HHS Public Access

Author manuscript

J Gen Virol. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

J Gen Virol. 2015 August ; 96(8): 2157–2164. doi:10.1099/vir.0.000153.

A re-evaluation of the origin of hepatitis C virus genotype 2 in West Africa

Michael A. Purdy[1], **Joseph C. Forbi**[1], **Amanda Sue**[1], **Jennifer E. Layden**[2], **William M. Switzer**[3], **Ohene K. Opare-Sem**[4], **Richard O. Phillips**[4], and **Yury E. Khudyakov**[1]

[1]Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

[2]Department of Public Health Sciences, Loyola University, Chicago, IL 60660, USA

[3]Division of HIV/AIDS, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

[4]Department of Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Abstract

Hepatitis C virus (HCV) is classified into seven genotypes based on genetic diversity, and most genotypes have been found in Africa. Infections with HCV genotype 2 (HCV2) are most prevalent in West Africa and it was suggested that HCV2 originated in West Africa. To better understand the evolutionary epidemiology of HCV2 in Africa, we examined new NS5B sequences of HCV2 strains obtained from Côte d'Ivoire, Ghana and Nigeria sequenced at the Centers for Disease Control and Prevention with those available from West, North and Central Africa. Bayesian phylogeographic analysis using a discrete trait model showed that Ghana was the most likely geographical region for the origin of HCV2. Spread of HCV2 from Ghana did not appear to be through diffusion to adjacent countries along the coast. Rather, it was transmitted from Ghana to many distant countries in Africa, suggesting that certain routes of geographical dissemination were historically more efficient than mere proximity and that the HCV2 epidemic history in West Africa is extremely complex.

INTRODUCTION

Infection with hepatitis C virus (HCV) is a serious global public health problem. Chronic HCV infection may progress to cirrhosis, leading to severe complications including hepatocellular carcinoma and end-stage liver failure (EASL, 1999; Itskowitz, 2007). It is estimated that ~170 million individuals or ~2.8% of the world's population are chronically infected and 3–4 million new cases of hepatitis C occur every year (WHO, 2012). Countries in Africa and Asia have the highest infection rates with estimated regional prevalences of 1.6–4.5% (Hanafiah *et al.*, 2013).

HCV is a small, positive-sense, ssRNA virus in the genus *Hepacivirus*, family *Flaviviridae* (Choo *et al.*, 1991; Lindenbach & Rice, 2005). The HCV genome is ~9600 nt and encodes a

Correspondence Michael A. Purdy MPurdy@cdc.gov.

One supplementary figure and one table are available with the online Supplementary Material.

single polyprotein with the following gene order: 5′-core–envelope (E1–E2)–non-structural proteins [NS1(p7)–NS2–NS3–NS4A–NS4B–NS5A–NS5B]-3′. The HCV genome is extensively heterogeneous. Currently, HCV is classified into seven genotypes that share between 65 and 70% nucleotide identity, each of which is further divided into subtypes and displays different global geographical distributions (Nakano *et al.*, 2012; Smith *et al.*, 2014).

Candotti *et al.* (2003) analysed the viral diversity of HCV genotype 2 (HCV2) sequences from Ghana. These sequences were found to have greater diversity than was found in other regions around the world, suggesting that HCV2 originated in West Africa. Furthermore, HCV2 sequences from Cameroon were found to be less diverse, suggesting an introduction of HCV2 from West Africa (Pasquier *et al.*, 2005). This finding was confirmed through the use of Bayesian analysis that showed HCV2 sequences in Cameroon had diverged more recently from their most recent common ancestor (MRCA) than had sequences from West Africa. Additionally, the Cameroon HCV2 sequences were inferred to have undergone exponential expansion between 1920 and 1960, suggesting rapid epidemic growth due to unsafe medical practices, whilst HCV2 spread relatively slowly in West Africa (Njouom *et al.*, 2007; Pouillot *et al.*, 2008). Phylogenetic analysis using an expanded set of sequences confirmed the eastward diffusion of HCV2 that took place over several centuries in West Africa. This spread appeared to be between adjacent countries, and sequences from Guinea-Bissau were most phylogenetically basal and genetically diverse, suggesting that HCV2 may have originated in this region of West Africa (Markov *et al.*, 2009).

