Phylogenetic estimation of the temporal spread of hepatitis C genotype 1a in North America

Jeffrey B. Joy[1]   Rosemary M. McCloskey[1]   Thuy Nguyen[1]   Richard H. Liang[1]   Yury Khudyakov[2]
Andrea Olmstead[3]   Mel Krajden[3,4]   John Ward[2]   P. Richard Harrigan[1,5]   Julio S. G. Montaner[1,5]
Art F. Y. Poon[1,5]
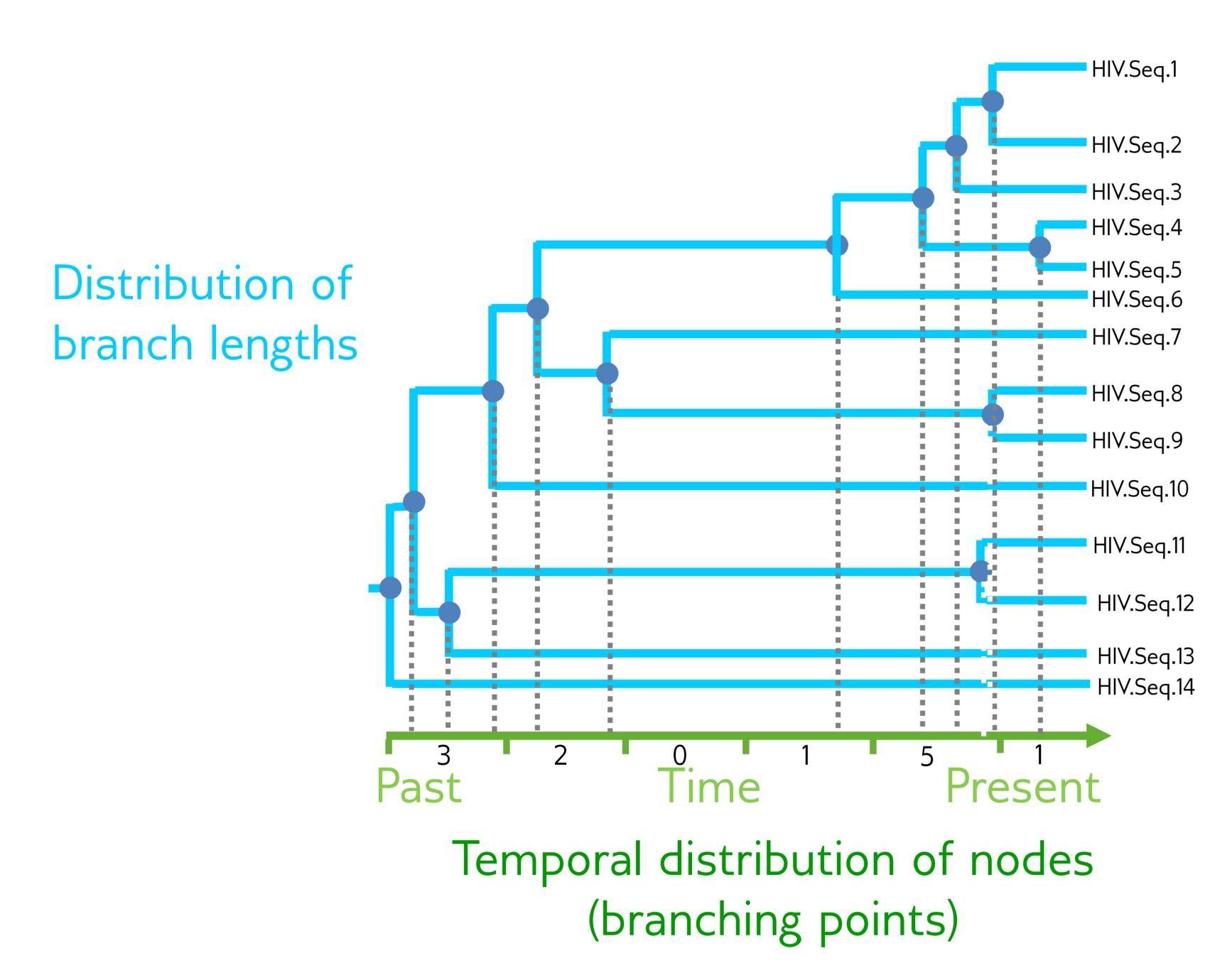
IAS, 2015
Vancouver, Canada

