

RESEARCH ARTICLE

# Phylogenetic Analysis and Epidemic History of Hepatitis C Virus Genotype 2 in Tunisia, North Africa

Mouna Rajhi<sup>1,7\*</sup>, Kais Ghedira<sup>2,6</sup>, Anissa Chouikha<sup>1</sup>, Ahlem Djebbi<sup>1</sup>, Imed Cheikh<sup>4</sup>, Ahlem Ben Yahia<sup>1</sup>, Amel Sadraoui<sup>1</sup>, Walid Hammami<sup>1</sup>, Msaddek Azouz<sup>5</sup>, Nabil Ben Mami<sup>3</sup>, Henda Triki<sup>1,6</sup>

**1** Pasteur Institute, Tunis, Tunisia; Laboratory of Clinical Virology, WHO Regional Reference Laboratory on Poliomyelitis and Measles, Tunis, Tunisia, **2** Pasteur Institute, Tunis, Tunisia; Laboratory of Bioinformatics, Mathematics and Statistics, Tunis, Tunisia, **3** Department of Gastroenterology, La Rabta Hospital, Tunis, Tunisia, **4** Department of Gastroenterology, Regional Hospital of Bizerte, Bizerte, Tunisia, **5** Department of Gastroenterology, Regional Hospital of Nabeul, Nabeul, Tunisia, **6** University of Tunis El Manar, Tunis, 1036, Tunisia, **7** University of Carthage, Faculty of Sciences, Bizerte, Tunisia

\* [rajhimouna@yahoo.fr](mailto:rajhimouna@yahoo.fr)



**OPEN ACCESS**

**Citation:** Rajhi M, Ghedira K, Chouikha A, Djebbi A, Cheikh I, Ben Yahia A, et al. (2016) Phylogenetic Analysis and Epidemic History of Hepatitis C Virus Genotype 2 in Tunisia, North Africa. PLoS ONE 11 (4): e0153761. doi:10.1371/journal.pone.0153761

**Editor:** Jason Blackard, University of Cincinnati College of Medicine, UNITED STATES

**Received:** December 22, 2015

**Accepted:** April 4, 2016

**Published:** April 21, 2016

**Copyright:** © 2016 Rajhi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** This study was totally supported by the Tunisian Ministry for High Education and Scientific Research (Grant LR11-IPT09).

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

HCV genotype 2 (HCV-2) has a worldwide distribution with prevalence rates that vary from country to country. High genetic diversity and long-term endemicity were suggested in West African countries. A global dispersal of HCV-2 would have occurred during the 20<sup>th</sup> century, especially in European countries. In Tunisia, genotype 2 was the second prevalent genotype after genotype 1 and most isolates belong to subtypes 2c and 2k. In this study, phylogenetic analyses based on the NS5B genomic sequences of 113 Tunisian HCV isolates from subtypes 2c and 2k were carried out. A Bayesian coalescent-based framework was used to estimate the origin and the spread of these subtypes circulating in Tunisia. Phylogenetic analyses of HCV-2c sequences suggest the absence of country-specific or time-specific variants. In contrast, the phylogenetic grouping of HCV-2k sequences shows the existence of two major genetic clusters that may represent two distinct circulating variants. Coalescent analysis indicated a most recent common ancestor (tMRCA) of Tunisian HCV-2c around 1886 (1869–1902) before the introduction of HCV-2k in 1901 (1867–1931). Our findings suggest that the introduction of HCV-2c in Tunisia is possibly a result of population movements between Tunisia and European population following the French colonization.

## Introduction

Hepatitis C virus (HCV) is one of the most important causes of chronic liver disease worldwide. Since its discovery in 1989 [1], several isolates from various regions of the world were sequenced and a high degree of genomic heterogeneity was observed. Thus, HCV strains are currently classified into seven genotypes (1 to 7) and several subtypes [2]. Genotype identification is clinically important to predict response to antiviral therapy, as some genotypes are more

resistant to treatment than others; it is now routinely performed before treatment and serves as a guideline to determine the duration of the therapy [3,4]. Genotyping is also useful for molecular epidemiology studies. The geographical distribution of HCV genotypes and subtypes varies significantly from a region to another in the world [5]. Genotypes 1, 2, and 3 have a worldwide distribution, while genotypes 4, 5 and 6 are mainly encountered in Egypt, Middle East and Central Africa, Southern Africa and Asia, respectively [6–9].

HCV genotype 2 (HCV-2) is worldwide distributed but is reported to be more frequent in West African countries and Martinique [10–13]. It is one of the HCV genotypes showing the widest genomic diversity. Twenty-one subtypes (designated a to u) are presently assigned in the HCV sequence database [14]. Subtypes 2a, 2b and 2c are ubiquitous while others subtypes are mainly found in specific regions [15].

In Tunisia, Genotype 1 is the most prevalent, representing 70 to 80% and Genotype 2 comes in the second position with approximately 10% of circulating isolates [16–18]. In a recently published study, we showed that most of HCV-2 isolates belong to subtype 2c with few co-circulating isolates from other subtypes, especially 2k [19]. The aim of the present study was to analyse the genetic diversity and to reconstruct the genetic history of the most prevalent HCV-2 subtypes. The NS5B sequences of 113 isolates from subtypes 2c and 2k, collected over a 10-years period in Tunisia, were analyzed by phylogeny and a Bayesian coalescent investigation was carried out to estimate the origin and the evolution of these two subtypes.

## Material and Methods

The samples used in the present study were obtained as part of the routine diagnostic activity of the Laboratory of Clinical Virology in Pasteur institute of Tunis and this did not need an ethical approval. HCV Viral load and genotyping were performed in response to the request of the physicians. Patient records/information was anonymized and de-identified prior to inclusion in the study.

## Study population

The study included serum samples from 113 Tunisian patients who were detected HCV RNA-positive and the isolated viruses belonged to HCV subtype 2c (N = 102) or HCV subtype 2k (N = 11). The serum samples were collected over a 10-years period, from 2003 to 2012.

