

The global origins of resistance-associated variants in the non-structural proteins 5A and 5B of the hepatitis C virus

Bradley R. Jones,^{1,†} Anita Y. M. Howe,¹ P. Richard Harrigan,^{1,2} and Jeffrey B. Joy^{1,2,*}

¹Laboratory Program, BC Centre for Excellence in HIV/AIDS, 608—1081 Burrard Street, Vancouver, BC, Canada V6Z 1Y6 and ²Department of Medicine, University of British Columbia, 2775 Laurel Street, Vancouver, BC, Canada V5Z 1M9

*Corresponding author: E-mail: jeffrey.b.joy@gmail.com

†<http://orcid.org/0000-0003-4498-1069>

Abstract

New, costly, fast acting, therapies targeting the non-structural proteins 5A and 5B (NS5A and NS5B) regions of the hepatitis C virus (HCV) genome are curative in the majority of cases. Variants with certain mutations in the NS5A and NS5B regions of HCV have been shown to reduce susceptibility to direct-acting NS5A and NS5B therapy and are found in treatment naïve patients. Despite this, the ease with which these variants evolve is poorly known, as are their evolutionary and geographic origins. To address this crucial gap we inferred the evolutionary and geographic origins of resistance-associated variants (RAVs) in the HCV NS5A and NS5B regions of subtypes 1a, 1b, and 3a sequences available from global databases. We found that RAVs in the NS5A region of HCV, when prevalent, were widely dispersed throughout the phylogenetic tree of HCV with multiple independent origins and that these variants are globally distributed. In contrast, most of the NS5B C316N variants came from one of two clades in the phylogenetic tree of HCV subtype 1b. The presence of serine (S) at codon 218 of HCV NS5B appears to facilitate the evolution of the C316N RAV. Other NS5B RAVs did not arise very frequently in our data set, except for S556G in subtype 1b and with respect to geography NS5B RAVs were also globally distributed. The inferred distribution of RAVs in the NS5A region and frequency of their origin suggest a low fitness barrier without the need for co-evolution of compensatory mutations. A low fitness barrier may allow rapid selection of *de novo* resistance to NS5A inhibitors during therapy.

Key words: hepatitis C virus; drug resistance; phylogenetics; phylogeography

1. Introduction

Understanding the origin and evolution of genetic variation conferring viral resistance to therapeutic agents is crucial to the long-term durability of therapy. Despite increasing access to novel, short-duration, curative, and direct-acting antiviral (DAA) treatment for hepatitis C virus (HCV), HCV associated mortality now exceeds all other infectious conditions in some developed countries (Ly et al. 2016); HCV contains a variety of naturally occurring

mutations in the non-structural proteins 5A and 5B (NS5A and NS5B) that reduce susceptibility to currently approved therapies targeting these regions of the HCV genome (Fridell et al. 2011; Lam et al. 2012; Gao 2013; Gentile, Buonomo, and Borgia 2014; Poordad et al. 2014; Liu et al. 2015; Hezode et al. 2016). These mutations occasionally rise in frequency during treatment with antiviral drugs that target the NS5A or NS5B regions to become dominant in the virus population causing treatment failure and virological rebound.

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1. The NS5A RAVs considered in this study.

Subtype	Variant	Daclatasvir	Ledipasvir	Velpatasvir	Ombitasvir	Elbasvir
1a	M28A	4,951 ¹	>1,000 ²			61 ³⁻⁵
1a	M28G		>1,000 ²			71,429 ^{3,4}
1a	M28T	205 ⁶⁻⁸	61 ^{2,8}		8,965 ⁸⁻¹¹	15 ^{3-5,12}
1a	M28V				58 ^{8,10,11}	
1a	Q30D					925 ^{4,5}
1a	Q30E	7,500 ⁶⁻⁸	5,458 ^{2,7,8}	37 ¹³		56 ^{4,12}
1a	Q30G		100–1,000 ²			84 ^{4,12}
1a	Q30H	435 ⁶⁻⁸	74 ^{2,6-8}			8 ³⁻⁵
1a	Q30K	24,545 ^{1, 8}	>1,000 ²		183 ^{10,11}	
1a	Q30R	365 ⁶⁻⁸	171 ^{2,6-8}		800 ^{8-11,14}	125 ^{3-5,12}
1a	L31F					96 ³⁻⁵
1a	L31I	50 ⁸				134 ⁴
1a	L31M	405 ⁶⁻⁸	141 ^{2,6-8}	12 ^{13,15}		10 ^{3-5,12}
1a	L31V	1,000 ^{6, 8}	>100 ^{2,6}			61 ³⁻⁵
1a	P32L		50–100 ²			
1a	H58D	500 ^{1,7}	1,127 ^{2,7,8}		243 ^{8, 10}	6 ^{3-5,12}
1a	Y93C	555 ^{1,6-8}	1,602 ^{2,6-8}	11 ¹³	1,675 ⁸⁻¹⁰	11 ^{4,5,12}
1a	Y93H	1,600 ⁶⁻⁸	1,677 ^{2,6-8}	609 ^{13,15}	41,383 ^{8-10,14}	220 ^{3-5,12}
1a	Y93N	14,100 ⁶⁻⁸	>14,706 ^{2,7,8}		366,740 ^{8-11,14}	929 ^{3-5,12}
1a	Y93S		>1,000 ²		1,013 ¹⁰	
1b	L28T				661 ^{9,10}	
1b	L31F	5 ⁸			10 ^{10,11}	15 ^{4,5}
1b	L31M					7 ^{4,5}
1b	L31V	28 ^{1,6,8}			8 ^{10,11}	13 ^{4,5}
1b	Y93H	24 ^{1,6,8}	1,325 ^{6,8}		77 ^{8,10,11,14}	
1b	Y93S				12 ¹¹	
3a	M28T				639 ¹⁰	
3a	A30K	44 ^{8,16,17}	14 ^{8,16}	16 ^{13,15}		50 ⁵
3a	L31F	603 ¹⁷			28 ¹⁰	143 ⁵
3a	L31M	537 ^{8,16}				
3a	L31V	1,181 ^{8,16}				
3a	Y93H	2,154 ^{8,16,17}	28 ^{8,16}	246 ¹³	6,728 ¹⁰	485 ⁵

Only RAVs deemed likely to cause resistance to at least one NS5A inhibitor according to the literature were considered. The values are the EC50 fold-shift of the variant with respect for the drug. A blank entry indicates that the RAV was not shown to cause resistance to the particular NS5A inhibitor.

Source: ¹Fridell et al. (2011); ²Sarrazin et al. (2014); ³Black et al. (2015); ⁴Jacobson et al. (2015); ⁵Liu et al. (2015); ⁶Gao (2013); ⁷Wong et al. (2013); ⁸Lontok et al. (2015); ⁹Gentile, Buonomo, and Borgia (2014); ¹⁰Krishnan et al. (2015a); ¹¹Krishnan et al. (2015b); ¹²Lahser et al. (2016b); ¹³Cheng et al. (2013); ¹⁴Poordad et al. (2014); ¹⁵Hezode et al. (2016); ¹⁶Hernandez et al. (2013); ¹⁷Wang et al. (2013).

