Associations Between Phylogenetic Clustering and HLA Profile Among HIV-Infected Individuals in San Diego, California

Sanjay R. Mehta,1 Sergei L. Kosakovsky Pond,1 Jason A. Young,1 Douglas Richman,1,2 Susan Little,1 and Davey M. Smith1,2

1Department of Medicine, University of California, San Diego, and 2Veterans Affairs San Diego Healthcare System, San Diego, California

**Background.** Specific sequence changes of human immunodeficiency virus type 1 (HIV-1) in the presence of specific HLA molecules may alter the composition and processing of viral peptides, leading to immune escape. Persistence of these mutations after transmission may leave the genetic fingerprint of the transmitter’s HLA profile. Here, we evaluated the associations between HLA profiles and the phylogenetic relationships of HIV sequences sampled from a cohort of recently infected individuals in San Diego, California.

**Methods.** We identified transmission clusters within the study cohort, using phylogenetic analysis of sampled HIV pol genotypes at a genetic distance of <1.5%. We then evaluated the association of specific HLA alleles, HLA homozygosity, HLA concordance, race and ethnicity, and mutational patterns within the clustering and nonclustering groups.

**Results.** From 350 cohort participants, we identified 162 clustering individuals and 188 nonclustering individuals. We identified trends for enrichment of particular alleles within individual clusters and evidence of viral escape within those clusters. We also found that discordance of HLA alleles was significantly associated with clustering individuals.

**Conclusions.** Some transmission clusters demonstrate HLA enrichment, and viruses in these HLA-associated clusters often show evidence of escape to enriched alleles. Interestingly, HLA discordance was associated with clustering in our predominantly MSM population.

The CD8+ T-cell response is important in controlling human immunodeficiency virus (HIV) infection, but HIV can evolve to alter the composition or the processing of viral peptides to evade these responses [1]. Specific viral mutations can be matched with particular HLA alleles and provide evidence of immune escape by the virus [2–5]. Since an individual’s HLA alleles determine the diversity of peptides that are recognized [6], individuals who are homozygous at HLA class 1 loci will have a reduced repertoire of peptides that are recognized by cytotoxic T lymphocytes (CTLs). Consistent with this, homozygosity at class I loci has been associated with increased mother-to-child transmission and more rapid progression of HIV [7, 8]. Similarly, HIV transmission that occurs between individuals who have concordant HLA alleles allows the transmission of a virus from the source that may have already adapted to the immune response repertoire that can develop in the recipient, and this scenario has been associated with a more rapid HIV disease progression in the recipient [7, 9–12].

HIV rapidly evolves in response to new selective pressures, resulting in distinct viral populations in each infected individual. The degree of similarity of the viral population between source and recipient partners in the transmission pairs permit networks to be discerned using phylogenetic analysis [13–15]. Therefore, sequence analysis of an individual’s HIV population can
identify footprints in the viral genome that are associated with HLA-induced immune pressure [2, 4, 16, 17], and sequence analysis of a community’s circulating HIV can characterize transmission networks in that location [13, 18, 19]. This study uses both sequence analysis concepts to evaluate the relationships between HLA, potential viral escape, and putative transmission networks in San Diego, California.

**METHODS**

**Study Population**

Subjects were individuals with acute and early HIV infection enrolled in the San Diego Primary Infection Cohort and were recruited using a traditional venue-based approach [20]. Written consent and demographic, HIV risk, and clinical data were obtained at the time of study entry. Duration of infection was estimated using established protocols [21]. Blood samples collected prior to treatment were used to obtain viral loads (COBAS Amplicor, Roche, Basel, Switzerland), CD4 T-cell counts, HIV population–based pol sequence (Virosel 2.0, Abbott, Abbott Park, IL), and HLA data for the HLA-A, HLA-B, HLA-C, and HLA-DRB1 loci as previously described [22].

**Association Between Clustering and HLA Alleles**

Patients were placed in a transmission cluster if the genetic distance of HIV pol between sequences was <1.5%, using a distance matrix based on synonymous nucleotide changes to reduce bias due to convergence, as previously described [13, 23]. First, the HLA profile of each individual was evaluated for the presence of homozygosity. The rates of homozygosity were compared between clustering and nonclustering individuals, using the Fisher exact test. To determine whether there was an association between clustering and specific HLA alleles, we first conservatively applied a Fisher exact test to evaluate whether any cluster was enriched for a particular HLA allele, compared with the nonclustering population. This analysis was performed using the HLA comparison tool on the Los Alamos National Laboratory (LANL) Web site [24], which calculates an exact 2-sided Fisher P value and a Storey q value to correct for the false discovery rate caused by multiple comparisons. Next, we determined the number of concordant HLA alleles between individuals in each cluster and compared the observed rate of HLA concordance to the expected rate on the basis of the prevalence of HLA alleles in our nonclustering population and then repeated the analysis with only the white individuals from our cohort. We then calculated the difference between observed and expected concordance for each cluster and evaluated the observations, using a Wilcoxon matched-pairs signed-ranks test. HLA disequilibrium in the cohort was evaluated using the LANL HLA Disequilibrium tool [25].