During the study period, 1430 serum samples were processed for genotyping, 1171 (81%), 135(9%), 63(4%) and 29(2%) belonged to genotypes 1, 2, 3 and 4, respectively. Subtyping of the HCV-2 isolates revealed that most of them belonged to subtype 2c (75%), followed by Subtype 2k (8%). All the isolates belonging to these two subtypes were included in the present study. The study population included 86 women and 27 men. All of them were Tunisians, they originated from all regions of Tunisia and were aged 24 to 70 years with a mean age of 51 years.

## PCR amplification and sequencing

RNA extraction was performed from 140µl of serum using the QIAmp viral RNA mini Kit (QIAGEN, Hilden-Germany) and the extracted RNA was then processed for reverse transcription (RT) and PCR amplification. The RT reaction mixture included 10µl of viral RNA, 10pmol of random hexamers, 1µl of 40U RNasin (Promega), 4µl of 5X transcriptase buffer, 2µl of 0,1M DTT, 1µl of dNTP mix (20mM) and 200U of the super script II Reverse Transcriptase (Invitrogen); in a final volume of 20µl. The RT reaction mixture was incubated at 42°C for 45min then heat/inactivated at 94°C for 5min. A 364-nucleotide fragment in the NS5B region (positions 8267 to 8630) was amplified by nested PCR using two pairs of primers specific to relatively conserved sequences in the NS5B genomic region: **Outer primers:** 1s (Forward (8256–

8275): 5'-TATGAYACCCGCTGYTTTGAC-3'), 2as (Reverse (8622–8641): 5'-GARTACCRTGT CATAGCCTC-3') and **inner primers** 3as (Forward (8267–8287): 5'-CTGYTTTGACTCMAC RGTAC-3'), 4as (Reverse (8611–8630): 5'-ATAGCCTCCGTGAAGRCTC-3'). The first round of PCR was performed with 10 $\mu$ l cDNA in a total reaction mixture of 50 $\mu$ l containing 5 $\mu$ l 10X polymerase buffer, 3 $\mu$ l MgCl<sub>2</sub>, 5U recombinant-Taq polymerase and 10pmol each forward and reverse outer primers. The same conditions were used in the second round of PCR using 5 $\mu$ l of first round PCR product and 10pmol of each inner primer. The cycling profile was identical for the two rounds and consisted in a pre-denaturation at 95°C for 5 min followed by 30 cycles at 94°C for 30s, 60°C for 30s and 72°C for 1min. A final extension was performed at 72°C for 7min. The PCR products were separated and visualized on a 1% agarose gel. The PCR products were excised from the gel and purified with the QIAquick Gel Extraction kit (QIAGEN). The purified templates DNA were sequenced in both directions using the Big-Dye terminator ready reaction cycle sequencing Kit (Applied Biosystems) and the same inner primers on an ABI Prism 3130 automated sequencer (Applied Biosystems).

## Phylogenetic analyses

Ancestral relationships among HCV isolates were inferred by reconstructing phylogenetic trees of Tunisian NS5B sequences, together with the HCV-2c and HCV-2k sequences available in GenBank database.

Phylogenetic analyses were performed using the Maximum composite likelihood methods included in the MEGA package (Version 5.1). Although the Tunisian sequences were more than 300-nucleotide-long, the phylogenetic analyses were performed on a fragment of 222 nucleotides of the NS5B region in the aim to include the maximum of sequences from other countries. The reliability of the phylogenetic constructions was estimated by bootstrap analysis with 1000 pseudo replicate data sets. Sixty-eight out of the 113 Tunisian sequences included in the present study were previously published in GenBank [19] whereas 45 sequences are new and reported for the first time in this study; they were submitted to GenBank under the accession numbers KJ206290 to KJ206333 and KJ206335. In addition, ninety-two HCV field sequences, published in GenBank and isolated from different countries in the world were included in these phylogenetic analyses: 53 isolates from Subtype 2c and 39 isolates from subtype 2k. The selected sequences represented the widest temporal and geographical range including 18 different countries and an isolation year ranging from 1993 to 2013.

## Calculation of the genetic distances

Estimates of evolutionary divergence between the sequences were conducted using the maximum likelihood method included in the MEGA 5.1 software.

## Coalescent analyses

The epidemic history of HCV subtypes 2c and 2k was investigated with a coalescent-based strategy. Briefly, HCV 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>) [20]. For this analysis, five population dynamics models were used: Bayesian, constant, exponential, expansion and logistic under the strict and relaxed clocks. In all cases, the Hasegawa-Kishino-Yano (HKY) nucleotide-substitution model [21] was used under the strict then relaxed molecular clock. All BEAST chains were run for 3x10<sup>8</sup> generations for HCV-2c and for 1.5x10<sup>8</sup> generations for HCV-2k in order to achieve an Effective Sample Size (ESS)>200 for the 2 subtypes. Since the time interval covered by the sequences of HCV-2c and HCV-2k for the NS5B region with known sampling date is about 10 years, the substitution rate parameter

could not be reliably estimated, and thus an external substitution rate was used. A substitution rate of  $5 \times 10^{-4}$  substitutions per site per year (s/s/y) was used according to previous estimation [22] for the NS5B region. Best fitting models were selected by calculation of a Bayes Factor (BF) [23], using Marginal Likelihoods Estimation (MLE) specifically path sampling (PS) [24,25] and stepping-stone (SS) [26] sampling implemented in Beast 1.8.1.

The effective number of infections through time was estimated by using the Bayesian skyline plot (BSP) approach [27] 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 two different datasets of NS5B sequences: subtype 2c (n = 102, 307nt, 8302 to 8608), subtype 2k (n = 11, 307nt, 8302 to 8608).