[1]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada [2]Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA, United States [3]BC Centre for Disease Control, Vancouver, Canada [4]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada [5]Department of Medicine, University of British Columbia, Vancouver, Canada

1081 Burrard Street
Vancouver, BC
CANADA V6Z 1Y6

jjoy@cfenet.ubc.ca
(604) 682-2344 x62466

Background

- 4.6 million hepatitis C virus (HCV) infections in North America, predominantly genotype 1a[1].
- Majority (75%) of infected individuals were born between 1946 and 1964.[1]
- Explanation for this high prevalence is still contested: often attributed to past sporadic behavioural risk (i.e.: experimentation with injection drug use, unsafe tattooing, high risk sex, travel to high endemic areas).
- Previous studies have implicated infected blood products and injection drug use.
- **Can we estimate the timing of the HCV 1a epidemic in North America using HCV RNA sequences?**
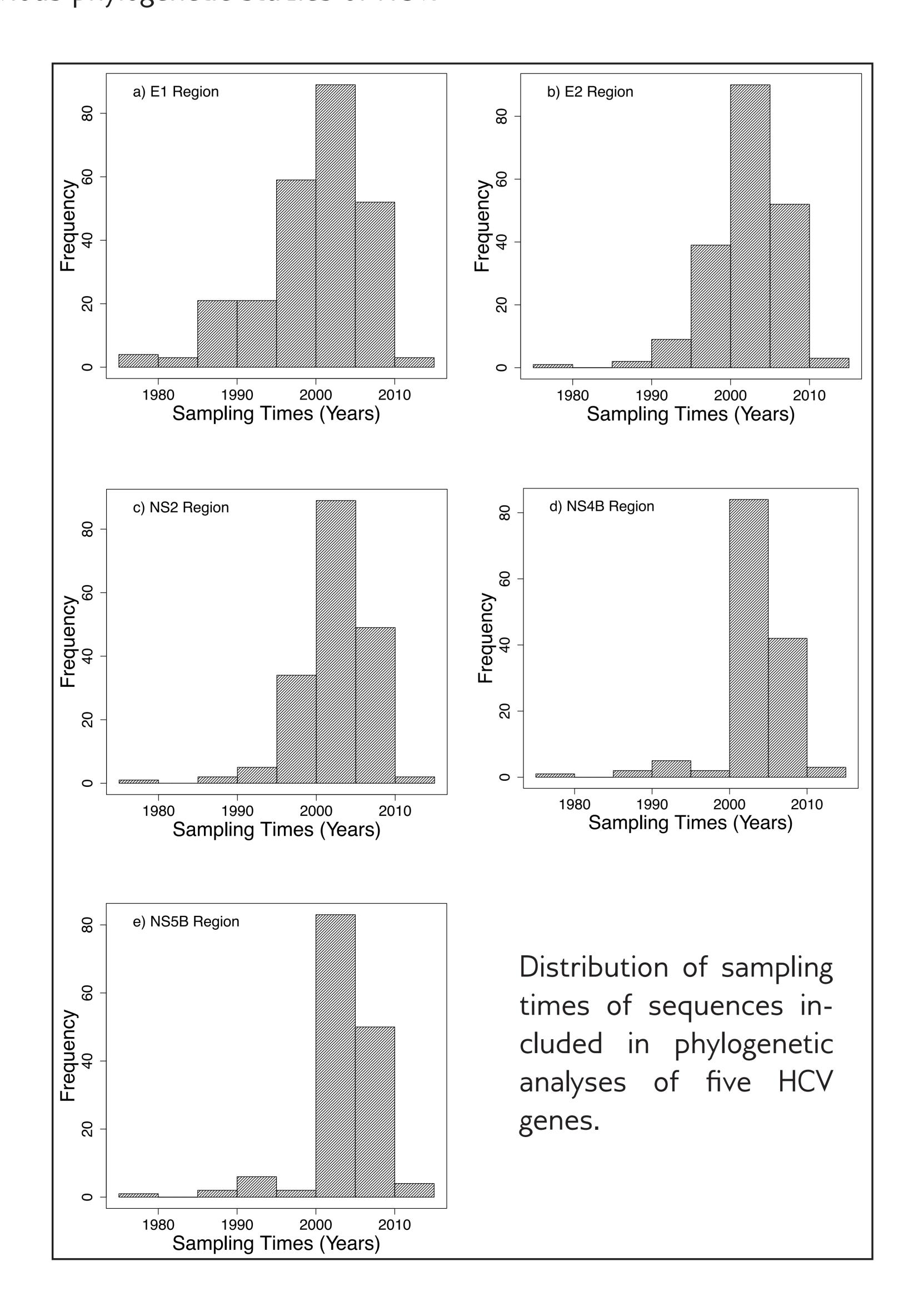
Phylodynamic methods use virus phylogenies to estimate epidemiological parameters

