Open Forum Infectious Diseases

MAJOR ARTICLE

Interplay Between Geography and HIV Transmission Clusters in Los Angeles County

Britt Skaathun,[1] Manon Ragonnet-Cronin,[1,2] Kathleen Poortinga,[3] Zhijuan Sheng,[3] Yunyin W. Hu,[3] and Joel O. Wertheim[1]

[1]Department of Medicine, University of California San Diego, La Jolla, California, USA, [2]Department of Infectious Disease Epidemiology, Imperial College London, London, UK, and [3]Division of HIV and STD Programs, Department of Public Health, Los Angeles County, California, USA

Background. Clusters of HIV diagnoses in time and space and clusters of genetically linked cases can both serve as alerts for directing prevention and treatment activities. We assessed the interplay between geography and transmission across the Los Angeles County (LAC) HIV genetic transmission network.

Methods. Deidentified surveillance data reported for 8186 people with HIV residing in LAC from 2010 through 2016 were used to construct a transmission network using HIV-TRACE. We explored geographic assortativity, the tendency for people to link within the same geographic region; concordant time–space pairs, the proportion of genetically linked pairs from the same geographic region and diagnosis year; and Jaccard coefficient, the overlap between geographical and genetic clusters.

Results. Geography was assortative in the genetic transmission network but less so than either race/ethnicity or transmission risk. Only 18% of individuals were diagnosed in the same year and location as a genetically linked partner. Jaccard analysis revealed that cis-men and younger age at diagnosis had more overlap between genetic clusters and geography; the inverse association was observed for trans-women and Blacks/African Americans.

Conclusions. Within an urban setting with endemic HIV, genetic clustering may serve as a better indicator than time–space clustering to understand HIV transmission patterns and guide public health action.

Keywords. cluster analysis; HIV infections/transmission; molecular epidemiology.

Trends in diagnosis of infectious diseases in time and space can identify outbreaks and clusters of transmission. Identification of these time–space clusters is an invaluable epidemiological tool for responding to outbreaks of infectious agents (eg, hepatitis A virus, HIV, pulmonary tuberculosis, legionellosis, etc.) [1–4].

Molecular epidemiology is the integration of pathogen genetic sequence data and classic epidemiologic information [5, 6]. It is a powerful means to determine the drivers, characteristics, geographic distribution, and dynamics of infectious disease transmission [7–9]. Genetic analyses draw inferences from the similarity between pathogen genetic sequences to reconstruct transmission histories and estimate epidemiologic parameters. This type of analysis is particularly useful in the context of HIV, where the high within-host diversity and divergence due to HIV replication errors result in a nearly unique HIV genetic sequence within each infected individual. This approach has been used to investigate recent HIV outbreaks in the United States primarily among people who inject drugs (PWID), such as in Scott County, Seattle, and Massachusetts, and can help public health officials prioritize real-time intervention strategies [10–12].

HIV incidence in the United States is declining, but concentrated efforts are required to achieve HIV elimination among populations where HIV transmission persists [13]. The US Department of Health and Human Services established the Ending the HIV Epidemic initiative, which aims to reduce HIV incidence in the United States by 90% by 2030 [14]. A pillar of this initiative involves the rapid detection of and response to emerging clusters of HIV diagnoses, as well as active partnerships between public health departments and providers [14]. The US Centers for Disease Control and Prevention (CDC) recommends investigating "priority" molecular clusters, defined as clusters characterized by rapid transmission history of recently diagnosed individuals [15]. In the absence of timely and complete sequence data, the CDC has assessed spatiotemporal increases in HIV diagnoses (above expected levels in distinct geographic areas) using case surveillance data [16]. It has been suggested that cluster detection would be most effective if molecular data were combined with traditional public health case surveillance methods [16].

Los Angeles County (LAC) in California is the most populous county in the United States, with over 10 million residents [17]. The city of Los Angeles comprises ~4 million of these residents [18]. LAC spans across 4083 square miles and is divided

Received 22 March 2021; editorial decision 19 April 2021; accepted 20 April 2021.

Correspondence: Britt Skaathun, MPH, PhD, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA (bskaathun@health.ucsd.edu).

Open Forum Infectious Diseases®2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

DOI: 10.1093/ofid/ofab211

into 507 ZIP codes, 26 Health Districts (HDs), and 8 Service Planning Areas (SPAs) [19]. At the end of 2017, there were 51 438 people with diagnosed HIV infection in LAC [20]. The 26 Health Districts in LAC align with census tracts and were designated to help guide public health resource allocation (Figure 1). In 2018, the Health District with the highest rate of new HIV diagnoses was Central, with 63 new HIV diagnoses per 100 000 population. On the other extreme, the Health District with the lowest rate of new HIV diagnoses was Alhambra, with only 7 new HIV diagnoses per 100 000 population. Here, we investigate how spatial and temporal clusters of HIV cases relate to the underlying transmission patterns in LAC, with a focus primarily at the Health District level.