## Results

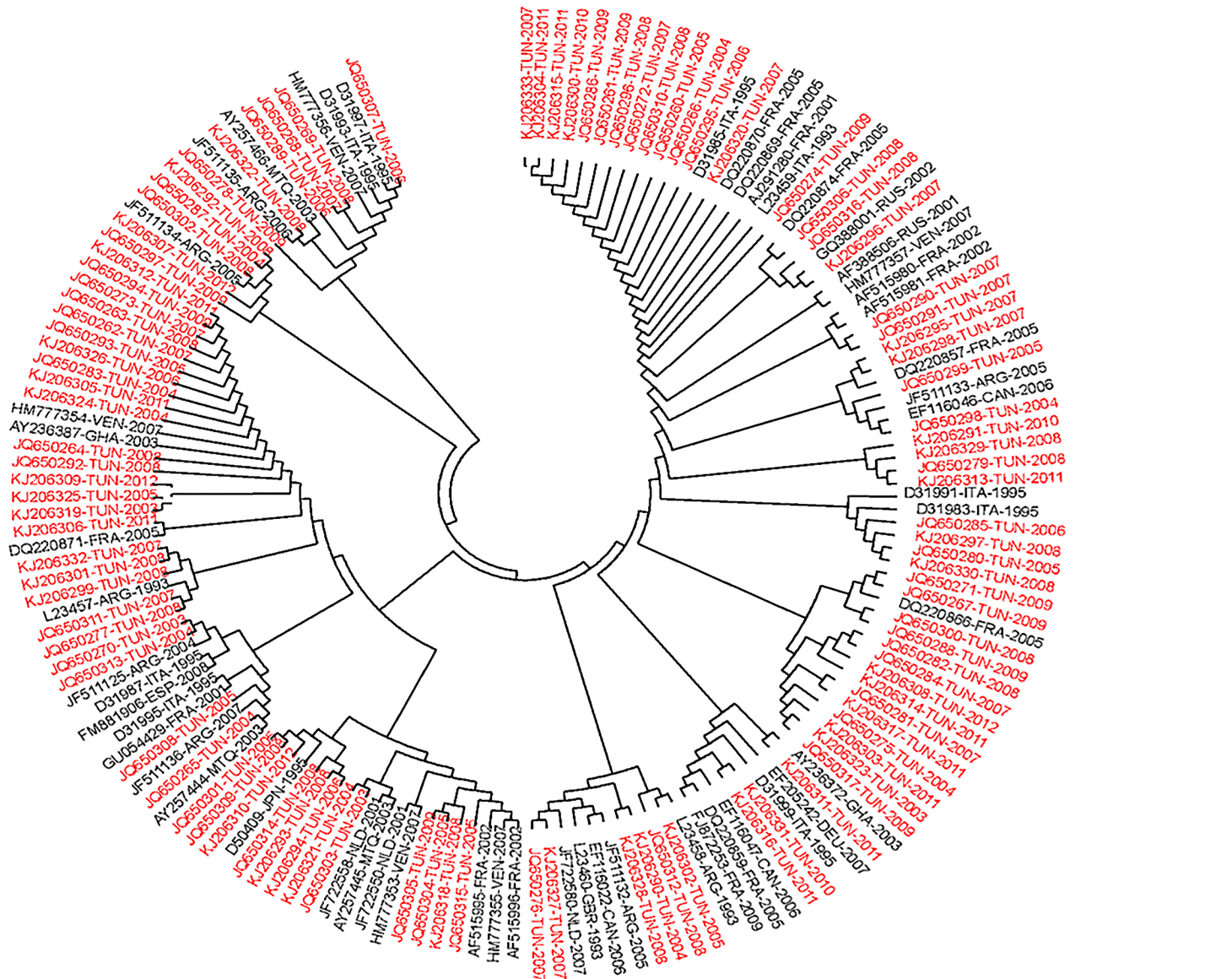
### Phylogenetic analyses

**Subtype 2c.** The phylogenetic tree in Fig 1 contains 102 Tunisian sequences that overlapped a fragment of 222 nucleotides in the NS5B genomic region; the corresponding virus isolates were already classified within subtype 2c. The tree also includes 53 field sequences published in GenBank and originating from 13 different countries: France (N = 14), Martinique (N = 3), Canada (N = 3), Italy (N = 9), Venezuela (N = 5), Ghana (N = 2), Russia (N = 2), Netherlands (N = 3), Spain (N = 1), Japan (N = 1), Germany (N = 1), Argentina (N = 8) and United Kingdom (N = 1). No consistent cluster supported by a high bootstrap value could be identified and the Tunisian sequences were randomly distributed throughout the phylogenetic tree, among all subtype2c sequences that formed a monophyletic group. The divergence rates between all the 2c sequences, in the studied region, varied from 0% to 10.7% with an overall mean distance of 2.6%.

**Subtype 2k.** The phylogenetic tree in Fig 2 includes 11 Tunisian sequences previously classified within subtype 2k and 39 field sequences of the same subtype reported from 10 different countries: France (N = 20), Canada(N = 4), Russia(N = 1), Martinique(N = 5), Moldova (N = 1), Uzbekistan(N = 1), Azerbaijan(N = 1), Morocco(N = 3), United kingdom(N = 1) and Iran(N = 2). All the sequences segregated into two clusters: Cluster1 comprised most of French sequences (18 out of 20), four out of the five sequences from Martinique and one out of the three sequences from Morocco. Cluster2 had wider geographic distribution and contained all the Tunisian sequences together with sequences from all the ten other countries. The divergence rates within subtype 2k, in the studied NS5B region, ranged from 0.6 to 13.6% with an overall mean distance of 6.9%.

