positivity are explored in the discussion. There was, however, a relationship between the ELISA signal to cut-off ratios and RNA-



**Fig. 5.** Maximum-likelihood phylogeny HPgV, estimated from partial NS3 sequences. New DRC isolates from this study are in red; other sequences are reference isolates from Genbank or the HCV Sequence Database. Numbers next to nodes represent maximum likelihood bootstrap support values; only values >0.5 are shown. Grey bars indicate HPgV genotypes 1–7; these labels are very approximate due to low phylogenetic resolution. Reference isolates are labeled with accession number and, where available, the location of sampling (ISO 3166 two-letter country codes, in parenthesis). The phylogeny is rooted using SPgVcpz (previously named GBV-C<sub>cpz</sub>) and the scale bar is in units of nucleotide substitutions per site.

positivity: no HCV RNA was recovered from samples with signal to cut-off ratios <3.6, whilst RNA was obtained from 42% (10 out of 24) samples with ratios >3.8. Four samples exhibited signal to cut-off ratios close to this threshold (between 3.0 and 3.8), of which one tested positive for HCV RNA.

For nine samples we obtained both a 1023 nt sequence from the 5'UTR-Core genome region and a 342 nt sequence from the NS5B genome region. For the remaining two samples (DRC2450 and DRC1424) we were only able to obtain a NS5B sequence. The age

distribution of HCV RNA-positive samples (Fig. 1b) matched that of the HCV seropositive samples (Fig. 1a).

### 3.3. HCV phylogenetic analysis

The HCV sequences obtained were combined with reference sequences and subjected to phylogenetic analysis. The HCV RNA-positive samples from the DRC grouped with various subtypes and lineages of genotype 4. The phylogenies estimated from the Core

(Fig. 2) and NS5B (Fig. 3) sequences gave concordant results. Three samples grouped within subtype 4c (DRC0327, DRC2141, DRC2518), three grouped within subtype 4k (DRC0095, DRC2492, DRC2497), and two grouped within subtype 4r (DRC1904, DRC1424). Isolate DRC0418 was placed immediately basal to subtype 4k in both trees (hereafter termed 4k-like). All of these clusters were supported with high maximum likelihood bootstrap scores ( $\geq 0.7$ ). In contrast the two remaining isolates (DRC2431 and DRC2450) were closely related to each other (Fig. 3; bootstrap score = 1.0) but distinct from all currently-identified subtypes. This pair may therefore represent a new unclassified lineage, although further samples are required for confirmation.

There was a significant difference in the relative frequency of the different HCV subtypes according to the age of the infected individuals (Fig. 4). Specifically, isolates classified as subtype 4k, 4l or 4c were only found in samples from individuals born before 1950, whereas those classified as subtype 4r were only found in samples from individuals born after 1956 (Fisher's exact test;  $p = 0.018$ ).

#### 3.4. HPgV RT-PCR, sequencing and analysis

The proportion of samples that were RNA-positive was higher for HPgV than for HCV, with 12.7% (33/259) of samples testing positive for HPgV RNA. For each of the positive samples we obtained a 268 nt sequence of the HPgV NS3 gene. The age distribution of HPgV RNA-positive samples (Fig. 1c) was different to that observed for HCV (Fig. 1b), as HPgV RNA was more likely to be found in samples from younger individuals than older ones. No HPgV RNA was found in individuals born before 1950 and the highest rate of HPgV positivity (>20%) was found in those born between 1976 and 1980. However, the HPgV prevalence results are difficult to interpret correctly, as only samples that were not seroreactive for HCV were available for analysis (see Discussion).

Fig. 5 shows the maximum likelihood phylogeny of the HPgV NS3 sequences obtained here. The bootstrap scores for most nodes in the tree were low and not all of the HPgV genotypes were resolved as reciprocally monophyletic groups, suggesting that the phylogenetic signal in the sequences is not strong. Most of the HPgV sequences from the DRC grouped together with genotype 1 strains. Four DRC samples were genetically distinct and clustered most closely with genotype 5 (DRC0101, DRC1220, DRC1520, DRC2908). However these groupings were not supported with high bootstrap support values.