- RNA viruses evolve rapidly and demonstrate significant within-host diversification.
- Transmission isolates a sub-population of virions in a new host, resulting in a branching point in the phylogeny.
- By assuming a molecular clock, speciation events in a viral phylogeny can be used to approximate dates of transmission events.
- Epidemiological information about an epidemic can therefore be inferred from the virus phylogeny.
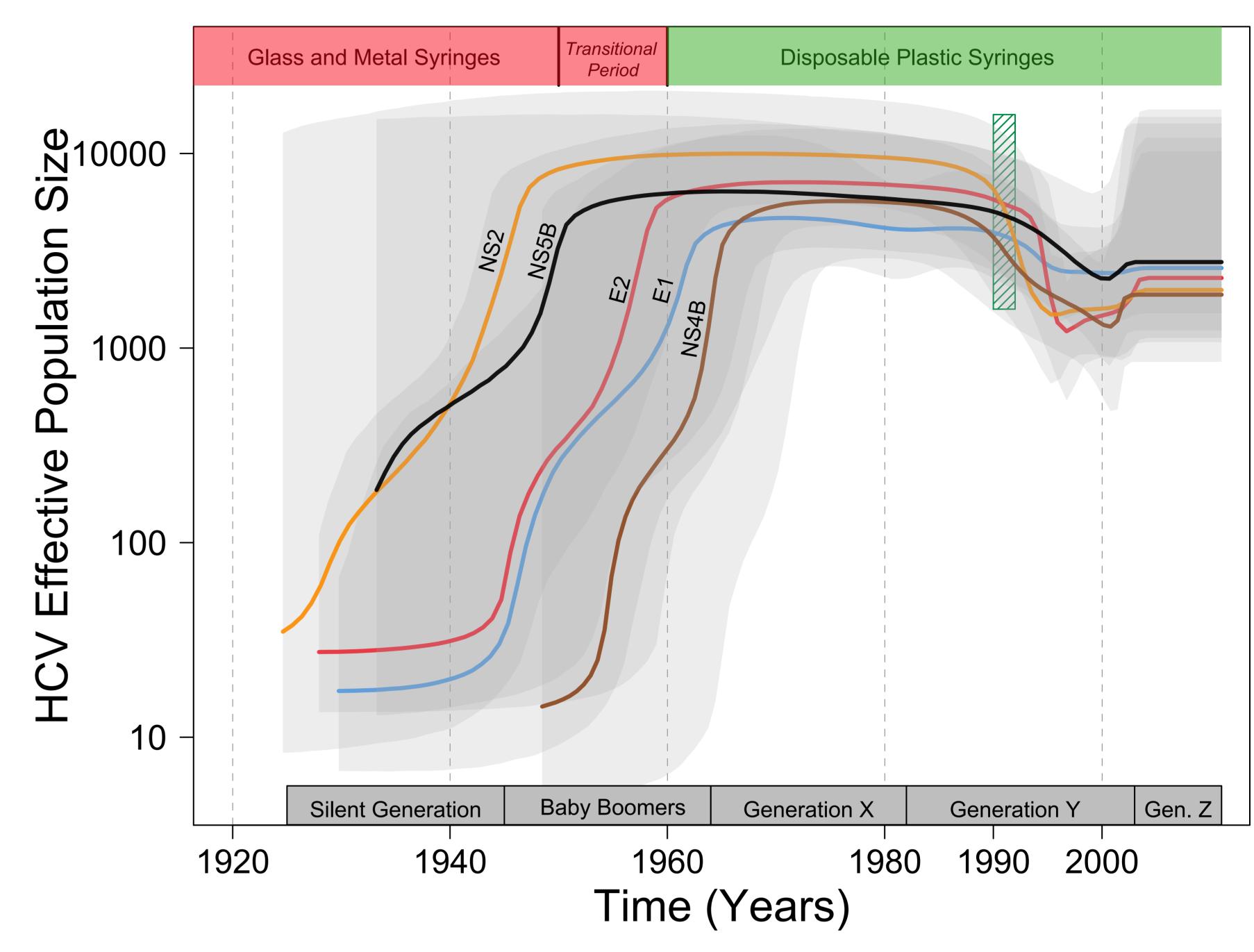

