Short communication

Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon

Régis Pouillot\(^a\), Guillaume Lachenal\(^b\), Oliver G. Pybus\(^c\), Dominique Rousset\(^a\), Richard Njouom\(^a,\)*

\(^a\) Centre Pasteur du Cameroun, Yaounde, Cameroon
\(^b\) Centre d’Etude des Mondes Africains, CNRS, Paris, France
\(^c\) Department of Zoology, University of Oxford, Oxford, UK

1. Introduction

Hepatitis C virus (HCV) infection is a major public health problem worldwide. The World Health Organization (WHO) estimates that about 3% of the world’s population (nearly 170 million people) is chronically infected with HCV (WHO, 1999). Africa is reported to have the highest regional HCV prevalence rate (5.3%) (Madhava et al., 2002). HCV is a highly genetically variable virus that has been classified into six phylogenetically distinct genotypes, each containing multiple subtypes (Simmonds et al., 2005). There are marked differences in the distributions of the genotypes and subtypes worldwide; some of them are distributed globally, whereas others are found only in specific geographic regions. Genotypes 1, 2, and 4 appear to be endemic to regions of West and Middle Africa and the Middle East, whereas divergent endemic strains of genotypes 3 and 6 are found in Southeast Asia (McOmish et al., 1994). It is not yet clear where the region of endemic infection for genotype 5 is located (Verbeeck et al., 2006).

Advances in the evolutionary analysis of sampled viral gene sequences have enabled the reconstruction of the past virus transmission trends from contemporary infections (Drummond et al., 2005; Nakano et al., 2004; Pybus et al., 2001, 2003; Tanaka et al., 2002, 2004). Recently, a Bayesian coalescent approach (Drummond et al., 2005) was used to estimate the dates of origin and the rates of virus spread through time for HCV genotypes 1, 2 and 4 found in Cameroon and Middle Africa (Njouom et al., 2007). By comparing the dates of the most recent common ancestor (MRCA) of the Cameroon HCV-2 strains, 1630 (95% highest posterior density interval: 1470–1760) was slightly more recent than that of West Africa, 1540 (95% highest posterior density interval: 1380–1680). Estimates of epidemic history indicate significant differences between the two strains. HCV-2 appears to have spread relatively slowly within the West African population from 1630 to 1900, whilst the Cameroon lineages exhibit rapid, exponential spread from 1920 to 1960. This comparative genetic analysis indicates that Cameroon HCV-2 strains are derived from West African strains and that HCV-2 has undergone radically different epidemiological histories in the two regions.

* Corresponding author at: Laboratoire de Virologie, Centre Pasteur du Cameroun, PO Box 1274, Yaounde, Cameroon.
E-mail address: njouom@pasteur-yaounde.org (R. Njouom).
2. Materials and methods

Two HCV-2 datasets of partial NS5B gene sequences were used in this study. The first dataset contains 58 sequences, 382 or 405 bp long, from HCV-infected patients from Cameroon (Ndjomou et al., 2003; Njouom et al., 2003, 2005; Pasquier et al., 2005). The second dataset was comprised of 40 sequences, 200, 270 or 340 bp long, from HCV-infected patients residing in four different West African countries (Fig. 1A) (Candotti et al., 2003; Jeannel et al., 1998; Ruggieri et al., 1996). No detailed information was available concerning the country of origin or migration history of the sampled patients for this dataset. Samples were collected over the year period 1993–1999 for the West African strains. Table 1 gives a description of the sequences used in this study.

