Hepatitis C virus genotype 1 and 2 recombinant genomes and the phylogeographic history of the 2k/1b lineage

Reilly Hostager,[1] Manon Ragonnet-Cronin,[1,†] Ben Murrell,[1,‡]
Charlotte Hedskog,[2] Anu Osinusi,[2] Simone Susser,[3] Christoph Sarrazin,[2,4]
Evguenia Svarovskaia,[2] and Joel O. Wertheim[1,*]

[1]Department of Medicine, University of California, San Diego, CA, USA, [2]Gilead Sciences, Foster City, CA, USA,
[3]Goethe-University Hospital, Medical Clinic, Frankfurt, Germany and [4]St. Josefs-Hospital, Medical Clinic 2,
Wiesbaden, Germany

[†]Present address: Department of Infectious Disease Epidemiology, Imperial College London, London, UK
[‡]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden.
[*]Corresponding author: E-mail: jwertheim@ucsd.edu

Abstract

Recombination is an important driver of genetic diversity, though it is relatively uncommon in hepatitis C virus (HCV).
Recent investigation of sequence data acquired from HCV clinical trials produced twenty-one full-genome recombinant
viruses belonging to three putative inter-subtype forms 2b/1a, 2b/1b, and 2k/1b. The 2k/1b chimera is the only known HCV
circulating recombinant form (CRF), provoking interest in its genetic structure and origin. Discovered in Russia in 1999, 2k/
1b cases have since been detected throughout the former Soviet Union, Western Europe, and North America. Although 2k/
1b prevalence is highest in the Caucasus mountain region (i.e., Armenia, Azerbaijan, and Georgia), the origin and migration
patterns of CRF 2k/1b have remained obscure due to a paucity of available sequences. We assembled an alignment which
spans the entire coding region of the HCV genome containing all available 2k/1b sequences ($>$500 nucleotides; $n = 109$)
sampled in ninteen countries from public databases (102 individuals), additional newly sequenced genomic regions (from
48 of these 102 individuals), unpublished isolates with newly sequenced regions (5 additional individuals), and novel com-
plete genomes (2 additional individuals) generated in this study. Analysis of this expanded dataset reconfirmed the mono-
phyletic origin of 2k/1b with a recombination breakpoint at position 3,187 (95% confidence interval: 3,172–3,202; HCV GT1a
reference strain H77). Phylogeography is a valuable tool used to reveal viral migration dynamics. Inference of the timed his-
tory of spread in a Bayesian framework identified Russia as the ancestral source of the CRF 2k/1b clade. Further, we found
evidence for migration routes leading out of Russia to other former Soviet Republics or countries under the Soviet sphere of
influence. These findings suggest an interplay between geopolitics and the historical spread of CRF 2k/1b.

Key words: phylogeography; circulating recombinant form; hepatitis C virus.

1. Introduction

Evolutionary investigations of pathogenic viruses such as human immunodeficiency virus (HIV), Ebola virus, and hepatitis c virus (HCV) via coalescent methods are an important tool in uncovering historical transmission patterns (Gilbert et al. 2007; Veras et al. 2011; Moureau et al. 2015; Al-Qahtani et al. 2017; Dellicour et al. 2018). HCV is an enveloped, single-stranded positive-sense RNA virus belonging to the *Hepacivirus* genus of the *Flaviviridae* family. The virus was first acknowledged as a non-A/non-B hepatitis form in 1975 during a transfusion study (Feinstone et al. 1975), and the first genome sequence, encoding a single 9.6 kb polyprotein, was published in 1989 (Choo et al. 1989). Global anti-HCV seroprevalence is estimated at 2.8 per cent, affecting more than 185 million people between 1990 and 2005 (Mohd Hanafiah et al. 2013; Petruzziello et al. 2016), though this value is likely an underestimate (Kauhl et al. 2015; Webster et al. 2015; Perez et al. 2019).

HCV is a blood-borne pathogen, primarily transmitted via people who inject drugs (PWID) and unscreened blood products administered during transfusions (Lauer and Walker 2001). The virus may be spontaneously cleared during the acute infection phase, though most cases progress to the chronic phase, where the majority of the disease burden lies (Chen and Morgan 2006). Nonetheless, both prognoses are treatable and curable by pharmacologic therapies (U.S. Food and Drug Administration 2017; Jaeckel et al. 2001; Webster, et al. 2015). Unlike hepatitis viruses A and B, no HCV vaccine is available, partially due to high variability between strains and a rapid mutation rate which varies considerably across the genome (Stumpf and Pybus 2002).

HCV has been classified into eight genotypes and eighty-six distinct subtypes (Simmonds 2004; Smith et al. 2014; Borgia et al. 2018). Improper classification of HCV genotypes and recombinants may result in suboptimal treatment regimens (Paolucci et al. 2017; Susser et al. 2017) or direct-acting antiviral therapy failure and relapse (Cuypers et al. 2016). Most studies of HCV variability are based on analyses of single sub-genomic regions, such as 'Heads or Tails' genotyping of Core and NS5B. Using this approach, intra-subtype recombinants will go undetected.

Although HCV has a high mutation rate, recombination is rare; recombinants seldom occur *in vivo* and are often nonviable *in vitro* (Giannini et al. 1999; Viazov et al. 2000). Of all published HCV sequences, only eight intra-genotype forms (1a/1b, 1a/1c, 1b/1a, 4d/4a, 6a/6o, 6e/6o, 6e/6h, and 6n/6o) and nine inter-genotype forms (2a/1a, 2b/1a, 2b/1b, 2b/6w, 2i/6p, 2k/1b, 2/5, 3a/1a, and 3a/1b) have ever been characterized (Kalinina et al. 2002; Colina et al. 2004; Cristina and Colina 2006; Kageyama et al. 2006; Noppornpanth et al. 2006; Legrand-Abravanel et al. 2007; Moreno et al. 2009; Lee et al. 2010; Bhattacharya et al. 2011; Calado et al. 2011; Yokoyama et al. 2011; Raghwani et al. 2012; Shi et al. 2012; Hedskog et al. 2015b; Gaspareto et al. 2016; Morel et al. 2016; Gupta et al. 2017; Kurata et al. 2018). Further, studies which have actively searched for evidence of recombination in large-scale datasets (Magiorkinis et al. 2007) and high-risk populations (Viazov et al. 2010) have consistently failed to detect recombinant HCV. Some of these recombinant forms have been detected in multiple individuals (e.g., 2b/1a); however, only the HCV recombinant 2k/1b is currently thought to represent a circulating recombinant form (CRF) in which sustained transmission of the same viral strain can be traced back via phylogenetic inference to a single homologous recombination event (Kalinina et al. 2002; Raghwani et al. 2012).

The 2k/1b strain was first detected within a cohort of injection drug users in St Petersburg, Russia in 1999 (Alter 1999), although it was retrospectively identified in an Estonian patient sample from 1998 (Tallo et al. 2007). The 2k/1b CRF is often detected in countries that were formerly part of the Soviet Union, typically with relatively low prevalence: Russia (2 per cent), Uzbekistan (1 per cent), Estonia (<1 per cent) (Tallo et al. 2007; Kurbanov et al. 2008). The highest prevalence of 2k/1b is observed in countries in the Caucasus mountain region (i.e., Armenia, Azerbaijan, and Georgia), particularly in Georgia where it is associated with 20 per cent of HCV cases (Zakalashvili et al. 2018). The evolutionary history of 2k/1b was first described by Raghwani et al. (2012), who performed a joint hierarchical analysis of the gene segments Core/E1 (of 2k origin) and NS5B (of 1b origin) from twenty-seven individuals in the same phylogeny as the corresponding pure 2k and 1b subtypes. They inferred a single recombinant origin event for 2k/1b with a time of most recent common ancestor (TMRCA) around 1946. However, this phylogeographic analysis was constrained by limited sampling, as all individuals in their study likely became infected with HCV in one of four former Soviet countries: Azerbaijan, Uzbekistan, Russia, and Georgia. Since this study, sequences have been now described from 109 people in sixteen sampling countries, many from outside the former Soviet Union, including France (Ramiere et al. 2014), the United States, Spain, and the Netherlands (Hedskog et al. 2015b).

