Characterizing HIV Transmission Networks Across the United States

Jeannette L. Aldous,[1] Sergei Kosakovsky Pond,[1] Art Poon,[8] Sonia Jain,[1] Huifang Qin,[1] James S. Kahn,[2] Mari Kitahata,[4] Benigno Rodriguez,[5] Ann M. Dennis,[6] Stephen L. Boswell,[7] Richard Haubrich,[1] and Davey M. Smith[1,3]

[1]University of California, San Diego, La Jolla, [2]University of California, San Francisco, and [3]Veterans Affairs Medical Center, San Diego, California, [4]University of Washington, Seattle, [5]Case Western Reserve University, Cleveland, Ohio, [6]University of North Carolina at Chapel Hill, [7]Fenway Health/Harvard Medical School, Boston, Massachusetts; and [8]BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada

Background. Clinically, human immunodeficiency virus type 1 (HIV-1) *pol* sequences are used to evaluate for drug resistance. These data can also be used to evaluate transmission networks and help describe factors associated with transmission risk.

Methods. HIV-1 *pol* sequences from participants at 5 sites in the CFAR Network of Integrated Clinical Systems (CNICS) cohort from 2000–2009 were analyzed for genetic relatedness. Only the first available sequence per participant was included. Inferred transmission networks ("clusters") were defined as ≥2 sequences with ≤1.5% genetic distance. Clusters including ≥3 patients ("networks") were evaluated for clinical and demographic associations.

Results. Of 3697 sequences, 24% fell into inferred clusters: 155 clusters of 2 individuals ("dyads"), 54 clusters that included 3–14 individuals ("networks"), and 1 large cluster that included 336 individuals across all study sites. In multivariable analyses, factors associated with being in a cluster included not using antiretroviral (ARV) drugs at time of sampling ($P < .001$), sequence collected after 2004 ($P < .001$), CD4 cell count >350 cells/mL ($P < .01$), and viral load 10 000–100 000 copies/mL ($P < .001$) or >100 000 copies/mL ($P < .001$). In networks, women were more likely to cluster with other women ($P < .001$), and African Americans with other African Americans ($P < .001$).

Conclusions. Molecular epidemiology can be applied to study HIV transmission networks in geographically and demographically diverse cohorts. Clustering was associated with lack of ARV use and higher viral load, implying transmission may be interrupted by earlier diagnosis and treatment. Observed female and African American networks reinforce the importance of diagnosis and prevention efforts targeted by sex and race.

Comparative analysis of human immunodeficiency virus type 1 (HIV-1) *pol* sequences can be used to define transmission networks, or "clusters," within a population [1–7]. Across geographically distant communities, identifying highly related clusters can evaluate HIV transmission networks within and between various populations. Such analyses could be particularly useful for developing strategies to interfere with HIV transmission among risk groups or among populations of interest, such as women and racial and ethnic minorities.

To date, most HIV molecular transmission studies have focused on cohorts that are limited by geography [6], HIV risk factor [2, 6, 8], recent infection [5, 7–11], or epidemiologic linkage [7, 12]. This leaves open the question of whether phylogenetic analysis could be useful to identify HIV transmission networks in larger, more diverse cohorts. Thus, we evaluated transmission networks within the CFAR Network of Integrated Clinical Systems (CNICS) HIV cohort [13]. We sought to determine (1) whether transmission networks (clusters) can

Received 13 January 2012; accepted 29 June 2012; electronically published 10 July 2012.

Correspondence. Jeannette L. Aldous, MD Division of Infectious Diseases, University of California, San Diego, 200 W Arbor Dr, MC 8208, San Diego, CA 92103 (jlaldous@ucsd.edu).

Clinical Infectious Diseases 2012;55(8):1135–43

DOI: 10.1093/cid/cis612

be identified in the CNICS cohort, consisting primarily of chronically infected patients; (2) whether the density of sample would be sufficient to study women and racial and ethnic minorities; and (3) which specific variables were associated with transmission networks.

METHODS

Study Population

CNICS is an observational HIV cohort at 8 US academic centers [13]. Five CNICS sites contributed to this study: University of Washington (UW), University of California, San Francisco (UCSF), Case Western Reserve University (CWRU), University of North Carolina at Chapel Hill (UNC), and Fenway Health/Harvard Medical School (FW). All participants who had an available HIV-1 *pol* nucleotide sequence were included. For those with multiple sequences, the first sequence was selected. Demographic and clinical data at time of sampling included age, sex, self-reported race and ethnicity, self-reported HIV risk factors, antiretroviral (ARV) use and ARV exposure history, CNICS site, year, CD4 cell count, and viral load. Resistance associated mutations in protease and reverse transcriptase were evaluated based on International AIDS Society (IAS-USA) definitions [14].