Whilst these earlier studies increased our understanding of the evolutionary history of HCV2, they may have been limited by undersampling from other African countries where HCV2 is prevalent. Here, we report the genetic analysis of HCV2 sequences from several African countries, including those recently obtained from Côte d'Ivoire, Ghana and Nigeria in West Africa, showing that Ghana is the most likely origin for HCV2.

RESULTS

Genetic diversity

To analyse the evolution of HCV2 in West Africa, sequences from Côte d'Ivoire, Ghana and Nigeria that were sequenced at the Centers for Disease Control and Prevention were added to West African sequences that were analysed previously (Candotti *et al.*, 2003; Markov *et al.*, 2009; Pasquier *et al.*, 2005; Pouillot *et al.*, 2008) (Table S1, available in the online Supplementary Material). As slightly different regions in the NS5B gene were sequenced in these studies, it was decided to create two sequence alignments for analysis. The first, trim1, contained the maximum number of sequences (n=300) compatible in length with the dataset from Markov *et al.* (2009), whilst the second, trim2 (n=340) included 40 additional sequences from Guinea-Bissau, but the length of the trim1 sequences were shortened 12 nt to align with the Guinea-Bissau sequences. As the results from trim1 were in agreement with those from trim2, only the trim2 results are presented in this paper.

Analysis of trim2 sequence diversity among HCV2 sequences showed that sequences from Ghana were the most diverse. Sequences from Burkina Faso, Cameroon, Guinea-Bissau, Morocco, Nigeria and Tunisia had a mean sequence diversity of 0.26 (SD 0.03), whilst those

from Ghana had a diversity of 0.45. In fact, the diversity seen in Ghana was greater than seen among all other countries combined (0.32).

Molecular clock

In order to determine the best molecular clock for this analysis, the strict, log-normal and exponential clocks were tested using fixed and estimated substitution rates with an initial rate of 5×10^{-4} substitutions per site per year. These models were compared against each other using Bayes factor analysis calculated using marginal-likelihood analysis. Among either the fixed rate models or the estimated rate models, the exponential clock had the highest Bayes factor; each exceeding a value of 3 when compared with the other clocks. The fixed rate, exponential clock model was better than the estimated rate, exponential clock model with a minimum Bayes factor value of 7 (Table 1) (Kass & Raftery, 1995). A fixed rate of 5×10^{-4} substitutions per site per year was chosen to make these analyses compatible with those of Markov *et al.* (2009).

Phylogeography

A discrete trait model was used to recreate a phylogeographic tree in a Bayesian framework. Initial runs showed that the posterior probabilities for some branches near the root of the tree were <0.7. To determine the variation in the position of these nodes, each model was replicated 10 times. In each run, Ghana was at the root of each tree with a posterior probability of 1.0 and a location probability of 0.997 (SD 6.4×10^{-4}) for the 300 trim1 sequence alignment and 0.97 (SD 6.7×10^{-3}) (Fig. 1) for the 340 trim2 sequence alignment. Some clades with posterior probabilities >0.7 appeared to be country specific: Guinea-Bissau, Cameroon, Madagascar, Morocco and Tunisia. These clades appeared in each tree and had posterior probabilities >0.7, except for the Guinea-Bissau clade; however, in spite of having a posterior probability as low as 0.24 on one tree, the Guinea-Bissau sequences consistently formed a clade in all replicate trees. The low posterior probability for the Guinea-Bissau clade appeared to be due to the inclusion/exclusion of a few sequences from Burkina Faso, Côte d'Ivoire, Ghana and Madagascar disrupting a fixed branching structure within the Guinea-Bissau clade. Although these clades appeared in all replicate trees, additional HCV2 sequences from these countries could be found variously distributed in the replicate trees. For example, sequences from Tunisia formed two to four additional clades in the replicates, but these clades were not found among all replicates and did not always have high posterior probabilities (Fig. 1). These alternative clades suggested multiple introductions of HCV2 to these countries, but the number of introductions and the timing of these introductions could not be determined from our analyses.

The program SPREAD was used to analyse the Bayesian phylogeographic reconstructions by incorporating spatial diffusion. This analysis indicated that Ghana was positioned at the centre of a web of links to other countries within Africa (Fig. 2). Most of the links between African countries included Ghana (n=10). The value n is the number of independent linkages calculated between a country, Ghana in this case, and other countries in West Central and North Africa. The next most connected countries were Burkina Faso (n=3) and Cameroon (n=3). The highest Bayes factors from this analysis were seen in linkages between Ghana and Burkina Faso, Côte d'Ivoire, Guinea-Bissau, Nigeria and Tunisia (Table

2). This finding was consistent with the basal location of the sequences from Ghana, and indicated a direct linkage between Ghana and all other countries with sequences in this study, except Algeria and Guinea. Furthermore, this was especially true for Morocco, Madagascar and Tunisia. However, a diffusion of HCV2 through neighbouring countries from Ghana to North Africa may have been missed due to lack of sequences from Mali, Mauritania and Niger.

