

# SCIENTIFIC REPORTS



OPEN

## The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia

Received: 14 October 2016  
Accepted: 15 February 2017  
Published: 21 March 2017

Ahmed A. Al-Qahtani<sup>1,2</sup>, Guy Baele<sup>3</sup>, Nisreen Khalaf<sup>1</sup>, Marc A. Suchard<sup>4,5</sup>, Mashael R. Al-Anazi<sup>1</sup>, Ayman A. Abdo<sup>6</sup>, Faisal M. Sanai<sup>7</sup>, Hamad I. Al-Ashgar<sup>8</sup>, Mohammed Q. Khan<sup>8</sup>, Mohammed N. Al-Ahdal<sup>1,2</sup>, Philippe Lemey<sup>3</sup> & Bram Vrancken<sup>3</sup>

The relatedness between viral variants sampled at different locations through time can provide information pertinent to public health that cannot readily be obtained through standard surveillance methods. Here, we use virus genetic data to identify the transmission dynamics that drive the hepatitis C virus subtypes 4a (HCV4a) and 4d (HCV4d) epidemics in Saudi Arabia. We use a comprehensive dataset of newly generated and publicly available sequence data to infer the HCV4a and HCV4d evolutionary histories in a Bayesian statistical framework. We also introduce a novel analytical method for an objective assessment of the migration intensity between locations. We find that international host mobility patterns dominate over within country spread in shaping the Saudi Arabia HCV4a epidemic, while this may be different for the HCV4d epidemic. This indicates that the subtypes 4a and 4d burden can be most effectively reduced by combining the prioritized screening and treatment of Egyptian immigrants with domestic prevention campaigns. Our results highlight that the joint investigation of evolutionary and epidemiological processes can provide valuable public health information, even in the absence of extensive metadata information.

The hepatitis C virus (HCV) imposes a considerable burden on societies and health-care systems throughout the world, and currently infects an estimated 80 (64–103) million people<sup>1</sup>. Despite dramatic improvements in response rates to antiviral therapy following the introduction of highly potent direct-acting antivirals, their global impact on the HCV prevalence is limited<sup>2</sup>. Consequently, a thorough understanding of the dynamics of HCV spread remains important to guide public health efforts.

Because the evolutionary and epidemiological processes of HCV occur on the same time scale, viral genetic data can be used to elucidate transmission patterns using phylodynamic tools<sup>3,4</sup>. When clinical, demographic and/or behavioral metadata is available these tools can be used to identify populations at increased risk (e.g. Jacka, *et al.*<sup>5</sup> for HCV). However, such richly annotated virus genetic data is usually not available, in which case researchers can often fall back on the more widely shared temporal and geospatial annotations to uncover epidemiologically relevant patterns of virus flow<sup>6–8</sup>.

HCV is classified into several genotypes and subtypes that are associated with varying geographical distributions. Some subtypes (e.g. 1a, 1b, 2a and 3a) are found abundantly across the globe, but the spread of others remains more confined to particular geographical areas<sup>1</sup>. An example of the latter is genotype 4, which is endemic in Central Africa from where multiple lineages were introduced into North Africa and the Middle East<sup>9</sup>. Among genotype 4 subtypes, subtype 4a (HCV4a) dominates the HCV epidemic in Egypt, which likely reflects the interplay between a founder effect and past unsafe injection practices during large public health campaigns<sup>10,11</sup>.

<sup>1</sup>Department of Infection and Immunity, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia.

<sup>2</sup>Department of Microbiology and Immunology, Alfaisal University School of Medicine, Riyadh, Saudi Arabia. <sup>3</sup>KU Leuven - University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, B-3000 Leuven, Belgium. <sup>4</sup>Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, USA. <sup>5</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, USA. <sup>6</sup>Section of Gastroenterology, Department of Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia. <sup>7</sup>Gastroenterology Unit, Department of Medicine, King Abdulaziz Medical City, Jeddah, Saudi Arabia. <sup>8</sup>Gastroenterology Unit, Department of Medicine, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia. Correspondence and requests for materials should be addressed to A.A.A.-Q. (email: aqahtani@kfshrc.edu.sa) or B.V. (email: bram.vrancken@kuleuven.be)

	HCV4a	HCV4d
Canada	14 (3)	2 (1)
China	—	1
Cyprus	6	—
DRC	2 (2)	—
Denmark	2	—
Egypt	366 (3)	3 (3)
France	1	2
India	—	1
Japan	1	—
Saudi Arabia	135*	110* (2)
Pakistan	2	—
Portugal	24 (1)	—
South Africa	2 (2)	—
Sri Lanka	2 (1)	—
USA	10	1
Uganda	1 (1)	—
<i>total</i>	568 (13)	120 (6)

**Table 1. Overview of the sequence dataset composition.** The number of sequences per location is given for both subtypes. Numbers between brackets indicate how many sequences of that location did not cluster in the diversity of interest, and were not retained in the final dataset (see main text). The multiple sequence alignments for HCV4a and HCV4d span 576 nt (corresponding to H77 nt positions 914–1490). The HCV4a isolates were collected between 2000 and 2014, and those for HCV4d between 2002 and 2014. \*All Saudi Arabia sequences were newly generated for this study.

Genotype 4 is also ubiquitously present in Saudi Arabia, a neighboring country of Egypt, where it accounts for 60% to 70% of the HCV infections<sup>12,13</sup>. Of these, about 50% can be attributed to HCV4a, and another ~40% are due to subtype 4d (HCV4d)<sup>13</sup>. While HCV4a is omnipresent in Egypt, only few HCV4d variants circulate in this country<sup>12</sup>. This suggests that whereas the HCV4a epidemics in Egypt and Saudi Arabia may overlap, HCV4d may have been introduced from elsewhere, or constitutes a more confined local epidemic.

Despite the substantial genotype 4 burden in Saudi Arabia, its population level patterns of spread have not been investigated to date. Here we make use of newly generated and publicly available virus genetic data to, for the first time, formally characterize the transmission dynamics of the Saudi Arabia HCV4a and HCV4d epidemics, with a focus on the spatial (HCV4a) and temporal (HCV4d) aspects of their transmission history, and frame them in their global context. Further, our results highlight how the patterns of virus flow revealed by statistical phylogenetic and phylogeographic analyses can assist in prioritizing interventions.