Cluster Analysis

We evaluated sequence relatedness using pairwise Tamura-Nei 93 (TN93) distances [15]. The TN93 distance corrects for substitution biases and unequal base composition in HIV [16] and is a biologically realistic model that permits rapid comparisons of 10^4–10^5 aligned sequences. Bulk *pol* sequences often contain mixed nucleotide bases [17], representing within-host polymorphisms, and 87% of our sequences contained ≥1 mixed base. We resolved mixed bases using a "partially derived" approach to maximize the number of nucleotide matches (see Supplementary Methods). To define clustering, a group of sequences formed a cluster at a given threshold (*D*), if and only if each sequence in the group had TN93 distance of D or less to *at least* 1 other sequence in the group. As an example, if for sequences A, B, and C, $(A, B) \leq 1.5\%$, $(A, C) \leq 1.5\%$, and $(B, C) > 1.5\%$, then the 3 sequences are in a cluster at $D \leq 1.5\%$.

The $D \leq 1.5\%$ genetic distance cutoff was selected based on the following: (1) the expected genetic distance for genetically *unrelated* sequences in the United States epidemic is >5% [18]; (2) 1.5% demarcated the 0.014 percentile of the TN93 distribution, making it very unlikely for a pair of randomly selected sequences to demonstrate <1.5% genetic distance from each other; and (3) 1.5% is the standard used by others in the field [3, 5, 19].

To ascertain that the largest cluster (cluster 3, comprising 336 individuals) was composed of related sequences, and not the result of "chaining" the links (whereby 2 individuals in a cluster are linked through several intermediaries but are themselves as distant as any 2 random sequences), we performed 2 checks. First, we computed the distribution of all pairwise distances in cluster 3 and compared it to the overall distribution from the entire data set. Second, we drew 100 random subsets of 336 sequences from the entire data set, computed all the pairwise distances between pairs of sequences in each random data set, evaluated the probability that a pairwise distance from cluster 3 was greater than a pairwise distance from a random cluster, and averaged this quantity over 100 random clusters. We also evaluated the overlap between the tails of the empirical pairwise TN93 distance distribution from cluster 3, and the corresponding distribution from the random subset.

Univariate analyses (χ^2) and multivariable logistic regression models were applied to determine associations between variables and clustering. Groups analyzed included (1) all patients and (2) clustering versus nonclustering patients. For the evaluation of resistance prevalence, we also included comparisons of patients who were ARV naive versus ARV experienced and clustering ARV naive versus nonclustering ARV naive. Spearman's rank test for correlation was used to evaluate associations between variables.

A subgroup analysis evaluated demographic characteristics of the population that fell into larger clusters of ≥3 sequences (termed "networks"), allowing for the identification of a predominant racial or sex characteristic of the cluster. For example, if >50% of the members were female, the cluster was defined as a "female network." Then the proportion of females who were in female networks was compared with the proportion of females who were not in female networks using Fisher's exact test. The same method was used to evaluate associations by race or ethnicity. To further confirm observed network associations by race and sex, we randomly permuted sex or race labels among subjects found in defined networks and evaluated the proportion of subjects expected to be in those networks if association was random. This procedure was repeated 1000 times to derive the expected proportions and *P* values.

RESULTS

Study Population

All participants who had ≥1 HIV-1 *pol* sequence available were evaluated: 1165 from UCSF, 1115 from UW, 666 from CWRU, 512 from UNC, and 244 from FW. Of 3640 sequences collected between 2000 and 2009, 5 contained errors and were treated as missing, but an additional 62 sequences from 1999 and 3 sequences from 2010 were available and were included, resulting in 3697 total sequences for the cluster analyses (Table 1). Approximately 98% of sequences were subtype B. The study population was 16% female, 33% African

Table 1. Demographic Characteristics

	Patients, No. (%)		
Characteristic	All	In a Cluster	Not in a Cluster
All	3697	885 (24)	2812 (76)
Sex			
Female	606	136 (22)	470 (78)
Male	3030	735 (24)	2295 (76)
Unknown	61	14 (23)	47 (77)
Race			
Black	1216	273 (22)	943 (78)
White	1855	457 (25)	1398 (75)
Asian/PI	91	25 (27)	66 (73)
Other	232	50 (22)	182 (78)
Unknown	303	80 (26)	223 (74)
Ethnicity			
Hispanic	468	126 (27)	342 (73)
Non-Hispanic	1499	341 (23)	1158 (77)
Unknown	1730	418 (24)	1312 (76)
Age, years[a]			
≤40	1919	496 (26)	1423 (74)
>40	1656	373 (23)	1283 (77)
Unknown	122	16 (13)	106 (87)
HIV risk factor			
MSM	1862	435 (23)	1427 (77)
Heterosexual	837	190 (23)	647 (77)
IDU	356	88 (25)	268 (75)
MSM/IDU	384	111 (29)	273 (71)
Unknown/other	258	61 (24)	197 (76)
CNICS site[b]			
UCSF	1161	331 (29)	830 (71)
UW	1114	249 (22)	865 (78)
CWRU	666	189 (28)	477 (72)
UNC	512	84 (16)	428 (84)
FW	244	32 (13)	212 (87)
Year[b]			
2000–2001	260	33 (13)	227 (87)
2002–2003	497	80 (16)	417 (84)
2004–2005	900	243 (27)	657 (73)
2006–2007	1190	308 (26)	882 (74)
2008–2009	723	210 (29)	513 (71)
Unknown/other	127	11 (8)	116 (92)
ARV history[b]			
Naive	1528	470 (31)	1058 (69)
Exposed	2047	399 (19)	1648 (81)
Unknown	122	16 (13)	106 (87)
ARV status[b]			
On ARV	1272	182 (14)	1090 (86)
Off ARV	2303	687 (30)	1616 (70)
Unknown	122	16 (13)	106 (87)
Viral load, copies/mL[b]			
<10 000	1006	170 (17)	836 (83)
10 000–100 000	1382	361 (26)	1021 (74)
>100 000	844	250 (30)	594 (70)
Unknown	465	104 (22)	361 (78)
CD4 cell count, cells/mL[c]			
<50	519	107 (21)	411 (79)
50–200	885	208 (24)	675 (76)
201–350	874	217 (25)	657 (75)
>350	1297	337 (26)	958 (74)
Unknown	122	16 (13)	111 (87)
Resistance[b]			
Yes	2004	371 (19)	1633 (81)
No	1693	514 (30)	1179 (70)

