Selection at the HIV Transmission Bottleneck

Jonathan Carlson, Microsoft Research
HIV research at MSR

People
David Heckerman
Nebojsa Jojic
Andrew Philips
Neil Daschau
Carl Kadie
Jennifer Listgarten
Nicolo Fusi

Research
Applying statistics and computation to better understand disease
10 yrs, >75 papers published in the HIV field

Selection bias at the heterosexual HIV-1 transmission bottleneck
July 11, 2014
HIV Mutates. Fast.

Infection is established by a single genetic variant.

The transmission bottleneck

Transmission is inefficient
1/300 exposures result in infection
Women are at 2x risk relative to men
Men with STI’s are at 2x risk

Hypothesis:
Could it be that many viruses make it across, but the replication rate is terminally low?
N > 1,000

Couple is identified as HIV serodiscordant

Couples counseling and condoms reduce transmissions by 2/3

HIV negative partner is tested once per month

HIV negative partner seroconverts

Plasma collected from Donor and Recipient, median 45 days post estimated infection

HIV sequencing of both partners

N = 137

Susan Allen

Eric Hunter
Estimating fitness with statistical power

Donor

ID1  ...EPRGS...ATTSNLQ...GWMTSNPP...PV...
ID2  ...DPRGS...TTSNLQ...GWMTNNP...PV...
ID3  ...EPR...LSDIAGTTSNLQ...GWMTNNP...PV...
ID4  ...EPRGSI...GGTTSLQ...IAWMTNNP...PV...
ID5  ...EPRLS...DIA...QGIGWMTNNP...PV...
ID6  ...EPR...GSDIAGTTSLQ...QGIGWMTNNP...PV...
ID7  ...DPRG...DIAGTTSNLQ...IAWMTNNP...PV...
ID8  ...EPRL...EIA...TTSLQ...GWMTNNP...PV...

Recipient

ID1  ...EPRGSDIA...TTSNLQ...GWMTSNPP...PV...
ID2  ...DPRGS...TATSLQ...GWMTNNP...PV...
ID3  ...EPRGS...DIAGTTSLQ...IAWMTNNP...PV...
ID4  ...EPRGSI...GGTTSLQ...IAWMTNNP...PV...
ID5  ...EPRLS...DIA...QGIGWMTNNP...PV...
ID6  ...EPR...GSDIAGTTSLQ...QGIGWMTNNP...PV...
ID7  ...DPRG...DIAGTTSNLQ...IAWMTNNP...PV...
ID8  ...EPRL...EIA...TTSLQ...GWMTNNP...PV...

99.8% identity
Estimating fitness with statistical power

<table>
<thead>
<tr>
<th>D</th>
<th>R</th>
<th>T</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1, pos1</td>
<td>E</td>
<td>E</td>
<td>1</td>
</tr>
<tr>
<td>ID1, pos2</td>
<td>P</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ID2, pos11</td>
<td>N</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>ID2, pos12</td>
<td>S</td>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td>ID2, pos13</td>
<td>T</td>
<td>T</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ID8, pos1711</td>
<td>P</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>ID8, pos1712</td>
<td>V</td>
<td>V</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \Pr(T = 1) \approx \frac{f_{AA}}{1 - f_{AA}} \times \frac{p_{AA}}{p_{AA}} \]

\[ \text{logodds}(T = 1) \approx \text{logodds}(f_{AA}) + \text{bias}_{AA} \]
Consensus := observed in ≥ 50% of Zambian HIV+ individuals
Estimating fitness with statistical power

### Approach

\[
\text{Pr}(T = 1) \approx \frac{f_{AA}}{1 - f_{AA}} \times \frac{p_{AA}}{p_{AA}}
\]

\[
\text{logodds}(T = 1) \approx \text{logodds}(f_{AA}) + \text{bias}_{AA}
\]

Estimate probability of transmission using a generalized linear mixed model

<table>
<thead>
<tr>
<th>D</th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1, pos1</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>ID1, pos2</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ID2, pos11</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>ID2, pos12</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>ID2, pos13</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ID8, pos1711</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>ID8, pos1712</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>