In addition, NS5A and NS5B ‘resistance-associated variants’ (RAVs), variants with mutations conferring resistance, have been found in viruses from treatment-naïve individuals, making treatment for those individuals by specific antiretroviral drugs ineffective.

Since NS5A and NS5B inhibitors are relatively new, research into the capacity of HCV to develop resistance to these drugs is just beginning, thus little is known about the mutability of RAVs with or without the presence of therapy. Some of these RAVs confer high EC50 fold-shift—the change in the concentration of drug needed to eradicate 50 per cent of the virus with the RAV relative to wildtype virus—such as the Y93H NS5A RAV for the NS5A inhibitor, daclatasvir (1600× in subtype 1a (Gao 2013; Wong et al. 2013, Lontok et al. 2015), 3× in subtype 1b (Lontok et al. 2015) and 537× in subtype 3a (Wang et al. 2013; Lontok et al. 2015)). Many RAVs exhibit increased resistance in the presence of one or more other RAVs in the same region. For example, a combination of S556G and C316N exhibits a 38× EC50 fold-shift for the NS5B inhibitor, dasabuvir, compared to the 5× EC50 fold-shift C316N and the 11× EC50 fold-shift of S556G in subtype 1b (Dietz et al. 2015). An NS5B RAV of note is the S282T variant. The S282T RAV is the only variant to confer resistance to the NS5B inhibitor, sofosbuvir, producing a 13× fold-shift in subtype 1a (Lam et al. 2012; Lontok et al. 2015) and a 7.3× fold-shift in subtype 1b (Lam et al. 2012).

Previous analyses of HCV RAVs in other gene regions, notably the Q80K variant in non-structural protein 3 (NS3), revealed that most resistant lineages were descended from a single origin and that they had a global common origin in North America (McCloskey et al. 2015; Cuyppers et al. 2017). Additionally, these RAVs were shown to be strongly coupled to compensatory mutations suggesting both a high fitness barrier and that it could be relatively difficult to select these RAVs *de novo* during therapy (McCloskey et al. 2015). However, the evolutionary and geographic origins of RAVs in NS5A and NS5B and their dependence on compensatory mutations in these gene regions are poorly known. Thus, the relative ease with which RAVs in NS5A/B arise and may be driven to higher frequency by broad scale selection from DAA therapy remains unclear. If NS5A and NS5B RAVs have been shaped by similar evolutionary dynamics as NS3 RAVs (e.g. Q80K), we expect the NS5A and NS5B phylogenies to contain few large clades of RAVs. Alternatively, we would expect many independent origins of NS5A and NS5B RAVs, indicating a high mutation rate and low fitness barrier for those RAVs. Finally, we may observe few or no instances of an RAV in our datasets. The lack of, or infrequent, observation of a particular RAV in our data could occur for several reasons: (1) the variant has a high fitness barrier and thus does not arise frequently in treatment naïve

Table 2. The NS5B RAVs considered in this study.

Subtype	Variant	EC50 fold-shift
1a	S282T	13 for sofosbuvir ^{1,2}
1a	C316Y	1,472 for dasabuvir ^{1,3,4}
1a	A395G	20 for dasabuvir ⁵
1a	M414I	17 for dasabuvir ^{1,5}
1a	M414T	32 for dasabuvir ^{1,3,5}
1a	M414V	18 for dasabuvir ⁵
1a	N444K	23 for dasabuvir ⁵
1a	Y448C	400 for dasabuvir ⁵
1a	Y448H	975 for dasabuvir ^{1,4,5}
1a	A553T	152 for dasabuvir ^{1,3}
1a	G554S	120 for dasabuvir ^{1,3}
1a	S556G	30 for dasabuvir ^{1,3,4,5}
1a	S556R	261 for dasabuvir ⁵
1b	S282T	7.8 for sofosbuvir ²
1b	C316N	5 for dasabuvir ^{4,6}
1b	C316Y	5 for dasabuvir ^{1,7,5}
1b	S368T	1,569 for dasabuvir ⁵
1b	N411S	54 for dasabuvir ⁵
1b	M414I	15 for dasabuvir ^{7,5}
1b	M414T	26 for dasabuvir ⁵
1b	M414V	18 for dasabuvir ⁵
1b	Y448C	160 for dasabuvir ⁵
1b	Y448H	37 for dasabuvir ⁵
1b	A553V	58 for dasabuvir ⁵
1b	S556G	11 for dasabuvir ^{1,5}
1b	D559G	100 for dasabuvir ⁵

Only RAVs deemed likely to cause resistance to at least one NS5B inhibitor according to the literature were considered. All substitutions are likely to cause resistance to dasabuvir, except S282T, which is likely to cause resistance to sofosbuvir.

Source: ¹Lontok et al. (2015); ²Lam et al. (2012); ³Krishnan et al. (2015b); ⁴Dietz et al. (2015); ⁵Koev et al. (2015); ⁶Kati et al. (2015); ⁷Poordad et al. (2014).

cases; (2) the variant is not readily transmissible; and (3) it may be geographically distributed non-randomly and thus its infrequency could occur as a result of sampling. To address these hypotheses we inferred the global phylogenetic history of HCV RAVs in subtypes 1a, 1b, and 3a from public databases. We then analysed the phylogenetic and geographic origins of RAVs in NS5A and NS5B of HCV. Finally, we investigated a possible permissive mutation for the C316N variant in NS5B.

2. Materials and methods

2.1 Data collection and curation

We collected all of the HCV sequences from GenBank using the query 'hepatitis + C + virus[orgn]' on 30 August 2016, receiving 200,863 sequences. We removed all records not annotated with year and country, resulting in a dataset composed of 71,590 records. Using MAFFT v7.300b (Kato and Standley 2013), we aligned each sequence to the HCV subtype 1a reference genome H77 (accession NC 004102). BioPython v1.67 (Cock et al. 2009) was used to strip insertions relative to H77 and clip the sequences to the NS5A and NS5B regions. Finally, we removed sequences with <50 per cent coverage over the NS5A/NS5B regions of H77 and removed duplicate sequences, retaining 4,916 NS5A sequences and 11,195 NS5B sequences.

The sequences were then genotyped by adding reference sequences for the HCV subtypes: 1a, 1b, 1c, 1g, 2, 3a, 3b, 3i, 3k,

4, 5, 6, and 7 from the Los Alamos National Laboratory HCV Database (LANL) to the NS5A and NS5B alignment. We inferred a distribution of 1,000 bootstrap replicates of the approximate maximum likelihood (ML) trees for each region (NS5A and NS5B) with a generalized time reversible substitution model as implemented in FastTree v2.1.7 (Price, Dehal, and Arkin 2010). To ascribe sequences to particular subtypes, we selected the largest clade in each tree with all of the reference sequences of a particular subtype and no other reference sequence. Sequences that were assigned different subtypes in different replicate trees were discarded. To validate our HCV genotype assignment, results were compared against subtypes assigned by the HCV genotype assignment tool, COMET HCV (Struck et al. 2014); we discarded each sequence whose subtype disagreed with COMET HCV. When both the NS5A and NS5B regions were available for a sequence, if either method assigned different subtypes to the NS5A and NS5B regions then the sequence was discarded. Each sequence in our dataset was then realigned to a reference sequence of the same subtype obtained from LANL and clipped to the NS5A and NS5B regions as above. Sequences with <75 per cent coverage over the NS5A/NS5B region were subsequently discarded. Supplementary Tables S1 and S2 present the number of sequences found per subtype—at this stage we retained 4,510 NS5A sequences and 1,462 NS5B sequences.

