HIV Recombination Overview

Presenter: Brian Foley

btf@lanl.gov

HIV Databases

Theoretical Biology and Biophysics,
Los Alamos National Laboratory

www.hiv.lanl.gov
seq-info@lanl.gov
The HIV databases contain comprehensive data on HIV genetic sequences and immunological profiles. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal Funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AIU 2007-001-0000. Our content is reviewed by an Editorial Board.

CATNAP: New features
CATNAP now provides an option to calculate geometric mean estimates including tests that were above threshold (padding a score at 100 IC50 or 20 OD50 for the purpose of the estimators). Also, we have introduced a “Trim and re-calculate” feature to the analysis which enables users to select data from specified papers instead of using the full set in CATNAP collections. This could be useful to reduce data redundancy or to address inconsistencies between studies (for instance, changes in pipette tips used for serial dilutions). 20 February 2019

Questions or comments? Contact us at info@lanl.gov

Sponsored by the National Security Education Program of the U.S. Department of Energy, National Nuclear Security Administration.
Workshop Topics

HIV Intersubtype Recombination and Unique Circulating Recombinant Forms

General introduction: HIV, like all retroviruses, is effectively diploid, packages two copies of the viral genome per virion.

Mechanism of Recombination: Template switching during reverse transcription, no DNA damage-repair enzymes needed.

Tools for detecting Intersubtype Recombination:

Overview of HIV-1 M group subtypes and CRFs:

Examples of what recombinants “look like”:
Recombinant viruses can be formed when one cell is infected with 2 viruses.

Distance or diversity between the two viruses can be large (intersubtype) or small (near identity intrapatient).

Recombination occurs by template switching during reverse transcription of heterodimorphic viruses.

FIG. 8. Requirements for generating recombinant genomes. (A) Coinfection with two genetically distinct viruses does not yield recombinants. However, a producer cell must be infected with two genetically distinct viruses (shown here as viral particles with two blue or two red RNAs) to produce viral particles with heterodimorphic gRNAs. (B) Recombination is observable in cells infected with heterodimorphic virions (particle containing one red and one blue RNA strand). Template switching during reverse transcription can generate a recombinant provirus.
Virus Recombination Detection Tools:

RIP: HIV-databases [https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html]
   Pros: Adjustable window size, prebuilt test set, allows user input test set,
         adjustable significance threshold, contains consensus for each subtype.
   Cons: Does not output numerical breakpoint locations.

jpHMMer: Gobics [http://jphmm.gobics.de/submission_hiv.html]
   Pros: Statistical support of precise breakpoints, outputs table of breakpoints.
   Cons: Does not have HMM models for CRFs, weak models for rare subtypes,
         does not include subsubtypes (A3, A4, A6, F2), window size not adjustable.

   Pros: Accepts input of many sequences, phylogeny as well as other methods,
         CRFs updated reasonably often.
   Cons: Window size not adjustable,

RDP3, RDP4: [http://web.cbio.uct.ac.za/~darren/rdp.html]


SimPlot: Stuart Ray [https://sray.med.som.jhmi.edu/SCRoftware/simplot/]

Highlighter: HIV Databases [https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html]
Many more recombination detection tools:
http://bioinf.man.ac.uk/robertson/recombination/programs.shtml

Links to Recombinant Sequence Analysis/Detection Programs

Welcome to the comprehensive list of recombination analysis software maintained by the Robertson Lab.

<table>
<thead>
<tr>
<th>Filter list by method or algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
</tr>
<tr>
<td>Algorithms</td>
</tr>
</tbody>
</table>

### 3seq
- **Description**: 3SEQ is a software program for identifying mosaic structure or recombination in nucleotide sequence data. 3SEQ takes as input a data set with a minimum of three aligned sequences, and it tests whether any sequence in the data set is a recombinant or mixture of any of the two sequences in the data set.
- **Implements algorithms**: None
- **Methods**: None
- **Additional notes**: None

### 4SIS
- **Description**: Two types of informative sites were distinguished, corresponding to the clustering of the putative recombinant with either of the parental representatives. The optimal breakpoint was located by maximizing a chi-square statistic.
- **Methods**: None
- **Additional notes**: Temporarily unavailable

### BARCE
- **Description**: BARCE is a C++ program for detecting recombination breakpoints in four-sequence alignments using hidden Markov models, Bayesian principles and Markov chain Monte Carlo sampling.
- **Methods**: None
- **Additional notes**: None

### Bellerophon
- **Description**: Bellerophon is a program for detecting chimeric sequences in multiple sequence datasets by an adaptation of partial timing analysis.
- **Implements algorithm**: None
- **Methods**: None

### BLAST Genotyping
- **Description**: This tool helps identify the genotype of a viral sequence. A window of size is slid along the query sequence and each window is compared by BLAST to each of the reference sequences for a particular virus.
- **Implements algorithm**: None

### cBrother
- **Description**: cBrother is software for inferring recombination when recombination is rare. This is a C version of the code originally written in Java available in DualBrothers.
- **Methods**: None

### DndIDP
- **Description**: DndIDP is a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DndIDP can estimate several measures of DNA-sequence variation within and between populations (e.g., homologous, synonymous or nonsynonymous sites), and in various sorts of codon positions, as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters.
- **Methods**: None

### DualBrothers
- **Description**: DualBrothers is a recombinant detection software based on the Dual Multiple Change-Point (MCP) model. This model allows for changes in topology and evolutionary rates across sites in a multiple sequence alignment.
- **Methods**: None

### Frag-dists
- **Description**: This program uses the principle of “Recombination has high similarity with its parental sequence” to visualize the location of breakpoints and gives out the potential parental sequences.
- **Methods**: None
How to decide “which tool”? 