Data collection and processing

- All publicly available HCV sequences with annotated sampling dates were downloaded from GenBank and aligned to HCV reference H77.
- Sequences from the same patient were removed by a phylogenetic distance cutoff, followed by manual curation.
- Sequence data sets were compiled for five regions of the HCV genome: E1, E2, NS2, NS4B, and NS5B.
- Emphasized collection of early sequences, which have not been included in previous phylogenetic studies of HCV.

Distribution of sampling times of sequences included in phylogenetic analyses of five HCV genes.

Changes in HCV effective population size align with blood product screening and transition to disposable syringes
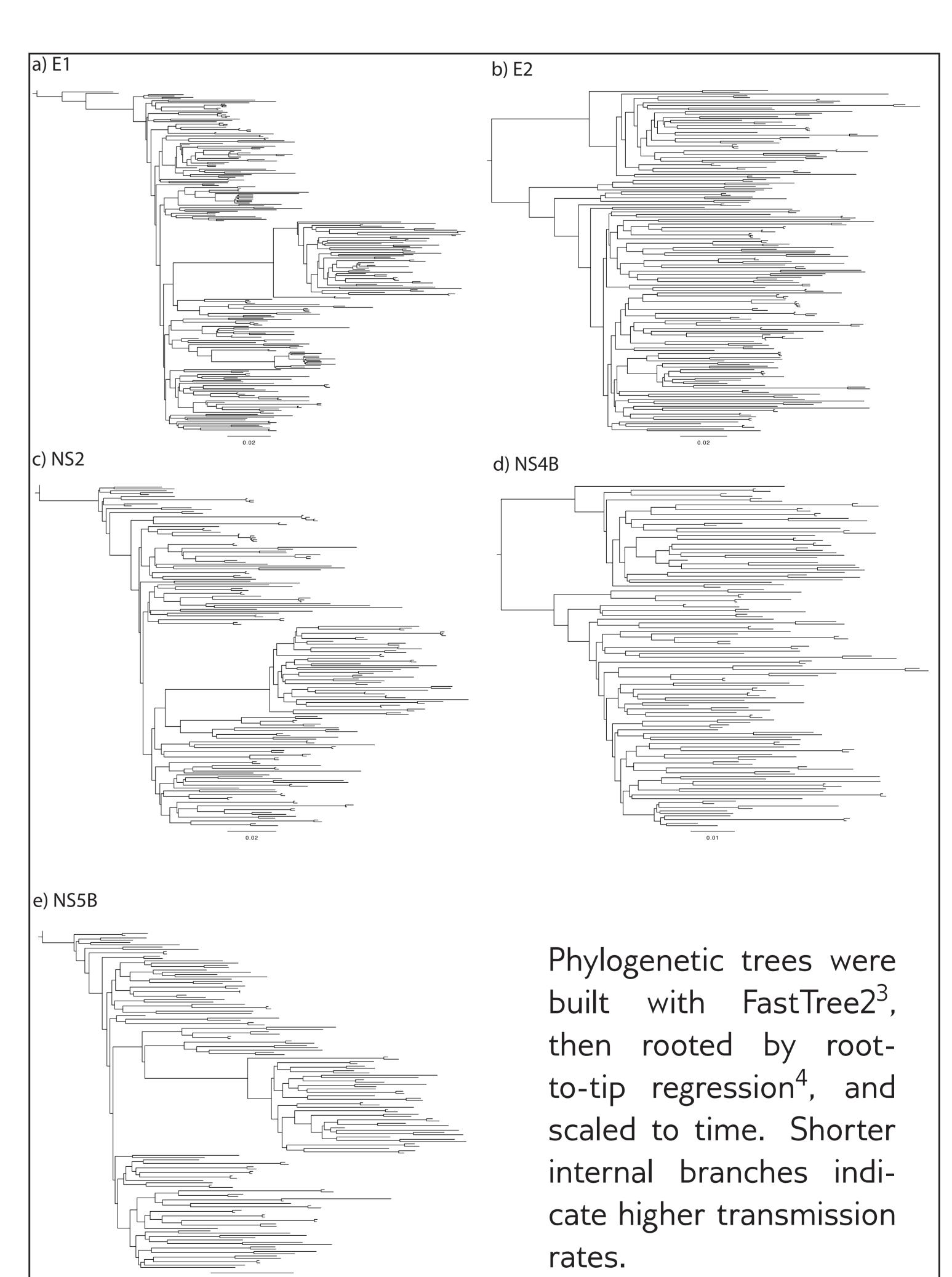
- The *effective number of infections* in an epidemic is:
  - expected to be proportional to the prevalence during the exponential growth phase[5], and
  - estimated using Bayesian skyline plots with a relaxed molecular clock, implemented in BEAST[2].
- All five genes indicate a period of exponential growth in the 1940s-60s, stability until the early 1990s, followed by a shorter decline.
- End of exponential growth period was co-incident with rollout of disposable plastic syringes in medical practice.
- Decline in effective number of infections occurred following the introduction of blood product screening[6] and widespread adoption of needle exchange programs[7] (green box).

