

HIV Transmission Chains Exhibit Greater HLA-B Homogeneity Than Randomly Expected

Huyen Nguyen, MSc,^{a,b} Christian W. Thorball, MSc,^{c,d} Jacques Fellay, MD,^{c,d} Jürg Böni, PhD,^b Sabine Yerly, PhD,^e Matthieu Perreau, PhD,^f Thomas Klimkait, PhD,^g Katharina Kusejko, PhD,^{a,b} Nadine Bachmann, MSc,^{a,b} Sandra E. Chaudron, MSc,^{a,b} Paolo Paioni, MD,^h Maria C. Thurnheer, MD,ⁱ Manuel Battegay, MD,^j Matthias Cavassini, MD,^k Pietro Vernazza, MD,^l Enos Bernasconi, MD,^m Huldrych F. Günthard, MD,^{a,b} and Roger Kouyos, PhD,^{a,b}
the Swiss HIV Cohort Study

Background: HIV's capacity to escape immune recognition by human leukocyte antigen (HLA) is a core component of HIV pathogenesis. A better understanding of the distribution of HLA class I in HIV-infected patients would improve our knowledge of pathogenesis in relation to the host HLA type and could better improve therapeutic strategies against HIV.

Materials and Methods: Three hundred one to 325 transmission pairs and 469–496 clusters were identified for analysis among Swiss HIV Cohort Study (SHCS) participants using HIV pol sequences from the drug resistance database. HLA class I data were compiled at 3 specificity levels: 4-digit, 2-digit alleles, and HLA-B supertype. The analysis tabulated HLA-I homogeneity as 2 measures: the proportion of transmission pairs, which are HLA concordant, and the average percentage of allele matches within all clusters. These measures were compared with the mean value across randomizations with randomly assorted individuals.

Results: We repeated the analysis for different HLA classification levels and separately for HLA-A, -B, and -C. Subanalyses by the risk group were performed for HLA-B. HLA-B showed significantly greater homogeneity in the transmission chains (2-digit clusters: 0.291 vs. 0.251, P value = 0.009; supertype clusters: 0.659 vs. 0.611, P value = 0.002; supertype pairs: 0.655 vs. 0.608, P value = 0.014). Risk group restriction caused the effect to disappear for men-who-have-sex-with-men but not for other risk groups. We also examined if protective HLA alleles B27 and B57 were under- or overrepresented in the transmission chains, although this yielded no significant pattern.

Conclusions: The HLA-B alleles of patients within HIV-1 transmission chains segregate in homogenous clusters/pairs, potentially indicating preferential transmission among HLA-B concordant individuals.

Key Words: HLA, coevolution, transmission chains

(*J Acquir Immune Defic Syndr* 2019;81:508–515)

Received for publication November 19, 2018; accepted March 27, 2019.

From the ^aDivision of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland; ^bInstitute of Medical Virology, Swiss National Center for Retroviruses, University of Zurich, Zurich, Switzerland; ^cSchool of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; ^dSwiss Institute of Bioinformatics, Lausanne, Switzerland; ^eLaboratory of Virology, Geneva University Hospital, University of Geneva, Geneva, Switzerland; ^fDivision of Immunology and Allergy, University Hospital Lausanne, University of Lausanne, Lausanne, Switzerland; ^gDepartment of Biomedicine, University of Basel, Basel, Switzerland; ^hDivision of Infectious Diseases and Hospital Epidemiology, Children's Research Center, University Children's Hospital Zurich, Zurich, Switzerland; ⁱUniversity Clinic of Infectious Diseases, University Hospital of Bern, University of Bern, Bern, Switzerland; ^jInfectious Diseases and Infection Control Clinic, Department of Medicine, University Hospital Basel, University of Basel, Basel, Switzerland; ^kDepartment of Infectious Diseases, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; ^lDivision of Infectious Diseases and Hospital Epidemiology, Kantonsspital St. Gallen, St. Gallen, Switzerland; and ^mDivision of Infectious Diseases, Regional Hospital, Lugano, Switzerland. Supported by the Swiss National Science Foundation (Grant numbers BSSGI0_155851, 33CS30_177499 to H.F.G.). H.F.G. was supported by the Swiss National Science Foundation (Grant number 179571). Furthermore, this study has been financed within the framework of the Swiss HIV Cohort Study, supported by the Swiss National Science Foundation (Grant number 148522), by the SHCS Research Foundation, by the Yvonne Jacob Foundation (to H.F.G.), by the clinical research priority program of the University of Zurich Viral infectious diseases, ZPHI (to H.F.G.), and the University of Zurich University Research Priority Program: Evolution in Action: From Genomes to Ecosystems (Grant number U-702-26-01, to R.K.). H.F.G. has received an unrestricted research Grant from Gilead to the SHCS Research Foundation. Presented at Conference on Retroviruses and Opportunistic Infections (CROI); March 4–7, 2018; Boston, MA.

