Robust phylogenetic method to reconstruct dates of infection in HIV-1 seroconverters from different risk groups

Rachel A McGovern¹, Zabrina L Brunne¹,², Calin Johnson¹, Robert Hogg¹,², Evan Wood¹, Thomas Kerr¹, Mc Milloy¹, P Richard Harrigan¹, Art FY Poon², and the Vanguard and VIDUS/ACCESS cohort studies

¹BC Centre for Excellence in HIV/AIDS, Vancouver, Canada; ²Faculty of Health Sciences, Simon Fraser University

RATIONALE

• Dates of HIV infection are important data for measuring HIV incidence.
• These dates are often unknown due to the long latency of HIV infection and barriers to HIV testing.
• We previously described a phylogenetic method for estimating dates of HIV infection from the within-host evolution of HIV (Poon et al. 2011).
• The most recent common ancestor (MRCA) of an infection can tend to coincide with the date of infection because of the HIV transmission bottleneck:

DATA COLLECTION

• Identified N = 125 HIV seroconverters from two prospective HIV cohorts, the Vancouver Injection Drug User Study (VIDUS1), and the Vanguard study of young MSM.
• HIV seroconversion dates estimated as midpoint between last seronegative and first seropositive visits. The data analyst was blinded from these dates until the analysis was complete.
• Extracted HIV RNA from frozen blood plasma sampled from 2 time points per patient (baseline visit and a followup within 2 years of baseline).
• RT-PCR amplification products targeting up to 2 regions of the HIV genome (env C2-V3-C3 and nef) for deep sequencing on Roche/454 GS Junior.
• To date, collected data from N = 8 injection drug users (IDUs) and N = 12 MSM. Additional samples from N = 71 other HIV seroconverters from VIDUS are currently being processed.

DATA PROCESSING

We screened for cross-contamination among samples from the same GS Junior run using a custom Python script that computed the alignment score of each read to the most common variant in each sample, which assigns a +5 to a nucleotide match, a −4 to a mismatch and a −10 penalty to open a gap. If a read was considered to be a potential cross-contaminant if its average same-sample score was below 4 and its maximum between-sample score was above 4. A read was discarded if none of its scores exceeded 3. Manual inspection of these discarded reads indicated that they were consistently attributable to excessive indel error.
Each set of reads was screened for indel errors using a codon-based alignment algorithm in HyPhy (hy454) that detects frame-shifts caused by indels relative to a reference sequence (HXB2). These reads were subsequently annotated by sample dates and grouped into data sets by patient. We randomly sampled 100 sequences per time point from each patient and generated multiple sequence alignments using MUSCLE (version 3.8.31, http://www.drive5.com/muscle). These were analyzed using BEAST by generating random chain samples under strict and relaxed molecular clock models and a Tamura-Nei (TN93) model of nucleotide substitution with a 4-category discretized gamma distribution to model rate variation across sites.

RESULTS

Estimated dates of infection are concordant with known dates of HIV seroconversion

Each point represents an HIV seroconverter from the VIDUS (triangles) or Vanguard (circles) study cohorts. Known dates of HIV seroconversion are plotted along the x-axis. Mean estimates of dates of HIV infection, plotted along the y-axis, were based on median times of MRCA's in sampled phylogenies. These estimates were averaged over genomic regions.

The mean discordance between dates was ±1.2 months, indicating that phylogenetic estimates were not measurably biased. We found strong concordance between known and estimated dates of infection (Spearman ρs = 0.92, 95% CI = 0.82-0.97) where ρs = 1 indicates perfect concordance. There was no significant difference in concordance by mode of HIV transmission (Student’s t = −0.22, P = 0.83) or by HIV genome region (t = −1.4, P = 0.18).

REFERENCES


ACKNOWLEDGEMENTS

The authors wish to recognize the important contributions from the participants and staff in the VIDUS1 and Vanguard study cohorts. This work was supported by an NIH (NIDA) R01 awarded to TK (2R01DA01199-09). AFYP is supported by a Canadian Institutes of Health Research (CIHR) Operating Grant (2010-09-HOP-235256) and by a Michael Smith Foundation for Health Research (MSFHR) / St. Paul's Hospital Foundation-Providence Health Care Research Institute (SPH-PRHC) Career Investigator Scholar Award. PRH is supported by a CHIKNSK Research Chair in Clinical Virology.