Infection, Genetics and Evolution 75 (2019) 103944

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Tracing the epidemic history of hepatitis C virus genotype 1b in Tunisia and in the world, using a Bayesian coalescent approach

Marwa Khedhiri[a,c,d,e,*], Kais Ghedira[b], Anissa Chouikha[a,c,d], Henda Touzi[a,d], Amel Sadraoui[a,d], Walid Hammemi[a,d], Henda Triki[a,c,d,f]

[a] Laboratory of Clinical Virology, Pasteur Institute of Tunis, University Tunis El Manar, Tunis, Tunisia
[b] Laboratory of Bioinformatics, Biomathematics and Biostatistics - LR16IPT09, Pasteur Institute of Tunis, University Tunis El Manar, Tunis, Tunisia
[c] Research Laboratory: "Transmission Controle et Immunobiologie des Infections" (LR11-IPT02), Pasteur Institute of Tunis, Tunisia
[d] Clinical Investigation Center (CIC), Pasteur Institute of Tunis, University Tunis El Manar, Tunis, Tunisia
[e] Faculty of Sciences of Tunis, University Tunis El Manar, Tunis, Tunisia
[f] Faculty of Medicine of Tunis, University Tunis El Manar, Tunis, Tunisia

1. Background

As a serious public health problem worldwide, hepatitis C virus (HCV) infects more than 177.5 million people or ~3% of the world's population (Li et al., 2017). HCV is the major cause of cirrhosis and hepatocellular carcinoma (Di Bisceglie, 1997; El-Serag, 2002). It has been classified into eight major genotypes (1 to 8) and 86 confirmed subtypes (Borgia et al., 2018). HCV genotyping is of great importance to choose therapeutic molecules, doses and duration of treatment (EASL, 2018). However, when the pan-genotypic HCV drug regimens is available, it would be possible to treat without genotyping and subtyping. It was used for the interpretation of resistance-associated substitutions (RAS) (Lontok et al., 2015) as well as for molecular epidemiology studies (Messina et al., 2015). The global distribution of HCV genotypes and subtypes changes over time as a result of novel transmission routes and population movements (Pawlotsky et al., 1995; Pol et al., 1995). Some of the HCV viruses are distributed globally such as subtypes 1a, 1b and 3a, termed 'epidemic' strains due to their high prevalence and lower genetic diversity. These subtypes started to spread quickly during the twentieth century with the transmission via non-sterile medical injections and Injecting Drug Use (IDU) (Silini et al., 1995; Pybus et al., 2001; Simmonds et al., 2005; Magiorkinis et al., 2009). In contrast, others 'endemic' strains found only in restricted geographic areas, represent a low-level endemic infection in particular populations. These later include mainly genotype 2 in West Africa, genotype 4 in Central Africa and the Middle East and genotype 6 in south-east Asia (Pybus et al., 2007; Messina et al., 2015).

In North Africa, the distribution of HCV genotypes differs between countries. It has been reported that the most prevalent subtype in Egypt is 4a (Iles et al., 2014), while, HCV-1b is the largely predominant subtype in Morocco (Benani et al., 1997) and Tunisia. Several studies showed an average prevalence of genotype 1b ranging from 79 to 90% of circulating genotypes in Tunisia (Ben Moussa et al., 2003; Othman et al., 2004; Mejri et al., 2005; Ezzikouri et al., 2013). Among the other co-circulating genotypes, subtypes 1a and 2c are the second most frequently isolated in Tunisia (Djebbi et al., 2003; Ezzikouri et al., 2013; Rajhi et al., 2014; Rajhi et al., 2016).

Here, we report a comprehensive phylogenetic and evolutionary study of the most prevalent HCV subtype 1b in chronically infected Tunisian patients together with subtype 1b sequences from other regions of the world. We employed a Bayesian coalescent-based method to investigate the migration history of HCV-1b by analyzing 1636 independent sequences covering part of the NS5B region, and to estimate the date of introduction of the most recent common ancestor (MRCA) of this subtype in Tunisia and in the world. To our Knowledge, this is the first description of historical trends associated to HCV epidemic for subtype 1b in Tunisia and the second study that is performing a phylodynamic and phylogeographic analysis of HCV-1b in the world based on a bayesian approach after Magiorkinis et al. (Magiorkinis et al., 2009).

2. Material and methods

2.1. HCV-NS5b sequences

The Tunisian sequences used in the present study were obtained from the routine diagnostic activity of the Laboratory of Clinical Virology in Pasteur Institute of Tunis. Partial sequencing in the NS5B region has been used as a reference method to genotype HCV as part of pretreatment investigation of the patients since 2012.

Total RNA was extracted from 140 µl of serum using the QIAmp viral RNA mini Kit (QIAGEN, Hilden-Germany) as indicated by the

* Corresponding author at: Laboratory of Clinical Virology, Pasteur Institute of Tunis, 13 Place Pasteur BP-74, 1002 Tunis Belvedere, Tunisia.
E-mail addresses: marwa.kehidiri@gmail.com (M. Khedhiri), kais.ghedira@pasteur.rns.tn (K. Ghedira), anissa.chouikha@pasteur.tn (A. Chouikha), Henda.Triki@pasteur.tn (H. Triki).

https://doi.org/10.1016/j.meegid.2019.103944
Received 10 August 2018; Received in revised form 25 June 2019; Accepted 27 June 2019
Available online 29 June 2019
1567-1348/ © 2019 Elsevier B.V. All rights reserved.

Table 1
Sequences, sense and positions of primers used in the present study.

Primers	5′-3′ Sequences	Sense	Position[a]
DM101	TTCTCR[b]TATGAY[c]ACCCGCTGY[c]TTTGA	Forward	8250–8275
DM100	TACCTV[d]GTCATAGCCTCCGTGAA	Reverse	8616–8638
1 s	TATGAY[c]ACCCGCTGY[c]TTTGAC	Forward	8256–8275

[a] Positions on the reference strain HCV 1a-H77-AF009606.
[b] R = A or G.
[c] Y = C or T.
[d] V = A or C or G.

manufacturers. Extracted RNA (10 µl) was reverse transcribed for 45 min at 42 °C using 1 mM of dNTP mix, 1× transcriptase buffer, 10 mM of DTT, 40 U RNasin (Promega), 200 U of the superscript II Reverse Transcriptase (Invitrogen) and 10 µM of reverse primer DM100 in a total volume of 20 µl. Amplification of a 383-nucleotide (nt) fragment in the NS5B region was carried out with a semi-nested Polymerase Chain Reaction (PCR) using primers specific to relatively conserved sequences in this region (Table 1). For the first round of PCR, 10 µl of cDNA were added to the mix reaction containing 0.2 µM of each outer primers (DM101 and DM100), 5 µl (10×) polymerase buffer, 1.5 mM of MgCl2 and 1 U recombinant-Taq polymerase in a final volume of 50 µl. The same conditions were used in the second round of PCR using 5 µl of the first round PCR product and 0.2 µM of DM100 and 1 s of the inner primers. The PCR cycles consisted of: initial denaturation at 94 °C for 5 min followed by 30 cycles at 94 °C for 30s, 60 °C for 30s and 72 °C for 1 min per cycle. The final elongation step was at 72 °C for 7 min. The PCR products were migrated on a 1% agarose gel and visualized under UV light. After purification on columns of the PCR products using the QIAquick PCR purification Kit (QIAGEN, GmbH, Hilden, Germany), direct sequencing was performed in both DNA senses using the same inner primers by the Big-Dye terminator ready reaction cycle sequencing Kit (Applied Biosystems) in an automated ABI Prism 3130- Genetic Analyser Applied Biosystems® (Applied Biosystems) sequencer.

To investigate the HCV strains genotypes and subtypes, sequence chromatograph files were initially analyzed by the «Molecular Evolutionary Genetic Analysis» MEGA.7.0.9 software (Kumar et al., 2016); the obtained consensus sequences; 271 nt fragment in the NS5B region (8307–8577: Positions on the reference *strain HCV 1a-H77-AF009606*) were finally aligned and compared with reference sequences representative of all HCV types and subtypes (Simmonds et al., 2005; Kuiken and Simmonds, 2009; Smith et al., 2014)

This sequencing activity increased since 2016 due to the implementation of a National Hepatitis C Elimination Program by 2023 in Tunisia (http://www.onmne.tn/fr/dossiers_thematiques.php?id_rub=4&id=122). The sequences from other regions from the world were retrieved from the NCBI GenBank (http://www.ncbi.nlm.nih.gov) database.

