

The Relatedness of HIV Epidemics in the United States–Mexico Border Region

Sanjay R. Mehta,¹ Wayne Delport,¹ Kimberly C. Brouwer,¹ Stephen Espitia,¹ Thomas Patterson,¹ Sergei Kosakovsky Pond,¹ Steffanie A. Strathdee,¹ and Davey M. Smith^{1,2}

Abstract

Phylogeography can improve the understanding of local and worldwide HIV epidemics, including the migration of subepidemics across national borders. We analyzed HIV-1 sequences sampled from Mexico and San Diego, California to determine the relatedness of these epidemics. We sampled the HIV epidemics in (1) Mexico by downloading all publicly available HIV-1 *pol* sequences from antiretroviral-naive individuals in GenBank ($n = 100$) and generating similar sequences from cohorts of injection drug users and female sex workers in Tijuana, Mexico ($n = 27$) and (2) in San Diego, California by *pol* sequencing well-characterized primary ($n = 395$) and chronic ($n = 267$) HIV infection cohorts. Estimates of population structure (F_{ST}), genetic distance cluster analysis, and a cladistic measure of migration events (Slatkin–Maddison test) were used to assess the relatedness of the epidemics. Both a test of population differentiation ($F_{ST} = 0.06$; $p < 0.01$) and a cladistic estimate of migration events (84 migrations, $p < 0.01$) indicated that the Tijuana and San Diego epidemics were not freely mixing. A conservative cluster analysis identified 72 clusters (two or more sequences), with two clusters containing both Mexican and San Diego sequences (permutation $p < 0.01$). Analysis of this very large dataset of HIV-1 sequences suggested that the HIV-1 epidemics in San Diego, California and Tijuana, Mexico are distinct. Larger epidemiological studies are needed to quantify the magnitude and associations of cross-border mixing.

Introduction

THE DEVELOPMENT OF PHYLOGENETIC and phylogeographic methodology has given researchers, clinicians, and epidemiologists a greater understanding of the worldwide HIV-1 epidemic, including the realization that it is actually a collection of subepidemics.¹ Understanding the epidemiology and transmission dynamics of these subepidemics may be critical for vaccine development and other interventions aimed at reducing HIV transmission.

Recent investigations of the spread of the HIV-1 subtype B epidemic in Europe and the emergence of HIV-1 in east Africa have applied phylogeographic approaches to better understand HIV transmission dynamics.^{2,3} For example, Gray *et al.* in 2009⁴ evaluated HIV-1 sequences of p24 (gag protein) and gp41 (envelope glycoprotein) coding regions in clades A and D in relation to geographic data and found that although the three major urban centers of the Democratic Republic of Congo (Kinshasa, Lubumbashi, and Kisangani) are located in the same country, the HIV epidemic in each city had evolved separately and was more closely related to nearby cities in

other countries that share transportation networks. These accessibility networks, or networks within which individuals can easily access other individuals or services, correlated better with phylogenetic HIV structure than structural risk factors such as migration, population growth, and armed conflict.^{4,5} In contrast, Paraskevis *et al.* in 2009³ analyzed 1337 *pol* sequences sampled from across Europe using a cladistic measure of gene flow and concluded that countries acted either as HIV-1 sources or sinks or had significant bidirectional migration patterns. Based on phylogeographic analysis, they proposed that intervention strategies targeting tourists, travelers, and migrants would be the key to decreasing the spread of HIV across Europe.³

The border crossing between San Diego, California and Tijuana, Baja California Mexico is the busiest in the world with 5–6 million registered crossings northward monthly,⁶ while easy access to street drugs and a thriving red light district in Tijuana (>9000 registered female sex workers) attract growing numbers of U.S. nationals southward.^{7–10} The prevalence of HIV-1 in Tijuana was estimated at 0.54% in 2006,¹¹ which is higher than both the CDC estimated

¹University of California San Diego, San Diego, California.