METHODS

Data Source and Phylogenetic Analyses

Deidentified HIV surveillance data were used for analyses. When a person is diagnosed with HIV in the United States, service providers report information about these persons to corresponding local health departments, and this information is subsequently passed on to the CDC. HIV-1 protease and reverse transcriptase (*pol*) genetic sequences are often generated for routine antiretroviral drug resistance testing and have been reportable to the LAC Department of Public Health since 2006. We analyzed data reported to the LAC Department of Public Health from 8186 people with an HIV diagnosis from 2010 to 2016 who resided in LAC at the time of diagnosis, were ≥13 years of age, and had a reported HIV *pol* gene sequence. These sequences were used to construct an HIV molecular transmission network using HIV-TRACE [21]. The first reported sequence for every person was used, unless this sequence was genetically linked to the HXB2 reference sequence (ie, potential laboratory contaminant) or had >5% nucleotide ambiguities (ie, problematic sequence). People were linked in our genetic network if the pairwise genetic distance between their viral sequences was ≤0.015 substitutions/site (ambiguity fraction of 0.015). This distance threshold is in accordance with the observed genetic distance between named HIV risk partners in a public health surveillance setting [22].

Transmission risk categories were defined as men who have sex with men (MSM), PWID (including transgender women and MSM who inject drugs), cisgender men who reported high-risk heterosexual activity or have unknown risk, cisgender women who reported high-risk heterosexual activity or have unknown risk, and transgender women. Race/ethnicity was categorized as Hispanic/Latino, White, Black/African American, and Other (which includes multiracial, American Indian/Alaska Native, Asian, Pacific Islander). We also included additional metadata including age at diagnosis, birth country, and $CD4^+$ count (cells/mm^3) at diagnosis. We compared the characteristics of people in these clusters with those of nonclustered people using the Mantel-Haenszel χ^2 test in Stata, version 15.

Statistical Analyses

Residential information at the time of diagnosis was examined at 3 different levels of granularity: ZIP code, Health District, and Service Planning Area. We characterized the relationship between geography and the molecular transmission network using 3 distinct approaches: (i) geographic assortativity, the tendency for people to be genetically linked in the transmission network to other people residing in the same geographic region; (ii) concordant time–space pairs, the proportion of genetically linked pairs from the same geographic region and diagnosis year; and (iii) the Jaccard Coefficient, a measure of the similarity between sets of geographic region at diagnosis and cluster membership. Assortativity was assessed using Newman's method, which ranges from –1, completely disassortative, to 1, completely assortative [23]. Significance for the geographic assortativity and concordant time–space pairs was determined using 1000 random permutations of the transmission network. Primary analyses were conducted on Health Districts using a genetic distance threshold of 0.015 substitutions/site. Sensitivity analyses were conducted at the level of ZIP code and Service Planning Area and at genetic distance thresholds of 0.010 and 0.005 substitutions/site, which represent more conservative markers of direct or indirect epidemiological relatedness. For the time–space analysis, we also considered time–space pairs diagnosed during the same calendar quarter. Assortativity analyses were conducted using RStudio, version 1.1.463, with the igraph package [24]. The Jaccard analysis was conducted at the Health District level to identify individual-level correlates of overlap between transmission clusters and geographic residence using a generalized linear model with the Jaccard Coefficient as the outcome. Variables were selected for inclusion in adjusted models based on a priori knowledge about their interrelationships with HIV transmission dynamics. All regression analyses were conducted in Stata, version 15. Maps were created using Carto [25].

The study was approved by both the University of California, San Diego, and the LAC Department of Public Health institutional review boards.

RESULTS

Of the 8186 HIV cases in our analysis, 4150 (50.7%) were clustered in the network at 0.015 substitutions/site. A plurality of those clustered were between the ages of 20 and 29 (45%), were MSM (75%), were Hispanic/Latino (53%), were born in the United States (52%), and had a $CD4^+$ count at diagnosis <500 cells/mm^3 (52%) (Table 1).

Network Assortativity

Geography at diagnosis, race/ethnicity, age at diagnosis, and transmission risk were significantly assortative ($P \leq .001$) (Table 2); however, the magnitude of the assortativity