Abbreviations: ARV, antiretroviral; CNICS, CFAR Network of Integrated Clinical Systems; CWRU, Case Western Reserve University; FW, Harvard/Fenway; HIV, human immunodeficiency virus; IDU, intravenous drug use; MSM, men who have sex with men; PI, Pacific Islander; UCSF, University of California, San Francisco; UNC, University of North Carolina, UW, University of Washington.

Univariate comparisons of cluster versus noncluster:

[a] $P < .05$.

[b] $P < .001$.

[c] Marginal association ($P < .1$).

American, and 13% Hispanic (although there was a 40% nonresponse rate for the Hispanic ethnicity variable). Self-reported HIV risk factors included 50% men who have sex with men (MSM), 23% heterosexual, 10% intravenous drug use (IDU), and 10% both MSM and IDU risk factor (MSM/IDU). At the time the sequence was generated, 41% were ARV naive and 55% were experienced (4% unknown). Treatment experienced patients may have been on or off therapy when the sequence was generated. During the study period, there was an increase in the total number of sequences over time (mean sequences per year, 190 for 2000–2003 vs 469 for 2004–2009) and an increase in the proportion of ARV-naive participants (mean, 15% in 2000–2003 vs 49% in 2004–2009).

Cluster Prevalence

The overall mean distribution of pairwise TN93 distances was 5.6% (median, 5.3%; interquartile range, 4.6%–6.3%; Supplementary Figure 1). Of the 3697 sequences, 885 (24%) fell into clusters at a genetic distance of ≤1.5%, resulting in 209 clusters ranging in size from 2 to 14 individuals, plus 1 outlier cluster, which encompassed 336 individuals (Figure 1).

Variables Associated With Clustering

In univariate analysis, individuals whose sequences clustered were more likely to be younger ($P = .02$), ARV naive

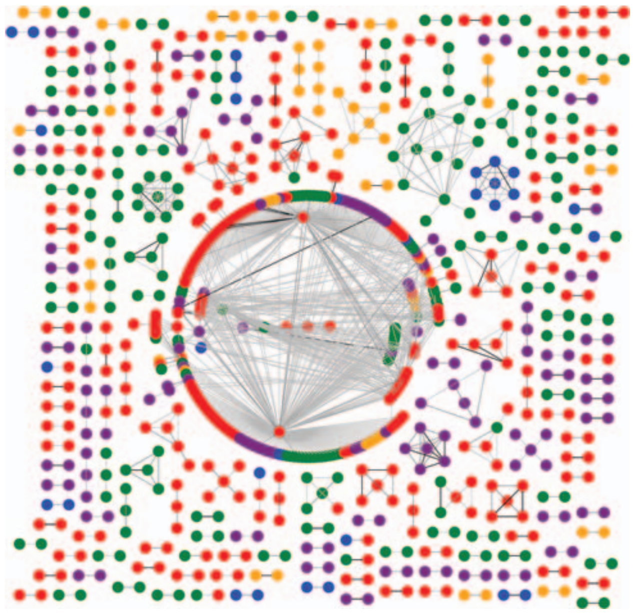

Figure 1. Cluster overview with pictorial representation of all patients who clustered at ≤1.5%. Each cluster patient (n = 885) is represented as a colored dot with lines connecting phylogenetically linked sequences. Black line connections represent genetic distances of <0.05%; dark gray lines, 0.5%–1%; and light gray lines, 1%–1.5%. Sequences in cluster 3 can be seen in the center of the figure. Red dots represent patients from the University of California, San Francisco; green, the University of Washington; blue, Harvard/Fenway; purple, Case Western Reserve University; and orange, the University of North Carolina.

(*P* < .001), sampled after 2004 (*P* < .01), had higher viral loads (*P* < .001), and were marginally more likely to have higher CD4 cell counts (*P* = .09) than those who did not cluster. Cluster patients were less likely to be from FW and UNC and less likely to be receiving ARV therapy (*P* < .001) and to have ARV resistance–associated mutations (*P* < .001; Table 1).