2.2. Sample selection

Initially, 11,407 sequences were retrieved from the NCBI GenBank (http://www.ncbi.nlm.nih.gov/nucleotide) database using keyword “HCV 1b NS5B”. Only sequences with known sampling time and location that were unambiguously subtyped as HCV-1b with subtyping tools Genome detective (Vilsker et al., 2019) and COMET (Struck et al., 2014), were retained. A total of 1266 HCV-NS5B-1b sequences from different countries of the world were used in the present study; their accession numbers are available in Supplementary Table S1. The sequences (n = 1266) were obtained over a 23-year period (1996–2018) and originated from 37 different countries Afghanistan (AFG, n = 2), Argentina (ARG, n = 2), Australia (AUS, n = 9), Belgium (BEL, n = 10), Cambodia (KHM, n = 82), Canada (CAN, n = 19), China (CHN, n = 340), Ethiopia (ETH, n = 3), Egypt (EGY, n = 1), France (FRA, n = 34), Germany (DEU, n = 30), Georgia (GEO, n = 5), Hungary (HUN, n = 1), Indonesia (IDN, n = 67), Iran (IRN, n = 5), Ireland (IRL, n = 74), Italy (ITA, n = 26), Japan (JPN, n = 15), Madagascar (MDG, n = 9), Malaysia (MYS, n = 12), Morocco (MAR, n = 16), Netherlands (NLD, n = 6), New Zealand (NZL, n = 13), Philippines (PHL, n = 2), Russia (RUS, n = 53), Romania (ROU, n = 39), Saudi Arabia (SAU, n = 31), South Africa (ZAF, n = 4), Tajikistan (TJK, n = 30), Thailand (THA, n = 62), Turkey (TUR, n = 1), United Kingdom (GBR, n = 19), Uruguay (URY, n = 5), USA (USA, n = 104), Uzbekistan (UZB, n = 5), Venezuela (VNM, n = 1), Vietnam (VNM, n = 129). These sequences were complemented with 370 Tunisian subtype 1b sequences which were collected from serologically HCV positive Tunisian patients between 2012 and 2018: 20, 46, 56, 18, 177, 48 and 5 sequences of HCV-NS5B-1b obtained in 2012, 2013, 2014, 2015, 2016, 2017 and 2018, respectively.

2.3. Bayesian evolutionary analysis

Based on the above mentioned HCV-1b segment NS5B sequences, an evolutionary timescale tree was estimated using the MCMC algorithm being part of the BEAST package (version 1.8.1) (Drummond and Rambaut, 2007). TreeAnnotator program was used to summarize and construct a tree from the resulting set of credible trees (10,000 trees) by choosing the one with the maximum product of posterior probabilities after a 10% burn-in. The final tree was denoted the MCC (Maximum Clade Credibility) tree. Phylogenetic tree was saved in Newick format then integrated into iTOL (Interactive Tree Of Life) web server (Letunic and Bork, 2016) to generate a graphic representation of the relationship between all HCV-1b investigated sequences. Two phylogenetic trees have been constructed including only HCV-1b Tunisian sequences (n = 370) then all HCV-1b sequences investigated in the present study (n = 1636).

2.4. Coalescent analysis

The epidemic history of HCV-1b was investigated with a coalescent-based strategy. Briefly, HCV-1b evolutionary history was inferred by Bayesian Markov Chain Monte Carlo (MCMC) analysis implemented in BEAST software version 1.8.1 (http://beast.bio.ed.ac.uk) (Drummond and Rambaut, 2007). For this analysis, five parametric models of coalescent population growth were used: bayesian skyline, constant population size, exponential growth, expansion growth and logistic growth under the strict clock, uncorrelated lognormal and exponential relaxed clocks. The SMS tool (Smart Model Selection) (Lefort et al., 2017) included in PhyML 3.0 (Guindon et al., 2010) was used to assess an automatic model selection using likelihood-based criteria (AIC) as a first step to conduct our coalescent analysis. The general time reversible (GTR) nucleotide-substitution model (Nei and Kumar, 2000) with gamma-distributed rates among sites and a proportion of invariable sites was used under the strict then relaxed molecular clocks (Table 2).

The MCMC chains were run for 20 10^8 generations in order to

Table 2
Smart model selection using PhyML.

Model	Decoration	K	Lik	AIC	BIC
GTR	+G + I	3279	−34,075,59,819	74,709,19,638	86,520,54,399
TN93	+G + I	3276	−34,106,58,259	74,765,16,718	86,565,70,843
GTR	+G	3278	−34,227,48,480	75,010,96,960	86,818,71,509
GTR	+I	3278	−39,071,30,799	94,698,61,598	96,506,36,147
GTR		3277	−42,003,92,925	90,561,85,850	102,366,00187

K: number of model free parameter.
Substitution model: GTR.
Equilibrium frequencies: ML optimized.
Proportion of invariable sites: estimated (0.332).
Number of substitution rates: 4.
Gamma shape parameter: estimated (0.583).

achieve an effective sample size (ESS) ≥200 and sampled every 2 million steps. A substitution rate of 9.13×10^{-4} substitutions per site per year (s/s/y) was used according to previous estimation (Al Qahtani et al., 2017) for the NS5B region. Best fitting models were selected by calculation of a Bayes Factor (BF) (Kass and Raftery, 1995), using Marginal Likelihoods Estimation (MLE) specifically path sampling (PS) (Ogata, 1989; Gelman and Meng, 1998) and stepping-stone (SS) (Xie et al., 2011) sampling implemented in Beast v1.8.1 (Baele et al., 2012).

The effective number of infections through time was estimated by using the Bayesian skyline plot (BSP) approach (Drummond et al., 2005) under the strict and relaxed molecular clock models. All BEAST output log files were analyzed with TRACER v1.6 program (available from http://tree.bio.ed.ac.uk/software/tracer/). The present analysis was run on one HCV-1b datasets of NS5B sequences: (n = 1636, n = 370, 271 nt).

2.5. Phylogeography analysis

In order to reduce computational burden in subsequent Bayesian analysis, we selected a subset of sequences to decrease their number. Briefly, from each sampling year and location the two most divergent sequences were selected with the phylogenetic diversity algorithm (PDA) (Chernomor et al., 2015) which selects the subtree of n taxa connected by the largest sum of branch lengths.

Subsequently, the phylogeographic analysis was performed using the BEAST package (Lemey et al., 2009; Drummond et al., 2012) to estimate the diffusion process of the HCV-1b in the world and particularly in Tunisia. The capital of each country was used as a discrete location state, and the sampling date was used to calibrate the time scale. As with the molecular clock analysis, the preliminary MCMC analysis were performed for each combination of clock and tree prior model, and MCMC analysis were run using the best-supported combination under Bayes factors. After removing 10% of the burn-in, those runs were combined to construct an MCC tree. The MCC tree was uploaded into SPREAD3 (Bielejec et al., 2016) and projected into a Geojson map format retrieved from https://geojson-maps.ash.ms/ to generate a keyhole markup language file (.kml). The selected sequences isolates were assigned to a total of 38 distinct geographic groups corresponding to each country. In order to provide a spatial projection, the migration routes indicated by the tree were visualized using Google Earth Pro version 7.3. To identify significant migrations, the Bayes factor testing was performed using the Bayesian stochastic search variable selection (BSSVS) procedure (Lemey et al., 2009). The migrations with Bayes factors that were greater than five were summarized as well-supported migrations (Kass and Raftery, 1995).

3. Results

3.1. Evolutionary analysis

The phylogenetic tree in Fig. 1 corresponds to the maximum clade credibility tree (MCC) including the 370 Tunisian HCV-1b-NS5B sequences obtained from different locations in Tunisia. The tree shows that at least two HCV-1b lineages are circulating in Tunisia with relatively high bootstrap support values; however, no consistent geographical or time related cluster could be identified. These lineages included sequences from different region of Tunisia with different years of isolation.

Similarly, two HCV-1b clades could be clearly distinguished in Fig. 2 forming two potential monophyletic groups with relatively high bootstrap values, the first (in green), including HCV-1b from mainly Asian countries and another (in red) grouping HCV-1b sequences from mainly European and African countries including Tunisia. The later phylogenetic tree shows the MCC including the 1636 HCV-1b-NS5B sequences used in the present study and originating from 38 countries. The tree shows that there were multiple independent sub-epidemics of HCV-1b circulating in Tunisia. Most of Tunisian sequences (labels with red background color) clustered in a large clade including sequences from Iran and Morocco. Several other Tunisian sequences are dispersed across the phylogenetic tree into several independent clades without an apparent pattern qua geographical origin and date of isolation.