²Veterans Affairs Medical Center, San Diego, California.

prevalence in U.S. adolescents and adults in 2006 of 0.45%,¹² and national estimates of HIV prevalence in Mexico of 0.3% in 2006.⁸ The higher HIV prevalence in Tijuana is characteristic of cities with a large migrant population that tends to engage in more risky behaviors.¹³ To better understand the relatedness of the HIV epidemics in San Diego and Mexico, we studied the epidemics on both sides of the border using a phylogeographic approach.

Materials and Methods

Study population

We obtained samples from the epidemic in (1) Mexico by downloading all publicly available HIV-1 *pol* sequences in GenBank on September 27, 2009 ($n = 100$ between years 2003 and 2005) and generating similar sequences from cohorts of injection drug users (IDU) and female sex workers (FSW) in Tijuana ($n = 27$ between years 2004 and 2006) and (2) in San Diego by *pol* sequencing of well-characterized primary ($n = 395$ between years 1996 and 2006) and chronic HIV infection cohorts ($n = 267$ between years 2006 and 2007). Given the published descriptions of the sampled cohorts, all sampled sequences were believed to originate from antiretroviral-naïve individuals except for 42 Mexican sequences downloaded from GenBank. HIV risk factors and location of sampling were unavailable for the majority of Mexican sequences downloaded from GenBank. Location of sampling was identified to be northeast Mexico for 32 sequences, and three sequences were noted to be from FSW and eight from IDU.

Viral sequencing

The *pol* region of HIV-1 was sequenced from all of the samples from the Tijuana cohorts by published methods using the Viroseq (Abbott, Chicago, IL) platform.¹⁴ Sequences were deposited in GenBank with accession numbers GU304608–GU304632.

Resistance profiling

All sequences from subtype B antiretroviral treatment (ART)-naïve individuals ($n = 741$) were uploaded into the Calibrated Population Tool on the Stanford University HIV Drug Resistance Database (accessed April 20, 2010) and resistance analysis of the sequences was performed.¹⁵

Phylogenetic analysis

Subtyping of each sequence was performed using Subtype Classification Using Evolutionary Algorithms (SCUEAL).¹⁶ All non-B sequences in the sample were excluded in the subsequent phylogeographic analyses. Sequences were aligned using HIV-specific alignment tools in HyPhy¹⁷ (sequence length = 453 nucleotides). A maximum likelihood tree was estimated using a TN93 model¹⁸ of sequence evolution using PhyML,¹⁹ and the among-site rate variation approximated with a 4 class gamma distribution. We further analyzed the relatedness of the epidemics in San Diego and Mexico (including Mexican sequences from GenBank and our sequences from Tijuana) using (1) estimates of population structure (F_{st} -statistics²⁰), (2) the identification of clusters based on the synonymous codon-based maximum likelihood

MG94xREV model²¹ with genetic distances of $< 0.5\%$, and (3) cladistic measures of migration events.²² A subanalysis was then performed comparing sequences from San Diego and Tijuana alone. All analyses were performed using the HyPhy Package.²³ The alignment and tree in NEXUS format can be downloaded from <http://www.hyphy.org/pubs/Mehta2009/>.

Results

From the naïve group, a total of 92 sequences (12.4%) had a resistance-associated mutation to a nucleoside reverse transcriptase inhibitor and 12.3% ($n = 91$) to a nonnucleoside reverse transcriptase inhibitor, similar to reported rates of transmitted drug resistance in the region.^{24–26} Notably, no baseline resistance-associated mutations were observed among the sequences from Tijuana.

