VIRUS EVOLUTION

Virus Evolution, 2022, **8(2)**, 1–9
DOI: https://doi.org/10.1093/ve/veac066
Advance access publication date: 21 July 2022
Research Article

Association of hepatitis C virus genotype 2 spread with historic slave trade and commerce routes in Western Africa

Ignacio Postigo-Hidalgo,[1] N'Faly Magassouba,[2] Barré Soropogui,[3] Elisabeth Fichet-Calvet,[4] and Jan Felix Drexler[1,5,*,†]

[1]Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Virology, Berlin 10117, Germany, [2]Université Gamal Abdel Nasser de Conakry, Conakry BP 1147, Guinea, [3]Direction préfectorale de la santé de Gueckedou, Laboratoire du Projet des Fièvres Hémorragiques de Guinée (PFHG), Conakry, Guinea, [4]Department of Virology, Bernhard Nocht Institute for Tropical Medicine, Hamburg 20359, Germany and [5]German Centre for Infection Research (DZIF), Associated Partner Site Charité – Universitätsmedizin Berlin, Braunschweig 38124, Germany
[†]https://orcid.org/0000-0002-3509-0232
*Corresponding author: E-mail: felix.drexler@charite.de

Abstract

The hepatitis C virus genotype 2 (HCV2) is endemic in Western and Central Africa. The HCV2 evolutionary origins remain uncertain due to the paucity of available genomes from African settings. In this study, we investigated the molecular epidemiology of HCV infections in rural Guinea, Western Africa, during 2004 and 2014. Broadly reactive nested reverse transcription polymerase chain reaction (RT-PCR)-based screening of sera from 1,571 asymptomatic adults resulted in the detection of 25 (1.5 per cent; 95 per cent confidence interval 0.9–2.3) positive samples, with a median viral load of 2.54E + 05 IU/ml (interquartile range 6.72E + 05). HCV-infected persons had a median age of 47 years, and 62.5 per cent were male and 37.5 per cent were female. The full polyprotein-encoding genes were retrieved by a combination of high throughput and Sanger sequencing from 17 samples showing sufficiently high viral loads. Phylogenetic analysis and sequence distances ≥13 per cent averaged over the polyprotein genes compared to other HCV2 subtypes revealed nine previously unknown HCV2 subtypes. The time to the most recent common ancestor of the Guinean HCV2 strains inferred in a Bayesian framework was 493 years (95 per cent Highest posterior density (HPD) 453–532). Most of the Guinean strains clustered poorly by location on both the level of sampling sites within Guinea and the level of countries in the phylogenetic reconstructions. Ancestral state reconstruction provided decisive support (Bayes factor > 100) for an origin of HCV2 in Western Africa. Phylogeographic reconstructions in a Bayesian framework pointed to a radial diffusion of HCV2 from Western African regions encompassing today's countries like Ghana, Guinea Bissau, or Burkina Faso, to Central and Northern African regions that took place from the 16th century onwards. The spread of HCV2 coincided in time and space with the main historic slave trade and commerce routes, supported by Bayesian tip-association significance testing (P = 0.01). Our study confirms the evolutionary origins of HCV2 in Western Africa and provides a potential link between historic human movements and HCV2 dispersion.

Key words: HCV; hepatitis C virus genotype 2; subtype; phylogeography; evolution; slave trade; Africa

1. Introduction

The hepatitis C virus (HCV) is a widely distributed human pathogen with an estimated 0.7 per cent global prevalence of viraemic infections (56.8 million cases) (Blach et al. 2022). Mainly due to liver cirrhosis and hepatocellular carcinoma resulting from chronic infections (Perz et al. 2006), HCV is responsible for substantial morbidity and mortality, causing approximately 5.5 million deaths each year (Blach et al. 2022). While direct-acting antivirals have become available in recent years, they are not universally accessible yet and global eradication is not yet within reach (Rasche et al. 2019).

HCV strains are classified into eight genotypes and further classified into ninety subtypes (Smith et al. 2014; ICTV 2021) that may differ in epidemiology, virulence, and treatment options (Zein 2000). A clear taxonomic distinction can be drawn between strains that differ by less than 13 per cent nucleotide content over their entire coding sequence (CDS) (members of the same subtype) and those that differ by more than 15 per cent (different subtypes) or by 17–36 per cent (different genotypes) (Smith et al. 2014). Some HCV genotypes are associated with distinct geographical regions. HCV genotype 2 (HCV2), the focus of our study, is frequently found in Western Africa, where a high diversity of subtypes has been found (Markov et al. 2009; Purdy et al. 2015; Petruzziello et al. 2016).

HCV2 is suggested to have originated in Western Africa. Previous studies reported Ghanaian HCV2 strains to be more diverse in comparison to other regions of the world (Candotti et al. 2003); this was later confirmed by an analysis of HCV2 sequences from Cameroon, which revealed younger common ancestors and less diversity in comparison to Western African strains and, thus,

Table 1. Number of samples collected by location and year and PCR detection by Guinean village. The 95 per cent CI corresponds to the proportion of PCR detections among the screened patients of each village.

Village	Year	Screened (*n*)	PCR detection (*n*)	PCR detection (%)	95% CI
Bantou	2004	135	7	5.19	1.20–4.90
Tanganya	2004	130	1	0.77	0.00–2.00
Brissa	2014	234	3	1.28	0.30–3.10
Dalafilani	2014	211	4	1.90	0.50–3.60
Damania	2014	203	0	0.00	–
Sokourala	2014	213	7	3.29	1.20–4.90
Sonkonya	2014	235	3	1.28	0.30–3.10
Yarawalia	2014	210	0	0.00	–
Total	–	1,571	25	1.59	1.03–2.34

CI: confidence interval.

pointed to an introduction from Western into Central Africa (Pasquier et al. 2005; Pouillot et al. 2008). A dataset including new sequences from Guinea Bissau confirmed an eastward spread from West African countries to Central Africa (Markov et al. 2009). However, the most recent re-evaluation of HCV2 including new sequences from regions corresponding to today's countries like Ghana, Côte d'Ivoire, and Nigeria showed that the spread of HCV2 was more complex and pointed to a radial diffusion from Western African regions to first Central African and later Northern African countries (Purdy et al. 2015). Overall, the available studies have been limited by the paucity of Africa-derived HCV2 genomic sequences, hindering definite assessments of the HCV2 evolutionary origins.

In this study, we screened a large cohort from Western Africa for HCV, identified new HCV2 subtypes, and analysed the evolutionary history of HCV2 in Africa.