Outgroup sequences

A maximum-likelihood tree published as Fig. 1 in Markov *et al.* (2009) contained HCV2 sequences that formed an outgroup to the main body of HCV2 sequences, including GenBank accession numbers AY236381 (Ghana), DQ345617 (Madagascar), FJ791094 (Central African Republic), GQ153866 (Guinea-Bissau) and GQ153902 (Guinea-Bissau). In the trees recreated here, none of these sequences appeared to form an outgroup (Fig. S1). Additionally, the two Guinea-Bissau sequences (GenBank accession numbers GQ153866 and GQ153902) clustered within the Guinea-Bissau clade in the trees inferred from our analyses.

Time to MRCA (TMRCA)

The inferred mean TMRCA to the root of the trees from the discrete location trait model runs was 602 years ago (SD 6.5 years) with a range of 514 (SD 6.5) to 695 years ago (SD 6.6 years) for the trim1 sequence alignment, and 618 years ago (SD 6.4 years) with a range of 540 (SD 6.3) to 698 years ago (SD 9.7 years) for the trim2 sequence alignment. These estimates were similar to those from other studies (Markov *et al.*, 2009; Pasquier *et al.*, 2005; Pouillot *et al.*, 2008). The discrete trait model was used to compare the results of this study with earlier studies. The reason for this was because TMRCA estimates from the discrete trait model and a model using the constant growth tree prior should yield similar results, but this study found that the trim2 mean TMRCA estimates with the constant growth prior were 7045 years ago [1693–17413 years ago, 95% highest posterior density (HPD)] and 2913 years ago (1385–5170 years ago, 95% HPD) using a fixed substitution rate of 5×10^{-4} substitutions per site per year for the uncorrelated exponential and log-normal molecular clocks, respectively. As a result of the disparity between the discrete trait and constant growth prior models, the time estimates in this paper are not considered to be accurate, but are used as a relative measure of events. TMRCA estimates for the country clades marked in Fig. 1 can be found in Table 3.

These results show that HCV2 had most likely spread throughout West, North and Central Africa along a temporal gradient with transmission to countries closer to Ghana at earlier times than countries further away from Ghana. Nigeria appeared to be an exception, but not all Nigerian HCV2 sequences were contained in a single clade, suggesting more than one introduction there. Sequences not belonging to this clade had TMRCAs back to Ghana in the range of 500–250 years ago, suggesting the potential for earlier introductions of HCV2 into Nigeria. More extensive sampling of HCV2 sequences from Nigeria will be required to obtain better estimates of these events.

DISCUSSION

Many studies investigating the evolution and molecular epidemiology of HCV2 over the last decade suggest that HCV2 originated in West Africa. A region encompassing Guinea and the Gambia was assumed to be the most probable site for the origin of HCV2 as sequences from this region were the most phylogenetically basal and more genetically diverse than sequences from other regions in West, North and Central Africa (Markov *et al.*, 2009). Recent research indicated that HCV2 was more likely to have spread between adjacent countries through diffusion from west to east over several centuries (Markov *et al.*, 2009; Pouillot *et al.*, 2008). To further explore the diversity and evolutionary history of HCV2, we re-examined HCV2 sequences collected from West, North and Central Africa, including 92 new sequences from Côte d'Ivoire, Ghana and Nigeria.

An analysis of sequence diversity shows that the sequences from Ghana were more genetically diverse (0.45) than sequences from any other country or all other countries combined (0.32). Discrete trait Bayesian analysis using the geographical location of the collected sequences demonstrated that Ghana was phylogenetically basal (Fig. 1). The location probability for Ghana at the root of the inferred trees was 0.97–0.997. Bayesian phylogenetic reconstruction of evolutionary dynamics to examine pairwise diffusion linkages between the countries studied showed that Ghana was most phylogenetically basal and had the most pairwise linkages to other countries, except Algeria and Guinea, which were linked to Morocco and Burkina Faso, respectively (Figs. 1 and 2).