## 4. Discussion

In an attempt to better understand the molecular epidemiology and genetic diversity of HCV in the DRC we screened 299 blood samples from the country for HCV using serological assays and PCR. An important finding of this study is the raised HCV-seropositivity and HCV RNA-positivity among males from the DRC aged 50 and over. Although overall HCV seropositivity (13.7%) was high compared to the global prevalence of HCV, previous studies have reported seropositivity of 6.4% in blood donors and 6.6% in commercial sex workers (CSWs) in Kinshasa, the capital of the DRC (Laurent et al., 2001; Tibbs et al., 1991). However, we observed a substantial difference between levels of HCV seropositivity (13.7%) and HCV RNA-positivity (3.7%), although the age profiles of these markers of infection were similar (Fig. 1a and b). A comparable difference was reported for HCV genotype 4 infections in Uganda (Biggar et al., 2006) and could result from non-specific ELISA reactivity, low detectable viraemia, or unusually high rates of clearance after infection. We did not undertake a confirmatory test (such as a RIBA immunoblot assay) to exclude non-specific reactivities

against other pathogens (Callahan et al., 1993), hence we hereafter discuss only the age distribution of HCV seroprevalence, not its absolute value.

Laurent et al. (2001) noted that HCV prevalence in CSWs increased with age, from 2.8% in those aged <20–21.3% in those aged >40. Although our sample set is comparatively small in size, its wide age range enabled us to observe that HCV prevalence in men in the DRC rises most rapidly in those aged >50 (Fig. 1a and b). Similar age profiles have been identified in neighboring countries: in several locations in south and south-eastern Cameroon HCV seroprevalence rises rapidly with age, surpassing 50% in those aged 50 and older (Nerrienet et al., 2005; Pépin et al., 2010a). HCV seropositivity also increases with age among pregnant women in Gabon, although at much lower levels (<6%; Ndong-Atome et al., 2008), and among Bantu populations in the Republic of the Congo (Cantaloube et al., 2010). Since HCV infection is far more likely to reduce than to increase survival, these results suggest that rates of HCV transmission in such populations were higher in the past. The strongest evidence that past HCV transmission can generate an age cohort-effect comes from Egypt, where parenteral anti-schistosomal therapy campaigns between 1930 and 1955 resulted in the repeated intravenous treatment of a large cross-section of the population with needles that were very likely incompletely sterilized (Strickland, 2010). Estimated levels of exposure to this treatment varied among locations and age groups, but closely matched HCV prevalence in each case (Frank et al., 2000). Today, approximately 10–20% of the Egyptian population is chronically infected by HCV and 90% of those infections are caused by subtype 4a (Arthur et al., 1997). The possible cause or causes of past iatrogenic HCV transmission in sub-Saharan Africa are more complex and less well understood (reviewed in Pépin, 2011). In Ebolowa, Cameroon, Pépin et al. (2010b) found that HCV seropositivity was associated with past intravenous treatment for malaria, but not with anti-treponemal treatment, as previously suspected (Pépin and Labbé, 2008). While HCV can be transmitted through sexual or intra-familial routes, past iatrogenic transmission better explains why HCV seropositivity varies greatly among locations within a country (Frank et al., 2000; Nerrienet et al., 2005) as well as its elevation in those aged >50 in affected locations.

The HCV sequences obtained in this study were genetically diverse; among only eleven RNA-positive isolates we found three previously-recognised HCV subtypes (4c, 4r, 4k) and an unclassified group (comprising isolates DRC2431 and DRC2450) that may constitute a new subtype. The HCV subtypes detected in our survey matched those previously observed for the DRC and surrounding countries. The three previously-reported Core gene sequences from the DRC (U10236, U10238 and U10239) belong to subtypes 4r and 4c (Fig. 2). Several other isolates classified as subtypes 4r, 4c and 4k were sampled in the Republic of Congo, Rwanda or Burundi, or were obtained from immigrants to Canada from those countries or from the DRC (Figs. 2 and 3). In contrast, HCV genotype 4 lineages described in the DRC are different from the genotype 4 lineages typically found in Egypt (e.g. subtypes 4a, 4m-o; Abdel-Hamid et al., 2007). Overall, HCV genotype 4 genetic diversity is greater in Central Africa than in Egypt and North Africa, strengthening the hypothesis that the former represents the region of origin of genotype 4 (Ndjomou et al., 2003; Pybus et al., 2007).

