Reconstructing the evolution of HIV within a patient

Art F. Y. Poon

BC Centre for Excellence in HIV/AIDS
Vancouver, Canada

Division of AIDS
University of British Columbia

Vancouver Bioinformatics User Group
Outline

The evolution of HIV
  Background
  $2^{nd}$ generation sequencing

Dating HIV infections
  The problem of HIV incidence

Ancestral reconstruction
  Background
  HIV coreceptor use

...oh, and one more thing
Genetics of HIV

- HIV is a retrovirus, encodes reverse transcriptase (RT), RNA → DNA
- ~ 9700 bp genome
- Two genome copies per virus
- Frequent recombination due to RT switching RNA templates
HIV is extremely diverse

- HIV RT is highly error-prone
- Average 1 mutation per 3 replications
- Short generation time (1-2 days)
- Life-long infection
The evolution of HIV

HIV within a single patient

HIV subtype B

Hepatitis B virus genotype C

Influenza A virus H3N2

Hepatitis C virus 6a

All trees depicted are approximately on the same scale.
The problem of HIV evolution

- HIV remains a global challenge due to its rapid evolution
- Virus eludes the host immune system
- A vaccine must protect against highly divergent (~30%) strains
- Adapts to drug therapy by accumulating resistance mutations
Measuring HIV genetic variation

- Capillary-based (‘1\textsuperscript{st}-generation’) sequencing

- Reverse-transcribe HIV RNA to complementary DNA (cDNA).

- Directly sequence the PCR product – average genotype of the HIV population.
HIV genotyping is standard-of-care

- Recommended at baseline and at treatment failure.

- Screen for HIV drug resistance mutations\(^1\).

<table>
<thead>
<tr>
<th>Nonnucleoside Analogue Reverse Transcriptase Inhibitors (NNRTIs)(^{a,m})</th>
<th>(L)</th>
<th>(K)</th>
<th>(K)</th>
<th>(V)</th>
<th>(V)</th>
<th>(Y)</th>
<th>(Y)</th>
<th>(G)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efavirenz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 101 103 106 108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etravirine(^n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Bulk HIV sequence data now ubiquitous.

---

Phylogenies from sequences

- A phylogeny is a hierarchical model (tree) of common ancestry.
- Ancestors are latent nodes inferred by similarity of observed descendants.
- Most recent common ancestor (MRCA) of sample represented by ‘root’.
Applications of HIV phylogenies. 1

- A phylogeny of HIV bulk sequences relates virus populations in different persons.
- The tree is shaped by the history of HIV transmissions.
Applications of HIV phylogenies. 2

- Reconstruct origins of HIV.
- Multiple transmissions from non-human primates.
- Reconstruct spread of global pandemic.

Going within the host

- Evolution of HIV within a host can exceed other viruses around the world.

- But this variation is ‘averaged out’ by bulk sequencing.
The evolution of HIV

Dating HIV infections

Ancestral reconstruction

...oh, and one more thing

1 base pair

400 bps 1 read

100 reads

approximate length of a bulk sequence

1 run

48 samples ~48,000,000 bps 1 run

~2500 reads

approximate size of a landmark HIV clonal study

BRITISH COLUMBIA CENTRE for EXCELLENCE in HIV/AIDS
A great opportunity

- 2GS can provide ideal raw material for studying HIV evolution within patients.
- ‘Ultra-deep’ sequencing – sequence the same genomic region from 1000’s of individual viruses.
- A less common application of 2GS, fewer developers.
A great bioinformatic challenge

- Almost every site is polymorphic
- Even rare variants can be clinically or phylogenetically significant
- Actual variation confounded by 2GS error
- What information can we extract from these data?
Outline

The evolution of HIV
  Background
  2\textsuperscript{nd} generation sequencing

Dating HIV infections
  The problem of HIV incidence

Ancestral reconstruction
  Background
  HIV coreceptor use

...oh, and one more thing
When was the patient infected with HIV?