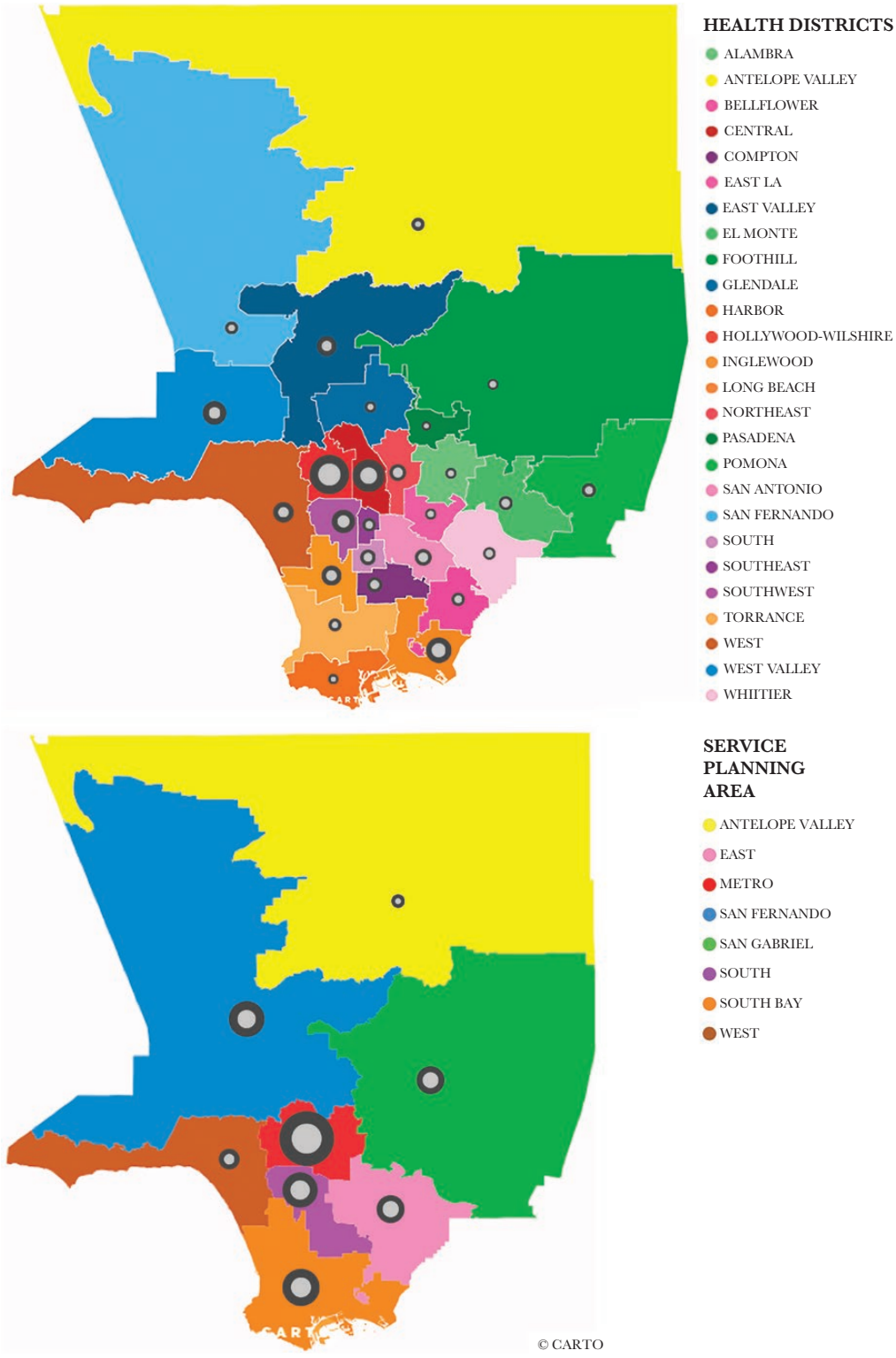

Figure 1. Los Angeles County Health Districts and Service Planning Area. Each Health District and Service Planning Area is shown in a unique color. Circles represent the number of new HIV diagnoses and sequences reported in each region, with the dark gray area proportional to the square root of the number of new HIV diagnoses and the light grey area proportional to the square root of the number of genomes reported in each region.

varied. Race/ethnicity was the most assortative attribute (assortativity, 0.27), and ZIP code was the least assortative attribute (assortativity, 0.02). Broader geographic categorizations were more assortative than smaller geographic categorizations: Service Planning Area, 0.15; Health District, 0.09; and ZIP code, 0.02.

Table 1. Breakdown of Individual-Level Characteristics by Genetic Distance Threshold Among Newly HIV-Diagnosed Persons in Los Angeles County, 2010–2016

	All	TN93 ≤0.015		TN93 ≤0.01		TN93 ≤0.005	
Characteristic	No. (%)	No. (%)	*P* Value	No. (%)	*P* Value	No. (%)	*P* Value
Total	8186 (100)	4150 (100)	–	3246 (100)	–	2120 (100)	–
Age			<.001		<.001		<.001
13–19 y	296 (4)	211 (5)		174 (5)		128 (6)	
20–29 y	3037 (37)	1858 (45)		1499 (46)		1004 (47)	
30–39 y	2310 (28)	1149 (28)		884 (27)		578 (27)	
40–49 y	1598 (20)	615 (15)		459 (14)		272 (13)	
50+ y	945 (12)	317 (8)		230 (7)		138 (7)	
Transmission risk/gender			<.001		<.001		<.001
Cisgender men	1713 (21)	713 (17)		521 (16)		303 (14)	
Cisgender women	200 (2)	64 (2)		51 (2)		33 (2)	
Transgender women	115 (1)	63 (2)		53 (2)		29 (1)	
MSM	5793 (71)	3123 (75)		2474 (76)		1664 (78)	
PWID	365 (4)	187 (5)		147 (5)		91 (4)	
Race/ethnicity			<.001				<.001
White	1848 (23)	888 (21)		682 (21)		428 (20)	
African American	1664 (20)	754 (18)		569 (18)		359 (17)	
Hispanic/Latino	4029 (49)	2193 (53)		1750 (54)		1192 (56)	
Other	645 (8)	315 (8)		245 (8)		141 (7)	
Country of birth			<.001		.001		<.001
US born/territory	4138 (51)	2175 (52)		1716 (53)		1129 (53)	
Foreign-born	2136 (26)	1017 (25)		793 (24)		515 (24)	
Unknown	1912 (23	958 (23)		737 (23)		476 (22)	
$CD4^+$ at diagnosis			<.001		<.001		<.001
<200	1666 (20)	559 (14)		389 (12)		224 (11)	
200–499	2897 (35)	1592 (38)		1268 (39)		868 (41)	
≥500	2113 (26)	1234 (30)		994 (31)		648 (31)	
Unknown	1490 (18)	757 (18)		590 (18)		376 (18)	