H.F.G. has received unrestricted research grants from Gilead Sciences and Roche; fees for data and safety monitoring board membership from Merck; consulting/advisory board membership fees from Gilead Sciences, Sandoz and Mepha; and travel reimbursement from Gilead. M.B. has received research or educational grants by AbbVie AG, Gilead Sciences Switzerland Sàrl, Janssen-Cilag AG, MSD Merck Sharp & Dohme AG, and ViiV Healthcare GmbH. E.B. has received fees for his institution for participation to advisory board from MSD, Gilead Sciences, ViiV Healthcare, AbbVie, and Janssen. M.C. has received research and travel grants for his institution from ViiV and Gilead. T.K. has received honoraria from Gilead Sciences and Roche Diagnostics. R.D.K. has received grants from the Swiss National Science Foundation and personal fees from Gilead Sciences, outside the submitted work. The remaining authors have no conflicts of interest to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Correspondence to: Huyen Nguyen, MSc, Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zürich, Rämistrasse 100, Zürich CH-8091 (e-mail: Huyen.nguyen@usz.ch).

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

INTRODUCTION

The pathogenesis of HIV varies vastly from individual to individual, largely determined by the infected individual's immune system's capacity to drive a clonal antigen-specific response and the virus' capacity for immune escape.^{1–3} This mechanism hinges on the human leukocyte antigen (HLA), a gene complex encoding for major histocompatibility complex proteins. Major histocompatibility complex class I proteins are found on the surface of all nucleated cells in vertebrates and present antigens from the interior of the cell to the exterior, allowing for binding to cytotoxic T-cells, thus mediating the identification and destruction of virally infected cells.^{4,5} The virus' capacity to develop escape mutations to elude HLA binding is crucial to the virus' ability to avoid long-term HIV control by the immune system, subsequently facilitating the infection of other individuals.³ The high diversity of the HLA genes is believed to be due to host-pathogen coevolution over multiple human generations, which drives the emergence of rarer HLA alleles that allow for binding to a broader variety of pathogen antigens.^{4–6} This coevolutionary dynamic is an important focus of recent HIV research.^{7–9}

HLA genes are known to affect viral progression—some HLA alleles (eg, Cw04) are associated with a more rapid progression to AIDS, whereas others such as B57 are demonstrated to be protective.^{1,10–12} Furthermore, we know that certain HLA-associated single-nucleotide polymorphisms (SNPs) are associated with HIV control. For example, the combined variation in both the HLA and CCR5 regions can explain 25% of the variability in viral load.² A statistical analysis conducted by Kaslow et al¹³ demonstrated that HIV-infected individuals' HLA-I profiles could predict the time to the onset of AIDS. An analysis of transmission pairs by Goepfert et al¹⁴ indicated that HLA-B restriction by cytotoxic T lymphocytes on escape mutations corresponded to lesser viral loads in the recipients within transmission pairs. An individual-level effect is observed where escape mutations specific to that host's HLA alleles are selected, and such mutations tend to reverse when transmitted to another host with a different HLA genotype.^{3,10,11,15,16} These dynamics may partially explain why mother-child transmission is more likely when there is HLA concordance.^{17,18} Evidence suggests that HLA-HIV dynamics do shape the HIV epidemic—eg, more escape mutations to protective HLA alleles are observed over time.¹⁹ It is, however, unclear if these processes impact the ability of the virus to be transmitted and how they could shape epidemiological patterns.

The increasing amount of genotypic resistance sequencing data from HIV+ patients in addition to burgeoning computational power has allowed us to construct transmission phylogenies.²⁰ In this study, we examined HIV transmission chains where the individuals' HLA type is known and assessed the distribution of HLA alleles in transmission chains, and if there is greater HLA homogeneity compared with if the individuals were randomly assorted. Such insights would improve our understanding of the coevolutionary dynamics between HIV and HLA at the population level.