As expected, the viral load variable was significantly correlated with CD4 cell counts (*P* < .001) and ARV use (*P* < .001). The "ARV exposed" and "on ARV" variables were also highly correlated (*P* < .001). Because those definitions overlapped, only "on ARV" was included in the multivariate analysis. In multivariate analysis (Table 2), later year of sampling, not currently receiving ARV therapy, higher viral load, and higher CD4 cell counts remained independently associated with clustering. Age ≤40 years and the MSM/IDU risk factor were marginally associated with clustering (*P* = .09 and .06). Participants from the FW and UNC sites were less likely to fall into clusters than participants from the other sites (*P* < .01), probably owing to fewer participants (lower sampling density) at those sites. There was no difference in proportion clustering by sex, race, or ethnicity.

Table 2. Multivariable Analysis for Likelihood of Clustering

Variable	Likelihood of Clustering, OR (95% CI)	*P*
Sex		
Male	…	…
Female	1.1 (.8–1.6)	.4
Race		
Black	…	…
White	0.9 (.7–1.1)	.33
Ethnicity		
Hispanic	1.1 (.6–1.7)	.7
Non-Hispanic		
Age, years		
≤40	1.2 (.9–1.4)	**.09**
>40	…	…
HIV risk factors (vs MSM)		
MSM	…	…
Heterosexual	1.0 (.7–1.3)	.8
IDU	1.1 (.8–1.6)	.5
MSM/IDU	1.3 (1–1.7)	**.06**
CNICS site (vs UCSF)		
UCSF	…	…
UW	1.0 (.6–1.7)	.9
CWRU	1.2 (.8–2.3)	.3
UNC	0.6 (.4–.9)	**.01**
FW	0.5 (.3–.8)	**.005**
Year (vs 2000–2001)		
2000–2001	…	…
2002–2003	1.0 (.6–1.8)	.9
2004–2005	2.0 (1.2–3.2)	**.009**
2006–2007	1.6 (.9–2.6)	**.08**
2008–2009	1.7 (1.0–2.9)	**.04**
ARV status	…	…
On ARV		
Off ARV	1.9 (1.5–2.4)	**<.0001**
Viral load, copies/mL (vs <10 000)		
<10 000	…	…
10–100 000	1.6 (1.2–2.0)	**.0002**
>100 000	2.0 (1.5–2.6)	**<.0001**
CD4 cell count, cells/mL (vs <50)		
<50	…	…
50–200	1.1 (.9–1.8)	.1
201–350	1.2 (1.0–1.8)	**.07**
>350	1.5 (1.1–2.1)	**.006**

Multivariable analysis included sex, race, ethnicity, age, HIV risk factors, ARV status, CNICS site, categorical CD4 cell count and viral load, and year of sampling. Boldface *P* values denote statistically significant variables.

Abbreviations: ARV, antiretroviral; CNICS, CFAR Network of Integrated Clinical Systems; CI, confidence interval; CWRU, Case Western Reserve University; FW, Harvard/Fenway; HIV, human immunodeficiency virus; IDU, intravenous drug use; MSM, men who have sex with men; OR, odds ratio; UCSF, University of California, San Francisco; UNC, University of North Carolina, UW, University of Washington.

Cluster Size and Clustering Across Sites

Of the 210 clusters, 155 (74%) included 2 participants ("dyads"), 54 clusters included 3–14 individuals ("networks"; n = 239), and 1 cluster ("cluster 3") had 336 persons (Figure 2). The majority of clusters were confined within geographic locations: only 22 of 210 clusters (and 9 of 54 networks) crossed 2 sites, and only cluster 3 spanned all sites.

Cluster 3 contained 336 individuals and arose from 7 sequences at the ≤0.5% genetic distance threshold, 65 at ≤1%, and 336 at ≤1.5%. Mean pairwise distance within the cluster was 0.032, significantly lower than expected by chance ($P < .01$; Supplementary Figure 2). Comparison of the mean distance between pairs of cluster 3 sequences and those of 100 random subsets of 336 sequences found that the mean value from the random subset was greater than that of cluster 3, highlighting the close distance between cluster 3 sequences (Supplementary Figure 3). Further, an evaluation of the tails of pairwise distance distributions for cluster 3 and those estimated from the random subsets found that the overlap was small (average 6.4%), again supporting the unusual relatedness of sequences in this cluster.

Cluster 3 included sequences from all study sites, but participants were more likely to be from CWRU or UCSF and less likely to be from UW or FW ($P < .001$). Cluster 3 patients were more likely to be African American ($P = .014$), be >40 years old ($P < .001$), have IDU as their HIV risk factors ($P = .04$), be ARV experienced ($P = .04$) and have higher viral load ($P = .002$) (Supplementary Table 1). These associations may represent characteristics specific for this network, because they contrast with the findings from the overall cohort, in which clustering was associated with ARV-naive status and younger age.

