Global origins of resistance-associated variants in the NS5A region of HCV

24th CROI – **Abstract 561**
February 13-16, 2017
Seattle, WA, USA

Bradley R. Jones[1], **Anita Y.M. Howe**[1], **P. Richard Harrigan**[1,2], **Jeffrey B. Joy**[1,2]

1. BC Centre for Excellence in HIV/AIDS, Vancouver, Canada; 2. University of British Columbia, Vancouver, Canada

Bradley R Jones
bjones@cfenet.ubc.ca

Background

HCV contains a variety of naturally occurring substitutions in the non-structural proteins 3, 5A and 5B (NS3, NS5A and NS5B) that reduce susceptibility to currently approved therapies targeting these regions of the HCV genome. A previous study [1], revealed that most HCV variants with the Q80K resistance mutation in NS3 had a **single common ancestor** in North America and the Q80K mutation was strongly coupled with compensatory mutations at codons 91 and/or 174 (Figure 1). A recent study [2] extended this to the NS5A and NS5B regions; however no geographic context was provided. Here we analyze, using all available HCV sequence data, from treatment naive individuals in the global databases, the phylogenetic and geographic origins and frequency of evolution of NS5A/B RASs.

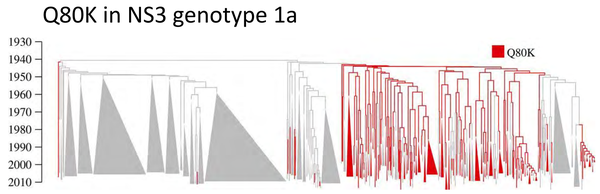


Figure 1: The phylogenetic history of the Q80K resistance mutation in the NS3 region of HCV genotype 1a. The Q80K resistance mutation in NS3 of HCV genotype 1a was found to have a global origin in North America as well as strong associations with the substitutions, A91S/T and S174N. The figure shown here is taken from [1].

Methods

We collected 200,863 HCV sequences from GenBank on August 30, 2016 (all but 2 of these samples predate the implementation of NS5A and NS5B inhibitors for treatment). After trimming sequences to the NS5A and NS5B regions and retaining only sequences from unique individuals with at least 75% coverage of the reference sequence and were genotype 1a, 1b or 3a, we were left with 1,498 N5SA sequences and 1,029 NS5B sequences.

Previously observed RASs were curated from the literature for the NS5A inhibitors: daclatasvir, ledipasvir, velpatasvir ombitasvir, and elbasvir and the NS5B inhibitors: dasabuvir and sofosbuvir. In total we considered 32 NS5A RASs and 26 NS5B RASs (See Tables 1 and 2).

Phylogenetic trees were constructed using FastTree v2.1.7 [3], then rooted and time-scaled using the ape v4.0 [4,5] package in R. We estimated ancestral states using a MG94xcREV codon model in HPHPY v2.22b [6,7] and tested for substitution association using Fisher's exact test in R. For each RAS we modeled the country of origin using a continent-aware model using R.

All analyses were compared amongst 100 replicate runs with no significant variation detected. All phylogenetic trees were plotted and displayed using the R package ggtree [8].

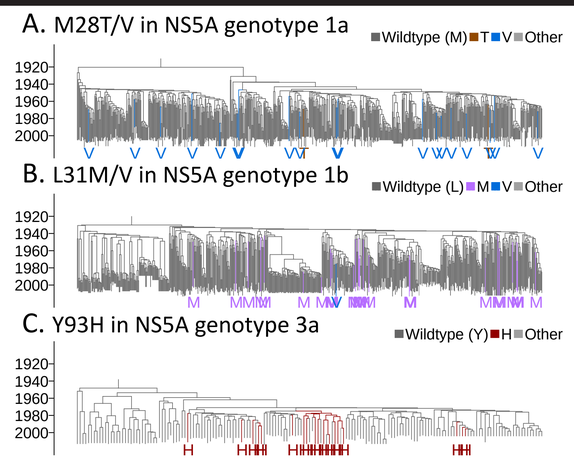


Figure 2: The phylogenetic history of (A) M28/T/V in genotype 1a; (B) L31M/V in genotype 1b; and (C) Y93H in genotype 3a in HCV NS5A. Each tree represents a time-scaled phylogeny of the sequences from a particular genotype. The edges are coloured according to the inferred codon at a given position in the sequence at the bottom of the edge. The RASs in the NS5A region of HCV were found to be widely dispersed throughout the phylogenetic trees, with few coalescing RASs in genotypes 1a and 1b. In genotype 3a (see Figure 2C), however there were small clades of RASs, but the sample size was also smaller.