At this point we removed all clonal sequences from our datasets by assessing equality in all nucleotide positions of the sequences with BioPython and identical sequences were censored from the dataset. Next, trees were pruned to remove epidemiologically linked sequences. To accomplish this, we created another 1,000 bootstrap replicates of the ML trees for each region and subtype using FastTree, as above, and pruned tips with pairwise distance <0.025 substitutions per base keeping the tip closest to the root. We kept each sequence that was retained in at least 50 per cent of the 1,000 bootstrap trees. Because there were less than 100 sequences in each of subtypes: 1c, 1g, 2, 3b, 3i, 3k, 4, 5, 6, 7, we did not consider them further and only analysed NS5A sequences from subtypes 1a, 1b, and 3a and NS5B sequences from subtypes 1a and 1b since there were fewer than 100 NS5B sequences from subtype 3a. After data curation, our final dataset was composed of 1,390 NS5A sequences and 986 NS5B sequences. A diagram illustrating our data filtering process is shown in Supplementary Fig. S1. Since the collection dates of these sequences pre-date the use of NS5A inhibitors and NS5B inhibitors, these sequences are assumed to be from individuals naïve to treatment with NS5A and NS5B inhibitors; though some individuals may have undergone other HCV treatment.

2.2 Library of RAVs

We curated a list of RAVs in the NS5A region from the literature that confer resistance to at least one the following NS5A inhibitors: daclatasvir, ledipasvir, velpatasvir, ombitasvir and elbasvir (see Table 1). The RAVs were selected on the basis of two criteria: (1) reduced drug susceptibility comparing to wild-type HCV (EC50 fold-change greater than 5- to 100-fold depending on the drug) and (2) the RAVs were observed in patients who failed DAA treatment in the clinic. We curated an analogous list of RAVs in the NS5B region for the NS5B inhibitors: dasabuvir and sofosbuvir (see Table 2).

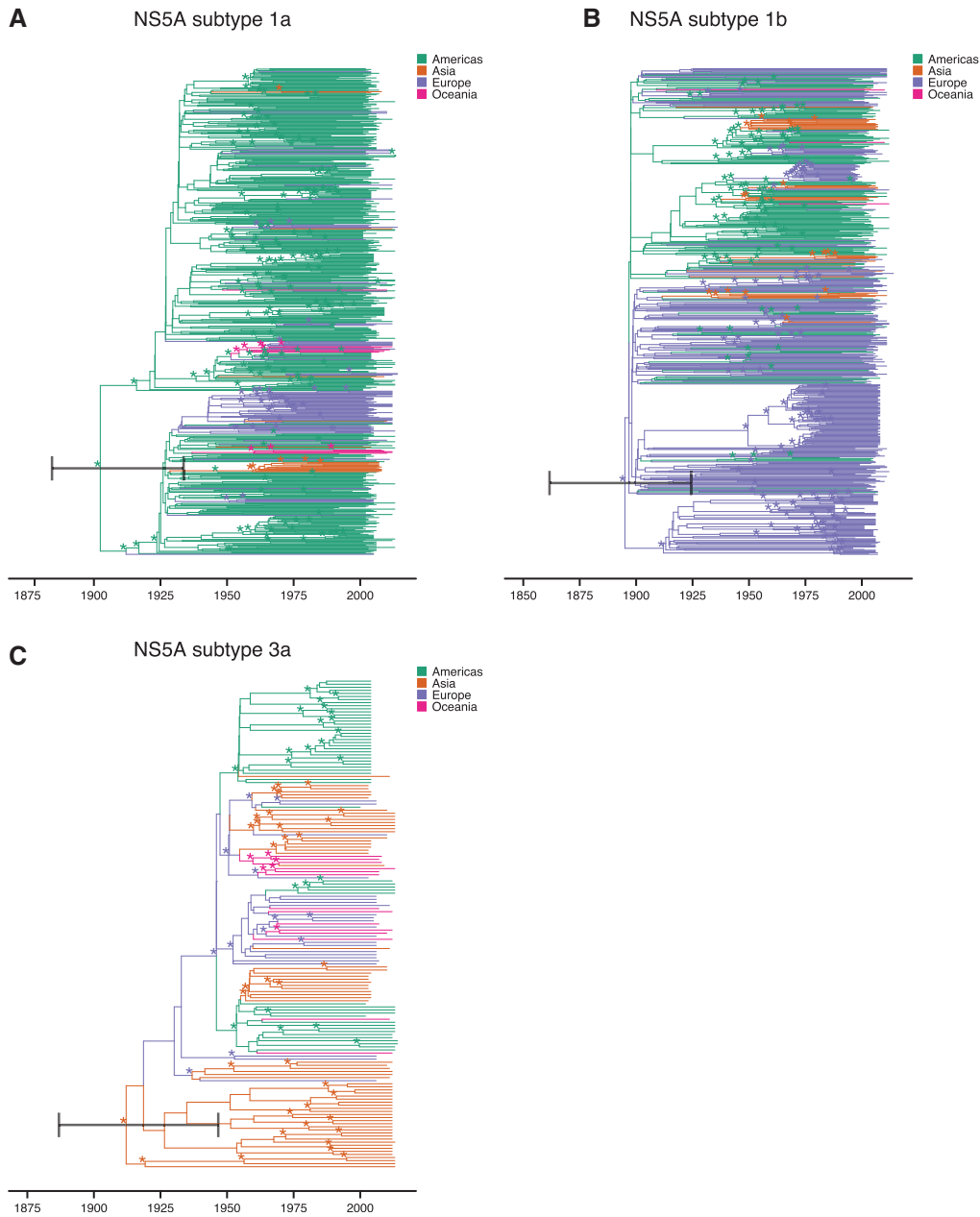


Figure 1. Geographic history of HCV NS5A. Each figure represents the phylogeographic history of the global HCV NS5A population for a particular subtype as an MCC tree. The trees are scaled to units of time with the dates in years common era (CE). Black bars show the 95 per cent HPD interval of the date of the root of the tree. Nodes with posterior probability >90 per cent are marked with an asterisk (*). The edges of the trees are coloured by the inferred continent of origin of the child of the edge. (A) NS5A subtype 1a, (B) NS5A subtype 1b and (C) NS5A subtype 3a.

2.3 Phylogenetic and phylogeographic inferences and ancestral genome reconstruction

For each subtype (1a, 1b, and 3a) and region (NS5A and NS5B), we built 1,000 bootstrap phylogenetic trees with RAxML v8.2.10 (Stamatakis 2014) and a GTR + Γ model of nucleotide substitution. Each tree was rooted with the *rt* function of the R package *ape* (Paradis, Claude, and Strimmer 2004) and time scaled with *node.dating* (Jones and Poon 2017).