Networks by Sex and Race

To investigate clustering by demographic group, networks were investigated in a subgroup analysis, representing 43 females, 194 males, and 2 of unknown sex and 58 blacks, 46 Hispanics, and 4 of unknown race. There were 9 majority-female networks, 16 "racial/ethnic minority" networks (14 majority African American, 2 equally mixed black/Hispanic), and 1 majority-Hispanic network. Of the 43 women, 79% identified their HIV risk factor as heterosexual contact, and 36 (84.3%) fell into majority-female networks ($P < .001$; Figure 3*A*). Similarly, African Americans were more likely to cluster with other African Americans: 77% were in majority-black networks, and 84.5% were in black and/or Hispanic networks ($P < .001$; Figure 3*B*). Based on 1000 replicates of a permutations test, the observed ratios for clustering for females and racial or ethnic minorities (83.7% for women, 84.5% for minorities) were significantly higher than the ratios expected by chance (15% and 32% respectively; $P < .001$). Although only 1 cluster included >50% Hispanics (all from UW), Hispanic participants were also more likely to cluster with African Americans; as a result, 58% of Hispanics were in "minority" networks.

Resistance

Resistance prevalence was relatively high given that we defined resistance as the presence of *any* IAS-USA major mutation,

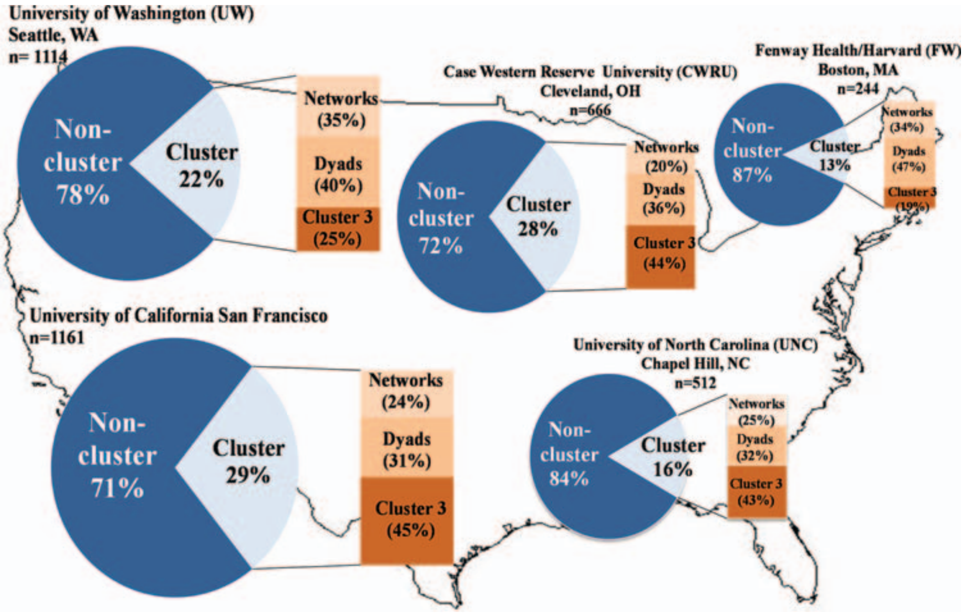

Figure 2. Map of cohort population, percentage of clustering by site, and proportion of cluster patients in "dyads" (clusters of 2 patients), "networks" (clusters of >2 patients), or in "cluster 3" (a unique cluster that spanned all sites and included 9% of cluster sequences).