3.2. Epidemic history of HCV1b

3.2.1. Bayesian skyline plots

Bayesian coalescent analysis was performed on the 1636 HCV-1b segment NS5B sequences from Tunisia and other countries. For each demographic and molecular clock models, chain lengths of 200 million were used and sampled every 2,000,000 states. The data was analyzed using the Bayesian Skyline Plot (BSP) that depicts the estimated change in the effective number of infected individuals over time for the HCV-1b virus. Fig. 3a exhibited an epidemic history, characterized by five phases of epidemic population growth including: an initial period of relatively constant population size, followed by a brief phase of exponential growth between 1936 and 1958, a third period of a relatively constant population size until 1970, a fourth period of high exponential growth from 1970 to 1975 and a final phase with constant population size from 1975 to, 2018. As the Tunisian HCV-1b sequences (370 sequences in total) did not originate from a single well-supported monophyletic cluster, a BSP was computed only for the 253 HCV-1b Tunisian sequences forming the large monophyletic clade to estimate the change in the effective number of infected individuals over the time. Fig. 3b exhibited an epidemic population growth history, characterized by five phases including: an initial period of a very slight decrease in population size, followed by a phase of exponential growth between 1960 and 1985, a third period of a relatively constant population size until 2012, followed by a short phase of exponential decrease until 2015 then a constant phase from 2015 to 2018.

3.2.2. Most recent common ancestor (tMRCA)

The time to the most recent common ancestor (tMRCA) was evaluated by using a fixed substitution rate for the NS5B region of 9.13×10^{-4} substitutions per site and per year according to Al-Qahtani study (Al Qahtani et al., 2017). The tMRCA of the progenitor virus that ultimately gave rise to all Tunisian HCV-1b was estimated around 1923 with a HPD (Highest Posterior Density) 95% [1904, 1942] and the one that ultimately gave rise to all HCV-1b circulating in the world was estimated around 1903 with a HPD95% [1841, 1930], respectively (Table 3).

3.2.3. Phylogeography analysis

3.2.3.1. Spatial and temporal dynamics of HCV-1b virus geographic dispersal. To gain insight into the spatial temporal dynamics of the diffusion process of HCV-1b in Tunisia and in the world, we mapped the spatial estimates annotated in the partial NS5B region MCC trees on Google Earth Pro version 7.3 (http://www.google.com/earth/download/ge/). This mapping allows to visualize the virus's geographic spread process over time. The links between different geographic regions represent branches in the MCC tree on which virus migration occur and circle areas reflect the number of branches maintaining a particular location at that time point. To reduce computational burden in phylogeographic analysis, we used PDA tool to select a subset of sequences corresponding to the most divergent sequences from each sampling location and year of isolation. Using this approach we have selected 219 HCV-1b sequences (among 1636 total HCV-1b sequences) from 38 countries and with isolation date ranging from 1996 to 2018. The panels in Fig. 4 show the temporal dynamics of HCV-1b spatial dispersal processes in the world.

Our study revealed that the most ancient HCV-1b virus date around 1903 in the world. Based on the investigated HCV-1b sequences the phylodynamic and phylogeographic analysis revealed that the virus started its spread in the Netherlands and its migration to China around

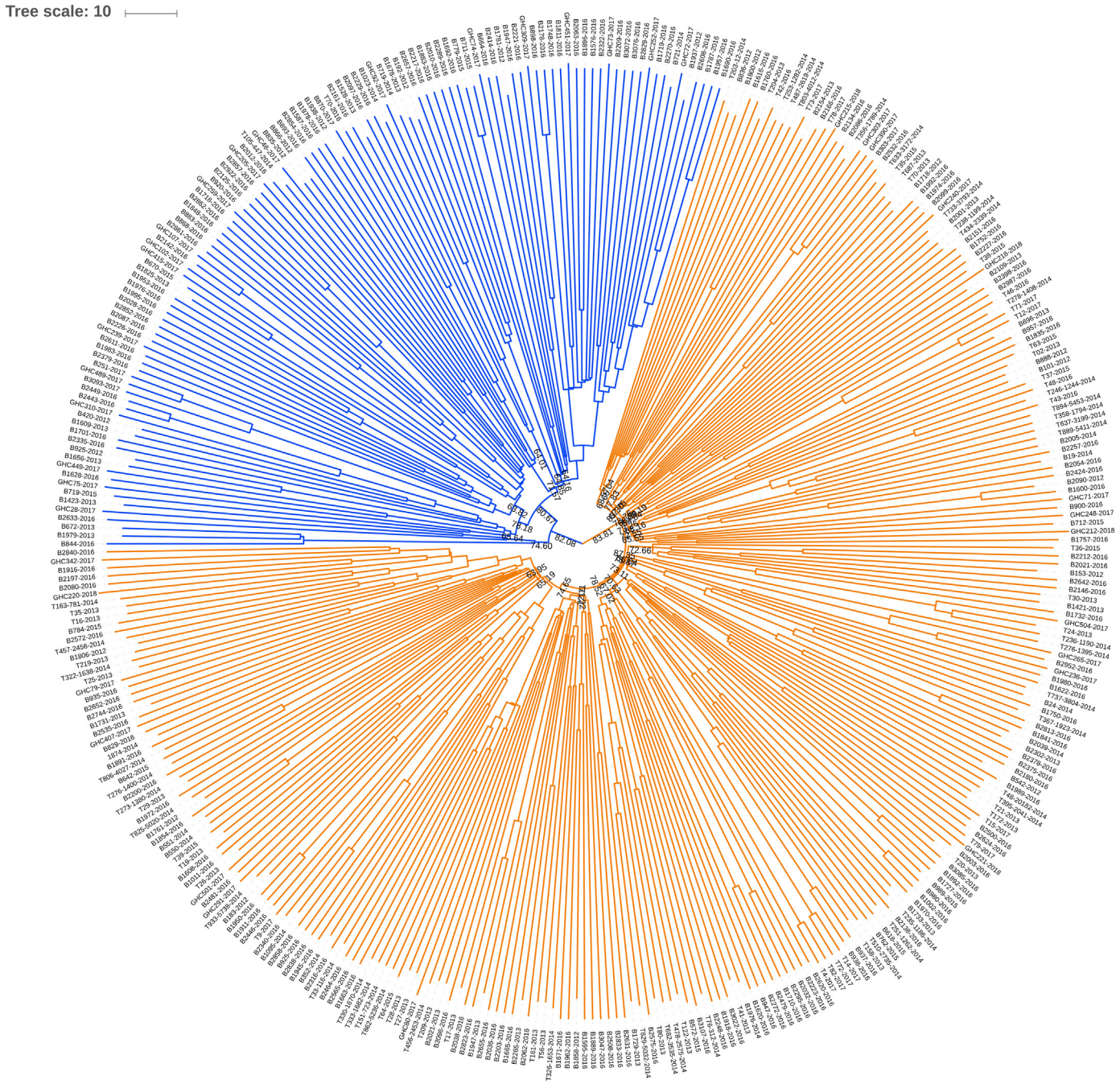

Fig. 1. Phylogenetic analysis of HCV 1b Tunisian sequences. Maximum clade credibility (MCC) tree representing the rooted genealogy of the 370 Tunisian HCV-1b sequences obtained by analyzing the 271-nucleotide NS5B segment sequences using a Bayesian evolutionary analysis. The main lineages are highlighted in different colors. Bootstrap support values were highlighted on the tree.

