Published in final edited form as:

AIDS. 2012 September 10; 26(14): 1813–1822. doi:10.1097/QAD.0b013e3283573244.

Phylogenetic insights into regional HIV transmission

Ann M. Dennis, MD MS¹, Stéphane Hué, PhD², Christopher B. Hurt, MD¹, Sonia Napravnik, PhD^{1,3}, Joseph Sebastian, PhD⁴, Deenan Pillay, MBBS PhD², and Joseph J. Eron, MD^{1,3}

¹Division of Infectious Diseases, University of North Carolina, Chapel Hill, NC, USA

²Department of Infection, University College London, London, UK

³Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

⁴Center for Molecular Biology and Pathology, Laboratory Corporation of America®, Research Triangle Park, NC, USA

Abstract

Objectives—Despite prevention efforts new HIV diagnoses continue in the Southern US, where the epidemic is characterized by significant racial/ethnic disparities. We integrated phylogenetic analyses with clinical data to reveal trends in local HIV transmission.

Design—Cross-sectional analysis of 1671 HIV-infected individuals each with one B-subtype *pol* sequence obtained during chronic (82%; UNC Center for AIDS Research Clinical Cohort) or acute/recent (18%; Duke/UNC Acute HIV Consortium) infection.

Methods—Phylogenies were inferred using neighbor joining to select related sequences then confirmed with Bayesian methods. We characterized transmission clusters (clades $n \geq 3$ sequences supported by posterior probabilities=1) by factors including race/ethnicity and transmission risk. Factors associated with cluster membership were evaluated for newly diagnosed patients.

Results—Overall, 72% were male, 59% black and 39% MSM. A total of 557 (33%) sequences grouped in either 108 pairs ($n=216$) or 67 clusters ($n=341$). Clusters ranged from 3–36 (median 4) members. Composition was delineated primarily by race, with 28% exclusively black, and to a lesser extent by risk group. Both MSM and heterosexuals formed discrete clusters though substantial mixing was observed. In multivariable analysis, patients with age ≥ 30 years ($P=0.009$), acute infection ($P=0.02$), local residence ($P=0.002$), and transmitted drug resistance ($P=0.02$) were more likely to be cluster members while Latinos were less likely ($P<0.001$).

Conclusions—Integration of molecular, clinical and demographic data offers a unique view into the structure of local transmission networks. Clustering by black race, youth and TDR and inability to identify Latino clusters will inform prevention, testing and linkage to care strategies.

Corresponding author: Ann M. Dennis, Division of Infectious Diseases, 130 Mason Farm Road Suite 2115, CB# 7030, Chapel Hill, NC 27599-7030, T 919-843-7760, F 919-966-6714, adennis@med.unc.edu.

Conflicts of Interest and Source of Funding: AD has received grant support from the Bristol-Myer Squibb Virology Fellows research training program. SN has received grant support from Pfizer, Bristol-Myers Squibb, and Merck. JS is an employee of Laboratory Corporation of America. JJE has received consulting fees from Tibotec, Bristol-Myers Squibb, Merck, GlaxoSmithKline and ViiV, lecture fees from Roche and Bristol-Myers Squibb, and grant support from GlaxoSmithKline, Merck, ViiV and BMS.

Parts of this work were presented at the 6th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention in Rome, Italy (Abstract #MOAC0205).

Author contributions: Patient data (SN, CBH, JJE), HIV *pol* nucleotide sequence data (CBH, JS), planning of analyses (AD, SH, DP, JJE), phylogenetic analyses (AD, SH), statistical analyses (AD, SN), preparation of tables and figures (AD), drafting manuscript (AD, JJE), manuscript review and edit (all).

Keywords

Molecular epidemiology; HIV transmission; Risk Factors; Acute Infection; Southeast US

INTRODUCTION

New HIV diagnoses continue in the Southern United States¹, the region with the largest number of AIDS cases in the nation and an above average rate of new diagnoses[1], despite the presence of comprehensive prevention efforts. As in other Southern states, the HIV epidemic in North Carolina (NC) is characterized by significant racial and ethnic disparities. In 2009, blacks accounted for 66% of new infections, with an infection rate over 9 times higher than whites[2]. Similarly, Latinos represent a growing proportion of the NC epidemic, with a significant rise from 1% to 8% of new cases between 1995 and 2005[3]. Although HIV testing efforts have increased following calls for routine screening in 2006[4], this has only modestly increased new diagnoses. Additionally, presentation late in the course of infection remains common[5], particularly among Latinos [6].

The failure to fully reach all individuals at risk for HIV infection, particularly disenfranchised minorities with limited social mobility, highlights the need for continued innovative approaches to better understand HIV transmission on a population level. Prevention programs must target factors contributing to onward HIV transmission locally in order to successfully diminish HIV incidence. Notably, HIV epidemics are heterogeneous, composed of a series of overlapping local sub or micro-epidemics defined by risk groups, temporal variation and localized geographic areas[7–9]. Uncovering these sub-epidemics early is challenging because HIV is often diagnosed years after transmission, making identification of the actual source or mode of exposure difficult. HIV sequence data can be used to characterize the structure of epidemics[10, 11] and when paired with epidemiologic data may provide unique insights into groups responsible for ongoing transmission [12]. This approach to HIV epidemiology is now possible through advances in genomic sequencing, statistical methods, and computational speed[13]. Utilizing HIV *pol* sequences, derived from genotypes sent during clinical care, results in increased sampling density to uncover trends in transmission at the population level[11, 14].