Interestingly, the relatively frequency of the lineages found in the DRC varied with the age of the infected individual: subtypes 4c and 4k were found in those born before 1950, whereas subtype 4r only was recovered from those born after 1956. In combination with the age-distribution results, this suggests that different routes or events (iatrogenic or otherwise) may explain HCV transmission among infected individuals of different ages. Larger sample sizes and detailed geographic information will be needed to investigate this hypothesis and we hope to address these issues in future sur-

veyes. Past rates of HCV transmission in Africa have been estimated from contemporary sequence data using coalescent-based methods (Pybus et al., 2003; Njouom et al., 2007; 2009; Ndong-Atome et al., 2008; Pépin et al., 2010a,b;). However this technique requires minimal sample sizes of ~15–20 isolates per subtype for reliable estimation and thus cannot be applied here. Historical research into specific parenteral treatment campaigns in the DRC will also be crucial in reconstructing the epidemic history of the HCV in the region (Pépin, 2011).

It is interesting to compare the HCV results with those obtained for human pegivirus (HPgV) from the same set of samples. Globally, 1–4% of healthy blood donors are viremic for HPgV and another 13% carry anti-HPgV antibodies (Blair et al., 1998; Gutierrez et al., 1997; Pilot-Matias et al., 1996; Tacke et al., 1997). Here, we found HPgV RNA in 12.7% of samples, consistent with a previous survey that reported viremia in 10.3% of pregnant women in Kinshasa (Liu et al., 1999). We could not assess the frequency of HCV/HPgV co-infection because due to limited sample volume only those samples that were HCV seronegative were screened for HPgV. This also prevents us from drawing firm conclusions from the age-distribution of HPgV viremia in our samples (Fig. 1c): HCV and HPgV are both efficiently transmitted via infected blood (Bhattarai and Stapleton, 2012) and therefore the removal of HCV seropositive samples may have biased downwards our estimate of HPgV prevalence in those aged >50. Most of the HPgV viruses we obtained clustered with genotype 1, which is typical of African HPgV strains, whilst a minority was more closely related to genotype 5. These classifications are not conclusive, however, because our HPgV phylogeny was not supported with high bootstrap values.

Although the work here represents the largest study to date of HCV molecular epidemiology in the DRC – the largest country in sub-Saharan Africa – it is still limited by a comparatively small sample size. Through virus genome sequencing, in addition to serological testing, we were able to provide insights into the epidemiology of HCV. Despite the limitations of sample size, differences in HCV positivity among those born before and after 1950 were significant and we expect this cohort effect to be observed in more comprehensive surveys undertaken in future. Lastly, this study illustrates that the importance of the HCV epidemic in sub-Saharan Africa and the need to plan for the aging cohort. The prevalence and diversity of HCV and related viruses in some of the largest and most populous countries in Africa (such as the DRC, Nigeria, Sudan, Ethiopia, Kenya and Angola) remain poorly studied, highlighting the need for broader surveys in this important region.

## Acknowledgements

We thank the Ministry of Public Health and the Ministry of National Defense of the Democratic Republic of Congo for permission to undertake this study and the U.S. Embassy of the DRC for their continued support. This project was financially supported by Global Viral. PK & GLAH were supported by The Wellcome Trust, Oxford Martin School, and NIHR Biomedical Research Centre, Oxford. CFD was supported in part by the NIH Fogarty International Center AIDS International Training and Research Program (2 D 43 TW000010-16/17). NDW is supported by the NIH Director's Pioneer Award (DP1-OD000370). Metabiota and Global Viral are graciously supported by the U.S. Department of Defense Armed Forces Health Surveillance Center, Division of Global Emerging Infections, Surveillance Operations (AFHSC GEIS), The Defense Threat Reduction Agency Cooperative Biological Engagement Program (DTRA-CBEP), Google.org, The Skoll Foundation, DoD HIV/AIDS Prevention Program (DHAPP), the Henry M. Jackson Foundation for the Advancement of Military Medicine, and the U.S. Agency for International Development (USAID) Emerging Pandemic

Threats Program PREDICT project, under the terms of Cooperative Agreement Number GHN-A-00-09-00010-00.