1927. Soon after, a large dispersal event occurred in which the virus migrated from Netherlands to neighbor European countries including Germany, Italy and Ireland around 1951, Georgia around 1964, Romania in 1993 and France later in 2016. Throughout this period, HCV-1b continues its migration from Netherlands to several all other locations worldwide including Indonesia in 1955, Canada and New Zealand around 1959, Madagascar around 1963, Uzbekistan in 1967, Cambodia and Saudi Arabia around 1970, South Africa around 1971, Thailand and Japan in 1974, Philippines in 1981 and Venezuela around 2004. The virus migrated later from Madagascar to Canada again in 2008. Meanwhile, several migrations of the virus occurred leading to the spread of the virus worldwide. Despite spreading from the primary epicenter in Netherlands starting from 1927, several other epicenters appeared including Ireland, Germany, Italy and Tunisia from 1951 to 1963 (Fig. 4). Regarding Tunisia and North Africa, the virus seems to migrate from the Netherland to Tunisia around 1953 corresponding to the time of its introduction in Tunisia. Apart of the first wave of migration from European countries, a second wave of migration started around 1951 from China to Turkey in 1956, from Turkey to Russia in 1962, then from Russia to reach Germany again in 2014. The virus migrated also from China to reach Thailand around 2007. HCV-1b knew its maximum spread in China from 1973 to 2005 followed by a significant decline. The virus appeared also to spread significantly from 1951 to 1977 in Europe, Asia and Africa contributing to its migration worldwide. Indeed, HCV-1b continued its migration from Netherlands and other European countries to reach Iran in 1954, Tajikistan in 1957

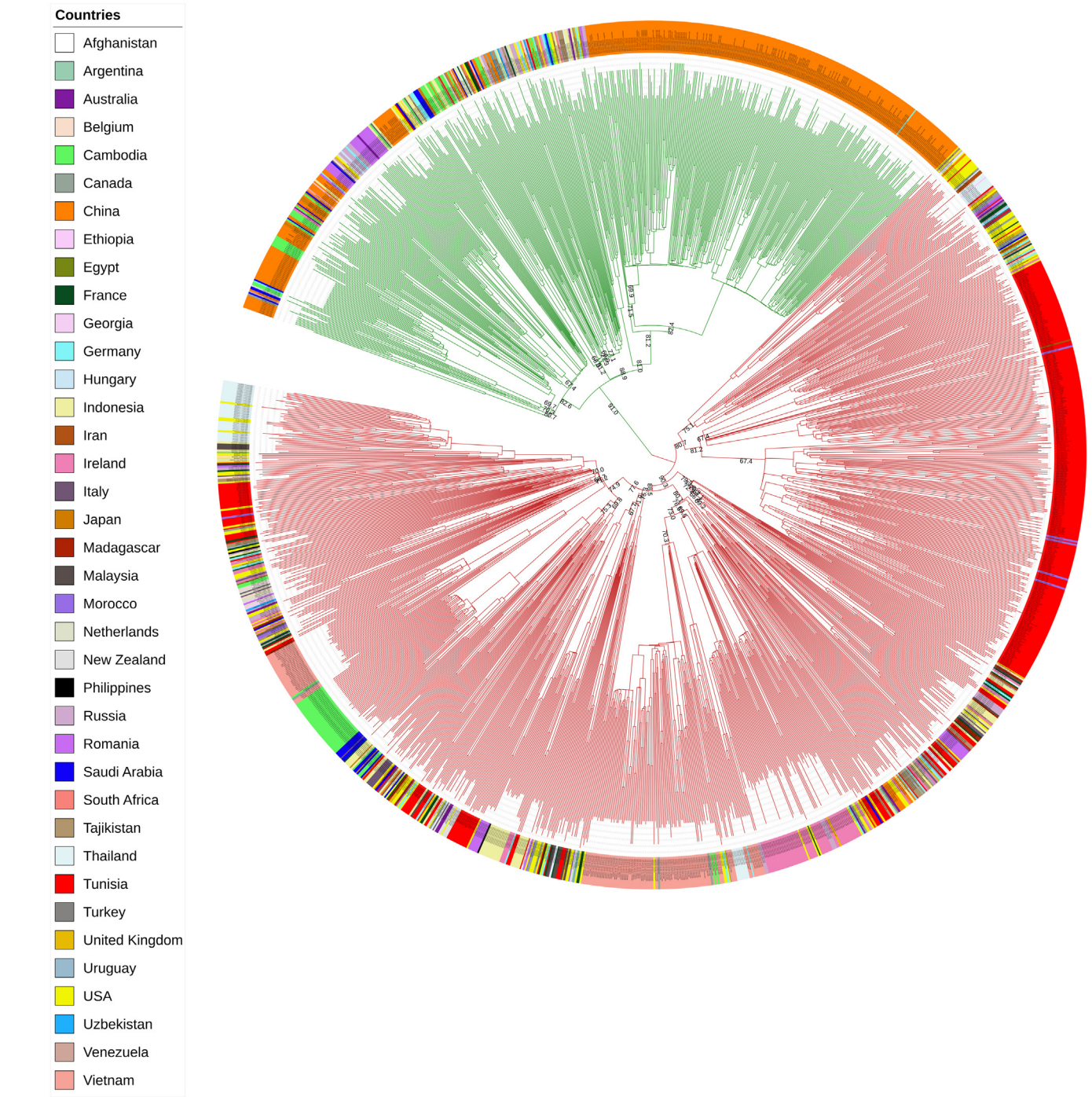

Fig. 2. Maximum clade credibility (MCC) tree. Maximum clade credibility tree representing the rooted genealogy of the 1636 HCV-1b sequences obtained by analyzing the 271-nucleotide NS5B segment sequences using a Bayesian evolutionary analysis. The main lineages are highlighted in different colors. Bootstrap support values were highlighted on the tree.

and Uzbekistan in 1970 where it knew a big spread until 1976 then migrated to Indonesia around 2006. HCV-1b knows a large spread in Tunisia in 1963 and started its migration to Morocco around this year. At the same time the virus knew a third wave of migrations from Italy to Russia, Uruguay, Germany in 2007 and Malaysia around 2010, from Germany to Belgium in 2007, from Ireland to Ethiopia in 2013 and several other countries worldwide. Remarkably, since 2003 the HCV-1b spread in Netherlands started to decline. Similarly, the decline concerned several other countries including Georgia in 2010, Romania in 2013, Italy and Ireland starting from 2012, Tunisia and Germany from 2015 to 2017.

To identify statistically significant transmission routes, we conducted Bayes factor (BF) test for significant non-zero rates between different locations. Only rates supported by a BF greater than 5 are indicated in Fig. 5.

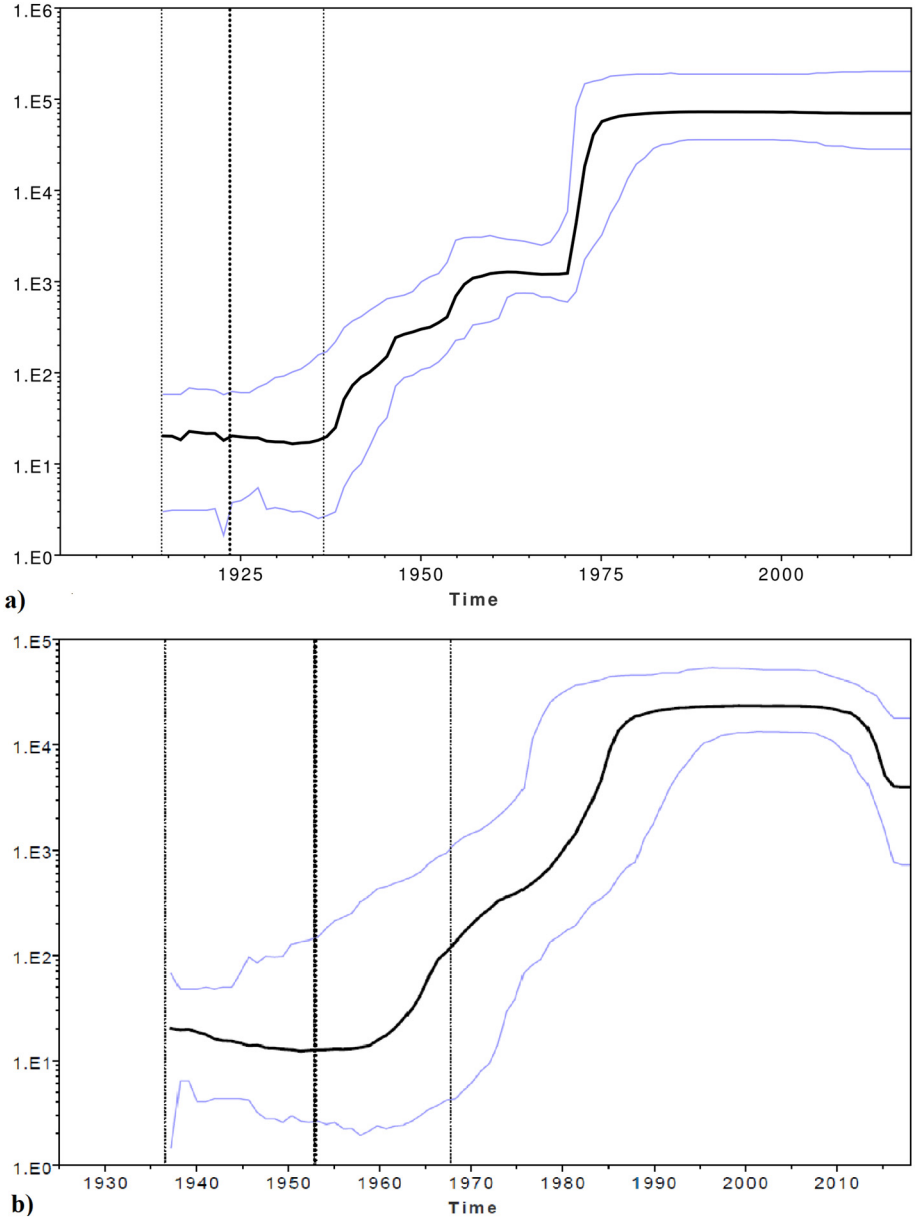

Fig. 3. Bayesian Skyline Plots for Demographic Reconstruction using NS5B sequences. A) In the world, B) In Tunisia. X-axis: Date in Years; Y-axis: Estimated effective number of infections; Bold Line: The median estimate of the estimated effective population size. Upper and lower Lines: Upper and Lower HPD95% (High Population Density) of Effective Number of viral population.