Table 1: RASs not found in GLOBAL HCV NS5A/NS5B. NS5A /NS5B RASs that we looked for, but did not find in our data set. The EC50 Fold Shift represents the range for daclatasvir, lepidasvir, velpatasvir ombitasvir, and elbasvir for NS5A RASs and dasabuvir for NS5B RASs, except S282T (sofosbuvir).

Region / Genotype	Mutation	EC50 Fold-Shift
NS5A 1a	M28A	61 – 4591 x
	M28G	1000 – 71429 x
	Q30D	925 x
	Q30E	37 – 7500 x
	Q30G	84 – 1000 x
	Q30K	183 – 24545 x
	L31F	96 x
	L31I	6 – 200 x
	Y93S	500 – 1013 x
NS5A 1b	L28T	661 x
	L31F	2 – 15 x
	Y93S	12 x
NS5A 3a	M28T	659 x
	L31F	28 – 603 x
	L31M	330 – 537 x
	L31V	1181 x

Region / Genotype	Mutation	EC50 Fold-Shift
NS5B 1a	S282T*	13 x
	C316Y	1472 x
	A395G	20 x
	M414I	17 x
	M414T	32 x
	M414V	18 x
	M444K	23 x
	Y448C	975 x
	S553T	152 x
	G554S	198 x
NS5B 1b	S282T*	7.8 x
	C316Y	1569 x
	N411S	54 x
	N414T	26 x
	N414V	18 x
	Y448C	160 x
	A553V	58 x
	D559G	100 x

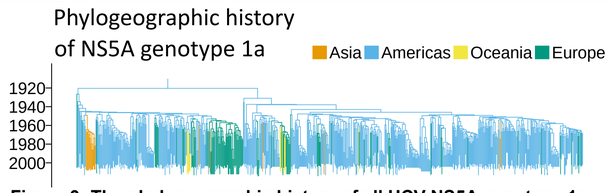


Figure 3: The phylogeographic history of all HCV NS5A genotype 1a sequences. The edges are coloured according to the inferred continent of origin at the bottom of the edge. The RASs of the NS5A region were found to be evenly distributed throughout the world as well, with the correlation between the geographic location of each RAS and the geographic location of wild type variants ranging from 0.939 and 0.998.

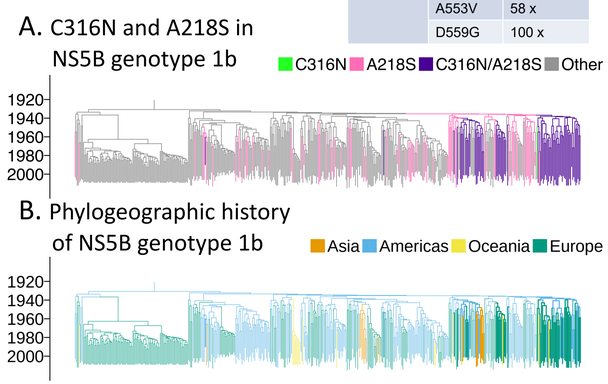


Figure 4: (A) The associated phylogenetic histories of C316N and A218S in HCV NS5B genotype 1b; (B) the phylogeographic history of HCV NS5B 1b (Bottom). The C316N RAS in NS5B was found in 3 large clades in the global phylogeny of HCV genotype 1b. When compared with the substitution, A218S, we see that the 3 clades are contained in a larger clade of A218S variants. The association between these substitutions is quite strong ($p = 2.8 \times 10^{-45}$). We also found a strong association of C316N with the RAS S556G ($p = 3.1 \times 10^{-8}$). Analyses suggest 2 of the clades of C316N variants have geographic origins in the Americas (USA) and the other clade has a geographic origin in Europe (Switzerland); the edges with predicted C316N substitutions are emboldened in Figure 4B. We also looked at NS5B RASs in genotype 1a (See Tables 1 and 2, no tree shown).