2. Materials and methods

2.1 Collection of the samples

Blood specimens were sampled during 2004 and 2014 within the Faranah prefecture in upper Guinea (Table 1). They were obtained from 1,571 asymptomatic adults sampled in eight villages: Bantou and Tanganya in 2004, and Brissa, Dalafilani, Damania, Sokourala, Sonkonya, and Yarawalia in 2014 (Fig. 1). Sampling in 2004 was done within a project termed 'New approaches to the treatment and control of Lassa fever and yellow fever in West Africa (TREATCONTROL-VHF)', and sampling in 2014 was done within a project termed 'Lassa fever in Guinea and Sierra Leone, rodent control and seasonality of human exposure to Rodents (LAROCS)'. Both studies investigated human–rodent interactions and infectious diseases such as those caused by Lassa or Hantaviruses (Fichet-Calvet et al. 2007; Klempa et al. 2013; Mari Saez et al. 2018).

For the 2004 cohort, all participants were self-recruited after a public announcement. Sampling took place in a health facility, where a doctor did consultancies after bleeding the patients. The participants were enrolled based on reported high fever during the last 3 months. For the 2014 cohort, forty households per village were recruited using an adapted and expanded programme of immunisation survey sampling strategy (Hadler et al. 2004): starting from the centre of the village, research teams walked in a randomly identified direction, counting and labelling all households until they reached the village border. This was repeated in the opposite and in the two perpendicular directions for a better geographic representation of the village population. Households were then selected with a constant sampling fraction per direction, and in each household, all individuals older than 5 years were recruited into the study.

Samples were collected and analysed in agreement with ethical permits from the Guinean National Institute of Public Health (066/INSP/12), the Guinean National Ethics Committee for Health Research (2003/PFHG/05/GUI and 12/CNERS/12), and the Charité – Universitätsmedizin Berlin (EA2/177/11).

2.2 PCR detection and viral load estimation

Nucleic acid was extracted from serum samples using the MagNA Pure 96 Viral NA small volume kit (Roche Molecular Systems, USA).

Specimens were screened for HCV RNA using a broadly reactive nested PCR assay targeting the nonstructural 3 (NS3)

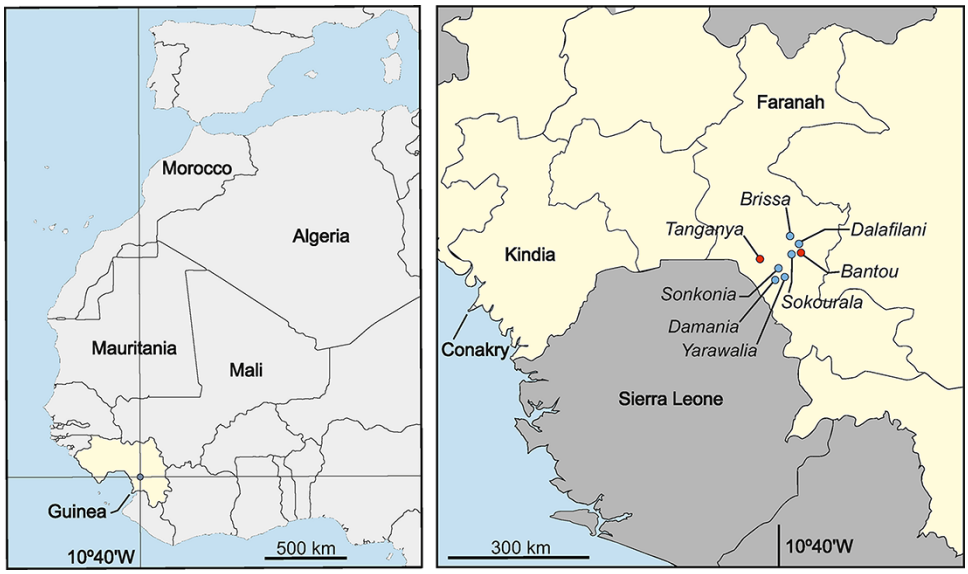

Figure 1. Map of Guinea and sampling locations. Dots in red (Bantou, Tanganya) correspond to sampling locations in 2004 and dots in blue (Brissa, Dalafilani, Damania, Sokourala, Sonkonia and Yarawalia), correspond to sampling locations in 2014.

region of the polyprotein gene, as published previously (Drexler et al. 2013), modified by a separate reverse transcription step using SuperScript III (Invitrogen, Karlsruhe, Germany), random hexamer primers, and subsequent PCR steps using Platinum Taq (Invitrogen, Karlsruhe, Germany) that involved the following cycling parameters: ten cycles of 94°C for 15 s, 60°C for 20 s (touch down −1°C/cycle), and 72°C for 45 s, followed by forty cycles of 95°C for 15 s, 50°C for 20 s, and 72°C for 45 s, combined with a final extension at 72°C for 1 min.

A highly sensitive reverse transcription-quantitative polymerase chain reaction assay targeting the HCV X-tail domain was used for confirmation and for viral load estimations, relying on an armoured RNA standard as a quantitative control, as described previously (Drexler et al. 2009).

From all positive samples, a 326 nucleotide (nt.) fragment from the NS5B polyprotein region corresponding to positions 8354 to 8678 of the HCV2 reference genome (GenBank accession number NC_009823) was amplified using primers and PCR conditions as previously described (Markov et al. 2009).

2.3 HCV2 sequencing

A mixed sequencing strategy of unbiased Illumina-based sequencing and amplicon Oxford Nanopore Technologies (ONTs)-based sequencing was employed.

To perform the amplicon ONT-based approach, we collected all available HCV2 genomes from GenBank (182 sequences; accessed on 2 August 2021) and aligned their full CDS, i.e. the polyprotein gene, using MAFFT (Katoh and Standley 2013). We then searched for conserved regions across the genome to design primers flanking overlapping regions of around 400 nt. A nested PCR approach was designed, resulting in two forward primers (one outer and one inner) and two reverse primers (one outer and one inner) for each amplicon. In total, two assays comprising 22 and 19 amplicons per round were designed (Tables S1 and S2). PCR conditions after reverse transcription were: a first round of forty cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 35 s combined with a final extension at 72°C for 1 min and a second round of thirty-five cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 35 s, followed by a final extension at 72°C for 1 min. Platinum Taq Polymerase (Invitrogen, Karlsruhe, Germany) was used for both rounds.

From each positive sample, amplicons were assigned a barcode using the ONT 'Native barcoding expansion 96 kit – EXP-NBD196' and pooled by sample. DNA repair and end-prep ensued to prepare the DNA for adapter ligation sequencing using a 'Ligation sequencing kit – SQK-LSK109'. After a final cleaning step and quality control, the final libraries were loaded into five different MinION flow cells, chemistry version 10.4. Raw reads were base-called and de-multiplexed simultaneously on a GridION (ONT, UK) system. Resulting sequences were mapped against the CDS of HCV2 subtypes retrieved from a reference set from the International Committee on Taxonomy of Viruses (ICTV 2021) using Bowtie2 (v.7.2.1) (Langmead and Salzberg 2012) implemented in Geneious Prime (2020.1.1).