Table 3
tMRCA estimations using the NS5B region for the HCV-1b.

HCV-1b geographic isolation	Number of isolates	MRCA (HPD[a] 95%) (year)
In the world	1636	1903 [1841,1930]
In Tunisia (All regions)	370	1923 [1904,1942]

[a] : HPD: highest posterior density.

4. Discussion

As one of the most prevalent HCV subtype worldwide, the historical movement of HCV-1b represents an interesting question in relation to HCV epidemiology and evolution. Previous reports suggested that the global dissemination of HCV subtype 1b initiated during and after the second world war through the widespread use of blood transfusion (Kendrick, 1964; Smith et al., 1997; Busch et al., 2003) and unsterilized needles and glass syringes (Drucker et al., 2001; Yerly et al., 2001; Tanaka et al., 2005). Furthermore, the observed epidemic development of HCV-1b matches with the international dissemination of contaminated blood and blood products during the twentieth century (Messina et al., 2015). This epidemic subtype has been described as the most prevalent HCV subtype in many countries around the world, such as Argentina (Di Lello et al., 2008), China (Lu et al., 2005), Italy (Ferraro et al., 2008), Algeria (Ezzikouri et al., 2013), Morocco (Benani et al., 1997) and Tunisia (Ben Moussa et al., 2003; Djebbi et al., 2003; Othman et al., 2004; Mejri et al., 2005; Ezzikouri et al., 2013).

Recent molecular epidemiological methods were used successfully to reconstruct the evolutionary history of HCV infection in some countries (Pybus et al., 2009; Lampe et al., 2010; Castells et al., 2015; Wu et al., 2016; Khan et al., 2017). Presently, a lot of studies that investigated the evolutionary and demographic history of HCV-1b are available. Indeed, some studies highlighted its time of introduction in different countries in the world including Argentina (Di Lello et al., 2008), China (Nakano et al., 2006), Italy (Ferraro et al., 2008) and

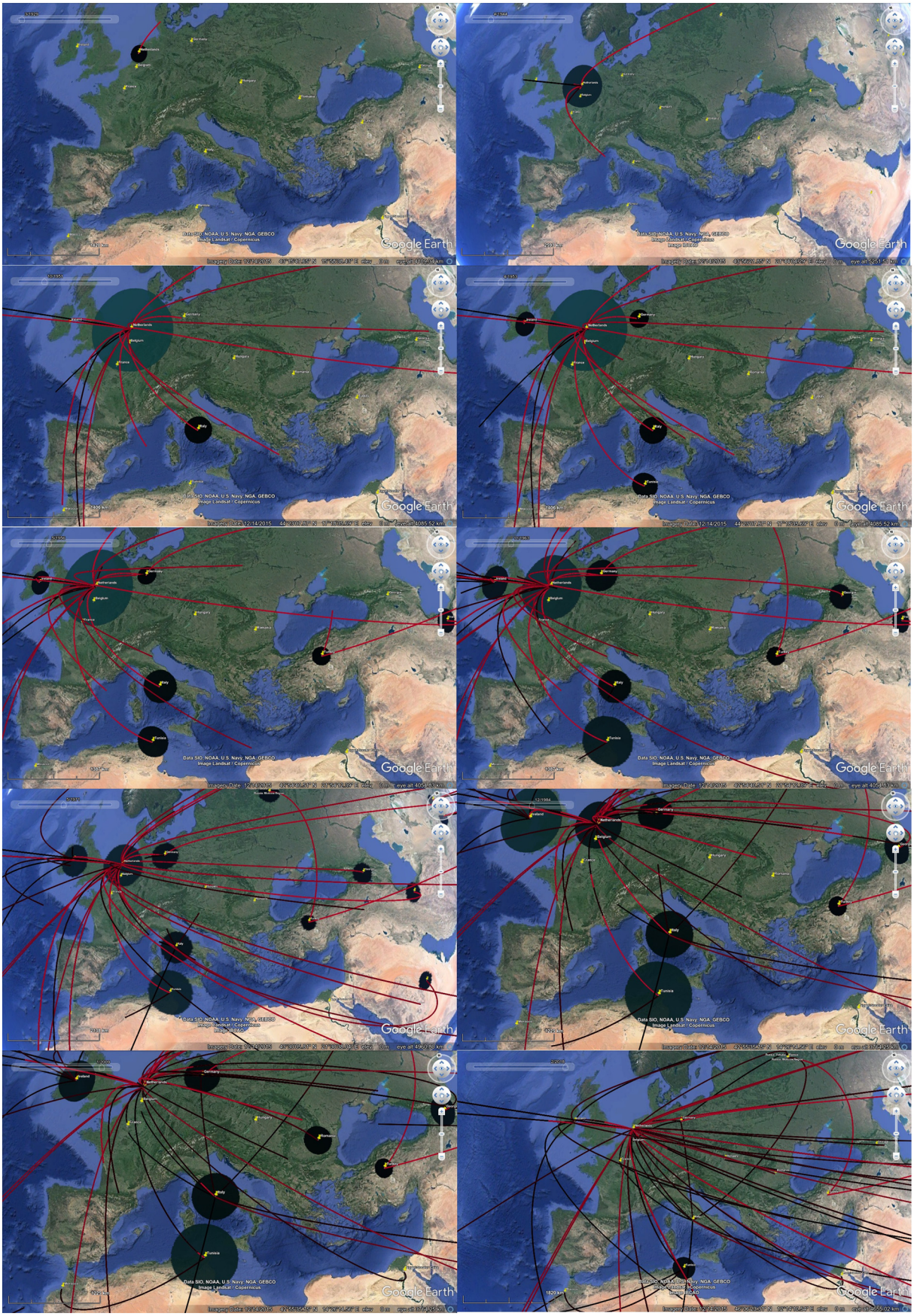

Fig. 4. Temporal dynamics of HCV-1b geographic dispersal in Tunisia. The snapshots show the dispersal pattern of HCV-1b virus for 1929, 1944, 1951, 1953, 1956, 1963, 1971, 1984, 2009 and 2018. Connections between different locations represent branches in the MCC tree along which the relevant location transition occurs. Location circle diameters are proportional to the number of lineages that maintain that particular location state at each time point. This map is produced by satellite pictures made available in Google Earth Pro.

