P-N4 656LB

Conference on Retroviruses and Opportunistic Infections
Boston, MA
March 3-6

Global Origin of HCV NS3 Substitution Q80K that is Associated with Lower Simeprevir Susceptibility

680-1081 Burrard Street
Vancouver, BC
CANADA V6Z 1Y6

apoon@cfenet.ubc.ca
(604) 682-2344 x66465

Art FY Poon[1], Richard H Liang[1], Jeffrey B Joy[1], Rosemary M McCloskey[1], Thuy Nguyen[1], Mel Krajden[2], Julio SG Montaner[1], and P Richard Harrigan[1]

[1]BC Centre for Excellence in HIV/AIDS, [2]BC Centre for Disease Control

Rationale

The mutation Q80K in the HCV protease gene NS3 is associated with reduced susceptibility to the direct-acting antiviral (DAA) inhibitor Simeprevir (Zeuzem et al. 2013).

Q80K is predominantly observed in HCV genotype 1a, especially in the United States (~50%, de Luca et al. 2013). It is relatively rare in Europe, and seldom observed in other genotypes.

Why is Q80K so prevalent in HCV 1a?

Did this mutation historically arise once or on multiple occasions?

Is Q80K stable and transmissible between hosts?

We performed a phylogenetic analysis of HCV NS3 sequence variation to reconstruct the evolutionary origins of this mutation.

Zeuzem, Stefan, Thomas Berg, Edward Gane, Peter Ferenci, Graham R. Foster, Michael W. Fried, Christophe Hezode et al. Simeprevir increases rate of sustained virologic response among treatment-experienced patients with HCV genotype-1 infection: a phase IIb trial. *Gastroenterology* 146, no. 2 (2014): 430-441.

De Luca, A, *et al.* Two distinct HCV genotype 1a clades: geographical distribution and association with natural resistance mutations to HCV NS3/4A inhibitors. *International Workshop on HIV & Hepatitis Virus Drug Resistance and Curative Strategies.* June 4-8 2013, Toronto, Canada.

Methods

We retrieved all published HCV NS3 sequences from the Genbank database (n=10,764) and used pair-wise alignment against a reference (H77) to extract the first 543 nucleotides.

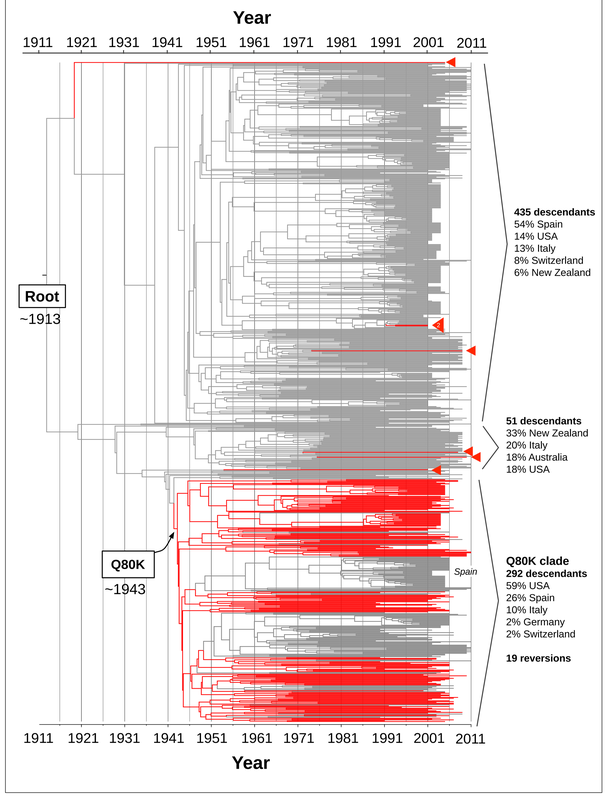
We used an approximate maximum likelihood (ML) phylogeny (Fasttree2) of these data to screen for genotype 1a (n = 5,021). Sequences were annotated by country and year of sample collection, or discarded if these data were not available.

In the absence of consistent source annotation, repeated samples from the same individual were filtered by generating a multiple sequence alignment (MUSCLE), reconstructing an ML phylogeny and pruning highly similar sequences by a branch length cutoff of 0.01 expected nucleotide substitutions per site. A ML phylogenetic tree was built from a multiple sequence alignment of the remaining *n* = 794 sequences.

We rooted this tree under a strict clock based on sample dates using a modified version of Path-O-Gen. Ancestral divergence times were estimated using a penalized likelihood method using a modified version of the *chronos* function in the R *ape* library.

We fit a Muse-Gaut model of codon substitution crossed with the generalized time-reversible model of nucleotide substitution to the alignment and tree in *HyPhy*. Amino acid substitutions were mapped to branches of the tree by joint ML ancestral sequence reconstruction.

Strict clock ML tree relating *n* = 794 HCV 1a NS3 sequences. Node heights were estimated using a penalized likelihood method. Lineages descending from ancestral Q80K mutations are indicated in **red**. This phylogeny was robust to censoring NS3 codon Q80.



Results

- Q80K was observed in 172 out of 794 (22%) HCV 1a NS3 sequences.
- The 1a phylogeny comprised two clades with different global distributions as previously described (de Luca *et al.* 2013).
- Root of phylogeny implied that genotype 1a emerged in the early 1910's, consistent with previous work.
- Ancestral sequence reconstructed at root encoded wild-type Q at codon 80.
- We mapped only 7 ancestral Q80K substitution events in the phylogeny.
- One Q80K substitution mapped deep in a predominantly US-based clade, in a lineage ancestral to 292 observed sequences (165 of which inherited this Q80K).

Conclusions

The vast majority (96%) of HCV 1a carrying Q80K have descended from a single substitution event from over 50 years ago.

The persistence of this ancestral substitution implies that Q80K is readily transmitted between hosts.

This evolutionary history may explained currently observed differences in the prevalence of Q80K between European and US populations.

This work was supported by an operating grant from the Canadian Institutes of Health Research (CIHR, HOP 111406) and by Genome British Columbia. AFYP is supported by a New Investigator Award from CIHR (Canadian HIV Vaccine Initiative for Vaccine Discovery and Social Research) and a Career Investigator Scholar Award from the Michael Smith Foundation for Health Research / St. Paul's Hospital Foundation - Providence Health Care Research Institute partnership. PRH is supported by a CIHR/GSK Research Chair in Clinical Virology.