*$P < .001$ for all characteristics.

Abbreviations: MSM, men who have sex with men; PWID, people who inject drugs; TN93, Tamura-Nei 93 genetic distance threshold.

We found little evidence for the recapitulation of Health Districts in the transmission network (Figure 2A). All 81 clusters with at least 10 individuals had at least 3 Health Districts represented, with a median of 9 Health Districts. The largest cluster comprised 54 people from 23 unique Health Districts. Similar results were found when assessing the recapitulation of Service Planning Areas in the transmission network. The median number of 5 Service Planning Areas was represented in these clusters, and the largest cluster included people from 7 of the 9 Service Planning Areas (Supplementary Figure 1). Assortativity coefficients for each individual Health District and Service Planning Area can be found in Supplementary Table 1.

Table 2. Assortativity of Individual Characteristics Ranked From Least Assortative to Most Assortative Among Newly HIV Diagnosed Persons in Los Angeles County, 2010–2016

Characteristic	Observed Assortativity Coefficient*	95% CI of Null Assortativity Expectation
ZIP code	0.0249	–0.0019 to 0.0009
Age	0.0704	–0.0052 to 0.0041
Health district	0.0916	–0.0042 to 0.0036
Transmission risk	0.1546	–0.0091 to 0.0017
Service planning area	0.1519	–0.0076 to 0.0060
Race/ethnicity	0.2677	–0.0089 to –0.0063

*$P < .001$ for all characteristics.

When examining the assortativity of individual Health Districts in the genetic transmission network, only 6 (23%) Health Districts had an assortativity coefficient >0.10, and the maximum assortativity coefficient for a Health District was <0.20 (Figure 2B). The 2 Health Districts with the highest assortativity coefficients are Long Beach and Hollywood-Wilshire. A report based on 2018 data found that Hollywood-Wilshire is ranked 2 out of 26 for the rate of new HIV diagnoses in LAC (42 HIV diagnoses per 100 000 population) [26], and 94% of people recently diagnosed with HIV in Hollywood-Wilshire are male. The second most assortative Health District is Long Beach, which ranks 5 out of 26 (tied with East Valley and Compton) for the rate of new HIV diagnoses in LAC (25 HIV diagnoses per 100 000 population), and 86% of people recently diagnosed with HIV in Long Beach are male [26]. The Health District with the lowest assortativity is Northeast. It ranks 12 out of 26 for the rate of new HIV diagnoses (16 HIV diagnoses per 100 000