Most of the sequences analyzed were subtype B (773/789), including all Mexican sequences; 16 of the San Diego sequences were non-subtype B including three subtype A, five subtype C, two unclassified strains, and six unique recombinant forms. Once the non-subtype B sequences were excluded, the population structure of the remaining sequences was assessed using measures of population differentiation. Mean pairwise synonymous codon-based genetic distance within the San Diego and Mexico regions was 4.54%, and 4.55% between regions. A simple nucleotide-based distance measure (TN93²⁷) yielded larger numbers: 5.6% and 6.1%, respectively. When only the sequences from San Diego (662) and Tijuana (27) were analyzed, the within-region synonymous diversity was 4.55% (TN93 diversity 5.6%), while the between-region diversity was 4.93% (TN93 6.1%). A measure of population differentiation as inferred from F_{st} statistics²⁸ was sensitive to the choice of genetic distance. The test using nucleotide TN93 distance indicated a significantly different population structure between all sequences from Mexico and San Diego ($F_{ST} = 0.07$, $p < 0.01$) and between Tijuana and San Diego ($F_{ST} = 0.08$, $p < 0.01$). However, when using codon-based synonymous distances, the Mexico vs. San Diego comparison was not significant ($F_{ST} = 0.01$, $p = 0.62$). For the San Diego vs. Tijuana, the population remained structured ($F_{ST} = 0.08$, $p = 0.02$), suggesting the epidemics are not freely mixing.

A cladistic measure of gene flow²⁹ also indicated that the Mexico and San Diego epidemics were not freely mixing, but nonetheless suggested that a number of cross-border transmissions have occurred (84 inferred migrations, $p < 0.01$). When the Tijuana and San Diego epidemics were compared, they were also found to be not freely mixing with only 14 inferred migrations ($p < 0.01$).

Finally, a very conservative clustering analysis (clusters defined as sequences with a synonymous genetic distance $< 0.5\%$ ³⁰) identified 72 clusters with ≥ 2 sequences (range 2 to 35 sequences), but only two contained sequences derived from both Mexico and San Diego; this degree of separation is statistically significant ($p < 0.01$ permutation test).

Discussion

Similar to previous work, we found that most Mexican sequences were subtype B and that transmitted drug resistance was present, although surprisingly not from our

sequences from Tijuana.³¹ We then applied a phylogeographic approach to the analysis of 773 subtype B HIV-1 *pol* sequences derived from 127 Mexican and 646 San Diego subjects.

To analyze the relatedness of the epidemics, we utilized synonymous distances as they are preferred for clustering analysis because they mitigate the confounding effect of convergent evolution due to immune and drug resistance selection. Furthermore, our estimation of genetic distance was performed using maximum likelihood methods, which can better handle the numerous ambiguous nucleotides present in bulk HIV sequences. We found that the San Diego and Mexican epidemics were distinct using cladistic measures of gene flow (Slatkin-Maddison test) and clustering analysis, but when using F_{ST} statistics, a more conservative distance-based approach, the Mexican and San Diego populations were not found to be phylogenetically distinct.

If the San Diego and Mexican epidemics were not distinct, the close proximity of San Diego and Tijuana would suggest that the subepidemics of these two cities would be freely mixing as well; however, when the population structure of the Tijuana and San Diego sequences was analyzed, these subepidemics were found to be phylogenetically distinct by all measures. Therefore these data suggest that even with the high levels of cross-border migration of people between San Diego and Tijuana, the HIV epidemics in each city have remained distinct, although interestingly, this was not demonstrated for the sampled HIV epidemics between San Diego and Mexico as a whole when using F_{ST} statistics. Potential explanations for this are that either the networks of HIV-positive individuals in San Diego have closer ties to other parts of Mexico than Tijuana, or that the lack of mixing observed between the Tijuana and San Diego epidemics was due to differences in the HIV risk factors of the sampled individuals.

Early evidence suggested that the Mexican HIV epidemic had its roots in the U.S. epidemic, as 100% of the HIV cases registered in Mexico in 1983 were from people who had lived in the United States.^{32,33} Today, although Mexico has the third highest absolute number of HIV cases in the Americas after the United States and Brazil, Mexico has only the twenty-third highest prevalence in the Americas.⁸ However, one-third of these infections are from individuals living in border states.³⁴ Pertinent to San Diego, the neighboring Mexican state of Baja California has the second highest cumulative HIV incidence in Mexico, second only to the federal district (Mexico City),⁸ with recent research estimating that 1 in 116 Tijuana residents between the ages of 15 and 49 are HIV infected.³⁵ Taken together, there is ample evidence that a separate more dynamic Mexican HIV subepidemic is occurring along its northern border with the United States.³⁶