- Many individuals have been infected for months or years before being diagnosed.
- Long latency of HIV infection.
- Social barriers to HIV testing.
Dates of infection are important

- Monitor the rate of new infections
- Assess the efficacy of HIV prevention measures
- Identify high-risk subpopulations to target measures
How do we estimate dates of HIV infection?

- Detuned serology assays (STARHS²), based on slow increase of anti-HIV antibodies in early infection.
- Predictive models on CD4 cell counts.
- Count polymorphisms (mixtures) in bulk HIV sequences — increasing genetic variation over time.

²Serologic Testing Algorithm for Recent HIV Seroconversion
Problems with these methods

- Limited to identifying early (≤ 6 months) HIV infections.
- Need a reference population to interpret results.
- Sensitive to variation among HIV infections.
A phylogenetic approach

- Mutation rate and time are confounded.

- Samples from the same patient at different points in time allows one to directly estimate the mean rate of HIV evolution.

- This ‘molecular clock’ lets us extrapolate dates from the tips to ancestors deeper in the tree\(^3\).

---

‘Dated-tips’ method

Estimate (1) rate of evolution and (2) times of ancestral splits in the tree.
HIV transmission bottleneck

- Majority of HIV infections are descended from a single virus.
- Implied by lack of variation at early phase of infection.
- Root of the tree\(^4\) can estimate time of infection.

\(^4\)most recent common ancestor, MRCA
Can we recover known dates of infection?

- Identified N~124 HIV seroconverters.
- Date of HIV seroconversion = midpoint between last seronegative and first seropositive visits\(^5\).
- Extracted HIV RNA from archived plasma specimens.
- Ultra-deep sequencing of 2 regions (HIV \textit{env} and \textit{nef})

\(^5\)Median 297 days between visits
Dated-tip analysis of VIDUS, Vanguard

Mean difference = +1.2 months
25%, 75% = –1.8, +6.5 months
Advantages of a phylogenetic method

- Uses only patient-specific data; does not require a reference population

- Robust to variation among HIV infections.

- Can potentially estimate actual dates, not just identifying early infections.
Summary

- Dates of HIV infection can be recovered by analysis of ultra-deep sequence data.
- Requiring only HIV RNA; retrospective study of frozen specimens.
- Scaling up study to reconstruct the historical trend of HIV infection rates in BC.
Outline

The evolution of HIV
  Background
  2\textsuperscript{nd} generation sequencing

Dating HIV infections
  The problem of HIV incidence

Ancestral reconstruction
  Background
  HIV coreceptor use

…oh, and one more thing
Can phylogenetics inform HIV vaccine design?

- HIV infections tend to descend from a single virus.
- A vaccine should protect against transmitted HIV genotypes.
- HIV infections are difficult to sample early.
- Large, expensive prospective cohort studies.
Ancestral reconstruction

• To estimate ancestral character states in a phylogeny.

• Evolution as a continuous-time Markov process.

\[
P(D|T) = \sum_{x \in \{A,C,G,T\}} P(A|x, t_1) P(G|x, t_2) P(x)
\]

• Find joint distribution that maximizes likelihood\(^6\).

\(^6\)Yang, Kumar and Nei. Genetics 1995; 141: 1641-1650.
A test of ancestral reconstruction

- Look for longitudinal data sets with early and ‘late’ clonal HIV sequences.

- Best estimate of transmitted genotype is consensus of early (< 6 mo.) sequences.

- Since we can build a within-host HIV tree, can we reconstruct this ancestor using only late sequences?
Phylogenetics can account for common ancestry
Data collection

• Criteria:
  • Estimated dates of infection available
  • An early sample within 6 months of infection
  • One or more follow-up samples
  • Clonal HIV sequences

• 336 longitudinal data sets (14,663 sequences)

• 232 unique patients
Bringing things together

- We can estimate the times of ancestral nodes.
- We can reconstruct ancestral genotypes.
- Suggests that we can reconstruct the complete history of HIV evolution in a patient over time.
- Map mutations where genotypes on either end of a branch differ.
The majority of HIV variants use CCR5.

