Assessing the HIV-1C Transmission Network in Botswana at Low Sampling Density

Poster 950

V. Novitsky[1,2], M. Zahralban-Steele[2], S. Moyo[1,2], D. Maruapula[1], B. Mokaleng[1], T. Nghisang[1], M.F. McLane[2], S. Madiba[2], E. van Widenfelt[1], T. Gaolathe[1], E. Kadima[1], S. Lockman[2], J. Makhema[1], S. Gaseitsiwe[1], and M. Essex[1,2]

[1]Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana [2]Harvard TH Chan School of Public Health, Boston, MA

CROI Conference on Retroviruses and Opportunistic Infections

ABSTRACT

Background: Phylodynamic assessment of HIV transmission clusters is essential for monitoring the HIV epidemic, better understanding of HIV transmission dynamics, and ultimately controlling the HIV epidemic. The phylogenetic structure of the HIV transmission network remains unknown.

Methods: Near full-length HIV-1C sequences were obtained from blood specimens collected within four Botswana Harvard AIDS Institute Partnership studies from 1996 - 2016. A total of 3,638 HIV-1C sequences originated from the South (48.3%; 15 communities), East (37.4%; 13 communities) and North (14.3%; 5 communities) of Botswana. The genotyping density (the number of analyzed HIV genomes as a proportion of the estimated total number of HIV-infected residents in the targeted area) varied between 0.2% and 31.1% (mean 10.6%, median 9.5%), which we consider to be a low sampling density. Near full-length genome HIV sequences were generated by Sanger sequencing (n=273) and next-generation sequencing (n=3,365). Phylogenetic relatedness among analyzed viral sequences was estimated by maximum likelihood using RAxML v.8 and the GTR+Γ4+I model. We defined an HIV cluster as a viral lineage that gives rise to a monophyletic subtree of the overall phylogeny with bootstrap support of splits >0.80 and median pairwised distance <10th quartile of the overall distribution of pairwise distances.

Results: We identified 553 phylogenetically distinct HIV-1C lineages circulating in Botswana and 496 (90%) of them included Botswana sequences. The identified HIV clusters had from 2 to 15 members. Among clusters with Botswana sequences, the proportion of local viral lineages (community-unique) seen in a single community was 25% (124 of 496). Among HIV-1C lineages spread across multiple communities, 49% (n=242) were found in two communities, 15% (n=74) were spread across three communities, and 11% (n=56) across 4+ communities. One viral lineage was spread across 11 communities. Regional analysis (South vs. East vs. North) demonstrated that 56% (279 of 496) of viral lineages were identified exclusively in the South, East, or North of the country. Among lineages seen in two communities, 129 were identified within, while 113 were spread between geographic regions.

Conclusions: The study revealed an HIV-1C transmission network with a complex structure. A substantial number of circulating phylogenetically distinct HIV-1C lineages were identified, although the genotyping density was relatively low. Twenty-five percent of viral lineages were local (community-unique), while about half of the identified lineages spread across two communities. The distribution of HIV-1C lineages within vs. between geographic regions (South, East and North) split 56% vs. 44%.

METHODS

Near full-length HIV-1C sequences were obtained from blood specimens sampled within four Botswana Harvard AIDS Institute Partnership (BHP) studies during 1996–2016. All studies included HIV genotyping as one of their core components and were approved by IRBs in Botswana and at the Harvard T.H. Chan School of Public Health. All participants provided written informed consent. All HIV-1C sequences in this study were represented by a single sequence per participant.

Near full-length HIV-1C genome sequences
We used sequence length of >7,000 bp. A total of 4,082 near full-length HIV-1C genome sequences included 3,638 (89%) sequences from Botswana and 444 (11%) sequences from outside of Botswana. The NGS was performed by the BioPolimers Facility (n=1,350) at Harvard Medical School and through collaboration with PANGEA HIV consortium (n=2,229) at the Welcome Trust Sanger Institute (Cambridge, UK).