The net flux of Mexicans migrating into the United States annually surpasses 400,000.³⁷ Migrants, who are separated from family and other supports,^{38,39} tend to engage in higher HIV risk behaviors,^{40,41} and those returning may carry the virus back home. In a study of male IDUs in Tijuana, those who had been deported from the United States to Mexico were four times more likely to be HIV positive than other male IDUs.⁴² In a study of FSW in Tijuana and Ciudad Juarez, two-thirds had clients from the United States, and those who did so were more likely to engage in other higher risk sexual behaviors, have active syphilis, and to inject drugs.^{43,44} More

than one-third of male IDUs in Tijuana had homosexual encounters, but these individuals tended not to self-identify as men who have sex with men (MSM) and had larger numbers of female partners than non-MSM.⁴⁵ Finally, highlighting the bidirectionality of social networks across borders, nearly 75% of San Diego MSMs and nearly 50% of MSMs from Tijuana reported partners from across the border.⁴⁶ Although many of the factors presented above suggest that U.S. and Mexican epidemics, and in particular the San Diego and Tijuana subepidemics, should be closely related, our data suggest something different. Our work shows that even with the tremendous amount of cross-border migration, the social networks of the majority of HIV-positive individuals in San Diego are not mixing with the social networks of HIV-positive individuals in Tijuana.

Limitations to this study include the possibility that sampling error affected the findings. For example, the bulk of the sequences from San Diego in this study came from MSM, whereas the bulk of the sequences from Tijuana came from cohorts of FSW and IDUs,^{47,48} and the HIV risk factors for non-Tijuana Mexican sequences was unknown. It is possible, therefore, that the observed separation in HIV transmission networks was because the networks sampled were distinct by HIV risk factor and not because of a national border. If more sequences from IDUs and non-MSM Hispanic men were obtained from San Diego County, more sequences from throughout Mexico were obtained with HIV risk factor data, or more sequences from self-identified MSMs were obtained from Tijuana, we might find that the epidemics are more interconnected.

Further efforts to better elucidate the relationships of the San Diego and Tijuana HIV subepidemics will need to focus on improving sampling of self-identified MSM social networks south of the border and Hispanic heterosexual and IDU networks north of the border. Future efforts to control the transmission of HIV in the border area may benefit from understanding how these transmission networks are related, so that suitable interventions can be targeted appropriately.

Acknowledgments

We thank Jean Carr for use of her previously published *pol* sequences from Mexico. This project was supported by Grants 5K01DA020364-05 (KCB), R01MHMH65849 (SAS), AI69432 (ACTG), AI080353 (R21), MH62512 (HNRC), MH083552 (Clade), AI077304 (Dual Infection), AI36214 (UCSD Center for AIDS Research), AI047745 (Dynamics), AI074621 (Transmission), AI07384 (AIDS Training Grant), AI080193 (ARRA) from the National Institutes of Health, DMS0714991 (NSF), the International AIDS Vaccine Initiative, and the California HIV/AIDS Research Program RN07-SD-702 (DMS).

Author Disclosure Statement

No competing financial interests exist.

References

1. Pieniazek D, Janini LM, Ramos A, *et al.*: HIV-1 patients may harbor viruses of different phylogenetic subtypes: Implications for the evolution of the HIV/AIDS pandemic. *Emerg Infect Dis* 1995;1(3):86–88.