\( N = 228,362 \)
A fitness bias

AA features related to fitness

Conservation
Predicted impact on structure
Interaction network properties
Relationship to immune escape
A fitness bias

AA features related to fitness

Conservation
Predicted impact on structure
Interaction network properties
Relationship to immune escape
A fitness bias

AA features related to fitness

Conservation

Predicted impact on structure

Interaction network properties

Relationship to immune escape
High-fitness amino acids are more likely to be transmitted

What about whole virus sequences?
## Transmission Index

<table>
<thead>
<tr>
<th>ID</th>
<th>pos</th>
<th>D</th>
<th>R</th>
<th>T</th>
<th>Ln-odds (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1, pos1</td>
<td></td>
<td>E</td>
<td>E</td>
<td>1</td>
<td>6.9</td>
</tr>
<tr>
<td>ID1, pos2</td>
<td></td>
<td>P</td>
<td>P</td>
<td>1</td>
<td>7.2</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>ID2, pos11</td>
<td></td>
<td>N</td>
<td>T</td>
<td>0</td>
<td>4.6</td>
</tr>
<tr>
<td>ID2, pos12</td>
<td></td>
<td>S</td>
<td>S</td>
<td>1</td>
<td>5.3</td>
</tr>
<tr>
<td>ID2, pos13</td>
<td></td>
<td>T</td>
<td>T</td>
<td>1</td>
<td>6.8</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>ID8, pos1711</td>
<td></td>
<td>P</td>
<td>P</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>ID8, pos1712</td>
<td></td>
<td>V</td>
<td>V</td>
<td>1</td>
<td>6.3</td>
</tr>
</tbody>
</table>

### Transmission Index

- 7.1
- 6.3
- 6.8
Individuals with weak viral populations are less likely to transmit to their partners
A new model of transmission

Frequent non-productive infection
Provides a window of opportunity for drugs and vaccines

What about risk factors?

\[
\frac{\Pr(T = 1)}{\Pr(T = 0)} \approx \frac{f_{AA}}{1 - f_{AA}} \times \frac{p_{AA} + c}{p_{AA} + c} \quad \text{as} \quad c \to \infty
\]

\[
\Pr(T = 1) \approx \frac{f_{AA}}{1 - f_{AA}}
\]

Joseph & Swanstrom, Science 2014
Men vs Women

Stronger bias in men

Less likely to be infected

More likely to be infected with a strong virus
Men vs Women

Stronger bias in men

Less likely to be infected

More likely to be infected with a strong virus

More evolutionary pressure on women’s virus

HR = 1.78
p = 0.016
Men vs Women

Stronger bias in men
Less likely to be infected
More likely to be infected with a strong virus
More evolutionary pressure on women’s virus

Stronger transmitted viruses lead to more severe disease
A paradigm:
Low biological risk leads to lower odds of infection, but more severe disease when infection happens.
Summary

Fitness bottleneck at transmission
Features consistent with viral fitness predict transmission
Viral sequences and populations with high predicted fitness are more likely to establish infection

Risk factors mitigate the fitness bottleneck
A mechanism for increased risk is a reduction in the fitness bottleneck
Possibly due to more target cells, or higher activation state of target cells

Clinical consequences
Anything that weakens the virus will reduce transmission rates
Vaccines and drugs that protect individuals from transmission may lead to more severe disease when breakthrough infection occurs
Acknowledgments

Eric Hunter Lab, Emory
Malinda Schaefer, Daniela Mónaco
Dan Claiborne, Jessica Prince
Martin Deymier, Zachary Ende, Nichole Klatt

ZEHRP
Susan Allen
William Kilembe
Mubiana Inambao
Staff & study participants

Microsoft Research
David Heckerman
Henry Lin
Jian Peng
Charlie DeZiel

Others
Todd Allen, Rebecca Batorsky & Aaron Seese, Ragon
Jill Gilmour & Matt Price, IAVI
Philipoulder & John Frater, Oxford
Thumhi Ndungu, U KwaZulu Natal
Roger Shapiro, Beth Israel Deaconess MC
James Tang & Paul Goepfert, UAB
Save the planet and return your name badge before you leave (on Tuesday)