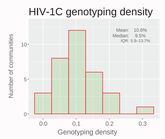
Multiple sequence alignment (MSA)
The near full-length HIV-1C sequences (n=4,082) were aligned by using *mafft* with the gap extension option *--op* 2. The poorly aligned regions across the MSA, such as the HIV-1 p2/p7/p1/p6 region of *gag*, V1, V2, V4 and V5 variable loops across HIV-1 *env* gp120, and variable regions within HIV-1 *vpu* and *nef* were extracted and re-aligned using *mafft* (*einsi*) with the gap extension options *--op* 0.01. All fragments were re-assembled into the final MSA by using the *phyloch c.genes()* in R ver. 3.4.0.

RESULTS

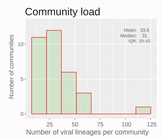
HIV genotyping density differed across the communities. Mean density was 10.6%, and median (IQR) genotyping density was 9.5% (5.9–13.7%).

Cluster size distribution. Overall, the identified HIV-1C clusters were relatively small and included from 2 to 15 members per viral lineage. The majority of identified clusters were dyads, accounting for 62% (309 of 496; 95% CI 58–67%) of clusters with Botswana sequences. Cluster size distribution followed the power law with an elongated tail of smaller numbers of relatively large clusters, including one cluster with 15 members, one with 14, and six clusters with 11 and 9 members.

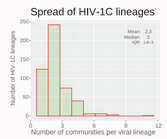
HIV-1C genotyping density



Community load



Spread of HIV-1C lineages

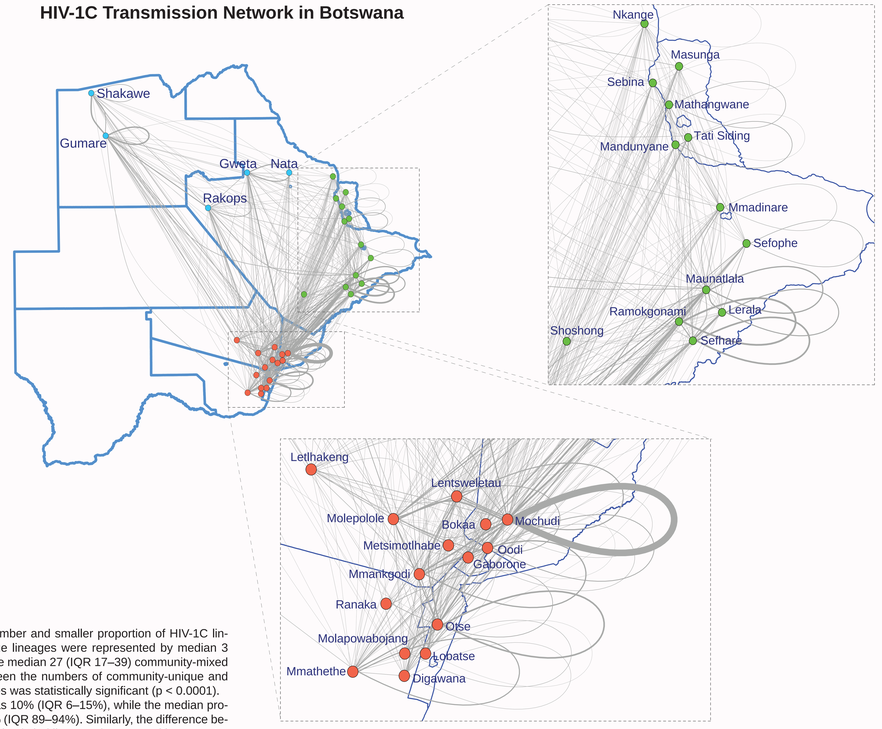


RESULTS

Phylogenetically distinct HIV-1C transmission clusters were identified based on bootstrap support of >=0.80 in the ML tree and median pairwise distances <10th quantile of the entire set of 4,082 near full-length HIV-1C genome sequences. A total of 553 phylogenetically distinct HIV-1C lineages were identified among those circulating in Botswana. The identified viral lineages included 1,604 (39%) near full-length HIV-1C genome sequences. The vast majority of identified HIV-1C lineages, 496 (90%), included HIV-1C sequences of Botswana origin (n=1,420; 89%). The majority of identified lineages, 482 (87%), included exclusively Botswana sequences, while 14 (3%) included both Botswana and non-Botswana sequences. The presented analysis is focused on HIV-1C transmission clusters (n=496) that include Botswana sequences (n=1,420).