Lastly, from each positive sample, library preparation and Illumina MiSeq sequencing were performed using the KAPA Frag Kit and KAPA Hyper Prep kit (Roche Molecular Diagnostics, Switzerland) and MiSeq reagent v2 chemistry (Illumina, USA) according to the manufacturers' protocols. Ribosomal RNA depletion was performed prior to cDNA synthesis with the QIAseq FastSelect Kit (Qiagen, Germany). Libraries were analysed with a Tape Station (Agilent Technologies, USA) and re-amplified when necessary, following a modified protocol from 'myBaits®' (ArborScience 2021). Generated sequences were mapped as described above. Eight samples with low viral loads did not yield enough reads to achieve full genome coverage.

2.4 HCV2 subtyping

The newly generated HCV2 CDS were added to a dataset of HCV2 subtypes retrieved from the International Committee on Taxonomy of Viruses (ICTV 2021). Separated datasets were created for the complete CDS and the NS3, NS5A, and NS5B polyprotein regions. Coding alignments were performed using MAFFT (v.1.4.0).

Approximately maximum-likelihood phylogenies of the complete CDS and the NS3, NS5A, and NS5B polyprotein regions were reconstructed using FastTree (v.2.1.12) (Price, Dehal, and Arkin 2009) with the Jukes–Cantor substitution model and 1,000 bootstrap replicates of the Shimodaira–Hasegawa approximate likelihood test (Supplementary Figs S1–S3). Sites containing only gaps or ambiguous positions were removed from the analyses. Resulting trees were visualised using FigTree (v.1.4.4) (Suchard et al. 2018).

Percentage sequence distance between the CDS of all new HCV2 sequences and the dataset of HCV2 subtypes retrieved from the ICTV was calculated in MEGA X (v.10.0.5) (Kumar et al. 2018). Sites containing only gaps or ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 9,428 positions in the final dataset.

2.5 Recombination analysis

Full CDS were analysed for recombination with RDP4 (v.4.101) (Martin et al. 2015), employing RDP, GENECONV, Bootscan, and MaxChi, all implemented in the programme. To identify possible recombination breakpoints in the typing datasets, an analysis was performed at the nucleotide level with all HCV2 reference subtypes
and the new genomes as query. Recombination events detected by less than two methods and those not statistically significant ($P > 0.05$) were ignored. Sequences with predicted recombination events by three or more methods were analysed individually and considered only if maximum likelihood phylogenies from adjacent genomic regions showed statistically supported divergent topologies.

2.6 Phylogeographic analysis

To enable comparison with previous studies (Markov et al. 2009; Purdy et al. 2015), HCV2 sequences, including those newly available in GenBank and our twenty-five newly characterised strains, were combined into a dataset covering the NS5B region of the polyprotein gene and termed 'Trim1' (252 nt.; $n = 341$) according to Purdy et al. (2015). A list with all the sequences used for this study is found in the Supplementary Materials (Supplementary File 1).

A maximum clade credibility tree (MCC) with discrete geographical traits was inferred using BEAST (v.1.10.4) (Suchard et al. 2018) with an HKY substitution model, four gamma rate categories, and a coalescent Bayesian Skyride (Minin et al. 2008) tree prior. A strict clock with a fixed clock rate of 5.0E – 4 substitutions per site per year was used for the main and the location partitions as previously described (Markov et al. 2009); we chose a strict clock over an exponential clock used in another study (Purdy et al. 2015) as we obtained better convergence behaviour and tree topologies matching those of prior studies. For the discrete trait substitution model, we used 'country' as discrete states, and a 'symmetric substitution model', inferring epidemiological links with the Bayesian stochastic search variable selection procedure (Lemey et al. 2009). Chain length was set to 30.0E + 7, logging to every 5,000 iterations.

Table 2. Tip-association significance test using slave/commerce routes as discrete geographic traits: traits, including timespan, were defined based on historical data. A 'West-Central' route was added to represent eastward commercial routes that would connect western ports with the interior of the African continent. Tip-association trait indices and P values were calculated with BaTS as described in the 'Materials and Methods' section. AI and PS statistics measure the degree to which taxa with the same traits cluster together. The 'null mean' is calculated from the null distribution that results from the randomisation of the tree to test the null hypothesis, which assumes no clustering of taxa by the given traits. The smaller the AI and PS, and the larger the MC, the higher degree of clustering in taxa with the same traits.

Statistic/ MC traits (timespan)	Observed mean	95% HDPI	Null mean	95% HDPI	P value
AI	2.3	1.6–2.9	29.3	27.3–31.0	0.0
PS	24.9	23.0–26.0	181.3	173.2–187.6	0.0
West-Central (1500–1900)	19.7	14.0–25.0	3.5	2.7–5	0.01
Central-West (1500–1900)	64.0	64.0–64.0	3.3	2.5–4.4	0.01
West-South (1700–1800)	6.0	6.0–6.0	1.1	1.0–1.4	0.01
West-North (1700–1900)	47.8	26.0–58.0	3.3	2.5–4.2	0.01

AI, association index; CI, confidence interval; MC, monophyletic clade size statistic; HDPI, highest posterior density interval; PS: Parsimony score.

Convergence was inspected in Tracer (version 1.6). A 10 per cent burn-in was applied to the log and tree files. Resulting trees were colour-coded and visualised using FigTree (v.1.4.4) (Suchard et al. 2018). Ancestral state reconstructions were done in BEAST to investigate the geographic origin of the root of MCC tree using two states to classify HCV2 strains ('Western Africa' vs. 'Not Western Africa'); the same substitution model, tree prior, clock, and discrete trait model settings were employed as above. Spread3 (v.0.9.7.1) (Bielejec et al. 2016) was used to visualise the spatio-temporal spread of HCV2 from the resulting MCC tree and to identify and visualise well-supported rates between locations using Bayes factor tests implemented in the software.

BaTS (v.0.9) (Parker et al. 2008) was used with dataset 'Trim1' to test for associations between phylogenetic clustering and historical slave/commerce routes (Eltis and Richardson 1995; Badie, Bernard, and Brachet et al. 2008) described as discrete character states composed by the route and an associated timespan (Table 2). We defined three different discrete character states: West-North—1,700–1,900, West-South—1,700–1,800, and Central-West—1,500–1,900, and an additional 'West-Central—1,500–1,900' character state to represent eastward commercial routes that would connect western ports with the interior of the African continent. The top 100 highest posterior trees from the BEAST analysis were selected and analysed using these 4 states and 100 null replicates per each tree. The null hypothesis under test was that a given historical slave or commerce route was not more likely to be shared with adjoining taxa than we would expect due to chance. Traits were assigned to tips belonging to a given phylogenetic cluster considering together their geographic origin, date, and the observed direction of the ancestral transitions across time as observed in the MCC tree (e.g. Tunisian tips with ancestors from Ghana would get a West-North—1,700–1,900 trait assignment) (Supplementary Fig. S4).