Using both datasets, initial multiple nucleotide alignments were obtained using ClustalX 1.81 (Thompson et al., 1997) and subsequently adjusted by hand. Maximum likelihood phylogenetic trees were inferred using the PHYML 2.4.4 software (http://atgc.lirmm.fr/phyml) (Guindon and Gascuel, 2003). A number of different substitution models were tested using an adaptation of the PHYMLTEST function implemented in the APE R library (http://pbil.univ-lyon1.fr/R/ape) (Paradis et al., 2004). The models tested included JC69 (Jukes and Cantor, 1969), K2P (Jin and Nei, 1990; Kimura, 1980), F81 (Felsenstein, 1981), F84 (Felsenstein and Churchill, 1996), HKY85 (Hasegawa et al., 1985), TN93 (Tamura and Nei, 1993) and GTR (Lanave et al., 1984), with or without account the variation in evolutionary rate among lineages and West Africa were investigated using Bayesian Monte Carlo Markov Chain (MCMC) analysis, as implemented in the BEAST 1.4.2 software (Drummond and Rambaut, 2007). Because the sequences were sampled over a short period of time, they contained insufficient information to accurately co-estimate their evolutionary rate. We therefore used an informative prior distribution, with a mean of $5.0 \times 10^{-4}$ and standard deviation of $1.7 \times 10^{-5}$. This prior distribution represents a best estimate of HCV NS5B evolutionary rates, as obtained from two independent prior analyses (Pybus et al., 2001; Tanaka et al., 2002). As recommended, a relaxed molecular clock approach (uncorrelated lognormal model) was used, thereby taking into account the variation in evolutionary rate among lineages (Drummond et al., 2006). The MCMC was run long enough to obtain an effective sampling size (ESS, Drummond et al., 2007) of at least $>300$ for all parameters. BEAST output files were analysed using TRACER 1.3. (http://tree.bio.ed.ac.uk/software/tracer/). Bayes Factors (BF) were used to choose the most statistically appropriate demographic model. A Bayesian Skyline Plot (BSP) with 5 linear steps performed better than the logistic and expansion parametric models (BF = 73.0 for West Africa; BF = 74.3 for Cameroon), and also performed better than a BSP with 10 linear steps (BF = 1226 for West Africa; BF > 6.7 for Cameroon).

3. Results and discussion

The model with the smallest AIC score was the GTR model with gamma-distributed rate variation plus invariant sites (Table 2). Consequently, this model was used in the rest of the study (phylogenetic and BEAST analyses). Fig. 1B shows the estimated phylogeny of the HCV NS5B sequences from Cameroon and West African countries. One sequence from Ghana (GH07-CA, referred in the original article as G4720, Candotti et al., 2003) is highly divergent and grouped with the outgroup strains. It has been omitted from Fig. 1B for clarity. Candotti et al. (2003) previously concluded this strain to be a divergent genotype 2 lineage with low bootstrap support. As the basal position of this strain cannot be reliably confirmed without longer sequence data, we chose to discard this strain from subsequent BEAST analyses.

The phylogeny in Fig. 1B shows two distinct clusters at the intercept of the Cameroon strain CA13-NJ and the West African strain BF10-JE. The West African samples form one cluster, and within this the Cameroon samples form a second, monophyletic group. The separation of these two groups was supported by a good

