Are high prevalence fishing communities on Lake Victoria the predominant source of new infections in Rakai, Uganda?

Oliver Ratmann, PhD
Lecturer in Statistics, Department of Mathematics, Imperial College London
or105@imperial.ac.uk
Oliver Ratmann

- 2009: PhD in Bayesian Statistics and Network Science, Imperial College London

- since 2013: research focus on viral phylogenetics and HIV (Dutch observational HIV cohort, PANGEA, HTEAM, HIV in cities: Seattle)

- since 2017: Lecturer in Statistics at Imperial; PhD students Xiaoyue Xi, Devrat Kaushal; post-doc Paul-Marie Grollemund

- two kids (Jonas and Timos)
Spatial heterogeneity in African HIV epidemic

- HIV epidemic markedly heterogeneous
- Foci of high HIV incidence and prevalence communities ("HOTSPOTS")
- Hypothesis: GEOGRAPHIC TARGETING of combination HIV prevention (CHP) to hotspots is ESSENTIAL
  - substantially ameliorate broader epidemic
  - cost-efficient use of resources

figure from Anderson et al., Lancet (2014)
Example: Lake Victoria Fishing Communities

- Lake Victoria fishing sites are hotspots
- Since 2014, targeting of CHP to fishing communities recommended
- Recommendation based on:
  - high incidence
  - assumption that fishing sites are major source of new infections among inland communities
• Confirm the assumption that fishing communities are a major source of new infections among inland communities.
Methods overview

• Two ways in which fishing sites can seed new infections in inland communities:
  ➢ transmission from individuals who reside in fishing sites
  ➢ migration to inland communities and subsequent spread of HIV

• Used a combination of techniques and data sources.
Acknowledgements

Rakai Health Sciences Program
David Serwadda
Fred Nalugoda
Joseph Kagaayi
Godfrey Kigozi
Gertrude Nakigozi
Tom Lutalo
Robert Ssekubugu
Grace Kigozi
Jeremiah Bazaale

Johns Hopkins School of Medicine
Aaron Tobian
Larry Chang

National Institute of Allergy and Infectious Diseases
Thomas Quinn
Andrew Redd
Steve Reynolds
Oliver Laeyendecker

Johns Hopkins School of Public Health
Mary Kate Grabowski
Ronald Gray
Maria Wawer
Justin Lessler
Caitlin Kennedy
Joseph Sekasanvu

Imperial College London
High performance computing center

PANGEA-HIV-II consortium
Christophe Fraser
Mary Kate Grabowski
Deenan Pillay
Andrew Rambaut

Rakai Health Science Program
Staff and Study participants

Oxford University
Christophe Fraser
Matthew Hall
Chris Wymant
Tanya Golubchik
Lucie Abeler-Dorner
• First, surveillance was extended to include 4 largest fishing sites in the area

<table>
<thead>
<tr>
<th></th>
<th>36 inland communities</th>
<th>4 fishing sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible</td>
<td>29,115</td>
<td>8,529</td>
</tr>
<tr>
<td>Present</td>
<td>20,089</td>
<td>6,152</td>
</tr>
<tr>
<td>Participated</td>
<td>19,799 (99%)</td>
<td>6,083 (99%)</td>
</tr>
<tr>
<td>Infected</td>
<td>2,703 (14%)</td>
<td>2,439 (40%)</td>
</tr>
<tr>
<td>ART- naïve</td>
<td>1,819 (67%)</td>
<td>2,059 (84%)</td>
</tr>
</tbody>
</table>
Methods overview

- Second, reconstructed likely HIV-1 transmission networks with viral phylogenetic analysis

- 36 inland communities
  - Infected: 2,703 (14%)
  - ART-naive: 1,819 (67%)
  - Sequenced: 1,138 (63%)

- 4 fishing sites
  - Infected: 2,439 (40%)
  - ART-naive: 2,059 (84%)
  - Sequenced: 1,514 (74%)
Methods overview

- Third, households and migration events were geo-coded.
Methods overview

- Third, households and migration events were geo-coded.

- This allowed us to interpret phylogenetic transmission networks at a resolution below 50km and while accounting for population movement.
Obtaining population-based samples of NGS data is feasible

- 5,142 HIV positive
- 3,878 ART naïve
- 3,758 deep sequenced
- 2,652 with sufficient sequence depth (30X)

- Overall sequence coverage 48.1% assuming individuals who did not participate were infected in proportion to surveyed population
Deep sequence phylogenetics enables to infer the direction of transmission

- Gives estimates on who might be the source case
- Validated against epidemiological data, found accuracy 83.7%
  - Wymant et al. Virus Evolution 2018;
  - Wymant et al. MBE 2017;
  - Ratmann et al. (in prep);
  - Rose et al. on behalf of HPTN052 #TUPEA001
Phyloscans across the genome

A

B

C

D

Phyloscans across the genome

Graphical representation with annotations and data points.