We simultaneously reconstructed the ancestral sequences and country of origin using BEAST v1.8.3 (Drummond et al. 2012) Markov chain Monte Carlo with two parallel runs per data set each with 10 million generations. For the BEAST runs, we drew trees from our distribution of 1,000 bootstrap phylogenetic

trees generated by RAxML, to reconstruct the ancestral sequences we used a relaxed lognormal molecular clock model and the GTR + Γ nucleotide substitution model, and to reconstruct the country of origin, we employed a symmetric substitution model and a strict clock model. The state trees from the chosen parallels runs were combined and down sampled to 1,000 trees with LogCombiner v.2.4.7 (Bouckaert et al. 2014). Finally, we inferred maximum clade credibility (MCC) trees with mean heights for each subtype with TreeAnnotator v2.4.7 (Bouckaert et al. 2014).

For each codon position with an RAV observed in at least ten sequences, we computed the distribution of continents of origin of sequences exhibiting RAVs at that codon position.

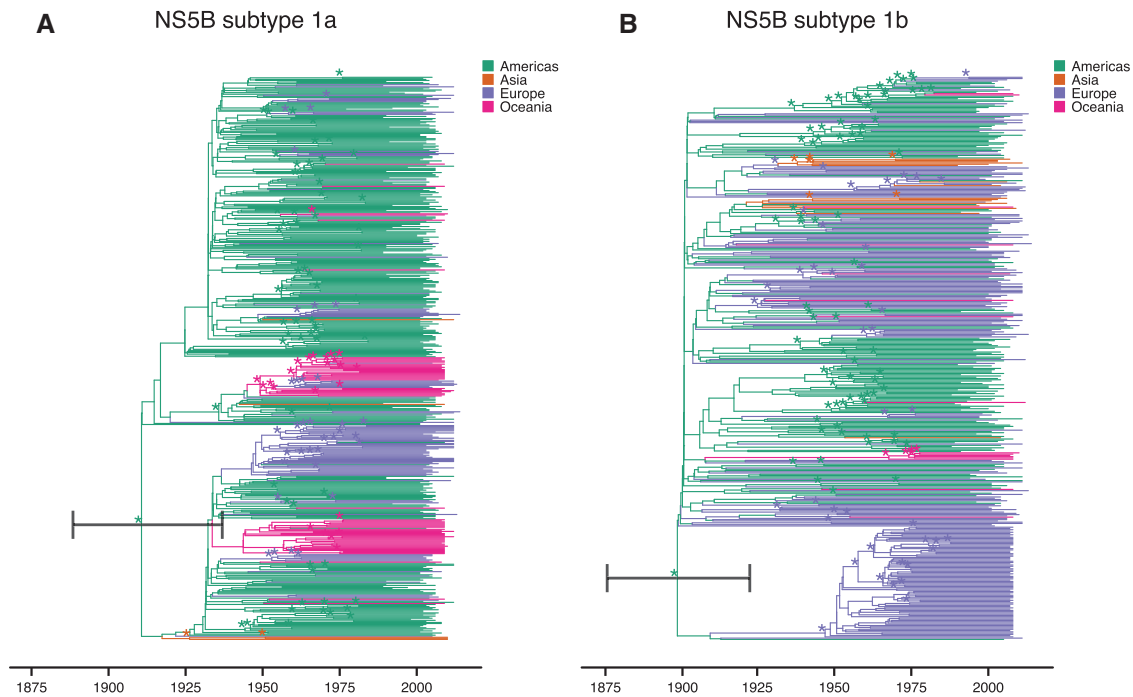


Figure 2. Geographic history of HCV NS5B. Each figure represents the phylogeographic history of the global HCV NS5B population for a particular subtype as a composite of the MCC tree. The trees are scaled to units of time with the dates in years CE. Black bars show the 95 per cent HPD interval of the date of the root of the tree. Nodes with posterior probability >90 per cent are marked with an asterisk (*). The edges of the trees are coloured by the inferred continent of origin of the child of the edge. (A) NS5B subtype 1a and (B) NS5B subtype 1b.

We compared this to the distribution of continents of sequences not exhibiting an RAV at that codon position by computing the concordance (Lin 1989) between the two distributions. We employed Fisher's exact test to test for associations between the presence of RAVs in a sequence and variants at other positions in the same sequence. We performed the concentrated changes test (CCT) (Maddison 1990) using a custom R script to verify whether the A218S variant influenced the evolution of the C316N RAV. Plots were created using the R package ggtree v1.11.0 (Yu et al. 2017) and DensiTree v2.0 (Bouckaert 2010).

3. Results

3.1 Genotypic root ages

We inferred a distribution of 1,000 BEAST trees for each subtype (1a, 1b and 3a) and region (NS5A and NS5B). Figures 1 and 2 show the MCC trees for the NS5A and NS5B data sets, respectively. See Supplementary Figs S1 and S2 for composite plots of the 1,000 BEAST trees with DensiTree (Bouckaert 2010). The roots of the trees were situated in the early twentieth century. In particular, for NS5A subtype 1a, the mean root date was 1902 with a 95 per cent highest posterior density (HPD) interval of (1885, 1935); for NS5A subtype 1b, the mean root date was 1896 with an HPD of (1862, 1924); and for NS5A subtype 3a, the mean root date was 1912 with an HPD of (1888, 1948). The mean root date for NS5B subtype 1a was 1910 with an HPD of (1889, 1934) and the mean root date for NS5B subtype 1b was 1900 with an HPD of (1876, 1923).

3.2 RAVs in NS5A

Ancestral reconstruction reveals that variants conferring reduced susceptibility to NS5A inhibitors (e.g. L31M, Y93H; Table 3) have

many independent origins. Figure 3 and Supplementary Fig. S4 show MCC trees detailing the evolutionary history of the RAVs in NS5A; only the trees for RAVs observed in the data are shown (see Table 3 for the list of NS5A RAVs found and see Supplementary Figs S5–S7 for composite plots).

In subtypes 1a and 1b, tips displaying an NS5A RAV rarely coalesced. Most of the phylogenies inferred using BEAST revealed no large clades of RAVs. What clades they had were composed of only two or three tips (see Table 3). Codons 28, 31, and 93 of NS5A in subtype 1a showed significant presence of RAVs with 1.2–3.3 per cent of the sequences containing an RAV at those codons. The other NS5A RAVs found in subtype 1a (Q30H and H58D) had a lower prevalence (<1%). The only NS5A RAVs found in subtype 1b were L31M and Y93H also at a significant prevalence (see Table 3). In contrast to RAVs in other gene regions, the high frequency of independent origins of NS5A RAVs and their shallow depth in the tree suggests that RAVs in NS5A have a low fitness barrier and that NS5A RAVs could evolve rapidly *in vivo*.

In subtype 3a, the A30K RAV displayed a different pattern. Most of 1,000 BEAST trees contained a single coalescing clade of A30K variants (see Fig. 3C and Table 3). The Y93H RAV, however, behaved the same in subtype 3a as in subtypes 1a and 1b, with many widely dispersed variants few of which coalesced (see Fig. 3D–F).