The objective of this study was to use phylogenetic analyses of HIV *pol* sequences in conjunction with clinical and demographic data to identify which groups in NC contribute to ongoing HIV transmission – a complementary approach to understanding HIV propagation which may otherwise be invisible by patient history alone. We characterized the composition of phylogenetically reconstructed “clusters”, or groups of people where multiple transmissions likely occurred, and assessed factors associated with membership in these clusters among patients diagnosed from 2000–2009.

METHODS

Study Population

We performed a cross-sectional evaluation of patients who had a HIV-1 genotype available for analysis and who participated in one of two cohorts based on the duration of HIV infection at diagnosis: (1) “unknown/chronic” or (2) “acute/recent”. The unknown/chronic cohort includes patients attending the University of North Carolina (UNC) Infectious Diseases Clinic who enrolled in the UNC Center for AIDS Research HIV Clinical Cohort

¹Southern region includes the following states: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia

(UCHCC). An estimated 98% of clinic patients participate in the UCHCC, providing an accurate representation of the HIV clinic population. To enroll in the UCHCC, participants must be 18 years old and provide written informed consent in English or Spanish.

The acute/recent cohort included patients in the Duke-UNC Acute HIV Consortium or NC Screening and Tracing Active Transmission (STAT), statewide programs tracking acute infections through publicly funded clinics with >99% participation; both have been previously described in detail [15–17]. Acute infection was defined as either a combination of non-reactive ELISA or an indeterminate western blot (WB) paired with a positive HIV RNA or p24 antigen test, or a negative ELISA and WB within 45 days of a positive ELISA or WB. A recent infection was defined as either (1) a documented negative ELISA or WB within 45–180 days of a documented positive test, or (2) a positive HIV RNA result, no detectable evidence of antiretroviral therapy (ART) on the basis of HPLC with ultraviolet detection (Clinical Pharmacology and Analytical Chemistry Laboratory, UNC CFAR, Chapel Hill, NC; lower limit of detection 10–25 ng/ml), and results consistent with a duration of infection <180 days on both a less sensitive ELISA (Vironostika, bioMérieux, Marcy-l'Étoile, France; standardized optical density cutoff <1.0) and an avidity-modified third-generation immunoassay (Bio-Rad Laboratories, Hercules, CA; avidity index cutoff <40).

Sequences and Phylogenetic Reconstruction

Full length protease (PR) and partial reverse transcriptase (RT) sequences were extracted from commercially performed genotypes obtained for clinical care. For UCHCC patients, 95% of assays were HIV GenoSure® or GenoSure® Plus (Laboratory Corporation of America®, Research Triangle Park, NC), which spans PR and RT codons 1–400. Sequences from the acute/recent cohort were derived from GenoSure® or the TRUGENE® HIV-1 assay (Siemens Healthcare Diagnostics, Tarrytown, NY), which spans PR and RT codons 1–250. For patients with multiple sequences, only the oldest sequence was used. Sequences were aligned and edited with Se-Al v2.0 [18]. Subtypes were identified using the Subtype Classification using Evolutionary Algorithms [19]. Only B-subtype sequences, which represent >98% of subtypes in the area, were included in final analyses. We deposited the sequences from the UCHCC patients which were used in these analyses into GenBank under accession numbers JX160108–JX161480. Codon positions associated with major drug resistance mutations (DRM) according to the IAS-USA 2009 list [20] were initially removed to avoid potential treatment-induced convergent evolution.

Using a neighbor-joining (NJ) phylogenetic tree, reconstructed under the HKY85 model of nucleotide substitution [21] in PAUP* v4.0 [22], we selected sequences that differed by <4.5% pairwise genetic distance from at least one other sequence [11] to scale the dataset to a manageable size. Transmission clusters were confirmed by Bayesian MCMC inference in Mr Bayes [23] using the general time reversible model of nucleotide substitution with a proportion of invariant sites (ι) and gamma distribution of rates (Γ), sample frequency of every 1000th generation, and a 10% burn-in. Convergence of the estimates was evaluated with generation vs. log probability plots in Tracer v.1.5 [24] using an Effective Sample Size >200. Due to size constraints, the NJ tree was divided in quarters and each run separately in Mr Bayes. Maximum clade credibility trees were generated with a 10% burn-in using Tree Annotator [25]. We further reconstructed a maximum-likelihood (ML) tree (undivided) under the same model conditions in RAxML v.7.0.4 [26], to ensure the quarter trees were robust. Finally, Bayesian analyses were repeated using complete sequences, evaluating the third codon position only, for additional confirmation of the phylogenetic clusters. These confirmatory analyses yielded trees of similar topology to the initial Bayesian trees (data not shown). To further support identification of “local” clusters, a ML tree was reconstructed in Fast Tree [27] with all study sequences and an additional 595 controls from the HIV

Genbank database using Viroblast[28] to identify the top ten related sequences for each study sequence. Clusters split by control sequences or with inconsistent topology to our Bayesian or RAXML trees were not considered robust.

Variables

Demographic, risk behavior, and clinical data were abstracted from medical records. The UCHCC has standardized data extraction methods from medical charts and institutional databases [29]. We evaluated age at diagnosis, sex, race/ethnicity, year of diagnosis, risk group, ART exposure, DRMs, and geographic residence. Geographic residence was defined as living within the primary catchment area for the UNC clinic, determined by a 16 contiguous county area (out of 100 NC counties) where ~75% of UCHCC patients reside. For ART-naïve individuals, mutations associated with transmitted drug resistance (TDR) were identified using the 2009 standardized surveillance list from the World Health Organization [30]. The CD4+ T lymphocyte (CD4) count and HIV RNA viral load closest to the date of diagnosis were recorded. Transmission risk was categorized as men who have sex with men (MSM), heterosexual, intravenous drug use (IDU) or other/unknown. Patients reporting IDU in addition to another risk were classified as IDU.

Outcomes and Definitions

We defined transmission “clusters” and “pairs” as phylogenetic clades with $n \geq 3$ and $n=2$ sequences respectively, and supported by Bayesian posterior probabilities=1. Because we utilized a relaxed genetic distance definition, we also conducted sensitivity analyses on clusters which satisfied both Bayesian posterior probabilities=1 and mean intra-cluster pairwise genetic distance difference of $\leq 1.5\%$ [31]. Cluster composition was evaluated by comparing characteristics among members, with “Predominant” clusters defined as those in which $\geq 2/3$ (66%) of members share the same characteristic. “Homogeneous” clusters were those in which all members shared a characteristic. Finally, we assessed factors associated with cluster and pair membership for patients diagnosed from 2000–2009. Although pairs represent observed transmission, we distinguished between pairs and clusters because many pairs are identified by partners who enter care together or through partner contact tracing, and may not necessarily indicate further onward transmission.

Statistical Analyses

For newly diagnosed patients, factors associated with membership in pairs and clusters were evaluated. Differences in categorical variables were tested with the Pearson’s χ^2 test and continuous variables with the Kruskal-Wallis test. Multivariable analyses were fit using logistic regression to identify independent predictors of cluster membership. We first fit a full model based on results of our bivariable analyses and then used backwards elimination to arrive at a final model that included only factors predictive of cluster membership based on a two-sided alpha of 0.05. All data were analyzed using STATA version 11.0 (StataCorp, College Station, TX).

RESULTS

Study Population

A total of 1694 sequences were available, acquired from 1997–2009, with each sequence from an individual patient. Of these, 23 (1.4%) sequences were non-B subtypes and eliminated. The final dataset ($n=1671$) consisted of 1373 (82%) sequences from the chronic cohort and 298 (18%) from the acute/recent cohort. In the UCHCC, these patients represent 57% of all participants, which consists of patients diagnosed in the past 20 years with $>50\%$ entering care after 2000. Although genotypes from ART-naïve individuals were not

common until 2007, 312 of these patients had genotypes analyzed from specimens archived prior to ART-exposure. The UCHCC patients with genotypes were similar to those without genotypes by sex (71% vs. 69% male; $P=0.39$), race (both 59% black), and risk (37% vs. 40% MSM, $P=0.10$). While our dataset includes a minority (estimated <20%) of sequences from all cases reported to the state in our primary catchment area from 2000–2009, it is highly representative of these cases by ethnicity (66% black; 7% Latino), sex (72% male), and risk (41% MSM). Additionally, among our study samples, the race/ethnic distribution per year (Supplemental Figure 1) was roughly similar to the clinic demographics.

The overall population was 59% black and 72% male, with over half receiving an HIV diagnosis after 2000 (Table 1). Over half (54%), were ART-naïve at the time of sequence acquisition and median time from diagnosis to sampling was 658 days (IQR 18–2959). The newly diagnosed patients consisted of slightly more Latinos (10% vs. 7%), and less patients reporting IDU (11% vs. 5%) but were otherwise similar to the entire population. Among the newly diagnosed patients who had a genotype before ART ($n=775$; 87%), TDR prevalence was 12%, and TDR was more common among patients diagnosed with acute or recent infection than those with an unknown duration (17.5% vs. 8.6%; $P<0.001$).

Transmission Cluster Composition

A total of 557 (33%) of sequences grouped in either 108 pairs ($n=216$; 13%) or 67 clusters ($n=341$; 20%). The phylogenetic tree is shown in Supplemental Figure 2. Clusters ranged from 3–36 sequences (median 4; IQR 3–5). The median intra-cluster pairwise genetic distance was 0.020 (IQR 0.012–0.029) nucleotide substitutions/site. Although the majority of clusters were small, two large clusters (13 and 18 members) with relatively short mean genetic distances were identified (Figure 1). While most clusters were composed of patients in the chronic cohort, 42 (63%) clusters included at least one member with acute or recent infection. Notably, 74% of these clusters included at least one chronic patient diagnosed prior to the acute/recent(s), suggesting that local transmission may not be dominated by acute-to-acute. The sensitivity analysis of the more strictly defined 1.5% cutoff yielded 103 pairs ($n=206$; 12%) and 33 clusters ($n=122$; 7%) with median 3 members (range 3–13). We examined cluster composition by various characteristics including race/ethnicity, risk, duration of infection, age at diagnosis, and presence of TDR. Overall trends were very similar between the two cluster definitions (Table 2).

Overall, clusters were defined by both racial composition and transmission risk. Most clusters ($n=45$; 67%) were predominant black and 19 (28%) were 100% black. A smaller proportion of clusters were predominant white and none were predominant Latino. Most Latino sequences (57%) were found in predominant black clusters.

By transmission risk, we found a substantial degree of mixing but also discrete transmission in both MSM and heterosexual groups. Of predominant risk clusters, 28 (42%) were MSM and 23 (34%) heterosexual; 12 and 10 of these clusters respectively were homogeneous (Table 2). Cluster sizes among both groups were similar (mean 3.7 MSM vs. 4.7 heterosexual members; $P=0.60$). No clusters were predominantly IDU and nearly 25% were mixed, mostly between heterosexuals and MSM. Among sequences that clustered, 102 (71%) from MSM grouped in predominant MSM clusters while only 5% fell in majority heterosexual clusters. Of heterosexuals, 91 (60%) grouped in predominant heterosexual clusters and 9% grouped in predominant MSM clusters. Both black and white women were equally likely to be members of heterosexual clusters (55% vs. 62%) but among men, blacks were more likely to be in heterosexual clusters than whites (21% vs. 9%; $P=0.02$). Among heterosexual men, 72% of whites and 57% of blacks were in predominant mixed or MSM clusters ($P=0.28$).

Transmission Clusters with Transmitted Drug Resistance

Of all clusters, 17 (25%) included at least one member with TDR. Of these, six had 60% of members with TDR, and in 5 of these, all members shared the same TDR mutation(s) (Table 3), and were among the newly diagnosed subset. These homogenous TDR clusters were smaller (mean 3.0 vs. 4.2 members; $P=0.16$) and had significantly smaller mean genetic distance (0.009 vs. 0.020 substitutions/site; $P=0.01$) compared to the clusters that had no members with TDR.

Factors Associated with Membership in Transmission Clusters

Of 889 patients who were newly diagnosed, 154 (17%) and 231 (26%) were members of pairs and clusters, respectively (Table 4). Generally, no statistically significant differences were found between patients in pairs and those who were not clustered, though a higher proportion of Latinos were in pairs. Cluster members were more likely to be younger (median age 30 [IQR 23–41] vs. 34 [IQR 26–43] years; $P<0.001$) than non-members. Latinos were the least likely of all racial/ethnic groups to be clusters members (9% vs. 30% blacks vs. 24% whites; $P<0.001$). Those with acute infections were also more likely to be in clusters (35% vs. 25% chronic; $P=0.02$) as were patients residing in UNC's primary catchment area (27% vs. 22%; $P=0.02$). Among the 775 ART-naïve patients, 92 (12%) had TDR. Of these, 33 (36%) were cluster members, a significantly higher proportion compared to 174 (25%) of cluster members without TDR ($P=0.03$). All these variables remained associated with cluster membership in the multivariable regression model (Table 4). Analysis of the more tightly defined 1.5% clusters showed similar overall trends, though with reduced levels of statistical significance for some variables due to lower sample size (Supplemental Table).

DISCUSSION

Our study represents the largest phylogenetic analysis of HIV *pol* sequences from the Southeastern US, a region plagued by a substantial burden of HIV/AIDS cases. We found that in our HIV cohort, representing both chronic and acute infections, 20% of *pol* sequences derived from genotypes formed transmission clusters involving three or more individuals and were delineated primarily by race and, to a lesser extent, risk groups. Although we observed substantial mixing between MSM and heterosexuals, both groups formed discrete clusters and were equally likely to contribute to onward transmission among newly diagnosed patients. In contrast to other studies [32, 33], we found similar cluster sizes among risk groups as well as several large heterosexual clusters, suggesting that the local transmission structure is different compared to regions where MSM transmission predominates. Further, among new patients, non-Latino ethnicity, younger age, acute infection, presence of TDR, and local residence were all independently associated with cluster membership. Our results not only offer a glimpse into the structure of local HIV transmission, but more importantly, pave the foundation for future work exploiting the now growing repositories of *pol* sequences in the US.

Within transmission clusters, we observed stronger trends with respect to race, especially black race, rather than risk groups. While 28% of clusters were exclusively black, only <18% were exclusively MSM or heterosexual. Few phylogenetic studies to date have explored racial/ethnic differences in transmission clusters [34, 35] though these methods have been used to evaluate migration and domestic transmission in Europe [33] and to assess risk groups [11, 14, 36]. The racial homogeneity of our clusters parallels alternative analyses showing high rates of assortative sexual mixing, or selecting partners of the same race/ethnicity, among US blacks [37] and black MSM [38, 39]. The weaker trends seen among risk groups may be due to several factors. Since risk groups are self-reported, they

may reflect sexual identity rather than actual routes of infection. Additionally, men who identify as MSM may still have sex with women[40]. The substantial mixing between risk groups in our study suggests possible underreporting of MSM or bisexual behavior, which has been noted in other phylogenetic studies [32]. Among heterosexually-identified men who were cluster members, the majority of both whites and blacks grouped in predominant MSM or mixed clusters. We did not find racial differences in the degree of risk group mixing, further countering the hypothesis that black MSM largely contribute to the heterosexual epidemic through undisclosed MSM activity [41].

Latinos were much less likely to be members of transmission clusters compared to both blacks and whites. This finding was surprising, as the sampling density was representative of our clinic demographics, and because Latinos in NC have nearly three times the HIV incidence rate compared to whites[2]. Members of potential clusters involving Latinos may have been missed in our analyses due to under sampling. Alternatively, many infections among Latinos may have been acquired elsewhere and thus do not cluster in our local cohort. Latinos in NC are more likely to be foreign-born compared to other regions in the US[42], and some are seasonal migrants who may have partners residing outside the state. Latinos did cluster in pairs similarly to other groups, possibly representing partners seeking care together following infection acquired elsewhere. Future studies tracking migration in conjunction with phylogenetic analyses may help delineate the structure of transmission among this hard to reach group.

Among newly diagnosed patients, we found that the presence of TDR, in addition to younger age and acute infection, was significantly associated with transmission cluster membership. Notably, the association between TDR and clustering remained significant even after controlling for infection duration in the multivariable model. These associations with clustering may simply be markers for very high risk behavior and rapid ongoing transmission as no data to date suggests TDR mutations make the virus more transmissible[43]. Additionally, we found several clusters with nearly all individuals harboring TDR and very few ART-experienced members, suggesting either that experienced individuals were missed by incomplete sampling or further supporting the role of drug naïve individuals contributing to onward spread of TDR [44–46]. Importantly, these clusters may reflect sexual networks that are reservoirs of drug resistance beyond ART-experienced individuals.

Notably, the reconstruction of transmission clusters on the population level represents an estimate of the local epidemic. Through incomplete sampling potential cluster members will be unidentified, either because they are undiagnosed, disengaged from care, or never had a genotype. Furthermore, the parameters of our analysis were not intended to only identify linked transmission between partners as there may be unrecognized third parties involved in the transmission chain and the directionality of transmission cannot be discerned. Although our use of *pol* sequences acquired during routine care represents an obvious and unavoidable selection bias, we had sufficient sampling density to uncover clusters and demonstrate both expected and unanticipated trends. We used robust statistical support to define our clusters which could potentially underestimate the number of actual transmission clusters. Although our less strict genetic distance cut off helps identify clusters where transmission events are spread out over several years, this method may reveal many historic clusters. Additionally, we are unable to determine when onward transmission events took place – whether before or after diagnosis. While we observed clusters predominantly composed of acute/recent patients [47], a substantial proportion included one or more chronic patients with diagnoses prior to the acute infection. Importantly, this indicates a failure of secondary prevention and suggests that acute-acute transmission may not dominate spread in NC; a significant amount of transmission may occur during chronic infection [48].

Our study demonstrates a unique view into the structure of local transmission in NC through the integration of molecular, clinical, and demographic data. These complementary methods have the potential to provide important insight into relationships that cannot be uncovered through traditional epidemiological methods alone. These methods may help identify gaps in case finding and transmission trends among high risk groups including hard to reach populations, such as Latinos in the Southeast. Ultimately, the integration of widespread genotypic sampling with epidemiologic data, time-scaling, and sophisticated statistical methods, could lead to the development of novel models predicating incident cases and onward transmission – ideal targets for prevention campaigns.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank patients and staff of the UNC Center for AIDS Research HIV Clinical Cohort, led by Oksana Zakhavora; patients and staff of the Duke-UNC Acute HIV Consortium and the NC Screening and Tracing of Acute Transmission Program including Julie Nelson, JoAnn Kuruc, and William Miller; and Myron Cohen for his review of the manuscript.

This work was supported by the University of North Carolina at Chapel Hill Center for AIDS Research (P30 AI50410) and the KL2 Multidisciplinary Clinical Research Development Award (5KL2RR025746-04) from the University of North Carolina at Chapel Hill.

References

- Centers for Disease Control and Prevention. HIV Surveillance Report. 2009; vol. 21 Available from <http://www.cdc.gov/hiv/topics/surveillance/resources/reports/>. Published February 2010.
- North Carolina Department of Health and Human Services. [Accessed March 27, 2011] Epidemiologic Profile for HIV/STD Prevention & Care Planning. Available from <http://www.epi.state.nc.us/epi/hiv/>. Published December 2010.
- North Carolina Department of Health and Human Services. [Accessed March 27, 2011] 1999 STD Surveillance Report. Available from <http://www.epi.state.nc.us/epi/hiv/>.
- Branson BM, Handsfield HH, Lampe MA, Janssen RS, Taylor AW, Lyss SB, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. Morbidity and Mortality Weekly Report Recommendations and Reports. 2006 Sep 22; 55(RR-14):1–17. [PubMed: 16988643]
- Althoff K, Gange S, Klein M, Brooks J, Hogg R, Bosch R, et al. Late Presentation for Human Immunodeficiency Virus Care in the United States and Canada. Clinical Infectious Diseases. 2010; 50(11):1512–1520. [PubMed: 20415573]
- Dennis AM, Napravnik S, Seña AC, Eron JJ. Late Entry to HIV Care Among Latinos Compared With Non-Latinos in a Southeastern US Cohort. Clinical Infectious Diseases. 2011; 53(5):480–487. [PubMed: 21844031]
- Piot P, Aggleton P. The global epidemic. AIDS Care. 1998 Jun; 10(Suppl 2):S201–S208. [PubMed: 9743741]
- Barcherini S, Cantoni M, Grossi P, Verdecchia A. Reconstruction of human immunodeficiency virus (HIV) sub-epidemics in Italian regions. International journal of epidemiology. 1999 Feb; 28(1):122–129. [PubMed: 10195676]
- Cantoni M, Cozzi Lepri A, Grossi P, Pezzotti P, Rezza G, Verdecchia A. Use of AIDS surveillance data to describe subepidemic dynamics. International journal of epidemiology. 1995 Aug; 24(4): 804–812. [PubMed: 8550279]
- Hue S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS. 2004 Mar 26; 18(5):719–728. [PubMed: 15075506]

11. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS medicine*. 2008 Mar 18;5(3):e50. [PubMed: 18351795]
12. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infectious Diseases*. 2010; 10(1):262. [PubMed: 20822507]
13. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics*. 2009 Aug; 10(8):540–550.
14. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS pathogens*. 2009 Sep;5(9):e1000590. [PubMed: 19779560]
15. Pilcher CD, Fiscus SA, Nguyen TQ, Foust E, Wolf L, Williams D, et al. Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*. 2005 May 5; 352(18):1873–1883. [PubMed: 15872202]
16. Hurt CB, McCoy SI, Kuruc J, Nelson JA, Kerkau M, Fiscus S, et al. Transmitted antiretroviral drug resistance among acute and recent HIV infections in North Carolina from 1998 to 2007. *Antiviral Therapy*. 2009; 14(5):673–678. [PubMed: 19704170]
17. Pilcher CD, McPherson JT, Leone PA, Smurzynski M, Owen-O'Dowd J, Peace-Brewer AL, et al. Real-time, universal screening for acute HIV infection in a routine HIV counseling and testing population. *Journal of the American Medical Association*. 2002 Jul 10; 288(2):216–221. [PubMed: 12095386]
18. Rambaut, A. Se-AI v.2.0. Oxford, UK: University of Oxford; 1996. Available from: <http://tree.bio.ed.ac.uk/software/seal>
19. Pond SLK, Posada D, Stawiski E, Chappey C, Poon AFY, Hughes G, et al. An Evolutionary Model-Based Algorithm for Accurate Phylogenetic Breakpoint Mapping and Subtype Prediction in HIV-1. *PLoS Computational Biology*. 2009 Nov;5(11)
20. Johnson VA, Brun-Vezinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D, et al. Update of the drug resistance mutations in HIV-1: December 2009. *Topics in HIV Medicine*. 2009 Dec; 17(5): 138–145. [PubMed: 20068260]
21. Hasegawa M, Kishino H, Yano T-a. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*. 1985; 22(2):160–174. [PubMed: 3934395]
22. Swofford, D. Paup*. *Phylogenetic Analysis Using Parsimony (* and other methods)*. Sunderland, Massachusetts: Version 4 Sinauer Associates; 2003.
23. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001 Aug; 17(8):754–755. [PubMed: 11524383]
24. Rambaut A, Drummond A. Tracer v1.4. 2007 Available from: <http://beast.bio.ed.ac.uk/Tracer>.
25. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*. 2007; 7:214. [PubMed: 17996036]
26. Stamatakis A. RAXML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006 Nov 1; 22(21):2688–2690. [PubMed: 16928733]
27. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular biology and evolution*. 2009 Jul 1; 26(7):1641–1650. 2009. [PubMed: 19377059]
28. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*. 2007 Sep 1; 23(17):2334–2336. [PubMed: 17586542]
29. Napravnik S, Eron JJ Jr, McKaig RG, Heine AD, Menezes P, Quinlivan EB. Factors associated with fewer visits for HIV primary care at a tertiary care center in the Southeastern U.S. *AIDS Care*. 2006; 18:45–50.
30. Bennett D, Camacho R, Otelea D, Kuritzkes D, Fleury H, Kiuchi M, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One*. 2009; 4(3):e4724. [PubMed: 19266092]

31. Pao, Da; Fisher, Ma; Hue, Sbe; Dean, Ga; Murphy, Ge; Cane, PAde, et al. Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. *AIDS*. 2005; 19(1):85–90. [PubMed: 15627037]
32. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis*. 2010 May 15; 201(10):1488–1497. [PubMed: 20384495]
33. von Wyl V, Kouyos RD, Yerly S, Boni J, Shah C, Burgisser P, et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *Journal of Infectious Diseases*. 2011; 204(7):1095–1103. [PubMed: 21881125]
34. Aldous, J.; Poon, A.; Kosakovsky Pond, S.; Jain, S.; Qin, H.; Kahn, J., et al. Identifying HIV Transmission Networks in the CNICS: Implications for the Study of Transmission in Women and Minorities in a Large Clinical Cohort. 18th Conference on Retroviruses and Opportunistic Infections [abstract 1045]; 27 February - 2 March; Boston, Massachusetts. 2011.
35. Ragonnet-Cronin M, Ofner-Agostini M, Merks H, Pilon R, Rekart M, Archibald CP, et al. Longitudinal Phylogenetic Surveillance Identifies Distinct Patterns of Cluster Dynamics. *Journal of Acquired Immune Deficiency Syndromes*. 2010; 55(1):102–108. [PubMed: 20622676]
36. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(12):4425–4429. [PubMed: 15767575]
37. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. *American Journal of Public Health*. 2009 Jun; 99(6):1023–1031. [PubMed: 19372508]
38. Berry M, Raymond HF, McFarland W. Same race and older partner selection may explain higher HIV prevalence among black men who have sex with men. *AIDS*. 2007 Nov 12; 21(17):2349–2350. [PubMed: 18090287]
39. Bingham TA, Harawa NT, Johnson DF, Secura GM, MacKellar DA, Valleroy LA. The effect of partner characteristics on HIV infection among African American men who have sex with men in the Young Men's Survey, Los Angeles 1999-2000. *AIDS Education and Prevention*. 2003 Feb; 15(1 Suppl A):39–52. [PubMed: 12630598]
40. Hightow LB, Leone PA, Macdonald PD, McCoy SI, Sampson LA, Kaplan AH. Men who have sex with men and women: a unique risk group for HIV transmission on North Carolina College campuses. *Sexually Transmitted Diseases*. 2006; 33(10):585–593. [PubMed: 16641826]
41. Bond L, Wheeler DP, Millett GA, LaPollo AB, Carson LF, Liao A. Black Men Who Have Sex With Men and the Association of Down-Low Identity With HIV Risk Behavior. *Am J Public Health*. 2009; 99(S1):S92–S95. [PubMed: 19218177]
42. Pew Hispanic Center. [Accessed October 31, 2010] Statistical Profiles of the Hispanic and Foreign-Born Population in the United States. 2008. Available from <http://pewhispanic.org/datasets/>
43. Leigh Brown AJ, Frost SD, Mathews WC, Dawson K, Hellmann NS, Daar ES, et al. Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population. *Journal of Infectious Diseases*. 2003 Feb 15; 187(4):683–686. [PubMed: 12599087]
44. Callegaro A, Svicher V, Alteri C, Lo Presti A, Valenti D, Goglio A, et al. Epidemiological network analysis in HIV-1 B infected patients diagnosed in Italy between 2000 and 2008. *Infection, Genetics, and Evolution*. 2011 Apr; 11(3):624–632.
45. Hue S, Gifford RJ, Dunn D, Fernhill E, Pillay D. on Behalf of the UK Collaborative Group on HIV Drug Resistance. Demonstration of Sustained Drug-Resistant Human Immunodeficiency Virus Type 1 Lineages Circulating among Treatment-Naive Individuals. *Journal of Virology*. 2009; 83(6):2645–2654. [PubMed: 19158238]
46. Yerly S, Junier T, Gayet-Ageron A, Amari E, von Wyl V, Gunthard H, et al. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. *AIDS*. 2009; 23(11):1415–1423. [PubMed: 19487906]

47. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. *The Journal of Infectious Diseases*. 2007 Apr 1; 195(7):951–959. [PubMed: 17330784]
48. Rieder P, Joos B, von Wyl V, Kuster H, Grube C, Leemann C, et al. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. *AIDS*. 2010 May 15; 24(8):1177–1183. [PubMed: 20386427]

Figure 1.

Cluster size distribution and example sections of Bayesian phylogenetic tree showing two large clusters characterized by risk group. Nodes with posterior probability=1 are indicated. Numbers at tips represent date of sequencing. Note. MSM, men who have sex with men; HET, heterosexual; OTH, Other/Unknown risk

A. Distribution of cluster sizes (n = 3 sequences). Median cluster size 4 (IQR 3-5) members.

B. Cluster n=18, Predominant MSM (83%) and Black (78%).

C. Cluster n=13, Homogenous HET (100%) and Predominant Black (69%).

Table 1

Characteristics of study population with B-subtype sequences for all patients in phylogenetic analyses (n=1671) and the subset of patients newly diagnosed from 2000-2009 (n=889).

Characteristic	All Patients	Patients Diagnosed from 2000–2009
Total n (%)	1671	889
Age at diagnosis, median (IQR)	32 (25–40)	33 (25–43)
Year of Diagnosis, median (IQR)	2000 (1995–2004)	2004 (2002–2006)
Race/Ethnicity		
Black	983 (59)	521 (59)
White	505 (30)	233 (26)
Latino	115 (7)	88 (10)
Other ^a /Unknown	68 (4)	47 (5)
Sex		
Male	1177 (72)	630 (73)
Female	460 (28)	226 (26)
Transmission Risk		
MSM	656 (39)	376 (42)
Heterosexual	688 (41)	377 (42)
IDU	176 (11)	46 (5)
Other/Unknown	151 (9)	90 (10)
Duration of infection		
Chronic	1373 (82)	603 (68)
Recent	129 (8)	125 (14)
Acute	169 (10)	161 (18)
Sequence < 1 year of diagnosis	777 (47)	712 (80)
CD4 count ^{bc} , median cells/mm ³ (IQR)	---	251 (49–449)
Viral load ^c , median log copies/ml (IQR)	---	4.8 (4.1–5.3)
Antiretroviral History (at the time of sequence)		
Experienced	763 (46)	114 (13)
Naïve	908 (54)	775 (87)
Resistance ^d		
Any DRMs	101 (11)	92 (12)
NRTIs	41 (5)	34 (4)
NNRTIs	59 (6)	56 (7)
PIs	16 (2)	16 (2)
2 classes	11 (1)	10 (1)
3 classes	2 (<1)	2 (<1)
Residence in Primary Catchment Area ^e	1243 (74)	626 (71)

MSM, men who have sex with men; DRMs, drug resistance mutations; NRTIs, nucleoside reverse transcriptase inhibitors; NNRTIs, nonnucleoside reverse transcriptase inhibitors; PI, protease inhibitors

^aOther race/ethnicity: 1.7% Native American, <1% Asian, <1% Mixed

^b CD4 counts included for patients in the chronic cohort only (n=603)

^c Values recorded at date closest to diagnosis

^d Resistance mutations for naïve sequences only

^e Primary catchment area includes 16 out of 100 NC counties that account for approximately 75% of the UNC clinic population. Address information unavailable for 13 patients.

Table 2

Cluster composition by selected characteristics among transmission clusters identified by phylogenetic analysis of 1671 sequences using two cluster definitions

Characteristic	Cluster composition			
	Bayesian Posterior Probability = 1 (n=67 clusters)		Bayesian Posterior Probability=1 and mean genetic distance ^a 1.5% (n=33 clusters)	
	Predominant n (%)	Homogenous n (%)	Predominant n (%)	Homogenous n (%)
Race/Ethnicity				
Black	45 (67)	19 (28)	23 (70)	9 (27)
White	15 (22)	7 (10)	6 (18)	4 (12)
Latino	0 (0)	---	0 (0)	---
Other	0 (0)	---	0 (0)	---
Mixed	7(11)	---	4 (12)	---
Transmission Risk				
MSM	28 (42)	12 (18)	16 (48)	8 (24)
Heterosexual	23 (34)	10 (15)	12 (36)	5 (15)
IDU/Other/Unknown	0 (0)	---	0 (0)	---
Mixed	16 (24)	---	5 (15)	---
Age at Diagnosis, years				
<30	26 (39)	9 (13)	14 (42)	10 (30)
30–44	13 (19)	3 (4)	6 (18)	2 (6)
45	4 (6)	1 (1)	4 (12)	0 (0)
Mixed	24 (36)	---	9 (27)	---
Duration of Infection				
Acute or Recent	14 (21)	5 (7)	9 (27)	4 (12)
Chronic	49 (73)	25 (37)	22 (67)	12 (36)
Mixed	4 (6)	---	2 (6)	---
TDR				
Yes	5 (7)	5 (7)	4 (12)	4 (12)
No	62 (93)	50 (75)	29 (88)	27 (82)

Predominant, 2/3 of members share same characteristic; Homogeneous, 100% of members share same characteristic; Mixed, no characteristic in cluster predominant. MSM, men who have sex with men; IDU, injection drug use; TDR, transmitted drug resistance

^aMean intra-cluster pairwise genetic distance (nucleotide substitutions per site)

Table 3

Characteristics of clusters with majority (>50%) of members sharing Transmitted Drug Resistance (TDR) among the 67 transmission clusters identified from phylogenetic inference of 1671 sequences.

Cluster ID No.	Cluster Size	TDR	Acute	MSM	Mutation(s) ^a
7	3	3	1	0	K103S, G190A
11	4	4	1	4	L90M
13	3	3	0	2	G190A
23	5	3	0	5	D67N, K219Q
65	3	3	2	3	K103N
70	3	3	0	1	K103N

TDR, transmitted drug resistance; MSM, men you have sex with men

^a All members of each cluster shared same mutation(s)

Comparison of factors associated with membership in phylogenetic transmission between pairs (n=2 sequences) and clusters (n=3) for patients diagnosed 2000–2009 (n=889). Logistic regression models show associations for membership in transmission clusters compared to those not in clusters or pairs.

Table 4

Characteristic	Not in Cluster	In Pair n=2	In Cluster n=3	Membership in Cluster n=3		
				Bivariable	Multivariable	
Total n (%)	504(57)	154 (17)	231 (26)	OR (95% CI)	P	OR (95% CI) P
Age 30 years	187 (52)	61 (17)	114 (31) ^{ab}	1.58 (1.16–2.15)	0.003	1.61 (1.13–2.30) 0.009
Year of Diagnosis >2004	199 (54)	70 (19)	98 (27)	1.07 (0.79–1.44)	0.68	NA
Race/Ethnicity						
Black	283 (54)	80 (15)	158 (30) ^{ab}	Ref		Ref
White	138 (59)	38 (16)	57 (24)	0.74 (0.52–1.06)	0.10	0.74(0.50–1.09) 0.13
Latino	55 (62)	25 (28)	8 (9)	0.23 (0.11–0.49)	<0.001	0.20 (0.09–0.46) <0.001
Other/Unknown	28 (60)	11 (23)	8 (17)	0.47 (0.21–1.03)	0.06	0.54 (0.14–2.07) 0.37
Sex and Transmission Risk						
Women	136 (60)	40 (18)	50 (22)	Ref		NA
Men – non-MSM	135 (55)	49 (20)	62 (25)	1.19 (0.77–1.82)	0.43	
MSM	209 (54)	61 (16)	114 (30)	1.48 (1.01–2.18)	0.04	
Duration of Infection						
Chronic	342 (57)	112 (19)	149 (25) ^{ab}	Ref		Ref
Recent	80 (64)	20 (16)	25 (20)	0.76 (0.47–1.23)	0.26	0.85 (0.46–1.57) 0.60
Acute	82 (51)	22 (14)	57 (35)	1.67 (1.15–2.42)	0.007	1.74 (1.10–2.73) 0.017
ART-naïve	431 (56)	137 (18)	207 (27)	1.37 (0.85–2.20)	0.20	NA
Genotype <1 year of diagnosis	397 (56)	127 (18)	188 (26)	1.12 (0.76–1.64)	0.57	NA
CD4 cell count >200 cells/mm ³ ^{cd}	178 (53)	70 (21)	89 (26)	1.23 (0.85–1.79)	0.28	NA
HIV RNA >5.0 log copies/ml ^d	192 (55)	60 (17)	94 (27)	1.16 (0.85–1.58)	0.34	NA
TDR	43 (47)	16 (17)	33 (36) ^b	1.64 (1.03–2.59)	0.04	1.80 (1.09–2.99) 0.02
Residence in primary catchment area ^e	337(54)	116 (19)	173 (28) ^{abf}	1.38 (0.98–1.96)	0.07	2.02 (1.29–3.15) 0.002

OR, odds ratio; NA not applicable; MSM, men who have sex with men; ART, antiretroviral therapy; TDR, transmitted drug resistance

^a $P < 0.05$ based on Pearson's χ^2 for group trend

- ^b $P < 0.05$ based on Pearson's χ^2 for membership in cluster n = 3 versus not in a cluster
- ^c CD4 cell counts only available for patients with chronic duration of infection (n=603)
- ^d Values recorded at date closest to diagnosis
- ^e Primary catchment area includes 16 out of 100 NC counties that account for approximately 75% of the UNC clinic population
- ^f $P < 0.05$ based on Pearson's χ^2 for membership in pair n=2 versus not in a cluster