Numerous instances of infection with recombinant HCV were identified over the course of Phase II/III Sofosbuvir/Ribavirin clinical trials conducting by Gilead Sciences across clinical sites in North America, Europe, Asia, and Africa, where trial participants received HCV genotyping following seropositive HCV diagnosis. Resulting investigations produced full-genome sequences for twenty-one HCV recombinants: 2b/1a ($n = 12$), 2b/1b ($n = 3$), and 2k/1b ($n = 6$). Here, we investigate the evolutionary history of these recombinant viruses by phylogenetic comparison to all publicly available complete HCV subtype genomes (i.e., 1a, 1b, 2b, and 2k). Our findings corroborate the single-origin hypothesis for 2k/1b and provide robust evidence for the diversification of CRF 2k/1b in the former Soviet Union prior to the collapse in 1991. Further, aside from CRF 2k/1b, we do not find evidence for circulation of other HCV recombinant viruses.

2. Materials and Methods

2.1 HCV full coding region recombinant data acquisition

Twenty-one recombinant viruses were detected across 12,615 patients enrolled in sixty-seven clinical trials run by Gilead Sciences (Foster City, CA, USA) worldwide. Study protocols followed the ethical guidelines set in place by the 1975 Declaration of Helsinki and were approved by the relevant institutional review board committees. Baseline samples collected from treatment naïve and experienced patients were genotyped with the Siemens VERSANT HCV Genotype INNO-LiPA 2.0 Assay (Innogenetics, Ghent, Belgium) and nonstructural (NS)5B sequencing. Isolates with inconsistent genotype assignments, deemed recombinant HCVs, belonged to three inter-subtype forms 2b/1a, 2b/1b, and 2k/1b following examination with a sequence-independent amplification approach and Illumina MiSeq sequencing of paired-end indexed libraries (Hedskog et al. 2015a). For further genotyping and sequencing procedures for these genomes, see Welzel et al. (2017) and Hedskog et al. (2015b). We supplemented these data with an additional fifteen

full-genome HCV recombinants available on GenBank and the Los Alamos National Laboratories (LANL) HCV sequence database (2b/1a, $n = 1$; 2b/1b, $n = 4$; 2k/1b, $n = 10$) (Kuiken et al. 2005). To assist in recombination breakpoint determination, all published full-genomes of the pure parental subtypes 1a ($n = 1,049$), 1b ($n = 463$), 2b ($n = 116$), and 2k ($n = 3$) were also downloaded from these databases. GenBank accession numbers for novel and previously published HCV genomes are included in Supplementary Table S1.

2.2 Breakpoint analysis

To infer recombination breakpoints, we aligned full-coding region sequences from 2b/1a, 2b/1b, and 2k/1b to their respective pure subtype genomes using the fast option in MAFFT version 7 in AliView (Katoh and Standley 2013; Larsson 2014). The Recombination Detection Program (RDP) v4.94 was run, applying six algorithms to each alignment: Geneconv, Bootscan, MaxChi, Chimaera, SiScan, and 3Seq (Martin et al. 2015). Each recombinant genome was scanned independently as a query sequence compared with pure subtype reference sequences; we also combined all 2k/1b full-genomes as a single query. Breakpoint positions and 95% confidence intervals were identified in relation to the subtype 1a H77 reference sequence (GenBank Accession Number NC_004102) in the centrally located NS2 and NS3 genes. Maximum likelihood phylogenies were constructed in IQ-TREE (Nguyen et al. 2015) using a General Time-Reversible substitution model with site-to-site rate variation ($GTR + \Gamma_4$) to examine the phylogenetic relationships between the non-recombinant parental genomes and the recombinant hemigenomes: the genomic region corresponding to the pure subtype, split at the RDP-inferred breakpoint. Branch support was measured with 1,000 replicates of ultrafast bootstrap approximation and Shimodaira–Hasegawa-like approximate likelihood ratio test (Anisimova et al. 2011; Hoang et al. 2018).

2.3 CRF 2k/1b genomic datasets

We assembled a dataset representing previously and newly published 2k/1b sequences from 109 patient samples spanning the entire coding region of the HCV genome. First, all 2k/1b sequences available on GenBank ($n = 102$ individuals) were downloaded. Core/E1 and NS5B sequences from the twenty-seven samples analyzed by Raghwani were included (2012), thirteen of which were extended to include the fully published sequence. To preserve the integrity of our inference, we restricted the sequences from GenBank to be at least 500 nucleotides across the genome and to have accompanying geographic data. Further, we included additional Sanger sequences from the Core, NS2, NS5A, and NS5B genomic regions for forty-eight of these individuals whose NS3 genomic regions were published (Susser et al. 2017). Finally, two previously unpublished full-genomes from the Gilead clinical trials and an additional previously unpublished Italian isolate with newly sequenced NS3, NS5A, and NS5B regions described in (Paolucci et al. 2017) were incorporated. The 2k/1b sequences were sampled in sixteen countries from people tracing back to nineteen countries of origin: Azerbaijan, Uzbekistan, Russia, Georgia, Spain, United Kingdom, Belgium, the Netherlands, Germany, Austria, Romania, Greece, Belarus, Ukraine, Armenia, Chechnya (treated as a distinct region in our analysis), Tajikistan, Kazakhstan, and the United States (see Supplementary Table S2).

Two alignments were created with these 2k/1b sequences (Fig. 1). The first alignment comprised the full coding region.
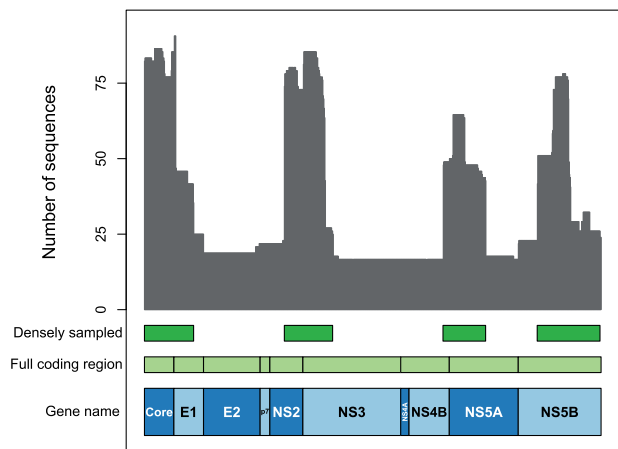
Figure 1. Density of sequence data available per gene segment in HCV genome. The full coding region of the HCV genome was divided into nine partitions: one for each gene, with NS4A and NS4B combined into a single partition. The densely sampled regions are sections of the genome which contain at least 25 sequences of the 109 included, totaling four partitions.

The second alignment comprised a concatenation of only the most densely sampled regions (i.e., ≥ 25 sequences in that region) in order to assess the robustness of our phylogeographic results. This concatenated alignment contained the four densely sampled genomic regions: Core/E1, NS2/NS3, NS4B/NS5A, and NS5B. Sequence alignments were performed using MUSCLE v2.0 in AliView and edited manually (Edgar 2004; Larsson 2014). GenBank accession numbers for newly reported 2k/1b sequences are MK527318–MK527509. Alignments are available as Supplementary Material.