2. Gray RR, Tatem AJ, Lamers S, *et al.*: Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS* 2009;23(14):F9–F17.
3. Paraskevis D, Pybus O, Magiorkinis G, *et al.*: Tracing the HIV-1 subtype B mobility in Europe: A phylogeographic approach. *Retrovirology* 2009;6:49.
4. Gray RR, Tatem AJ, Lamers S, *et al.*: Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS* 2009;23(14):F9–F17.
5. Salathe M, Jones JH: Dynamics and control of diseases in networks with Community Structure. *PLoS Comput Biol*. 2010 Apr 8; 6(4):e1000736.
6. US DOT: Transportation Capacity Building Training Program. US DOT website, 11-13-2009.
7. Ojeda VD, Strathdee SA, Lozada R, *et al.*: Associations between migrant status and sexually transmitted infections among female sex workers in Tijuana, Mexico. *Sex Transm Infect* 2009;85(6):420–426.
8. Secretaria De Salud: Consejo Nacional para la Prevencion y Control del VIH/SIDA. Panorama epidemiologico del VIH/SIDA e ITS en Mexico. <http://www.salud.gob.mx/conasida/estadis.htm>, 2007. 2-3-2008.
9. Bucardo J, Semple SJ, Fraga-Vallejo M, Davila W, and Patterson TL: A qualitative exploration of female sex work in Tijuana, Mexico. *Arch Sex Behav* 2004;33(4):343–351.
10. Patterson TL, Semple SJ, Fraga M, *et al.*: Comparison of sexual and drug use behaviors between female sex workers in Tijuana and Ciudad Juarez, Mexico. *Subst Use Misuse* 2006;41(10–12):1535–1549.
11. Iniguez-Stevens E, Brouwer KC, Hogg RS, *et al.*: [Estimating the 2006 prevalence of HIV by gender and risk groups in Tijuana, Mexico]. *Gac Med Mex* 2009;145(3):189–195.
12. CDC: HIV/AIDS Surveillance Report 2006. CDC Website, 8-1-2008. 11-23-2009.
13. Borges G, Medina-Mora ME, Orozco R, *et al.*: The Mexican migration to the United States and substance use in northern Mexico. *Addiction* 2009;104(4):603–611.
14. Butler DM, Pacold ME, Jordan PS, Richman DD, and Smith DM: The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. *J Virol Methods* 2009;162(1–2):280–283.
15. Gifford RJ, Liu TF, Rhee SY, *et al.*: The calibrated population resistance tool: Standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* 2009;25(9):1197–1198.
16. Kosakovsky Pond SL, Posada D, Stawiski E, Chappey C, Poon AF, Hughes G, *et al.*: An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comp Biol* 2009;11(Nov 5):e1000581.
17. Pond SL, Frost SD, and Muse SV: HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 2005;21(5):676–679.
18. Tamura K and 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(3):512–526.
19. Guindon S, Lethiec F, Duroux P, and Gascuel O: PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005;33(Web Server issue):W557–W559.
20. Hudson RR, Slatkin M, and Maddison WP: Estimation of levels of gene flow from DNA sequence data. *Genetics* 1992;132(2):583–589.
21. Muse SV: Estimating synonymous and nonsynonymous substitution rates. *Mol Biol Evol* 1996;13(1):105–114.
22. Slatkin M and Maddison WP: A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 1989;123(3):603–613.
23. Pond SL, Frost SD, and Muse SV: HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 2005;21(5):676–679.
24. Little SJ, Holte S, Routy JP, *et al.*: Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002;347(6):385–394.
25. Liu L, May S, Richman DD, *et al.*: Comparison of algorithms that interpret genotypic HIV-1 drug resistance to determine the prevalence of transmitted drug resistance. *AIDS* 2008;22(7):835–839.
26. Smith DM, Wong JK, Shao H, *et al.*: Long-term persistence of transmitted HIV drug resistance in male genital tract secretions: Implications for secondary transmission. *J Infect Dis* 2007;196(3):356–360.
27. Tamura K and 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(3):512–526.
28. Hudson RR, Slatkin M, and Maddison WP: Estimation of levels of gene flow from DNA sequence data. *Genetics* 1992; 132(2):583–589.
29. Slatkin M and Maddison WP: A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 1989; 123(3):603–613.
30. 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(2):225–232.
31. Eyzaguirre L, Brouwer KC, Nadai Y, *et al.*: First molecular surveillance report of HIV type 1 in injecting drug users and female sex workers along the U.S.-Mexico border. *AIDS Res Hum Retroviruses* 2007;23(2):331–334.
32. Valdespino-Gomez JL, Garcia-Garcia ML, del Rio-Zolezzi A, *et al.*: [The epidemiology of AIDS/HIV in Mexico: From 1983 to March 1995]. *Salud Publica Mex* 1995;37(6):556–571.
33. Strathdee SA and Magis-Rodriguez C: Mexico's evolving HIV epidemic. *JAMA* 2008;300(5):571–573.
34. Magis-Rodriguez C, Gayet C, Negroni M, *et al.*: Migration and AIDS in Mexico: An overview based on recent evidence. *J Acquir Immune Defic Syndr* 2004;37(Suppl 4):S215–S226.
35. Iniguez-Stevens E, Brouwer KC, Hogg RS, *et al.*: [Estimating the 2006 prevalence of HIV by gender and risk groups in Tijuana, Mexico]. *Gac Med Mex* 2009;145(3):189–195.
36. Strathdee SA and Magis-Rodriguez C: Mexico's evolving HIV epidemic. *JAMA* 2008;300(5):571–573.
37. Magis-Rodriguez C, Lemp G, Hernandez MT, *et al.*: Going north: Mexican migrants and their vulnerability to HIV. *J Acquir Immune Defic Syndr* 2009;51(Suppl 1):S21–S25.
38. Parrado EA, Flippen CA, and McQuiston C: Use of commercial sex workers among Hispanic migrants in North Carolina: Implications for the spread of HIV. *Perspect Sex Reprod Health* 2004;36(4):150–156.
39. Apostolopoulos Y, Sonmez S, Kronenfeld J, *et al.*: STI/HIV risks for Mexican migrant laborers: Exploratory ethnographies. *J Immigr Minor Health* 2006;8(3):291–302.
40. Magis-Rodriguez C, Gayet C, Negroni M, *et al.*: Migration and AIDS in Mexico: An overview based on recent evidence. *J Acquir Immune Defic Syndr* 2004;37(Suppl 4):S215–S226.
41. Viani RM, Araneta MR, Ruiz-Calderon J, *et al.*: Migration and risk factors for HIV acquisition in pregnant women in Baja California, Mexico. *MedGenMed* 2005;7(2):69.
42. Strathdee SA, Lozada R, Ojeda VD, *et al.*: Differential effects of migration and deportation on HIV infection among male