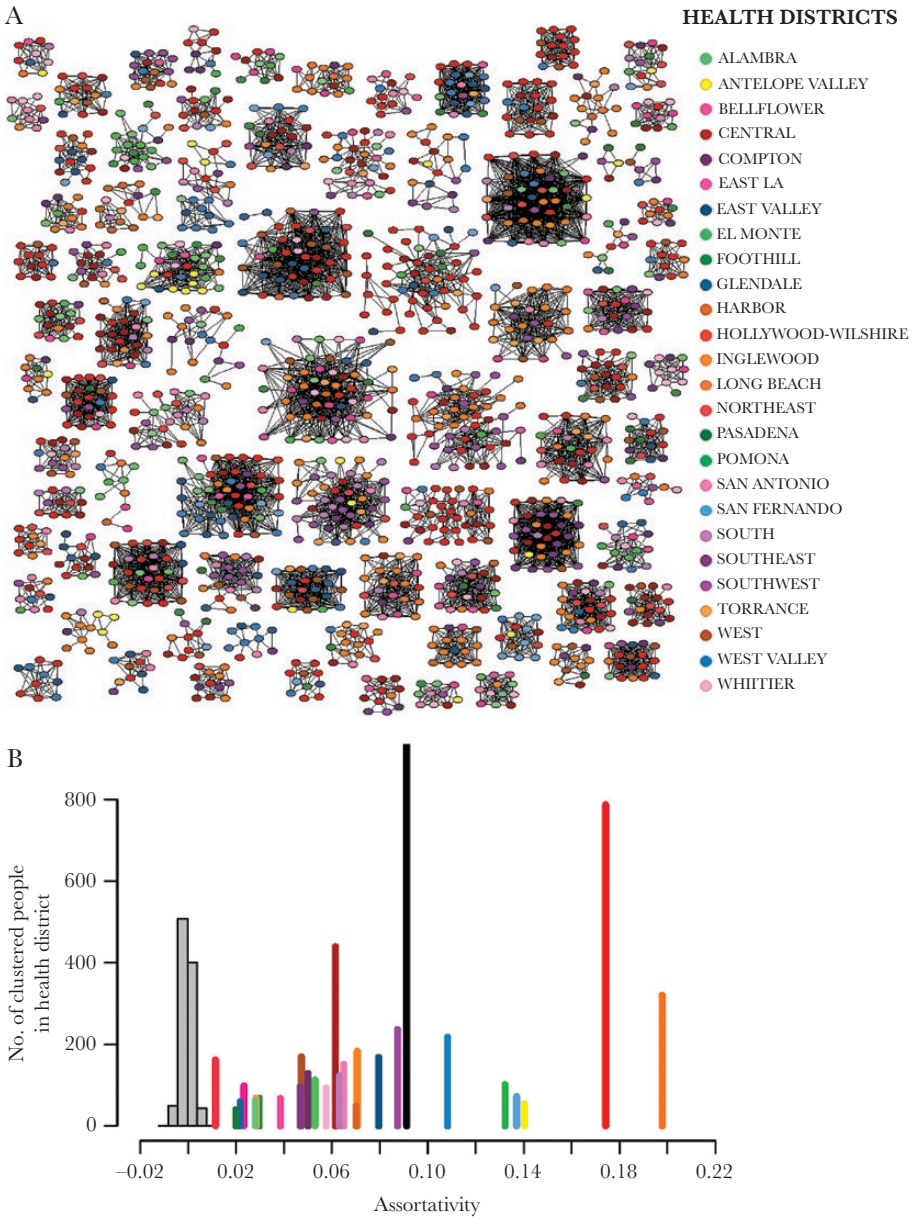

Figure 2. Assortativity of the Los Angeles County molecular transmission network by Health District residence at diagnosis. A, Clusters comprising ≥10 people. Color indicates Health District. Edges denote viral sequences ≤0.015 substitutions/site divergent. B, Assortativity of each Health District. The height of the bars indicates the number of individuals with a reported HIV genetic sequence from each Health District included in our analysis. The gray bars indicate the null expectation (from 1000 permutations) if individuals were sorting at random.

population), with 98% of people recently diagnosed with HIV in the Northeast Health District being male [26]. The highest-prevalence (Central) and lowest-prevalence (Alhambra) Health Districts were both unremarkable with regards to assortativity (Figure 2B).

Time–Space Linkage

To determine the relationship between HIV molecular transmission clusters and temporal and spatial patterns in HIV diagnoses, we counted concordant time–space pairs (ie, the proportion of genetically linked pairs from the same geographic region and/or diagnosis year). Between 2010 and 2016, 2328 people (56% of 4150 linked in the network at 0.015 substitutions/site) were diagnosed in the same year as those genetically linked to them. Over that same time period, 696 people (17%) were linked to other newly diagnosed cases in the same ZIP code, 1802 (44%) people in the same Health District, and 2644 (64%) people in the same Service Planning Area. Notably, few people were genetically linked to an individual who overlapped in both space and time. Only 275 people (7%) were linked to a

person with a diagnosis in the same year and ZIP code. As expected, this overlap was greater for larger geographical areas: 773 (19%) were linked in the same diagnosis year and Health District, and 1255 (30%) were linked in the same diagnosis year and Service Planning Area. This overlap was markedly smaller when we considered shorter time intervals (quarterly rather than yearly): 134 (5%) for ZIP code, 337 (8%) for Health District, and 557 (14%) for Service Planning Area. Similar patterns were seen for more conservative genetic distance thresholds of 0.010 and 0.005 substitutions/site (Table 3).

Genetically linked pairs that were diagnosed in the same time and space were more likely to include at least 1 PWID than would be expected by chance. Of the 773 genetically linked pairs (at the 0.015 substitutions/site threshold) diagnosed in the same Health District in the same year, 35 (4.5%) pairs included at least 1 PWID (Supplementary Table 2). This observed number of pairs with a PWID is 1.8 times larger than would be expected based on permutation analysis ($P = .001$). Genetically linked pairs in the same Service Planning Area diagnosed in the same year (n = 71; 5.6%) were 1.4 times more frequent than expected ($P = .009$). We did not, however, observe an excess of genetically linked time–space pairs in which both members were PWID, likely owing to their relative rarity and likely underreporting of PWID in the data set: 2 pairs in Health Districts and 3 pairs in Service Planning Areas.