## Methods

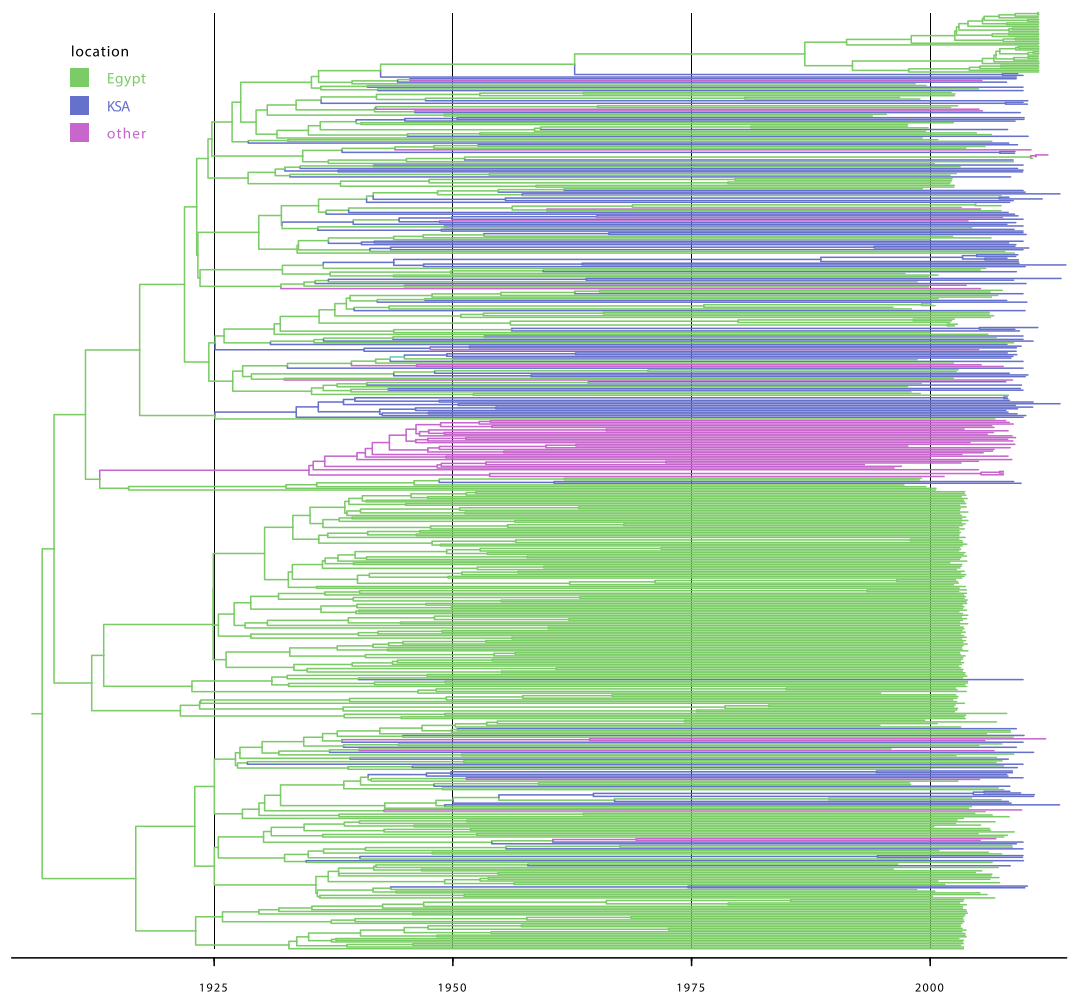
**Study patients.** The study included 245 Saudi Arabian nationals who were chronically infected with HCV as determined by anti-HCV serology and a detectable HCV RNA for more than six months. The patients were recruited between 2007 and 2014 from three major hospitals in Riyadh city including the King Faisal Specialist Hospital & Research Center, the Prince Sultan Military Medical City, and the King Khalid University Hospital. The study was approved by the institutional review board of all participating hospitals (Project # RAC 2090001), and conducted in accordance with the Helsinki Declaration of 1975. Informed consents were obtained from all patients.

**Sequence dataset compilation.** New E1 gene region sequence data (H77 nucleotide positions 914–1490) were complemented with a selection of publicly available sequences to assemble the most comprehensive dataset for investigating the evolutionary history of HCV4a and HCV4d. The new sequence data are available from Genbank under accession numbers KY655493–KY655737. Further methodological details are provided in Supplementary Information.

**Timed phylogenetic inferences.** Phylogenetic trees were estimated using the Bayesian evolutionary analysis by sampling trees (BEAST) software<sup>14</sup>. Because accurate evolutionary rates could not be estimated based on the differences in sampling date of the HCV4a isolates, we capitalized on previously published data with strong temporal signal to calibrate the molecular clock<sup>15</sup> (see Supplementary Information for details). We specified a codon position partitioning model (the SRD06 model<sup>16</sup>) for both subtype data sets. A flexible non-parametric model<sup>17</sup> was fitted to describe demographic changes for HCV4d, but this parameter-rich model could not be informed by the HCV4a data. Because preliminary analysis showed that Egypt is the source of most HCV4a lineages, and Egyptian data dominate the HCV4a dataset (Table 1), we follow Pybus, *et al.*<sup>18</sup> and resorted to a constant-logistic (con-log) growth model for which we needed to specify an informative prior distribution on the  $t_{50}$  parameter. The con-log model assumes a period of exponential growth (starting at the time of the most recent common ancestor ( $t_{MRCA}$ )) that is preceded by and stabilizes into a phase with constant population size. In

location		posterior probability (%)
from	to	
Egypt	Saudi Arabia	79.12
Egypt	other	15.37
Saudi Arabia	Egypt	1.95
Saudi Arabia	other	2.96
other	Egypt	0.46
other	Saudi Arabia	0.15