---

Table 1

<table>
<thead>
<tr>
<th>Reference</th>
<th>Origin</th>
<th>n</th>
<th>Year of sampling</th>
<th>Sequence length (bp)</th>
<th>GenBank accession number</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Africa (n = 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ruggieri et al., 1996)</td>
<td>Republic of Guinea</td>
<td>5</td>
<td>1993</td>
<td>200</td>
<td>X93323, X93324, X93325, X93326, X93327</td>
<td>GC01-RU to GC05-RU</td>
</tr>
<tr>
<td>(Candotti et al., 2003)</td>
<td>Ghana</td>
<td>19</td>
<td>1999</td>
<td>270</td>
<td>AY236367, AY236368, AY236369, AY236371, AY236372, AY236373, AY236374, AY236375, AY236376, AY236378, AY236379, AY236380, AY236381, AY236382, AY236383, AY236384, AY236385, AY236386, AY236387</td>
<td>GH01-CA to GH19-CA</td>
</tr>
<tr>
<td>(Jeannel et al., 1998)</td>
<td>Burkina Faso</td>
<td>10</td>
<td>1994–1995</td>
<td>340</td>
<td>AF037240, AF037245, AF037246, AF037247, AF037248 to, AF037254, BF01-JE to BF10-JE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benin</td>
<td>5</td>
<td>1994–1995</td>
<td>340</td>
<td>AF037239, AF037240, AF037241, AF037242, AF037243, AF037244</td>
<td>BN01-JE to BN05-JE</td>
</tr>
<tr>
<td></td>
<td>Republic of Guinea</td>
<td>1</td>
<td>1994–1995</td>
<td>340</td>
<td>AF037247</td>
<td>GC06-JE</td>
</tr>
<tr>
<td>Cameroon (n = 58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Njouom et al., 2003, 2005)</td>
<td>Cameroon</td>
<td>51</td>
<td>2000–2003</td>
<td>382</td>
<td>AY265420, AY265422, AY265423, AY265426, AY265427, AY265433, AY265436, AY265437, AY265442, AY265443, AY265444, AY265445, AY265451, AY685018, AY685033, AY685048, AY632154 to AY632199, AY257103, AY257086, AY257100, AY257101, AY257089, AY257093, AY257082</td>
<td>CA01-NJ to CA51NJ</td>
</tr>
<tr>
<td>(Ndjomou et al., 2003)</td>
<td>Cameroon</td>
<td>7</td>
<td>1998</td>
<td>405</td>
<td></td>
<td>CA52-NJ to CA58-NJ</td>
</tr>
</tbody>
</table>
Fig. 1. (A) Map of West Africa and Cameroon, showing sequence sampling locations (grey stars for West Africa, black stars for Cameroon). The white circles indicate the locations of smallpox inoculation practices, as identified by written sources from Europe dating from the 18th, 19th and 20th centuries (adapted from Herbert, 1975) and including the sites specified in Jubé-Beaulaton and Lainé, 2005; Grischow, 2004; Houillon, 1905). (B) Maximum likelihood phylogeny, representing the West African (black) and Cameroonian (green) partial NS5B sequences (267 nucleotides). Sequence names begin with two letters that denote the country of sequence isolation: BF: Burkina Faso,
bootstrap score. However, one Cameroonian strain (CA38-NJ) was more closely related to West African samples than to the remaining Cameroonian samples. Epidemiological information indicates a West African origin for this patient. It is therefore likely that this patient acquired the infection in West Africa before moving to Cameroon. Since this strain is more typical of West African infection, it was not included in the subsequent BEAST analyses of the Cameroonian sequences. Conversely, two West African strains (strains GC01-RU and GC05-RU) grouped within the “Cameroonian” cluster. In the absence of epidemiological information of these patients, no definitive conclusion could be made on these strains. They were kept for subsequent analyses as “West African strains”.

To investigate the origin and spread of HCV-2 in both populations more carefully, we estimated divergence dates and epidemic history using a Bayesian coalescent approach. The Cameroonian and the West African strains were analysed separately. The date of the MRCA of the West African strains was estimated to be 1536 (95% highest posterior density interval (HPD): 1384–1675). This supports the notion of long term endemic HCV-2 transmission in West Africa (Candotti et al., 2003; Jeannel et al., 1998; Ruggieri et al., 1996; Wansbrough-Jones et al., 1998). The date of the MRCA of the Cameroonian strains was more recent, estimated at 1629 (HPD: 1465–1755). This confirms the previous suggestion that this genotype was introduced to Middle Africa from West Africa (Njouom et al., 2005, 2007). The date of the MRCA of the Cameroonian strains is earlier than that obtained in our previous study (Njouom et al., 2007), which incorrectly included the CA38-NJ sequence of West African origin. It could be argued that our evolutionary rate value, which was estimated from genotype 1 datasets, may not exactly match that of genotype 2. However, previous work demonstrates that genotype 1 evolutionary rates correctly date known historical events when applied to other genotypes (Pybus et al., 2003). In addition, the absolute evolutionary rate used will not affect our main result, of relatively earlier appearance of HCV-2 in West Africa.

The Bayesian Skyline Plots (Fig. 2) depict the estimated change in the effective number of infections through time. There are obvious differences between the patterns obtained for Cameroon and for West Africa. In West Africa, HCV seems to have spread relatively slowly from 1630 to 1900, after which the effective population size remained stable until present. Any apparent slight decrease is not significant, given the size of the confidence intervals. The BSP fits better than a parametric logistic growth model (Bayes Factor: 73). In Cameroon, as already noted for genotype 2 (but also confirmed for genotypes 1 and 4) (Njouom et al., 2007), the population size initially seems stable, before switching to exponential increase between 1920 to 1960. After 1960, the growth rate appears to decrease. The BSP fits better than a parametric expansion growth model (Bayes Factor: 74). These results indicate that the epidemiological histories of HCV-2 in Cameroon and West Africa are substantially different.