Spread of HIV lineages
The distribution of identified HIV-1C lineages across Botswana communities/villages was analyzed using the number of viral lineages per community and the number of communities that the identified viral lineages were spread across. Median (IQR) of circulating HIV-1C lineages per community was 31 (19–43), ranging from 9 (Ranaka) to 112 (Mochudi).
The identified HIV-1C lineages were spread across a median (IQR) of 2 (1.8–3) communities. About 25% (124 of 496) of viral lineages were circulating in a single community (community-unique viral lineages). Among HIV-1C lineages spread across multiple communities, 49% (n=242) were found in two communities, 15% (n=74) were spread across three communities, and 11% (n=56) across 4+ communities. There was only one broadly spread viral lineage, which was found in 11 communities.
The number of identified lineages per community was associated with genotyping density (Spearman's rho = 0.61, p<0.001). We also found a strong association between the cluster size and the number of communities HIV-1C lineages were spread across (Spearman's rho = 0.71, p<0.0001). It should be noted that clusters with small numbers of members cannot be spread across a large number of communities. Hence, we do not expect to have many points in the upper-left triangle of the scatter plot.

HIV-1C Transmission Network in Botswana

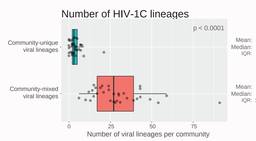


Community-unique vs. -mixed lineages
The community-unique viral lineages accounted for a low number and smaller proportion of HIV-1C lineages circulating across communities. The community-unique lineages were represented by median 3 (IQR 2–5) viral lineages per community. In contrast, there were median 27 (IQR 17–39) community-mixed lineages circulating in each community. The difference between the numbers of community-unique and community-mixed viral lineages circulating across communities was statistically significant (p < 0.0001).
The median proportion of community-unique viral lineages was 10% (IQR 6–15%), while the median proportion of community-mixed lineages in communities was 90% (IQR 89–94%). Similarly, the difference between the proportions of community-unique and community-mixed viral lineages in communities was statistically significant (p < 0.0001).

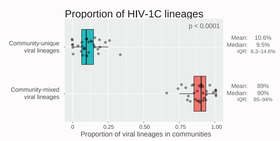
Geographic distribution of circulating HIV lineages
We analyzed the distribution of identified HIV lineages within and between three major geographic areas in Botswana: south, east, and north. Among the identified HIV-1C lineages with Botswana sequences, 33% (n=162) were found exclusively in the South, 18% (n=88) in the East and 6% (n=29) in the North. Among 496 clusters with Botswana sequences, 56% (n=279), were region-specific and were observed exclusively in the South, East or North of Botswana. The remaining 44% (n=217) of viral lineages were found in two (n=187) and three (n=16) geographic regions of Botswana, or included sequences of non-Botswana origin (n=14).
Among 242 lineages that were found in two communities, 50% (n=120) of clusters were identified within the same geographic region, while 47% (n=113) were spread across two regions. In addition, 9 HIV-1C lineages included both Botswana and non-Botswana sequences.