**Table 2. Posterior probabilities for all possible types of migration events.**

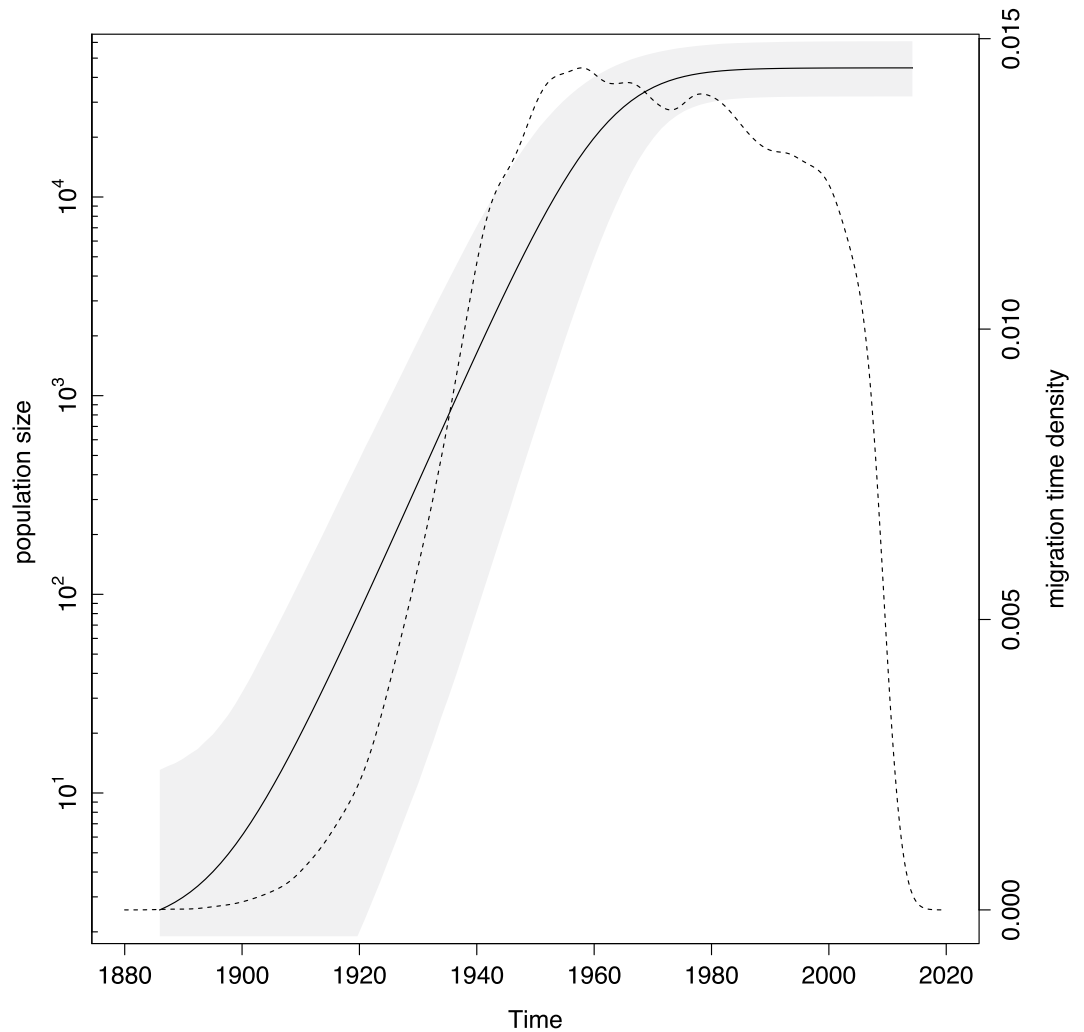


**Figure 1. Bayesian maximum clade credibility tree of the HCV4a dataset.** Tips and internal branches are colored according to the most probable reconstructed ancestral state (location). The correspondence between the colors and locations is as in the legend.

this model  $t_{50}$  represents the point in time where the population size is half the size of the (constant) population size at present. Our prior choice on  $t_{50}$  was informed by the timing of the parenteral anti schistosomiasis (PAT) campaigns in Egypt<sup>10</sup>. Specifically, we centered  $t_{50}$  at 1975 and specified the standard deviation on the normal distribution such that the 2.5 and 97.5 percentiles of the distribution correspond to 1961 to 1989.

We obtained estimates of the posterior distribution using Markov chain Monte Carlo (MCMC) sampling and ran several independent MCMC analyses that were combined after removal of the burn-in and visual inspection of convergence and mixing using Tracer v1.6.0 (<http://tree.bio.ed.ac.uk>).

**Phylogeographic inference.** Prior to extensive phylogeographic reconstructions, we tested whether the data contain signal that can inform phylogeographic model parameter estimations, and assessed which model is most appropriate for describing the patterns of spread (see Supplementary Information for details). Following these results we adopt a model that allows for different rates to and from a particular location<sup>19</sup>. To identify the



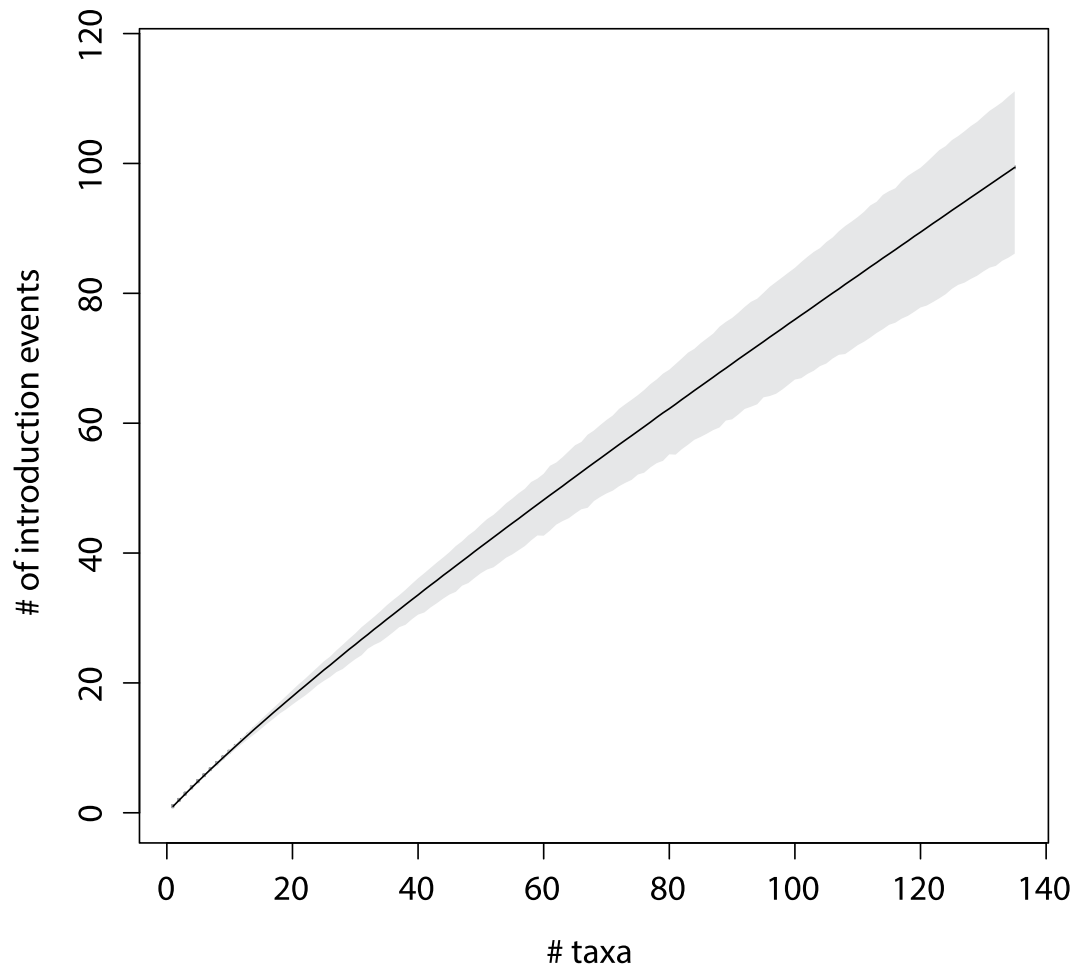
**Figure 2. Combined plot of HCV4a population size growth and migration time density.** The solid black line represents the estimated product of the effective population size and generation time ( $N_e \cdot \tau$ ) trajectory on a  $\log_{10}$  scale (left axis), which reflects the number of individuals that contribute offspring to the descendent generation. The grey-shaded area marks the associated 95% HPD interval. The striped line shows the migration density (right axis). After a steep increase in the number of migrations in the period between 1920 and 1950, the migration intensity remained fairly steady.