**Evaluation of HIV CTL Escape**

To evaluate for CTL escape in identified clusters, we first identified all HLA alleles with a trend toward overrepresentation within a cluster defined by a Fisher P value of <.1 and then mapped all putative CTL epitopes on the HIV pol sequences contained within these clusters [26]. We then used the Immune Epitope Database (IEDB) major histocompatibility complex (MHC) 1 prediction tool [27–31] to evaluate for potential viral CTL escape in the clustered HIV pol sequences. We defined putative CTL escape as present if there was a decrease in the average total processing score of >0.5 and an average increase of 0.5 log_{10} in the MHC binding affinity score. We then took the latest available HIV pol sequence for each individual within that cluster and obtained a combined MHC binding and total processing score for each of those same epitopes across that cluster. Finally, we evaluated transmission pairs with both epidemiologic and phylogenetic linkage for evidence of reversion of escape mutations in recipients’ virus.

**RESULTS**

**Study Population**

Between July 1996 and May 2010, 510 individuals were enrolled in the San Diego Primary Infection Cohort, of whom 350 had HLA testing data. More than 90% of these individuals were male, and sexual exposures with other men was the most reported HIV risk factor, consistent with demographic characteristics of the HIV epidemic in San Diego [32]. The estimated mean duration of HIV infection at the time of sampling among participants was 78 days. By use of a genetic distance of <1.5% to define a cluster, sequences from 162 of these individuals clustered together with another individual, while 188 of them were deemed to be nonclustering (Table 1).

**Relationship Between Clustering and HLA Alleles**

We did not find an association between homozygosity and clustering when comparing all 4 loci together (P = .25) or independently. When we compared our entire clustering population with our nonclustering population, we identified a trend for 2 particular HLA alleles to be enriched within clusters (A32, P = .03; and B57, P = .15) and a trend toward underrepresentation for 3 alleles (A29, P = .07; DRB01, P = .04; and DRB10, P = .06). It should be noted that these univariate analyses were not corrected for multiple comparisons. When clusters (>3 individuals) were individually evaluated for enrichment, 19 alleles were found enriched within the 22 clusters, with P values of <.05; however, after use of the Storey q test to correct for multiple comparisons, none of the alleles had q values of <.01, and only 3 alleles (A23, A32, and B18) were enriched within individual clusters, as defined by q values of <.02. We also found 6 pairs (P < 5.144 × 10^{-5}) and 2 triplets (P < 4.167 × 10^{-4}) of alleles that were in linkage
Evidence of HIV CTL Escape and Clustering

We next evaluated HIV pol sequences obtained from all of the individuals within a cluster for evidence of escape mutations from the HLA alleles enriched within that cluster. Escape was determined by a change in the predicted MHC binding affinity and epitope processing score [28, 30, 31] and was determined for each putative HIV pol epitope specific for the particular HLA allele enriched within the cluster. Although we identified 19 alleles with a trend toward enrichment, only 11 of these could be analyzed with the IEDB MHC class 1 processing prediction tool. Computational modeling predicted that, of the 47 defined epitopes mapped within HIV pol for these 11 alleles, 6 showed evidence of escape in the respective cluster; all were associated with HLA alleles A2, A30, B18, and B50. Only one epitope, associated with HLA A30, demonstrated putative improved immunogenicity. Interestingly, in the largest cluster (cluster 62; n = 350), we found that, in 4 of the 9 identified pairs, there was evidence of reversion of potential escape mutations from well-characterized epitopes between the source and the recipient partners.

DISCUSSION

We hypothesized that there may be increased transmission of HIV between individuals with similar HLA profiles [33], which would be seen as increased prevalence of shared HLA alleles among individuals in phylogenetically defined clusters. To evaluate this, we examined the relationship of HLA alleles with clustering in a cohort of recently HIV-infected individuals in San Diego. This study found that HLA homozygosity was not associated with overall clustering. However, there were trends that some HLA alleles may be enriched within certain clusters (ie, concordance), but these results were not significant when corrected for multiple comparisons. Despite the evaluation of a large cohort, this is likely a limitation attributable to the sample size. Interestingly, when evaluating clusters with HLA enrichment, we found some degree of potential
disequilibrium. However, none of these alleles were enriched within clusters, suggesting that accompanying alleles did not affect the analysis.