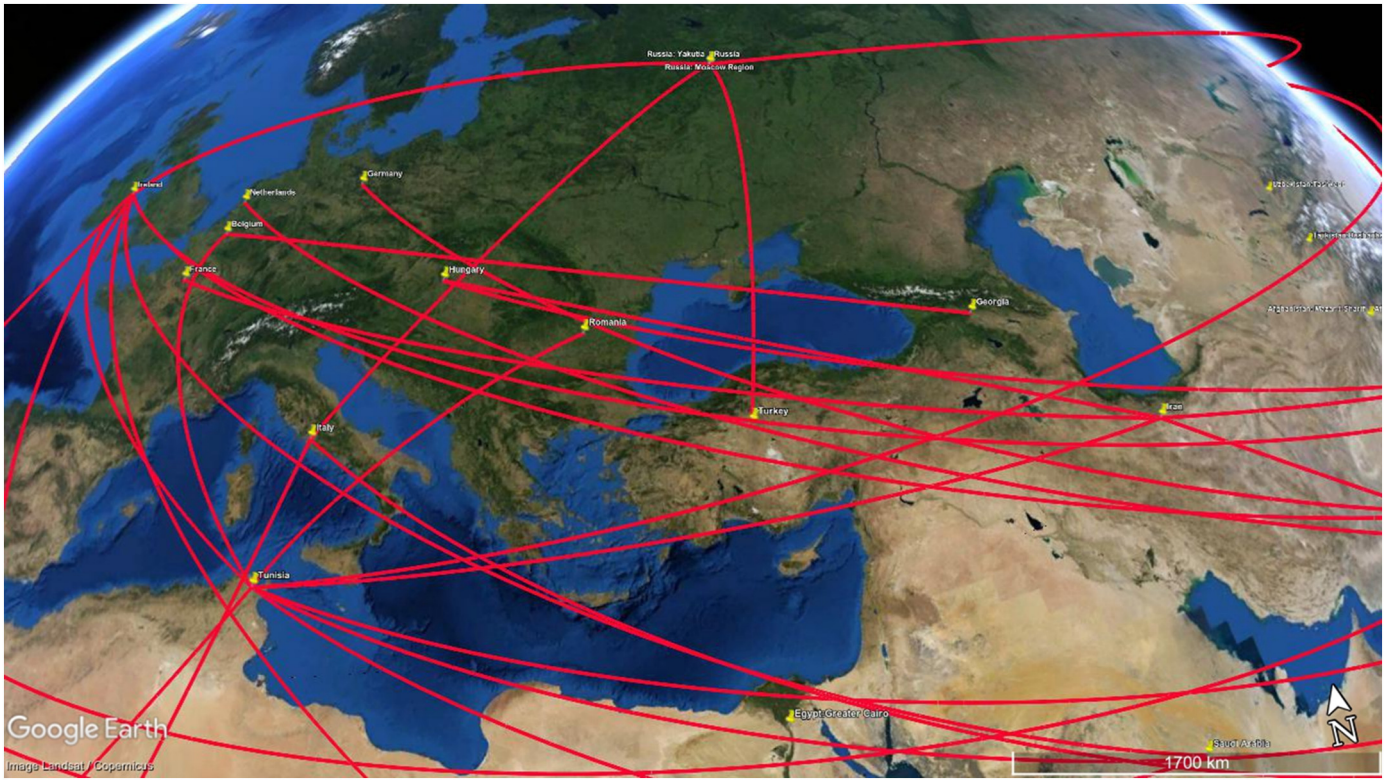

Fig. 5. Significant dispersion routes for HCV-1b using a Bayes factor (BF) test. Lines between different locations indicate transmission routes supported by a BF values. Only rates supported by a BF greater than 5 are indicated.

Morocco (Benani et al., 1997). Another study investigated its emergence all over the world (Magiorkinis et al., 2009).

In Tunisia, Rajhi and co-workers (Rajhi et al., 2016) investigated the HCV genotype 2 epidemic history and estimated the time of its introduction in Tunisia based on a Bayesian study. In the present work, we investigated the epidemic and phylogeographic history of HCV subtype 1b in Tunisia and in the world. Our phylogenetic analysis performed on 370 HCV-NS5B-1b sequences, obtained between 2012 and, 2018 shows that Tunisian sequences are at least divided into two distinct monophyletic groups (Fig. 1). The phylogenetic tree including all the 1636 HCV-1b sequences (Fig. 2) allowed to delineate two HCV-1b lineages or clusters including HCV-1b mainly from China (Labels with Orange background color) and another one grouping HCV-1b sequences from over the world including HCV-1b sequences from Tunisia (Labels with Red background color) (Fig. 2). This tree suggests the existence of several independent HCV-1b variants that are circulating all over the world including Tunisia.

Using a Bayesian coalescent inference of past population dynamics from the 1636 HCV-NS5B-1b sequences circulating in the world, the estimated age of the most recent common ancestor (MRCA) was around 1903 with a HPD 95% 1841–1930 in the world. The estimated age of the progenitor virus that ultimately gave rise to all Tunisian HCV-1b was around 1923 with a HPD 95% 1904–1942 while the date of introduction of HCV-1b in Tunisia was estimated around 1953. This estimated date for HCV-1b in Tunisia followed what has been suggested by Rajhi et al. in 2016 for HCV-2c and HCV-2 k, for which the date of introduction in Tunisia was estimated in 1886 and in 1901, respectively (Rajhi et al., 2016). While Magiorkinis et al. estimated the HCV-1b tMRCA around 1911 with a confidence interval ranging from 1806 to 1959, our study revealed a nearly similar origin of the HCV-1b in the world around 1903 with a confidence interval ranging from 1841 to 1930. Despite the difference in the number of sequences and the number of represented countries the two studies are highly similar. Thus, we believe that the present study as well as Magiorkinis study may well reflect the global HCV epidemic. Based on the sequences included in this study, our phylodynamic and phylogeographic analysis showed that the most ancient HCV-1b sequences are from the Netherlands. The Netherlands is among the countries in Europe with the most diverse geographical origin, moreover, because of its policies, the Netherlands attracts foreign drug users and male homosexuals, two populations known to be at higher risk for viral infection (Paraskevis et al., 2009). However, due to the non availability of HCV-1b sequences from other countries, especially from Sub Saharan and West Africa, the initial site of emergence of HCV-1b may be different. Previous studies suggested that HCV genotype 1 would have originated from West Africa (Smith et al., 1997; Simmonds, 2004; Ezzikouri et al., 2013). Regarding North Africa, the epidemic history of HCV-1b is still insufficiently documented. To our knowledge, the unique report is the one of Brahim et al. who estimated the date of introduction of subtype 1b in the Moroccan population around 1910 (Brahim et al., 2011). Our results report that the date of introduction of HCV-1b in Morocco occurs longtime after. Magiorkinis and co-workers showed that the most likely origin of the currently circulating lineages of HCV-1b is not in Europe but in the USA (Magiorkinis et al., 2009). Our inferred origin is most likely the result of the sampling in combination with the sensitivity of the used discrete phylogeographic model to uneven sampling across locations (Baele et al., 2018). Indeed discrete trait analyses approaches compared to structured coalescent and continuous diffusion models, make a number of restrictive assumptions that might be inappropriate when applied to the migration of lineages between geographic locations by under representing ancestral trait uncertainty and are known to be sensitive to biased sampling of subpopulations (De Maio et al., 2015).

While our study presents some strengths based on the facts that: i) all investigated sequences were subtyped HCV-1b using two distinct tools (Genome detective and COMET), ii) the sequences are representative of 38 countries and iii) we used a strong temporal signal (Al Qahtani et al., 2017), we should notice that it has, however, some limitations. Indeed, relying on a short stretch of the HCV-1b genome

comes at the cost of potentially missing spatial linkages (Vrancken et al., 2018). To avoid this problem, the use of full genome sequences would be better, however due to the lack of full genome sequences from several countries; the focus on small sequences or regions is most commonly performed.

The two BSPs (Fig. 3a and Fig. 3b) summarize the spread and epidemic growth of the globally prevalent HCV genotypes 1b in the world and in Tunisia, respectively. Fig. 3a clearly shows that subtype 1b epidemic was in a steady non expanding phase from 1850 to 1936 followed by an exponential phase from 1936 to 1958. This period overlaps with the second world war (1939–1945) that corresponds to a high rate of population movement from European countries all over the world, escaping the war, and the spread of the HCV-1b worldwide through the widespread use of blood transfusion. This phase was followed by another steady non expanding phase from 1958 to 1970, then a significant exponential growth from 1970 to 1975 and a final phase with constant population growth from 1975 to 2018. However, Fig. 3b shows a slight decrease in population size from 1940 to 1960, followed by a small phase of exponential growth between 1960 and 1985, a third period of a relatively constant population size until 2012, followed by a short phase of exponential decrease in population size between 2012 and 2015 and a constant population size from 2015 to 2018. During this first exponential period (1960 to 1985), Tunisia would have experienced a significant rate of population movements with European countries just after the independence date (1956). Same trend has been observed in the world between 1970 and 1975. The two BSPs seems to share a constant population growth from 1985 to 2012. However, there is a possibility that the posterior estimate may be unintentionally biased by parameters having little or no prior information. Bayesian skyride method could be more appropriate as it places less emphasis on priors based on Gaussian Markov random fields to smooth the effective population size over time (De Silva et al., 2012).

The population growth in Tunisia knew a decline from around 2012. This date may correspond to the Tunisian revolution known as spring revolution in which Tunisia knew a decrease in population movements. Regarding the other subtypes of HCV, our Tunisian HCV-1b tMRCA estimation around 1923 (1904–1942) followed the introduction of HCV-2c in Tunisia with tMRCA around 1886 (1869–1902), HCV-2 k tMRCA 1901 (1867–1931) (Rajhi et al., 2016).

