247 Phylodynamic analysis of HIV sub-epidemics in Mochudi, Botswana

Vlad Novitsky[1], Denise Kühnert[2], Sikhulile Moyo[3], Erik van Widenfelt[3], Lillian Okui[3], and Max Essex[1,3]

[1]Harvard School of Public Health, Boston, MA; [2]Institute of Integrative Biology, Zürich, Switzerland; [3]Botswana-Harvard AIDS Institute Partnership, Gaborone, Botswana

ABSTRACT

Southern Africa continues to be the epicenter of the HIV/AIDS epidemic. This HIV-1 subtype C epidemic has a predominantly heterosexual mode of virus transmission and high (>15%) HIV prevalence among adults. The epidemiological dynamics of the HIV-1C epidemic in southern Africa are still poorly understood. Here, we aim at a better understanding of HIV transmission dynamics by analyzing HIV-1 subtype C sequences from Mochudi, a peri-urban village in Botswana.

HIV-1C env gene sequences (gp120 V1C5) were obtained through enhanced household-based HIV testing and counseling in Mochudi. More than 1,200 sequences were generated and phylogenetically distinct sub-epidemics within Mochudi identified. The Bayesian birth-death skyline plot was used to estimate the effective reproductive number, R, and the timing of virus transmission, to classify sub-epidemics as “acute” (those with recent viral transmissions) or “historic” (those without recent viral transmissions).

We identified two of the 15 sub-epidemics as “acute.” The median estimates of R among the clusters ranged from 0.72 to 1.77. The majority of HIV lineages, 11 out of 15, appear to have been introduced to Mochudi between 1996 and 2002. The median peak duration of viral transmissions was 7.1 years (range 2.9–9.7 years). The median life span of identified HIV sub-epidemics, i.e. the time between the inferred epidemic origin and its most recent sample, was 13.1 years (range 10.2–22.1 years). Most viral transmissions within the sub-epidemics occurred between 1997 and 2007. The time period during which infected people are infectious appears to have decreased since the introduction of the national ART program in Botswana.

Real-time HIV genotyping and breaking down local HIV epidemics into phylogenetically distinct sub-epidemics may help to reveal the structure and dynamics of HIV transmission networks in communities, and aid in the design of targeted interventions for members of the acute sub-epidemics that likely fuel local HIV/AIDS epidemics.

METHODS

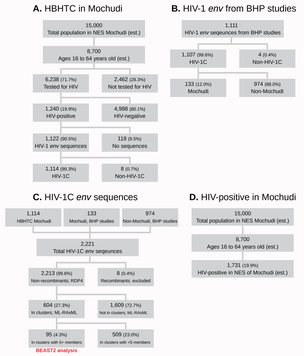
Study Subjects. The study subjects participating in the Mochudi Prevention Project (MPP) have been previously described (Novitsky et al., 2013). Three rounds of home-based HIV testing and counseling (HTC) were conducted in Mochudi during May 2010 – August 2013.

HIV-1C *env* gp120 sequences. Nucleotide sequences spanning the HIV-1C *env* gp120 V1C5 region were generated by population-based (bulk) Sanger sequencing. A total of 1,122 sequences were generated from 1,240 HIV-positive individuals from Mochudi. A total of 1,111 HIV-1 env V1C5 sequences with known geographic origin and time of sampling were utilized from multiple Botswana-Harvard AIDS Institute Partnership (BHP) studies performing viral genotyping.

The overall HIV-1C genotyping coverage for Mochudi was estimated at 72.0% (95% CI 69.8% to 74.1%).

Definition of HIV cluster: a viral lineage that gives rise to a monophyletic sub-tree of the overall phylogeny with strong statistical support in the context of high sampling density. The bootstrap support of 0.80 was used as statistical support of monophyletic clades (subtrees, viral lineages, clusters).

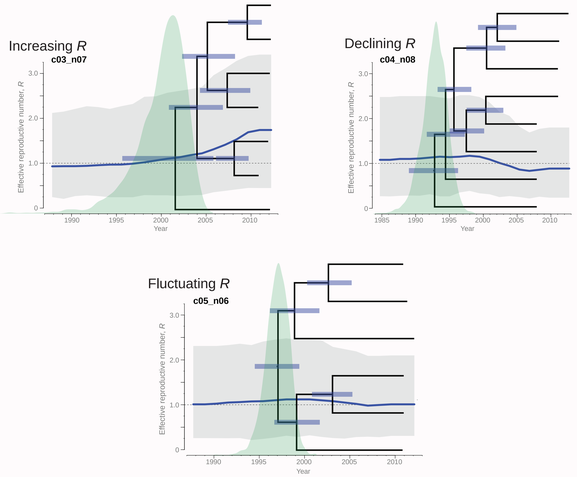
HIV-1C clusters with 5+ members identified by ML screening were grouped for phylodynamic analysis. The Bayesian Markov Bayesian Markov Cain Monte Carlo (MCMC) phylogenetic inference implemented in the BEAST package v.2.1.3 (Bouckaert et al., 2014) was utilized to estimate timing of HIV transmissions and temporal changes in trajectories of the HIV-1C lineage-specific effective reproductive number,*R*. The analysis employed a general time-reversible substitution model with a gamma distributed rate variation and proportion of invariant sites (GTR+Γ4+I), and uncorrelated log-normal relaxed molecular clock model and a Birth-Death Skyline Serial model (BDSKY) as a model for viral transmission (Stadler et al., 2013).