2.4 Bayesian molecular dating and sensitivity analysis

Markov chain Monte Carlo (MCMC) coalescent analyses were separately conducted on the full-coding region and densely sampled region alignments in BEAST v1.10.4, using the BEAGLE library for enhanced performance (Suchard et al. 2018; Ayres et al. 2019). For all runs, a $GTR + \Gamma_4$ nucleotide substitution model with empirical base frequencies was selected as well as an uncorrelated relaxed lognormal (URL) clock (Drummond et al. 2006). We calibrated the molecular clock based explicitly on the previously inferred 2k/1b TMRCA by placing a normal prior (mean = 1946.0; standard deviation = 0.5) on the twenty-seven taxa included in the Raghwani analysis (2012). We did not enforce a monophyly constraint on these taxa, which allowed us to assess whether the newly incorporated viral sequences descended from viruses that predate the previously inferred TMRCA. MCMC analyses were performed in duplicate with a length of 100 million, sampling every 10,000 generations, yielding 10,000 posterior trees. Convergence, demonstrated by effective sample sizes >200, was assessed in Tracer v1.7 (Rambaut et al. 2018), specifying the burn-in as the first 10 per cent of the total number of states. Maximum clade credibility (MCC) trees were generated using TreeAnnotator (Suchard et al. 2018).

We performed non-parametric Skyride coalescent model analyses. The Skyride model, which employs temporal smoothing via a Gaussian Markov random field prior (Minin et al. 2008), is a development of the a Skyline model, which approximates population dynamics with a user-selected volume of change-point bins (Drummond et al. 2005). Sampling year for ten sequences without published collection dates were estimated using random walk operators and uniformly distributed priors

(Shapiro et al. 2011) with lower and upper bounds set to the range between the first detection of the CRF_2k/1b strain in 1999 and the corresponding paper's publication year (identified in Supplementary Table S2). The Skyride model did not converge using these operators, yielding a negative branch length error. Therefore, we used a Skyline model to estimate sampling dates to fix in the Skyride model. The Skyline model was used in the original 2k/1b study (Raghwani et al. 2012). For our Skyline analysis, we designated twenty-one bins for our analysis to satisfy roughly five coalescent events per bin. The final Skyride analyses (detailed below) were completed using sampling dates for these ten samples with unknown collection years fixed to the mean date estimates from the Skyline analysis.

Using this Skyride coalescent framework, we explored various partitioning schemes involving the $GTR + \Gamma_4$ nucleotide substitution model and URL molecular clock in the full coding region (nine partitions) and densely sampled region (four partitions) (Fig. 1). For the full coding region analyses, these partitions were designated based on gene regions, except for the short NS4A gene which was combined with NS4B. For the densely sampled region analysis, these partitions were based on the sampled regions. We explored partitioning schemes in which (i) each partition shared a single nucleotide substitution model and URL molecular clock model, (ii) each partition received its own nucleotide substitution model but shared a single URL molecular clock model, and (iii) each partition received its own nucleotide substitution model and its own URL molecular clock model.

2.5 Phylogeography

We sampled 9,000 trees from the posterior distribution of time-resolved phylogenies (20,000 trees less 10 per cent burn-in from each MCMC, sampled at half frequency) from each partition strategy in an ancestral state reconstruction analysis in BEAST to infer viral lineage movement between discrete locations (Lemey et al. 2009). We employed a strict clock in a continuous-time Markov chain with asymmetric variation between countries and a Bayesian Stochastic Search Variable Selection (BSSVS) procedure to parsimoniously explore all potential migration pathways (Lemey et al. 2009; Edwards et al. 2011). Each model was run for 10 million generations, with sample outputs produced every 10,000 steps, less 10 per cent burn-in. Bayes Factors (Kass and Raftery 1995) for nonzero rates were calculated according to Lemey et al. (2009). Timing of country-to-country movement was assessed by post-processing of Markov Jumps (Minin et al. 2008).

We also performed a sensitivity analysis to examine the hypothesis that 2k/1b originated in the Caucasus mountain region, the region with highest 2k/1b prevalence. This analysis treated the location of taxa from the Caucasus mountain countries (i.e., Azerbaijan, Armenia, and Georgia) as a single geographic location. Combining these locations increases the probability that the 2k/1b ancestor is from that region by increasing the observation frequency.

3. Results

3.1 Recombination breakpoint determination

The inferred position of the recombination breakpoints for 2b/1a ($n = 13$), 2b/1b ($n = 7$), and 2k/1b ($n = 16$) full-genomes all fell around the NS2/NS3 junction (Supplementary Table S1; Fig. 2).
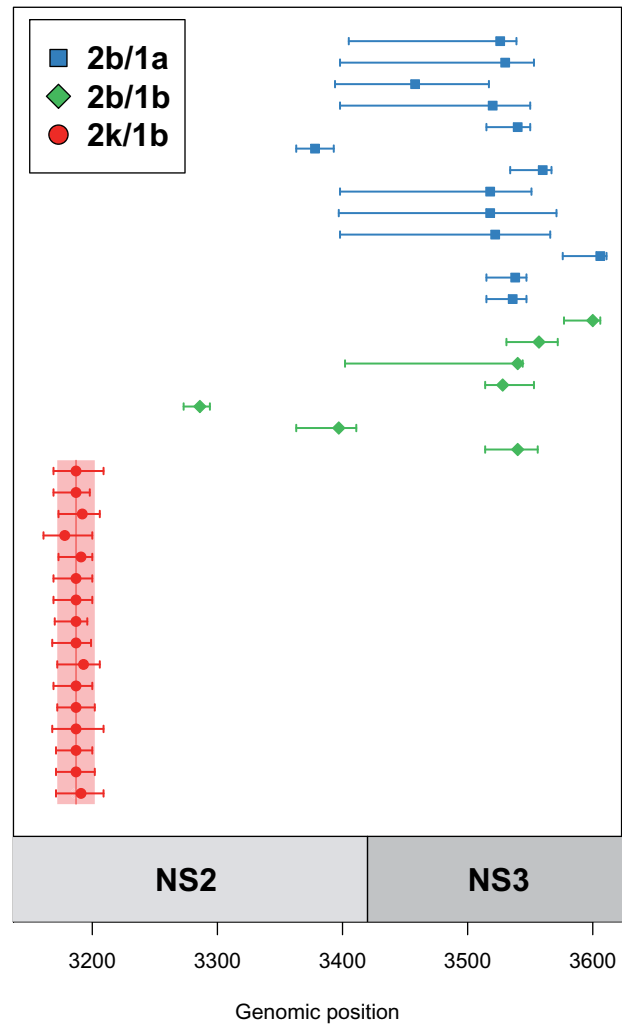

Figure 2. Inferred recombinant breakpoints for 2b/1a, 2b/1b, and 2k/1b full-genomes. Breakpoint estimates are shown with 95% confidence intervals. For 2k/1b, the results of the combined genomic analysis are depicted as a shaded red box (95% confidence intervals), with the breakpoint point estimate as a solid vertical line. HCV recombinant viruses follow the naming convention: 5′ genomic segment/3′ genomic segment. Genomic positions are shown relative to the genotype 1a H77 reference strain.

However, 2b/1a and 2b/1b breakpoints were usually distinct, and some were up to 200–300 nucleotides apart with non-overlapping 95% confidence intervals, suggesting independent recombinant origins. In contrast, the inferred positions of the 2k/1b breakpoints were consistent with each other and with the joint recombination breakpoint estimate, indicating a single origin event for 2k/1b. This joint estimate for the 2k/1b recombination breakpoint was at position 3,187 of the reference strain (95% confidence interval: 3,172–3,202), in accordance with previous reports (Kalinina et al. 2002; Hedskog et al. 2015b; Susser et al. 2017).