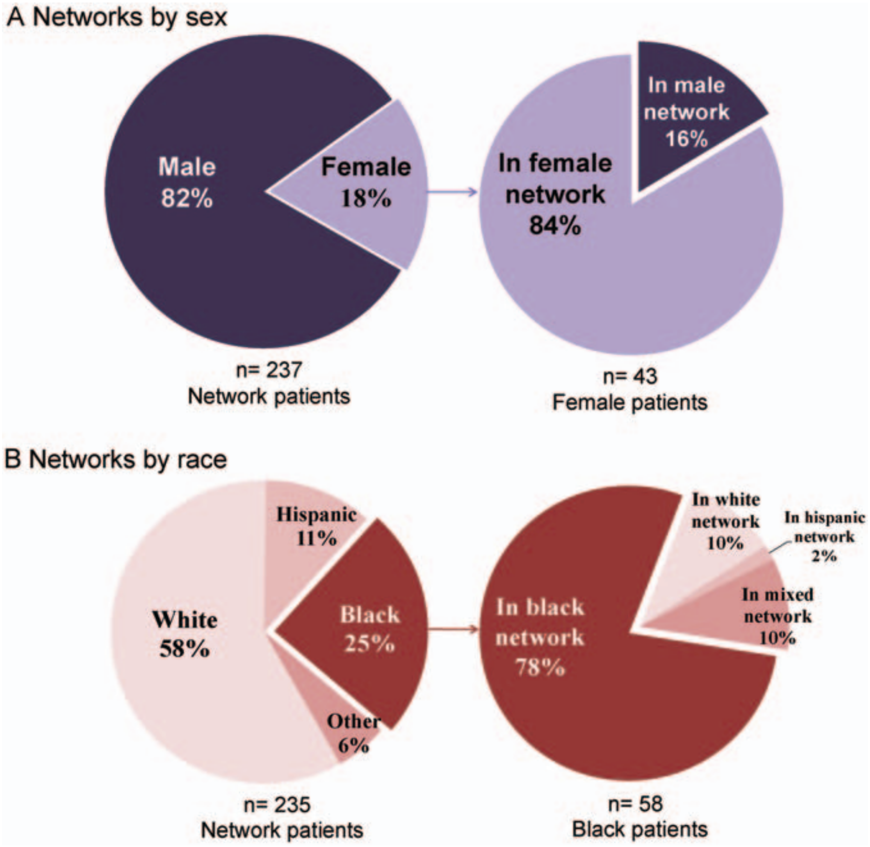

Figure 3. Analysis of networks. Clusters containing ≥3 patients were defined by the predominant demographic characteristic of patients in that network. For example, a network was defined as "female" if >50% of the patients in the network were female. *A,* Among women, 36 of 43 (84%) clustered with other women. *B,* Among blacks, 45 of 58 (78%) clustered with other blacks.

but sequences that clustered were less likely to have resistance compared to sequences that did not cluster (42% vs 58%; $P < .001$; Table 1 and Supplementary Table 2*A*). As expected, ARV-naive participants were less likely to have resistance mutations than those who were ARV-experienced (37% vs 66%; $P < .001$). Because ARV-naive participants were independently more likely to cluster than the ARV-experienced group, this may explain in part the lower prevalence of resistance among those who clustered. When we analyzed only the ARV-naive population, there was no difference in the prevalence of resistance among those who clustered and those who did not (36% vs 38%; $P = .4$), providing further evidence that viral evolution associated with the development of drug resistance was not associated with clustering (Supplementary Table 2*B*).

DISCUSSION

Traditional epidemiologic techniques for identifying and interrupting HIV transmission networks are limited by individual recall and delays between infection and diagnosis. This is underscored by a number of studies that have shown discordance between patient-reported epidemiologic and phylogenetic linkage [7, 19–21]. Although it is important to note that phylogenetic analysis is not an appropriate tool to identify direct transmissions [22], it may provide accurate information on transmission networks. This study used phylogenetic techniques to investigate transmission networks in the United States by including all available CNICS sequences without limitation by stage of HIV infection, use of ARVs, demographics, or geographic location.

Because previous studies using phylogenetic analysis have assessed HIV transmission more narrowly by focusing on early stages of infection [3, 6, 9, 11], specific risk groups [23, 24], drug resistance [5,11,25,26], movement across regions [2,27,28], and distribution of HIV-1 subtypes [4, 27, 29], our first aim was to evaluate if inclusion of a broader population would provide sufficient sample density to identify networks. We found that the inclusion of chronically infected patients from geographically distant sites across the United States demonstrated a high degree of clustering (24% clustered with at least

1 other person). The majority of these clusters were contained within each site, providing evidence that identified networks were likely true epidemiologic networks.

Our inferred transmission networks did not greatly overlap across racial and ethnic boundaries, a finding also observed in smaller studies of black men [30, 31] and a recent study of Los Angeles County Service Planning Areas [32]. Interestingly, our study also found networking by sex, with women much more likely to fall into transmission networks with other women. Although women represent 25% of the US epidemic [33], we are not aware of any other large study of female transmission networks using molecular epidemiology. In this retrospective study we cannot determine specific epidemiologic relationships between these women, but we propose 2 hypotheses: (1) because most of the women self-identified as heterosexual, this finding may point to a significant underdiagnosis of HIV infection in men who have sex with women, and (2) these clusters may represent female social networks with shared risk. If so, this supports the possibility that HIV testing focused on the social networks of HIV-positive women could be effective in finding other undiagnosed HIV-infected women [34, 35].

Clustering was associated with lack of ARV use, higher viral load, and higher CD4 cell counts; factors also observed among persons with newly acquired HIV infection. Higher CD4 cell counts are also associated with ARV therapy, however ARV-exposed patients were less, not more, likely to cluster. Thus, a more likely explanation for higher CD4 cell counts in the cluster group is that these participants were earlier in their course of infection or had ARV therapy deferred based on CD4 cell count guidelines [36, 37]. These associations are similar to findings in other reports of HIV transmission among recently infected individuals and those with higher viral loads [3, 9, 38], and are consistent with other molecular epidemiology studies that demonstrated phylogenetic clustering in both ARV-naive and experienced patients [39, 40]. Together, these data add to the growing body of evidence that early treatment may have a considerable impact on the spread of the HIV epidemic [41, 42].

Cluster 3 was a significant outlier in size from the rest of the observed clusters, encompassing 38% of all sequences that clustered and 9% of sequences overall, and was overrepresented among MSM and IDU risk factors. Such a large cluster raises questions that the cluster represents overlap of loosely related smaller clusters. However, analysis of maximal distances within cluster 3 showed that its sequences were closely related (Supplementary Figure 2) and that the chances of such a cluster occurring by chance were quite low (Supplementary Figure 3).