The routes of long-term endemic HCV transmission in tropical regions are not known; current plausible hypotheses include vertical transmission, exposure to blood through cultural practices, or mechanical insect transmission (Pybus et al., 2007). In contrast, more recent epidemic transmission in Africa, such as that reported for Egypt (Pybus et al., 2003), has been linked to unsafe medical injections (Frank et al., 2000). Our findings should be interpreted in the context of these patterns. In Cameroon, it has been suggested that the rapid increase in HCV effective population size during the early twentieth century is linked to widespread vaccination and/or treatment campaigns during the colonial period. HCV could have been massively transmitted by non-sterile injections and/or blood sampling during campaigns for the treatment of trypanosomiasis (Nerrienet et al., 2005; Njouom et al., 2007) and of yaws and syphilis (Pepin and Labbe, 2008). The absence of recent exponential growth in West Africa could be explained by differential exposure to such campaigns. Trypanosomiasis was
only sporadic in Cotounou (South of Dahomey/Benin) and Conakry (Coastal French Guinea/Republic of Guinea) (Bado, 1996). In British Gold Coast/Ghana, the fight against sleeping sickness typically used entomological (rather than medical) methods (Grischow, 2004; Grischow and Morris, 2006; Patterson, 1981). It seems likely that these regions have been comparatively less exposed to injection campaigns than the populations of South, Central and East Cameroon. However, trypanosomiasis treatment campaigns were quite intense from 1930 to 1960 in the region of Bobo-Dioulasso (Haute-Volta/ Burkina-Faso) (Bado, 1996). Following Pepin and Labbe’s emphasis on the role of yaws treatment in the spread of HCV in Cameroon (Pepin and Labbe, 2008), it can also be hypothesized that a lower incidence of treponematoses (or a lesser availability of arsenical drugs) in the regions of West Africa we studied could account for the absence of HCV exponential growth in the period 1930–1960. Only further epidemiological investigation and quantitative comparison of exposure history will help to resolve these issues.

The second major difference between the two regions is the apparent higher growth rate in West Africa in the 18th and 19th centuries. Several possible explanations can be put forward for this pattern. First, the difference could reflect differential human population growth rates in the past. Second, past HCV-2 spread in West Africa may represent the natural, gradual increase in endemic transmission which only ends once equilibrium prevalence is reached, as previously suggested for HCV genotypes 4 and 6 (Pybus et al., 2001). Third, the regions may differ in the frequency of historical practices that involved blood exposure. For example, arm-to-arm smallpox inoculation was extensively practiced in most areas of West Africa, whereas the technique was unknown in south Cameroon and in the Congo Basin area (Herbert, 1975). Smallpox inoculation involved serial exchange of blood and lymph between individuals, thus constituting a potentially efficient mean of transmitting HCV. Written evidences document collective practice of smallpox inoculation between 1700 and 1900 in what is now Benin, Ghana, Guinea and Burkina-Faso (Grischow, 2006; Herbert, 1975; Houillon, 1905; Juhe-Baualton and Lainé, 2005; Patterson, 1981) (Fig. 1A). The practice was gradually abandoned in the 20th century when colonial medical services introduced smallpox vaccination (Herbert, 1975). Further research on the epidemiological history of HCV in Africa is needed to understand the significance of this putative transmission route, which may have contributed – along with other routes – to the pre-20th century spread of HCV.

The high prevalence of HCV infection in Africa (Madhava et al., 2002) highlights the importance of public health efforts and plans to limit HCV spread. Future spread could be limited by ensuring the safety of blood products and controlling nosocomial risk of transmission. We also note that the unrecognized routes of transmission that maintained endemic infection prior to the 20th century have yet to be identified. There is thus a need of comprehensive studies of risk factors for acute hepatitis C, both in Cameroon and elsewhere.

In conclusion, a comparative genetic analysis of HCV-2 in West Africa and Cameroon shows that the Cameroon lineage was derived from West African strains, and that these two regions exhibit radically different epidemic histories of HCV that require further investigation.

Acknowledgements

We thank all the scientists who provided free all the software used in this study. This work was supported by the “Centre Pasteur du Cameroun”.

References


Jin, L., Nei, M., 1990. Limitations of the evolutionary parsimony method of phylo-