3.2 Recombinant phylogenetics

We inferred maximum likelihood phylogenetic trees using sequences from the pure subtypes 1a, 1b, and 2b, and the corresponding recombinant hemigenomes, separated at the RDP-inferred breakpoint (Fig. 3). Of the three recombinant types, only 2k/1b formed a clade in the phylogeny (Fig. 3B). This
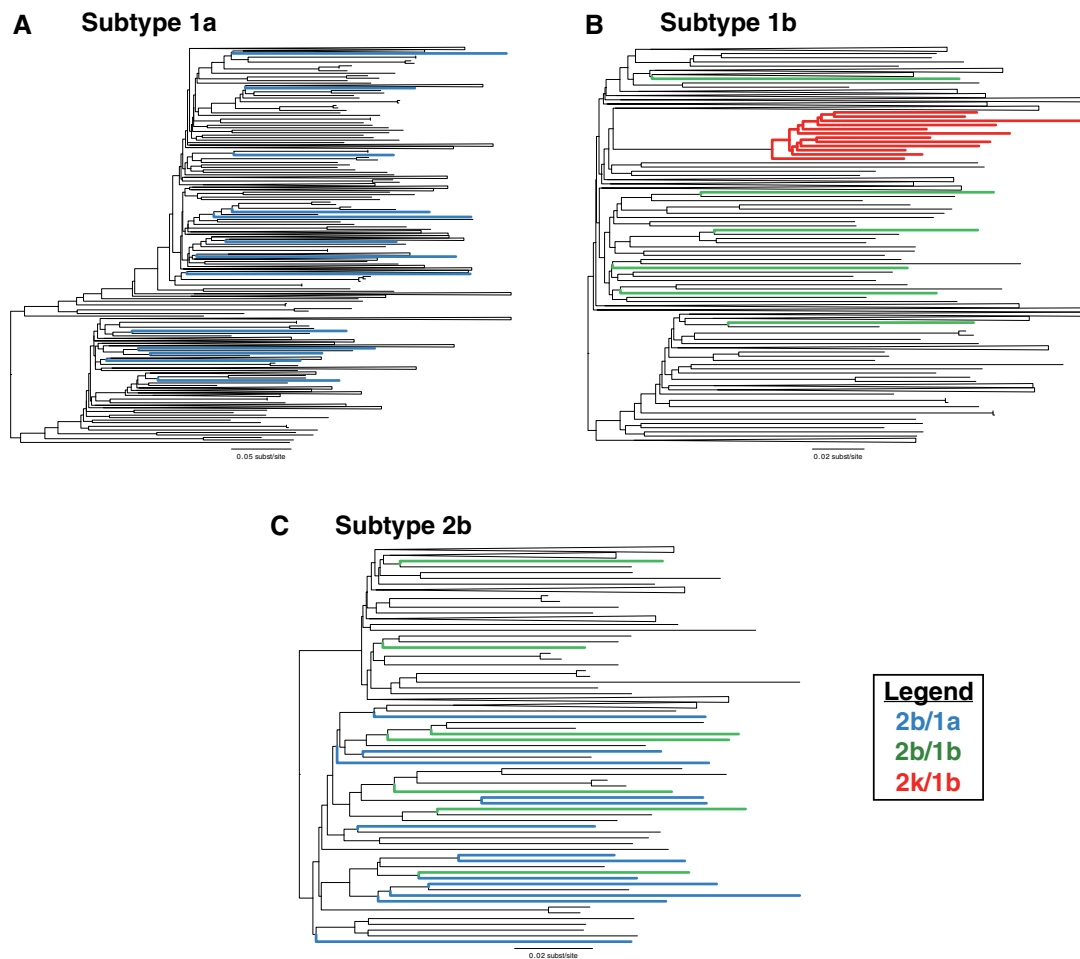
Figure 3. Maximum likelihood phylogenetic trees depicting position of recombinant HCV taxa. (A) 2b/1a recombinant hemigenomes relative to subtype 1a taxa ($n = 1,049$ taxa). (B) 2b/1b and 2k/1b recombinant hemigenomes relative to subtype 1b reference taxa ($n = 463$ taxa). (C) 2b/1a and 2b/1b recombinant hemigenomes relative to subtype 2b reference taxa ($n = 116$ taxa). All non-recombinant clades comprising four or more taxa are collapsed. Recombinant genomes are bolded and colored. Complete phylogenies including bootstrap support values are available as Supplementary Material.

observation affirms the recombination breakpoint analysis, as well as previous work suggesting a single origin of the 2k/1b recombinant strains. None of the 2b/1b strains were sister taxa (i.e., most closely related to each other) in either subtype 1b or subtype 2b phylogenies. For 2b/1a, two strains were most closely related to each other in the 2b tree (Fig. 3C), an arrangement consistent with a shared recombinant origin; however, this tree included only 116 taxa. In the larger 1a phylogeny (1,049 taxa; Fig. 3A), no 2b/1a genomes were sister taxa, indicating that each of the analyzed 2b/1a and 2b/1b genomes arose from independent recombination events. Our results indicate that all sequenced HCV 2b/1a and 2b/1b recombinant viruses are each of independent origin, whereas all 2k/1b strains share a single common ancestor with ultrafast bootstrap support of 100 per cent (Fig. 3B).

3.3 CRF 2k/1b temporal analysis

We observed low posterior node support for the MCC trees from both the full coding region and densely sampled region alignments, regardless of which nucleotide substation model or URL clock partitioning strategy was used, consistent with previous work (Raghwani et al. 2012). Here, we focus on results from the MCMC analysis using a single nucleotide substitution model

and URL clock for the entire alignment; however, as we will show, our findings are consistent across the simple and more complex partition strategies. Coalescent demographic reconstructions from each alignment using Bayesian Skyride show a consistent increase in the 2k/1b population size over time, though this growth is gradually tapering (Supplementary Fig. S1). The 2k/1b MCC trees did not contain lineages basal to the tree inferred in the Raghwani analysis. Therefore, we recovered our TMRCA prior of 1946 (95% HPD: 1944.7–1947.1).

The inferred substitution rate for the full coding region analysis, given in substitutions/site/year, was 8.25×10^{-4} (95% HPD: 7.78×10^{-4}–8.77×10^{-4}) and 7.78×10^{-4} (95% HPD: 7.29×10^{-4}–8.28×10^{-4}) for the densely sampled region analysis. These full-genome rates were slightly lower than the previously reported rates for Core/E1 (1.56×10^{-3}) and NS5B (0.90×10^{-3}) (Raghwani et al. 2012). By partitioning the URL clock across the full coding region, we were able to characterize gene-specific substitution rates for 2k/1b (Table 1).

3.4 Geographic origin of 2k/1b recombinant

Russia was the inferred ancestral geographic source of 2k/1b in the phylogeographic analyses with a posterior probability of 0.99 in both the full coding region phylogeny (Fig. 4) and the

Table 1. Region-specific evolutionary rates for CRF 2k/1b from full coding region analysis with partitioned GTR+Γ_4 nucleotide substitution models and partitioned URL molecular clocks.