### Epidemic history of HCV-2c and HCV-2k circulating in Tunisia

**Bayesian Skyline Plots.** Bayesian coalescence analysis was performed on all the NS5B sequences from Tunisia, separately for the subtype 2c dataset (102 sequences) and the subtype 2k dataset (11 sequences). For each demographic and molecular clock models, chain lengths of 300 million for HCV-2c and of 150 million for HCV-2k were used and sampled every 30000 and 15000 states, respectively. For HCV-2c, the best-fitted model was the exponential growth with relaxed molecular clock. For HCV- 2k, the best-fitted model was the logistic growth with strict molecular clock. The data was analysed using the Bayesian Skyline Plot (BSP) that depicts the estimated change in the effective number of infected individuals over time for each subtype. -2c and HCV-2k (Fig 3a and 3b) exhibited a similar epidemic history, characterized by three phases of epidemic population growth including: an initial period of relatively constant population size, followed by a well-defined phase of exponential growth ranging from 1912 to 1970

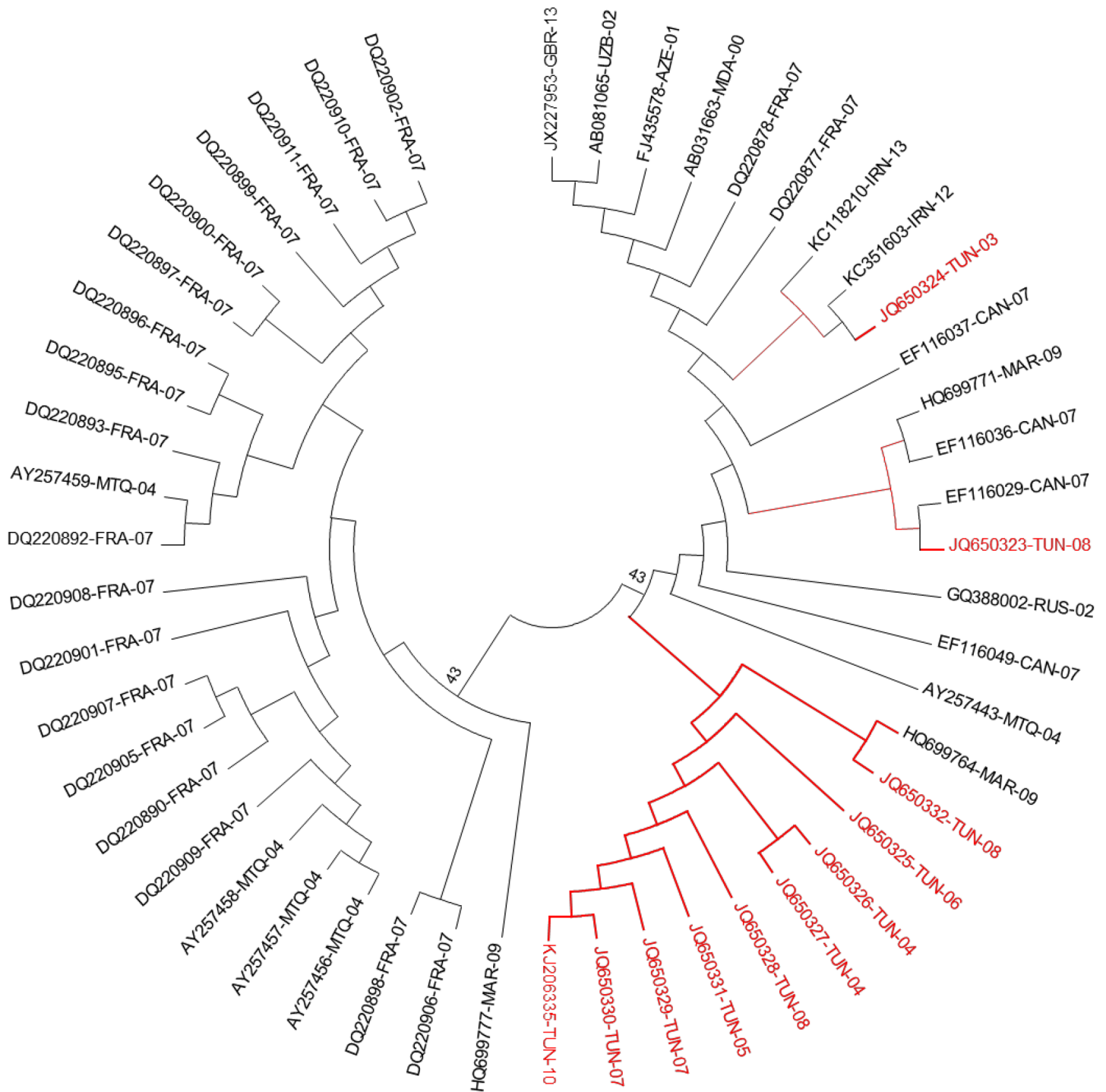


**Fig 1. Phylogenetic analysis of isolates from subtype 2c.** Phylogenetic analyses were performed using the Maximum Composite Likelihood method included in the MEGA package (version 5.1). The reliability of the phylogenetic constructions was estimated by bootstrap analysis with 1000 pseudo replicate data sets. The tree is based on the analysis of a 222-nucleotide-long fragment in the NS5B region (nucleotides 8316 to 8537 according to the H77-1a prototype strain, AF009606). It includes the 102 studied Tunisian sequences (in red) and 53 HCV-2c field sequences from other countries (in black). All the sequences were indicated by their Accession Numbers followed by the country code and the collection date. The country codes according to the standard abbreviation are: France (FRA), Martinique (MTQ), Canada (CAN), Italy (ITA), Venezuela (VEN), Ghana (GHA), Russia (RUS), Netherlands (NLD), Spain (ESP), Japan (JPN), Germany (DEU), Argentina (ARG) and United Kingdom (GBR).