3. Results

3.1 HCV2 prevalence in Guinea

The median age of the study participants was 26.0 years [interquartile range (IQR) 33.0]; 51.9 per cent were females and 48.1 per cent were males. In our cohort, 25 adults presented HCV2 infection, corresponding to 1.8 per cent (95 per cent CI 1.1–2.6) of the study cohort. Persons with a positive HCV2 result had a median age of 47.0 years (IQR 24.0), which is in agreement with the upper age range limit of HCV-infected people from West African countries like Nigeria or Ghana, but 19 years above the median age reported from Guinea (Riou et al. 2016); 62.5 per cent were males and 37.5 per cent females, which is consistent with global estimates of the sex ratio of HCV infections. Potential transmission routes including cultural and religious practices such as tattooing, scarification, or direct contact with blood (Rasche et al. 2019) could not be elucidated in our study. The HCV2 detection rate was consistent with a meta-analysis on HCV2 seroprevalence in Africa, which included more than 200 studies evaluating the general population, that reported an HCV2 seroprevalence in Guinea close to 1.5 per cent (95 per cent CI 0.5–9.5) (Riou et al. 2016), suggesting that our sample may in part represent larger populations from Guinea.

3.2 High diversity of HCV2 subtypes

HCV2 is especially diverse in West Africa; however, only limited numbers of African HCV2 sequences are available (Fig. 2). From all twenty-five HCV2-positive samples, we sequenced a region of the NS5B polyprotein region previously used for phylogenetic analysis (Markov et al. 2009; Purdy et al. 2015) (GenBank accession numbers ON256660–ON256667), and we recovered seventeen complete CDS (GenBank accession numbers ON314812–ON314828). The latest classification of HCV published in 2014 (Smith et al. 2014) agrees on a 13 per cent nucleotide divergence between HCV2 subtypes across the polyprotein gene. An approximately maximum-likelihood phylogeny including our seventeen newly characterised sequences showed nine well-supported phylogenetic subtypes, independent from other subtypes (Fig. 3). This was supported by sequence distance comparisons, as the distance within all members of each putative subtype was 13.0 per cent or less (Supplementary Table S3). All potentially new subtypes of HCV2 that formed monophyletic clades in the ML tree based on the full CDS were still represented in those ML trees based on the partial NS5B region (Supplementary Fig. S3), which suggests robustness of our NS5B-based analyses. Our recombination analysis showed no significantly supported recombination events, which was consistent with low frequency of recombination in HCV (Viazov et al. 2000).

Most of the Guinean strains clustered poorly by location on both the level of sampling sites within Guinea and the level of countries in the phylogenetic reconstructions (Fig. 4). Likewise, repeated transmission events were observed at the local level (Supplementary Fig. S5), with only few samples clustering according to the geographic level of villages.

3.3 Western Africa remains the origin of HCV2

An ancestral state reconstruction based on the expanded Trim 1 dataset showed a posterior probability support of 1.0 for the root state corresponding to Western Africa, with a calculated Bayes factor >100, which is considered decisive evidence (Jeffreys

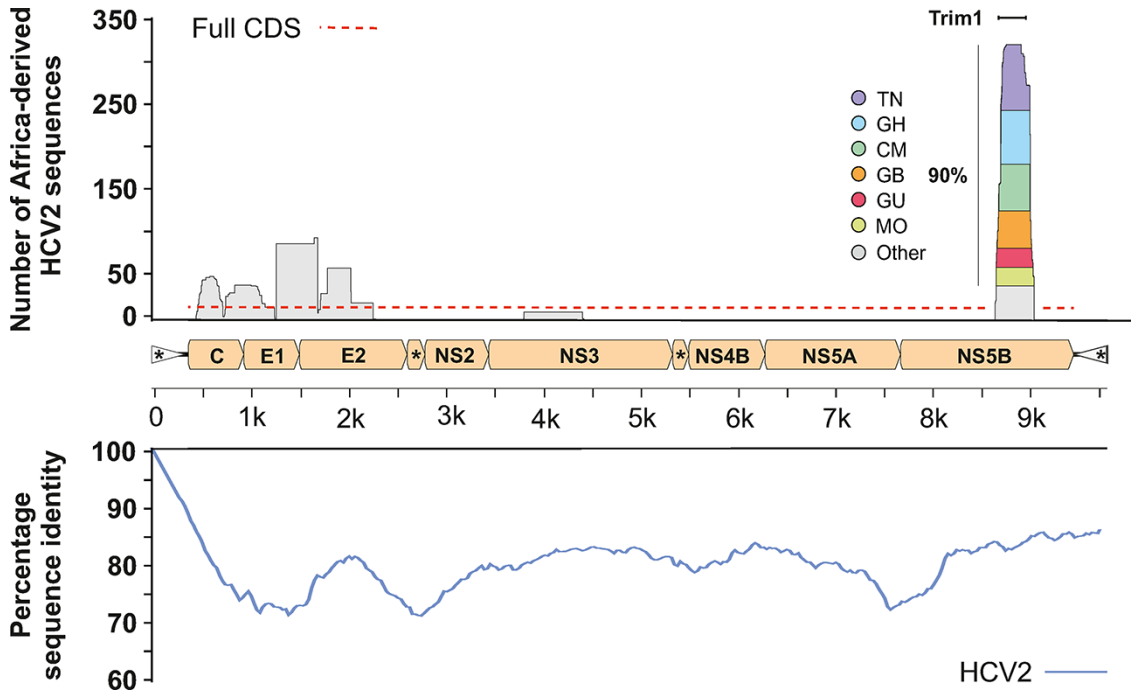

Figure 2. Africa-derived hepatitis C virus genotype 2 sequence availability. Upper portion of the figure: number of HCV2 Africa-derived HCV2 sequences from the different HCV polyprotein regions available in GenBank (accessed: 27 April 2021) and mapped to the full genome of NC_009823. P7, NS4A regions, and Untranslated regions are represented as '*' due to space limitations. The coloured 100 per cent stacked bar represens country diversity of the Trim1 dataset. Lower portion of the figure: genomic identity within all HCV2 genomes available. The identity plot was calculated in SSE (Simmonds 2012), using a fragment length of 400 nt. and an increment between fragments of 25 nt. Abbreviations: TN, Tunisia; GH, Ghana; CM, Cameroon; GB, Guinea Bissau; GU, Guinea; MO, Morocco.