Genomic region	Mean rate[a]	95% HPD[a]
Core	0.45	0.38–0.52
E1	5.53	4.71–6.43
E2	1.42	1.22–1.63
P7	1.74	1.27–2.21
NS2	0.70	0.62–0.79
NS3	1.40	1.27–1.54
NS4A/NS4B	0.62	0.50–0.75
NS5A	1.09	0.98–1.22
NS5B	0.71	0.61–0.81
Core	0.45	0.38–0.52

[a]Rates shown as 10^{-3} substitutions/site/year.

densely sampled region phylogeny (Supplementary Fig. S2). Russian sequences were interspersed throughout the tree, and ancestral state reconstruction suggests that virus from other countries are all nested within Russian lineages (Fig. 4). We also observed clusters of sequences from Germany, Chechnya, Azerbaijan, Belarus, and Georgia. Posterior support for these lineages varies (Supplementary Figs S3 and S4). Whereas location of sampling informed us of where the virus is present at specific time points, the ancestral state reconstruction analysis inferred the presence of 2k/1b in several former Soviet regions based on the trait assignment of internal nodes, including Georgia (in the 1950s through 1990s), Belarus (in the 1970s), Azerbaijan (in the 1970s), and Chechnya (in the 1980s) (Fig. 4). Further, despite the inclusion of sequences from seven non-Soviet countries, we found no instances of non-Soviet sister taxa in the phylogeny before the Soviet Union dissolved. Presence of clades whose country of origin lies outside of the Soviet bloc prior to 1991 would indicate sustained non-Soviet transmission clusters. Our phylogeny reveals no non-Soviet clades before this date, and a sole pairing of German sister taxa in the mid-1990s, though we have no context on whether these may be from the former East or West Germany. We also estimated the timing of migration events out of the former Soviet Union into non-Soviet countries using Markov Jumps. All of these jumps have a 95% HPD that includes the era post-1991, after the collapse of the Soviet Union. Combined, these observations indicate that 2k/1b circulated across the Soviet sphere of influence during its existence and argue against the introduction of 2k/1b outside of the former Soviet bloc prior to 1991.

3.5 Migration of 2k/1b

Across the full coding and densely sampled datasets and partitioning schemes, we inferred six strongly supported migration paths out of Russia from the BSSVS network (BF ≥ 20). All six of these migration paths emanated from Russia to regions in the former Soviet bloc or its sphere of influence: Armenia, Azerbaijan, Germany, Chechnya, Ukraine, and Georgia (Fig. 5 and Supplementary Table S3). Viral migration from Russia to Georgia appears to have been the most frequent, with a median of four inferred migration events across the phylogeny (95% HPD: 2–7 migration events). Support for migration from Russia to Chechnya was reduced when multiple nucleotide substitution models and URL clocks were applied across partitions in the full-coding region analysis.

In addition, we found positive evidence (BF ≥ 3.3) for significant migration routes from Armenia to Romania, Azerbaijan to Russia, Georgia to Belgium, Russia to Kazakhstan, Russia to the Netherlands, and Russia to Uzbekistan (Supplementary Table S3). These findings were consistent between the full coding and densely sampled datasets and across partition strategies (Supplementary Table S3).

The phylogeographic analysis identified three instances of clustered sequences with matched risk factors and location (Supplementary Fig. S5). A paraphyletic group which includes six PWID-associated Azerbaijani isolates is the largest cluster in our analysis. The cluster was monophyletic in a previous analysis (Raghwani et al. 2012) but now includes a seventh sequence isolated from a woman in Italy who received a blood transfusion in Russia following a miscarriage in 1988 (Paolucci et al. 2017). A pair of Azerbaijani men within this cluster who reported injection drug use shared identical HCV sequences, likely representative of acute infection (Matthews et al. 2011). A second phylogenetic cluster comprised two injection drug users of Georgian descent, whose viruses were identified and sequenced in 2007 while they were residing in France and the Netherlands.

3.6 Caucasus region as alternative origin of 2k/1b

Given the high prevalence of 2k/1b in the Caucasus mountain region, and to account for the changing international borders in this region over the last century, we performed a phylogeographic sensitivity analysis in which these three Caucasus mountain countries were joined as a single region. Nonetheless, this analysis found an 80 per cent posterior probability that the ancestor of 2k/1b was found in Russia, even when more taxa were included from the Caucasus region ($n = 40$) than from Russia ($n = 30$). Only 20 per cent of the posterior samples placed the 2k/1b ancestor in the Caucasus region.

4. Discussion

The 2k/1b CRF of HCV extends the opportunity to assess the origin and dissemination of a recombinant virus and expands our current understanding of the evolutionary history of HCV. By including newly sequenced viruses, our findings confirm the single-origin hypothesis for 2k/1b from a common source in Russia during the mid-twentieth century and provide robust evidence for the diversification of CRF 2k/1b in the former Soviet Union prior to the fall of the Iron Curtain in 1991.

We report that the most frequent route of migration out of Russia was to Georgia, the country with the highest documented prevalence of 2k/1b. All of the six highly supported 2k/1b migration routes were inferred to originate in Russia. Importantly, these findings suggest a source-sink dynamic in which Russia served as a source. Among the countries that have evidence of sustained transmission after introduction, all were part of the former Soviet Union, and many of these introductions date back to the 1970s. These observations, combined with a lack of evidence for transmission to non-Soviet countries prior to 1991, suggest that 2k/1b migration was restricted by the Iron Curtain. An epoch analysis aimed at discerning migration events before and after this time point [as in Vasylyeva et al. (2018)] was not feasible due to the dearth of coalescent events in the phylogeny post-1991. As more 2k/1b strains are discovered and sequenced, a more nuanced picture of the recent evolution of 2k/1b will likely emerge.
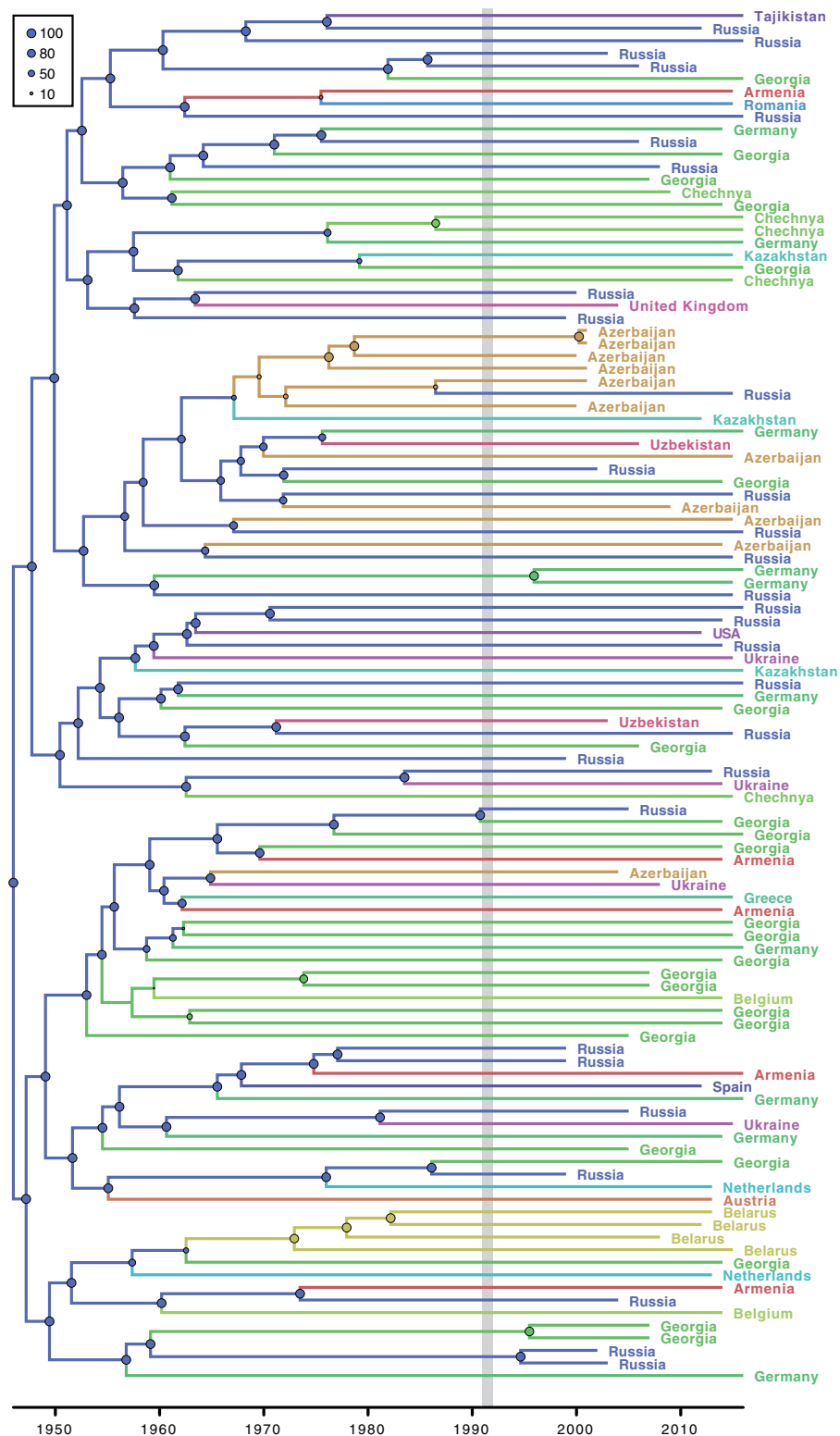
Figure 4. Phylogeographic reconstruction of the HCV 2k/1b clade from full coding region analysis. Phylogenetic inference performed using a single substitution model and relaxed molecular clock. Node and branch colors correspond to the inferred most probable ancestral state location, and each lineage is labeled with country name. Posterior probability support for inferred geographic location is depicted as the area of the circle on each internal node (larger circles indicate greater support; scale shown in insert). Gray bar at 1991 represents the collapse of the Soviet Union.