Although the results of the study were robust, some limitations remain. It should be stressed again that we inferred linkage based on genetic relatedness, which does not definitively determine epidemiologic linkage or imply direct transmission. However, the findings of clustering by demographic and clinical factors provide substantial evidence that the phylogenetically inferred networks are probably true epidemiologic networks. We found that ARV use and viral load were significantly associated with clustering, which fits with expectations of transmission risks. However, certain clinical variables (ARV use, viral load, CD4 cell count) are highly correlated, which may influence the power to see independent effects, so there may have been other associations that we did not have power to detect.

Including ARV-experienced participants may raise concerns about falsely interpreting clustering secondary to drug resistant mutations; however, measures of genetic distance removed the codons associated with amino acid changes owing to resistance [7, 43]. Similarly, including chronically infected individuals may raise questions about viral evolution between time of transmission and generation of the sequence [44]. However, unlike other HIV genes (eg, *env*) *pol* evolves relatively slowly (<1.5% over 5 years) [18].

Selection bias may arise from studying patients who underwent genotyping and excluding those who did not; however, whereas differences are possible (eg, patients who underwent genotyping may have had higher viral loads than those who did not undergo genotyping), it is unclear how they would specifically affect the validity of inferred transmission networks. Self-reporting of HIV risk factors and ethnicity may have also introduced some bias. For example, the cohort had a high number of nonresponses for Hispanic ethnicity, so there were probably more Hispanic patients in our networks than were identified. Finally, the female and African American networks were small, making conclusions preliminary.

To our knowledge, this the first study to use molecular epidemiology to identify transmission networks in such a diverse population, across such a wide geographic distribution, and without limitations by stage of infection or ARV status. The relatively high degree of clustering, and the fact that clusters could be identified for demographic groups, confirms that, even in large and diverse populations, relatedness of HIV-1 *pol* sequences can be used within and across communities to assess variables associated with transmission networks. These observations may also be important for public health interventions, which could use assessments of transmission networks to identify risks for expanding epidemics and target effective public health responses.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online (http://cid.oxfordjournals.org). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We are grateful to Joe Eron, MD, and Sonia Napravnik, PhD, from the University of North Carolina, for contribution of site data and for manuscript review.

Author contributions. Contribution of data from CNICS study sites: J. S. K., M. K., B. R., A. M. D., S. L. B.; data compilation: J. L. A., H. Q., S. K. P.; planning of the analysis: J. L. A., D. M. S., A. P., S. K. P., R. H.: statistical analysis: H. Q., S. K. P., S. J., J. L. A.; preparation of graphs and tables: J. L. A.; drafting of the manuscript: J. L. A., D. M. S.; manuscript review and comment: all authors; final manuscript approval: all authors. J. L. A. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Financial support. This project was supported by the CFAR-Network of Integrated Clinical Systems (CNICS) funding from the National Institute of Allergy and Infectious Diseases (NIAID), National Heart, Lung, and Blood Institute (grant 5R24AI067039), the National Institutes of Health (NIH; grants K24 AI064086, K24AI100665, R21AI080397, AI043638, MH62512, MH083552, AI077304, AI047745, AI74621, AI080353, and DA034978-01); the James B. Pendleton Charitable Trust; the University of California San Diego Center for AIDS Research (grant AI36214); and the San Diego AIDS Clinical Trials Group (AI69432). J. L. A. receives support from NIAID (supplement to grant R24AI067039) and from the National Center on Minority Health and Health Disparities; A. P. receives support from the Canadian Institutes of Health Research (fellowships award and operating grant); J. S. K. receives support from the National Center for Research Resources (grants HHSN268200425213 and RR024369); and A. M. D. receives support from NIAID (grant T32 AI007001-33).

