HIV Phylodynamics in North Carolina: Detecting "Active" Clusters for Intervention

Conference on Retroviruses and Opportunistic Infections
Boston, Massachusetts
February 22-25, 2016

Ann M. Dennis[1], Stéphane Hué[2], Joseph Sebastian[3], Victoria Mobley[4], William C. Miller[1], Joseph J. Eron[1]

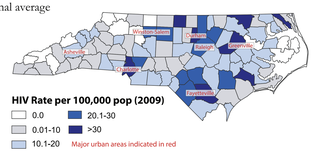
[1]University of North Carolina, Chapel Hill, NC, [2]London School of Hygiene and Tropical Medicine, London, UK, [3]Laboratory Corporation of America, Research Triangle Park, NC, [4]Communicable Diseases Branch, North Carolina Department of Health and Human Services, Raleigh, NC

Ann M. Dennis, MD
Division of Infectious Diseases
adennis@med.unc.edu

Background

- Phylogenetic analyses of HIV sequences can be used to monitor epidemic trends and help target prevention through the identification of transmission clusters [1]
- Transmission clusters can help identify local HIV epidemic drivers by revealing links in subepidemics overlapping in geography, time, and risk behaviors.
- Characterizing such clusters is needed in North Carolina, where >1,500 new HIV diagnoses are reported annually despite widespread HIV prevention [2]

Fig. 1. Rate of new HIV diagnoses. Several counties have rates above the national average



Objectives

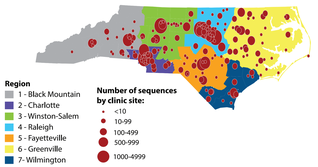
- To characterize HIV transmission clusters across NC using HIV-1 pol sequences collected statewide
- To evaluate transmission cluster expansion and epidemic growth to detect "active" clusters as potential intervention targets

Methods

Study Population and Sequences

- Cross-sectional analyses of HIV-1 infected individuals ≥18 years receiving care in NC who had at least one genotype performed at LabCorp
- LabCorp is the largest commercial laboratory in NC and conducts the majority of HIV genotyping for clinics across the state (Fig 2)
- Full length PR and partial RT sequences were extracted from genotypes collected between 1997 and mid-2014
- Sequences were aligned with MUSCLE [3] and edited manually in BioEdit [4] with stripping of gapped positions
- Subtypes were identified with SCUEAL [4] and REGA. [5] Major drug resistance mutations identified using the Stanford HIVdb Program [6]

Fig. 2. Locations clinics and number of sequences sent by clinic site.



Identifying Transmission Clusters

- First available sequences per individual used in all analyses
- A maximum-likelihood (ML) tree was constructed in FastTree [7] with the GTR model and support clades assessed with local support values (Shimodaira-Hasegawa [SH] test)
- Putative transmission clusters were defined as clades with <3.5% maximum pairwise genetic distance difference and branch support ≥0.90
- Clusters were identified using the automated tool ClusterPicker v1.3 [8]
- Individuals who had genotypes sent from more than one clinic were identified by 1) identical birth dates 2) membership in same clusters

Time-scaled Analyses

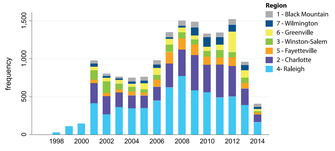
- Clusters identified in ML tree were then analyzed in BEAST v.1.8.2 [10] for confirmation and to estimate the time of origin (tMRCA)
- Initial ML tree was split into 39 sequence files to decrease computation time
- Used SRD06 substitution model, a lognormal relaxed molecular clock, and the Bayesian Skyline coalescent prior
- Bayesian Markov chain Monte Carlo (MCMC) chains were run from 50-100 million generations with 1-5 runs per file to achieve an effective sample size (ESS) >200 (assessed using Tracer v1.6.0)
- Tree files were combined using LogCombiner v1.8.2 with 10% burn-in
- Maximum-clade credibility trees summarized with TreeAnnotator v.1.8.2
- Basic reproductive number (R0) was estimated for large clusters (≥22 members) using the birth-death model in BEAST v.1.8.2 [11]

Results

Study population

- 24,973 pol sequences suitable for analysis from 15,247 individuals (only first available sequence analyzed)
- 71% men, median age 40 years (IQR 32-48)
- Median sample year 2008 (IQR 2005-2011)
- Most sequences sampled in Region 4 (Raleigh, 43%) and Region 2 (Charlotte, 25%)

