RESEARCH ARTICLE

Molecular characterization and epidemic history of hepatitis C virus using core sequences of isolates from Central Province, Saudi Arabia

Medhat K. Shier[1,2]*, James C. Iles[3], Mohammad S. El-Wetidy[1], Hebatallah H. Ali[1], Mohammad M. Al Qattan[1,4]

1 College of Medicine Research Center, King Saud University, Riyadh, Saudi Arabia, **2** Department of Medical Microbiology and Immunology, College of Medicine, Menoufia University, Menoufia, Egypt, **3** Faculty of Medicine, School of Public Health, Imperial College, London, United Kingdom, **4** Department of Surgery, College of Medicine, King Saud University, Riyadh, Saudi Arabia

* mshier3@gmail.com

Abstract

The source of HCV transmission in Saudi Arabia is unknown. This study aimed to determine HCV genotypes in a representative sample of chronically infected patients in Saudi Arabia. All HCV isolates were genotyped and subtyped by sequencing of the HCV core region and 54 new HCV isolates were identified. Three sets of primers targeting the core region were used for both amplification and sequencing of all isolates resulting in a 326 bp fragment. Most HCV isolates were genotype 4 (85%), whereas only a few isolates were recognized as genotype 1 (15%). With the assistance of Genbank database and BLAST, subtyping results showed that most of genotype 4 isolates were 4d whereas most of genotype 1 isolates were 1b. Nucleotide conservation and variation rates of HCV core sequences showed that 4a and 1b have the highest levels of variation. Phylogenetic analysis of sequences by Maximum Likelihood and Bayesian Coalescent methods was used to explore the source of HCV trans-mission by investigating the relationship between Saudi Arabia and other countries in the Middle East and Africa. Coalescent analysis showed that transmissions of HCV from Egypt to Saudi Arabia are estimated to have occurred in three major clusters: 4d was introduced into the country before 1900, the major 4a clade's MRCA was introduced between 1900 and 1920, and the remaining lineages were introduced between 1940 and 1960 from Egypt and Middle Africa. Results showed that no lineages seem to have crossed from Egypt to Saudi Arabia in the last 15 years. Finally, sequencing and characterization of new HCV isolates from Saudi Arabia will enrich the HCV database and help further studies related to treatment and management of the virus.

Introduction

Hepatitis C virus (HCV) displays very high levels of genetic diversity which have been used to differentiate seven major genotypes and about 80 subtypes [1]. Although the different genotypes and subtypes share basic biological and pathogenic features, they differ in clinical outcomes, response to treatment and epidemiology [2]. The HCV genome is described to have an elevated percentage of sequence variation. When differences at the nucleotide level are between 31% and 33%, HCV strains are divided into genetically different genotypes. Differences between 20% and 25% specify different subtypes and variation below these levels indicate quasispecies [3]. Different HCV genotypes show a considerable variation in geographical distribution where genotypes 1, 2 and 3 have a worldwide distribution. "According to WHO report in 2000, it was found that HCV levels of infection in Egypt are the highest worldwide (15%) where HCV genotype 4 is the most predominant [4]. Unlike Egypt, only 1–3% of Saudi population are infected HCV from which 62% are HCV genotype 4 patients [5]. Genotype 5 is localized in South Africa, France, and Belgium, where, genotype 6 is found Indonesia, Thailand, and Singapore. [6].

One useful HCV region for subtype differentiation is the core that contains 573 nucleotides. This region contains higher mutation rate than the 5′ UTR region. The main function of the core protein is to form the capsid that protects HCV genomic RNA during viral infection. However, through interaction with many cellular proteins, HCV core protein also changes many host pathways [7]. The nonstructural 5B viral region (NS5B) contains 1777 bp and the sequence motifs are highly conserved among all the known RNA-dependent RNA polymerases (RdRps) [8]. RdRp is considered an important target for drug development [9,10], as well as the useful region for HCV genotypes and subtypes identification.

Studying evolution and genetic history of HCV genotype 4 is essential for several reasons. HCV genotype 4 responds less efficiently to interferon-based anti-HCV drug treatment than genotypes 2 and 3, especially in patients of African origin [11, 12]. Recently, the new treatment of HCV by Direct Acting Antiviral agents (DAA) provides pan-genotypic regimens that may be the most effective treatment of most of HCV patients [13, 14]. It has been hypothesized that this phenotype is the result of the long-term presence of genotypes 1 and 4 in Central African populations [12]. There are a number of query concerns about the origin and spread of HCV genotype 4 within Africa. Previous studies have not clearly demonstrated the current and past distribution of HCV genotype 4 strains among different countries or detect the geographic source of the HCV lineages present in the Middle East [12, 15–17]. In this study, we focus on the most common genotypes found in Saudi Arabia (genotype 1 and 4) and mainly in the central province including the capital Riyadh which is considered, from our point of view, an excellent environment for investigating the HCV epidemic history as the population size and number of infected patients are the most as reported before [18, 19]. Similar studies were done on NS5B region in the western province (Jeddah, Saudi Arabia) [20–22]. Our aim is to investigate a number of HCV isolates from Saudi Arabia, sequenced in the core region with analysis of nucleotide variation present. We also examined the epidemiological history of these isolates related to other neighbor countries.

Materials and methods

Ethics and consent statements

Subjects, including human material or human data, in addition to all written informed consents, have been obtained, documented and provided by Pathology Department, College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia (KSA). The project and

data forms were approved by the Ethics Committee at College of Medicine and King Khalid University Hospital, King Saud University, Riyadh, KSA (Ref. no. 16/0172/IRB) in compliance with the Helsinki Declaration (https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/).

HCV isolates

We used 95 serum samples from different HCV chronic patients, collected from King Khaled University Hospital (KKUH) according to ethical regulations. KKUH is a well representative model for this study as it is visited by most of HCV patients from the central province of Saudi Arabia which includes either male or female patients, patients of different ages, origins and from different locations including the capital Riyadh (where the higher population exist) or surrounding areas [23]. We used the same patients' sera and data used previously to compare the 5'UTR genotyping results with those of HCV core [24]. The results of only 54 samples were used in this study because these samples gave a strong signal of PCR products. The unused 40 samples were not suitable for this study because of their lower viral titer, less than 50,000 copies/ml, where these samples gave no or bad results with either the universal primer of HCV 5'UTR region or HCV core primers despite they were positive to HCV antibodies. Few samples gave no results with any of the three primer sets detecting HCV core in either HCV genotype 4 or genotype 1 which prove that their genotyping results by 5'UTR (only genotype 1 or 4 were used) were incorrect. The samples in these mentioned cases were excluded. The viral load was detected using real-time PCR technique and COBAS AmpliPrep /COBAS TaqMan HCV Quantitative Test, version 2.0 Instrument (Roche Molecular Diagnostics, California, USA) and ranged from 200,000 to 25,000,000 copies / ml.

HCV genotyping

The methods for viral RNA extraction, RT-PCR amplification, sequencing reaction and analysis of all HCV isolates have been described previously [24]. Both of HCV core and NS5B primers were designed at the conserved regions that would give the longest sequence and cover the most known genotype 4 and 1 sequences, according to references obtained from HCV database (http://hcv.lanl.gov/content/sequence/NEWALIGN/align.html). All sequences were aligned together by MEGA 6 software [25], http://www.megasoftware.net/, where primers locations were selected. Four primers were designed for amplification of HCV core region in which one forward primer (at position 426–446 relative to H77) works with either one of three reverse primers (at position 732–751 relative to H77). For HCV core primers (nt 426 and nt 751 related to H77), sequences are as follows: The forward primer (CF) 5′-CAG ATC GTT GGC GGA GTT TAC-3′, the reverse primer (CR) 5′- ATRTATCCCATGAGRTCGGC-3′ where sequences of NS5B primers (nt 7952 and nt 8617 related to H77) are as follows: The forward primer (NF) 5′ACCACATCARCTCCGTGTGG′3, the reverse primer (NR) 5′ CTCGT CATRGCYTCCGTGAA′3. The products size of HCV core and NS5B are 326 bp and 684 bp, respectively, that is sufficient for variation detection in these regions. Purified PCR products were sequenced directly with Big Dye Terminator V 3.1 kit (Applied Biosystems, Foster City, CA) using GA-3130 DNA automated sequencer.

Sequence collation

All available HCV genotype 4 sequences were downloaded from GenBank, National Center for Biotechnology Information (NCBI) [26]. Sequences were retained if extended over the sub-genomic region sequenced in this study (positions 426–751 relative to H77) and originated from the Middle East or Africa, resulting in 486 reference sequences. Sequences were

combined with the new sequences gathered in this study and aligned by hand using AliView 1.17.1 [27]. Three alignments were created; all genotype 4 Saudi Arabian sequences (140 sequences), all genotype 1 Saudi Arabian sequences (54 sequences), and all gathered samples (473 sequences). For each isolate, two pieces of data were collected from online databases and the primary literature: year of sampling and country of origin. To study the geographic distribution, we used isolate sequences of Middle Eastern and African countries (S1 Table).