subset of migration rates that is most informative to reconstruct the phylogeographic history we combine this model with the Bayesian stochastic search variable selection (BSSVS) procedure and use *Spread3*<sup>20</sup> to calculate Bayes factor (BF) support for all possible types of location exchanges. Estimates of the posterior expected number of migration events between all pairs of locations (Markov jumps) were computed through stochastic mapping techniques<sup>21,22</sup>. All phylogeographic reconstructions were run on empirical tree distributions obtained from the timed phylogenetic inferences<sup>23</sup>.

**Phylogeographic rarefaction curve.** While a visual inspection of a phylogeographic tree may provide indications of the migration intensity, it may provide little insight to what extent this is sensitive to the sampling. It also does not scale well to larger sets of trees such as those now routinely obtained using Bayesian phylogenetics. We tackle this with what we refer to as the phylogeographic rarefaction curve, which plots the expected number of introduction events into a specific location (y-axis) as a function of the number of taxa randomly selected from all taxa sampled at that location (x-axis). For this we implemented a script that uses functions from several R packages<sup>24–26</sup>. The y-axis values for the Saudi Arabia HCV4a epidemic were computed by repeating the procedure we detail in Supplementary Information for 100 random subsamples of size  $n$ , with  $1 \leq n \leq N$  and  $N$  the total number of Saudi Arabian taxa.

## Results

**Sequence datasets.** An overview of the dataset composition is provided in Table 1. The complete sequence datasets of HCV4a and 4d contained 568 and 120 taxa respectively. There were substantially more publicly available sequences from non-Saudi Arabia locations for HCV4a ( $n = 433$ ) than for HCV4d ( $n = 9$ ). These taxa were



**Figure 3. Phylogeographic rarefaction curve.** The solid line represents the expected number of introduction events in Saudi Arabia for  $n$  randomly chosen Saudi Arabian taxa. The grey-shaded area marks the 95% credible interval.

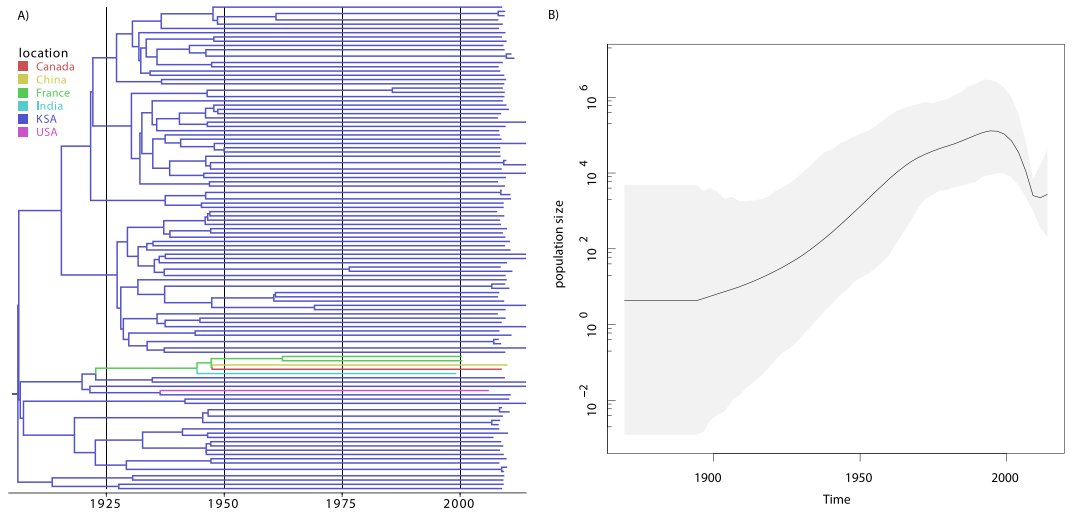
from 13 and 6 countries for the HCV4a and HCV4d dataset respectively. Most of the non-Saudi Arabia HCV4a taxa are from Egypt ( $n = 366$ ), which is not surprising because of a history of explosive spread of this subtype in Egypt<sup>18</sup>.

Initial analyses of both datasets revealed that a small number of taxa clustered as an outgroup separated by long branches. Because this study focused on the more recent phylogeographic events these were removed from the dataset, leaving 555 taxa for HCV4a and 114 taxa for HCV4d. As only a limited number of HCV4a taxa were sampled in countries other than Egypt or Saudi Arabia (10.8%), these were grouped under a single location state ('other') in the phylogeographic analyses.