Fig. 3. Number of sequences sampled per year by clinic location (n=15,247)



Putative Transmission Clusters

- 7,647 (50%) sequences were identified in 2,318 clusters
- Median cluster size 3 members

Fig. 4. Number of individual cases by cluster size

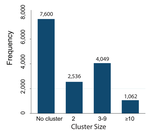
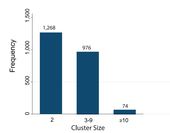


Fig. 5. Number of clusters by size



Transmission Cluster Characteristics

- Composition of clusters varied by cluster size. Sequences in clusters were significantly more likely to sampled more recently, from younger patients, and from Raleigh and Charlotte metropolitan areas.

Characteristic	Total n (%)	No Cluster n (%)	In Clusters 2-9 members, n (%)	In Clusters 10-36 members, n (%)
Total	15,247	7,600	6,585	1,062
Sex				
Men	10,643 (71)	5,215 (70)	4,582 (70)	846 (81)*
Women	4,359 (29)	2,232 (30)	1,923 (30)	204 (19)
Age				
18-30	3,417 (22)	1,111 (15)	1,786 (27)	520 (49)*
31-50	9,193 (60)	4,936 (65)	3,808 (58)	449 (42)
>50	2,637 (17)	1,553 (20)	991 (15)	93 (8)
Sample Year				
1997-2002	2,080 (14)	1,342 (18)	706 (11)	32 (3)
2003-2008	6,114 (40)	3,142 (41)	2,630 (40)	342 (32)
2009-2014	7,053 (46)	3,116 (41)	3,249 (49)	688 (65)*
Region of Sampling				
1 –Black Mountain	732 (5)	388 (5)	319 (5)	25 (2)
2 – Charlotte	3,850 (25)	1,907 (25)	1,709 (26)	234 (22)
3 – Winston-Salem	1,084 (7)	620 (8)	428 (7)	36 (3)
4 – Raleigh	6,573 (43)	3,176 (42)	2,821 (43)	576 (54)*
5 – Fayetteville	1,182 (8)	543 (7)	532 (8)	107 (10)
6 – Greenville	1,075 (7)	587 (8)	461 (7)	27 (2)
7 – Wilmington	751 (5)	379 (5)	315 (5)	57 (5)
Subtype				
B	14,920 (98)	7,360 (97)	6,498 (99)	1,062 (100)*
Non-B	327 (2)	240 (3)	87 (1)	0
Major Drug Resistance†				
Any (WHO list)	5,528 (36)	3,261 (43)	2,036 (31)	231 (22)*

*P-value <0.001 in bivariate analyses across groups comparing in cluster vs. not in cluster using chi-square test
†Antiretroviral drug exposure data unavailable

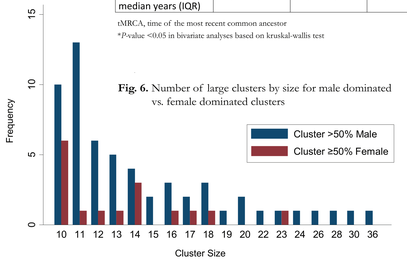
Time Scaled Clusters

- All clusters ≥10 members were confirmed with high support in BEAST (posterior probability >0.99).
- 74 large clusters identified involving 1,062 individuals
- All were subtype-B and contained more men than smaller clusters (81% vs. 70%)
- Over one fifth of large clusters (22%) were ≥50% women; these clusters had earlier origin dates and longer time spans compared to male dominated clusters

Characteristic	Total	Male Clusters (>50% men)	Female Clusters (≥50% women)
Number of clusters	74	58	16
Cluster size, median (IQR)	13 (11-16)	13 (11-17)	13 (10- 15)
tMRCA, median year (IQR)	1998 (1995-2001)	1999 (1995-2002)*	1996 (1993-1997)*
Cluster time span, median years (IQR)	15 (12-19)	15 (12-19)*	18 (16-19)*

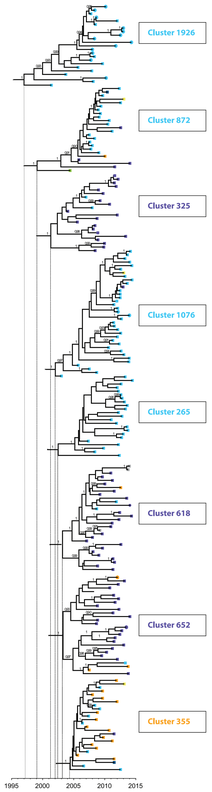
tMRCA, time of the most recent common ancestor
*P-value <0.05 in bivariate analyses based on kruskal-wallis test