3.3 RAVs in NS5B

Most RAVs in NS5B had low prevalence or were not present in our data set at all (see Table 4 and Supplementary Figs S8–S11). Interestingly, the S282T RAV, the only known RAV conferring resistance to sofosbuvir, was not found in subtype 1a nor subtype 1b in our data set. However, S556G and C316N RAVs were found in many samples in subtype 1b and were likely to belong

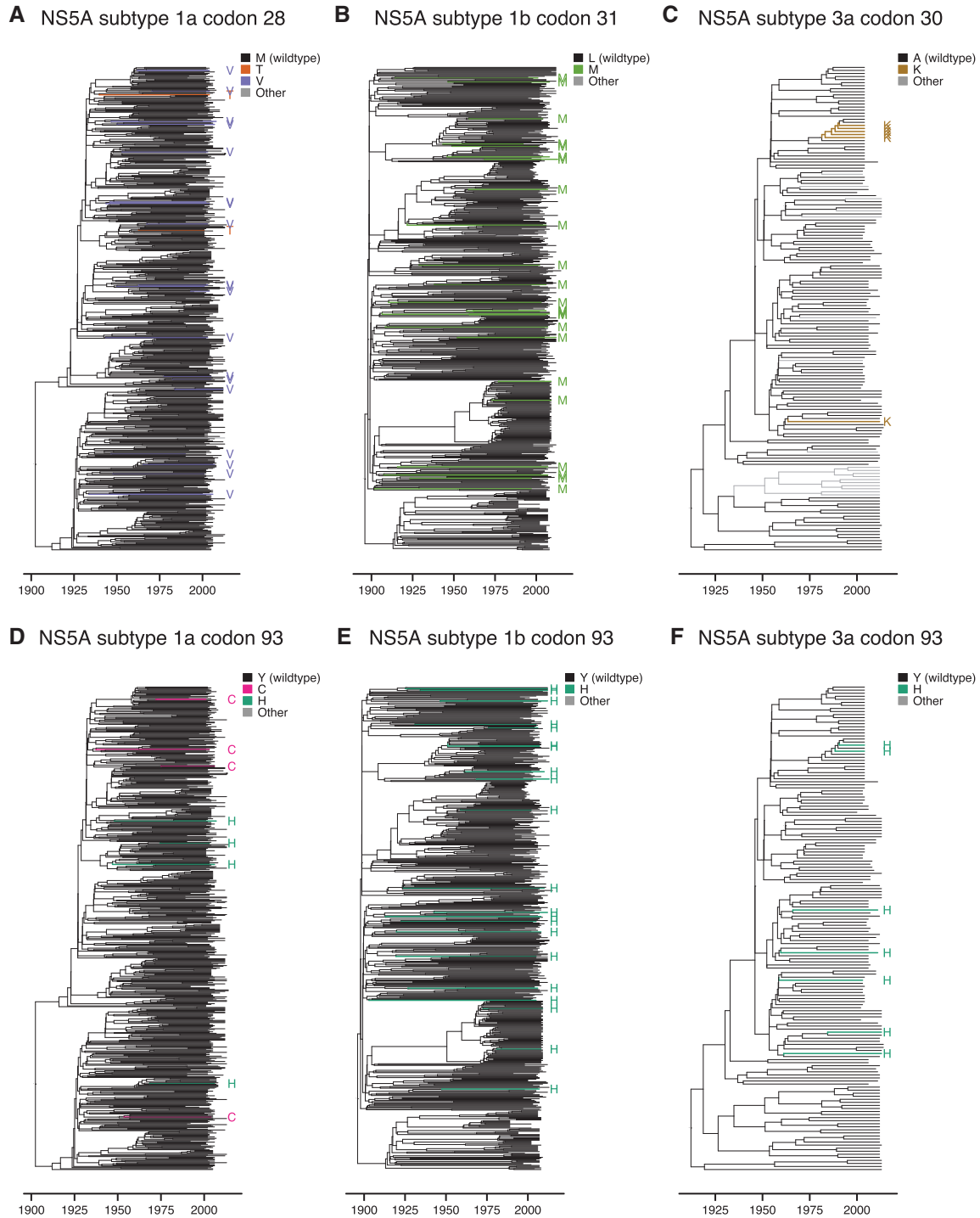


Figure 3. The phylogenetic history of select RAVs in NS5A. Each figure represents the phylogenetic history of the global HCV NS5A population for a particular subtype as an MCC tree. The trees were scaled to units of time with dates in years CE. The edges of the trees are coloured by the inferred state of the specified codon of the child of the edge. (A) M28T/V in subtype 1a, (B) L31M in subtype 1b, (C) A30K in subtype 3a, (D) Y93C/H in subtype 1a, (E) Y93H in subtype 1b and (F) Y93H in subtype 3a.

to large clades of variants with that RAV (see Table 4). In contrast, we found only four samples with the S556G RAV and no samples with the C316 RAV in subtype 1a.

3.4 The C316N RAV in NS5B subtype 1b

The C316N RAV in NS5B subtype 1b produces a $5 \times$ EC50 fold-shift in dasabuvir (Dietz et al. 2015). Two large clades of C316N are present in most of the BEAST trees of NS5B subtype 1b

inferred by our analysis (see Fig. 4 and Supplementary Fig. S12). After compiling co-occurring mutations and testing for association with C316N, we discovered very strong associations of C316N to the variants L159F, A207T and A218S (all $P < 10^{-3}$). Whether or not these variants confer resistance to NS5B inhibitors is unknown.

Only one of the seventy-one sequences containing the C316N variant did not contain the A218S variant. However, this

Table 3. The distribution of NS5A RAVs.

Subtype	Variant	Prevalence (%)	Singletons	Clades	Largest Clade
1a	M28T	0.31 [2]	2 (0)	0 (0)	
1a	M28V	3.0 [19]	19 (0.91)	0.28 (0.52)	2.3 (0.49)
1a	Q30H	0.79 [5]	5.0 (0.083)	0.007 (0.083)	2 (0)
1a	L31M	1.3 [8]	4.8 (0.41)	1.2 (0.40)	3.0 (0.84)
1a	H58D	0.31 [2]	2 (0)	0 (0)	
1a	Y93C	0.63 [4]	4.0 (0.083)	0.007 (0.083)	2 (0)
1a	Y93H	0.63 [4]	4.0 (0.032)	0.001 (0.032)	2 (0)
1b	L31M	4.0 [24]	20 (1.4)	1.9 (0.67)	2.2 (0.55)
1b	Y93H	3.5 [21]	18 (1.2)	1.4 (0.61)	2.1 (0.57)
3a	A20K	3.8 [6]	1.4 (0.66)	1.0 (0.055)	3.6 (1.4)
3a	Y93H	4.4 [7]	6.4 (0.83)	0.44 (0.57)	2.1 (0.22)

Prevalence is the percentage of tips that exhibit the RAV (quantity shown in square brackets). Singletones are the average number of RAV tips that did not have an inferred RAV ancestor. Clade sizes are the average number of clades with an inferred RAV ancestor. Largest clade is the average size of the largest clade with an inferred RAV ancestor (size is calculated by the number of tips in the clade; this may include tips in the clade that do not exhibit an RAV). Largest clade is calculated over all replicates that contain at least one clade. Values in parentheses indicate standard deviations of the data. RAVs not found in the data set are excluded from this table.

sequence appears in most of BEAST trees in a clade containing other C316N variants. Figure 4 shows the evolutionary relationships between C316N and A218S. In 80 per cent of the BEAST trees, each of the clades of C316N variants is contained within clades that exhibit A218S variants, suggesting that A218S serves as a permissive mutation for C316N. This is reinforced by the fact that there are isolated C316N variants and each of these variants has the A218S variant. Further support for this pattern is provided by the CCT (Maddison 1990), which reveals a significant effect of the state of codon 218 on the mutability of codon 316 ($P < 10^{-2}$).