Phylogenetic analysis

Each sequence was run through the HCV genotyping interface, using BLAST and HCV database, to obtain basic genotype information [28, 29]. This was used to place sequences in an alignment with reference genomes from the respective genotype. The table of confirmed HCV genotypes/ subtypes is available at https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sgflavi/35/table-1-confirmed-hcv-genotypessubtypes-may-2015. For HCV core trees, both genotype 1 and genotype 4, we used a maximum-likelihood phylogeny which was generated using Garli 0.951 [30]. We used a general time-reversible substitution model with estimated parameters, estimated base frequencies, no invariable sites and variable sites estimated with a gamma distribution with 4 rate categories. Confidence in the phylogeny was assessed by using 400 bootstrap replicates that were created and summarized with TreeAnnotator 1.8 [31]. Sequences were subtyped according to being placed in a clade with a subtype reference genome with bootstrap support > 0.7.

Bayesian evolutionary analysis

We used the Bayesian Markov Chain Monte Carlo (MCMC) inference method implemented in BEAST v1.8 [31]. The epidemic history of each HCV alignment was estimated. Each analysis used an SDR06 nucleotide substitution model (two independent HKY+Γ substitution models–one for the first and second codon positions, and one for the third), a gamma model of among-site rate heterogeneity with 4 categories, and an uncorrelated log normal relaxed molecular clock. A Bayesian skyline coalescent model was used to estimate epidemic history, with 10 population groups. The effective population size for each group was given an uninformative uniform prior distribution, bounded at 0 and 10^{100}. A normal prior distribution was placed upon the mean evolutionary rate parameter, with a mean of 7.21 x 10^{-4} and standard deviation of 1.6 x 10^{-4} chosen according to evolutionary rates seen in the core region in previous studies [32]. Each MCMC run contained 200 million states, sampled once every 5000 states; trees were sampled every 50,000 states. Multiple MCMC runs were calculated to ensure convergence and were combined to increase the accuracy of parameter estimates. Tracer software version 1.5, http://beast.bio.ed.ac.uk/tracer, was used to monitor MCMC convergence and effective sample sizes. Maximum clade credibility trees were calculated and annotated using TreeAnnotator 1.8 [31]. FigTree v1.3.1, http://beast.bio.ed.ac.uk/figtree, was used to color lineages according to their sampling location using the parsimony criterion.

Results

HCV genotypes and subtypes

For studying HCV sequences, we used serum samples from HCV chronic patients and extracted viral RNA, which was then subjected to direct sequencing. The results of 54 isolates showed that there were two major genotypes; 46 isolates were HCV genotype 4 and 8 isolates were genotype 1. Both genotypes produced PCR products with one or more of the three primer sets designed for HCV core region amplification (Fig 1).
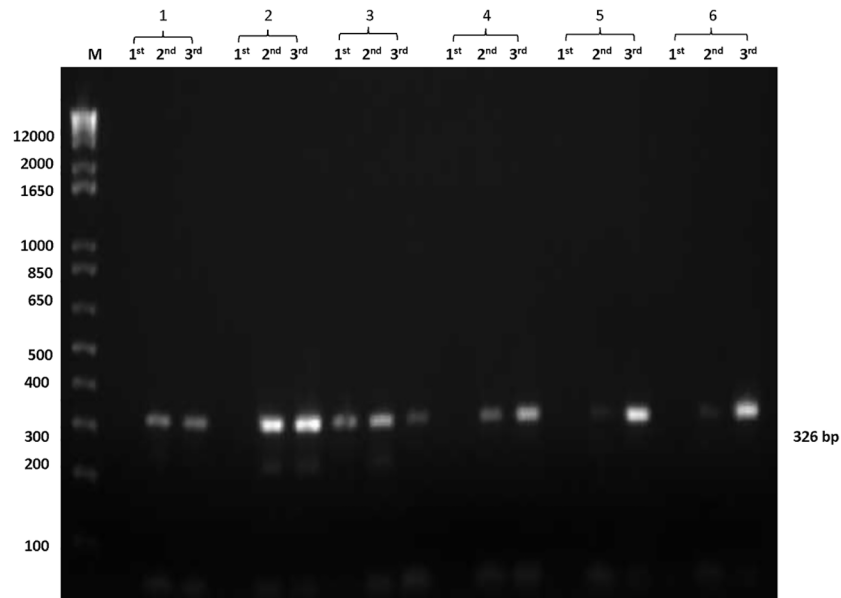
Fig 1. PCR products of HCV core region. PCR products were separated on 1.5% agarose gel. Three different primer sets were used to amplify 326 bp of HCV core region. The isolates appear in this gel were as follows: 1) EGSC1511 {4d}, 2) EGSC1512 {4d}, 3) EGSC1515 {4a}, 4) EGSC1520 {4d}, 5) EGSC1522 {4n} and 6) EGSC1556 {4d}.

https://doi.org/10.1371/journal.pone.0184163.g001

All isolates that were used in this study have been investigated before by direct sequencing of HCV 5'UTR region [15]. Genotyping of HCV using core sequences revealed that some of the isolates have different HCV genotypes like EGSC1503, EGSC1525, EGSC1526, EGSC1543, EGSC1545, EGSC1547 and EGSC1552 that were genotyped as 4 previously while HCV core sequencing showed these isolates are genotype 1. Two other isolates, EGSC1519 and EGSC1521, couldn't be genotyped accurately by 5'UTR sequencing only while core sequencing confirmed that these isolates are genotype 1 and 4, respectively.

Genotype identification of our HCV sequences, using blast search, relied on 95–99% and 85–93% identity for HCV core and NS5B regions, respectively. These identities were calculated against sequences of complete genome references according to subtype and country of origin [S2 Table]. These results were similar for both NCBI BLAST database and the specific HCV database indicating the same genotype. The nucleotide sequences of the HCV isolates have been deposited in the GenBank database under accession numbers from KU883372 to KU883417 for HCV core genotype 4 sequences, KU883418 to KU883425 for HCV core genotype 1 sequences and KU883426 to KU883436 for HCV NS5B genotype 4 sequences.

The HCV genotype 4 core isolates were subtyped into six main groups. 25 sequences were subtyped as 4d, while nine sequences were 4a and six sequences were 4r. Other subtypes such as 4m, 4n, and 4o divided the remaining six sequences equally. For HCV core genotype 1, two isolates were 1a and six were 1b. We picked nine isolates of those were genotype 4 by core sequences to be sequenced in NS5B region, which is known to be more reliable for subtyping according to higher variation rates, confirmed different subtypes resulted [33], to confirm subtyping results from HCV core. The Basic Local Alignment Search Tool (BLAST) search of the resulted sequences of both regions agreed in genotype classification, confirming all to be HCV genotype 4. The nine isolates selected for HCV NS5B sequencing showed similarity to subtypes detected by HCV core except for isolate EGSN1541 that was closer to 4a than 4d (data not shown).

Table 1. HCV genotypes and subtypes compared in respect to nucleotides conservation and variation rates.

	HCV core genotype 1		HCV core genotype 4					
Subtypes	1a	1b	4a	4d	4m	4n	4o	4r
No. of isolates	2	6	9	25	2	2	2	6
Total	8		46					
Conserved nucleotides	277 (97%)	250 (88%)	247 (87%)	259 (91%)	279 (98%)	281 (98.5%)	270 (95%)	275 (96.5%)
Varied nucleotides	8 (3%)	35 (12%)	38 (13%)	26 (9%)	6 (2%)	4 (1.5%)	15 (5%)	10 (3.5%)
Variation to HCV database References	1–3%	2–4%	1–7%	0–8%	1–3%	0–5%	2–6%	1–4%
No. of HCV database References used	44	70	81	43	5	8	4	7
Isolate with highest variation rate	EGSC1543	EGSC1526	EGSC1536	EGSC1504	EGSC1509 EGSC1521	EGSC1522	EGSC1524	EGSC1531
Total	285 (100%)							

HCV nucleotide conservation and variation

Calculating the nucleotide variation percentage was done by hand and MEGA software for HCV isolates of each subtype separately. We calculated the percentages of variation and conservation between resulted sequences and reference sequences related to their estimated subtype and also we compared sequences of the same subtype to each other to estimate how far they are from each other or from consensus sequences. The references used were as follows: H77.NC_004102 (1a), EU781827 (1b), Y11604 (4a), DQ418786 (4d), FJ462433 (4m), FJ462441 (4n), FJ462440 (4o) and FJ462439 (4r) (Table 1). HCV core genotype 4 subtypes revealed that 4a is the most variable with 13% of nucleotides had variation in the same position of aligned sequences. For the most common subtype 4d, the variation was lesser than 4a; 9%. For HCV genotype 1 sequences, subtype 1a showed only 3% variation whereas subtype 1b showed 12% variation of the aligned nucleotides. Comparison of our isolates to all previously subtyped HCV sequences obtained from GenBank, shows high variation rate that reaches up to 8% for 4d samples, particularly isolate EGSC1504 and up to 7% for 4a samples, particularly isolate EGSC1536, where 4d and 4a represent most of the isolates used in this study (Table 1).