Progression to AIDS is associated with switching to CXCR4.
HIV pathogenesis and the coreceptor switch

- A switch to using CXCR4:
  - is associated with accelerated decline of CD4 cell count.
  - is associated with increased rate of progression to AIDS.
  - contraindicates coreceptor antagonist-based therapy.

- Why does the switch occur in some patients and not others?
Fitness valley hypothesis\textsuperscript{7}

- CCR5- and CXCR4-using genotypes are separated by low-fitness intermediates.
- Nearly all mutations removed by selection.
- Crossing this ‘fitness valley’ is a rare chance event (rapid succession of mutations).

Simulations under HIV coreceptor switch models

Fitness valley

Gradual

Replications (x1000)
How do we test this hypothesis?

- Need to be able to measure the rate that intermediate genotypes spread within a patient.
- Determine what mutations accumulate in specific virus lineages.
- Use ancestral reconstruction in a dated-tips phylogeny.
Reconstructing the HIV coreceptor switch

Data collection

- Amsterdam Cohort Studies on HIV-1 infection and AIDS (ACS)
- 8 study participants (DS1-DS8) with chronic untreated HIV infections (1988-1994)
- Known to have undergone HIV coreceptor switch
- Blood draws obtained at ~3 month intervals (median 7 time points per patient)
Reconstructing the HIV coreceptor switch

Ultra-deep sequencing

- Ultra-deep sequencing of the HIV env V3 region.
- Reconstruct dated-tip phylogenies.
- Predict HIV coreceptor usage using \textit{geno2pheno} SVM \(^8\)

Figure 3 from AFY Poon et al. (2012) PLoS Comput. Biol.
Dynamics of HIV coreceptor switch vary among patients

Figure 5 from AFY Poon et al. (2012) PLoS Comput. Biol.
Summary

- Whether an HIV infection has to cross a fitness valley may depend on the transmitted HIV genotype.
- Implies that the mode of switching may be ‘heritable’.
- May also depend on host’s immune response.
- A new way of visualizing NGS data.
Outline

The evolution of HIV
  Background
  $2^{nd}$ generation sequencing

Dating HIV infections
  The problem of HIV incidence

Ancestral reconstruction
  Background
  HIV coreceptor use

...oh, and one more thing
What information can we extract from tree shape?

- Virus phylogeny could be shaped by:
  - Epidemiology of virus
  - Immune response of host
  - Modes of transmission

- Tree shape very difficult to quantify.

- Imbalance statistics, Robinson-Foulds distance…
My group

- Richard H. Liang, post-doc
- Jeffrey Joy, post-doc
- **Looking for grad students!**
- Rosemary McCloskey, undergrad co-op
- Heather Murray, undergrad co-op
- Several Macs and Linux clusters
Acknowledgements

• BC Centre for Excellence in HIV/AIDS
  • Dr. Julio Montaner, Director
  • Dr. P. Richard Harrigan, Program Director, Molecular Laboratory

• Simon Fraser University
  • Dr. Zabrina L. Brumme
This work was made possible through collaborations with Mark Wainberg and Bluma Brenner at the **Lady Davis Institute** (McGill); and Angélique van ’t Wout and Hanneke Schuitemaker of the **Amsterdam Cohort Studies on HIV infection and AIDS.**
This work was supported by an operating grant from the Canadian Institutes of Health Research to AFYP, and funding from the National Institute on Drug Abuse (USA).

AFYP is supported by a Michael Smith Foundation for Health Research (MSFHR) / St. Paul’s Hospital Foundation – Providence Health Care Research Institute (SPHF-PHCRI) Career Investigator Scholar Award.