doi:10.1371/journal.pone.0153761.g001

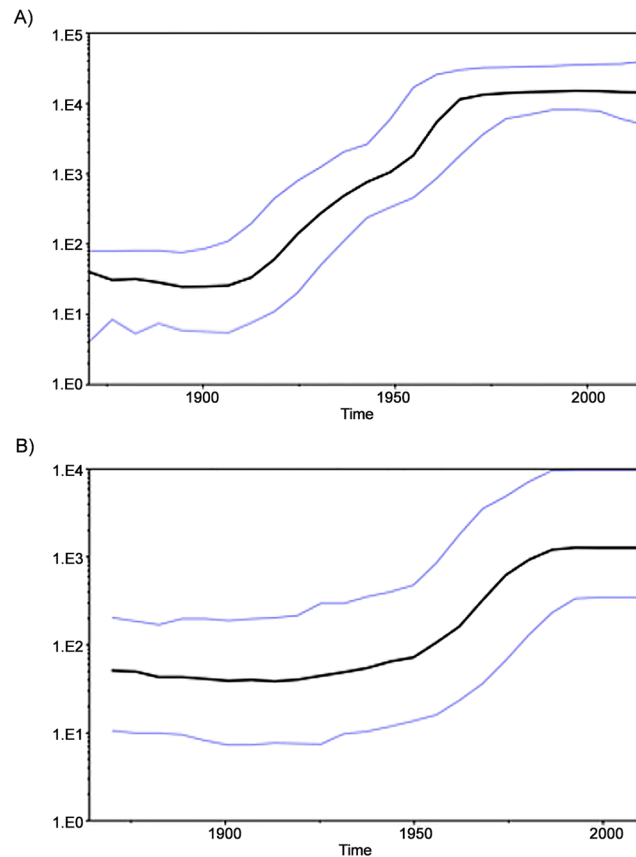
for HCV-2c and from 1940 until 1985 for HCV-2k; the final stabilization phase starts around 1970 for subtype 2c and around 1985 for subtype 2k.

**Most Recent Common Ancestor (tMRCA).** The most recent common ancestor (tMRCA) was evaluated by using a fixed substitution rate for the NS5B region of  $5.10^{-4}$  substitutions per site and per year according to a previous study [22]. The tMRCA of subtype 2c was evaluated around 1886 (1869–1902), prior to the introduction of subtype 2k estimated in 1901 (1867–1931) (Table 1).



**Fig 2. Phylogenetic analysis of isolates from Subtype 2k.** Phylogenetic analyses were performed using the Maximum Composite Likelihood method included in the MEGA package (version 5.1). The reliability of the phylogenetic constructions was estimated by bootstrap analysis with 1000 pseudo replicate data sets. The tree is based on the analysis of a 222-nucleotide long fragment in the NS5B region (nucleotides 8346 to 8568 according to the H77-1a prototype strain, AF009606). It includes the 11 studied Tunisian sequences (in red) and 39 HCV-2k field sequences from other countries (in black). All the sequences were indicated by their Accession Numbers followed by the country code and the collection date. The country codes according to the standard abbreviation are: France (FRA), Martinique (MTQ), Canada (CAN), Russia (RUS), United Kingdom (GBR), Uzbekistan (UZB), Azerbaijan (AZE), Morocco (MAR), Iran (IRN) and Moldova (MDA).

doi:10.1371/journal.pone.0153761.g002



**Fig 3. Bayesian Skyline Plots for Demographic Reconstruction using NS5B sequences.** A) subtype 2c, B) subtype 2k; X-axis: Date in Years; Y-axis: Estimated effective number of infections; Bold Line: Mean Effective Number of viral population. Upper and lower Lines: Upper and Lower HPD95% (High Population Density) of Effective Number of viral population.

doi:10.1371/journal.pone.0153761.g003

## Discussion

HCV-2 has a worldwide distribution with prevalence rates that vary from country to country. Previous studies suggested that it originated from Africa where it would have endemically persisted for 500 to 600 years [28–31]. A global dispersal of HCV-2 would have occurred during the 20<sup>th</sup> century and in some countries such as Venezuela, a reduction in the circulation of HCV-1 was observed in the last decade with an increase of circulation of genotype 2 [32]. Recently, the genetic diversity of HCV-2 and its epidemic history have been assessed in several countries, mainly from Europe and Latin America [33–36] but such investigations remain very scarce for HCV-2 isolates circulating in North African countries.

**Table 1. tMRCA estimations using the NS5B region for the HCV- 2c and HCV-2k in Tunisia.**

HCV subtype	Number of isolates	Substitution rate	MRCA (HPD <sup>1</sup> 95%) (year)
2c	102	5.10 <sup>-4</sup> s/site/year	1886 (1869–1902)
2k	11	5.10 <sup>-4</sup> s/site/year	1901 (1867–1931)

(<sup>1</sup>: HPD: Highest Posterior Density).

doi:10.1371/journal.pone.0153761.t001

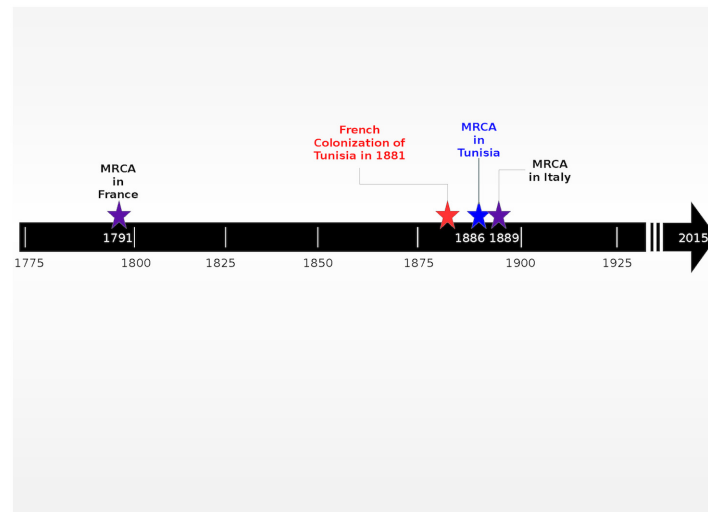
Among the different HCV-2 subtypes, HCV-2c is reported to be ubiquitous worldwide. It was identified in several countries, contrary to the other subtypes, which seem to have more restricted geographical distribution [15]. The frequency of HCV-2c varies from one country to another. Although not very common in Brazil, Canada, Chile, Mexico, or the USA [37,38], HCV-2c is the second most frequently isolated after genotype 1 in Argentina [39], Italy [40] and in Tunisia [19]. In the present work, we analyzed the genetic diversity and the phylogenetic relatedness between the NS5B sequences within HCV-2c subtype. These analyses were performed on 102 Tunisian sequences and 53 representative sequences from 13 other countries. The genetic distance found between all HCV-2c sequences did not exceed 10.7%, with an overall mean distance of 2.6%. Several Tunisian strains were identical to each other and to other sequences from France and Italy. Previous studies reported low genetic diversity of HCV-2c and suggested that it is an epidemic subtype [22,41]. Furthermore, all HCV-2c sequences formed a monophyletic group; the Tunisian sequences were intermixed with the other sequences suggesting the absence of country-specific or time-specific variants. Similar results were also reported in previous studies and the authors suggested that the low genetic diversity within HCV-2c sequences from various geographical origins would be a reflection of its quick spread during the 20<sup>th</sup> century [22,35,41].