Multinetwork Jaccard Analysis

To identify characteristics associated with overlap between molecular clusters and geographic clusters, we conducted generalized linear regression analysis with the Jaccard coefficient as the outcome (ie, the union of molecular clusters and geographic clusters divided by the intersection of molecular clusters and geographic clusters). We found that the Jaccard coefficient differed by individual demographic and HIV risk characteristics. Greater geographic similarity within clusters (ie, higher Jaccard coefficient) was associated with younger age at diagnosis (13–19 years vs 20–29 years; adjusted odds ratio [aOR], 1.21; 95% CI, 1.04–1.40), cisgender men who do not report sex with men (vs MSM; aOR, 1.22; 95% CI, 1.11–1.34), and $CD4^+$ <200 cells/mm^3 at diagnosis (vs ≥500 cells/mm^3; aOR, 1.16; 95% CI, 1.04–1.30). Less geographic similarity within clusters (lower Jaccard coefficient) was associated with older at diagnosis (30–39 years vs 20–29 years; aOR, 0.90; 95% CI, 0.18–0.97), transgender women (vs MSM; aOR, 0.60; 95% CI, 0.48–0.58), Black/African American (vs Hispanic; aOR, 0.84; 95% CI, 0.77–0.92), and foreign born (vs US born; aOR, 0.90; 95% CI, 0.82–0.98) (Table 4). Similar results were found for more conservative genetic distance thresholds (Supplementary Tables 3 and 4).

DISCUSSION

HIV molecular transmission clusters and spatial–temporal patterns in HIV diagnoses represent 2 complementary approaches to understanding HIV transmission patterns. In Los Angeles County, we found statistically significant but weak associations between these 2 approaches. Overlap between geographic region of residency at HIV diagnosis and HIV transmission clusters increased as the size of the geographic region increased (ie it was higher for Service Planning Areas than ZIP codes). The genetic transmission network is modestly assortative for these large geographic regions, but this assortativity is weaker than has been reported for race/ethnicity and transmission risk [27]. These findings indicate that genetic clustering may be a better indicator of HIV transmission patterns than time–space clusters of diagnosis in urban HIV-endemic areas. Time–space

Table 3. Relationship Between HIV Molecular Transmission Clusters and Temporal and Spatial Patterns in HIV Diagnoses

	TN93 ≤0.015	TN93 ≤0.010	TN93 ≤0.005
Level of Association	No. (%)	No. (%)	No. (%)
All			
No. of linked nodes in network	4150	3224	2109
Geography-only			
No. of linked nodes in same ZIP code	696 (17)	499 (15)	306 (15)
No. of linked nodes in same HD	1802 (44)	1324 (41)	786 (37)
No. of linked nodes in same SPA	2644 (64)	1978 (61)	1181 (56)
Yearly			
No. of linked nodes in same year	2328 (56)	1819 (56)	1150 (55)
No. of linked nodes in same year & ZIP code	275 (7)	216 (8)	155 (9)
No. of linked nodes in same year & HD	773 (19)	583 (18)	382 (18)
No. of linked nodes in same year & SPA	1255 (30)	976 (30)	594 (28)
Quarterly			
No. of linked nodes in same quarter	1202 (29)	883 (27)	556 (26)
No. of linked nodes in same quarter & ZIP code	134 (5)	108 (6)	92 (10)
No. of linked nodes in same quarter & HD	337 (8)	274 (9)	205 (10)
No. of linked nodes in same quarter & SPA	557 (14)	438 (14)	295 (14)

Abbreviations: HD, Health District; SPA, Service Planning Area; TN93, Tamura-Nei 93 genetic distance (substitutions/site); ZIP, Zone Improvement Plan.

Table 4. Multivariate Generalized Linear Model Assessing Relationship Between Individual Characteristics and the Jaccard Coefficient, at the Level of Health District

Characteristic	Adjusted Odds Ratio	95% CI	*P* Value
Age			
13–19 y	1.208	1.043 to 1.399	**.011**
20–29 y	Ref.	–	-
30–39 y	0.896	0.182 to 0.966	**.004**
40–49 y	0.901	0.809 to 1.003	.056
50+ y	–0.06	0.808 to 1.087	.394
Transmission risk/gender			
Cisgender men[a]	1.222	1.111 to 1.342	**<.001**
Cisgender women	1.049	0.861 to 1.279	.633
Transgender women	0.602	0.479 to 0.575	**<.001**
MSM	Ref.	–	-
PWID	0.972	0.812 to 1.165	.760
Race/ethnicity			
White	0.964	0.872 to 1.065	.466
African American	0.837	0.765 to 0.915	**<.001**
Hispanic/Latino	Ref.	–	-
Other	0.874	0.789 to 0.968	**.010**
Country of birth			
US born/territory	Ref.	–	-
Foreign-born	0.899	0.821 to 0.984	**.020**
Unknown	0.958	0.879 to 1.043	.324
$CD4^+$ at diagnosis			
<200	1.162	1.037 to 1.302	**.010**
200–499	1.081	0.996 to 1.174	.061
≥500	Ref.	–	-
Unknown	1.106	1.006 to 1.215	**<.001**

Bold values indicate *P* value <.05. Model includes all variables present in the table.

Abbreviations: MSM, men who have sex with men; PWID, people who inject drugs.

[a]Cisgender men who do not report sex with men.

clustering may be a better indicator of changing transmission dynamics in larger areas, such as at the county or state level, or in nonendemic areas where the HIV burden is lower [28].