Age and gender
The distribution of participants age at enrollment revealed a gender gap among both clustered and non-clustered individuals. Men in clusters (median age 42 years old; IQR 35–49 years old) were older than women (median age 35 years old; IQR 29–44 years old), p<0.0001. Similarly, men outside of identified clusters (median age 43 years old; IQR 37–51 years old) were older than women (median age 38 years old; IQR 31–47 years old), p<0.0001.
Without taking gender into account, participants in clusters were younger that non-clustered participants (median age 37 years old; IQR 30–46 years old for clustered participants vs. median age 39 years old; IQR 32–48 years for non-clustered individuals; p<0.0001). Women in clusters were younger than women outside of clusters (median age difference 3 years; p<0.0001), while age difference between clustered and non-clustered men did not reach statistical significance (median age difference 1 year; p = 0.051).
Similar patterns of participants' age distribution—men older than women—were observed among clustered and non-clustered participants in the South, East and North.

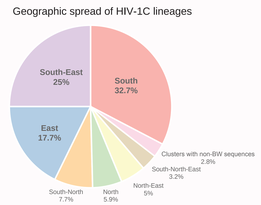
Number of HIV-1C lineages



Proportion of HIV-1C lineages



Geographic spread of HIV-1C lineages

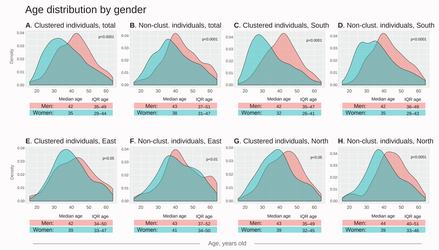


Age difference between men and women within clusters

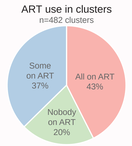


RESULTS

Age distribution by gender



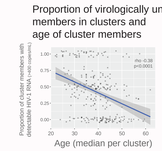
ART use in clusters
n=482 clusters



HIV-1 RNA in clusters and age of cluster members



Proportion of virologically unsuppressed members in clusters and age of cluster members



CONCLUSIONS

- The study assessed the country-wide structure of the HIV transmission network using thousands (n=4,082) of near full-length HIV genomes generated by next-generation sequencing.
- A total of 496 phylogenetically distinct HIV-1C lineages circulating in 33 Botswana communities from 1996 to 2016 have been identified at a mean genotyping density of 10%.
- The community load—the number of circulating HIV-1C lineages per community— median (IQR) was 31 (19–43) HIV-1C lineages per community.
- The spread of virus lineages—the number of communities each HIV-1C lineages was found in—was 2 (1.8–3) communities.
- About ¾ of circulating HIV-1C lineages in Botswana were spread across one or two communities, while ¼ of lineages circulate across 3+ communities.
- The community-unique lineages—lineages found in a single community—accounted for 25% of circulating lineages, while about a half of virus lineages, 49%, were found in exactly two communities.
- The proportion of community-unique viral lineages per community was 10% vs. 90% of community-mixed lineages.
- The distribution of HIV-1C lineages in three geographic regions—South, East and North—revealed that 56% of viral lineages were region-specific and were found in a single region, while 44% of lineages were spread across multiple geographic regions.
- The gender gap in clusters: men (median age 42) were older than women (median age 35). In about ⅓ of identified clusters, women were older than men.
- All cluster members were on ART in 43% of identified clusters, while nobody was receiving ART in 20% of viral lineages.
- Our results are in line with previous studies demonstrating higher HIV-1 RNA levels among clustered than among non-clustered individuals. The proportion of individuals with detectable HIV-1 RNA was higher among clustered than among non-clustered individuals.
- All cluster members were virologically unsuppressed in 20% of clusters.
- All cluster members were virologically suppressed in 30% of clusters.
- Levels of HIV-1 RNA in clusters were not associated with cluster size or the extent of viral lineage spread (the number of communities in which the viral lineage was found).
- Viral load in clusters and the proportion of cluster members with detectable viral load were negatively associated with age suggesting higher levels of HIV-1 RNA in young HIV-infected people.
- Levels of HIV-1 RNA were not associated with age difference in clusters.

This study was supported by the US President's Emergency Plan for AIDS Relief (PEPFAR) through CDC under the terms of cooperative agreement U01 GH000447, in part by NIH-funded AI47067, AI092715 and AI083036, and through collaboration with PANGEA HIV consortium.