The other highly associated variants, L159F and A207T, only appear alongside C316N in the presence of A218S; in fact, A207T was found exclusively in sequences with the A218S variant. There is also a strong association between C316N and the S556G RAV ($P < 10^{-3}$). It has been shown that the combination of C316N and S556G has a 38× EC50 fold-shift for dasabuvir compared to the 5× EC50 fold-shift of C316N and the 11× EC50 fold-shift of S556G in subtype 1b (Dietz et al. 2015). The large clusters of S556G and C316N variants in NS5B subtype 1b contain sequences from multiple continents. Interestingly, in subtype 1a, cysteine (C) was found at codon 316 and serine (S) was found at codon 218 in every sequence considered in this study.

3.5 RAVs and geography

The phylogenetic trees of each subtype and region had a variety of inferred origins. For the NS5A data sets, most (99.9%) of the subtype 1a BEAST trees had a geographic origin in the Americas, most (96%) of the subtype 1b BEAST trees had a geographic origin in Europe and most (97%) of the subtype 3a BEAST trees had a geographic origin in Asia. For the NS5B data sets, most of the subtype 1a (99.9%) and subtype 1b (95%) BEAST trees had a geographic origin in the Americas. The disagreement of geographic origin between NS5A subtype 1b and NS5B subtype 1b is probably due to oversampling of European sequences in NS5A subtype 1b and American sequences in NS5B subtype 1b.

Table 4. The distribution of NS5B RAVs.

Subtype	Variant	Prevalence (%)	Singletons	Clades	Largest clade
1a	Y448H	0.19 [1]	1 (0)	0 (0)	
1a	S556G	0.75 [4]	4.0 (0.063)	0.005 (0.071)	2.2 (0.49)
1a	S556R	0.19 [1]	0.802 (0.40)	0.20 (0.40)	2.6 (0.53)
1b	C316N	116 [71]	0.23 (0.46)	1.7 (0.56)	58 (22)
1b	S368T	0.22 [1]	1 (0)	0 (0)	
1b	M414I	0.22 [1]	1 (0)	0 (0)	
1b	Y448H	0.22 [1]	1 (0)	0 (0)	
1b	S556G	9.2 [41]	30 (2.4)	3.3 (1.2)	13 (30)

Prevalence is the percentage of tips that exhibit the RAV (quantity shown in square brackets). Singletones are the average number of RAV tips that did not have an inferred RAV ancestor. Clade sizes are the average number of clades with an inferred RAV ancestor. Largest clade is the average size of the largest clade with an inferred RAV ancestor (size is calculated by the number of tips in the clade; this may include tips in the clade that do not exhibit an RAV). Largest clade is calculated over all replicates that contain at least one clade. Values in parentheses brackets indicate standard deviations of the data. RAVs not found in the data set are excluded from this table.

The RAVs in the NS5A region are widely dispersed around the world and were evenly dispersed geographically within our dataset (Fig. 1). The concordance of the continents of origin of RAVs compared to the continents of origin of other variants at the same codon was high in the NS5A region—in subtype 1a (M28T/V: 0.96, Q30H: 0.89, L31M: 0.89, Y93C/H: 0.93), in subtype 1b (L31M: 0.89, Y93H: 0.91) and in subtype 3a (Y93H: 0.97). However, in subtype 3a, the NS5A RAV, A30K, had a concordance of 0.22. This was due to there being more American and Oceanic samples exhibiting the A30K variant. As highlighted previously, the A30K RAV behaves differently than the other RAV in NS5A in that it is highly localized in one clade.

The geographic distribution of RAVs in NS5B subtype 1b is similar to those in NS5A (Fig. 2), with an even dispersion and global distribution of RAVs. Similarly, the concordance of the continents of origin of RAVs compared to the continents of origin of other variants at the same codon was high in the NS5B region (C316N: 0.84, S556G: 0.83). However, in subtype 1a, the concordance for the NS5B RAV, S556G/R, was 0.69. This was due to more sequences exhibiting the S556G/R RAVs having originated from Europe and Oceania.

4. Discussion

The RAVs in the NS5A region have a strikingly different distribution relative to that of the Q80K RAV in NS3 (see Fig. 1 in McCloskey et al. (2015)). While most of the Q80K variants in NS3 with requisite compensatory mutations are descended from a common ancestor, the RAVs in the NS5A region show multiple origins dispersed throughout the HCV phylogeny. The low prevalence of coalescing RAVs combined with the high prevalence of RAVs, suggest that RAVs mutate readily in this gene region and have a low-fitness barrier. The plasticity of the development of RAVs supports the notion that selection for these variants could occur rapidly during treatment with NS5A inhibitors (Lahser et al. 2016a, 2016b). Furthermore, these RAVs could be maintained after treatment is withdrawn and be readily transmitted between individuals (Dvory-Sobol et al. 2015).

Our results suggest that in the NS5B region, the A218S variant acts as a permissive mutation for the C316N RAV and in particular, the A218S variant seeds the C316N RAV (see Fig. 4). This relationship could confer a higher probability of HCV evolving the