Within LAC, greater geographic similarity within clusters was associated with younger age at diagnosis, cisgender men, and having a $CD4^+$ count <200 cells/mm^3 at diagnosis. In contrast, less geographic similarity within clusters was associated with older age at diagnosis, transgender women, Blacks/African Americans, and being born outside of the United States. The greater geographic similarity among younger individuals and cisgender men might be the result of more frequent geospatial app use among these groups, as younger individuals (<21 years of age) are not permitted in venues that serve alcohol and thus seek partners closer to their home since they have lower mobility [29]. Likewise, we observed less geographic similarity among those aged 30–39, which may represent higher income and therefore increased access to mobility, although we were not able to assess this relationship with our data. Interestingly, a previous analysis using the same data in LAC found that transgender women were more likely to cluster than other risk groups and were more likely than expected to cluster with cisgender men [30]. Our findings show that transgender women, who themselves are highly assortative in the genetic network, tend to reside in geographic areas that are different from their genetically linked partners, indicating that clusters that include transgender women are more geographically diverse. These findings illustrate how higher-risk minority groups may be geographically overdispersed. Consequently, time–space approaches for identifying trends in transmission and diagnosis within these groups may not be productive. On the other hand, our finding of an excess of PWID in genetically linked time–space pairs suggests that a geographic approach may be relevant to HIV prevention and response efforts among this risk group.

In the United States, the relative lack of clustering by geography was also found in Chicago (Cook County) [31]. The Chicago study used Kulldorff's spatial scan statistic on 920 *pol* sequences sampled between 2008 and 2011 in Cook County and did not find any significant geographic groups, even after limiting to cluster sizes of >5. An earlier analysis in Mississippi of 799 *pol* sequences sampled between 2005 and 2008 also found that clusters that included Black MSM (their population of interest) were geographically heterogeneous. However, growing clusters were found to be assortative by Field Services Region in North Carolina; Field Services Regions are larger geographic

areas that group counties across the state for public health response [32]. The accumulation of these results indicates that small-scale geographic boundaries may not be useful for the direction of public health resources.

Our analyses and conclusions are restricted to only those individuals with an HIV diagnosis and a viral resistance genotype reported to the LAC Department of Public Health. This type of comparison is only feasible for individuals with a reported viral sequence. Although genotype reporting completeness does vary by geography, undersampling is not expected to strongly bias assortativity estimates in the genetic network [27]. Notably, HIV infection and diagnosis may occur in different time periods due to infrequent testing. Future analyses that incorporate social network data could enhance the completeness of the overall network by incorporating both HIV-uninfected individuals and people with HIV who do not have a reported viral genotype. These data may also give us a better understanding of transmission patterns and identify those at risk of HIV infection who are eligible for pre-exposure prophylaxis (PrEP) and people with HIV who could benefit from linkage to care or other support services.

In conclusion, the low magnitude of the associations found between HIV transmission clusters and geographic residence at diagnosis indicate that within an urban setting with endemic HIV, genetic clustering may serve as a better indicator of HIV transmission patterns than time–space clustering to inform public health action.

Supplementary Data

Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

Financial support. This work was supported by the San Diego Center for AIDS Research (SD CFAR), an National Institutes of Health (NIH)-funded program (P30 AI036214). B.S. was also funded by an NIH-NIDA K01 (DA049665). M.R.C. acknowledges joint center funding from the UK Medical Research Council and Department for International Development (MR/R015600/1). J.O.W. was funded by a California HIV/AIDS Research Program IDEA Award (ID15-SD-052), an NIH-NIAID K01 Career Development Award (K01AI110181), and an NIH-NIAID R01 (AI135992).

Potential conflicts of interest. B.S., M.R.C., K.P., Z.S., and Y.W.H. have no conflicts of interest to report. J.O.W. has additional funding from the CDC through contracts to his institution. 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.

Patient consent. The study was approved by the University of California (UC) San Diego and LAC DPH Institutional Review Boards. All data were collected through standard surveillance protocols, and the study therefore did not necessitate written patient consent. Per the terms of a Data Use Agreement between LAC DPH and UC San Diego, all data were deidentified before sharing with UC San Diego researchers. To further preserve individual privacy, only month and year for the date of diagnosis and lab tests were shared.

Prior presentation. This information has been presented in part at the 27th Conference on Retroviruses and Opportunistic Infections and the EPIDEMICS[7] conference.