1961). Consistent with previous studies (Candotti et al. 2003; Purdy et al. 2015), common HCV2 ancestors from a region corresponding to today's Ghana remained at the root and basal nodes of the tree (Fig. 4). Several countries formed well-supported nodes (Cameroon, Ghana, Guinea Bissau, Madagascar, Morocco, Nigeria, and Tunisia). In contrast, the samples from Guinea formed clades representing several introduction events from Ghana, one major event from Burkina Faso, and two introduction events to Madagascar and Tunisia as represented in the MCC tree by coloured branches (Fig. 4). The most recent common ancestor of all Guinean strains dated back to 1523 (95 per cent HPD 1485–1563). The largest amount of diverse HCV2 Guinean lineages came from a single introduction event from Burkina Faso dating back to 1610 (95 per cent HPD 1515–1612). However, we could not find a single solid historic movement that would support this introduction, other than possible migrations that started during the 16th century, such as the movements of the Yarse people that settled on the valley of the Black Volta River, and slave raids that were frequent in the area of today's Burkina Faso (McFarland and Lawrence 1998).

3.4 HCV2 dispersal may have been shaped by historic slave and commerce routes

We projected the MCC tree in time and space (Supplementary Video). The diffusion process revealed early transmission events during the 16th century within Western African regions and later, an eastward expansion, even to distant countries like Madagascar during the 17th and 18th centuries and with sporadic westward influx (Supplementary Fig. S6). More recent events during the 19th and 20th centuries accounted for the higher amount of lineage counts, which may have occurred because of a northwards expansion from the West African region to the Maghreb (Morocco, Tunisia, and Algeria). The radial diffusion from Ghana was strongly supported, both to adjacent and more distant countries (Fig. 5A). The highest support was found between country pairs where diffusion occurred more recently like Morocco and Tunisia or Guinea Bissau and Guinea and corresponds to well-supported apical branches in the tree. The remaining events yielded positive Bayes factors and corresponded to long-distance transitions that occurred from the 18th century onwards (Fig. 5A).

Overall, the diffusion process of HCV2 coincided temporally with transatlantic slave trade, during the 16th and 19th centuries (Eltis and Richardson 1995) and coincided geographically with the main African slave trade and commerce routes defined in Table 2 as 'Central-West', 'West-South', and 'West-North', as exemplified in (Badie, Bernard, and Brachet et al. 2008) (Fig. 5B). The Bayesian association test between the phylogeny and the defined slave trade routes showed significant association indices and parsimony scores for all defined traits ($P = 0.01$), yielding strong evidence for geographical grouping along these routes (Table 2) and supporting the existence of a geographic overlap of these routes with the main transitions that drove the dispersion of HCV2 (Fig. 5A, B).

4. Discussion

In this study, we screened individuals from rural Guinea for HCV infections, and we generated new HCV2 sequences that were used to evaluate the evolutionary history of HCV2 in Africa.

Many studies suggest that there is an enormous HCV diversity yet to be discovered in West Africa (Candotti et al. 2003; Pasquier et al. 2005; Smith et al. 2014). The ICTV guidelines for subtype assignment require (1) absence of recombination, (2) independent phylogenetic groups, and (3) representation by three or more

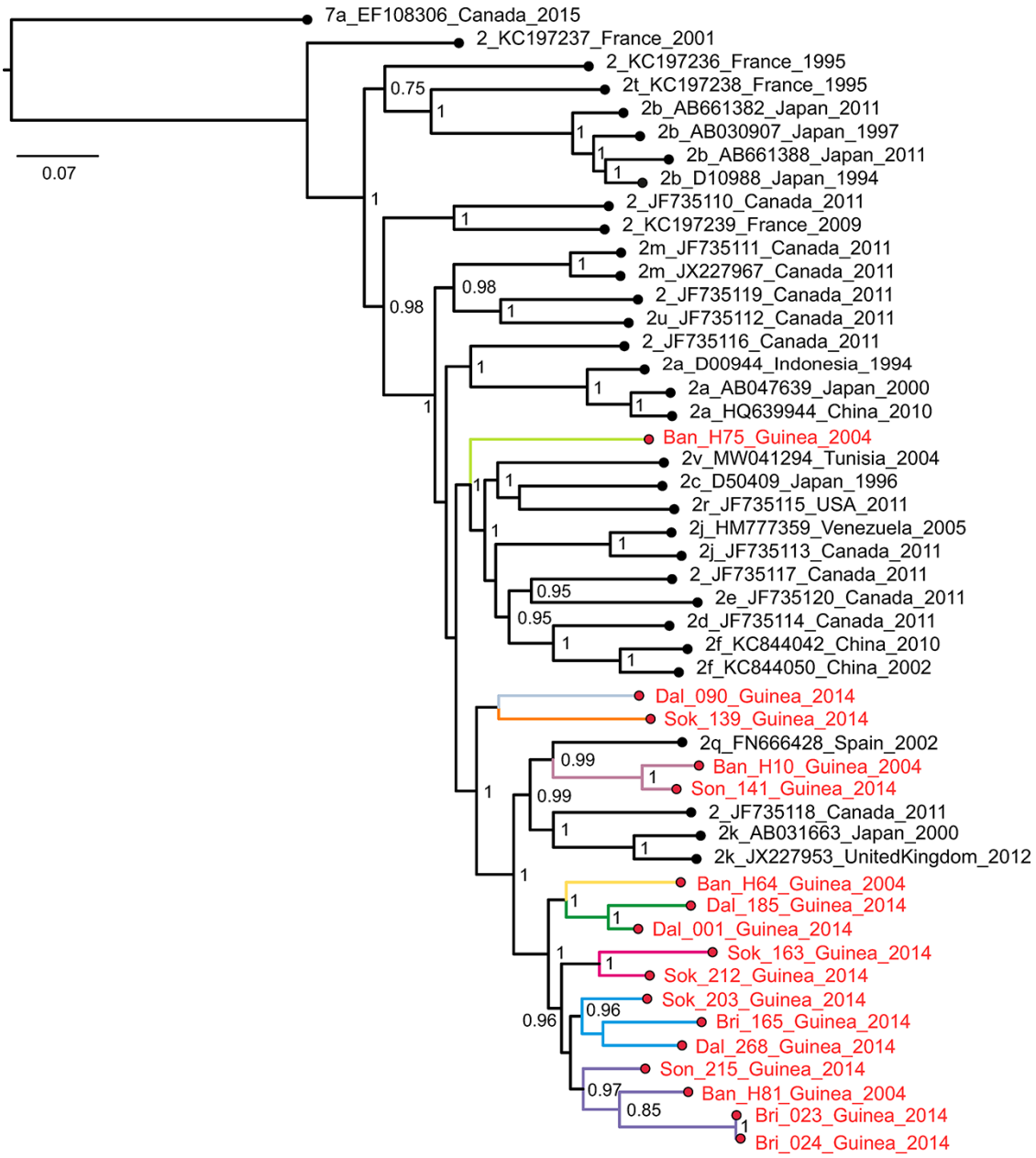

Figure 3. Evolutionary relationships of HCV2 strains. Approximately maximum-likelihood tree of the complete CDS calculated with FastTree. Node support is expressed as local support based on the Shimodaira–Hasegawa test; only support values above 0.75 are shown.

epidemiologically unlinked isolates. In the last decade, thirty-six new HCV2 subtypes have been described in previous studies (Forbi et al. 2012; Rajhi et al. 2021); however, only eleven have been officially recognised as the sequences were either not publicly shared or the ICTV criteria were not met (ICTV 2021). Our phylogenetic analysis allowed us to identify nine new subtypes, of which two met the ICTV criteria.