5. Conclusion

The present study combined demographic, historical and phylogeographic analysis to reconstruct the transmission history of the HCV-1b in the world and particularly in Tunisia. To our knowledge, it is the first study to describe the HCV-1b evolutionary history in Tunisia based on a Bayesian Coalescent Approach and the second analysis performed in North Africa after Brahim et al. study. In conclusion, our findings showed that the origin of the HCV-1b is around 1903 (the beginning of the 20th Century) and confirmed what have been suggested by several studies, underlined the strong epidemiological link between Maghrebian countries (Tunisia and Morocco) with European countries.

Funding statement

This study was partially funded by the Ministry of Health and the Ministry of High Education and Scientific Research, Tunisia (PRF-4; CIC2016IPT02, LR16IPT09).

Authors' contributions

Conceived and designed the experiments: MK, AC, HTo, AS and WH. Performed the experiments: MK and AC. Analyzed the data: KG and MK. Contributed reagents/materials/analysis tools: MK, KG, AC, HTo, AS and WH. Wrote the paper: MK and KG. Performed the Bayesian coalescent analysis: KG. Supervised the work: HTr. Revised the manuscript: MK, KG and HTr.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.103944.

References

Al Qahtani, A.A., Baele, G., Khalaf, N., Suchard, M.A., Al-Anazi, M.R., Abdo, A.A., Sanai, F.M., Al-Ashgar, H.I., Khan, M.Q., Al-Ahdal, M.N., Lemey, P., Vrancken, B., 2017. The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia. Sci. Rep. 7, 44947. https://doi.org/10.1038/srep44947.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 2157–2167. https://doi.org/10.1093/molbev/mss084. (Epub 2012 Mar 7).

Baele, G., Dellicour, S., Suchard, M.A., Lemey, P., Vrancken, B., 2018. Recent advances in computational phylodynamics. Curr. Opinion Virol. 31, 24–32. https://doi.org/10.1016/j.coviro.2018.08.009.

Ben Moussa, M., Barguellil, F., Bouziani, A., Amor, A., 2003. Comparison of two hepatitis C virus typing assays in a Tunisian population. Ann. Biol. Clin. 61 (2), 234–238 (French).

Benani, A., El-Turk, J., Benjelloun, S., Sekkat, S., Nadifi, S., Hda, N., Benslimane, A., 1997. HCV genotypes in Morocco. J. Med. Virol. 52 (4), 396–398.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol. Biol. Evol. 33 (8), 2167–2169. https://doi.org/10.1093/molbev/msw082.

Borgia, S.M., Hedskog, C., Parhy, B., Hyland, R.H., Stamm, L.M., Brainard, D.M., Subramanian, M.G., McHutchison, J.G., Mo, H., Svarovskaia, E., Shafran, S.D., 2018. Identification of a novel hepatitis C virus genotype from Punjab, India: expanding classification of hepatitis C virus into 8 genotypes. J. Infect. Dis. 218 (11), 1722–1729. https://doi.org/10.1093/infdis/jiy401. 20.

Brahim, I., Akil, A., Mtairag El, M., Pouillot, R., Malki, A.E., Nadir, S., Alaoui, R., Njouom, R., Pineau, P., Ezzikouri, S., Benjelloun, S., 2011. Morocco underwent a drift of circulating hepatitis C virus subtypes in recent decades. Arch. Virol. 157 (3), 515–520. https://doi.org/10.1007/s00705-011-1193-7.

Busch, M.P., Kleinman, S.H., Nemo, G.J., 2003. Current and emerging infectious risks of blood transfusions. JAMA 289 (8), 959–962.

Castells, M., Bello, G., Ifrán, S., Pereyra, S., Boschi, S., Uriarte, R., Cristina, J., Colina, R., 2015. Epidemic history of major genotypes of hepatitis C virus in Uruguay. Infect. Genet. Evol. 32, 231–238. https://doi.org/10.1016/j.meegid.2015.03.021.

Chernomor, O., Minh, B.Q., Forest, F., Klaere, S., Ingram, T., Henzinger, M., von Haeseler, A., 2015. Split diversity in constrained conservation prioritization using integer linear programming. Methods Ecol. Evol. 6 (1), 83–91.

De Maio, N., Wu, C.H., O'Reilly, K.M., Wilson, D., 2015. New routes to phylogeography: a Bayesian structured coalescent approximation. PLoS Genet. 11(8):e1005421. Doi: https://doi.org/10.1371/journal.pgen.1005421. (eCollection 2015 Aug).

De Silva, E., Ferguson, N.M., Fraser, C., 2012. Inferring pandemic growth rates from sequence data. J. R. Soc. Interface 9 (73), 1797–1808. https://doi.org/10.1098/rsif.2011.0850.

Di Bisceglie, A.M., 1997. Hepatitis C and hepatocellular carcinoma. Hepatology 26, 34S–38S (Review).

Di Lello, F., Garcia, G., Kott, V., Sookoian, S., Campos, R., 2008. Diversity of hepatitis C virus genotype 1b in Buenos Aires, Argentina: description of a new cluster associated with response to treatment. J. Med. Virol. 80 (4), 619–627. https://doi.org/10.1002/jmv.21112.

Djebbi, A., Triki, H., Bahri, O., Cheikh, I., Sadraoui, A., Ben Ammar, A., Dellagi, K., 2003. Genotypes of hepatitis C virus circulating in Tunisia. Epidemiol. Infect. 130 (3), 501–505.

Drucker, E., Alcabes, P.G., Marx, P.A., 2001. The injection century: massive unsterile injections and the emergence of human pathogens. Lancet 358 (9297), 1989–1992 (Review).

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22 (5), 1185–1192.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973. https://doi.org/10.1093/molbev/mss075.

EASL, 2018. Recommendations on treatment of Hepatitis C. J. Hepatol. 69 (2), 461–511. https://doi.org/10.1016/j.jhep.2018.03.026.

El-Serag, H.B., 2002. Hepatocellular carcinoma: an epidemiologic view. J. Clin. Gastroenterol. 35, S72–S78 (Review).

Ezzikouri, S., Pineau, P., Benjelloun, S., 2013. Hepatitis C virus infection in the Maghreb region. J. Med. Virol. 85 (9), 1542–1549. https://doi.org/10.1002/jmv.23643.

Ferraro, D., Genovese, D., Argentini, C., Giordano, V., Pizzillo, P., Stroffolini, T., Craxì, A., Rapicetta, M., Di Stefano, R., 2008. Phylogenetic reconstruction of HCV genotype 1b dissemination in a small city Centre: the Camporeale model. J. Med. Virol. 80 (10), 1723–1731. https://doi.org/10.1002/jmv.21276.

Gelman, A., Meng, X.L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Stat. Sci. 13, 163–185.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

Iles, J.C., Raghwani, J., Harrison, G.L.A., Pepin, J., Djoko, C.F., Tamoufe, U., LeBreton, M., Schneider, B.S., Fair, J.N., Tshala, F.M., Kayembe, P.K., Muyembe, J.J., Edidi-Basepeo, S., Wolfe, N.D., Simmonds, P., Klenerman, P., Pybus, O.G., 2014. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology 464-465, 233–243. https://doi.org/10.1016/j.virol.2014.07.006.

Kass, R.E., Raftery, A.E., 1995. Bayes factor. J. Am. Stat. Assoc. 90, 773–795.

Kendrick, D., 1964. Blood program in world war II. In: Supplemented by experiences in the Korean war. Office of the Surgeon General Department of the Army, Washington, DC Boyd Coates J, MC, ed.

Khan, A., Al Balwi, M., AlAyyar, L., AlAbdulkareem, I., Albekairy, A., Aljumah, A., 2017. Tracing the epidemic history of hepatitis C virus genotypes in Saudi Arabia. Infect. Genet. Evol. 52, 82–88. https://doi.org/10.1016/j.meegid.2017.04.024.

Kuiken, C., Simmonds, P., 2009. Nomenclature and numbering of the hepatitis c virus. Methods Mol. Biol. 510, 33–53.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874. https://doi.org/10.1093/molbev/msw054.

Lampe, E., Espirito-Santo, M.P., Martins, R.M., Bello, G., 2010. Epidemic history of hepatitis C virus in Brazil. Infect. Genet. Evol. 10 (7), 886–895. https://doi.org/10.1016/j.meegid.2010.05.010.

Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34 (9), 2422–2424. https://doi.org/10.1093/molbev/msx149. 1.

Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5 (9), e1000520. https://doi.org/10.1371/journal.pcbi.1000520.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44 (W1), W242–W245. https://doi.org/10.1093/nar/gkw290.

Li, Z., Chen, Z.W., Li, H., Ren, H., Hu, P., 2017. Prevalence of hepatitis C virus-resistant association substitutions to direct-acting antiviral agents in treatment-naïve hepatitis C genotype 1b-infected patients in western China. Infect. Drug Res. 10, 377–392. https://doi.org/10.2147/IDR.S146595. 31.

Lontok, E., Harrington, P., Howe, A., Kieffer, T., Lennerstrand, J., Lenz, O., McPhee, F., Mo, H., Parkin, N., Pilot-Matias, T., Miller, V., 2015. Hepatitis C virus drug resistance-associated substitutions: state of the art summary. Hepatology 62 (5), 1623–1632. https://doi.org/10.1002/hep.27934.

Lu, L., Nakano, T., He, Y., Fu, Y., Hagedorn, C.H., Robertson, B.H., 2005. Hepatitis C virus genotype distribution in China: predominance of closely related subtype 1b isolates and existence of new genotype 6 variants. J. Med. Virol. 75 (4), 538–549.

Magiorkinis, G., Magiorkinis, E., Paraskevis, D., Ho, S.Y., Shapiro, B., Pybus, O.G., Allain, J.P., Hatzakis, A., 2009. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Med. 6 (12), e1000198. https://doi.org/10.1371/journal.pmed.1000198.

Mejri, S., Salah, A.B., Triki, H., Alaya, N.B., Djebbi, A., Dellagi, K., 2005. Contrasting patterns of hepatitis C virus infection in two regions from Tunisia. J. Med. Virol. 76 (2), 185–193.

Messina, J.P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G.S., Pybus, O.G., Barnes, E., 2015. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology 61 (1), 77–87. https://doi.org/10.1002/hep.27259.

Nakano, T., Lu, L., He, Y., Fu, Y., Robertson, B.H., Pybus, O.G., 2006. Population genetic history of hepatitis C virus 1b infection in China. J. Gen. Virol. 87 (Pt 1), 73–82.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Ogata, Y., 1989. A Monte Carlo method for high dimensional integration. Numer. Math. 55, 137–157.

Othman, S.B., Trabelsi, A., Monnet, A., Bouzgarrou, N., Grattard, F., Beyou, A., Bourlet, T., Pozzetto, B., 2004. Evaluation of a prototype HCV NS5b assay for typing strains of hepatitis C virus isolated from Tunisian haemodialysis patients. J. Virol. Methods 119 (2), 177–181.

Paraskevis, D., Pybus, O., Magiorkinis, G., Hatzakis, A., Wensing, A.M., van de Vijver, D.A., Albert, J., Angarano, G., Asjö, B., Balotta, C., Boeri, E., Camacho, R., Chaix, M.L., Coughlan, S., Costagliola, D., De Luca, A., de Mendoza, C., Derdelinckx, I., Grossman, Z., Hamouda, O., Hoepelman, I., Horban, A., Korn, K., Kücherer, C., Leitner, T., Loveday, C., Macrae, E., Maljkovic-Berry, I., Meyer, L., Nielsen, C., Op de Coul, E.L., Ormaasen, V., Perrin, L., Puchhammer-Stöckl, E., Ruiz, L., Salminen, M.O., Schmit, J.C., Schuurman, R., Soriano, V., Stanczak, J., Stanojevic, M., Struck, D., Van Laethem, K., Violin, M., Yerly, S., Zazzi, M., Boucher, C.A., Vandamme, A.M., 2009. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology 6, 49. https://doi.org/10.1186/1742-4690-6-49.

Pawlotsky, J.M., Tsakiris, L., Roudot-Thoraval, F., Pellet, C., Stuyver, L., Duval, J., Dhumeaux, D., 1995. Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. J. Infect. Dis. 171 (6), 1607–1610.

Pol, S., Thiers, V., Nousbaum, J.B., Legendre, C., Berthelot, P., Kreis, H., Brechot, C., 1995. The changing relative prevalence of hepatitis C virus genotypes: evidence in hemodialyzed patients and kidney recipients. Gastroenterology 108 (2), 581–583.

Pybus, O.G., Charleston, M.A., Gupta, S., Rambaut, A., Holmes, E.C., Harvey, P.H., 2001. The epidemic behavior of the hepatitis C virus. Science 292 (5525), 2323–2325 22.

Pybus, O.G., Markov, P.V., Wu, A., Tatem, A.J., 2007. Investigating the endemic transmission of the hepatitis C virus. Int. J. Parasitol. 37 (8–9), 839–849.

Pybus, O.G., Barnes, E., Taggart, R., Lemey, P., Markov, P.V., Rasachak, B., Syhavong, B., Phetsouvanah, R., Sheridan, I., Humphreys, I.S., Lu, L., Newton, P.N., Klenerman, P., 2009. Genetic history of hepatitis C virus in East Asia. J. Virol. 83 (2), 1071–1082. https://doi.org/10.1128/JVI.01501-08.

Rajhi, M., Mejri, S., Djebbi, A., Chouaieb, S., Cheikh, I., Ben Yahia, A., Sadraoui, A., Hammami, W., Azouz, M., Ben Mami, N., Triki, H., 2014. Subtyping genotype 2 hepatitis C viruses from Tunisia: identification of two putative new subtypes. Virus Genes 48 (2), 209–217. https://doi.org/10.1007/s11262-013-1010-3.

Rajhi, M., Ghedira, K., Chouikha, A., Djebbi, A., Cheikh, I., Ben Yahia, A., Sadraoui, A., Hammami, W., Azouz, M., Ben Mami, N., Triki, H., 2016. Phylogenetic analysis and epidemic history of hepatitis C virus genotype 2 in Tunisia, North Africa. PLoS ONE 11 (4), e0153761. https://doi.org/10.1371/journal.pone.0153761.

Silini, E., Bono, F., Cividini, A., Cerino, A., Maccabruni, A., Tinelli, C., Bruno, S., Bellobuono, A., Mondelli, M., 1995. Molecular epidemiology of hepatitis C virus infection among intravenous drug users. J. Hepatol. 22 (6), 691–695.

Simmonds, P., 2004. Genetic diversity and evolution of hepatitis C virus–15 years on. J. Gen. Virol. 85 (Pt 11), 3173–3188 (Review).

Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., Feinstone, S., Halfon, P., Inchaupse, G., Kuiken, C., Maertens, G., Mizokami, M., Murphy, D., Okamoto, H., Pawlotsky, J.M., Penin, F., Salbon, E., Shin-I, T., Stuver, L., Thiel, H.J., Viazov, S., Weiner, A., Widell, A., 2005. Consensus proposals for a unified system of nomenclature of hepatitis c virus genotype. Hepatology 42, 962–973.

Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L., Simmonds, P., 1997. The origin of hepatitis C virus genotypes. J. Gen. Virol. 78 (Pt 2), 321–328.

Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T., Simmonds, P., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource, Hepatology 59(1):318-327. Doi: https://doi.org/10.1002/hep.26744. (European Association for Study of liver).

Struck, D., Lawyer, G., Ternes, A.M., Schmit, J.C., Bercoff, D.P., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42 (18), e144. https://doi.org/10.1093/nar/gku739. (Epub 2014 Aug 12).

Tanaka, Y., Hanada, K., Orito, E., Akahane, Y., Chayama, K., Yoshizawa, H., Sata, M., Ohta, N., Miyakawa, Y., Gojobori, T., Mizokami, M., 2005. Molecular evolutionary analysis implicate injection treatment for schistosomiasis in the initial hepatitis C epidemics in Japan. J. Hepatol. 42 (1), 47–53.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L.C., Vanden Eynden, E., Vandamme, A.M., Deforche, K., De Oliveira, T., 2019. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 35 (5), 871–873. https://doi.org/10.1093/bioinformatics/bty695. 1.

Vrancken, B., Alavian, S.M., Aminy, A., Amini-Bavil-Olyaee, S., Pourkarim, M.R., 2018. Why Comprehensive Datasets Matter when Inferring Epidemic Links or Subgenotyping.

Wu, T., Xiong, L., Wang, F., Xu, X., Wang, J., Lin, F., Li, C., Lu, L., Zhou, Y., 2016. A unique pattern of HCV genotype distribution on Hainan Island in China revealed by evolutionary analysis. Cell. Physiol. Biochem. 39 (1), 316–330. https://doi.org/10.1159/000445626.

Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.H., 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60 (2), 150–160. https://doi.org/10.1093/sysbio/syq085.

Yerly, S., Quadri, R., Negro, F., Barbe, K.P., Cheseaux, J.J., Burgisser, P., Siegrist, C.A., Perrin, L., 2001. Nosocomial outbreak of multiple bloodborne viral infections. J. Infect. Dis. 184 (3), 369–372.