The answer depends on many factors such as number of sequences which need to be screened, short or long sequences, diversity in the local population, etc…

No single tool does everything very well.

There are trade-offs in speed vs accuracy, rate of false-positive vs false negative results, etc…

Very fast methods can be simple, such as BLAST of each sequence against a small subtype reference set local database. Parsing and interpreting the results may be the most difficult part.
HIV-1 subtypes and CRFs

- Subtypes and subsubtypes: A1 – A6, B, C, D, F1 – F2, G, H, J, K

- Circulating Recombinant Forms: CRF01_AE – CRF98_06B

- Both require a minimum of 2 complete genomes and at least one more partial genome with sequences from regions that can confirm the structure of the first 2.
CRF03_AB is recombinant between A6, the subsubtype of A found in the former Soviet Union region and B.

In this plot 2 reference genomes of CRF03_AB are included, along with A1, A2, B, C, F1, F2, and G genomes.

Window size 150 bases and step of 50 bases
CRF03_AB Maps

From publication describing CRF03_AB

From jpHMMer analysis at Gobics
FIG. 2. Recombinant structure of the Kaliningrad IDU-associated HIV-1 strain. Similarity (top) and bootscanning (bottom) analyses were used to map the complete genome sequence of the AB-98RU001 HIV-1 isolate. In both analyses a window of 400 bases and an increment of 50 bases were used. Gap regions in the alignment were excluded from the analyses. The Kimura 2-parameter model with 100 replicates was used as the algorithm for bootscanning. Subtype C isolate ETH2220 was used as an outgroup. Similarity and bootstrap value are shown on the y axis and positions on the full genome alignment are shown on the x axis. Vertical lines indicate the recombination points. The subtype origin of the different genome regions is indicated (middle) as gray (subtype A) and white (subtype B) regions. The small region in the pol–vif region, which seems to be derived from subtype A, could not be reliably verified by separate phylogenetic analysis and is therefore shown as uncertain (striped).
AliView (free for Mac) or BioEdit (free for Windows) multiple sequence alignment editor view of CRF03_AB plus reference genomes.

http://ormbunkar.se/aliview/
http://www.mbio.ncsu.edu/BioEdit/bioedit.html
Aliview Control-Click on Sequence to highlight differences from that sequence
Positions or sites in HIV-1 genomes are numbered using alignment to the HXB2 reference genome as the standard. The HIV Map drawing tool can be used to create maps of the genome colored by region. 

https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html
jpHMM-HIV at Gobics gives best recombination site location numbering.

vs input query sequence location

vs HXB2 standard sequence location
HIV recombination Map
Drawing tool at HIV-DB customizable to change colors used, and input your own numbers in cases where jpHMM (or another tool) did not give the correct sites.

https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html
SimPlot - Query: CRF03_A1B

Filename: Y:\Documents\SIMPLOT-Mgroup\CRF03_AB Ukraine\CRF03_PlusRefsSTRPD.FASTA

Window: 300 bp, Step: 20 bp, GapStrip: On, Kimura (2-parameter), T/t: 2.0

https://sray.med.som.jhmi.edu/SCRoftware/simplot/
Even a very small region of misalignment, hypermutation, or poor sequence quality can have a large impact on similarity plots, phylogenetic trees, and other analyses. Similarity plots can be quite useful for identifying sites in a multiple sequence alignment that should be scrutinized, and corrected if in error, as this example shows.
Zoomed in on “A-like” central region, shows that it truly is most likely recombination and not just an aberrant region.

Other factors should also be considered too. In this case for example we know that the recombinant was formed in a person dual-infected with A6 and B viruses, so a region of A6 is not at all unexpected.
Phylogenetic tree of A-like central region of CRF03_AB

B-FSU relatively diverse in this small region of the genome

F1 and F2, B and D widely separated in this small region, indicates perhaps a lack of solid phylogenetic information

CRF03 well within the A6 subclade
REGA HIV-1 & 2 Automated Subtyping Tool (Version 2.0)

This tool is designed to use phylogenetic methods in order to identify the subtype of a specific sequence. The sequence is analysed for recombination using bootscanning methods.

Note for batch analysis: The REGA subtype tool accepts up to 1000 sequences at a time.

Enter here your input data as FASTA format.

Choose a mirror to subtype your sequences

https://www.genomedetective.com/app/typingtool/hiv

Developed by: Tulio de Oliveira, Koen Dehoche, Sharon Cassol, Andrew Rambaut and Anne-Mieke Vandamme.

Developed in cooperation with the Evolutionary Biology Group at University of Oxford, UK, the HIV-1 Pathogenesis and Immunotherapeutics Program at the University of Pretoria, South Africa, and the REGA Institute at the Katholieke Universiteit Leuven, Belgium.

Funded by the Marie Curie Fellowship, Flemish Bilateral Cooperation Program and the Wellcome Trust (grant # 061238).

For questions, suggestions or problems please contact: Dr. Tulio de Oliveira.
HighLighter for intra-patient recombination

https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html

Dual-infected patient.

Infected with two strains of the same subtype.
RAPR recombination analysis program

https://www.hiv.lanl.gov/content/sequence/RAP2017/