In the other hand and contrary to HCV-2c, the phylogenetic analysis of HCV-2k sequences showed different features. The genetic diversity within the subtype was higher and the phylogenetic analysis of the 11 sequences from Tunisia and 39 sequences representative of 10 other countries showed the existence of two major genetic clusters that may represent two distinct circulating variants. All Tunisian isolates belonged to the same variant. Subtype 2k was reported from several countries; France, Canada, Russia, Martinique, Moldova, Uzbekistan, Vietnam, Madagascar, Azerbaijan, Morocco, United Kingdom and Iran. In southwestern France, subtype 2k is the second most frequent subtype of circulating HCV-2 isolates with a prevalence of 22% [34]. However, it remains much less investigated compared to HCV-2c and, to our knowledge, beyond prevalence rates, its molecular epidemiology still insufficiently documented.

Recent works highlighted the role of human migration process in the global and rapid dissemination of HCV genotypes [35,42,43]. Some studies suggested the introduction of HCV-2 in western countries from the African continent probably in the 17<sup>th</sup>-18<sup>th</sup> centuries through the slave trade and colonialism [28,33,36, 44]. Studies conducted on the genetic history of HCV genotype 2 mainly concerned subtype 2c and proposed the following dates for the tMRCA: 1791 in France [42], 1871 in Argentina [35], 1889 in Italy [43]. The tMRCA of subtype 2j in Venezuela was evaluated around 1785 [33]. The epidemic history of HCV is rarely studied in North Africa. To our knowledge, the only report in this topic is the one of Brahim et al. [45] who established the genetic history of HCV-1b and HCV-2i in Morocco. In the present work, the epidemic history of HCV-2c and HCV-2k Tunisian strains was estimated with a coalescent approach. The results suggest that the subtype 2c would have been introduced in the country at the end of the 19<sup>th</sup> century, with an estimated tMRCA date in 1886 (1869–1902) and that HCV-2k would have been introduced in the beginning of the 20<sup>th</sup> century, around 1901 (1867–1931). Also, the Bayesian Skyline Plot shown in Fig 3 suggests that the 2c infected population increased exponentially between 1912 and 1970 while the 2k infected population experienced exponential growth later, between 1940 and 1985.

Looking on the history of Tunisia, the introduction of the HCV-2c may have been caused by the contact of the Tunisian population with European populations following the French colonization. Between the 19<sup>th</sup> and the early 20<sup>th</sup> century, the Italian community in Tunisia represented the largest group of European origin. At the beginning of the 19<sup>th</sup> century, it included several thousands of political activists and intellectuals from the central and northern regions





**Fig 4. The timeline of HCV subtype 2c introduction in Tunisia.**

doi:10.1371/journal.pone.0153761.g004

of Italy, sheltering in the Regency of Tunisia. At the end of the 19th century, the economic difficulties and the social crisis in the southern regions of the newly established Kingdom of Italy contributed to a continuous increase in the Italian population living in Tunisia with the arrival of migrants mainly from Sicily, Pantelleria, Sardinia and Procida [45]. Tunisia became a French protectorate in 1881 and the French population in Tunisia increased. However, the Italian community continued to grow and represented the main European community in the French protectorate with more than 100,000 Italians in the first years of the 20th century; Sicilians made up 72.5% of the community's population, while 16.3% were from central Italy, 3.8% from Sardinia and only 2.5% from northern Italy [46]. Thereafter, the colonization of Tunisia by French expatriates was followed by the return-back of Italians from Tunisia to Italy. Between 1943 and 1970, the Italian community almost disappeared from Tunisia.

Thus, if we consider the dates of the HCV-2c MRCA estimated in 1791 in France [42], in 1886 in Tunisia (the present work) and in 1889 in Italy [44], we may conclude that HCV-2c would have been introduced first in France and then in Tunisia and Italy, 90 to 100 years later, possibly as a result of population movements between Tunisia and European Countries after the French colonization (Fig 4). The massive return back of Italians from Tunisia to Italy during the 20<sup>th</sup> century may have contributed to the high prevalence of HCV-2c especially in southern Italy [40, 43].

## Conclusion

In conclusion, this study combined demographic, historical and viral genetic data to reconstruct the transmission history of HCV-2 in Tunisia. To our knowledge, it is the first to describe the evolutionary history of subtype 2c at the international level and is the first study on the diversity and genetic history of HCV-2c in a country from North Africa. Our results indicate that the epidemiology of HCV-2 in Tunisia was mainly influenced by population movements between countries of the Mediterranean basin especially those from the north bench of the Mediterranean Sea. Through a better understanding of past HCV-2 transmission events, this study helps to better explain the worldwide distribution of HCV-2 and to define risk factors and preventive measures for HCV-related diseases.

## Acknowledgments

This study was supported by the Tunisian Ministry for Scientific Research and Technology, (LR11-IPT09).

## Author Contributions

Conceived and designed the experiments: MR HT AC AD IC MA NBM. Performed the experiments: MR. Analyzed the data: MR KG. Contributed reagents/materials/analysis tools: ABY AS WH. Wrote the paper: MR KG HT. Performed the Bayesian coalescent analyses: KG.