**Subtype 4a.** Our investigation of the epidemiological relationships of HCV4a between Saudi Arabia, Egypt and the 'other' locations using phylogeographic methods confirms a dominant role for Egypt in the HCV4a spread. Specifically, the posterior inclusion probabilities of the virus migration rates from Egypt to Saudi Arabia and to the 'other' locations were indistinguishable from 1, which is reflected in the very high BF support for these rates of  $\geq 1105$ . None of the other migration rates was well supported (BF support  $< 3$ ). We also quantified the expected number of migration events between all possible pairs of locations (Markov jumps, Table 2). This corroborates what is evident from Fig. 1, namely that Egypt is the most important source location, and most of the movements out of Egypt (83.73%) are to Saudi Arabia. To some extent, migration out of Saudi Arabia also occurs, but this involves only a small fraction of all virus movements. The fact that only few migrations are directed towards Egypt further highlights Egypt as a net exporter in the history of HCV4a.

Although reconstructing the HCV4a demographic history is not the focus of this work, it is reassuring that the estimated increase in population size precedes the migration density through time (Fig. 2).

**Visualizing the transmission network size structure.** We built a phylogeographic rarefaction curve to determine how the Saudi Arabia HCV4a epidemic is shaped by cross-border migration (Fig. 3). Because each introduction event marks the start of a local transmission network, and the probability of detecting a previously unrecognized introduction event hence depends both on the number of local transmission networks in the total



**Figure 4.** Left panel: Bayesian maximum clade credibility tree of the HCV4d dataset. Tips and internal branches are colored according to the most probable reconstructed ancestral state (location). The correspondence between the colors and locations is as in the legend. Right panel: Plot of the demographic history for HCV4d. The black line represents the change in mean effective number of infections through time on a log<sub>10</sub> scale. The surrounding grey shaded area indicates the associated 95% HPD interval. The mean estimate is largest in 1994 and the timings of the maximum values for the upper and lower boundaries of the 95% HPD interval are 1992 and 1997.

sample and their relative abundance, the gradient of the slope of such curves has an epidemiologically relevant and intuitive interpretation. For example, if the sample consists of a few larger clusters complemented with many smaller ones, the curve is expected to have a mild slope because most of the additional isolates represent the already detected larger transmission networks, and vice versa.

The detection of new subpopulations at a high rate (Fig. 3) reveals that the Saudi Arabia HCV4a epidemic consists of many similar-sized small transmission clusters rather than being dominated by one or a few large local circulation networks. This agrees with the observation that only 32/135 Saudi Arabian taxa group in 14 well-supported clusters (posterior root node support  $\geq 0.95$ ) of size 2, 3 or 5 in the maximum clade credibility (MCC) summary tree (Fig. 1). Of these, only 6 transmission pairs likely represent recent transmission events ( $t_{\text{mrc}} \leq 10$  years and maximum within cluster pairwise genetic distance  $\leq 1\%$ ). The fragmented nature of the epidemic is also reflected by the high number of introductions into Saudi Arabia (centered around 99.4, Supplementary Information, Figure S1). Further, the constancy of the slope indicates that the current sampling only reveals a restricted number of introduction events.

**Subtype 4d.** Unlike for HCV4a, the historical epidemic dynamics of HCV4d are not well characterized. We therefore used our Saudi Arabia-focused sample to estimate changes in viral population size through time (Fig. 4). In contrast to the HCV4a dataset there was a sufficient accumulation of evolutionary divergence over the sampling time period to calibrate the molecular clock. The mean evolutionary rate was estimated at  $9.57 \times 10^{-4}$  (95% Highest Posterior Density (HPD) interval:  $6.38\text{--}13.0 \times 10^{-4}$ ), which closely matches the subtype 1a and 1b E1 rate estimate (Supplementary Information, Table S1), and reassures that the latter are appropriate external calibrations to estimate timings of past evolutionary and epidemiological events. The plot of changes in population size through time (Fig. 4, right panel) reveals an approximately exponentially growing population from the beginning of the 20<sup>th</sup> century to about 1967, followed by a period of less rapid growth up to the beginning of the sampling time period.

## Discussion

We made use of statistical phylogenetics tools to investigate the epidemiological history of the dominant HCV subtypes in Saudi Arabia using newly generated virus genetic data. The availability of sufficient HCV4a reference sequences enabled us to directly investigate the patterns of virus flow relating Saudi Arabia to other locations, Egypt in particular. For HCV4d on the other hand, no representative reference data are currently available which is why we focused on the temporal aspects of the transmission dynamics for this subtype.

Our results show that recent HCV4a transmissions between countries likely have an origin in Egypt (Table 2), which is in line with previous findings concerning a central role of this country in the global HCV4a epidemic<sup>27–29</sup>. Virus spread between Saudi Arabia and Egypt is predominantly directed towards Saudi Arabia, showing that both HCV4a epidemics are connected through a source-sink relation, perhaps linked to the large flow of Egyptian migrant workers<sup>13</sup>. The virus is also transmitted from Egypt to countries from geographically diverse locations that were grouped in the ‘other’ category (Table 1). Saudi Arabia plays a limited secondary role in HCV4a exportation and viral flow out of the country is only partly directed to Egypt (Table 2). This can point to limited remigration of Egyptian migrant workers or it can be related to the large Asian and other African communities in

Saudi Arabia and to the large numbers of foreign pilgrims that visit the country<sup>30,31</sup>. Unfortunately, the lack of a comprehensive geographic sampling and the focus on a limited genome region prevented us from directly evaluating this hypothesis and from corroborating previously found links. For example, migrant workers have also been invoked to account for the local westward spread of HCV4a out of Egypt<sup>27</sup>, but our sampling does not include Libyan nor Tunisian isolates. Likewise, Egyptian immigrants have been linked to the spread of HCV4a to Europe<sup>28,29</sup>.

We also found that the large number of introduction events shapes the structure of the Saudi Arabia HCV4a epidemic (Fig. 3), for which we rely on a transmission cluster analysis using phylogenetic trees. Because the sampling density impacts the extent of clustering, this provides a lower boundary on the introduction count. The large number of introduction events into Saudi Arabia translates to many independent local transmission networks. Most of these are single lineages, and of the remainder only six pairs represent well-supported recent transmission links. Combined, these results suggest limited onwards HCV4a transmission within Saudi Arabia and promote the targeted screening and treatment of (former) immigrant workers as an effective means to lessen the HCV4a burden.