Subgraph distance

Genomic position (relative to HXB2)

topological subgraph relationship

ancestor m→f

ancestor f→m

intermingled

mating

disconnected

MSP

Male spouse

FSP1

Female spouse

First partner

FSP2

Female spouse

Second partner

OM

Other male

??
Phyloscans across the genome

![Diagram showing genomic position and topological subgraph relationship]
Phylocans across the genome

![Graph showing genomic position and subgraph distance](image-url)

- **Subgraph distance (subst/site)**
- **Genomic position (relative to HXB2)**

- **Topological subgraph relationship**
  - ancestral m→f
  - ancestral f→m
  - sibling
  - disconnected
  - intermingled

- **Relationships and Percentages**
  - 3.14/6.4 = 49.1%
  - 1.40/6.4 = 21.8%
  - 1.16/6.4 = 18.2%
  - 4.59/7.6 = 60.4%
  - 1.03/7.6 = 13.5%
  - 1.87/7.6 = 24.6%
Highly parallelized phylogenetic inference based on NGS data

2,803 individuals
2,803*2,802/2 = 4.5m pairings
335 sliding phylogenies per pair
Phylogenetically reconstructed transmission networks

- 446 reconstructed transmission networks, containing 888 phylogenetic linkages
Phylogenetically reconstructed transmission networks

- 446 reconstructed transmission networks, containing 888 phylogenetic linkages
  - 351 linkages had low statistical support
  - 80 further FF
  - 81 further MM
- 376 highly supported male-female linkages
- 293 (78%) linkages had one direction of transmission highly supported
Direct transmission cannot be proven in population-based sample

- Phyloscanner inference among couples
- Couples were either phylogenetically close or distant
- 76% of close couples had strong evidence that transmission occurred in one direction
Direct transmission cannot be proven in population-based sample

- Phyloscanner inference on the population-based sample
- Phylogenetic linkage not clear-cut
- Up to 35% of phylogenetically linked pairs in transmission network may not represent direct transmission events
- 78% of phylogenetically close pairs had strong evidence that transmission occurred in one direction
Inferred direction of transmission has low false-discovery rate

- We cross-validated our results against clinical data from which the direction of transmission could be inferred.

**Sero-discordant**

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CD4-discordant**

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- CD4 measurement:
  - <400
  - 400–799
  - 800+

years relative to date both partners positive
Inferred direction of transmission has low false-discovery rate

<table>
<thead>
<tr>
<th>Classification</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>classified as phylogenetically linked</td>
<td>71</td>
</tr>
<tr>
<td>direction consistent with clinical evidence</td>
<td>46</td>
</tr>
<tr>
<td>direction ambiguous</td>
<td>16</td>
</tr>
<tr>
<td>direction inconsistent with clinical evidence</td>
<td>9</td>
</tr>
<tr>
<td>false discovery rate</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

Pairs in population-based sample with clinical evidence for transmission in one direction
Conclusions Part 1

- Deep sequence phylogenetic analyses can be conducted at the scale of population-based surveys with the phyloscanner method.

- Linkage analyses using deep sequence reads have similar limitations compared to using consensus sequences.

- Phyloscanner analysis cannot prove transmission of HIV between two individuals.

- The direction of transmission could be inferred in the large majority of closely related pairs, and with small error.
Results
Reconstructed transmission flows

- 293 source-recipient pairs reconstructed
• 265/293 (91%) were not between inland and fishing sites.
• Suggesting largely distinct transmission networks.
Reconstructed transmission flows

- 7/293 (2.3%) from fishing to inland.
- 21/293 (7.2%) from inland to fishing.
One more complication

- Substantial differences in participation and sequencing rates
Adjusting for participation and sequencing biases

- Adjusted analysis for known biases with Bayesian data augmentation (Givens et al. 1997, Stat Sci)
Fishing sites are net sinks despite being hotspots
Implications for geographic targeting

- Cautious case study:
- Hotspots are not necessarily the main sources of new infections in neighboring low risk areas.
- Populations in hotspots will directly benefit from geographic targeting of CHP.
- Indirect benefits of geographic targeting to neighboring low risk areas may be limited.
Going forward

• Mapping may not be enough.
• Comparative assessment of hotspots on their contribution to the broader spread of HIV should be a priority.
• Similar analyses could help allocation of limited resources, by prioritizing those hotspots for geographic targeting that have a large impact on the epidemic.
Acknowledgements

Rakai Health Sciences Program
David Serwadda
Fred Nalugoda
Joseph Kagaayi
Godfrey Kigozi
Gertrude Nakigozi
Tom Lutalo
Robert Ssekubugu
Grace Kigozi
Jeremiah Bazaale

Johns Hopkins School of Medicine
Aaron Tobian
Larry Chang

National Institute of Allergy and Infectious Diseases
Thomas Quinn
Andrew Redd
Steve Reynolds
Oliver Laeyendecker

Johns Hopkins Bloomberg School of Public Health
Mary Kate Grabowski
Ronald Gray
Maria Wawer
Justin Lessler
Caitlin Kennedy
Joseph Sekasanvu

Imperial College London
High performance computing center

PANGEA-HIV-II consortium
Christophe Fraser
Mary Kate Grabowski
Deenan Pillay
Andrew Rambaut

Rakai Health Science Program Staff and Study participants

Bill and Melinda Gates Foundation
NIH / NIAID
Center For AIDS Research