## References

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*. 1989; 244: 359–62. PMID: [2523562](#)
2. Smith DB, Bukh J, Kuiken C, Scott Muerhoff A, Rice C M, Stapleton JT, et al. Expanded Classification of Hepatitis C Virus Into 7 Genotypes and 67 Subtypes: Updated Criteria and Genotype Assignment Web Resource. *Hepatology*. 2014; 59:318–27. doi: [10.1002/hep.26744](#) PMID: [24115039](#)
3. Ezequiel R. Predictors of Response to Chronic Hepatitis C Treatment. *Future Virology*. 2012; 7 (11):1089–1101.
4. Cavalcante LN, Lyra AC. Predictive factors associated with hepatitis C antiviral therapy response. *World J Hepatol*. 2015; 7(12):1617–31. doi: [10.4254/wjh.v7.i12.1617](#) PMID: [26140082](#)
5. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*. 2015; 61(1):77–87. doi: [10.1002/hep.27259](#) PMID: [25069599](#)
6. Daw MA, Dau AA. Hepatitis C virus in Arab world: a state of concern. *Scientific World Journal*. 2012; 2012:719494. doi: [10.1100/2012/719494](#) PMID: [22629189](#)
7. Khattab MA, Ferenci P, Hadziyannis SJ, Colombo M, Manns MP, Almasio PL, et al. Management of hepatitis C virus genotype 4: recommendations of an international expert panel. *J Hepatol*. 2011; 54 (6):1250–62. doi: [10.1016/j.jhep.2010.11.016](#) PMID: [21316497](#)
8. Wantuck JM, Ahmed A, Nguyen MH. Review article: the epidemiology and therapy of chronic hepatitis C genotypes 4, 5 and 6. *Aliment Pharmacol Ther*. 2014; 39(2):137–47. doi: [10.1111/apt.12551](#) PMID: [24251930](#)
9. Al Naamani K, Al Sinani S, Deschênes M. Epidemiology and treatment of hepatitis C genotypes 5 and 6. *Can J Gastroenterol*. 2013; 27(1):e8–12. PMID: [23378985](#)
10. Ruggieri A, Argentini C, Kouruma F, Chionne P, D'Ugo E, Spada E, et al. Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *J Gen Virol*. 1996; 77: 2073–6. PMID: [8811005](#)
11. Jeannel D, Fretz C, Traore Y, Kohdjo N, Bigot A, Pê Gamy E, et al. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol*. 1998; 55 (2): 92–7. PMID: [9598927](#)
12. Candotti D, Temple J, Sarkodie F, Allain JP. Frequent Recovery and Broad Genotype 2 Diversity Characterize Hepatitis C Virus Infection in Ghana, West Africa. *J Virol*. 2003; 77(14): 7914–23. PMID: [12829831](#)
13. Martial J, Morice Y, Abel S, Cabié A, Rat C, Lombard F, et al. Hepatitis C virus (HCV) genotypes in the Caribbean island of Martinique: Evidence for a large radiation of HCV-2 and for recent introduction from Europe of HCV-4. *J Clin Microbiol*. 2004; 42:784–91. PMID: [14766854](#)
14. International Committee on Taxonomy of Viruses. Available: [http://talk.ictvonline.org/ictv\\_wikis/w/sg\\_flavi/56.hcv-classification.aspx](http://talk.ictvonline.org/ictv_wikis/w/sg_flavi/56.hcv-classification.aspx).
15. Li C, Cao H, Lu L, Murphy D. Full-length sequences of 11 hepatitis C virus genotype 2 isolates representing five subtypes and six unclassified lineages with unique geographical distribution and genetic variation patterns. *J Gen Virol*. 2012; 93:1173–84. doi: [10.1099/vir.0.038315-0](#) PMID: [22357752](#)
16. Mejri S, Salah AB, Triki H, Alaya NB, Djebbi A, Dellagi K. Contrasting patterns of hepatitis C virus infection in two regions from Tunisia. *J Med Virol*. 2005; 76:185–93. PMID: [15834884](#)
17. Djebbi A, Triki H, Bahri O, Cheikh I, Sadraoui A, Ben Ammar A, et al. Genotypes of hepatitis C virus circulating in Tunisia. *Epidemiol Infect*. 2003; 130: 501–5. PMID: [12825736](#)