As was observed earlier, the frequency of detection of HCV2 versus other HCV genotypes varies among countries in West Africa. However, the HCV2 frequency does not decline steadily from the most western to central coastal regions of the subcontinent. Instead, its frequency fluctuates. Among all HCV strains from West Africa, HCV2 was found in 85% of cases in Guinea, Guinea-Bissau (Markov *et al.*, 2009) and Ghana (Layden *et al.*, 2015), and only in 33, 25 and 15% (Forbi *et al.*, 2012, 2014; Markov *et al.*, 2009) of cases in Côte d'Ivoire, Cameron and Nigeria, respectively. Thus, the countries of Guinea and Ghana with the highest HCV2 frequency are separated by Côte d'Ivoire with a low HCV2 frequency, which is inconsistent with the west-to-east gradient of HCV2 frequency proposed earlier. Although diffusion of HCV2 between neighbouring countries is possible, irregular geographical distribution of this genotype contradicts its gradual dissemination from West to Central African countries along the coast of the subcontinent. However, the introduction of other HCV genotypes, e.g. genotype 1 in Côte d'Ivoire (Forbi *et al.*, 2014) and Nigeria (Forbi *et al.*, 2012), might influence HCV2 prevalence in these countries by displacing HCV2 as the dominant genotype.

Combined, the aforementioned observations suggest direct introduction of HCV2 from Ghana to most other countries in this study rather than diffusion through chains of linkages such as Ghana–Côte d'Ivoire–Guinea–Guinea-Bissau. The inferred TMRCAs suggest a temporal diffusion of HCV2 out of Ghana (Table 3), although this sampling of country-specific clades is not conclusive as country-specific sequences are not monophyletic (Fig. 1). Instead, the tree topologies suggest multiple introduction events of HCV2 to countries such as Nigeria and Tunisia. The strength of these introduction events are difficult to

estimate because of the low posterior probabilities associated with these sequences. More extensive sampling for HCV2 in these countries may allow better estimates to be made in the future.

The data obtained here indicate that strain sampling is very important for identification of the HCV2 geographical origin. Sampling of sequences in West Africa has been inconsistent, as noted here. Some countries are under-represented and other countries are not sampled. This could influence our study. Originally, Candotti *et al.* (2003) suggested HCV2 originated in Ghana with limited sampling. Markov *et al.* (2009) were able to expand the West African HCV sequence set and their results indicated that Guinea-Bissau might be the origin for HCV2. With the further expansion of West African HCV2 sequences presented here, it appears that Ghana is once again the location of the origin. With additional sampling the origin of HCV2 should be better estimated. Additional sequences allowed for resolving phylogenetic relationships of a small set of sequences that appeared to form an outgroup (Fig. 1 in Markov *et al.*, 2009) by showing that these sequences were not an actual outgroup (Fig. S1). These observations suggest that the HCV2 epidemic history in West Africa is extremely complex and can be accurately elucidated using only a very substantial sample of HCV2 sequences from all countries in this geographical region. HCV2 did not radiate from Ghana evenly to all countries in West Africa, with distant countries such as Guinea and Guinea-Bissau having experienced a significant influx, whilst neighbouring countries such as Côte d'Ivoire and Nigeria have a low frequency of HCV2. Such an erratic spatial distribution suggests that certain routes of geographical dissemination may have been more efficient than mere proximity.

Unlike earlier research, the linkages between Ghana and countries in North and Central Africa suggest a direct linkage between these countries and Ghana rather than diffusion between adjacent countries. It is conceivable that dissemination of HCV2 along trade routes, such as those dominating sub-Saharan Africa from the eighth to the sixteenth century, could have contributed to the modern patterns of HCV2 in Africa. Nonetheless, sampling of HCV2 sequences from unrepresented countries like Mali, Mauritania and Niger may help in understanding the contemporary patterns of HCV2 distribution in the region. Considering that northern Ghana is in geographical proximity to the major ancient trading centres at the southern edge of the Sahara, such as Timbuktu, Jenne and Gao, but has never served as a major trading centre itself, it is conceivable that HCV2 was continuously dispersed from Ghana through these trading centres to other geographical regions over many centuries, with the rate of dispersion having been affected by the extent of human travel related to the trade, which may potentially explain variation in the introduction of HCV2 to different African regions. In this respect, it is intriguing that a few HCV2 sequences sampled from Burkina Faso are scattered across HCV2 clusters (Fig. 1), suggesting that this country either is a part of the geographical origin or experienced numerous introductions of HCV2 from Ghana. A massive introduction of HCV2 to the most western countries of the subcontinent, such as Guinea-Bissau, may reflect a certain ease of communication through the Niger River along the ancient trading routes.