## References

- Abdel-Hamid, M., El-Daly, M., Molnegren, V., El-Kafrawy, S., Abdel-Latif, S., Esmat, G., Strickland, G.T., Loffredo, C., Albert, J., Widell, A., 2007. Genetic diversity in hepatitis C virus in Egypt and possible association with hepatocellular carcinoma. *Journal of General Virology* 88, 1526–1531.
- Alter, M.J., Kuhnert, W.L., Finelli, L., 2003. Guidelines for laboratory testing and result reporting of antibody to hepatitis C virus. Centers for Disease Control and Prevention – Morbidity and Mortality Weekly Report Recommendations and Reports 52, 1–13.
- Arthur, R.R., Hassan, N.F., Abdallah, M.Y., El-Sharkawy, M.S., Saad, M.D., Hackbart, B.G., Imam, I.Z., 1997. Hepatitis C antibody prevalence in blood donors in different governorates in Egypt. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 91, 271–274.
- Basaras, M., Santamaría, A., Sarsa, M., Gutiérrez, E., de Olano, Y., Cisterna, R., 1999. Seroprevalence of hepatitis B and C, and human immunodeficiency type 1 viruses in a rural population from the republic of equatorial Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 93, 250–252.
- Batina Agasa, S., Dupont, E., Kayembe, T., Molima, P., Malengela, R., Kabemba, S., Andrien, M., Lamberton, M., Cotton, F., et al., 2010. Multiple transfusions for sickle cell disease in the Democratic Republic of Congo: the importance of the hepatitis C virus. *Transfusion Clinique et Biologique* 17, 254–259.
- Berkes, J., Cotler, S.J., 2005. Global epidemiology of HCV infection. *Current Hepatitis Reports* 4, 125–130.
- Bhattarai, N., Stapleton, J.T., 2012. GB virus C: the good boy virus? *Trends in Microbiology* 20, 124–130.
- Biggar, R.J., Ortiz-Conde, B.A., Bagni, R.K., Bakaki, P.M., Wang, C.D., Engels, E.A., Mbulaiteye, S.M., Ndugwa, C.M., 2006. Hepatitis C virus genotype 4 in Ugandan children and their mothers. *Emerging Infectious Diseases* 12, 1440–1443.
- Blair, C., Davidson, F., Lycett, C., McDonald, D., Haydon, G., Yap, P.L., Hayes, P., Simmonds, P., Gillon, J., 1998. Prevalence, incidence, and clinical characteristics of hepatitis G virus/GB virus C infection in Scottish blood donors. *Journal of Infectious Diseases* 178, 1779–1782.
- Burbelo, P.D., Dubovi, E.J., Simmonds, P., Medina, J.L., Henriquez, J.A., Mishra, N., Wagner, J., Tokarz, R., Cullen, J.M., et al., 2012. Serology-enabled discovery of genetically diverse hepaciviruses in a new host. *Journal of Virology* 86, 6171–6178.
- Callahan, J.D., Constantine, N.T., Kataaha, P., Zhang, X., Hyams, K.C., Bansal, J., 1993. Second generation hepatitis C virus assays: performance when testing African sera. *Journal of Medical Virology* 41, 35–38.
- Candotti, D., Temple, J., Sarkodie, F., Allain, J.P., 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *Journal of Virology* 77, 7914–7923 (Am. Soc. Microbiol.).
- Cantaloube, J.-F., Gallian, P., Bokilo, A., Jordier, F., Biagini, P., Attoui, H., Chiaroni, J., de Micco, P., 2010. Analysis of hepatitis C virus strains circulating in Republic of the Congo. *Journal of Medical Virology* 82, 562–567.
- Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., Houghton, M., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244, 359–362.
- Delaporte, E., Thiers, V., Dazza, M.C., Romeo, R., Mlika-Cabanne, N., Aptel, I., Schrijvers, D., Bréchet, C., Larouzé, B., 1993. High level of hepatitis C endemicity in Gabon, equatorial Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 87, 636–637.
- Djoko, C.F., Rimoin, A.W., Vidal, N., Tamoufe, U., Wolfe, N.D., Butel, C., LeBreton, M., Tshala, F.M., Kayembe, P.K., et al., 2011. High HIV type 1 group M pol diversity and low rate of antiretroviral resistance mutations among the uniformed services in Kinshasa, Democratic Republic of the Congo. *AIDS Research and Human Retroviruses* 27, 323–329.
- Felsenstein, J., 1989. PHYLIP – phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.
- Frank, C., Mohamed, M.K., Strickland, G.T., Lavanchy, D., Arthur, R.R., Magder, L.S., El Khobay, T., Abdel-Wahab, Y., Aly Ohn, E.S., et al., 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* 355, 887–891.
- Gutierrez, R.A., Dawson, G.J., Knigge, M.F., Melvin, S.L., Heynen, C.A., Kyk, C.R., Young, C.E., Carrick, R.J., Schlauder, G.G., et al., 1997. Seroprevalence of GB virus C and persistence of RNA and antibody. *Journal of Medical Virology* 53, 167–173.
- Jeannel, D., Fretz, C., Traore, Y., Kohdjo, N., Bigot, A., Pê Gamy, E., Jourdan, G., Kourouma, K., Maertens, G., et al., 1998. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *Journal of Medical Virology* 55, 92–97.
- Kapoor, A., Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J., Firth, C., Hirschberg, D., Rice, C., Shields, S., Lipkin, W., 2011. Characterization of a canine homolog of hepatitis C virus. *Proceedings of the National Academy of Sciences of the United States of America* 108, 11608–11613.
- Kane, A., Lloyd, J., Zaffran, M., Simonsen, L., Kane, M., 1999. Transmission of hepatitis B, hepatitis C and human immunodeficiency viruses through unsafe injections in the developing world: model-based regional estimates. *Bulletin of the World Health Organization* 77, 801–807.