The following prior distribution of the BDSKY model parameters was used: LogNorm(0; 0.5) for effective reproductive number *R*; LogNorm(-0.5; 1) for the rate of becoming non-infectious *δ*; and Beta(35;15) for sampling rate *s*. The evolutionary rate for the analyzed V1C5 region of gp120 was set to 7.86E-03. The BEAST2 analyses were run until all relevant parameters converged, with 20% of the MCMC chains discarded as burn-in.

RESULTS

Trajectories of the effective reproductive number, *R*. The median estimates of R were within the range 0.72–1.77. The uncertainty illustrated by the 95% HPD intervals in grey varied across the identified HIV sub-epidemics in Mochudi.. Four HIV clusters (c03_n07, c10_n05, c13_n05, and c14_n05) had increasing *R*. These HIV lineages with the effective reproductive number *R* above the 1.0 threshold during the most recent time are interpreted as HIV sub-epidemics on the rise. The *R* trajectories in four other HIV clusters (c02_n09, c04_n08, c08_n06, and c09_n05) had decreasing R below the 1.0 threshold at present; hence these HIV sub-epidemics appear to have peaked in the past and to be declining..



HIV-1C V1C5 clusters with 5+ members

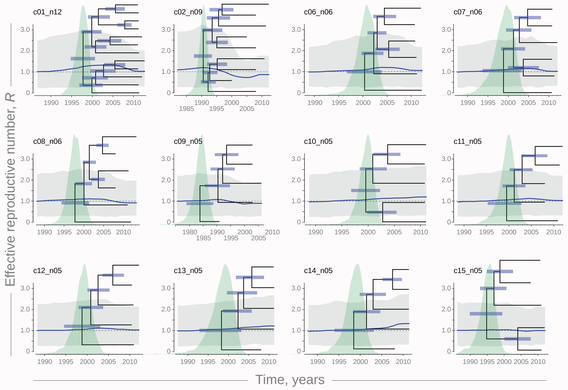
Cluster ID	Members per cluster			ML bootstrap support	Internode Certainty*	The most recent HIV transmission	Time to the most recent transmission	Acute, or historic sub-epidemic**
	Total	From Mochudi	From outside Mochudi					
c01_n12	12	12	0	0.92	0.71	Nov 2007	4.8	historic
c02_n09	9	3	6	0.88	0.88	Oct 1996	15.5	historic
c03_n07	7	7	0	1	0.99	Aug 2009	2.7	acute
c04_n08	8	6	2	0.8	0.84	Feb 2002	10.6	historic
c05_n06	6	6	0	0.99	0.98	Jan 2003	9.1	historic
c06_n06	6	6	0	1	1	May 2005	6.9	historic
c07_n06	6	6	0	0.86	0.53	Oct 2004	7.4	historic
c08_n06	6	6	0	0.9	0.89	Jun 2004	8.8	historic
c09_n05	5	1	4	0.94	0.96	Aug 1998	12.6	historic
c10_n05	5	5	0	0.92	0.72	Sep 2003	7.4	historic
c11_n05	5	5	0	1	1	Nov 2005	6.1	historic
c12_n05	5	5	0	1	1	Jan 2006	6.1	historic
c13_n05	5	5	0	1	1	Sep 2005	5.6	historic
c14_n05	5	5	0	0.8	0.72	Jun 2007	3.9	acute
c15_n05	5	4	1	0.81	0.62	Feb 2004	7.9	historic

* - most prevalent conflicting bipartitions

** - HIV sub-epidemic with (acute), or without (historic) HIV transmission over last 5 years

RESULTS

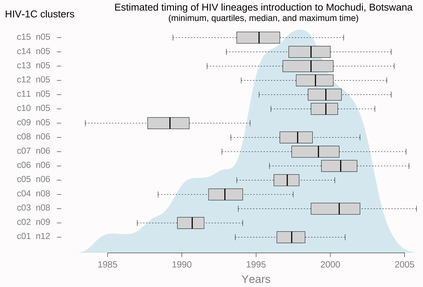
Trajectories of effective reproductive number R (blue curves) with 95% HPD (gray polygons), and MCC time-trees within analyzed HIV sub-epidemics with 5+ members.