NS5B subtype 1b codons 316 and 218

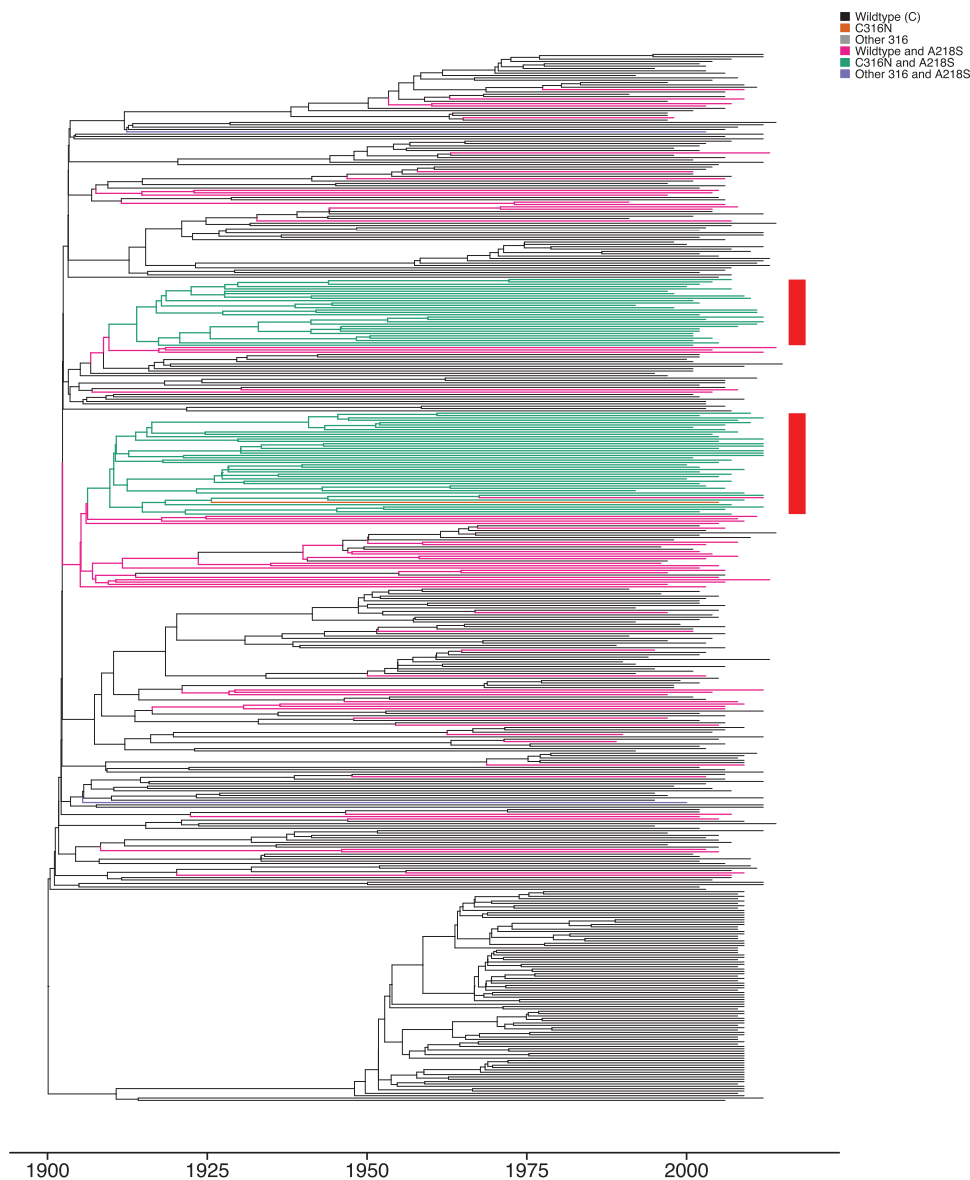


Figure 4. The phylogenetic history of the C316N and A218S variants in NS5B subtype 1b. An MCC tree shows the phylogenetic history of the global HCV NS5B population in subtype 1b. The tree was scaled to units of time with dates in years CE. The edges of the tree are coloured by the inferred states of codons 316 and 218 in NS5B of the child of the edge. Note the two clades of green variants with both the C316N and A218S variants—red bars mark these clades. These clades are each contained in separate clades of A218S variants.

C316N variant in patients on therapy with dasabuvir if the A218S variant is present and thus lead to resistance and ultimately treatment failure. Future work should investigate the EC₅₀ fold-shift on dasabuvir with respect to the A218S variant in combination and apart from the C316N variant. We recommend that the effect of the A218S variant on the fitness of the C316N RAV and mutability of codon 316 in the presence of dasabuvir be explored *in vivo*. Future research would also profitably focus on the interaction between the C316N and S556G RAVs in NS5B. Our analyses reveal a high prevalence and strong association of these two RAVs in HCV subtype 1b. This is clinically significant as this combination exhibits a 38× EC₅₀ fold-shift for dasabuvir compared to the 5× EC₅₀ fold-shift of C316N and the 11× EC₅₀ fold-shift of S556G in subtype 1b (Dietz et al. 2015).

Overall, the NS5A and NS5B trees of each subtype have roots that date to the beginning of the twentieth century. Between regions the subtype 1a and 1b trees have root dates that agree within 10 years. However, there is disagreement between the geographic origin of subtype 1b in the NS5A and NS5B trees. The NS5A origin is in Europe and the NS5B origin is in the Americas. Considering that the most sampled continent was always the inferred originating continent of the tree this discrepancy is likely due to oversampling of locations. There is no discordance in the geographic origin of subtype 1a between NS5A and NS5B. To account for the oversampling we ran our pipeline through multiple trial runs subsampling the sequences to obtain a more representative distribution of continents sampled. All of our subsampling trials in subtype 1a agreed with our main results

in that the roots of the majority of trees were inferred to be American. However, our NS5A subtype 1b trials produced trees whose roots were in the Americas and our NS5B subtype 1b trials produced trees whose roots were in Europe—the opposite of our main result. Due to this we conclude that the geographic origin of subtype 1b ancestor could not be resolved (see [Supplementary Materials](#)).

In conclusion, we found that NS5A RAVs, when found in a subtype, were widely dispersed geographically and phylogenetically, in contrast to the behaviour of the Q80K mutation in NS3. The RAVs in NS5B subtype 1a were scarce; however, C316N NS5B RAVs in subtype 1b formed two large clades and their evolution appears to be facilitated by the A218S variant in NS5B.

Acknowledgement

We would like to thank Rosemary McCloskey for assistance with the software.

Data availability

Data are available through GitHub repository: <https://github.com/brj1/HCVRAVOrigins>.

Supplementary data

[Supplementary data](#) are available at [Virus Evolution](#) online.

Conflict of interest: None declared.