In conclusion, our results suggest that Ghana was the geographical place of origin for HCV2 in West Africa. A more extensive sampling of additional HCV2 sequences from

unrepresented and under-represented countries in West, North and Central Africa will help to further improve our understanding of the epidemic history, evolution and molecular epidemiology of HCV2.

METHODS

Patient samples

Cote d'Ivoire—In total, 608 plasma samples were obtained in 1995 from pregnant women in various hospitals throughout Côte d'Ivoire as part of human immunodeficiency virus surveillance studies. This resulted in the isolation of six HCV2 sequences with GenBank accession numbers KC012573, KC012577, KC012578, KC012585, KC012587 and KC012588. Additional details can be found in Forbi *et al.* (2014).

Ghana—In total, 363 blood samples from a population of recalled blood donors from the Komfo Anokye Teaching Hospital (KATH) blood bank, Kumasi, Ghana, were collected in 2012–2013. The HCV2 sequences from that study have the GenBank accession numbers KM213527–KM213591. Additional details can be found in Layden *et al.* (2015).

Nigeria—In total, 519 serum specimens were randomly obtained from asymptomatic indigenes from two rural communities in Keffi, Nigeria in 2007. Nine HCV2 sequences were isolated from this study with GenBank accession numbers JQ679028–JQ679036. Additional details can be found in Forbi *et al.* (2012).

Nucleic acid extraction and DNA sequencing—Total nucleic acid was extracted from all samples using an automated MagNA Pure LC robot and a MagNA Pure LC Total Nucleic Acid Isolation kit (Roche), and eluted with 50 μl elution buffer according to the manufacturer's instructions. cDNA was generated using a high-temperature capability SuperScript VILO cDNA Synthesis kit (Invitrogen) on an ABI PRISM 9700 PCR system. The PCR conditions following reverse transcriptase were: 25°C for 10 min, 42°C for 90 min and 85°C for 5 min.

Amplification of the NS5B gene region was done as described previously (Forbi *et al.*, 2012). Nested NS5B amplicons derived from the PCR amplification were purified (PCR purification kit; Qiagen) and sequenced with their respective nested primers using a BigDye v3.1 chemistry sequencing kit (Applied Biosystems) with an automated sequencer (ABI 3130xl; Applied Biosystems) as described previously (Forbi *et al.*, 2012).

Sequence diversity—Sequence diversity was calculated from sequence similarity using BioEdit (version 7.1.3.0) (Hall, 1999). Sequence diversity was calculated as 1 minus the fraction of sequence identity between sequences.

Sample analysis—A sequence database was created with all the HCV2 sequences from these studies combined with additional West African sequences, with dates of collection, downloaded from Gen-Bank. Sequences were aligned visually and the termini were trimmed in BioEdit (version 7.1.3.0) (Hall, 1999) to remove gaps resulting in an alignment of 246 nt containing 340 sequences (Table S1). The sequences aligned between positions 8371 to

8616 using GenBank accession number NC_009823 as reference. Two sequence alignments were created from this alignment. In the first alignment, all sequences <246 nt were removed, resulting in an alignment containing 300 sequences (trim1). In the second alignment, 12 bases were trimmed from the 5′ end of the alignment resulting in an alignment of 340 sequences covering positions 8383–8616 (234 nt). The first alignment maximized the length of the sequences, whilst the second alignment maximized the number of sequences from Guinea-Bissau, which were under-represented in the first dataset (trim2).

Coalescent analysis—BEAST (version 1.8.1) (Drummond *et al.*, 2012) was used to estimate the TMRCA and reconstruct phylogeographic trees using a discrete trait model. The HKY substitution model was used with four Γ rate categories and invariant sites. An initial clock rate of 5.0×10^{-4} substitutions per site per year as determined by Pybus *et al.* (2001) was used so these analyses could be compared with the results of Markov *et al.* (2009). For TMRCA analyses, a coalescent constant size prior was used for the analysis. Clock rates were either fixed or estimated using strict, log-normal or exponential clocks. A chain length of 1×10^{8} was used, and parameters and trees were logged every 5000 iterations. Convergence was inspected using the trace option in Tracer (version 1.6). A 10% burn-in was applied to the log and tree files. Bayes factors were calculated using marginal-likelihood analysis in Tracer (version 1.5) with 1000 bootstrap replicates (Kass & Raftery, 1995).