Region / Genotype	Mutation	EC50 Fold-Shift	Prevalence	Number of Isolates	Number of Clades	Largest Clade
NS5A 1a	M28T	5 – 8965 x	0.30% (2)	2	0	N/A
	M28V	0.6 – 58 x	2.9% (19)	19	0	M/A
	Q30H	2 – 435 x	1.1% (7)	7	0	N/A
	Q30R	2 – 800 x	0.15% (1)	1	0	N/A
	L31M	**2 – 141 x**	**1.4% (9)**	**6**	**1**	**3**
	L31V	61 – 1000 x	0.30% (2)	2	0	N/A
	P32L	44 – 100 x	0.15% (1)	1	0	N/A
	H58D	6 – 1127 x	0.45% (3)	3	0	N/A
	Y93C	11 – 1675 x	0.60% (4)	4	0	N/A
	Y93H	220 – 41383 x	0.60% (4)	4	0	N/A
	Y93N	100 – 66740 x	0.15% (1)	1	0	N/A
NS5A 1b	**L31M**	**0.9 – 10 x**	**3.7% (25)**	**19**	**3**	**2**
	L31V	1 – 50 x	0.15% (1)	1	0	N/A
	Y93H	**1.2 – 1325 x**	**3.3% (22)**	**20**	**1**	**2**
NS5A 3a	**A30K**	**14 – 50 x**	**3.0% (5)**	**1**	**1**	**6**
	Y93H	**28 – 6728 x**	**11% (18)**	**7**	**2**	**15**
NS5B 1a	Y448H	975 x	0.18% (1)	1	0	N/A
	S556G	30 x	0.91% (5)	5	0	N/A
	S556R	261 x	0.18% (1)	1	0	N/A
NS5B 1b	**C316N**	**5 x**	**15% (74)**	**5**	**3**	**42**
	S368T	110 x	0.21% (1)	1	0	N/A
	M414I	15 x	0.21% (1)	1	0	N/A
	Y448H	37 x	0.21% (1)	1	0	N/A
	S556G	**100 x**	**8.9% (43)**	**32**	**2**	**9**

Table 2: The prevalence of RASs in HCV NS5A/NS5B. The EC50 Fold Shift represents the range of EC50 fold-shifts for daclatasvir, lepidasvir, velpatasvir ombitasvir, and elbasvir NS5A RASs and the EC50 fold-shift for dasabuvir for NS5B RASs. The number of clades are the are number of groups of individuals with a predicted RAS ancestor in the drug-naïve population. The number of isolates were the number of individual sequences with an RAS that were not a member of a clade. Largest clade is given in the number of tips in the clade. Rows in bold represent RASs with observed clades.

Conclusions

- Given that RASs in NS5A are widely spread throughout their phylogenetic trees with no common ancestor, we predict that these RAS have a low fitness-barrier and therefore can easily arise during therapy on NS5A inhibitors.
- Similar to Q80K in NS53, NS5B has three large clades of C316N variants and the C316N RAS is highly associated with a presumably compensatory mutation, A218S.
- C316N is also highly associated with the RAS S556G, which combined with C316N has a 38x EC50 fold-shift for dasabuvir compared to the 5x EC50 fold-shift of C316N and 11x EC50 fold-shift of S556G alone [9].

1. McCloskey, Liang, *et al. J Infec Dis* **211**:1288-95 (2015). 2. Patiño-Galindoa, Salvatierra, *et al. Antimicrob Agents Chemother* **60**:2402-16 (2016). 2. Price, Dehal, *et al. PLoS One* **5**:e9490 (2010). 3. Paradis, Claude, *et al. Bioinformatics* **20**:289-90 (2004).4. Jones and Poon. *Bioinformatics* (2016). accepted. doi:10.1093/bioinformatics/btw744 5. Pond, Frost, *et al. Bioinformatics* **21**:676-9 (2005). 6. Pond and Muse. *Mol Biol Evol* **22**:2375-85 (2005). 7. Yu, Smith, *et al. Methods Ecol Evol* **8**:28-36 (2017). 8. Dietz, Susser, *et al. PLoS One* **10**:e013495 (2015).