In the mid-twentieth century, the majority of HCV infections were due to blood transfusions and medical interventions (Liakina et al. 2009). Now, in developed nations, HCV transmission is largely attributed to PWID (Kalinina et al. 2001; Pybus et al. 2001; Rodrigo et al. 2016). In our dataset, 78 per cent of sequences for which we have risk factor information were from people who reported injection drug use. The remainder of our sequences from people with identified risk factors reported
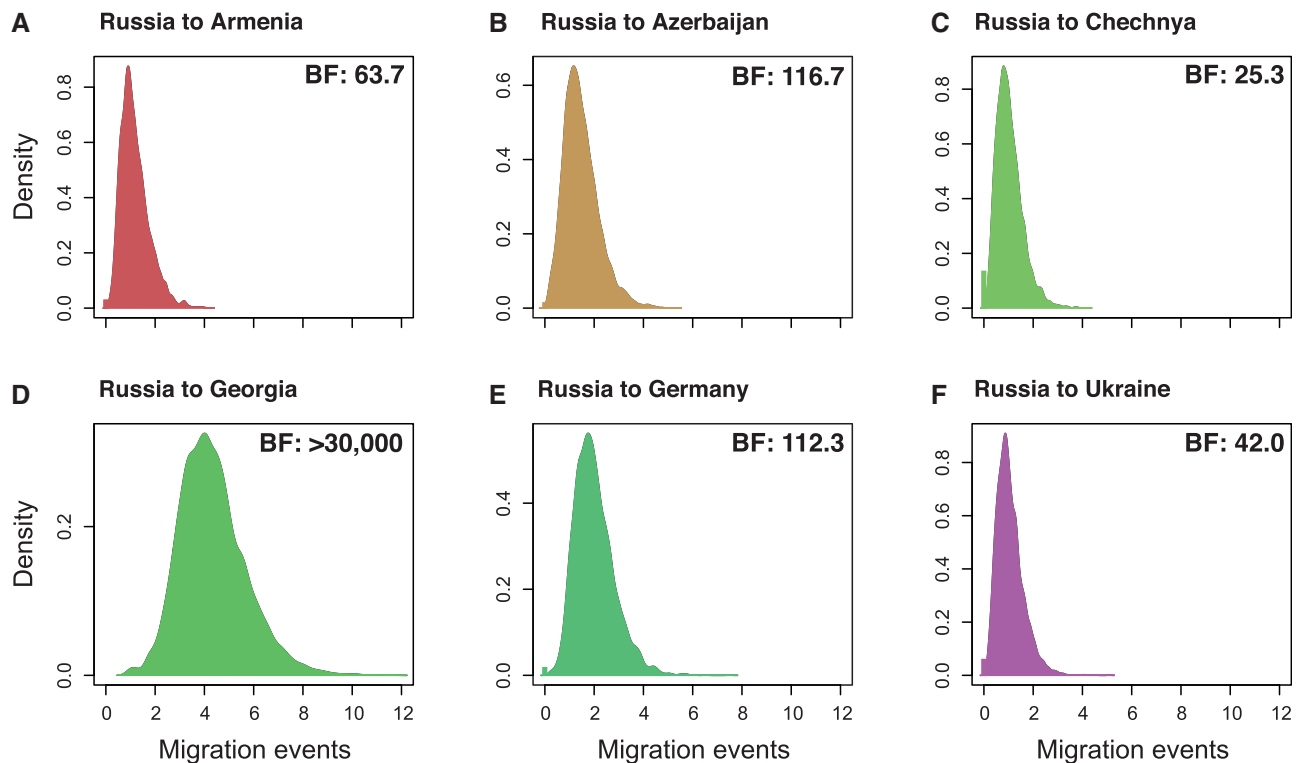
Figure 5. Density plots of the relevant migration events between Russia and six other countries (A–F) estimated from full coding region analysis with a single substitution model and molecular clock partition. Probability that the number of migration events is zero is shown as a bar at the 0 on the X-axis. Only migration routes with Bayes Factor (BF) ≥20 in the phylogeographic analysis are reported here. See Supplementary Table S3 for results from other partition strategies and densely sampled region analyses.

acquisition through iatrogenic means (e.g., tattoo exposure, blood transfusion, nosocomial infection from dialysis, or surgical intervention) though no two individuals with a reported iatrogenic risk were most closely related in the phylogeny. Despite the change in transmission routes, coalescent demographic reconstructions reveal overall reduced transmission rates compared with the initial phase of the 2k/1b epidemic (Supplementary Fig. S1). Bayesian Skyride methods from both the full coding and densely sampled regions show that the overall growth of 2k/1b is slowing, a finding which is in agreement with the previous reconstructions completed in 2012 (Raghwani et al. 2012). Notably, the rise in 2k/1b CRF transmission occurred while the Soviet Union was developing the earliest centralized blood donation system, when centers were surging in popularity (Huestis 2002). Moreover, the stabilization period of incidence (Volz et al. 2009) of 2k/1b infections from demographic reconstructions in the early 1990s (Supplementary Fig. S1) is concurrent with widely implemented measures to increase HCV screening in blood donors (van der Poel 1999). However, we note that coalescent-based population dynamic reconstructions can produce erroneous inference of reduced growth and stabilization (de Silva et al. 2012; Volz and Didelot 2018).

The circumstances surrounding the origin of 2k/1b remain obscure. In the 1930s, prior to the start of World War II, the Soviet Union developed a highly advanced and sophisticated Soviet blood banking and transfusion network. Stockpiled blood from cadavers, placentas, and donors were 'canned' and stored at nearly 600 blood storage centers throughout the Soviet Union (Starr 1999). The estimated emergence of 2k/1b in 1946 coincides temporally with the utilization of this blood bank network (Huestis 2002) and with the prolific expansion of subtype 1b

from developed to developing nations (Kendrick 1964; Magiorkinis et al. 2009). Subtype 2k is suspected to have been present in the former Soviet Union, and the 2k strains most closely related to the ancestor of 2k/1b were from persons in Moldova and Azerbaijan (Raghwani et al. 2012). These conditions may have resulted in the opportunity for the formation of a transmissible HCV recombinant form.