Phylogeography—A phylogenetic tree with discrete geographical location traits (http://beast-classic.googlecode.com/files/ARv2.0.1.pdf) associated with the tips of the tree used the substitution model listed above for the nucleotide sequences and the symmetrical substitution model with social network inferred with Bayesian stochastic search variable selection was used for the location trait. The clock for the nucleotide sequences used the rate and clocks listed above, and the location trait rate was estimated used a strict clock with the default rate, the continuous time Markov chain rate reference prior distribution and the GMRF Bayesian Skyride tree prior (Minin *et al.*, 2008). The analysis was run 10 times to determine the variation in results. Parameter estimates are reported as mean±SD. The tree in Fig. 1 was chosen at random.

Pairwise diffusion—The discrete Bayes factors for each pairwise location indicator between locations were calculated using SPREAD (version 1.0.6) with a Bayes factor cut-off of 3.0 (Bielejec *et al.*, 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr Chong-Gee Teo for his discussion and comments on this paper. This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. It does not represent and should not be construed to represent any agency determination or policy. Use of trade names is for identification only and does not imply endorsement by the US Department of Health and Human Services, the Public Health Service or the Centers for Disease Control

and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011; 27:2910–2912. [PubMed: 21911333]

Candotti D, Temple J, Sarkodie F, Allain JP. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. J Virol. 2003; 77:7914–7923. [PubMed: 12829831]

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby R, Barr PJ. Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci U S A. 1991; 88:2451–2455. [PubMed: 1848704]

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the beast 1.7. Mol Biol Evol. 2012; 29:1969–1973. [PubMed: 22367748]

EASL. EASL International Consensus Conference on Hepatitis C. Paris, 26–27 February 1999. Consensus statement. J Hepatol. 1999; 31(Suppl 1):3–8. [PubMed: 10622553]

Forbi JC, Purdy MA, Campo DS, Vaughan G, Dimitrova ZE, Ganova-Raeva LM, Xia GL, Khudyakov YE. Epidemic history of hepatitis C virus infection in two remote communities in Nigeria, West Africa. J Gen Virol. 2012; 93:1410–1421. [PubMed: 22456613]

Forbi JC, Campo DS, Purdy MA, Dimitrova ZE, Skums P, Xia GL, Punkova LT, Ganova-Raeva LM, Vaughan G, et al. Intra-host diversity and evolution of hepatitis C virus endemic to Côte d'Ivoire. J Med Virol. 2014; 86:765–771. [PubMed: 24519518]

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999; 41:95–98.

Hanafiah KM, Groeger J, Flaxman AD, Wiersma ST. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. Hepatology. 2013; 57:1333–1342. [PubMed: 23172780]

Itskowitz MS. Hepatitis C: epidemiology, diagnosis, and management. Compr Ther. 2007; 33:87–93. [PubMed: 18004020]

Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995; 90:773–795.

Layden JE, Phillips RO, Owusu-Ofori S, Sarfo FS, Kliethermes S, Mora N, Owusu D, Nelson K, Opare-Sem O, et al. High frequency of active HCV infection among seropositive cases in west Africa and evidence for multiple transmission pathways. Clin Infect Dis. 2015; 60:1033–1041. [PubMed: 25477425]

Lindenbach BD, Rice CM. Unravelling hepatitis C virus replication from genome to function. Nature. 2005; 436:933–938. [PubMed: 16107832]

Markov PV, Pepin J, Frost E, Deslandes S, Labbé AC, Pybus OG. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. J Gen Virol. 2009; 90:2086–2096. [PubMed: 19474244]

Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008; 25:1459–1471. [PubMed: 18408232]

Nakano T, Lau GM, Lau GM, Sugiyama M, Mizokami M. An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. Liver Int. 2012; 32:339–345. [PubMed: 22142261]

Njouom R, Nerrienet E, Dubois M, Lachenal G, Rousset D, Vessière A, Ayouba A, Pasquier C, Pouillot R. The hepatitis C virus epidemic in Cameroon: genetic evidence for rapid transmission between 1920 and 1960. Infect Genet Evol. 2007; 7:361–367. [PubMed: 17137845]

