248: Exploring transmission dynamics of HIV in rural KwaZulu-Natal, using phylogenetics

Eduan Wilkinson[1], Siva Danaviah[1], Justen Manasa[1], Frank Tanser[1,2], Kobus Herbst[1], Deenan Pillay[1,3] and Tulio de Oliveira[1,2]
on behalf of the PANGEA_HIV Consortium [1]Africa Centre for Health and Population Studies, University of KwaZulu-Natal, South Africa; [2]College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa [3]School of Life and Medical Sciences, University College London, U.K.

Background: South Africa is home to the largest number of people living with HIV/AIDS, with an estimated 5,6 million people infected. The epidemic is dominated by the subtype C strain of HIV-1 (>95% of infections).

In our study setting, at the Africa Centre for Health and Population Studies, a large degree of geographical heterogeneity have been observed with epidemiological analysis, which indicates that transmission is not uniform across the study population. Furthermore, the transmission of the virus still continuous at alarming rates. Key to devise the most effective interventions will be a detailed knowledge on how HIV-1 is transmitted at a population level.

In recent years, the use of phylogenetic and phylodynamic methods have been used with great success in complementing traditional epidemiological methods. To date the majority of these studies have only been conducted in developed countries. With the newly established PANGEA_HIV consortium we aim to generate large datasets of HIV-1 whole genomes and utilize such methods in the study of the epidemic in generalized HIV epidemics in sub-Saharan Africa.

Aims and Objectives:

1) **Generate a comprehensive genotypic** dataset of the locally infected population.

2) **To establish the molecular epidemiological characteristics of HIV-1 transmission in the study population.**

A. Characterize clusters of genetically linked HIV-1 infections.
B. Determine how clusters evolve over time with increasing ART coverage.

3) **To identify the influence of high-transmission networks have on the dynamics of a generalized epidemic with increasing coverage of antiretroviral therapy (ART).**

A. Identify high-transmission hotspots change over time.
B. Determine the impact of ART interventions on HIV transmission by the application of a number of phylogenetic methods to detect changes in incidence (see poster by Hodcroft, Ratmann and colleagues).

More info: bioafrica.net | pangea-hiv.org

Methods: RNA is extracted and sequenced from dried blood spots (DBS), with success rate of ~ 86% (VL > 10,000) and plasma samples, success rate ~ 95%.

Current dataset contains 1,320 Pol genotypes and 450 complete genomes. Dataset increases at approximately 200 genotypes a month (see poster from Danaviah, Manasa and colleagues).

Current coverage ≈ 5% of infections (Objective is to increase coverage to : by end 2016).



Figure 1A: Surveillance uses Dried blood spot (DBS).



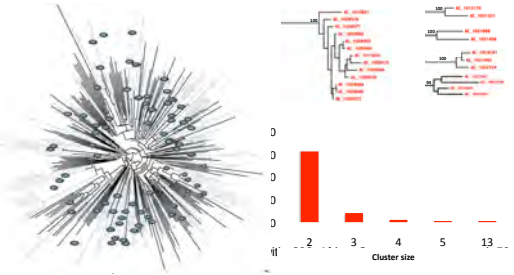
Figure 1B: Sampling is complemented with patients failing ART at local clinics.

Phylogenetics and Phylodynamics analyses:

1. Africa Centre (AC) genotypes were analyzed against all available homologous reference strains (n=1,693) from South Africa.
2. A NJ tree was inferred with 500 bootstraps in order to identify clusters of low genetic diversity (<0.02) and high bootstrap support (>99%).
3. AC sequences in clusters (n =290) were aligned with a random selection of 50 South African C sequences. JModeltest was used to select evolutionary model. ML trees were constructed in PhyML and RaxML with 100 bootstrap replicates.
4. Clusters were identified using previously mentioned criteria. In addition, we also identified clusters with >95% and >90% bootstrap and diversity less than 0.03 and 0.04. Clusters in both phylogenies were retained.
5. Individuals in clusters were characterized based on demographic, clinical and behavioral characteristics.
6. Phylogeographic analysis with detailed GIS coordinates were used to determine the geographical spread of networks of transmission.
7. Lastly, We applied phylodynamics models, such as Bayesian Skyline Plot and Birth-death skyline models in order to estimate epidemiological parameters.

Results: Through the phylogenetic analyses a total of 74 linked infections comprising 174 individuals were uncovered (ML trees with bootstrap > 90 and diversity < 0.04).

The majority of linked infections contained two individuals (n=68), while a small number of larger clusters was observed (Figure 2).



numbered by the Phylotype software application (http://www.phylotype.org). Four of 74 clusters are highlighted in red, bootstrap support is shown. Graph illustrates the distribution of cluster size.

Characteristics of linked infections:

Variable	Linked (n=174)	Unlinked (n=1146)	P-value
Demographic			
Median age	28.775	34.332	**0.0001**
Employed % (n)	17.81%	14.74%	0.6855
Clinical			
Mean viral load (copies/ml)	336 991	194 880	**0.0040**
Median viral load (copies/ml)	119 800	59 590	**< 0.0001**
ART naïve % (n)	69.54%	49.97%	**< 0.0001**
Seroconverted in last year % (n)	12.06%	4.53%	**< 0.0001**
Failing ART % (n)	18.39%	42.23%	**0.0005**
Behavioral			
Knowledge of HIV+ status % (n)	52.3%	35.1%	0.3760
Mean current relationship count % (n)	1.041	0.951	0.1890
Mean lifetime partners % (n)	2.709	3.138	0.3970

Geographical linkage of infections:

GIS was used to evaluate the geographical distribution of linked infections.

High concentration of linked infections could be traced back to a peri-urban settlement adjacent to the national highway, which is a hotspot of incidence and prevalence of HIV-1 (see poster by Tanser, de Oliveira and colleagues).



Figure 3: Dot represents genotyped individuals. Line represents linkage. Intentional random error added to the locations for confidentiality reasons.

Dynamic modeling:

Phylodynamic modeling shows a stable effective population size or Ne over the past decade with a slightly decrease in recent years (Figure 4).

We are currently parameterizing Bayesian Skyline Plots (BSP) that were run under Birth-Death tree priors in order to estimate the basic reproductive number (R0) and the rate of becoming non-infections (δ). Early results show that becoming non-infectious has increased in recent years due to successful treatment.

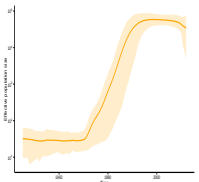


Figure 4: BSP plot.

Conclusion: 74 clusters with 174 linked infections were identified even though the sampling coverage in our study population is still fairly low. **Viral load**, **age**, **treatment failure**, **being not on treatment**, and **recently infected individuals** were all significantly associated with linked individuals. GIS analyses of the linked infections confirm the large degree of **geographical heterogeneity** in incidence previously observed.

Funding: PANGEA_HIV is funded by the Bill & Melinda Gates Foundation, the genomics unit was funded by PANGEA_HIV, South African Medical Research Council, and South Africa National Research Foundation. Africa Centre is funded by a strategic award of the Wellcome Trust.