Potential conflicts of interest. D. M. S. has received grant support from Pfizer and served as a consultant for Gen-Probe. S. K. P. has served as a consultant for Gen-Probe and Monogram Biosciences. S. L. B. has served as a consultant for Merck. A. M. D. has received grant support from the Bristol-Myers Squibb Virology Fellowship program. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. 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**; 18:719–28.
2. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci U S A **2005**; 102:4425–9.
3. Brenner BG, Roger M, Routy JP, et al. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis **2007**; 195:951–9.
4. Gifford RJ, de Oliveira T, Rambaut A, et al. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. J Virol **2007**; 81:13050–6.
5. Brenner BG, Roger M, Moisi DD, et al. Transmission networks of drug resistance acquired in primary/early stage HIV infection. AIDS **2008**; 22:2509–15.
6. 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.
7. 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**; 23:225–32.
8. Drumright L, Little SJ, Richman DD, Frost SD. Age discordance and drug resistance predict clustering of HIV among recently-infected MSM in San Diego, CA. In: 14th Conference on Retroviruses and Opportunistic Infections. Vol 654. Los Angeles, CA, **2007**.
9. Yerly S, Vora S, Rizzardi P, et al. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. AIDS **2001**; 15:2287–92.
10. Pao D, Fisher M, Hue S, et al. Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. AIDS **2005**; 19:85–90.
11. Yerly S, Junier T, Gayet-Ageron A, et al. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS **2009**.
12. Trask SA, Derdeyn CA, Fideli U, et al. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol **2002**; 76:397–405.
13. Kitahata MM, Rodriguez B, Haubrich R, et al. Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol **2008**; 37:948–55.
14. Johnson VA, Brun-Vezinet F, Clotet B, et al. Update of the drug resistance mutations in HIV-1: December 2010. Top HIV Med **2010**; 18:156–63.
15. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol **1993**; 10:512–26.
16. Hue S, Clewley JP, Cane PA, Pillay D. Investigation of HIV-1 transmission events by phylogenetic methods: requirement for scientific rigour. AIDS **2005**; 19:449–50.
17. Poon AF, Kosakovsky Pond SL, Bennett P, Richman DD, Leigh Brown AJFrost SD. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. PLoS Pathog **2007**; 3:e45.
18. Hightower GK, May SJ, Mehta SR, et al. The velocity of HIV-1 pol evolution following primary infection. In: 18th Conference on Retroviruses and Opportunistic Infections. Boston, MA, **2011**.
19. Bao L, Vidal N, Fang H, et al. Molecular tracing of sexual HIV Type 1 transmission in the southwest border of China. AIDS Res Hum Retroviruses **2008**; 24:733–42.
20. Yirrell DL, Pickering H, Palmarini G, et al. Molecular epidemiological analysis of HIV in sexual networks in Uganda. AIDS **1998**; 12: 285–90.
21. Resik S, Lemey P, Ping LH, et al. Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. AIDS Res Hum Retroviruses **2007**; 23:347–56.
22. Hecht FM, Wolf LE, Lo B. Lessons from an HIV transmission pair. J Infect Dis **2007**; 195:1239–41.
23. Cachay ER, Moini N, Kosakovsky Pond SL, et al. Active methamphetamine use is associated with transmitted drug resistance to non-nucleoside reverse transcriptase inhibitors in individuals with HIV infection of unknown duration. Open AIDS J **2007**; 1:5–10.
24. Bezemer D, van Sighem A, Lukashov VV, et al. Transmission networks of HIV-1 among men having sex with men in the Netherlands. AIDS **2010**; 24:271–82.
25. Kaye M, Chibo D, Birch C. Phylogenetic investigation of transmission pathways of drug-resistant HIV-1 utilizing pol sequences derived from resistance genotyping. J Acquir Immune Defic Syndr **2008**; 49:9–16.
26. Chan PA, Tashima K, Cartwright CP, et al. Short communication: transmitted drug resistance and molecular epidemiology in antiretroviral naive HIV type 1-infected patients in Rhode Island. AIDS Res Hum Retroviruses **2011**; 27:275–81.
27. Paraskevis D, Pybus O, Magiorkinis G, et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology **2009**; 6:49.

28. Mehta SR, Wertheim JO, Delport W, et al. Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. Infect Genet Evol **2011**; 11:975–9.
29. Veras NM, Santoro MM, Gray RR, et al. Molecular epidemiology of HIV type 1 CRF02_AG in Cameroon and African patients living in Italy. AIDS Res Hum Retroviruses **2011**; 27:1173–82.
30. Oster A, Mena L, Wejnert C, Heffelfinger J. Network analysis among HIV-infected young black MSM demonstrates high connectedness around few venues. In: 18th Conference on Retroviruses and Opportunistic Infections. Boston, MA, **2011**.
31. Oster A, Pieniazek D, Switzer W, et al. Phylogenetic analysis shows insularity with respect to HIV transmission of young black men in Mississippi who have sex with men. In: 17th Conference on Retroviruses and Opportunistic Infections. San Francisco, CA, **2010**.
32. Sayles J, Rurangirwa J, Aldous J, Pond S, King J, Smith D. Public health applications of molecular epidemiology: use of HIV-1 pol sequences to identify HIV transmission networks in Los Angeles County. In: National HIV Prevention Conference. Atlanta, GA, **2011**.
33. Centers for Disease Control and Prevention. Fact Sheet: HIV Among Women. August 2011. Available at: http://www.cdc.gov/hiv/topics/women. **2010**. Accessed 15 July 2012.
34. Use of social networks to identify persons with undiagnosed HIV infection—seven U.S. cities, October 2003-September 2004. MMWR Morb Mortal Wkly Rep **2005**; 54:601–5.
35. Kimbrough LW, Fisher HE, Jones KT, Johnson W, Thadiparthi S, Dooley S. Accessing social networks with high rates of undiagnosed HIV infection: the social networks demonstration project. Am J Public Health **2009**; 99:1093–9.
36. US Department of Health and Human Services. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. US Department of Health and Human Services, **2008**.
37. Hammer SM, Eron JJ Jr., Reiss P, et al. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. JAMA **2008**; 300:555–70.
38. Quinn TC, Wawer MJ, Sewankambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med **2000**; 342:921–9.
39. Callegaro A, Svicher V, Alteri C, et al. Epidemiological network analysis in HIV-1 B infected patients diagnosed in Italy between 2000 and 2008. Infect Genet Evol **2011**; 11:624–32.
40. 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**; 204:1463–9.
41. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet **2009**; 373:48–57.
42. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med **2011**.
43. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics **2005**; 21:676–9.
44. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol **1999**; 73: 10489–502.