Pasquier C, Njouom R, Ayouba A, Dubois M, Sartre MT, Vessière A, Timba I, Thonnon J, Izopet J, Nerrienet E. Distribution and heterogeneity of hepatitis C genotypes in hepatitis patients in Cameroon. J Med Virol. 2005; 77:390–398. [PubMed: 16173014]

Pouillot R, Lachenal G, Pybus OG, Rousset D, Njouom R. Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon. Infect Genet Evol. 2008; 8:676–681. [PubMed: 18599380]

Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. Science. 2001; 292:2323–2325. [PubMed: 11423661]

Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014; 59:318–327. [PubMed: 24115039]

WHO. Hepatitis C. Fact Sheet 164. 2012. http://www.who.int/mediacentre/factsheets/fs164/en/

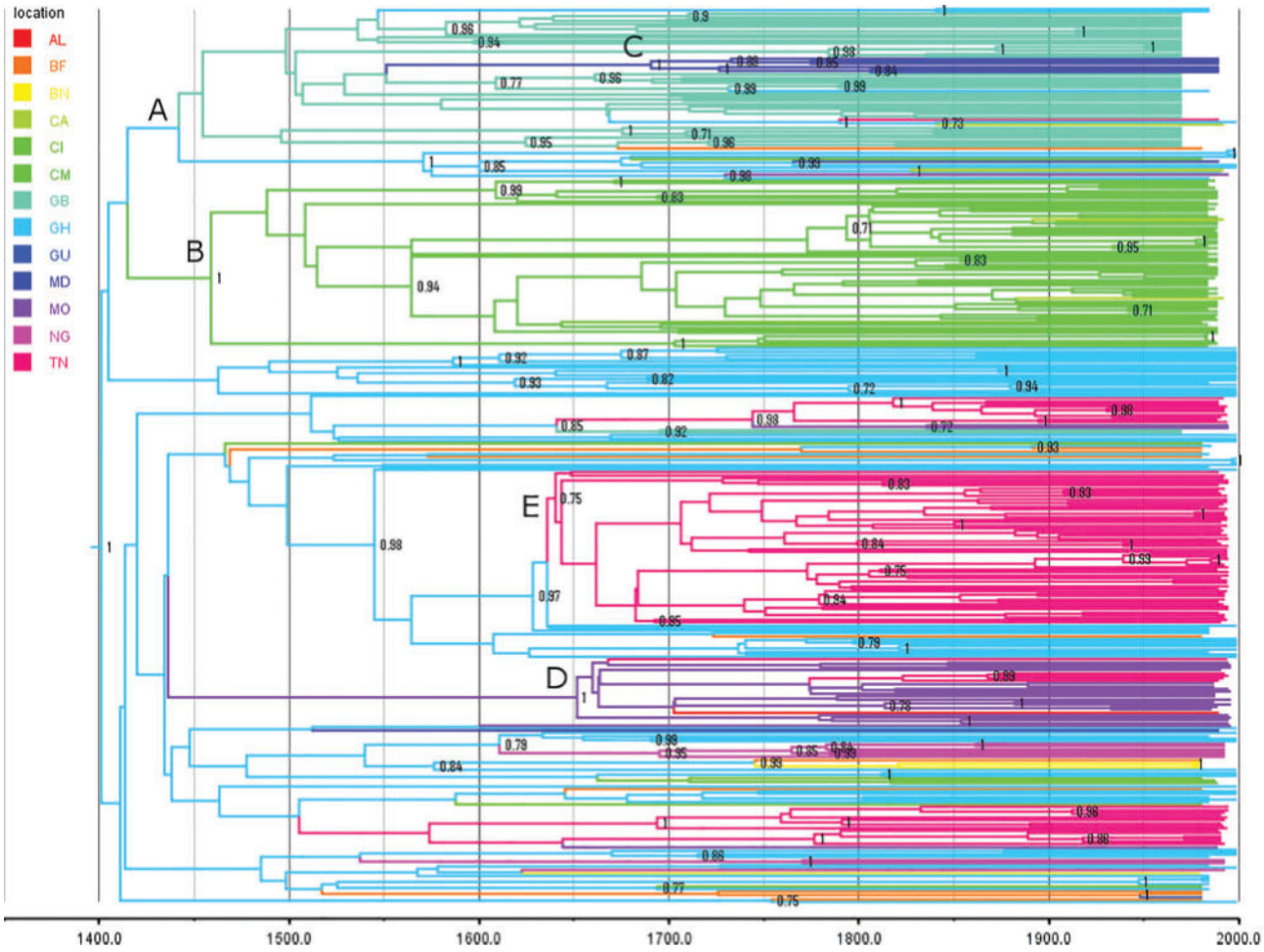

Fig. 1.