When examining the HCV core genotype 4 isolates sequence logo, it was noticeable that variation among different subtypes occurred mostly between two nucleotides in 58 sites, while triplet variation occurred in 17 sites and only 3 sites showed 4 nucleotide variation (Fig 2). The double variation revealed that most of the variations are either pyrimidines (Y = C or T) in 27 nucleotide positions or purines (R = A or G) in 19 positions. Similar results occurred for HCV core genotype 1 and HCV NS5B genotype 4. These results demonstrated that core sequences of genotype 1 were the most conserved as only 34 out 285 sites (~12%) had variation compared to those of genotype 4 with 78 sites (~27%) showed variation (Fig 3) and (Table 2).

Phylogenetic trees

The phylogenetic analysis of HCV core genotype 4 isolates showed the same genotype and subtype identity as BLAST and HCV database. Pairwise nucleotide alignment divided sequences into six different subtypes. The majority of isolates (n = 25) were grouped as one cluster with higher bootstrap values to references from the USA and Canada that were identified as 4d and lower values to references of 4t and 4p. Two isolates came closer to 4m references FJ462433 and JX227972 from Canada and UK, respectively. Nine isolates were gathered in subtype 4a cluster with references Y11604 and DQ988074 from Egypt and

Fig 2. Sequence logo diagram of HCV core sequences. This was created by the alignment tool MEGA 5.05 Software and the WebLogo3 software (http://weblogo.threeplusone.com/create.cgi). The primers' sequences at both ends were deleted from all represented sequences. The vertical bar is 2 bits high. The nucleotides overlap persistent in one position indicates the polymorphic sites. The height of each letter represents the relative proportion of each nucleotide at that position. (A) HCV core genotype 4 logo using forty-six isolates. (B) HCV core genotype 1 logo using eight isolates.

https://doi.org/10.1371/journal.pone.0184163.g002
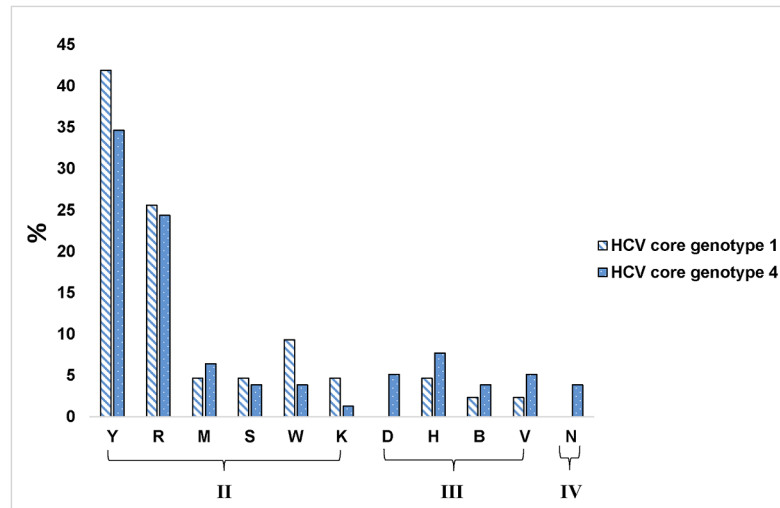
Fig 3. Bar graph representation of the nucleotide variation, expressed as mixed bases, among HCV genotype 1 and 4 sequences in the core region. The Greek numbers indicate how many nucleotides are called for the same positions where II, III, and IV indicates two, three and four substitutions, respectively, for the same nucleotide position. For the outline of the mixed bases, refer to Table 2.

DQ418789 from USA. Two sequences showed higher bootstrap values to subtype 4o references JX227977 and FJ462440 from UK and Canada, respectively. Two other clusters of subtype 4r with 6 isolates and subtype 4n with two isolates showed up at the bottom of the phylogenetic tree with corresponding references. Other subtypes of genotype 4 didn't pairwise with any of our sequences (Fig 4). Sequences thought to be HCV genotype 1 by BLAST were confirmed by phylogenetic analysis. Six sequences gave pairwise alignment with 1b references which are mainly from Europe and North America whereas only two sequences matched 1a references from USA (Fig 5).

Table 2. Types and rates of nucleotides substitution among HCV genotypes 1 and 4.

Nucleotides substitution type	Mixed bases symbols*	HCV genotype 1 core		HCV genotype 4 core	
		No.	**%**	**No.**	**%**
II	Y	18	41.86	27	34.615
	R	11	25.581	19	24.359
	M	2	4.6512	5	6.4103
	S	2	4.6512	3	3.8462
	W	4	9.3023	3	3.8462
	K	2	4.6512	1	1.2821
III	D	0	0	4	5.1282
	H	2	4.6512	6	7.6923
	B	1	2.3256	3	3.8462
	V	1	2.3256	4	5.1282
IV	N	0	0	3	3.8462
	Total	34	100	78	100