The location clustering patterns of the Guinean samples matched studies on Lassa Fever Virus (LASV) at a local level, where the samples did not cluster by villages of the same area (Fichet-Calvet et al. 2016; Marien et al. 2020). However, in those studies, LASV showed clustering at a regional level, in contrast to the Guinean samples, that clustered by different countries, suggesting repeated introduction events as opposed to movements between the villages. This may be explained by the transmission routes of these viruses, where LASV is likely transmitted through movement of infected rodents, contaminated food, or from humans to rodents by reverse zoonosis (Asogun et al. 2019); in comparison, HCV transmission is more restricted, as the virus is predominantly blood-borne and limited to human-to-human transmission (Rasche et al. 2019).

Slave trade and commerce routes could have served as the backbone of the HCV2 diffusion process in Africa. Enslaved people were fleeted to the Americas from the bright of Benin and surrounding areas since the 16th century; but most importantly, internal African routes converged in these West African ports, serving as potential hubs for infection; this could explain the transitions from Western to Central Africa and vice versa that we

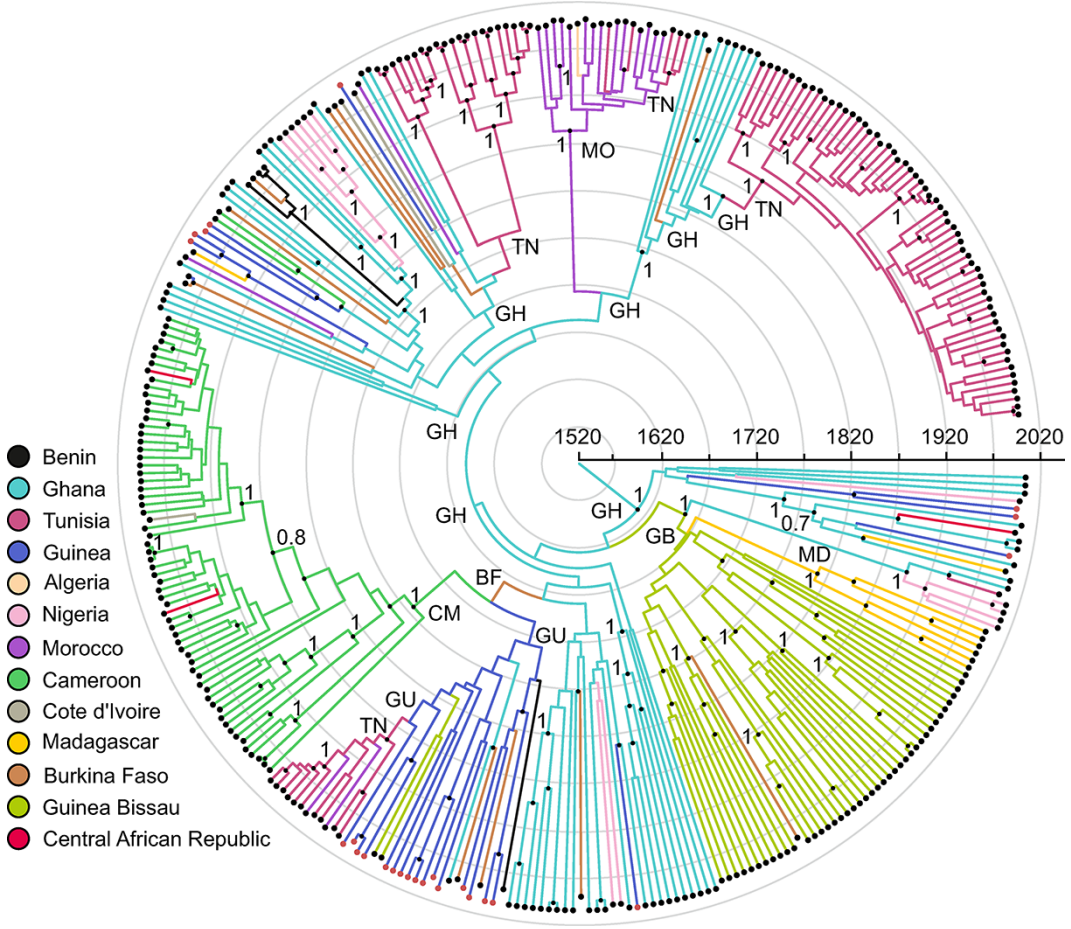

Figure 4. Time-calibrated phylogeny of HCV2 across Africa. Black dots at nodes represent posterior support >0.7. Red dots at the end of branches show the new Guinean sequences. Abbreviations: TN, Tunisia; GH, Ghana; CM, Cameroon; GB, Guinea Bissau; GU, Guinea; MO, Morocco; BF, Burkina Faso; MD, Madagascar.