References

1. Mentasti M, Afshar B, Collins S, et al. Rapid investigation of cases and clusters of Legionnaires' disease in England and Wales using direct molecular typing. J Med Microbiol **2016**; 65:484–93.
2. Murray M, Nardell E. Molecular epidemiology of tuberculosis: achievements and challenges to current knowledge. Bull World Health Organ **2002**; 80:477–82.
3. Chudy M, Budek I, Keller-Stanislawski B, et al. A new cluster of hepatitis A infection in hemophiliacs traced to a contaminated plasma pool. J Med Virol **1999**; 57:91–9.
4. Yirrell DL, Robertson P, Goldberg DJ, et al. Molecular investigation into outbreak of HIV in a Scottish prison. BMJ **1997**; 314:1446–50.
5. Eybpoosh S, Haghdoost AA, Mostafavi E, et al. Molecular epidemiology of infectious diseases. Electron Physician **2017**; 9:5149–58.
6. Vineis P. Commentary: First steps in molecular epidemiology: Lower *et al*. 1979. Int J Epidemiol **2007**; 36:20–22.
7. Grenfell BT, Pybus O, Gog J, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science **2004;** 303:327–32.
8. Honardoost M, Rajabpour A, Vakili L. Molecular epidemiology; New but impressive. Med J Islam Repub Iran **2018**; 32:53.
9. Schulte P, Perera F. Molecular Epidemiology: Principles and Practice. San Diego, CA: Academic Press Inc; **1993**.
10. Oster AM, Wertheim JO, Hernandez AL, et al. Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. J Acquir Immune Defic Syndr **2015**; 70:444–51.
11. Golden MR, Lechtenberg R, Glick SN, et al. Outbreak of human immunodeficiency virus infection among heterosexual persons who are living homeless and inject drugs — Seattle, Washington, 2018. MMWR Morbid Mortal Wkly Rep **2018**; 68:334–49.
12. Alpren C, Dawson EL, John B, et al. Opioid use fueling HIV transmission in an urban setting: an outbreak of HIV infection among people who inject drugs-Massachusetts, 2015–2018. Am J Public Health **2020**; 110:37–44.
13. Centers for Disease Control and Prevention. HIV in the United States: at a glance. Available at: https://www.cdc.gov/hiv/statistics/overview/ataglance.html. Accessed 7 September 2018.
14. Fauci AS, Redfield RR, Sigounas G, et al. Ending the HIV epidemic: a plan for the United States. JAMA **2019**; 321:844–5.
15. Centers for Disease Control and Prevention. Detecting and responding to HIV transmission clusters: a guide for health departments. Draft version 2.0 ed. 2018. Available at: https://www.cdc.gov/hiv/pdf/funding/announcements/ps18-1802/CDC-HIV-PS18-1802-AttachmentE-Detecting-Investigating-and-Responding-to-HIV-Transmission-Clusters.pdf. Accessed 11 October 2019.
16. Fitzmaurice AG, Linley L, Zhang C, et al. Novel method for rapid detection of spatiotemporal HIV clusters potentially warranting intervention. Emerg Infect Dis **2019**; 25:988–91.
17. United States Census Bureau. QuickFacts: Los Angeles County, California; California. Available at: https://www.census.gov/quickfacts/fact/table/losangelescountycalifornia,CA/PST045218. Accessed 11 October 2019.
18. US Census Bureau. Annual estimates of the resident population for incorporated places of 50 000 or more, ranked by July 1, 2018 population: April 1, 2010 to July 1, 2018. Available at: https://factfinder.census.gov/. Accessed 13 October 2019.
19. County of Los Angeles. Statistics. Available at: https://lacounty.gov/government/geography-statistics/statistics/. Accessed 13 October 2019.
20. Los Angeles County Department of Public Health. 2017 annual HIV surveillance report. **2018**. http://publichealth. lacounty.gov/dhsp/Reports/HIV/2019Annual_HIV_Surveillance_Report_08202020_Final_revised_Sept2020.pdf. Accessed 11 October 2019.
21. Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRAnsmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. Mol Biol Evol **2018**; 35:1812–9.
22. Wertheim JO, Kosakovsky Pond SL, Forgione LA, et al. Social and genetic networks of HIV-1 transmission in New York City. PLoS Pathog **2017**; 13:e1006000.
23. Newman MEJ. Mixing patterns in networks. Physical Review E **2003**; 67:026126.
24. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal **2006**. Complex Systems, 1695. https://igraph.org
25. Location Intelligence Platform. https://carto.com/
26. Division of HIV and STD Programs, Department of Public Health, County of Los Angeles. HIV Surveillance Annual Report, 2019. Published May 2020. http://publichealth.lacounty.gov/dhsp/Reports.htm. Accessed 11 October 2019.

27. Ragonnet-Cronin M, Benbow N, Hayford C, et al. Sorting by race/ethnicity across HIV genetic transmission networks in three major metropolitan areas in the United States. AIDS Res Hum Retroviruses **2021**; doi:10.1089/AID.2020.0145
28. Gonsalves GS, Crawford FW. Dynamics of the HIV outbreak and response in Scott County, IN, USA, 2011-15: a modelling study. Lancet HIV **2018**; 5:e569–77.
29. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe right: dating website and app use among men who have sex with men. AIDS Behav **2018**; 22:1265–72.
30. Ragonnet-Cronin M, Hu YW, Morris SR, et al. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. Lancet HIV **2019**; 6: e164–72.
31. Lubelchek RJ, Hoehnen SC, Hotton AL, et al. Transmission clustering among newly diagnosed HIV patients in Chicago, 2008 to 2011: using phylogenetics to expand knowledge of regional HIV transmission patterns. J Acquir Immune Defic Syndr **2015**; 68:46–54.
32. Dennis AM, Hué S, Billock R, et al. Human immunodeficiency virus type 1 phylodynamics to detect and characterize active transmission clusters in North Carolina. J Infect Dis **2020**; 221:1321–30.