Estimated effective number of HCV infections from early 20th century until present day. Grey regions are 95% highest probability density. Green hatched box indicates introduction of blood product screening[6] and widespread adoption of needle exchange programs[7] in North America.

Maximum likelihood phylogenies of five HCV genes confirm correctly filtered data set

- Few closely related groups of tips, indicating successful screening of sequences from the same patient.
- Lack of phylogenetic outliers (long terminal branches) indicates all included sequences were correctly identified as HCV genotype 1a.

Phylogenetic trees were built with FastTree2[3], then rooted by root-to-tip regression[4], and scaled to time. Shorter internal branches indicate higher transmission rates.

Conclusions: end of exponential HCV spread was co-incident with improvements in medical practice

- The end of the exponential growth period of the HCV genotype 1a epidemic in North America was contemporary with the introduction of disposable syringes.
- Subsequent improvements in medical practice led to declines in prevalence.
- The prevailing view that the North American HCV epidemic is predominantly attributable to past sporadic behavioural risk is not supported by our data.
- Early transmission dynamics are consistent with nosocomial and iatrogenic causes as drivers for early HCV 1a spread in North America.

[1]Cohen, Jon. "Calling All Baby Boomers: Get Your Hepatitis C Test." Science 337.6097 (2012): 903-903. [2]Drummond, Alexei J., and Andrew Rambaut. "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evolutionary Biology 7.1 (2007): 214. [3]Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. "FastTree 2 - approximately maximum-likelihood trees for large alignments." PLoS ONE 5.3 (2010): e9490. [4]Rambaut, Andrew. "Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies." Bioinformatics 16.4 (2000): 395-399. [5]Frost, Simon DW, and Erik M. Volz. "Viral phylodynamics and the search for an ﬁeffective number of infectionsﬁ." Philosophical Transactions of the Royal Society B: Biological Sciences 365.1548 (2010): 1879-1890. [6]Preiksaitis JK, Rivet C. "A history of the evolution of hepatitis C testing of blood donors: implications for the Canadian blood supply system." Transfusion 1995;35:348-52. [7]Des Jarlais DC, Perlis T, "Arasteh K, et al. Reductions in hepatitis C virus and HIV infections among injecting drug users in New York City, 1990-2001." AIDS 2005;19:S20-S5.