Fig. 6. Number of large clusters by size for male dominated vs. female dominated clusters



Dynamics of Large Clusters

Fig. 7. Phylogenies of the 8 largest clusters detailed with tMRCA (1997-2004)



Characteristic	Cluster 1926	Cluster 872	Cluster 325	Cluster 1076	Cluster 265	Cluster 618	Cluster 652	Cluster 355
Size	23	24	22	36	23	30	28	26
% Male	**43%**	100%	100%	94%	100%	97%	93%	100%
Dominant Region	4	4	2	4	4	2	2	5
Major DRM	---	---	---	---	---	---	---	**Y188L**
Avg. genetic distance	1.9%	1.2%	1.2%	1.8%	1.6%	1.7%	2.0%	0.9%
tMRCA (year)	1997	1999	2001	2002	2002	2003	2003	2004
Cluster time span (years)	17.4	15.0	12.1	12.4	12.0	11.3	10.9	8.8
Transmissions per year	1.3	1.6	1.8	2.9	1.9	2.7	2.6	3.0
R_{0t}, median (95% HPD)	1.2 (0.7-2.1)	1.6 (0.8-3.1)	1.2 (0.7-2.3)	**1.9 (1.0-3.5)**	**4.1 (1.1-12.6)**	**1.9 (1.0-3.9)**	**3.4 (1.0-10.3)**	2.0 (0.9-4.7)

DRM, drug resistance mutation
tMRCA, time of the most recent common ancestor

- 8 clusters ranged in size from 22-36 members
- Originated between 1997-2004
- Clusters spanned mean of 12.5 years and with mean 2.2 transmissions/year
- Most clusters homogenous by geographic sampling region
- 75% of clusters had highly supported internal nodes 2010-2014 indicating cluster expansion
- One cluster 57% women (Cluster 926) and one cluster with all members with major drug resistance (Cluster 355)



Conclusions

- A high proportion of sequences were identified in transmission clusters with high branch support indicating significant local HIV transmission
- While this dataset offers denser sampling across the population, limitations include lack of clinical data such as antiretroviral exposure and risk behavior
- However, important insights into transmission dynamics can be gained from these sequence data such as distinguishing between "active" and historic clusters. Of the largest clusters analyzed, all originated over a decade ago but most (75%) had expanded since 2010
- Transmission dynamics of male dominated clusters differed compared to clusters with women, with shorter time spans and more recent origin
- Targeting interventions towards active clusters which include more recent transmissions and a high potential for further onward transmission could more effectively reduce HIV incidence
- Enhanced prevention interventions to active clusters are planned to halt further progression

References

1. Dennis, A.M., et al., Phylogenetic Studies of Transmission Dynamics in Generalized HIV Epidemics: An Essential Tool Where the Burden is Greatest? J Acquir Immune Defic Syndr, 2014. 67(2): p. 181-95.
2. North Carolina Department of Health and Human Services. 2013 HIV/STD Surveillance Report, http://epi.publichealth.nc.gov/cd/stds/figures/std13rpt.pdf.
3. Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 2004. 32(5): p. 1792-1797.
4. Hall, T., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 98/98/NT. Nucl Acids Symp Ser, 1999. 41: p. 95 - 98.
5. Pond, S.L.K., et al., An Evolutionary Model-Based Algorithm for Accurate Phylogenetic Breakpoint Mapping
6. http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/
7. http://sierra2.stanford.edu/sierra/servlet/JSierra
8. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular biology and evolution, 2009. 26(7): p. 1641-1650.
9. Ragonnet-Cronin, M., et al., Automated analysis of phylogenetic clusters. BMC Bioinformatics, 2013. 14: p. 317.
10. Drummond, A.J. and A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology, 2007. 7: p. 214.
11. Stadler, T., et al., Estimating the basic reproductive number from viral sequence data. Mol Biol Evol, 2012. 29(1): p.

Acknowledgements

The project described was supported by the National Institute of Allergy and Infectious Diseases (NIAID), NIH, through Grant Award Number K08AI112432-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH."