HHS Public Access

Author manuscript

J Acquir Immune Defic Syndr. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

J Acquir Immune Defic Syndr. 2015 December 1; 70(4): 444–451. doi:10.1097/QAI.0000000000000809.

Using Molecular HIV Surveillance Data to Understand Transmission between Subpopulations in the United States

Alexandra M. Oster[1], **Joel O. Wertheim**[2,3], **Angela L. Hernandez**[1], **M. Cheryl Bañez Ocfemia**[1], **Neeraja Saduvala**[2], and **H. Irene Hall**[1]

[1]Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, United States

[2]ICF International, Atlanta, United States

[3]Department of Medicine, University of California, San Diego, United States

Abstract

Background—Studying HIV transmission networks provides insight into the spread of HIV and opportunities for intervention. We identified transmission dynamics among risk groups and racial/ethnic groups in the United States.

Methods—For HIV-1 *pol* sequences reported to the U.S. National HIV Surveillance System during 2001–2012, we calculated pairwise genetic distance, identified linked pairs of sequences (those with distance ≤1.5%), and examined transmission category and race/ethnicity of these potential transmission partners.

Results—Of 40,950 sequences, 12,910 (32%) linked to ≥1 other sequence. Of men who have sex with men (MSM) who linked to ≥1 sequence, 88% were linked to other MSM and only 4% were linked to heterosexual women. Of heterosexual women for whom we identified potential transmission partners, 29% linked to MSM, 21% to heterosexual men, and 12% to persons who inject drugs. Older and black MSM were more likely to be linked to heterosexual women. Assortative mixing was present for all racial/ethnic groups; 81% of blacks/African Americans linked to other blacks.

Conclusions—This analysis is the first use of U.S. surveillance data to infer an HIV transmission network. Our data suggest that HIV infections among heterosexual women predominantly originate from MSM, followed by heterosexual men. Although few MSM were linked to women, suggesting that a minority of MSM are involved in transmission with heterosexual women, these transmissions represent a substantial proportion of HIV acquisitions by heterosexual women. Interventions that reduce transmissions involving MSM are likely to also reduce HIV acquisition among other risk groups.

Correspondence: Alexandra M. Oster, MD, 1600 Clifton Rd NE, MS E-47, Atlanta, GA 30329, Phone 404-639-6141, FAX 404-639-2980, ; Email: AOster@cdc.gov

The authors do not have commercial or other associations that might pose a conflict of interest.

Portions of these data were presented at the Conference on Retroviruses and Opportunistic Infections, Boston, Massachusetts, United States, March 2014 (Abstract 211 and 213) and at the Texas HIV/STD Conference, Austin, Texas, United States, August 2014.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Keywords

HIV; Molecular epidemiology; surveillance; risk factors; men who have sex with men; intravenous drug users; women

Introduction

The study of HIV transmission networks provides insight into the spread of HIV, and thus into opportunities for intervention [1–5]. By using HIV nucleotide sequence data, we can construct transmission networks that link persons who are infected with genetically similar HIV variants [5–7]. Linked persons are likely to have a direct or indirect epidemiologic connection and represent potential transmission partners. The extent to which transmission occurs between people who are similar or different with respect to race/ethnicity and risk characteristics is unclear, and network data can aid in understanding transmission dynamics (reviewed in [1]). For example, understanding the links between different populations can be particularly useful in elucidating transmission routes for heterosexual women, who are often unaware of the HIV risk behaviors of their partners [8, 9]. Moreover, persons whose HIV sequences are similar to (i.e., are linked to) many other sequences are of particular interest for prevention, as these persons may be part of a dense section of a transmission network, and, therefore, most likely to transmit HIV. These data can help HIV researchers and public health professionals move beyond understanding which groups are affected by HIV to understanding how HIV spreads within and between populations.

Previous studies have investigated the contribution of demographic and risk characteristics of transmission networks [6, 7, 10–17]. The few studies that have explored transmission networks in the United States focused on a single geographic location; one exception, a 5-site study of U.S. transmission networks, did not explore transmission between risk groups [14]. Since 2001, as part of the U.S. National HIV Surveillance System, the Centers for Disease Control and Prevention (CDC) has funded projects in 21 states and the District of Columbia to collect HIV sequence data from persons living with HIV infection. This data set, which is 1–2 orders of magnitude larger than previous investigations into HIV in the United States, incorporates a rich set of variables collected by the National HIV Surveillance System, including demographic and clinical characteristics and risk information.

The aim of this analysis was to use U.S. surveillance data to construct an HIV transmission network, explore HIV transmission between different demographic and risk groups, and describe persons with many potential transmission links. These findings will supplement current understanding of how HIV spreads between risk and racial/ethnic populations in the United States and can be used to tailor prevention interventions.

Methods

Sources of Sequence Data and Inclusion Criteria

We analyzed HIV-1 *pol* sequences reported through the National HIV Surveillance System during 2001–2012. These sequences were collected in various jurisdictions (Figure 1) through three drug resistance surveillance projects: Antiretroviral Drug Resistance Testing

(ARVDRT, 2001–2007); the Dried Blood Spots project (DBS, 2004–2009); and Variant, Atypical, and Resistant HIV Surveillance (VARHS, 2004–2012) [18].[1] Remnant diagnostic specimens (ARVDRT and VARHS) and dried blood spots created from remnant specimens or new fingersticks (DBS) were shipped to a contract laboratory for CDC-funded antiretroviral drug resistance testing. During the last 2 years of VARHS, sequence data were instead collected from routinely performed genotypic resistance tests conducted at commercial, private, and public laboratories that were reported to the National HIV Surveillance System.[2] We did not include sequences from other databases (e.g., GenBank), because we are unable to ensure that these sequences do not originate from persons already included in our surveillance system.

Our analysis included one sequence per person aged ≥ 13 years. When more than one sequence was available (10% of persons), we selected the longest sequence. We excluded sequences that were less than 500 nucleotides in length because our transmission network approach has not been validated for very short sequences.

Sequence Analysis and Construction of Network

To construct the transmission network, we followed the protocol outlined by Wertheim *et al* [6] using HIV-TRACE (HIV TRAnsmission Cluster Engine; available at hivtrace.org). We aligned each of the 41,304 HIV sequences in a pairwise fashion to the HIV reference sequence HXB2 (positions 2253 to 3869), then computed pairwise distances for all sequences (>850 million comparisons). Tamura-Nei 93 (TN93) genetic distance ≤1.5% was considered evidence of possible linkage. The choice of nucleotide substitution model has been shown not to affect genetic distance estimation for short genetic distances [19]. We excluded 10 sequences that were closely related to HXB2, a common laboratory control, as we deemed them possible contaminants. We described characteristics of the network, including the number of sequences (nodes), links (edges) and clusters (groups of linked sequences). All resulting data sets were analyzed in SAS version 9.3 (SAS Institute, Cary, NC).

Measures

Data reported to the National HIV Surveillance System do not include personally identifiable information such as name, address, or phone number. Data on age at HIV diagnosis, sex, race/ethnicity, transmission category, area of residence at diagnosis, and country of birth are reported using standard CDC classification schemes. Race/ethnicity was categorized as white; black or African American (hereafter referred to as ‘black’); Hispanic

[1]Jurisdictions participating in ARVDRT included Colorado, Illinois, Maryland, and Seattle/King County (Washington). DBS included Chicago, Los Angeles County, Minnesota, and New York (excluding New York City). The first period of VARHS (2004–2007) included Chicago, the District of Columbia, Florida, Indiana, Louisiana, Massachusetts, Michigan, Mississippi, New Jersey, New York (including New York City), North Carolina, Pennsylvania (excluding Philadelphia), Puerto Rico, South Carolina, Texas (excluding Houston), Virginia, and Washington. The second period of VARHS (2008–2012) included Chicago, Colorado, Connecticut, Florida, Los Angeles County, Louisiana, Michigan, New York City, South Carolina, Texas (excluding Houston), and Washington.

[2]States and U.S. territories collect and report demographic, behavioral, clinical, and laboratory data on persons diagnosed with HIV infection in accordance with public health disease reporting requirements in their jurisdictions and voluntarily share those data with CDC as part of the NHSS for national level analyses. Therefore, each state has primary authority for determining whether their laws and regulations permit the submission to GenBank or other open databases, and CDC is not able to submit sequences to these databases.

or Latino; Asian, Native Hawaiian, and Other Pacific Islander; or other (including American Indian/Alaska Native and multiple races). Transmission categories included male-to-male sexual contact (men who have sex with men, MSM), MSM who inject drugs, persons who inject drugs (PWID, stratified into male PWID and female PWID), heterosexual contact (stratified into heterosexual men and heterosexual women), and other (includes hemophilia, blood transfusion, perinatal exposure, and risk factor not reported or not identified). Missing data for transmission category were assigned using multiple imputation procedures employed for standard National HIV Surveillance System reports; these procedures have been described in detail elsewhere [20]. Multiple imputation was required for 15% of persons included in this analysis.

Analysis of Mixing by Transmission Category and Race/Ethnicity

To understand the extent to which HIV transmission occurs between persons of the same or different race/ethnicity or transmission category, we analyzed the characteristics of potential transmission partners. To account for bias arising from persons with multiple links, each potential transmission partner for a person is weighted based on the number of links associated with the person (k). The weight is simply 1/k for each of the k links. To further explore links between MSM and women, we examined (i) the proportion of MSM that linked to heterosexual women and (ii) the proportion of heterosexual women that linked to MSM. Both analyses were stratified by age, race/ethnicity, population of area of residence, and U.S. census region; chi-square tests were used to assess differences between subgroups.

Analysis of Persons with Multiple Potential Transmission Links

Next, we compared three groups: persons who did not link, persons who linked to 1–3 other persons, and persons who linked to ≥4 other persons (i.e., those in the highest quartile of those persons with any links). We used multivariable logistic regression to identify characteristics associated with having ≥4 links, compared with having 0–3 links. The model included age at HIV diagnosis, race/ethnicity, transmission category, population of area of residence, and country of birth.

Sensitivity Analyses

To assess the sensitivity of our findings to assumptions such as distance threshold and the inclusion of codon positions associated with drug resistance mutations, we repeated all of the above analyses with the following changes in parameters from our baseline analysis (which had a TN93 distance threshold of 1.5% and retained all codons): (i) distance threshold of 1.0%, (ii) distance threshold of 2.0%, and (iii) removal of 48 codons associated with drug resistance mutations [18].

Results

We had 47,821 HIV-1 nucleotide sequences belonging to 41,549 persons. In all, 599 sequences were excluded (245 belonged to persons aged <13 years at diagnosis, 344 were shorter than 500 nucleotides, and 10 were possible contaminants). In all, 40,950 sequences were included in this analysis. Of the persons with sequences included in this analysis, 78% were male, 62% were aged 20–39 years, 47% were black, 61% were MSM, 84% lived in

areas with population ≥500,000, and 64% were known to be born in the United States (Table 1).

In all, 12,910 (32%) sequences linked to at least one other sequence and were included in the network, which contained a total of 29,493 links (edges). The number of links per person ranged from 1 to 83 (median: 1, interquartile range: 1–4). There were 3,584 distinct clusters, with the number of sequences per cluster ranging from 2–85 (median: 4, interquartile range: 2–9). The inferred proportion and distribution of linked sequences conforms to expectations under a preferential attachment, scale-free network [5, 6].

Mixing by Race/Ethnicity

For each racial/ethnic group, the percentage of persons linked to other persons of the same race/ethnicity was higher than the percentage of persons included in this analysis of that race/ethnicity. Of whites in the transmission network, 62% were linked to other whites, 20% to Hispanics/Latinos, and 12% to blacks (Figure 2a). Of blacks, 81% were linked to other blacks, and only 9% to whites and 8% to Hispanics/Latinos. Of Hispanics/Latinos, 52% were linked to other Hispanics/Latinos, 28% to whites, and 15% to blacks. Asians, Native Hawaiians, and other Pacific Islanders were more commonly linked to whites (40%) and to Hispanics/Latinos (29%) than to other Asians, Native Hawaiians, and other Pacific Islanders (17%).

Mixing by Transmission Category

Of MSM, 92% were linked to other MSM or MSM who inject drugs (Figure 2b). MSM who inject drugs were more commonly linked to MSM (76%) than to other MSM who inject drugs (12%) or male PWID (2%) or female PWID (1%). Male PWID were commonly linked to MSM (49%), female heterosexuals (25%), and other male PWID (8%) or female PWID (9%). Male heterosexuals were most commonly linked to female heterosexuals (49%) and MSM (34%). Approximately one-third of both female PWID and female heterosexuals were linked to other female heterosexuals. Female PWID were commonly linked to MSM (27%), male PWID (10%) and female PWID (14%), and female heterosexuals were commonly linked to MSM (29%) and male heterosexuals (21%). Few female heterosexuals were linked to male PWID (6%) or female PWID (6%).

We further examined links between MSM and heterosexual women. Overall, 4% of MSM were linked to a heterosexual woman (Table 2). However, this percentage varied significantly by age; MSM aged 50 years or older had the highest percentage (8%) linked to a heterosexual woman. We also identified significant differences in the percentage of MSM who were linked to heterosexual women by race/ethnicity: the percentage linked to heterosexual women was lowest for whites (2%) and highest for blacks (7%). Further, there were significant geographic differences; a lower percentage of MSM in metropolitan statistical areas (4%) than in smaller metropolitan areas (6%) or nonmetropolitan areas (7%) were linked to heterosexual women. Finally, a higher percentage of MSM in the Midwest (6%) and South (6%) than the Northeast (2%) and West (2%) were linked to heterosexual women.

Although only 4% of MSM were linked to a heterosexual women, these links represented 29% of links among heterosexual women. This percentage varied significantly by race/ethnicity, and was lower among black (25%) than among white (39%) and Hispanic/Latino (37%) women (Table 2). The percentage of heterosexual women who were linked to MSM also varied significantly by region; a higher percentage of women in the West (52%) were linked to MSM than among those in the Midwest (36%), Northeast (24%), or South (24%). The percentage linking to MSM did not vary by age or population of area of residence.

Persons with Multiple Potential Transmission Links

Overall, 10% of persons included in our analysis had ≥4 links in the network (Table 1). The percentage of persons with ≥4 links was higher for persons aged 13–19 years (21%) and 20–29 years (15%), whites (13%), MSM (14%), persons living in metropolitan statistical areas (10%), and persons born in the United States or with unknown country of birth (both 10%). All of these associations were statistically significant in multivariable analysis. Persons with ≥4 links represented only 31% of persons with any links but were involved in 26,378 (80%) of the 29,493 links.

Sensitivity Analyses

Sensitivity analyses indicate that these results are generally robust to assumptions regarding genetic distance and inclusion of codons associated with drug resistance (see Table, Supplemental Digital Content 1, which shows key results under various assumptions). As expected, the absolute number of clusters and links varied depending on the TN93 threshold selected and whether sites associated with drug resistance mutations were excluded; however, none of our major findings were sensitive to these changes in threshold.

Discussion

This analysis represents the first use of U.S. National HIV Surveillance System data to infer an HIV transmission network. Our analysis of mixing by transmission category yields important new insights into HIV transmission among various populations. Not surprisingly, MSM are disproportionately connected in the network, and the majority of MSM were linked to other MSM [13]. MSM who inject drugs were more commonly linked to MSM who did not inject drugs than to other MSM who inject drugs or to PWID, suggesting that transmission among MSM who inject drugs may be more commonly sexual than parenteral. Interestingly, other risk groups, including PWID and both male and female heterosexuals, were also commonly connected to MSM. As a result, interventions that reduce transmissions involving MSM are likely to reduce HIV acquisition among other risk groups as well.

Our results suggest that HIV infection among heterosexual women likely originates from two predominant sources. First, a large proportion of this group’s links were to MSM (29% of all links, or approximately half of non-female links). Although few MSM were linked to women, suggesting that a minority of MSM are involved in transmission with heterosexual women, these transmissions represent a substantial proportion of HIV acquisitions among heterosexual women. However, we also found that many heterosexual women were linked to heterosexual men (21% of all links, or approximately one third of non-female links).

Although we are not able to establish directionality of transmission from these data, evidence indicates that vaginal and anal transmission more commonly occur from males to females, and we can infer that the majority of these links likely represent transmissions from males to females. Interestingly, although a larger percentage of black MSM than MSM of other races and ethnicities were linked to heterosexual women, fewer black heterosexual women than other heterosexual women were linked to MSM. This finding provides particular support for ongoing heterosexual transmission among blacks and indicates that black women are at risk for HIV infection through multiple routes.

Certain groups of MSM were more likely to link to women. Consistent with a prior analysis of HIV sequence data from Mississippi, older MSM were more likely to link to heterosexual women [13]. This finding may be due to a generational change in attitudes toward homosexuality, as has been found in population-based data [21], and resultant decreases in bisexual behavior among younger generations of MSM for whom homosexuality is more widely accepted. Moreover, MSM in smaller metropolitan and nonmetropolitan areas and those in the South and Midwest were more likely to link to heterosexual women, which may suggest that they are more likely to engage in bisexual behavior than MSM in large metropolitan areas and MSM in the Northeast and West, respectively; a previous study found that MSM in smaller metropolitan and nonmetropolitan areas were far less likely to identify as gay [22].

We found a large number of links between women and other women. Although female-to-female HIV transmission has been documented [23], it is likely that the vast majority of the links between women represent indirect connections, due to missing sequences from one or more intermediate partners. Additional work is needed to examine whether we are able to determine the likely risk characteristics of these missing links. Expansion of our molecular HIV surveillance efforts, from 11 jurisdictions (representing 43% of new HIV diagnoses in the United States) in 2008–2012 to 27 jurisdictions (representing 72% of new HIV diagnoses) in 2013–2017 and continued improvements in completeness of molecular surveillance data will improve our ability to characterize the transmission network.

Assortative mixing in HIV transmission was present for all racial/ethnic groups. Notably, the vast majority (81%) of blacks were linked to other blacks. This type of assortative mixing, which has also been found in molecular analyses in Mississippi and in a 5-site study [13, 14], suggests that the majority of HIV transmissions involving blacks occur with other blacks and provides additional support for network level factors as an important cause of disparities in HIV infection among blacks in the United States [24]. Furthermore, prevention approaches that successfully reach HIV-infected blacks have the potential to reduce HIV incidence among HIV-negative blacks, whereas prevention delivered to other racial/ethnic groups is unlikely to have a substantial impact on incidence among blacks.

We also found evidence of assortative mixing among other racial/ethnic groups, as the percentage of potential transmission partners that were of the same race/ethnicity was at least twice what would be expected (based on the race/ethnicity of those in the network) for whites and Hispanics/Latinos and more than 8 times what would be expected for Asians, Native Hawaiians and other Pacific Islanders. Nevertheless, the extent of racial/ethnic

mixing suggests that HIV transmission commonly occurs between these populations and that efforts to disrupt acquisition among any one of these groups must be delivered broadly.

We found that MSM and young persons (aged 13–29 years) more commonly had multiple potential transmission partners, suggesting that they are disproportionately involved in clusters associated with HIV transmission. These findings corroborate other analyses demonstrating high HIV incidence among these groups [25] and suggest that strategies that prioritize MSM and young persons are likely to be most effective in preventing spread of HIV. A prior study has shown that using network characteristics to target prevention efforts can be very effective at reducing transmission [7].

Limitations

Our analysis may be affected by selection/sampling bias due to a limited number of areas being funded for collection of HIV-1 sequences and differences in completeness of the data by site and by other characteristics. As we expand the number of participating areas and the completeness of data, we expect this potential bias will diminish. Further, transmission category is based on self-reported data, and failure to report certain HIV risk behaviors could affect inferences about routes of transmission. When missing, transmission category was imputed, which could affect our estimates; however, imputation was only required for 15% of persons.

We have identified potential transmission partners, but we cannot definitively draw a link between any two persons or assess directionality of transmission. However, although the links we identify may not be direct, we can still draw conclusions from a large number of potentially indirect links. Moreover, because our sequence results are from bulk sequencing, in which multiple strains are sequenced simultaneously, we may have missed some potential connections.

Finally, these are national results, and findings might be different when examined at the regional or jurisdictional level, or when further stratified by demographic or risk characteristics. In the near future, we plan to conduct more granular analyses to understand how these data can be useful at the local level.

Conclusion

Molecular surveillance and epidemiology offer tools to understand HIV transmission dynamics. These analyses provide a more nuanced picture of how HIV spreads between populations in the United States, and we anticipate being able to make even stronger inferences as the reach and coverage of molecular HIV surveillance continues to improve. In the meantime, these data can be used to target and tailor prevention interventions with the ultimate goal of reducing HIV incidence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the local and state health department staff instrumental in collecting HIV sequence data; other present and past members of CDC Incidence and Molecular Epidemiology Team, including Qian An, David Kim, Magan Pearson, Richard Kline, Danuta Pieniazek, Joseph Prejean, and Rebecca Ziebell; and William Switzer.

Sources of Support: The Centers for Disease Control and Prevention

References

1. Dennis AM, Herbeck JT, Brown AL, 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
2. Hue S, Brown AE, Ragonnet-Cronin M, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. Aids. 2014
3. Fitch WM. Networks and viral evolution. Journal of molecular evolution. 1997; 44(Suppl 1):S65–S75. [PubMed: 9071014]
4. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS pathogens. 2009; 5:e1000590. [PubMed: 19779560]
5. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. J Infect Dis. 2011; 204:1463–1469. [PubMed: 21921202]
6. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. J Infect Dis. 2014; 209:304–313. [PubMed: 24151309]
7. Little SJ, Kosakovsky Pond SL, Anderson CM, et al. Using HIV Networks to Inform Real Time Prevention Interventions. PLoS One. 2014; 9:e98443. [PubMed: 24901437]
8. Fleming PL, Lansky A, Lee LM, Nakashima AK. The epidemiology of HIV/AIDS in women in the southern United States. Sex Transm Dis. 2006; 33:S32–S38. [PubMed: 16794553]
9. Haverkos HW, Chung RC, Norville Perez LC. Is there an epidemic of HIV/AIDS among heterosexuals in the USA? Postgraduate medical journal. 2003; 79:444–448. [PubMed: 12954955]
10. Holmes EC, Zhang LQ, Robertson P, et al. The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. J Infect Dis. 1995; 171:45–53. [PubMed: 7798682]
11. Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis. 2010; 201:1488–1497. [PubMed: 20384495]
12. Smith DM, May SJ, Tweeten S, et al. A public health model for the molecular surveillance of HIV transmission in San Diego, California. Aids. 2009; 23:225–232. [PubMed: 19098493]
13. Oster AM, Pieniazek D, Zhang X, et al. Demographic but not geographic insularity in HIV transmission among young black MSM. Aids. 2011; 25:2157–2165. [PubMed: 21866038]
14. Aldous JL, Pond SK, Poon A, et al. Characterizing HIV transmission networks across the United States. Clin Infect Dis. 2012; 55:1135–1143. Epub 2012 Jul 10. [PubMed: 22784872]
15. Dennis AM, Hue S, Hurt CB, et al. Phylogenetic insights into regional HIV transmission. Aids. 2012; 26:1813–1822. [PubMed: 22739398]
16. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med. 2008; 5:e50. [PubMed: 18351795]
17. Frentz D, Wensing AM, Albert J, et al. Limited cross-border infections in patients newly diagnosed with HIV in Europe. Retrovirology. 2013; 10:36. [PubMed: 23551870]
18. Wheeler WH, Ziebell RA, Zabina H, et al. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. Aids. 2010; 24:1203–1212. [PubMed: 20395786]
19. Wertheim JO, Kosakovsky Pond SL. Purifying selection can obscure the ancient age of viral lineages. Molecular biology and evolution. 2011; 28:3355–3365. [PubMed: 21705379]

20. Harrison KM, Kajese T, Hall HI, Song R. Risk factor redistribution of the national HIV/AIDS surveillance data: an alternative approach. Public Health Rep. 2008; 123:618–627. [PubMed: 18828417]
21. Glick SN, Golden MR. Persistence of racial differences in attitudes toward homosexuality in the United States. J Acquir Immune Defic Syndr. 2010; 55:516–523. [PubMed: 20838226]
22. Oster AM, Sternberg M, Nebenzahl S, et al. Prevalence of HIV, sexually transmitted infections, and viral hepatitis by Urbanicity, among men who have sex with men, injection drug users, and heterosexuals in the United States. Sex Transm Dis. 2014; 41:272–279. [PubMed: 24622641]
23. Chan SK, Thornton LR, Chronister KJ, et al. Likely female-to-female sexual transmission of HIV--Texas, 2012. MMWR Morb Mortal Wkly Rep. 2014; 63:209–212. [PubMed: 24622284]
24. Sullivan PS, Peterson J, Rosenberg ES, et al. Understanding racial HIV/STI disparities in black and white men who have sex with men: a multilevel approach. PLoS One. 2014; 9:e90514. [PubMed: 24608176]
25. Centers for Disease Control and Prevention. Estimated HIV Incidence in the United States, 2007–2010. HIV Surveillance Supplemental Report. 2012; 17

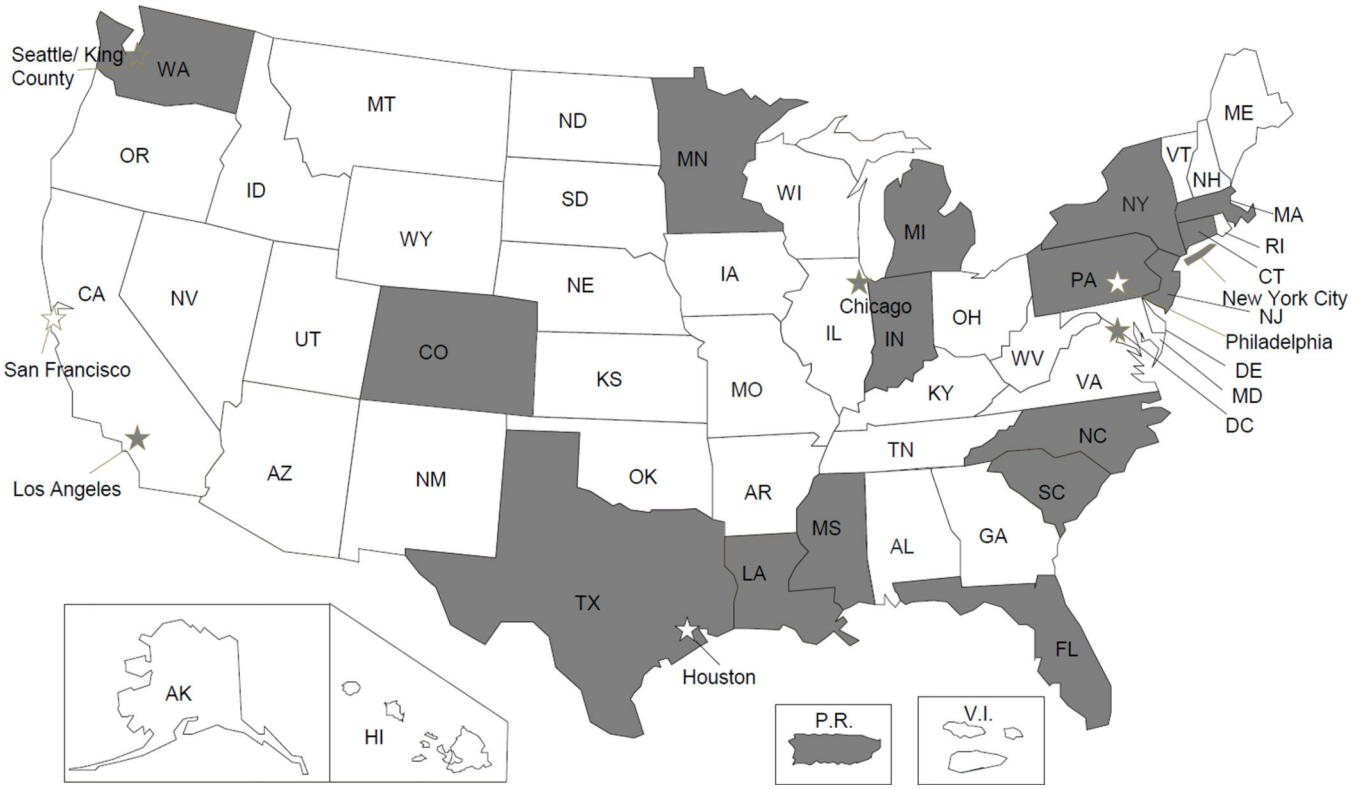

Figure 1.

Jurisdictions participating in collection of HIV-1 nucleotide sequence data, 2001–2012. Jurisdictions shaded in gray have participated in collection of nucleotide sequence data.

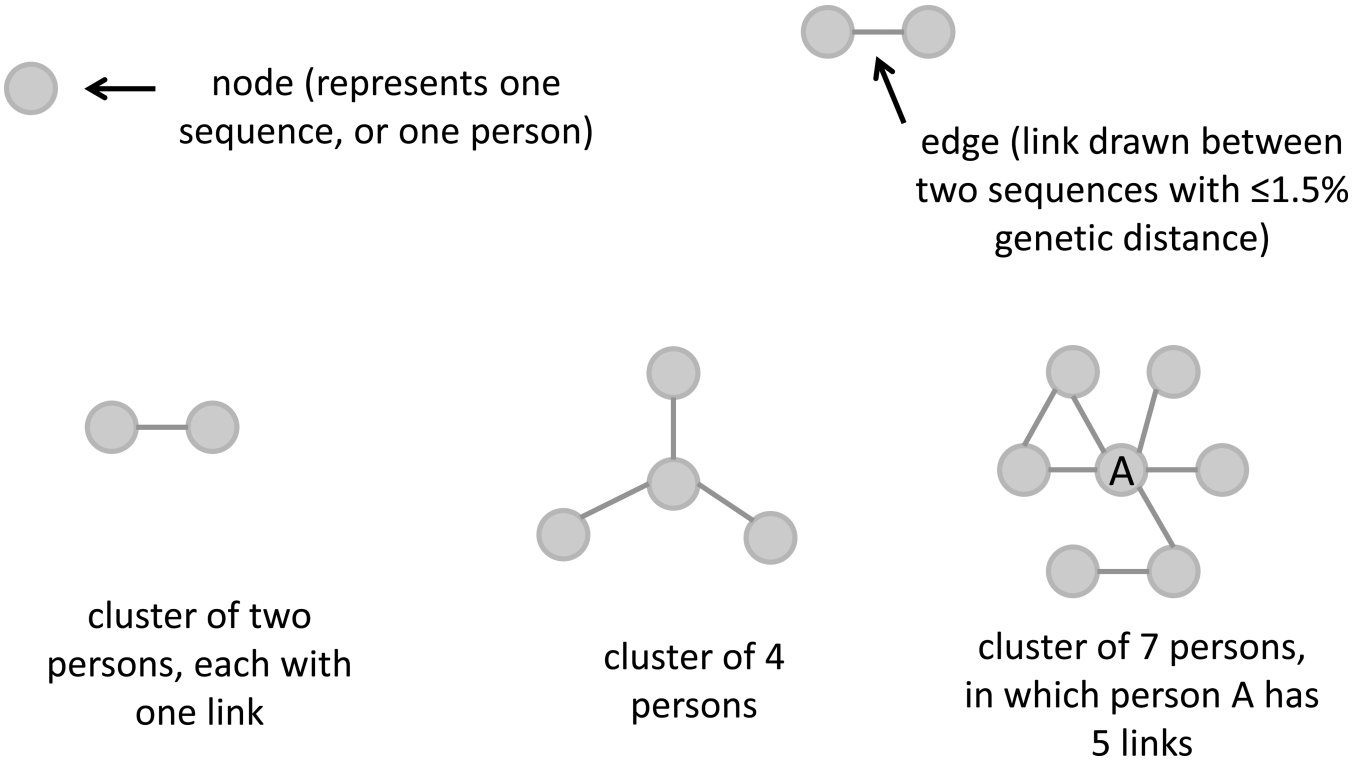

Figure 2.
Depiction of various terms used to describe the transmission network.

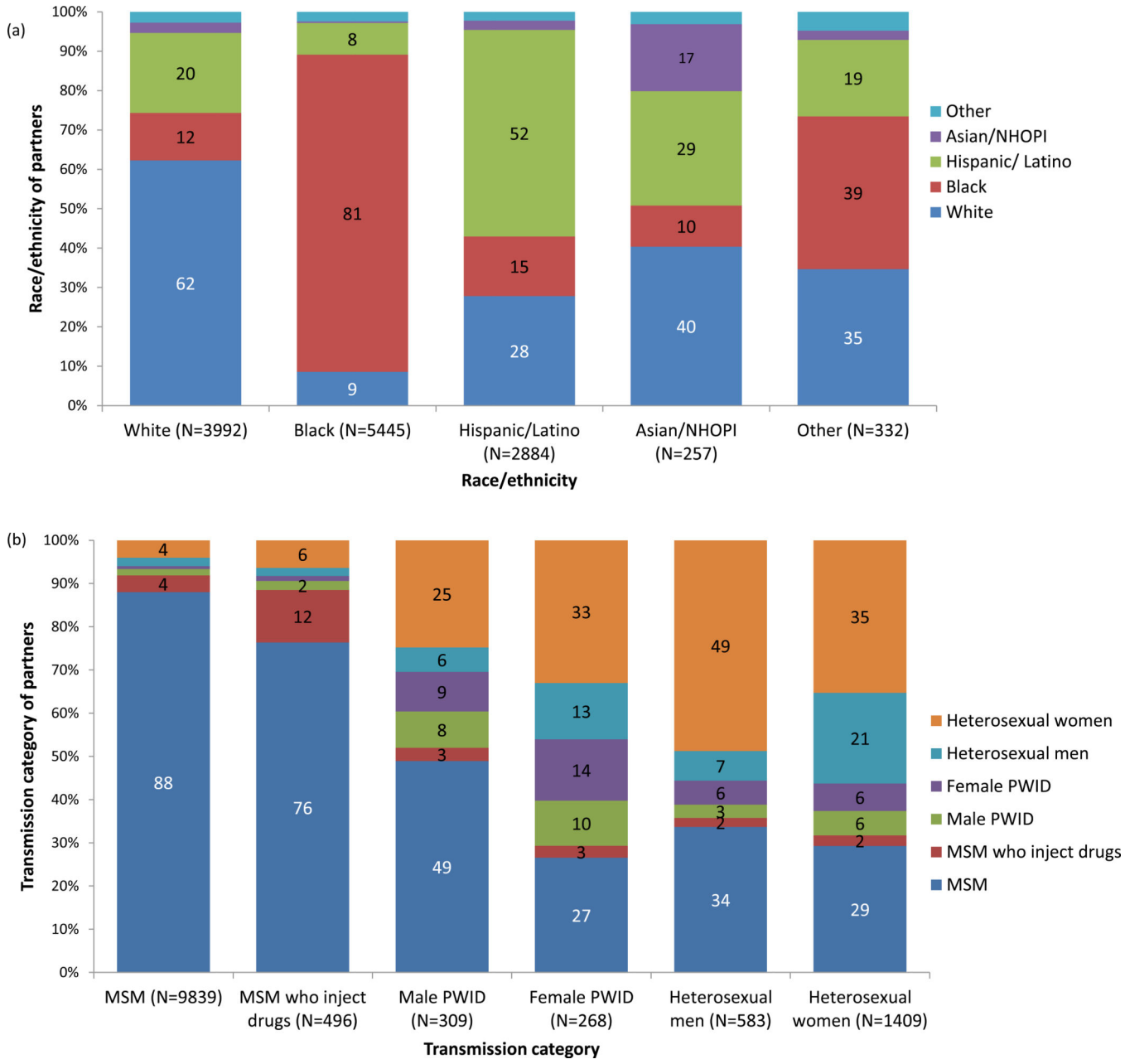

Figure 3.

(a) Race/ethnicity and (b) transmission category of potential transmission partners. These figures show, for persons in the transmission network of a given (a) race/ethnicity or (b) transmission category, the (a) race/ethnicity or (b) transmission category of their potential transmission partners. MSM: men who have sex with men; PWID: persons who inject drugs; NHOPI: Native Hawaiian or Other Pacific Islander.

Table 1

Characteristics of persons with HIV-1 sequences, by number of potential transmission links, 25 U.S. jurisdictions, 2001–2012.

	Total		0 links		1–3 links		4+ links		
	n	Column %	n	Row %	n	Row %	n	Row %	multivariate p-value**
Total	40950	100%	28040	68%	8928	21%	3982	10%	n/a
Sex									
Male	32033	78%	20801	65%	7530	24%	3702	12%	
Female	8917	22%	7239	81%	1398	16%	280	3%	
Age at HIV Diagnosis (yrs)									<0.0001
13–19	2258	6%	1096	49%	676	30%	486	22%	
20–29	13850	34%	7896	57%	3919	28%	2035	15%	
30–39	11414	28%	8242	72%	2276	20%	896	8%	
40–49	8667	21%	6826	79%	1426	16%	415	5%	
50–59	3701	9%	3045	82%	526	14%	130	4%	
≥60	1060	3%	935	88%	105	10%	20	2%	
Race/Ethnicity									<0.0001
Asian/NHOPI	774	2%	517	67%	184	24%	73	9%	
Black	19269	47%	13824	72%	3830	20%	1615	8%	
Hispanic/Latino	9225	23%	6341	69%	2050	22%	834	9%	
White	10645	26%	6653	63%	2628	25%	1364	13%	
Other*	1037	3%	705	68%	236	23%	96	9%	
Transmission Category									<0.0001
MSM	24961	61%	15122	61%	6431	26%	3408	14%	
MSM who inject drugs	1605	4%	1108	69%	353	22%	143	9%	
Male PWID	2204	5%	1894	86%	248	11%	61	3%	
Female PWID	1678	4%	1410	84%	215	13%	53	3%	
Heterosexual men	3229	8%	2646	82%	494	15%	89	3%	
Heterosexual women	7219	18%	5810	80%	1182	16%	227	3%	
Other/Unknown	54	0%	49	89%	5	9%	1	2%	
Population of Area of Residence									<0.0001
Nonmetropolitan Areas (<50,000)	2284	6%	1603	70%	526	23%	155	7%	

	Total		0 links		1–3 links		4+ links		
	n	Column %	n	Row %	n	Row %	n	Row %	multivariate p-value**
Metropolitan Areas (50,000–499,999)	3908	10%	2698	69%	859	22%	351	9%	
Metropolitan Statistical Areas (≥500,000)	34556	84%	23567	68%	7520	22%	3469	10%	
Unknown	202	0%	172	85%	23	11%	7	3%	
Country of Birth									<0.0001
United States	26072	64%	17366	67%	5980	23%	2726	10%	
U.S. Dependency	571	1%	469	82%	77	13%	25	4%	
Other Country	6085	15%	4651	76%	1069	18%	365	6%	
Unknown	8222	20%	5554	68%	1802	22%	866	11%	

NHOPI: Native Hawaiian or Other Pacific Islander; MSM: men who have sex with men; PWID: persons who inject drugs

*Includes American Indian/Alaska Native, other race, multiple races, or unknown race

**P-value from multivariate model predicting likelihood of having 4+ links (compared with 0–3 links). Model includes all variables listed except sex, because transmission category is stratified by sex.

Table 2

Links between men who have sex with men (MSM) and heterosexual women, by selected characteristics, 25 U.S. jurisdictions, 2001–2012.

	Number of MSM	% linking to heterosexual women	chi-square p-value	Number of heterosexual women	% linking to MSM	chi-square p-value
Total	9839	4		1409	29	
Age at HIV Diagnosis (yrs)			p<0.0001			p=0.4
13–19	898	4		179	35	
20–29	4811	3		586	29	
30–39	2349	5		337	27	
40–49	1318	5		188	28	
50+	462	8		119	30	
Race/Ethnicity			p<0.0001			<0.0001
White	3351	2		189	39	
Black	3678	7		953	25	
Hispanic/Latino	2335	4		229	37	
Asian/NHOPI	226	2		13	52	
Other	249	3		27	33	
Population of Area of Residence			p=0.0004			p=0.4
Nonmetropolitan Areas (<50,000)	431	7		136	24	
Metropolitan Areas (50,000–499,999)	824	6		191	29	
Metropolitan Statistical Areas (≥500,000)	8559	4		1082	30	
U.S. Census Region			p<0.0001			p<0.0001
Northeast	2508	2		275	24	
Midwest	1866	6		331	36	
South	2684	6		679	24	
West	2771	2		122	52	

NHOPI: Native Hawaiian or Other Pacific Islander