**I-160**
Conference on Retroviruses and Opportunistic Infections
March 3-6 2013
Atlanta, USA

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

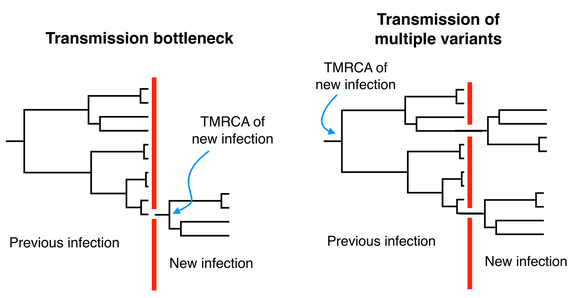
Rachel A McGovern[1], Zabrina L Brumme[1,2], Caitlin Johnson[1], Robert Hogg[1,2], Evan Wood[1], Thomas Kerr[1], MJ Milloy[1], P Richard Harrigan[1], Art FY Poon[1], and the Vanguard and VIDUS/ACCESS cohort studies
[1]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada; [2]Faculty of Health Sciences, Simon Fraser University

613-1081 Burrard St.
Vancouver, BC
CANADA V6Z 1Y6

apoon@cfenet.ubc.ca
(604) 682-2344 x66364

## 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**:




- Modes of transmission that bypass the mucosal barrier (*e.g.*, injection drug use) may introduce multiple HIV founder variants (Keele *et al.*, 2008, Bar *et al.*, 2010).
- Transmission of multiple HIV variants may disrupt the link between the MRCA and date of infection.
- **How robust are phylogenetic estimates of dates of HIV infection to modes of HIV transmission?**

## 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.

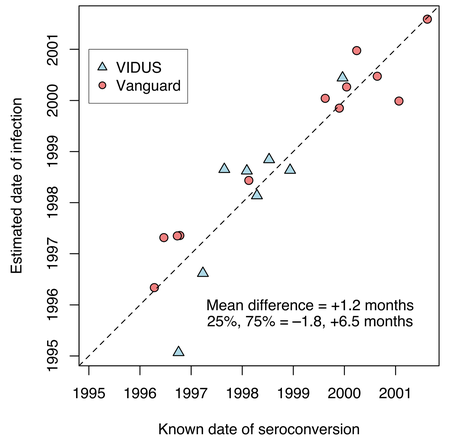
## DATA PROCESSING

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. 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 MRCAs 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 (Lin's $\rho_c = 0.92$, 95% C.I = 0.82-0.97) where $\rho_c = 1$ 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 ($t = -1.4$, $P = 0.18$).

## SUMMARY

- Phylogenetic estimates of dates of HIV infection appear to be robust to modes of HIV transmission.
- Suggests that variation in multiplicity of HIV has limited impact on phylogenetic inference.
- If not screened for, sample cross-contamination in a Roche/454 run (median rate = 0.25%, IQR = 0.14%-0.44%) may lead one to overestimate the prevalence of HIV superinfection.
- "Deep" sequencing only one region of the HIV genome (*env* C2-V3-C3) may be sufficient for accurate estimation of infection dates.

## REFERENCES

Bar, KJ, et al. *Wide variation in the multiplicity of HIV-1 infection among injection drug users.* J Virol 2010, 84: 6241-6247.
Keele, BF, et al. *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.* Proc Natl Acad Sci USA 2008, 105: 7552-7557.
Poon, AFY, et al. *Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data.* AIDS 2011, 25: 2019-2026.

## 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 (2R01DA011591-09). 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 (SPHF-PHRCI) Career Investigator Scholar Award. PRH is supported by a CIHR/GSK Research Chair in Clinical Virology.

BRITISH COLUMBIA CENTRE *for* EXCELLENCE *in* HIV/AIDS
MICHAEL SMITH FOUNDATION FOR HEALTH RESEARCH Discover. Connect. Engage.
St Paul's Hospital FOUNDATION
PROVIDENCE HEALTH CARE Research Institute
CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada
NIDA NATIONAL INSTITUTE ON DRUG ABUSE
UBC
SFU SIMON FRASER UNIVERSITY THINKING OF THE WORLD