HCV sequence datasets often lack a clear temporal signal, making it common practice to calibrate the molecular clock with other empirical evolutionary rate estimates (e.g. this study and refs 9, 18, 32–36). These are usually not from the same genotype or subtype (e.g. this study and refs 9, 18, 42, 33–35), which might raise questions about their suitability. Our finding that the subtype 4d substitution rate - obtained under a tip-date informed model - closely matches that of the frequently used estimates reported by Pybus, *et al.*<sup>35</sup> and Gray, *et al.*<sup>37</sup> (Table S1), offers reassurance that these external calibrations are appropriate to estimate timings of past evolutionary and epidemiological events. Nonetheless, it remains important to keep in mind that substitution models may not perfectly correct for multiple substitutions, and their effect will manifest more rapidly on divergence time estimates based on faster-evolving regions compared to more conserved genomic regions. Because the reported divergence timings are based on the rapidly evolving E1 gene region<sup>37</sup>, they represent lower bounds. More accurate results could of course be obtained by longer genomic markers, preferably complete genome sequences. Of note, the highly similar evolutionary rates result in largely overlapping estimates of the subtype 4a (1905, 95% HPD: 1885–1926) and 4d (1907, 95% HPD: 1872–1943)  $t_{\text{MRCA}}$ . These timings are compatible with a recent origin of genotype 4 (1733, 95% HPD 1650–1805, ref. 9), and show it took only around 170 years for genotype 4 viruses to cross the continent.

Previous phylogenetic analyses have demonstrated that HCV genotype 4 emerged in Central Africa, from where it spread to other regions<sup>9,38</sup>. The current genotype 4 prevalence and geographic distribution likely still reflects to a large extent the early historical diversification and dissemination events, with Central and Eastern sub-Saharan Africa and the Middle East now experiencing the highest burden<sup>1</sup>. The absence of subtype 4d in surveillance studies in West Africa<sup>38–40</sup> and Central Africa<sup>9,41–43</sup> suggests that this subtype may have emerged along a more eastward trajectory. Although subtype-level data from this region are scarce, the detection of HCV4d in 22% of seropositive samples in a small-scale Ethiopian survey<sup>44</sup> supports such a scenario, as do the likely introduction of this subtype in southern Italy by returning colonists, of whom 82% were stationed in the Horn of Africa (Eritrea, Ethiopia and Somalia, the others inhabited Libya)<sup>45</sup>, and its high prevalence in Saudi Arabia<sup>12,13</sup>.

The likely eastward spread of subtype 4d to the Middle East may be linked to human mobility, as is usually the case for virus dispersal<sup>6,32,46,47</sup>. This variant is rarely reported (<1%) in the well-surveyed Egyptian epidemic<sup>13</sup>, but the large population of migrant workers from the Horn of Africa<sup>30,31</sup> may have facilitated HCV4d dispersion in a similar way as for HCV4a from Egypt. Therefore, whereas the Saudi Arabian samples dominate the subtype 4d dataset, the reconstructed HCV4d demographic history (Fig. 4) may at least partly reflect the regional epidemiological dynamics. Although it remains difficult to determine the precise extent of this, the expectation that incidence trends will be reflected in effective population size patterns<sup>48–51</sup>, often allowing demographic trends to be linked to important societal changes<sup>18,35,36,52,53</sup>, may help discriminate between competing hypotheses of the dominant source(s) of transmissions. The absence of consistent screening of blood donors in African countries other than Zimbabwe and South Africa before 2002<sup>54</sup>, and the start of the World Health Organisation's programme to reduce the transmission of blood-borne pathogens through the promotion of safe injection practices (the SIGN initiative) in 1999<sup>55</sup>, which occurred after the start of the decline of the HCV4d effective population size (Fig. 4), indicate that unsafe injection practices and blood transfusions in countries from the Horn of Africa may not explain the observed temporal pattern. On the other hand, the declining subtype 4d epidemic coincides remarkably well with the period of increasing awareness for safe clinical practices and the start of the screening for HCV in blood products in Saudi Arabia<sup>56,57</sup>. Together, these indirect lines of evidence suggest that our subtype 4d sample to a large extent reflects the Saudi Arabian epidemic, which in turn points towards a limited importation of HCV4d viruses through migrant workers from the Horn of Africa. Further, as the timing of the decrease in effective population size does not match with a concomitant increase in substance abuse in Saudi Arabia<sup>56</sup>, the pre-1994 expansion was most likely not heavily impacted by injecting drug use (IDU). Nowadays, good screening practices impede the spread of HCV in Saudi Arabia through infected blood products<sup>56,57</sup>, and sporadic transmission through various modes, including but not limited to IDU and needle-stick injuries<sup>13</sup>, maintain the spread of HCV. We note, however, that our subtype 4d dataset only offers limited insights and additional sampling efforts, in particular in the Horn of Africa (Sudan, Ethiopia, Eritrea and Somalia), are needed to shed light on the spatial dispersal patterns. Additional sampling is also required to evaluate the role of Saudi Arabia in the global HCV4d spatial circulation, and may prove useful in establishing the presence and nature of links with the 4d epidemics in Turkey<sup>58</sup> and with the European 4d epidemic among injecting drug users and HIV-1 co-infected MSM<sup>29,59</sup>.

As noted by Iles, *et al.*<sup>9</sup>, the presence of other subtypes in a region (e.g. subtypes 4n, 4m, 4l, 4r and 4o in Saudi Arabia<sup>13</sup> and 4f, 4l, 4m, 4n and 4o in Egypt<sup>60</sup>) indicates that among all variants introduced, HCV4d was the one that by chance was able to benefit most from changes in human behaviour in Saudi Arabia, as was the case for HCV4a in Egypt. The remarkable difference in absolute scale of the HCV epidemics in both regions, with

prevalences of respectively ~1.5% in Saudi Arabia versus ~15% in Egypt<sup>1</sup>, is likely explained by the PAT campaigns that amplified the Egyptian subtype 4a<sup>10,11</sup> but not the subtype 4d reservoir. While the subtype 4a growth plateaus from the 1980s onwards, the growth period of subtype 4d extends into the 1990s (Figs 2 and 4). Because informative prior distributions were specified for the substitution rate and the t50 parameter of the coalescent model for the subtype 4a dataset, we refrain from speculating on possible causes that can underlie this difference.