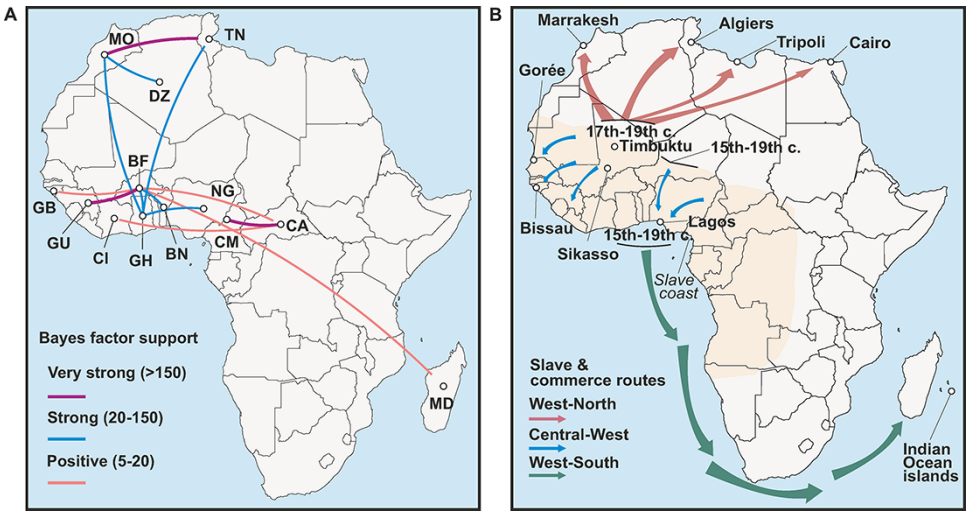

Figure 5. HCV2 spread in Africa. (A) Spatial representation of the transitions that dominated the diffusion process expressed as computed Bayes factor for each significant transition pair. Lines are colour-coded by Bayes factor support. (B) Representation of the main internal historic slave and commerce routes in the African continent. The beige area represents the slave efflux from inside the continent to the coast to join transatlantic slave trade routes. Base maps were made with free vector and raster map data from Natural Earth. Depicted routes and timespans were compiled from the following sources: Deschamps 1972; Ki-Zerbo 1994; Eltis and Richardson 1995; Pétré-Grenouilleau 1998a, 1998b; Coquery-Vidrovitch 1999; Lugan 2001; Badie, Bernard, and Brachet et al. 2008. Abbreviations: c., century; TN, Tunisia; GH, Ghana; CM, Cameroon; GB, Guinea Bissau; GU, Guinea; MO, Morocco; DZ, Algeria; BF, Burkina Faso; CI, Cote d'Ivoire; MD, Madagascar; CA, Central African Republic; BN, Benin; NG, Nigeria.

observed between the 16th and 18th centuries but also the distant later transitions from Western to Northern Africa and Madagascar, as found in our study, or to distant Caribbean locations such as Surinam or Martinique, as reported in previous studies (Markov et al. 2012). Comparatively, similar dispersal histories following slave trade routes through the African continent have been found for other viruses like yellow fever virus, human T cell leukaemia virus 1, or hepatitis B virus genotype A (Bryant, Holmes, and Barrett 2007; Afonso, Cassar, and Gessain 2019; Wolf et al. 2021).

There are some limitations to this study. The median viral load of some of the samples positive for HCV2 was low, which prevented the generation of full CDS in eight samples. However, for all HCV-positive samples, an NS5B fragment was amplified allowing for representative phylogenetic analyses. This region is typically used for screening, but it is genetically not highly divergent, resulting in low phylogenetic resolution as evidenced by this and past studies (Markov et al. 2009; Purdy et al. 2015). Moreover, HCV2 sampling in West Africa has been inconsistent; 90 per cent of the samples belong to six countries only (Tunisia, Ghana, Cameroon, Guinea Bissau, Guinea, and Morocco) and the remaining 10 per cent belong to other West and Central African countries. This has a direct influence on the assessment of HCV2 origins, as shown by Markov et al. (2009), whose results suggested that HCV2 originated in the western area of today's Guinea Bissau due to insufficient sampling. With the addition of twenty-five new Guinean strains to all available Africa-derived HCV2 sequences, our evaluation of the HCV2 evolutionary history supports the hypothesis of a Western African region encompassing current countries like Ghana, Guinea Bissau, or Burkina Faso as the area of origin, as suggested by Candotti et al. (2003) and supported by Purdy et al. (2015).

In conclusion, this study expands the known HCV2 diversity in Africa, confirms the relevance of Western Africa as the origin for HCV2, and describes its phylodynamic history following human movement while acknowledging defined historical routes across the whole African continent. Our results highlight the value of using rapidly evolving viruses to re-assess dispersal histories of their hosts (Ishak et al. 2017; Lukashev et al. 2017; Fisher, Streicker, and Schnell 2018).

Data availability

Viral genomic data are available in GenBank under accession numbers ON256660–ON256667 and ON314812–ON314828. Data on sequences used in the phylogenetic analyses are available in the Supplementary Materials.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

We are particularly indebted to Dr Pema Grovogui, Dr Oumar Sylla, Amara Camara, Fodé Kourouma, Amadou Doré, Fanta Bérété, and Mohamed Touré who participated in the blood collections on the field. We also thank Matthias Borchert and Almudena Mari Saez for their contribution to the sampling procedure and for commenting on an earlier version of the draft.

Funding

This work was supported by Host Switching Pathogens, Infectious Outbreaks and Zoonosis; the European programme International Cooperation and Development (grant ICA4-CT2002-10050); the European Union Horizon 2020 Innovative Training Networks (grant 721367); and the German Research Foundation (DFG project number 214436063, BO 3790/1-1, and FI 1781/1-1).

Conflict of interest: None declared.

References

Afonso, P. V., Cassar, O., and Gessain, A. (2019) 'Molecular Epidemiology, Genetic Variability and Evolution of HTLV-1 with Special Emphasis on African Genotypes', *Retrovirology*, 16: 1–15.

ArborScience. (2021), *myBaits Hybridization Capture for Targeted NGS - User Manual - Version 5.01. User Manual* <https://arborbiosci.com/wp-content/uploads/2021/03/myBaits_v5.01_Manual.pdf> accessed 15 Aug 2021.

Asogun, D. A. et al. (2019) 'Lassa Fever: Epidemiology, Clinical Features, Diagnosis, Management and Prevention', *Infectious Disease Clinics of North America*, 33: 933–51.

Badie, B. et al. (2008) *L'atlas Des Migrations: Le Monde Hors-série*, p. 186. Paris: Le Monde - Sciences Po.

Bielejec, F. et al. (2016) 'SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes', *Molecular Biology and Evolution*, 33: 2167–9.

Blach, S. et al. (2022) 'Global Change in Hepatitis C Virus Prevalence and Cascade of Care between 2015 and 2020: A Modelling Study', *The Lancet Gastroenterology & Hepatology*, 7: 396–415.

Bryant, J. E., Holmes, E. C., and Barrett, A. D. T. (2007) 'Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas', *PLoS Pathogens*, 3: e75.

Candotti, D. et al. (2003) 'Frequent Recovery and Broad Genotype 2 Diversity Characterize Hepatitis C Virus Infection in Ghana, West Africa', *Journal of Virology*, 77: 7914–23.

Coquery-Vidrovitch, C. (1999) *L'Afrique Et Les Africains Au XIX E Siècle. Mutations, Révolutions, Crises*. Paris: ARMAND COLIN.

Deschamps, H. (1972) *Histoire de la Traite Des Noirs, de l'Antiquité À Nos Jours*. Paris: Fayard.

Drexler, J. F. et al. (2009) 'A Novel Diagnostic Target in the Hepatitis C Virus Genome', *PLoS Medicine*, 6: e31.

——— et al. (2013) 'Evidence for Novel Hepaciviruses in Rodents', *PLoS Pathogens*, 9: e1003438.

Eltis, D., and Richardson, D. (1995) *Atlas of the Transatlantic Slave Trade*. UK: Yale University Press.