Like all phylogeographic analyses, our study was constrained by non-random sampling of viral sequences. Nonetheless, the Caucasus region sensitivity analysis revealed that although the Caucasus countries were the most common geographic region sampled, our finding that Russia is the geographic source of the sampled 2k/1b strains was not unduly influenced by this sampling bias. Our genome-wide, full coding region analysis encompassed the entirety of available 2k/1b sequence diversity and increased the number of geographic locales compared to previous investigations. However, this approach comes with a trade-off, whereby the relationship between strains with non-overlapping sequences can only be inferred through related strains with overlapping sequences. Even so, the same highly supported migration routes were inferred using only the densely sampled genomic regions.

Globalization has facilitated the movement of pathogens through trade and travel routes. Evolutionary inferences must be taken together with these societal factors which contribute to disease dynamics (Grenfell et al. 2004; Holmes 2004). Ecological interpretations of host availability and mobility provide context for temporal and spatial relationships (Pybus, Tatem, and Lemey 2015). Our findings reinforce the power of phylodynamics in revealing the geopolitical context of viral evolution history. The phylogeographic results detailed here

explain the origin and dissemination of HCV 2k/1b recombinants from a probable source in Russia during the 1940s. Our demographic reconstructions provide a feasible timeline in which the extensive development of blood banks and transfusions in the Soviet Union preceding World War II coincide with increased incidence of 2k/1b, followed by the implementation of blood screening procedures and the fall of the Iron Curtain in the early 1990s.

Acknowledgements

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

References

Al-Qahtani, A. A. et al. (2017) 'The Epidemic Dynamics of Hepatitis C Virus Subtypes 4a and 4d in Saudi Arabia', *Scientific Reports*, 7: 44947.

Alter, M. J. (1999) 'Hepatitis C Virus Infection in the United States', *Journal of Hepatology*, 31 Suppl 1: 88–91.

Anisimova, M. et al. (2011) 'Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-Based Approximation Schemes', *Systematic Biology*, 60: 685–99.

Ayres, D. L. et al. (2019) 'BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics', *Systematic Biology*.

Bhattacharya, D. et al. (2011) 'Naturally Occurring Genotype 2b/1a Hepatitis C Virus in the United States', *Virology Journal*, 8: 458.

Borgia, S. M. et al. (2018) 'Identification of a Novel Hepatitis C Virus Genotype from Punjab, India—Expanding Classification of Hepatitis C Virus into 8 Genotypes', *The Journal of Infectious Diseases*, 218: 1722–9.

Calado, R. A. et al. (2011) 'Hepatitis C Virus Subtypes Circulating among Intravenous Drug Users in Lisbon, Portugal', *Journal of Medical Virology*, 83: 608–15.

Chen, S. L., and Morgan, T. R. (2006) 'The Natural History of Hepatitis C Virus (HCV) Infection', *International Journal of Medical Sciences*, 3: 47–52.

Choo, Q. L. et al. (1989) 'Isolation of a cDNA Clone Derived from a Blood-Borne non-A, non-B Viral Hepatitis Genome', *Science*, 244: 359–62.

Colina, R. et al. (2004) 'Evidence of Intratypic Recombination in Natural Populations of Hepatitis C Virus', *The Journal of General Virology*, 85: 31–7.

Cristina, J., and Colina, R. (2006) 'Evidence of Structural Genomic Region Recombination in Hepatitis C Virus', *Virology Journal*, 3: 53.

Cuypers, L. et al. (2016) 'Impact of HCV Genotype on Treatment Regimens and Drug Resistance: A Snapshot in Time', *Reviews in Medical Virology*, 26: 408–34.

de Silva, E., Ferguson, N. M., and Fraser, C. (2012) 'Inferring Pandemic Growth Rates from Sequence Data', *Journal of the Royal Society, Interface*, 9: 1797–808.

Dellicour, S. et al. (2018) 'Phylodynamic Assessment of Intervention Strategies for the West African Ebola Virus Outbreak', *Nature Communications*, 9: 2222.

Drummond, A. J. et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

—— et al. (2005) 'Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences', *Molecular Biology and Evolution*, 22: 1185–92.

Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.

Edwards, C. J. et al. (2011) 'Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline', *Current Biology*, 21: 1251–8.

Feinstone, S. M. et al. (1975) 'Transfusion-Associated Hepatitis Not Due to Viral Hepatitis Type a or B', *The New England Journal of Medicine*, 292: 767–70.

Gaspareto, K. V. et al. (2016) 'HCV Inter-Subtype 1a/1b Recombinant Detected by Complete-Genome Next-Generation Sequencing', *Archives of Virology*, 161: 2161–8.

Giannini, C. et al. (1999) 'Prevalence of Mixed Infection by Different Hepatitis C Virus Genotypes in Patients with Hepatitis C Virus-Related Chronic Liver Disease', *Journal of Laboratory and Clinical Medicine*, 134: 68–73.

Gilbert, M. T. et al. (2007) 'The Emergence of HIV/AIDS in the Americas and Beyond', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 18566–70.

Grenfell, B. T. et al. (2004) 'Unifying the Epidemiological and Evolutionary Dynamics of Pathogens', *Science (New York, NY)*, 303: 327–32.

Gupta, D. et al. (2017) 'Recombination in Hepatitis C Virus Is Not Uncommon among People Who Inject Drugs in Kolkata, India', *Infection, Genetics and Evolution*, 48: 156–63.

Hedskog, C. et al. (2015a) 'Genotype- and Subtype-Independent Full-Genome Sequencing Assay for Hepatitis C Virus', *Journal of Clinical Microbiology*, 53: 2049–59.

—— et al. (2015b) 'Characterization of Hepatitis C Virus Intergenotypic Recombinant Strains and Associated Virological Response to Sofosbuvir/Ribavirin', *Hepatology (Baltimore, MD)*, 61: 471–80.

Hoang, D. T. et al. (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35: 518–22.

Holmes, E. C. (2004) 'The Phylogeography of Human Viruses', *Molecular Ecology*, 13: 745–56.

Huestis, D. W. (2002) 'Russia's National Research Center for Hematology: Its Role in the Development of Blood Banking', *Transfusion*, 42: 490–4.

Jaeckel, E. et al. (2001) 'Treatment of Acute Hepatitis C with Interferon Alfa-2b', *The New England Journal of Medicine*, 345: 1452–7.

Kageyama, S. et al. (2006) 'A Natural Inter-Genotypic (2b/1b) Recombinant of Hepatitis C Virus in the Philippines', *Journal of Medical Virology*, 78: 1423–8.

Kalinina, O. et al. (2002) 'A Natural Intergenotypic Recombinant of Hepatitis C Virus Identified in St. Petersburg', *Journal of Virology*, 76: 4034–43.

—— et al. (2001) 'Shift in Predominating Subtype of HCV from 1b to 3a in St. Petersburg Mediated by Increase in Injecting Drug Use', *Journal of Medical Virology*, 65: 517–24.

Kass, R. E., and Raftery, A. E. (1995) 'Bayes Factors', *Journal of the American Statistical Association*, 90: 773–95.

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

Kauhl, B. et al. (2015) 'The Spatial Distribution of Hepatitis C Virus Infections and Associated Determinants–an Application of a Geographically Weighted Poisson Regression for Evidence-Based Screening Interventions in Hotspots', *PLoS One*, 10: e0135656.

Kendrick, D. B. 1964. *Blood Program in World War II*. Washington, DC: Office of the Surgeon General/Department of the Army.

Kuiken, C. et al. (2005) 'The Los Alamos Hepatitis C Sequence Database', *Bioinformatics (Oxford, England)*, 21: 379–84.

Kurata, H. et al. (2018) 'Chronic Hepatitis Caused by Hepatitis C Virus Showing a Discrepancy between Serogroup and Genotype Because of Intergenotypic 2b/1b Recombination: A Pitfall in Antiviral Therapy with Direct-Acting Antivirals', *Hepatology Research: The Official Journal of the Japan Society of Hepatology*, 48: E372–E378.