18. Ben Othman S, Trabelsi A, Monnet A, Bouzgarrou N, Grattard F, Beyou A, et al. Evaluation of a prototype HCV NS5b assay for typing strains of hepatitis C virus isolated from Tunisian haemodialysis patients. *J Virol Methods*. 2004; 119:177–81. PMID: [15158600](#)
19. Rajhi M, Mejri S, Djebbi A, Chouaieb S, Cheikh I, Ben Yahia A, et al. Subtyping genotype 2 hepatitis C viruses from Tunisia: identification of two putative new subtypes. *Virus Genes*. 2014; 48: 209–217. doi: [10.1007/s11262-013-1010-3](#) PMID: [24272697](#)
20. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7:214. PMID: [17996036](#)
21. Hasegawa M, Kishino H, Yano T. Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*. 1985; 22: 160–174. PMID: [3934395](#)
22. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behaviour of the hepatitis C virus. *Science*. 2001; 292: 2323–5. PMID: [11423661](#)
23. Kass RE, Raftery AE. Bayes Factor. *Journal of the American Statistical Association*. 1995; 90: 773–95.
24. Ogata Y. A Monte Carlo method for high dimensional integration. *Numer Math*. 1989; 55: 137–57.
25. Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat Sci*. 1998; 13:163–85.
26. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol*. 2011; 60:150–60. doi: [10.1093/sysbio/syq085](#) PMID: [21187451](#)
27. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005; 22: 1185–92. PMID: [15703244](#)
28. Markov PV, Pepin J, Frost E, Deslandes S, Labbe AC, Pybus OG. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol*. 90, 2086–2096 (2009) doi: [10.1099/vir.0.011569-0](#) PMID: [19474244](#)
29. Njouom R, Frost E, Deslandes S, Mamadou-Yaya F, Labbe AC, Pouillot R, et al. Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. *J Gen Virol*. 2009; 90: 2452–6. doi: [10.1099/vir.0.011981-0](#) PMID: [19535500](#)
30. Njouom R, Nerrienet E, Dubois M, Lachenal G, Rousset D, Vessiere A, et al. The hepatitis C virus epidemic in Cameroon: genetic evidence for rapid transmission between 1920 and 1960. *Infect Genet Evol*. 2007; 7: 361–7. PMID: [17137845](#)
31. Pouillot R, Lachenal G, Pybus O, Rousset D, Njouom R. Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon. *Infect Genet Evol*. 2008; 8: 676–81. doi: [10.1016/j.meegid.2008.06.001](#) PMID: [18599380](#)
32. Pujol FH, Loureiro CL. Replacement of hepatitis C virus genotype 1b by genotype 2 over a 10-year period in Venezuela. *J Clin Gastroenterol*. 2007; 41(5):518–20. PMID: [17450037](#)
33. Sulbaran MZ, Di Lello FA, Sulbarán Y, Cosson C, Loureiro CL, Rangel HR, et al. Genetic History of Hepatitis C Virus in Venezuela: High Diversity and Long Time of Evolution of HCV Genotype 2. *PLoS ONE*. 2010; 5(12): e14315. doi: [10.1371/journal.pone.0014315](#) PMID: [21179440](#)
34. Thomas F, Nicot F, Sandres-Sauné K, Dubois M, Legrand-Abravanel F, Alric L, et al. Genetic diversity of HCV genotype 2 strains in south western France. *J Med Virol*. 2007; 79: 26–34. PMID: [17133555](#)
35. Ré VE, Culasso ACA, Mengarelli S, Fariás AA, Fay F, Pisano MB, et al. Phylodynamics of Hepatitis C Virus Subtype 2c in the Province of Córdoba, Argentina. *PLoS One*. 2011; 6(5): e19471. doi: [10.1371/journal.pone.0019471](#) PMID: [21611129](#)
36. Markov PV, van de Laar TJ, Thomas XV, Aronson SJ, Weegink CJ, van den Berk GE, et al. Colonial history and contemporary transmission shape the genetic diversity of hepatitis C genotype 2 in Amsterdam. *J Virol*. 2012; 86:7677–87. doi: [10.1128/JVI.06910-11](#) PMID: [22573865](#)
37. Cristina J. Genetic diversity and evolution of hepatitis C virus in the Latin American region. *J Clin Virol*. 2005; 34(Suppl 2): S1–S7. PMID: [16461234](#)
38. Campiotto S, Pinho JR, Carrilho FJ, Da Silva LC, Souto FJ, Spinelli V, et al. Geographic distribution of hepatitis C virus genotypes in Brazil. *Braz J Med Biol Res*. 2005; 38: 41–9. PMID: [15665987](#)
39. Quarleri JF, Robertson BH, Mathet VL, Feld M, Espínola L, Requeijo MP, et al. Genomic and phylogenetic analysis of hepatitis C virus isolates from Argentine patients: a six-year retrospective study. *J Clin Microbiol*. 2000; 38:4560–8. PMID: [11101596](#)
40. Ansaldi F, Bruzzone B, Salmaso S, Rota MC, Durando P, Gasparini R, et al. Different seroprevalence and molecular epidemiology patterns of hepatitis C virus infection in Italy. *J Med Virol*. 2005; 76(3):327–32. PMID: [15902713](#)
41. Simmonds P. Genetic diversity and evolution of hepatitis C virus— 15 years on. *J Gen Virol*. 85,3173–88 (2004) PMID: [15483230](#)

42. Cantaloube JF, Gallian P, Laperche S, Elghouzzi MH, Piquet Y, Bouchardeau F, et al. Molecular characterization of genotype 2 and 4 hepatitis C virus isolates in French blood donors. *J Med Virol*. 2008; 80: 1732–9. doi: [10.1002/jmv.21285](https://doi.org/10.1002/jmv.21285) PMID: [18712846](https://pubmed.ncbi.nlm.nih.gov/18712846/)
43. Marascio N, Ciccozzi M, Equestre M, Lo Presti A, Costantino A, Cella E, et al. Back to the origin of HCV 2c subtype and spreading to the Calabria region (Southern Italy) over the last two centuries: a phylogenetic study. *Infect Genet Evol*. 2014; 26:352–8. doi: [10.1016/j.meegid.2014.06.006](https://doi.org/10.1016/j.meegid.2014.06.006) PMID: [24973737](https://pubmed.ncbi.nlm.nih.gov/24973737/)
44. Jordier F, Deligny ML, Barré R, De Micco P, Cantaloube JF. Evidence for two phylogenetic clusters within hepatitis C virus (HCV) genotype 2 inferred from analysis of complete coding sequences of 15 HCV strains. *J Med Virol*. 2013; 85: 1754–64. doi: [10.1002/jmv.23674](https://doi.org/10.1002/jmv.23674) PMID: [23861126](https://pubmed.ncbi.nlm.nih.gov/23861126/)
45. Brahim I, Akil A, Mtairag EM, Pouillot R, El Malki A, Nadir S, et al. Morocco underwent a drift of circulating hepatitis C virus subtypes in recent decades. *Archives of virology*. 2012; 157 (3):515–20. doi: [10.1007/s00705-011-1193-7](https://doi.org/10.1007/s00705-011-1193-7) PMID: [22160625](https://pubmed.ncbi.nlm.nih.gov/22160625/)
46. Wikipedia. Italian Tunisians. Available: [http://en.wikipedia.org/wiki/Italian\\_Tunisians](http://en.wikipedia.org/wiki/Italian_Tunisians).