Fichet-Calvet, E. et al. (2007) 'Fluctuation of Abundance and Lassa Virus Prevalence in *Mastomys natalensis* in Guinea, West Africa', *Vector-Borne and Zoonotic Diseases*, 7: 119–28.

——— et al. (2016) 'Spatial and Temporal Evolution of Lassa Virus in the Natural Host Population in Upper Guinea', *Scientific Reports*, 6: 21977.

Fisher, C. R., Streicker, D. G., and Schnell, M. J. (2018) 'The Spread and Evolution of Rabies Virus: Conquering New Frontiers', *Nature Reviews. Microbiology*, 16: 241–55.

Forbi, J. C. et al. (2012) 'Epidemic History of Hepatitis C Virus Infection in Two Remote Communities in Nigeria, West Africa', *Journal of General Virology*, 93: 1410–21.

Hadler, S. et al. (2004) 'Chapter 55: Vaccines'. In: *Vaccination Programs in Developing Countries*. Amsterdam: Elsevier Inc.

ICTV. (2021), *Updated HCV Classification. ICTV Online*. <https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification> accessed 2 August 2021.

Ishak, R. et al. (2017) 'Infectious Agents as Markers of Human Migration toward the Amazon Region of Brazil', *Frontiers in Microbiology*, 8: 1663.

Jeffreys, H. (1961) *The Theory of Probability*. UK: OUP Oxford.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Ki-Zerbo, J. (1994) *Histoire de l'Afrique Noire, D'hier À Demain*. Paris: Hatier.

Klempa, B. et al. (2013) 'Seroepidemiological Study Reveals Regional Co-occurrence of Lassa- and Hantavirus Antibodies in Upper Guinea, West Africa.', *Tropical Medicine & International Health : TM & IH*, 18: 366–71.

Kumar, S. et al. (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Molecular Biology and Evolution*, 35: 1547–9.

Langmead, B., and Salzberg, S. L. (2012) 'Fast Gapped-Read Alignment With Bowtie 2', *Nature Methods*, 9: 357–9.

Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.

Lugan, B. (2001) *Atlas Historique de l'Afrique, Des Origines À Nos Jours*. Paris: Éditions du Rocher.

Lukashev, A. N. et al. (2017) 'Close Genetic Relatedness of Picornaviruses from European and Asian Bats', *Journal of General Virology*, 98: 955–61.

Mari Saez, A. et al. (2018) 'Rodent Control to Fight Lassa Fever: Evaluation and Lessons Learned from a 4-year Study in Upper Guinea', *PLOS Neglected Tropical Diseases*, 12: e0006829.

Marien, J. et al. (2020) 'Households as Hotspots of Lassa Fever? Assessing the Spatial Distribution of Lassa Virus-infected Rodents in Rural Villages of Guinea', *Emerging Microbes & Infections*, 9: 1055–64.

Markov, P. V. et al. (2012) 'Colonial History and Contemporary Transmission Shape the Genetic Diversity of Hepatitis C Virus Genotype 2 in Amsterdam', *Journal of Virology*, 86: 7677–87.

——— et al. (2009) 'Phylogeography and Molecular Epidemiology of Hepatitis C Virus Genotype 2 in Africa', *Journal of General Virology*, 90: 2086–96.

Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.

McFarland, D. M., and Lawrence, R. (1998) 'Historical Dictionary of Burkina Faso African'.

Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008) 'Smooth Skyride through a Rough Skyline: Bayesian Coalescent-based Inference of Population Dynamics', *Molecular Biology and Evolution*, 25: 1459–71.

Parker, J., Rambaut, A., and Pybus, O. G. (2008) 'Correlating Viral Phenotypes with Phylogeny: Accounting for Phylogenetic Uncertainty', *Infection, Genetics and Evolution*, 8: 239–46.

Pasquier, C. et al. (2005) 'Distribution and Heterogeneity of Hepatitis C Genotypes in Hepatitis Patients in Cameroon', *Journal of Medical Virology*, 77: 390–8.

Perz, J. F. et al. (2006) 'The Contributions of Hepatitis B Virus and Hepatitis C Virus Infections to Cirrhosis and Primary Liver Cancer Worldwide', *Journal of Hepatology*, 45: 529–38.

Pétré-Grenouilleau, O. (1998a) *L'Argent de la Traite. Milieu Négrier, Capitalisme Et Développe Ment: Un Modèle*. Paris: Aubier HIstoires.

——— (1998b) *La Traite Des Noirs*. Paris: Qu e sais-je?

Petruzziello, A. et al. (2016) 'Global Epidemiology of Hepatitis C Virus Infection: An Up-date of the Distribution and Circulation of Hepatitis C Virus Genotypes', *World Journal of Gastroenterology*, 22: 7824–40.

Pouillot, R. et al. (2008) 'Variable Epidemic Histories of Hepatitis C Virus Genotype 2 Infection in West Africa and Cameroon', *Infection, Genetics and Evolution*, 8: 676–81.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009) 'FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix', *Molecular Biology and Evolution*, 26: 1641–50.

Purdy, M. A. et al. (2015) 'A Re-evaluation of the Origin of Hepatitis C Virus Genotype 2 in West Africa', *Journal of General Virology*, 96: 2157–64.

Rajhi, M. et al. (2021) 'Identification of Two Novel Hepatitis C Virus Subtype 2 from Tunisia (2v and 2w)', *PLoS One*, 16: e0248249.

Rasche, A. et al. (2019) 'Evolutionary Biology of Human Hepatitis Viruses', *Journal of Hepatology*, 70: 501–20.

Riou, J. et al. (2016) 'Hepatitis C Virus Seroprevalence in Adults in Africa: A Systematic Review and Meta-analysis', *Journal of Viral Hepatitis*, 23: 244–55.

Simmonds, P. (2012) 'SSE: A Nucleotide and Amino Acid Sequence Analysis Platform', *BMC Research Notes*, 5: 1–10.

Smith, D. B. et al. (2014) 'Expanded Classification of Hepatitis C Virus into 7 Genotypes and 67 Subtypes: Updated Criteria and Genotype Assignment Web Resource', *Hepatology*, 59: 318–27.

Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Virus Evolution*, 4: vey016.

Viazov, S., Widell, A., and Nordenfelt, E. (2000) 'Mixed Infection with Two Types of Hepatitis C Virus Is Probably a Rare Event', *Infection*, 28: 21–5.

Wolf, J. M. et al. (2021) 'Temporal and Geographic Spreading of Hepatitis B Virus Genotype A (HBV-A) in Brazil and the Americas', *Journal of Viral Hepatitis*, 28: 1130–40.

Zein, N. N. (2000) 'Clinical Significance of Hepatitis C Virus Genotypes', *Clinical Microbiology Reviews*, 13: 223–35.