Evaluation of the rates of HLA allele concordance in our study population revealed that concordance in the clustering group was less than what would be expected by chance. We first analyzed rates of concordance between individuals within single clusters (≥2 individuals) and found that, of 162 individuals in 42 clusters, 54.6% of HLA-A alleles, 40.1% HLA-B alleles, 62.0% HLA-C alleles, and 55.9% HLA-DRB1 alleles were concordant with those of another individual within that cluster. Next, we compared these observed rates of concordance to the expected rates of concordance that were based on the distribution of HLA alleles in our nonclustering population and found that our observed rates of concordance within clusters were significantly less than the expected rates, when analyzing all 4 HLA loci together (P = .042), HLA-A alone (P = .029), or HLA-B alone (P = .003). When evaluating only the transmission pairs from our cohort with known epidemiologic and phylogenetic linkage, we also found evidence that rates of concordance were significantly less than expected (HLA-A P = .004, HLA-B P < .001, and HLA-DRB1 P = .004). When evaluating only white individuals, significantly less concordance was found only at the HLA-C locus (P = .02).
viral escape within 6 of 47 epitopes, demonstrating viral escape across a cluster. Although subtle, these results provide evidence that alleles enriched within transmission clusters have footprints on the viral sequence circulating within that cluster and are consistent with results of studies that have identified increased transmission between individuals who share HLA alleles [33].

When we compared the rates of concordance within our clusters with what would be expected by chance on the basis of the prevalence of the HLA alleles in our population, concordance between HLA alleles within clusters was statistically less than expected by chance. Although the biologic function of this is unclear, previous work has suggested that heterosexual individuals are more sexually attracted to others of the opposite sex who have a different HLA profile than their own [34, 35]. This is presumably due to the evolutionary benefit for potential offspring, as a more diverse HLA repertoire could allow the offspring to respond to a greater diversity of pathogens. This study provides some small evidence that this phenomenon may not be restricted to heterosexual individuals, since the vast majority of study participants were men who reported sex with other men as their HIV risk factor.

This study has several limitations, including a relatively small sample size and the use of convenience venue-based testing to identify recent infections. A more thorough and dense sampling of our local population could have better characterized clusters and reduced selection bias. This is demonstrated by the difference in the clustering and nonclustering population numbers from subjects sampled prior to 2000 and those sampled after 2000. After 1 January 2000, the Primary Infection program began a much more intensive sampling of our local HIV population and therefore was able to identify a greater percentage of the acute HIV infections in the area. This increased sampling translated into increased identification of transmission clusters in the cohort. Of note, the phylogenetic clusters evaluated in this study do not necessarily represent direct transmissions but rather individuals linked along transmission chains. Finally, our evaluations of viral escape were based on computational models of the effects of consensus sequence changes on MHC binding affinity and peptide processing. Methods that evaluate CTL recognition and response directly, like the enzyme-linked immunospot assay, should be used in future studies to better characterize putative escape mutations.

In conclusion, this study found evidence for HLA enrichment in particular HIV transmission clusters and viral escape from these enriched alleles within those clusters. These findings could represent either that there is either reduced viral evolution among transmissions between individuals with similar HLA alleles or that HIV transmissions are more likely between individuals with similar HLA alleles. However, since this study also found evidence that HLA discordance was robustly associated with clustering, it is more likely that there is reduced viral evolution occurring between individuals with similar HLA alleles. Although this study included 350 recently infected, well-characterized individuals in one community with a large proportion of clustering (46%), an even larger and more densely sampled cohort will be needed to better determine the full extent of the relationship between HLA profile and transmission dynamics within a population.

### Notes

**Acknowledgments.** We thank Josue Perez-Santiago and Anya Umlauf for assistance with R coding.

**Financial support.** This work was supported by the National Institutes of Health (grant AI093163 to S. R. M. and grants MH083552, AI077304, AI69432, MH62512, AI27670, AI38858, AI43638, AI43752, AI047745, NS51132, AI6214 (UCSD Center for AIDS Research Translational Virology Core) AI36214, AI080353, AI29164, AI064086, A157167, and AI087164 to D. M. S.), the Pendleton Foundation (to D. M. S.), and the San Diego Veterans Affairs Healthcare System (to D. M. S. and S. R. M).

**Potential conflicts of interest.** All authors: No reported 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