Kurbanov, F. et al. (2008) 'Molecular Epidemiology and Interferon Susceptibility of the Natural Recombinant Hepatitis C Virus Strain RF1_2k/1b', *The Journal of Infectious Diseases*, 198: 1448–56.

Larsson, A. (2014) 'AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets', *Bioinformatics (Oxford, England)*, 30: 3276–8.

Lauer, G. M., and Walker, B. D. (2001) 'Hepatitis C Virus Infection', *The New England Journal of Medicine*, 345: 41–52.

Lee, Y. M. et al. (2010) 'Molecular Epidemiology of HCV Genotypes among Injection Drug Users in Taiwan: Full-Length Sequences of Two New Subtype 6w Strains and a Recombinant Form_2b6w', *Journal of Medical Virology*, 82: 57–68.

Legrand-Abravanel, F. et al. (2007) 'New Natural Intergenotypic (2/5) Recombinant of Hepatitis C Virus', *Journal of Virology*, 81: 4357–62.

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

Liakina, V. et al. (2009) 'Changes in Hepatitis C Virus Infection Routes and Genotype Distribution in a Lithuanian Cohort with Chronic Hepatitis C', *Medical Science Monitor*, 15: PH17–23.

Magiorkinis, G. et al. (2009) 'The Global Spread of Hepatitis C Virus 1a and 1b: A Phylodynamic and Phylogeographic Analysis', *PLoS Medicine*, 6: e1000198.

—— et al. (2007) 'Analysing the Evolutionary History of HCV: Puzzle of Ancient Phylogenetic Discordance', *Infection, Genetics and Evolution*, 7: 354–60.

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

Matthews, G. V. et al. (2011) 'Patterns and Characteristics of Hepatitis C Transmission Clusters among HIV-Positive and HIV-Negative Individuals in the Australian Trial in Acute Hepatitis C', *Clinical Infectious Diseases*, 52: 803–11.

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

Mohd Hanafiah, K. et al. (2013) 'Global Epidemiology of Hepatitis C Virus Infection: New Estimates of Age-Specific Antibody to HCV Seroprevalence', *Hepatology (Baltimore, MD)*, 57: 1333–42.

Morel, V. et al. (2016) 'Phylogenetic Analysis of a Circulating Hepatitis C Virus Recombinant Strain 1b/1a Isolated in a French Hospital Centre', *Infection, Genetics and Evolution*, 40: 374–80.

Moreno, P. et al. (2009) 'Evidence of Recombination in Hepatitis C Virus Populations Infecting a Hemophiliac Patient', *Virology Journal*, 6: 203.

Moureau, G. et al. (2015) 'New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences', *PLoS One*, 10: e0117849.

Nguyen, L. T. et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

Noppornpath, S. et al. (2006) 'Identification of a Naturally Occurring Recombinant Genotype 2/6 Hepatitis C Virus', *Journal of Virology*, 80: 7569–77.

Paolucci, S. et al. (2017) 'HCV Intergenotype 2k/1b Recombinant Detected in a DAA-Treated Patient in Italy', *Antiviral Therapy*, 22: 365–8.

Perez, A. B. et al. (2019) 'Increasing Importance of European Lineages in Seeding the Hepatitis C Virus Subtype 1a Epidemic in Spain', *Euro Surveill*, 24: 1800227.

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

Pybus, O. G. et al. (2001) 'The Epidemic Behavior of the Hepatitis C Virus', *Science (New York, NY)*, 292: 2323–5.

——, Tatem, A. J., and Lemey, P. (2015) 'Virus Evolution and Transmission in an Ever More Connected World', *Proceedings. Biological Sciences*, 282: 20142878.

Raghwani, J. et al. (2012) 'Origin and Evolution of the Unique Hepatitis C Virus Circulating Recombinant Form 2k/1b', *Journal of Virology*, 86: 2212–20.

Rambaut, A. et al. (2018) 'Posterior Summarisation in Bayesian Phylogenetics Using Tracer 1.7', *Systems Biology*, 67: 901–4.

Ramiere, C. et al. (2014) 'Recent Evidence of Underestimated Circulation of Hepatitis C Virus Intergenotypic Recombinant Strain RF2k/1b in the Rhone-Alpes Region, France, January to August 2014: Implications for Antiviral Treatment', *Euro Surveill*, 19: 20944.

Rodrigo, C. et al. (2016) 'Historical Trends in the Hepatitis C Virus Epidemics in North America and Australia', *The Journal of Infectious Diseases*, 214: 1383–9.

Shapiro, B. et al. (2011) 'A Bayesian Phylogenetic Method to Estimate Unknown Sequence Ages', *Molecular Biology and Evolution*, 28: 879–87.

Shi, W. et al. (2012) 'Recombination in Hepatitis C Virus: Identification of Four Novel Naturally Occurring Inter-Subtype Recombinants', *PLoS One*, 7: e41997.

Simmonds, P. (2004) 'Genetic Diversity and Evolution of Hepatitis C Virus–15 Years On', *The Journal of General Virology*, 85: 3173–88.

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

Starr, D. 1999. *Blood: An Epic History of Medicine and Commerce.* New York, NY: Harper Perennial.

Stumpf, M. P., and Pybus, O. G. (2002) 'Genetic Diversity and Models of Viral Evolution for the Hepatitis C Virus', *FEMS Microbiology Letters*, 214: 143–52.

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

Susser, S. et al. (2017) 'Origin, Prevalence and Response to Therapy of Hepatitis C Virus Genotype 2k/1b Chimeras', *Journal of Hepatology*, 67: 680–6.

Tallo, T. et al. (2007) 'Genetic Characterization of Hepatitis C Virus Strains in Estonia: Fluctuations in the Predominating Subtype with Time', *Journal of Medical Virology*, 79: 374–82.

U.S. Food and Drug Administration. (2017) 'FDA approves Vosevi for Hepatitis C', fda.gov.

van der Poel, C. L. (2007) 'Hepatitis C Virus and Blood Transfusion: Past and Present Risks', *Journal of Hepatology*, 31 Suppl 1: 101–6.

Vasylyeva, T. I. et al. (2018) 'Molecular Epidemiology Reveals the Role of War in the Spread of HIV in Ukraine', *Proceedings of the National Academy of Sciences of the United States of America*, 115: 1051–6.

Veras, N. M. et al. (2011) 'High-Resolution Phylogenetics and Phylogeography of Human Immunodeficiency Virus Type 1 Subtype C Epidemic in South America', *Journal of General Virology*, 92: 1698–709.

Viazov, S. et al. (2010) 'Hepatitis C Virus Recombinants Are Rare Even among Intravenous Drug Users', *Journal of Medical Virology*, 82: 232–8.

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

Volz, E. M., and Didelot, X. (2018) 'Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance', *Systematic Biology*, 67: 719–28.

—— et al. (2009) 'Phylodynamics of Infectious Disease Epidemics', *Genetics*, 183: 1421–30.

Webster, D. P., Klenerman, P., and Dusheiko, G. M. (2015) 'Hepatitis C', *Lancet (London, England)*, 385: 1124–35.

Welzel, T. M. et al. (2017) 'Global Epidemiology of HCV Subtypes and Resistance-Associated Substitutions Evaluated by Sequencing-Based Subtype Analyses', *Journal of Hepatology*, 67: 224–36.

Yokoyama, K. et al. (2011) 'Identification and Characterization of a Natural Inter-Genotypic (2b/1b) Recombinant Hepatitis C Virus in Japan', *Archives of Virology*, 156: 1591–601.

Zakalashvili, M. et al. (2018) 'Identification of Hepatitis C Virus 2k/1b Intergenotypic Recombinants in Georgia', *Liver International*, 38: 451–7.