* **M**: A or C, **R**: A or G, **W**: A or T, **S**: C or G, **Y**: C or T, **K**: G or T, **V**: A or C, or G, **H**: A or C or T, **D**: A or G or T, **B**: C or G or T and **N**: any base (*Cornish-Bowden (1985) Nucl. Acids Res. 13*: 3021–3030).
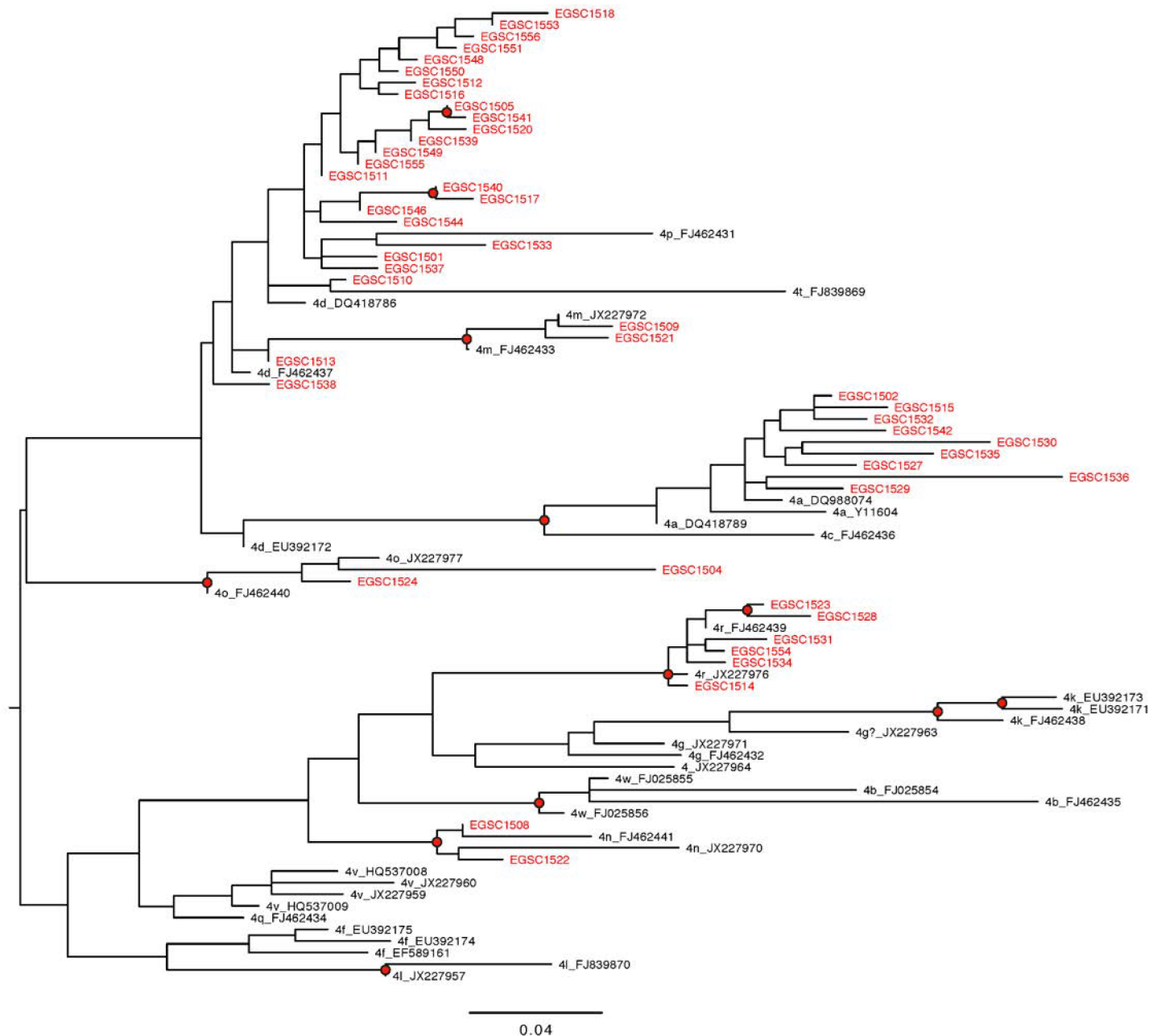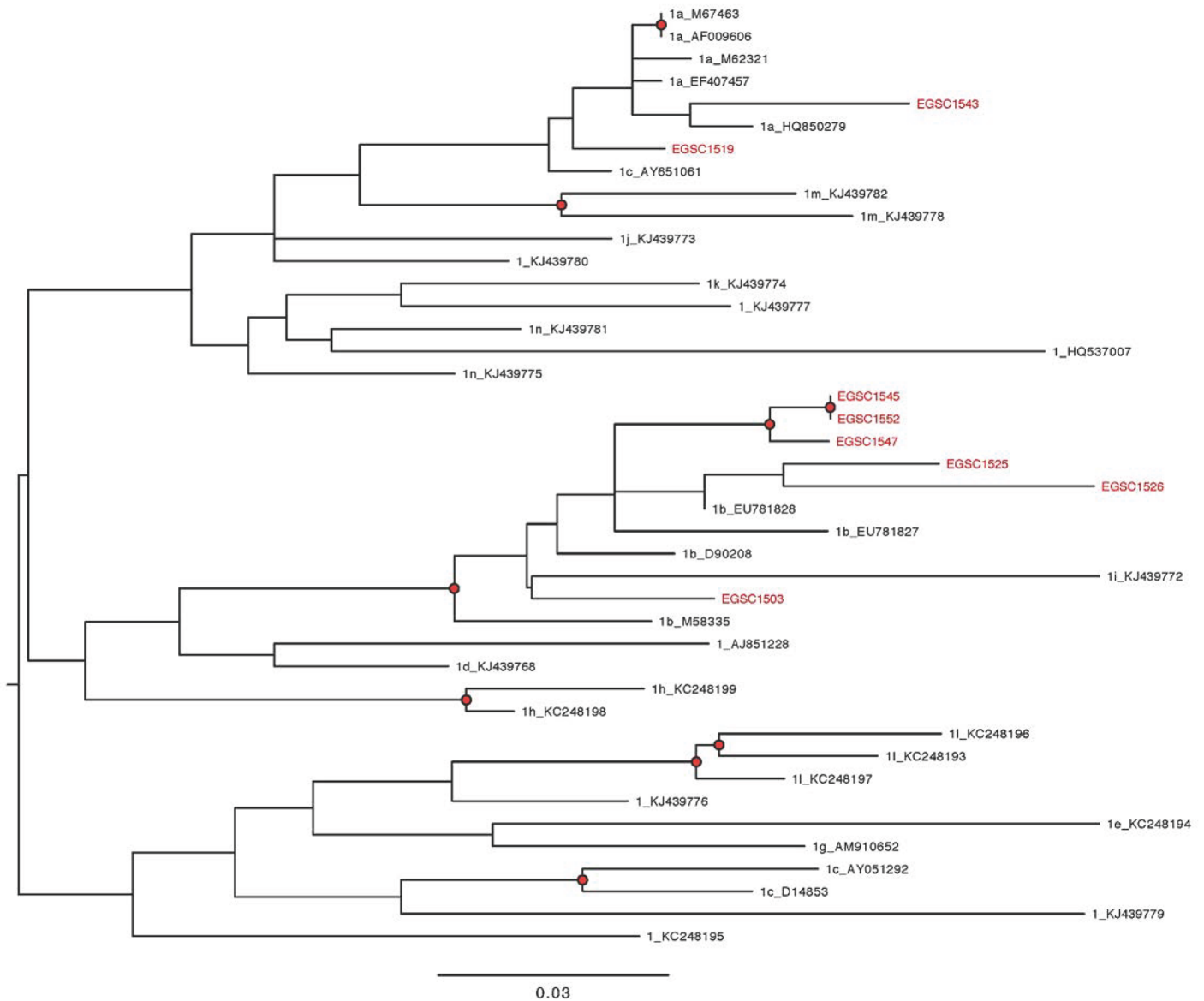
Fig 4. Maximum-Likelihood phylogeny of HCV core genotype 4 sequences. Maximum-Likelihood phylogenetic trees were used to subtype HCV core samples gathered in this study. Phylogenies were estimated from a 326-nt long alignment of the HCV core gene using a Generalized Time Reversible (GTR) substitution model with a gamma distribution model of among-site rate variation. Internal nodes are highlighted with a black circle if they have bootstrap support >70% (400 maximum likelihood replicates). Scale bar shows the expected number of substitutions per site.

https://doi.org/10.1371/journal.pone.0184163.g004

Recent transmission in Saudi Arabia

The maximum-clade credibility tree was generated from all genotype 4 and genotype 1 samples gathered in Saudi Arabia (Figs 6 and 7). The samples gathered in this study are dispersed across each tree, and aside from one clade in genotype 1b and one in 4d, do not seem to form clusters of recent transmission. Many previous studies used a closer number of samples to be

Fig 5. Maximum-Likelihood phylogeny of HCV core genotype 1 sequences. The phylogenetic tree, for the ten HCV genotype 1 sequences from HCV core region, was constructed in the same manner like that of HCV genotype 4.

used in phylogenetic analysis [34] therefore, the samples found in this study are likely a representative sample of HCV strains in Saudi Arabia.

Previous study by The Bayesian skyline plot (BSP) analysis showed that HCV subtype 4d, which dominates the current HCV infections in Saudi Arabia since ~1970, was found since 1900 and increases slowly till 1950 then rapidly till 1980 [16]. The skyline plot produced by these analysis, colored according to genotype, shows how the virus' epidemic size (calculated as effective population size x generation time) has changed over time (Fig 8), and seems to be divided into three distinct time periods: before 1960, when the virus circulated at a low level; between 1960 and 2005, where the epidemic size grew by two orders of magnitude; and the time between 2005 and the present, when the epidemic size fell by an order of magnitude. This
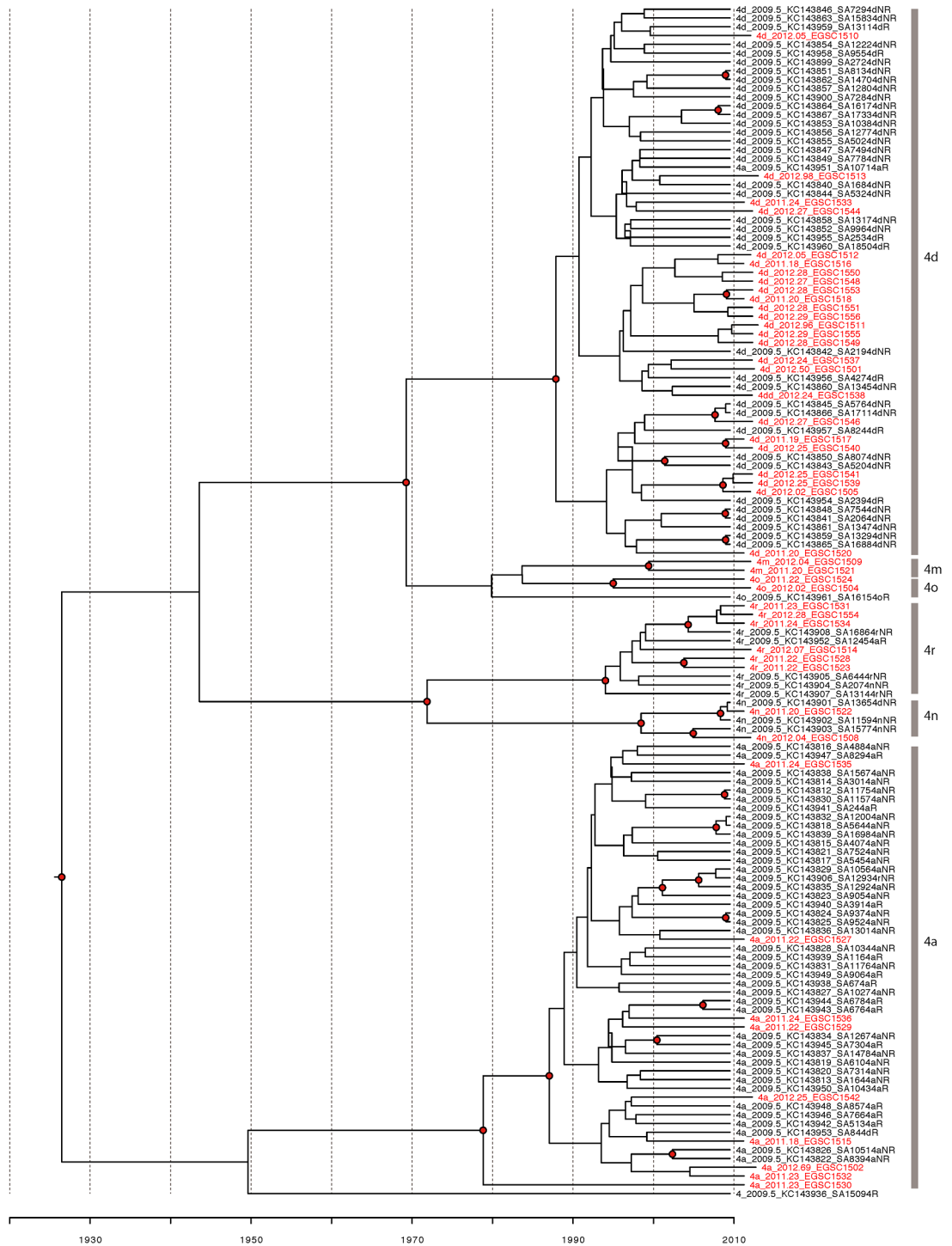
Fig 6. Maximum clade credibility molecular clock phylogenies of HCV genotype 4 isolates circulating in Saudi Arabia. Phylogenies were estimated from the combined alignments of all genotype 4 samples. Trees were generated using BEAST 1.8 and summarized using TreeAnnotator (http://beast.bio.ed.ac.uk). Branch lengths represent time-see scale bar at the bottom. Sequences originating in this study are colored red and labeled with their subtype, collection date and isolate name; other sequences are labeled with their subtype, collection date, accession number and isolate name.

https://doi.org/10.1371/journal.pone.0184163.g006
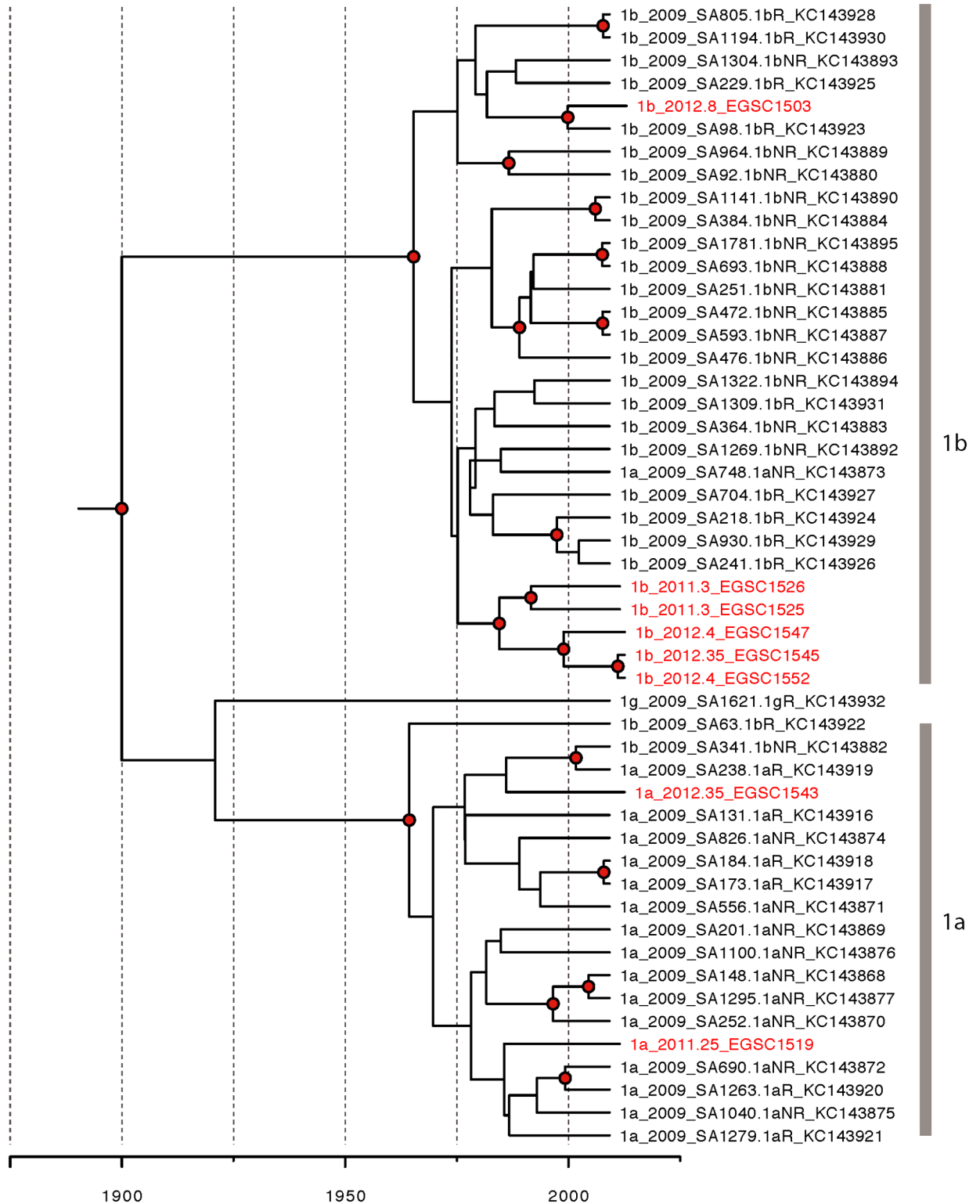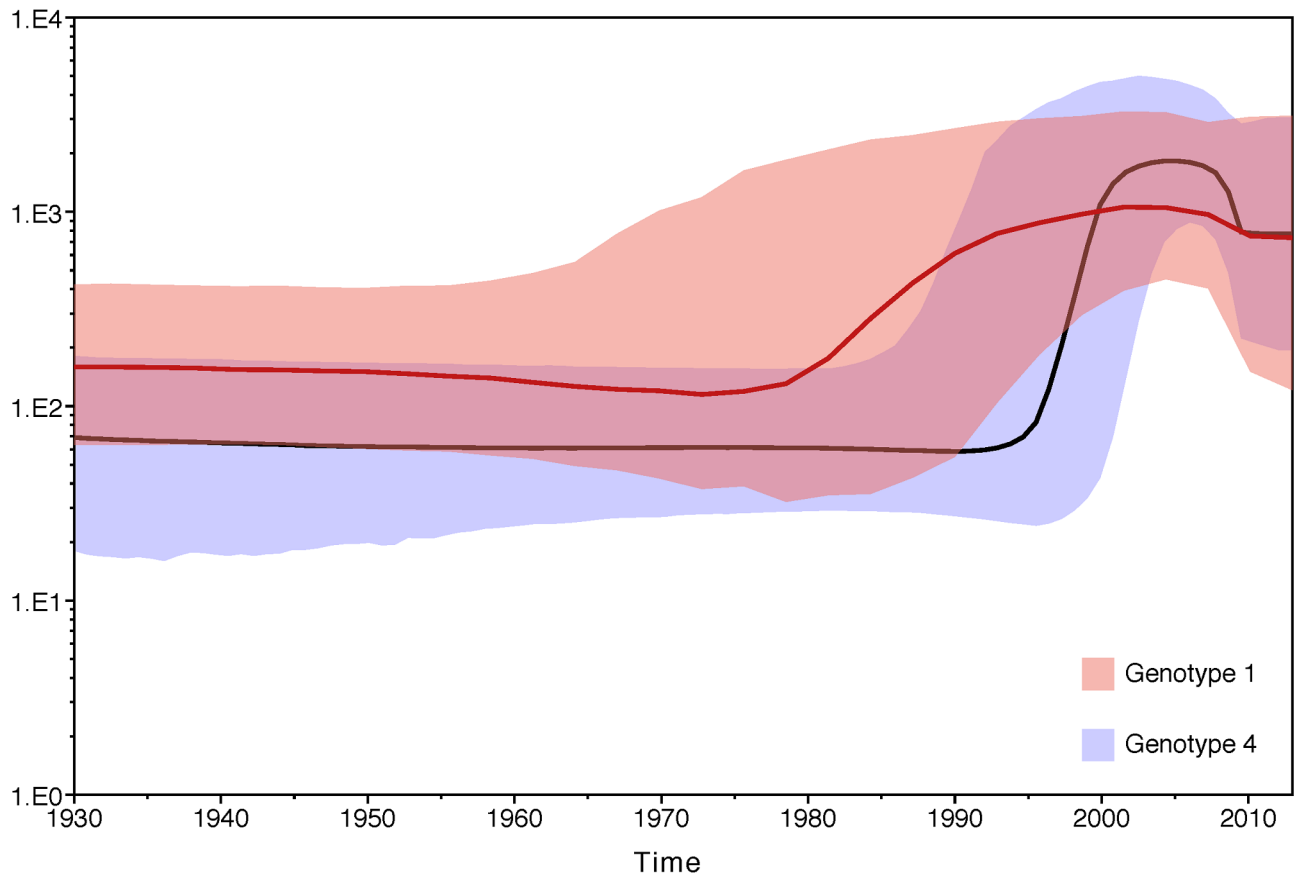
Fig 7. Maximum clade credibility molecular clock phylogenies of HCV genotype 1 isolates circulating in Saudi Arabia. The tree construction is similar to that of HCV genotype 4.