Discrete location trait Bayesian tree for the trim2 sequence alignment. The legend in the upper left corner shows the colour coding for the branches in the tree. The numbers at the nodes of the tree are posterior probabilities. Posterior probabilities <0.7 were excluded. The timescale is specific for this tree. A, Guinea-Bissau clade; B, Cameroon clade; C, Madagascar clade; D, Morocco clade; E, Tunisia clade. Two-letter country codes: AL, Algeria; BF, Burkina Faso; BN, Benin; CA, Central African Republic; CI, Côte d'Ivoire; CM, Cameroon; GB, Guinea-Bissau; GH, Ghana; GU, Guinea; MD, Madagascar; MO, Morocco; NG, Nigeria; TN, Tunisia.

Fig. 2.

Spatial diffusion of HCV2. The lines show calculated spatial diffusion between pairs of countries in West Africa. Thicker lines indicate the higher Bayes factor support for a linkage pair (see also Table 2). The dashed line indicates the linkage between Burkina Faso and Benin was only seen in two of the 10 replicate runs. Two-letter country codes are the same as in Fig. 1.

Table 1

Bayes factor comparisons of clock models

	Estimated rate, exponential clock	Estimated rate, log-normal clock	Estimated rate, strict clock	Fixed rate, exponential clock	Fixed rate, exponential clock	Fixed rate, strict clock
trim1 alignment (246 nt)						
Estimated rate, exponential clock		10	88	−7	5	78
Estimated rate, log-normal clock	−10		78	−18	−6	68
Estimated rate, strict clock	−88	−78		−96	−84	−10
Fixed rate, exponential clock	7	18	96		12	86
Fixed rate, exponential clock	−5	6	84	−12		74
Fixed rate, strict clock	−78	−68	10	−86	−74	
trim2 alignment (234 nt)						
Estimated rate, exponential clock		4	93	−14	−4	72
Estimated rate, log-normal clock	−4		88	−18	−8	68
Estimated rate, strict clock	−93	−88		−107	−96	−20
Fixed rate, exponential clock	14	18	107		11	86
Fixed rate, exponential clock	4	8	96	−11		76
Fixed rate, strict clock	−72	−68	20	−86	−76	

Table 2

Discrete Bayes factors for pairwise rate of diffusion between locations

Data are listed as mean (SD) of 10 replicate analyses; square brackets indicate the number of estimates per location pair if <10 estimates were obtained. Two-letter country codes are the same as in Fig. 1.

	BN	CM	GH	GU	MD	MO
trim1 alignment (246 nt)						
AL						133 (61)
BF	3 (0) [2]		46 300 (0)	178 (77)		
BN			14 (2)			
CA		1977 (775)	44 (7)			
CI		9 (1) [9]	41 670 (9762)			
CM			21 (4)			
GB			37 (5)			
MD			8892 (13 391)			
MO			264 (58)			
NG			20 186 (18 429)			
TN			17 514 (15 942)			5159 (5479)
trim2 alignment (234 nt)						
AL						117 (45)
BF	3 (0) [2]		45 945 (750)	185 (132)		
BN			14 (3)			
CA		2865 (1087)	20 (5)			
CI		7 (2)	45 945 (750)			
CM			26 (11)			
GB			45 945 (750)		6 (1)	
MD			48 (5)			
MO			427 (130)			
NG			16 774 (20 409)			
TN			20 043 (14 558)			1384 (1151)

Table 3

TMRCA and lower/upper 95 % HPD data (years ago) for select sequence clades

Data are listed as mean (SD) of 10 replicate runs. Two-letter country codes are the same as in Fig. 1.

	TMRCA	Lower 95 % HPD	Upper 95 % HPD
GB	561 (13.5)	481 (9.6)	641 (8.0)
CM	519 (4.7)	431 (4.5)	608 (8.5)
MO	390 (7.1)	291 (6.8)	491 (9.4)
TN	362 (16.4)	289 (10.3)	434 (37.7)
MD	320 (3.6)	210 (4.5)	435 (5.6)
NG	262 (19.6)	160 (15.1)	371 (27.2)