References

- Black, S. et al. (2015) 'Resistance Analysis of Virologic Failures in Hepatitis C Genotype 1 Infected Patients Treated with grazoprevir/elbasvir +/- ribavirin: The C-worthy Study', *The International Liver Congress 2015: 50th Annual meeting of the European Association for the Study of the Liver*, Vienna, Austria, abstract P0891. doi: 10.1016/S0168-8278(15)31094-1.
- Bouckaert, R. et al. (2014) 'BEAST 2: A Software Platform for Bayesian Evolutionary Analysis', *PLoS Computational Biology*, 10: e1003537.
- Bouckaert, R. R. (2010) 'DensiTree: Making Sense of Sets of Phylogenetic Trees', *Bioinformatics*, 26: 1372–3.
- Cheng, G. et al. (2013) 'GS-5816, a Second Generation HCV NS5A Inhibitor with Potent Antiviral Activity, Broad Genotypic Coverage and a High Resistance Barrier', *The International Liver Congress 2013: 48th Annual meeting of the European Association for the Study of the Liver*, Amsterdam, Netherlands, abstract P1191. doi: 10.1016/S0168-8278(13)61192-7.
- Cock, P. J. A. et al. (2009) 'Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics', *Bioinformatics*, 25: 1422–3.
- Cuypers, L. et al. (2017) 'Implications of Hepatitis C Virus Subtype 1a Migration Patterns for Virus Genetic Sequencing Policies in Italy', *BMC Evolutionary Biology*, 17: 70.
- Dietz, J. et al. (2015) 'Consideration of Viral Resistance for Optimization of Direct Antiviral Therapy of Hepatitis C Virus Genotype 1-Infected Patients', *PLoS One*, 10: e013495.
- Drummond, A. J. et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29: 1969–73.
- Dvory-Sobol, H. et al. (2015), 'Long-term persistence of HCV NS5A variants after treatment with NS5A inhibitor ledipasvir', *The International Liver Congress 2015: 50th Annual Meeting of the European Association for the Study of the Liver*, Vienna, Austria, abstract O059. doi: 10.1016/S0168-8278(15)30073-8.
- Fridell, R. A. et al. (2011) 'Genotypic and Phenotypic Analysis of Variants Resistant to Hepatitis C Virus Nonstructural Protein 5A Replication Complex Inhibitor BMS-790052 in Humans: In Vitro and in Vivo Correlations', *Hepatology*, 54: 1924–35.
- Gao, M. (2013) 'Antiviral Activity and Resistance of HCV NS5A Replication Complex Inhibitors', *Current Opinion in Virology*, 3: 514–20.
- Gentile, I., Buonomo, A. R., and Borgia, G. (2014) 'Ombitasvir: A Potent Pan-Genotypic Inhibitor of NS5A for the Treatment of Hepatitis C Virus Infection', *Expert Reviews in Anti-Infection Therapy*, 12: 1033–43.
- Hernandez, D. et al. (2013) 'Natural Prevalence of NS5A Polymorphisms in Subjects Infected with Hepatitis C Virus Genotype 3 and Their Effects on the Antiviral Activity of NS5A Inhibitors', *Journal of Clinical Virology*, 57: 13–8.
- Hezode, C. et al. (2016). 'Resistance Analysis in 1284 Patients with Genotype 1 to 6 HCV Infection Treated with Sofosbuvir/Velpatasvir in the Phase 3 Astral-1, Astral-2, Astral-3 and Astral-4 Studies', *EASL, The International Liver Congress 2016*, Barcelona, Spain, abstract THU-216. doi: 10.1016/S0168-8278(16)00629-2.
- Jacobson, I. M. et al. (2015), 'Prevalence and Impact of Baseline NS5A Resistance Associated Variants (RAVs) on the Efficacy of Elbasvir/Grazoprevir (EBR/GZR) Against GT1a Infection', *The 66th Annual Meeting of the American Association for the Study of Liver Diseases: The Liver Meeting 2015*, San Francisco, California, abstract LB-22. doi: 10.1002/hep.28313.
- Jones, B. R., and Poon, A. F. Y. (2017) 'Node.dating: Dating Ancestors in Phylogenetic Trees in R', *Bioinformatics (Oxford, England)*, 33: 932–4.
- Kati, W. et al. (2015) 'In Vitro Activity and Resistance Profile of Dasabuvir, a Nonnucleoside Hepatitis C Virus Polymerase Inhibitor', *Antimicrobial Agents and Chemotherapy*, 59: 1505–11.
- Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Microbiology and Evolution*, 30: 772–80.
- Koev, G. et al. (2015), 'Characterization of Resistance Mutations Selected In Vitro by the Non-nucleoside HCV Polymerase Inhibitors ABT-333 and ABT-072', *The International Liver Congress 2009: 44th Annual Meeting of the European Association for the Study of the Liver*, Copenhagen, Denmark, abstract 953. doi: 10.1016/S0168-8278(09)60955-7.
- Krishnan, P. et al. (2015a) 'In Vitro and in Vivo Antiviral Activity and Resistance Profile of the Hepatitis C Virus NS3/4A Protease Inhibitor ABT-450', *Antimicrobial Agents and Chemotherapy*, 59: 979–87.
- et al. (2015b) 'Resistance Analysis of Baseline and Treatment-Emergent Variants in Hepatitis C Virus Genotype 1 in the AVIATOR Study with Paritaprevir-Ritonavir, Ombitasvir, and Dasabuvir', *Antimicrobial Agents and Chemotherapy*, 59: 5445–54.
- Lahser, F. C. et al. (2016a). 'Interim Analysis of a 3-Year Follow-up Study of NS5A and NS3 Resistance-Associated Variants (RAVs) After Treatment With Grazoprevir-Containing Regimens in Patients With Chronic Hepatitis C Virus (HCV) Infection', *The 67th Annual Meeting of the American Association for the Study of Liver Diseases: The Liver Meeting 2016*, Boston, Massachusetts, abstract 61. doi: 10.1002/hep.28796.
- et al. (2016b) 'The Combination of Grazoprevir, a HCV NS3/4A Protease Inhibitor, and Elbasvir, a HCV NS5A Inhibitor, Demonstrates a High Genetic Barrier to Resistance in HCV

- Genotype 1a Replicons', *Antimicrobial Agents and Chemotherapy*, 60: 2954–64.
- Lam, A. M. et al. (2012) 'Genotype and Subtype Profiling of PSI-7977 as a Nucleotide Inhibitor of Hepatitis C Virus', *Antimicrobial Agents and Chemotherapy*, 56: 3359–68.
- Lin, L. I.-K. (1989) 'A Concordance Correlation Coefficient to Evaluate Reproducibility', *Biometrics*, 45: 255–6.
- Liu, R. et al. (2015) 'Susceptibilities of Genotype 1a, 1b, and 3 Hepatitis C Virus Variants to the NS5A Inhibitor Elbasvir', *Antimicrobial Agents and Chemotherapy*, 59: 6922–9.
- Lontok, E. et al. (2015) 'Hepatitis C Virus Drug Resistance-Associated Substitutions: State of the Art Summary', *Hepatology*, 62: 1623–32.
- Ly, K. N. et al. (2016) 'Rising Mortality Associated with Hepatitis C Virus in the United States, 2003–2013', *Clinical Infectious Diseases*, 62: 1287–8.
- Maddison, W. P. (1990) 'A Method for Testing the Correlated Evolution of Two Binary Characters: Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic Tree?', *Evolution; International Journal of Organic Evolution*, 44: 539–57.
- McCloskey, R. M. et al. (2015) 'Global Origin and Transmission of Hepatitis C Virus Nonstructural Protein 3 Q80K Polymorphism', *The Journal of Infectious Diseases*, 211: 1288–95.
- Paradis, E., Claude, J., and Strimmer, K. (2004) 'APE: Analyses of Phylogenetics and Evolution in R Language', *Bioinformatics*, 20: 289–90.
- Poordad, F. et al. (2014) 'ABT-450/r-Ombitasvir and Dasabuvir with Ribavirin for Hepatitis C with Cirrhosis', *The New England Journal of Medicine*, 370: 1973–82.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) 'FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS One*, 5: e9490.
- Sarrazin, C. et al. (2014), 'Baseline and Post-baseline Resistance Analyses of Phase 2/3 Studies of Ledipasvir/Sofosbuvir +- RBV', *The 65th Annual Meeting of the American Association for the Study of Liver Diseases: The Liver Meeting 2014*, Boston, Massachusetts, abstract 1926. doi: 10.1002/hep.27533.
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies', *Bioinformatics (Oxford, England)*, 30: 1312–3.
- Struck, D. et al. (2014) 'COMET: Adaptive Context-based Modeling for Ultrafast HIV-1 Subtype Identification', *Nucleic Acids Research*, 42: e144.
- Wang, C. et al. (2013) 'In Vitro Activity of Daclatasvir on Hepatitis C Virus Genotype 3 NS5A', *Antimicrobial Agents and Chemotherapy*, 57: 611–3.
- Wong, K. A. et al. (2013) 'Characterization of Hepatitis C Virus Resistance from a Multiple-Dose Clinical Trial of the Novel NS5A Inhibitor GS-5885', *Antimicrobial Agents and Chemotherapy*, 57: 6333–40.
- Yu, G. et al. (2017) 'GGTREE: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data', *Methods in Ecology and Evolution*, 8: 28–36.