In conclusion, our analyses reveal tight links between the Saudi Arabia HCV4a epidemic and the one in Egypt, indicating that intervention and treatment efforts directed against HCV4a should focus on immigrant populations from Egypt. In contrast, the HCV4d epidemic may be more locally confined, which implies that public health efforts targeting domestic modes of transmission may be more efficient in this case. This study is a powerful example of the value for public health planning of the evolutionary analysis of temporally and geospatially referenced virus genetic data, which are becoming increasingly available owing to upscaled genotyping efforts that accompany the roll out of direct acting antivirals based therapies<sup>61</sup>.

## References

- Gower, E., Estes, C., Blach, S., Razavi-Shearer, K. & Razavi, H. Global epidemiology and genotype distribution of the hepatitis C virus infection. *Journal of hepatology* **61**, S45–S57, doi: 10.1016/j.jhep.2014.07.027 (2014).
- Hajarizadeh, B., Grebely, J. & Dore, G. J. Epidemiology and natural history of HCV infection. *Nature reviews. Gastroenterology & hepatology* **10**, 553–562, doi: 10.1038/nrgastro.2013.107 (2013).
- Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* **10**, 540–550, doi: 10.1038/nrg2583 (2009).
- Baele, G. S., Marc A., Rambaut, A. & Lemey, P. Emerging Concepts of Data Integration in Pathogen Phylodynamics. *Systematics in Biology* in press (2016).
- Jacka, B. *et al.* Transmission of hepatitis C virus infection among younger and older people who inject drugs in Vancouver, Canada. *Journal of hepatology*, doi: 10.1016/j.jhep.2016.02.031 (2016).
- Pybus, O. G., Tatem, A. J. & Lemey, P. Virus evolution and transmission in an ever more connected world. *Proc Biol Sci* **282**, doi: 10.1098/rspb.2014.2878 (2015).
- Bloomquist, E. W., Lemey, P. & Suchard, M. A. Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol* **25**, 626–632, doi: 10.1016/j.tree.2010.08.010 (2010).
- Faria, N. R., Suchard, M. A., Rambaut, A. & Lemey, P. Toward a quantitative understanding of viral phylogeography. *Curr Opin Virol* **1**, 423–429, doi: 10.1016/j.coviro.2011.10.003 (2011).
- Iles, J. C. *et al.* Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. *Virology* **464–465**, 233–243, doi: 10.1016/j.virol.2014.07.006 (2014).
- Frank, C. *et al.* The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**, 887–891 (2000).
- Strickland, G. T. Liver disease in Egypt: hepatitis C superseded schistosomiasis as a result of iatrogenic and biological factors. *Hepatology* **43**, 915–922, doi: 10.1002/hep.21173 (2006).
- Kamal, S. M. & Nasser, I. A. Hepatitis C genotype 4: What we know and what we don't yet know. *Hepatology* **47**, 1371–1383, doi: 10.1002/hep.22127 (2008).
- Al Ashgar, H. I. *et al.* Hepatitis C genotype 4: genotypic diversity, epidemiological profile, and clinical relevance of subtypes in Saudi Arabia. *Saudi journal of gastroenterology: official journal of the Saudi Gastroenterology Association* **19**, 28–33, doi: 10.4103/1319-3767.105920 (2013).
- Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973 (2012).
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88 (2006).
- Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23**, 7–9 (2006).
- Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* **25**, 1459–1471, doi: 10.1093/molbev/msn090 (2008).
- Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**, 381–387 (2003).
- Edwards, C. J. *et al.* Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current biology: CB* **21**, 1251–1258, doi: 10.1016/j.cub.2011.05.058 (2011).
- Bielejec, F. *et al.* Spred3: interactive visualisation of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol*, doi: 10.1093/molbev/msw082 (2016).
- Minin, V. N. & Suchard, M. A. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci* **363**, 3985–3995, doi: 10.1098/rstb.2008.0176 (2008).
- Minin, V. N. & Suchard, M. A. Counting labeled transitions in continuous-time Markov models of evolution. *Journal of mathematical biology* **56**, 391–412, doi: 10.1007/s00285-007-0120-8 (2008).
- Lemey, P. *et al.* Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog* **10**, e1003932, doi: 10.1371/journal.ppat.1003932 (2014).
- Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131, doi: 10.1093/bioinformatics/btm538 (2008).
- Jombart, T. *et al.* OutbreakTools: a new platform for disease outbreak analysis using the R software. *Epidemics* **7**, 28–34, doi: 10.1016/j.epidem.2014.04.003 (2014).
- Fadlalla, F. A., Mohamoud, Y. A., Mumtaz, G. R. & Abu-Raddad, L. J. The epidemiology of hepatitis C virus in the Maghreb region: systematic review and meta-analyses. *PLoS One* **10**, e0121873, doi: 10.1371/journal.pone.0121873 (2015).
- Eriksen, M. B. *et al.* Molecular and epidemiological profiles of hepatitis C virus genotype 4 in Denmark. *Journal of medical virology* **82**, 1869–1877, doi: 10.1002/jmv.21896 (2010).
- de Bruijne, J. *et al.* Emergence of hepatitis C virus genotype 4: phylogenetic analysis reveals three distinct epidemiological profiles. *J Clin Microbiol* **47**, 3832–3838, doi: 10.1128/JCM.01146-09 (2009).
- Bel-Air, F. o. D. *Demography, Migration and Labour Market in Saudi Arabia, Explanatory Note No. 1/2014, Gulf Labour Market and Migration (GLMM) programme (GLMM) programme of the Migration Policy Center (MPC) and the Gulf Research Center (GRC)*. 1–15 (2014).
- Kapiszewski, A. *Arab versus asian migrant workers in the gcc countries.* (United nations expert group meeting on international migration and development in the Arab region, 2006).