- Kuiken, C., Yusim, K., Boykin, L., Richardson, R., 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics* 21, 379–384.
- Laurent, C., Henzel, D., Mulanga-Kabeya, C., Maertens, G., Larouze, B., Delaporte, E., 2001. Seroepidemiological survey of hepatitis C virus among commercial sex workers and pregnant women in Kinshasa, Democratic Republic of Congo. *International Journal of Epidemiology* 30, 872–877.
- Liu, H.-F., Muyembe-Tamfum, J.-J., Dahan, K., Desmyter, J., Gouba, P., 1999. High prevalence of GB virus C/hepatitis G virus in Kinshasa, Democratic Republic of Congo: a phylogenetic analysis. *Journal of Medical Virology* 60, 159–165.
- Lyons, S., Kapoor, A., Sharp, C., Schneider, B.S., Wolfe, N.D., Culshaw, G., Corcoran, B., McGorum, B.C., Simmonds, P., 2012. Nonprimate Hepaciviruses in Domestic Horses, United Kingdom. *Emerging Infectious Diseases* 18, 1976–1982.
- Markov, P.V., Pépin, J., Frost, E., Deslandes, S., Labbe, A.-C., Pybus, O.G., 2009. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *Journal of General Virology* 90, 2086–2096.
- Murphy, D.G., Willems, B., Deschenes, M., Hilzenrat, N., Mousseau, R., Sabbah, S., 2007a. 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. *Journal of Clinical Microbiology* 45, 1102–1112.
- Murphy, D., Chamberland, J., Dandavino, R., Sablon, E., 2007b. A new genotype of hepatitis C virus originating from Central Africa. *Hepatology* 46, 623A.
- Ndjomou, J., Pybus, O.G., Matz, B., 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *Journal of General Virology* 84, 2333–2341.
- Ndongo-Atome, G.-R., Makuwa, M., Njouom, R., Branger, M., Brun-Vézinet, F., Mahé, A., Rousset, D., Kazanji, M., 2008. Hepatitis C virus prevalence and genetic diversity among pregnant women in Gabon, Central Africa. *BMC Infectious Diseases* 8, 82.
- Nerrienet, E., Pouillot, R., Lachenal, G., Njouom, R., Mfoupouendoun, J., Bilong, C., Mauclere, P., Pasquier, C., Ayouba, A., 2005. Hepatitis C virus infection in Cameroon: a cohort-effect. *Journal of Medical Virology* 76, 208–214.
- Njouom, R., Frost, E., Deslandes, S., Mamadou-Yaya, F., Labbé, A.-C., Pouillot, R., Mbélessso, P., Mbadingai, S., Rousset, D., Pépin, J., 2009. Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. *Journal of General Virology* 90, 2452–2456.
- 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. *Infection, Genetics and Evolution* 7, 361–367.
- Pépin, J., 2011. The Origins of AIDS. Cambridge University Press, UK.
- Pépin, J., Labbé, A.-C., 2008. Noble goals, unforeseen consequences: control of tropical diseases in colonial Central Africa and the iatrogenic transmission of blood-borne viruses. *Tropical Medicine and International Health* 13, 744–753.
- Pépin, J., Labbé, A.-C., Mamadou-Yaya, F., Mbélessso, P., Mbadingai, S., Deslandes, S., Locas, M.-C., Frost, E., 2010a. Iatrogenic transmission of human T cell lymphotropic virus type 1 and hepatitis C virus through parenteral treatment and chemoprophylaxis of sleeping sickness in colonial equatorial Africa. *Clinical Infectious Diseases* 51, 777–784.