Fig 8. Transmission history and citation of HCV isolates in Saudi Arabia. HCV genotypes 1 and 4 in Saudi Arabia were analyzed by Bayesian Evolutionary Analysis, skyline plot. Each plot shows the change through time in epidemic size (effective population size x generation time) on a log scale. Thick lines indicate the median estimate and the shaded areas show the 95% highest posterior density credible region. The red curve was estimated from HCV genotype 1 sequences, while the blue curve was estimated from HCV genotype 4 sequences.

pattern is more clearly seen in the genotype 4 skyline, likely due to the greater number of sequences used in that analysis.

This pattern is in contrast to the dynamic seen in Egypt and the Democratic Republic of the Congo, where the epidemic size stayed at the level reached by the exponential growth period or even rose intermittently [32, 35, 36]. This implies that measures to control the HCV epidemic have been more effective in Saudi Arabia than in those other countries, although HCV transmission seems to have reached an equilibrium resulting in an epidemic size of 5×10^2.

Sources of transmission into Saudi Arabia

To determine if a recent transmission from other countries was the source of recent epidemics in Saudi Arabia, we analyzed an alignment containing all genotype 4 samples gathered across Africa and the Middle East. The maximum-clade credibility tree, resulting from this analysis, was created with each branch colored according to its most probable country of origin (Fig 9). Strains associated with Saudi Arabia fall into two categories; clades that transferred into the country more than 50 years ago and have spread within the country ever since (e.g. 4d, 4r, the larger clade in 4a) and individual strains transmitted into the country over the past 50 years. While the former strains have both a central African and Egyptian origin, the latter strains all

Fig 9. Comparison of HCV isolates sequences with those from African and Middle Eastern countries.
The Maximum Clade Credibility molecular clock phylogenies were used to cite the HCV isolates from Saudi
Arabia among isolates from the surroundings countries. Phylogeny was estimated from the alignment of
genotype 4 sequences collected in this study as well as sequences from online databases. Tips are colored
according to positions of sequence origin (see key). The locations of internal branches were deduced using

parsimony and are colored similarly. Branch lengths represent time–see scale bar at the bottom. Nodes with a posterior probability >0.7 are labeled with a white circle. Samples gathered in this study are labeled with red stars; subtypes are indicated on the right side of the diagram.

originate from Egypt. Even the most recent introduction of HCV from Egypt to Saudi Arabia had a coalescence time of 1974 (95% highest posterior distribution: 1966, 2002), implying that cross-border transmission makes up a very small proportion of HCV replication within Saudi Arabia in the modern day. Transmissions of HCV from Egypt to Saudi Arabia are estimated to have occurred in three major clusters: 4d was introduced into the country before 1900, the major 4a clade's MRCA was introduced between 1900 and 1920, and the remaining lineages were introduced between 1940 and 1960. This suggests that the majority of HCV replicating in Saudi Arabia belongs to lineages that have been present in the country for half a century or more, and that modern migration presents a comparatively small source of HCV diversity. Most of the HCV transmission from other countries to Saudi Arabia originate in Egypt, but this is a relatively small number compared to the large quantity of HCV transmission occurring within the country. In fact, there are no lineages in the tree that seem to have crossed from Egypt to Saudi Arabia in the last 15 years, and in that time 40 branches were created within Saudi Arabia.

Discussion

The overall HCV prevalence in Saudi Arabia reflects how the disease burden is difficult to be controlled without remarkable changes in the available treatment paradigm [37]. HCV was the seventh (95% UI, Uncertainty Interval) leading cause of death worldwide in 2013, compared with tenth in 1990 as reported by WHO [38]. DAAs that proved more effective and less toxic on genotype 1, however, more studies are needed for genotype 4 patients [39]. HCV genotyping is essential for identifying different types of HCV and thus determining the best approach for treatment. HCV genotyping can be done either by Real Time PCR assay of 5'UTR or by direct sequencing of different HCV viral regions. Genotyping that depends on a targeting 5'UTR is not the most accurate due to lack of nucleotide polymorphisms in 5'UTR [40, 41]. Direct sequencing analysis is considered the most accurate method. Both HCV core and NS5B viral regions are considered the most reliable regions for genotyping because of sufficient genetic diversity and absence of recombination events [42]. Genotyping of HCV by BLAST analysis depends on the comparison of newly obtained sequences to previously genotyped sequences in the HCV database relying on the percentage of identities among sequences. Continuous submitting of more genotyped sequences to database reinforce the usage of BLAST analysis as a more reliable method [43].

Mutation rate defines the ability of a virus to maintain important information while surviving with different environmental challenges [44, 45]. Previous studies showed the manifestation of more variability in sequences of HCV NS5B in comparison to the core sequences [46, 47]. The HCV core region has been proved to be better for HCV subtyping due to subtype-specific variation in a small proportion of sites [47, 48]. The use of the more variable region, NS5B, can be relied upon for accurate discrimination between and within subtypes [49]. Although NS5B is the preferred region for subtyping, amplification of this region is not always possible due to primer-target mismatch within the highly variable NS5B sequences. Unsuccessful sequencing of the NS5B gene has been documented by some investigators despite multiple attempts to amplify cDNA with the NS5B primers [50]. In the present study, three primer sets were used to amplify HCV core region but NS5B amplification required the use of seven sets.

Our HCV core genotype 1 and 4 sequences showed conservation rates of 94% and 73%, respectively. The high percentage of NS5B protein variation causes the RNA polymerase enzyme to make significant errors that result in nucleotide changes introduced in the viral genome [51, 52]. Previous studies indicated that variation in other viruses like HIV or HBV can enhance drug resistance variants development [53–55], similar to the development of HCV protease inhibitor resistance changes [56].

Since the sequence similarities in the core and NS5B regions between tested isolates and strains from the HCV database were >93 and 87.8%, respectively [57, 58], genotypes of sequences at the top BLAST outputs were considered as identical to the tested isolates. Based on the BLAST result, genotypes identified by core and NS5B sequence analysis were compared and our results showed that genotyping of both regions were identical. HCV core genotype 4 sequences of 46 isolates and the reference strains of genotypes 4a, 4d, 4m, 4n, 4o, and 4r were grouped into six clusters in the phylogenetic tree. The bootstrap values of each cluster exceeded 70%, indicating that the topology of core sequences was highly reliable. Those infected with HCV genotype 4 are likely to be Middle Eastern or South Asian, however, some references from USA, Canada, and Europe show some combination with our isolates especially 4d sequences. Only eight isolates were identified as HCV genotype 1. Those sequences were compared to different references representing different genotype 1 subtypes. Similar to genotype 4, Phylogenetic tree results of genotype 1 confirm that the two clusters, 1a and 1b, occurred and indicated Middle Eastern population.

Sequences of HCV core were compared to all published sequences from Saudi Arabia and Egypt because of more traveling incidents between the two neighbor countries. Egypt has a uniquely high HCV prevalence, resulting from collective campaigns to treat bilharziasis with injection therapy during the period from 1960 to 1987 [59]. Genetic analysis suggests that HCV was present at low prevalence in Egypt around the 1920s [35]. The majority of the Egyptian epidemic arrived from Central Africa in time between 1860 and 1925 (the estimated date of the MRCA of subtype 4a). It is clear that the Egyptian HCV epidemic arose from more than one Central Africa strain, as subtypes 4a, 4m, and 4o each transferred to Egypt separately [60]. A previous study of isolates collected in Kinshasa, the capital of the DRC during 2007, show that there was a lengthy period of endemic transmission, during which the effective number of infections was low, prior to the twentieth century. There was a severe switch from endemic infection to rapid exponential growth starting about the 1950s. The molecular clock and phylogeographic results clearly indicated that HCV genotype 4 epidemics in the Middle East, and North Africa, all originated from Central Africa [15].

The possibility of infection routes other than blood transmission including previous surgeries and dental procedures were well-documented. Contact-associated transmissions, such as toothbrushes, tattooing, folk body piercing, and non-medicalized circumcision, as well as exposure to skin lesion exude, are also significant risk factors [61]. Sharing of equipment used for parenteral practices is a significant way of HCV transmission in rural regions of Africa [62]. However, such methods as body scarification, clitodectomy using unsterilized tools, circumcision, and shaving by local barbers are known to be accountable for HCV transmission in Africa. Previously studied strains showed similarity with Asian, European and American strains, potentially due to migration events. That study reports a large community of South Asians settled in European countries [49]. This indicates that migration from one place to another might contribute to the distribution of HCV subtypes that would not, otherwise, exist in Saudi Arabia.