Acute and historic HIV sub-epidemics. We distinguish two types of HIV sub-epidemics based on the presence or absence of recent HIV transmissions, e.g., within the last 5 years, within targeted clusters. HIV lineages with recent HIV transmissions and *R* trajectories on the rise were considered acute HIV sub-epidemics. HIV lineages without evidence of recent HIV transmissions and declining or fluctuating R trajectories were considered historic HIV sub-epidemics. In this study, a five-year period (from the most recent sampling date in the cluster) was chosen as the recency period..

Using a combination of the 5-year recency threshold and the shape of R trajectories, we identified 2 acute and 13 historic HIV sub-epidemics with 5+ members in Mochudi.

Two acute HIV sub-epidemics are c03_n07 and c14_n05. In these sub-epidemics, time between the most recent HIV transmission and the most recent sample is 2.7 and 3.9 years, respectively, and both demonstrate a clear increase of *R* trajectories towards the present.



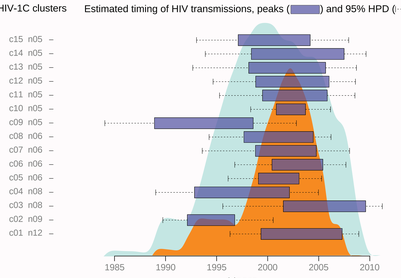
Introduction of HIV-1C lineages to Mochudi. The time HIV-1C lineages were introduced to Mochudi was estimated through the posterior distribution of the tree heights obtained for each HIV lineage. The barplots show the estimated time at which each sub-epidemic was introduced to Mochudi. The shaded density plot is a smoothed summary over all sub-epidemics.This summary suggests that the majority of HIV-1C lineages were introduced to Mochudi between 1996 and 2002.

RESULTS

Timing of viral transmissions within HIV-1C sub-epidemics in Mochudi. The internal nodes in the MCC time tree suggest time of viral transmissions. The projection of internal nodes on the time scale gives an idea on potential time of transmission events. The uncertainty of transmission time is reflected in the 95% HPD. Overall, this analysis provides estimates for the time of HIV transmissions within each HIV sub-epidemic in Mochudi.

The time of viral transmissions within these 15 HIV-1C sub-epidemics in Mochudi was estimated as the time interval between the oldest and the most recent viral transmission based on projection of internal nodes associated with terminal branches to the time line in the MCC time-trees. The estimated time interval for each sub-epidemic is interpreted as peaks of HIV-1C transmission for each HIV lineage (shown as boxes). The 95% HPD values of flanking nodes correspond to confidence intervals (shown as dashed lines).

The analysis suggests that the majority of HIV transmissions in Mochudi (within the 15 sub-epidemics analyzed) peaked over about a decade from about 1997 to 2007. The 95% HPD intervals spread across a broader time interval from about 1993 to 2009.



CONCLUSIONS

- We used phylogenetic linkage as a tool for breaking down the HIV epidemic in a single peri-urban village in Botswana into a series of phylogenetically distinct HIV sub-epidemics. The ultimate goal of breaking down a local epidemic into HIV sub-epidemics is to trace virus spread and to extinguish HIV transmission chains, one by one, by targeted interventions. If this approach works at the community level, it could be expanded to the global HIV epidemic.
- The important clinical and public health relevance of the HIV sub-epidemics approach lies in its ability to track the main driver of HIV transmissions in communities. To identify circulating HIV lineages in real time, spread of the virus in communities should be traced proactively.
- Knowledge of HIV spread is time-sensitive. Targeted interventions, such as enhanced HTC, linkage to care, and initiation of ART, could be applied to “acute” HIV sub-epidemics, i.e., sub-epidemics with recent viral transmissions. Targeted treatment-as-prevention (TasP) for the members of acute sub-epidemics could be more efficient and cost-effective than uniformly applied TasP, particularly if the scale-up of TasP takes time.
- We showed differential transmission dynamics in HIV sub-epidemics caused by different HIV lineages. We identified *acute* HIV sub-epidemics with evidence of recent HIV transmissions and distinguished them from *historic* sub-epidemics without recent HIV transmissions. In the context of high sampling density, our results suggest that only a small fraction of circulating HIV lineages are acute (with recent HIV transmissions). It is likely that only the acute sub-epidemics fuel HIV spread in communities at any given time.

REFERENCES

1. Novitsky V, Bussmann H, Logan A, Moyo S, van Widenfelt E, Okui L, Mmalane M, Baca J, Buck L, Phillips E, Tim D, McLane MF, Lei Q, Wang R, Makhema J, Lockman S, DeGruttola V, Essex M. Phylogenetic Relatedness of Circulating HIV-1C Variants in Mochudi, Botswana. *PLoS One*. 2013;8(12):e80589.
2. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T.G., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biogogy* 10, e1003537.
3. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A*. Jan 2 2013;110(1):228-233.