- Pépin, J., Lavoie, M., Pybus, O.G., Pouillot, R., Fouppouapouognigni, Y., Rousset, D., Labbé, A.-C., Njouom, R., 2010b. Risk factors for hepatitis C virus transmission in colonial Cameroon. *Clinical Infectious Diseases* 51, 768–776.
- Pilot-Matias, T.J., Carrick, R.J., Coleman, P.F., Leary, T.P., Surowy, T.K., Simons, J.N., Muerhoff, A.S., Buijk, S.L., Chalmers, M.L., et al., 1996. Expression of the GB virus C E2 glycoprotein using the semliki forest virus vector system and its utility as a serologic marker. *Virology* 225, 282–292.
- de Pondé, R.A., 2011. Hidden hazards of HCV transmission. *Medical Microbiology and Immunology* 200, 7–11.
- 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. *Molecular Biology and Evolution* 20, 381–387.
- Pybus, O.G., Charleston, M.A., Gupta, S., Rambaut, A., Holmes, E.C., Harvey, P.H., 2001. The epidemic behavior of the hepatitis C virus. *Science* 292, 2323–2325.
- Pybus, O.G., Markov, P.V., Wu, A., Tatem, A.J., 2007. Investigating the endemic transmission of the hepatitis C virus. *International Journal for Parasitology* 37, 839–849.
- Ray, S.C., Arthur, R.R., Carella, A., Bukh, J., Thomas, D.L., 2000. Genetic epidemiology of hepatitis C virus throughout Egypt. *Journal of Infectious Diseases* 182, 698–707.
- Sharp, C.P., Vermeulen, M., Nébié, Y., Djoko, C.F., LeBreton, M., Tamoufe, U., Rimoïn, A.W., Kayembe, P.K., Carr, J.K., et al., 2010. Epidemiology of human parvovirus 4 infection in sub-Saharan Africa. *Emerging Infectious Diseases* 16, 1605–1607.
- Shepard, C.W., Finelli, L., Alter, M.J., 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infectious Diseases* 5, 558–567.
- Simmonds, P., 2004. Genetic diversity and evolution of hepatitis C virus—15 years on. *Journal of General Virology* 85, 3173–3188.
- Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L., Simmonds, P., 1997. The origin of hepatitis C virus genotypes. *Journal of General Virology* 78, 321–328.
- Stapleton, J.T., Foung, S., Muerhoff, A.S., Bukh, J., Simmonds, P., 2011. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *Journal of General Virology* 92, 233–246.
- Strickland, G.T., 2010. An epidemic of hepatitis C virus infection while treating endemic infectious diseases in equatorial Africa more than a half century ago: did it also jump-start the AIDS pandemic? *Clinical Infectious Diseases* 51, 785–787.
- Tacke, M., Schmolke, S., Schlueter, V., Sauleda, S., Esteban, J.I., Tanaka, E., Kiyosawa, K., Alter, H.J., Schmitt, U., et al., 1997. Humoral immune response to the E2 protein of hepatitis C virus is associated with long-term recovery from infection and reveals a high frequency of hepatitis C virus exposure among healthy blood donors. *Hepatology* 26, 1626–1633.
- Tibbs, C.J., Palmer, S.J., Coker, R., Clark, S.K., Parsons, G.M., Hojvat, S., Peterson, D., Banatvala, J.E., 1991. Prevalence of hepatitis C in tropical communities: the importance of confirmatory assays. *Journal of Medical Virology* 34, 143–147.
- Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.