Saudi Arabia is an open community that embraces different cultures. In our study, only 16% of isolates were from non-African/Middle Eastern subtypes. Indeed, the coalescent tree shows evidence that the great majority of HCV infections within Saudi Arabia are acquired

from and passed on to Saudi Arabian residents. Despite the different non-hygienic practices that cause the spread of different HCV subtypes in Saudi Arabia, the improvement of health services, investigation, and treatment that emerged in the last fifteen years have decreased the rate of transmission. The newly identified and published HCV sequences will enrich the online sequence databases which will increase the possibilities for more precise analysis and identification for any new HCV transmission to Saudi Arabia.

Supporting information

S1 Table. Geographic origin of sequences used in phylogenetic analysis. Number of sequences African and Middle Eastern countries that were used in the MCC tree of HCV core isolates from Saudi Arabia.
(DOCX)

S2 Table. List of references used in the construction of maximum likelihood trees and phylogenetic analysis of both HCV Core region. All references express HCV complete genomes and were chosen according to the subtype and country of origin.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Medhat K. Shier, Mohammad M. Al Qattan.

Data curation: Medhat K. Shier, Hebatallah H. Ali.

Formal analysis: James C. Iles, Mohammad S. El-Wetidy.

Investigation: Mohammad S. El-Wetidy.

Methodology: Mohammad S. El-Wetidy, Hebatallah H. Ali.

Project administration: Mohammad S. El-Wetidy, Mohammad M. Al Qattan.

Resources: Mohammad S. El-Wetidy, Hebatallah H. Ali.

Software: James C. Iles, Mohammad S. El-Wetidy.

Supervision: Medhat K. Shier, Mohammad M. Al Qattan.

Validation: James C. Iles, Mohammad S. El-Wetidy, Hebatallah H. Ali.

Writing – original draft: Medhat K. Shier, Mohammad S. El-Wetidy, Hebatallah H. Ali, Mohammad M. Al Qattan.

Writing – review & editing: Medhat K. Shier, Mohammad S. El-Wetidy, Hebatallah H. Ali, Mohammad M. Al Qattan.

References

1. Smith D, Bukh J, Kuiken C, Muerhoff A, Rice C, Stapleton J, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014; 59:318–327. https://doi.org/10.1002/hep.26744 PMID: 24115039

2. González-Candelas F, López-Labrador FX, Bracho MA. Recombination in hepatitis C virus. Viruses. 2011; 3(10): 2006–2024. https://doi.org/10.3390/v3102006 PMID: 22069526

3. Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005; 42(4): 962–973.

4. World Health Organization. Hepatitis C. Geneva, Switzerland: World Health Organization, 2000 WHO fact sheet 164. http://www.who.int/mediacentre/factsheets/fs164/en/print.html. Accessed 1 August 2010.

5. Nguyen MH, Keeffe EB. Prevalence and treatment of hepatitis C virus genotypes 4, 5, and 6. 2005; Clin Gastroenterol Hepatol, 3 (Suppl. 2) (2005), pp. S97–S101.

6. González-Horta EE, Marante J, Amador-Cañizares Y, Alvarez-Lajonchere L, Guerra I, Martínez-Donato G, et al. Analysis of hepatitis C virus core encoding sequences in chronically infected patients reveals mutability, predominance, genetic history and potential impact on therapy of Cuban genotype 1b isolates. Eur Rev Med Pharmacol Sci. 2011; 15(11): 1320–1327. PMID: 22195367

7. Polyak SJ, Klein KC, Shoji I, Miyamura T, Lingappa JR. Assemble and Interact: Pleiotropic Functions of the HCV Core Protein. In: Tan SL, editor. Hepatitis C Viruses: Genomes and Molecular Biology. Norfolk (UK): Horizon Bioscience; 2006. Chapter 3. http://www.ncbi.nlm.nih.gov/books/NBK1625/

8. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J.1989; 8(12): 3867–3874. PMID: 2555175

9. Beaulieu PL, Tsantrizos YS. Inhibitors of the HCV NS5B polymerase: new hope for the treatment of hepatitis C infections. Curr Opin Investig Drugs. 2004; 5(8): 838–850. PMID: 15600240

10. Wu J, Hong Z. Targeting NS5B RNA-dependent RNA polymerase for anti-HCV chemotherapy. Curr Drug Targets Infect Disord. 2003; 3(3): 207–219. PMID: 14529354

11. Chen Y, Xu HX, Wang LJ, Liu XX, Mahato RI 554, Zhao YR. Meta-analysis: IL28B polymorphisms predict sustained viral response in HCV patients treated with pegylated interferon-α and ribavirin. Aliment Pharmacol Ther. 2012; 36(2): 91–103. https://doi.org/10.1111/j.1365-2036.2012.05131.x PMID: 22591106

12. Rose R, Markov PV, Lam TT, Pybus OG. Viral evolution explains the associations among hepatitis C virus genotype, clinical outcomes, and human genetic variation. Infect Genet Evol. 2013; 20: 418–421. https://doi.org/10.1016/j.meegid.2013.09.029 PMID: 24140473

13. Kohli A, Osinusi A, Sims Z, Nelson A, Meissner EG, Barrett LL, et al. Virological response after 6 week triple-drug regimens for hepatitis C: a proof-of-concept phase 2A cohort study. Lancet 2015; 385:1107–1113. https://doi.org/10.1016/S0140-6736(14)61228-9 PMID: 25591505

14. Hathorn E, Elsharkawy AM. Management of hepatitis C genotype 4 in the directly acting antivirals era. BMJ Open Gastroenterol. 2016; 3(1): e000112. eCollection 2016. https://doi.org/10.1136/bmjgast-2016-000112 PMID: 27752338

15. Iles JC, Raghwani J, Harrison GL, Pepin J, Djoko CF, Tamoufe U et al. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology. 2014; 464–465: 233–243. https://doi.org/10.1016/j.virol.2014.07.006 PMID: 25105489

16. Khan A, Al Balwi M, AlAyyar L, AlAbdulkareem I, Albekairy A, Aljumah A.Tracing the epidemic history of hepatitis C virus genotypes in Saudi Arabia.Infect Genet Evol. 2017; 52:82–88. https://doi.org/10.1016/j.meegid.2017.04.024 PMID: 28458032

17. Al Traif I, Al Balwi MA, Abdulkarim I, Handoo FA, Alqhamdi HS, Alotaibi M, Aljumah A, Al Ashgar HI, Bzeizi K, Al Quaiz M, Alalwan A, Al Hamoudi W, Sanai F, Abdo A. HCV genotypes among 1013 Saudi nationals: a multicenter study. Ann Saudi Med. 2013; 33(1):10–2. https://doi.org/10.5144/0256-4947.2013.10 PMID: 23458933

18. Abdel-Moneim AS, Bamaga MS, Shehab GM, Abu-Elsaad AA, Farahat FM. HCV infection among Saudi population: high prevalence of genotype 4 and increased viral clearance rate. PLoS One. 2012; 7 (1): e29781. https://doi.org/10.1371/journal.pone.0029781 PMID: 22253780 Epub 2012 Jan 13.

19. Al Ashgar HI, Khan MQ, Al-Ahdal M, Al Thawadi S, Helmy AS, Al Qahtani A, Sanai FM. Hepatitis C genotype 4: genotypic diversity, epidemiological profile, and clinical relevance of subtypes in Saudi Arabia. Saudi J Gastroenterol. 2013; 19(1):28–33. https://doi.org/10.4103/1319-3767.105920 PMID: 23319035

20. Osoba AO, Ibrahim M, Abdelaal MA, Al-Mowallad A, Al Shareef B, Haj Hussein B Al, Hepatitis C Virus Genotyping by Polymerase Chain Reaction and DNA Enzyme Immunoassay among Saudi Patients in the Western Province, Saudi Arabia. 2000; 20(5–6): 394–397

21. El Hadad S, AlHamdan H, Linjawi S. Partial sequencing analysis of the NS5B region confirmed the predominance of hepatitis C virus genotype 1 infection in Jeddah, Saudi Arabia. PLoS One. 2017; 12 (5): e0178225 https://doi.org/10.1371/journal.pone.0178225 PMID: 28552946

22. Fakeeh M, Zaki AM. Hepatitis C: Prevalence and common genotypes among ethnic groups in Jeddah, Saudi Arabia. Am J Trop Med Hyg. 1999; 61:889–92. PMID: 10674665

23. Al-Faleh FZ, Ramia S. Hepatitis C virus (HCV) infection in Saudi Arabia: A review. Ann Saudi Med. 1997; 17(1):77–82.

24. Shier MK, El-Wetidy MS, Ali HH, Al-Qattan MM. Characterization of Hepatitis C Virus Genotypes by Direct Sequencing of HCV 5´UTR Region of Isolates from Saudi Arabia. PLoS One. 2014; 9(8): e103160. https://doi.org/10.1371/journal.pone.0103160 PMID: 25099694 eCollection 2014.

25. Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 2013; 30:2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

26. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. Bioinformatics. 2005; 21(3): 379–384. https://doi.org/10.1093/bioinformatics/bth485 PMID: 15377502

27. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014; 30(22): 3276–3278. https://doi.org/10.1093/bioinformatics/btu531 PMID: 25095880

28. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics. 2005; 21(19): 3797–3800. https://doi.org/10.1093/bioinformatics/bti607 PMID: 16076886

29. Alcantara LC, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res. 2009; 37(Web Server issue): W634–42. https://doi.org/10.1093/nar/gkp455 PMID: 19483099

30. Zwickl, DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin. 2006. https://repositories.lib.utexas.edu/handle/2152/2666.

31. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29(8): 1969–1973. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

32. Hogan CA, Iles J, Frost EH, Giroux G, Cassar O, Gessain A et al. Epidemic History and Iatrogenic Transmission of Blood-borne Viruses in Mid-20th Century Kinshasa. J Infect Dis. 2016. pii: jiw009

33. Lamoury FM, Jacka B, Bartlett S, Bull RA, Wong A, Amin J et al. The Influence of Hepatitis C Virus Genetic Region on Phylogenetic Clustering Analysis. PLoS One. 2015; 10(7):e0131437. https://doi.org/10.1371/journal.pone.0131437 PMID: 26192190 eCollection 2015.

34. Simmonds P. Genetic diversity and evolution of hepatitis C virus—15 years on. J Gen Virol. 2004; 85 (Pt 11): 3173–3188. https://doi.org/10.1099/vir.0.80401-0 PMID: 15483230

35. Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol. 2003; 20(3): 381–387.

36. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005; 22(5): 1185–1192. https://doi.org/10.1093/molbev/msi103 PMID: 15703244

37. Aljumah AA, Abaalkhail F, Al-Ashgar H, Assiri A, Babatin M, Al Faleh Fet al. Epidemiology, disease burden, and treatment strategies of chronic hepatitis C virus infections in Saudi Arabia in the new treatment paradigm shift. Saudi J Gastroenterol. 2016 22(4):269–81. https://doi.org/10.4103/1319-3767.187609 PMID: 27488321

38. Stanaway JD, Flaxman AD, Naghavi M, Fitzmaurice C, Vos T, Abubakar I, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. Lancet. 2016; 388(10049):1081–8. https://doi.org/10.1016/S0140-6736(16)30579-7 PMID: 27394647

39. Albenmousa A, Al Obary E, Bzeizi K. Treatment Options for HCV Genotype-4. J Infect Dis Ther. 2016; 4:266. https://doi.org/10.4172/2332-0877.1000266

40. Benedet M, Adachi D, Wong A, Wong S 600, Pabbaraju K, Tellier R et al. The need for a sequencing-based assay to supplement Abbott m 2000 RealTime HCV Genotype II assay: a 1 year analysis. J Clin Virol. 2014; 60(3): 301–304. https://doi.org/10.1016/j.jcv.2014.04.005 PMID: 24794397

41. Pollicita M, Cento V, Paba P, Perno CF, Ciotti M. Nucleotide polymorphisms in the 5´UTR region of HCV can affect the ability of two widely used assays to assign an HCV genotype. J Virol Methods. 2013; 193(1): 205–208. https://doi.org/10.1016/j.jviromet.2013.06.002 PMID: 23791962

42. McCormick AL, Macartney MJ, Abdi-Abshir I, Labbett W, Smith C, Irish D et al. Evaluation of sequencing of HCV core/E1, NS5A and NS5B as a genotype predictive tool in comparison with commercial assays targeting 5'UTR. J Clin Virol. 2015; 66: 56–59. https://doi.org/10.1016/j.jcv.2015.03.006 PMID: 25866338

43. Cai Q, Zaho Z, Liu Y, Shao X, Gao Z. Comparison of three different HCV genotyping methods core, NS5B sequence analysis and line probe assay. Int J Mol Med. 2013; 31(2): 347–352. PMID: 23241873

44. Domingo E, Holland JJ. (1994) Mutation rates and rapid evolution of RNA viruses. In: Morse SS, editor. The Evolutionary Biology of Viruses. New York (USA): Raven Press; 1994. pp. 161–184.

45. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science. 1995; 267(5197): 483–489. PMID: 7824947

46. Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, Payan C, Goudeau A, Goudeau A et al. Genetic diversity of the hepatitis C virus: impact and issues in the antiviral therapy. World J Gastroenterol. 2007; 13(17): 2416–2426. PMID: 17552024

47. Nagarapu R, Rahamathulla S,Vishwakarma S 621, Vinjimuri S, Ramachandra Rao VV. A meta-analysis of genotyping methods and nucleotide variation in Hepatitis C Virus. Int J Adv Res. 2013; 1(8): 439–449.

48. Alhamlan FS, Al-Ahdal MN, Khalaf NZ, Abdo AA, Sanai FM, Al-Ashgar HI, ElHefnawi M, Zaid A, Al-Qahtani AA. Genetic variability of the core protein in hepatitis C virus genotype 4 in Saudi Arabian patients and its implication on pegylated interferon and ribavirin therapy. J Transl Med. 2014; 12:91. https://doi.org/10.1186/1479-5876-12-91 PMID: 24708767

49. Kanwal S, Mahmood T. Occurrence of genetic modifications in core, 5′UTR and NS5B of HCV associated with viral response to treatment. Virol J. 2014; 11: 171. https://doi.org/10.1186/1743-422X-11-171 PMID: 25270660

50. Tamalet C, Colson P, Tissot-Dupont H, Henry M, Tourres C, Tivoli N, et al. Genomic and phylogenetic analysis of hepatitis C virus isolates: A survey of 535 strains circulating in Southern France. J Med Virol. 2003; 71(3): 391–398. https://doi.org/10.1002/jmv.10505 PMID: 12966544

51. Bukh J, Miller RH, Purcell RH. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liver Dis. 1995; 15(1): 41–63. https://doi.org/10.1055/s-2007-1007262 PMID: 7597443

52. Di Maio V, Cento V, Mirabelli C, Artese A, Costa G. Alcaro S et al. Hepatitis C Virus Genetic Variability and the Presence of NS5B Resistance-Associated Mutations as Natural Polymorphisms in Selected Genotypes Could Affect the Response to NS5B Inhibitors. Antimicrob Agents Chemother. 2014; 58(5): 2781–2797. https://doi.org/10.1128/AAC.02386-13 PMID: 24590484

53. Svicher V, Cento V, Salpini R, Mercurio F, Fraune M. Beggel B et al. Role of hepatitis B virus genetic barrier in drug-resistance and immune-escape development. Dig Liver Dis. 2011; 43(12): 975–983. https://doi.org/10.1016/j.dld.2011.07.002 PMID: 21831732

54. Gotte M. The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. Curr Opin Virol. 2012; 2(5): 644–650. https://doi.org/10.1016/j.coviro.2012.08.004 PMID: 22964133

55. Maïga AI, Malet I, Soulie C, Derache A, Koita V, Amellal B et al. Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02_AG subtypes. Antivir Ther. 2009; 14(1): 123–129. PMID: 19320246

56. Cento V, Mirabelli C, Salpini R, Dimonte S 644, Artese A, Costa G et al. HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. PLoS One. 2012; 7(7): e39652. https://doi.org/10.1371/journal.pone.0039652 PMID: 22792183

57. Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, Urdea MS et al. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. J Gen Virol. 1994; 75 (Pt 5): 1053–1061.

58. Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, Irvine B et al. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol. 1993; 74 (Pt 11): 2391–2399.

59. Frank C, Mohamed MK, Strickl GT, Lavanchy D, Arthur RR, Magder LS et al. The role of parenteral anti-schistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet. 2000; 355(9207): 887–891. PMID: 10752705

60. Nigel T. Foreign Volunteers of the Allied Forces 1939–45. London: Osprey press; 1991.

61. Abd EL- Wahab E., Mikhaeal A., Sidkey F., Shatat H. Factors associated with hebatitis C infection among chronic HCV Egyptian patients. Iranian J Health. 2014; 43:1510–1518.

62. Ndong-Atome GR, Njouom R, Padilla C, Bisvigou U, Makuwa M, Kazanji M. Absence of interfamilial transmission of hepatitis C virus and low risk for sexual transmission in rural central Africa indicate a cohort effect. J Clin Virol. 2009; 45(4): 349–353. https://doi.org/10.1016/j.jcv.2009.04.017 PMID: 19473878