- and female injection drug users in Tijuana, Mexico. *PLoS One* 2008;3(7):e2690.
43. Ojeda VD, Strathdee SA, Lozada R, *et al.*: Associations between migrant status and sexually transmitted infections among female sex workers in Tijuana, Mexico. *Sex Transm Infect* 2009;85:420–426.
44. Strathdee SA, Lozada R, Semple SJ, *et al.*: Characteristics of female sex workers with US clients in two Mexico–US border cities. *Sex Transm Dis* 2008;35(3):263–268.
45. Deiss RG, Brouwer KC, Loza O, *et al.*: High-risk sexual and drug using behaviors among male injection drug users who have sex with men in 2 Mexico–US border cities. *Sex Transm Dis* 2008;35(3):243–249.
46. Ritieni A, Facer M, Ruiz J, Meneses-Imay MC, Magis-Rodriguez C, and Monitor F: Prevalence of HIV infection and related risk behaviors among young Latino men who have sex with men: San Diego–Tijuana border region. www.dhs.gov, 2006. 2–4–2008.
47. Strathdee SA, Lozada R, Pollini RA, *et al.*: Individual, social, and environmental influences associated with HIV infection among injection drug users in Tijuana, Mexico. *J Acquir Immune Defic Syndr* 2008;47(3):369–376.
48. Strathdee SA, Lozada R, Semple SJ, *et al.*: Characteristics of female sex workers with US clients in two Mexico–US border cities. *Sex Transm Dis* 2008;35(3):263–268.

Address correspondence to:
Sanjay R. Mehta
9500 Gilman Drive MC#0711
La Jolla, California 92093
E-mail: srmehta@ucsd.edu