32. Markov, P. V. *et al.* Colonial history and contemporary transmission shape the genetic diversity of hepatitis C virus genotype 2 in Amsterdam. *J Virol* **86**, 7677–7687, doi: 10.1128/JVI.06910-11 (2012).
33. Markov, P. V. *et al.* Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *The Journal of general virology* **90**, 2086–2096, doi: 10.1099/vir.0.011569-0 (2009).
34. Pybus, O. G. *et al.* Genetic history of hepatitis C virus in East Asia. *J Virol* **83**, 1071–1082, doi: 10.1128/JVI.01501-08 (2009).
35. Pybus, O. G. *et al.* The epidemic behavior of the hepatitis C virus. *Science* **292**, 2323–2325, doi: 10.1126/science.1058321 (2001).
36. Nakano, T., Lu, L., Liu, P. & Pybus, O. G. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J Infect Dis* **190**, 1098–1108, doi: 10.1086/422606 (2004).
37. Gray, R. R. *et al.* The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol* **11**, 131, doi: 10.1186/1471-2148-11-131 (2011).
38. Ndjomou, J., Pybus, O. G. & Matz, B. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *The Journal of general virology* **84**, 2333–2341, doi: 10.1099/vir.0.19240-0 (2003).
39. Cantaloube, J. F. *et al.* Analysis of hepatitis C virus strains circulating in Republic of the Congo. *Journal of medical virology* **82**, 562–567, doi: 10.1002/jmv.21724 (2010).
40. Pasquier, C. *et al.* Distribution and heterogeneity of hepatitis C genotypes in hepatitis patients in Cameroon. *Journal of medical virology* **77**, 390–398, doi: 10.1002/jmv.20468 (2005).
41. Ndong-Atome, G. R. *et al.* Hepatitis C virus prevalence and genetic diversity among pregnant women in Gabon, central Africa. *BMC infectious diseases* **8**, 82, doi: 10.1186/1471-2334-8-82 (2008).
42. Njouom, R. *et al.* Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. *The Journal of general virology* **90**, 2452–2456, doi: 10.1099/vir.0.011981-0 (2009).
43. Njouom, R. *et al.* Phylogeography, risk factors and genetic history of hepatitis C virus in Gabon, central Africa. *PLoS One* **7**, e42002, doi: 10.1371/journal.pone.0042002 (2012).
44. Abreha, T. *et al.* Genotypes and viral load of hepatitis C virus among persons attending a voluntary counseling and testing center in Ethiopia. *Journal of medical virology* **83**, 776–782, doi: 10.1002/jmv.21788 (2011).
45. Ciccozzi, M. *et al.* Hepatitis C virus genotype 4d in Southern Italy: reconstruction of its origin and spread by a phylodynamic analysis. *Journal of medical virology* **84**, 1613–1619, doi: 10.1002/jmv.23384 (2012).
46. Bryant, J. E., Holmes, E. C. & Barrett, A. D. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* **3**, e75, doi: 10.1371/journal.ppat.0030075 (2007).
47. Faria, N. R. *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61, doi: 10.1126/science.1256739 (2014).
48. Faria, N. R. *et al.* Phylodynamics of the HIV-1 CRF02\_AG clade in Cameroon. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 453–460, doi: 10.1016/j.meegid.2011.04.028 (2012).
49. Gill, M. S., Lemey, P., Bennett, S. N., Biek, R. & Suchard, M. A. Understanding Past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates. *Syst Biol* **65**, 1041–1056, doi: 10.1093/sysbio/syw050 (2016).
50. Frost, S. D. & Volz, E. M. Viral phylodynamics and the search for an ‘effective number of infections’. *Philos Trans R Soc Lond B Biol Sci* **365**, 1879–1890, doi: 10.1098/rstb.2010.0060 (2010).
51. Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J. & Frost, S. D. Phylodynamics of infectious disease epidemics. *Genetics* **183**, 1421–1430, doi: 10.1534/genetics.109.106021 (2009).
52. Joy, J. B. *et al.* The spread of hepatitis C virus genotype 1a in North America: a retrospective phylogenetic study. *Lancet Infect Dis* **16**, 698–702, doi: 10.1016/S1473-3099(16)00124-9 (2016).
53. Magiorkinis, G. *et al.* The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med* **6**, e1000198, doi: 10.1371/journal.pmed.1000198 (2009).
54. Madhava, V., Burgess, C. & Drucker, E. Epidemiology of chronic hepatitis C virus infection in sub-Saharan Africa. *Lancet Infect Dis* **2**, 293–302 (2002).
55. Luby, S. *Safe Injection Global Network (SIGN) - Initial Meeting Report.* (1999).
56. Madani, T. A. Hepatitis C virus infections reported over 11 years of surveillance in Saudi Arabia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 132–136, doi: 10.1016/j.trstmh.2008.08.001 (2009).
57. Tamimi, W. *et al.* Expansion of Saudi blood donor pool by better screening and vaccination practices. *Clinical and diagnostic laboratory immunology* **10**, 1159–1160 (2003).
58. Ciccozzi, M. *et al.* Phylogenetic analysis of HCV-4d in Turkey: the curious case of Kayseri province. *Journal of medical virology* **86**, 454–460, doi: 10.1002/jmv.23787 (2014).
59. van Asten, L. *et al.* Spread of hepatitis C virus among European injection drug users infected with HIV: a phylogenetic analysis. *J Infect Dis* **189**, 292–302, doi: 10.1086/380821 (2004).
60. Elkady, A. *et al.* Genetic variability of hepatitis C virus in South Egypt and its possible clinical implication. *Journal of medical virology* **81**, 1015–1023, doi: 10.1002/jmv.21492 (2009).
61. Cuypers, L. *et al.* Mapping the genomic diversity of HCV subtypes 1a and 1b: Implications of structural and immunological constraints for vaccine and drug development. *Virus evolution* **2**, vew024, doi: 10.1093/ve/vew024 (2016).

## Acknowledgements

This work was supported by the Bijzonder Onderzoeksfonds KU Leuven (BOF) No. OT/14/115, the National Science Foundation grant DMS 1264153, by the National Institutes of Health grant R01 AI117011 and by a grant from King Abdulaziz City for Science and Technology [ARP-30-38]. The Research Advisory Council (RAC) of King Faisal Specialist Hospital and Research Centre (KFSHRC) approved this work (RAC # 2090001). The support of the Research Center administration at King Faisal Specialist Hospital and Research Center is highly appreciated. The VIROGENESIS project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 634650.

## Author Contributions

Designed the experiments: A.A.A.-Q., H.I.A.-A., M.Q.K., M.N.A.-A. Performed the experiments: A.A.A.-Q., N.K., M.R.A.-A., A.A.A., F.M.S., H.I.A.-A., M.Q.K., M.N.A.-A. Analyzed the data: A.A.A.-Q., B.V., G.B., P.L., M.A.S. Contributed reagents/materials/analysis tools: A.A.A.-Q., N.K., M.R.A.-A., A.A.A., F.M.S., H.I.A.-A., M.Q.K., M.N.A.-A. Wrote the paper: B.V., M.A.S., A.A.A.-Q., G.B., P.L.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Al-Qahtani, A. A. *et al.* The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia. *Sci. Rep.* 7, 44947; doi: 10.1038/srep44947 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017