This study is the first that overlays patient HLA data over HIV transmission data at this scale.

Better understanding the dynamics of how HIV obtains escape mutations to avoid HLA binding and then reverts to the wild type has implications for vaccine development as this would help better assess the potential of T-cell-based vaccines.²¹ Insights into how the individual makeup of HLA class I in HIV+ individuals is distributed within transmission chains are important to understanding HIV evolutionary dynamics, contributing to efforts in vaccine design and other therapeutic strategies.

METHODS

Cohort

The Swiss HIV Cohort Study (SHCS) is a prospective multicenter study with continuing enrollment of people living with HIV in Switzerland since 1988 and covers about half of all people ever declared to be living with HIV to the Swiss health authorities and 75% of those on antiretroviral treatment in Switzerland.²² As of January 2018, the SHCS has a cumulative total of 20,089 patients. Demographic information, mode of HIV acquisition, information of risk behavior, treatment, and other data are collected at 6-month intervals following the standard protocol.

Ethics Statement

The SHCS was approved by the participating institutions' ethics committees (Kantonale Ethikkommission Bern, Ethikkommission des Kantons St. Gallen, Comité Départemental d'Éthique des Spécialités Médicales et de Médecine Communautaire et de Premier Recours, Kantonale Ethikkommission Zurich, Repubblica e Cantone Ticino—Comitato Etico Cantonale, Commission Cantonale d'Éthique de la Recherche sur l'Être Humain, Ethikkommission beider Basel; all approvals are available at <http://www.shcs.ch/206-%0Dethic-committee-approval-and-informed-consent>). Written informed consent was obtained from all participants.²²

Phylogenetic Tree

Viral pol gene sequences were obtained from patients' blood as part of HIV-1 routine resistance testing and through systematic retrospective sequencing of stored plasma samples. The sequences consisted of the partial pol region, specifically all of protease and a minimum of reverse transcriptase codons 28–225. The earliest sequence of adequate length of each patient (n = 11,922) was used to construct a maximum-likelihood phylogenetic tree along with 11,390 Los Alamos HIV database background sequences, obtained by blasting the SHCS sequences against the database to help identify which clusters likely involved transmission within Switzerland.²³ The phylogenetic tree was constructed using MUSCLE for sequence alignment against the HIV reference sequence HXB2. Insertions related to HXB2 and common resistance mutation positions were removed by the program TrimAL.^{24–26} FastTree, which infers a maximum-

likelihood phylogeny, was used for the final phylogeny reconstruction.^{27,28}

HLA Data

HLA allelic information was obtained from the SNP genotype data, using SNP2HLA with the type 1 Diabetes Genetics Consortium reference panel for HLA imputation techniques from the exome/SNP data.^{29–31} The data used in the analyses were limited to those of HLA class I molecules (HLA-A, -B, and -C) because of the literature supporting the key role of HLA class I peptides in HIV control.^{32,33} The homogeneity analyses were performed for each HLA gene and for each classification level. The HLA allelic information is classified in our data at 3 different levels: 4-digit, 2-digit, and supertype. The 4-digit classification refers to a specific peptide sequence of the HLA peptide, whereas 2-digit groups to serologically similar peptides, or an “allele group.” For HLA-B, a broader classification scheme by supertype was also used, following main peptide anchor specificity.³⁴

Pairs Analysis

For the pairs analysis of HLA homogeneity, pairs were identified as follows: (1) 2 individuals with sequences being adjacent tips in the phylogeny (“cherries”), (2) having a maximum cophenetic distance between the 2 individuals’ sequences of 0.03, (3) both being SHCS patients, (4) both having data on at least a single HLA allele, and (5) not having any perinatally infected individual.

For each pair analysis (for HLA-A, -B, and -C at 2 levels of allelic specificity in addition to supertype for HLA-B), the SHCS pairs were scrambled 10,000 times to obtain a distribution of the average proportion of pairs with a match on at least a single HLA allele. The proportion of pairs with an allelic match in the actual SHCS pairs was compared with the distribution from the randomizations to obtain a *P* value, which was used to evaluate whether there is significantly greater homogeneity in the SHCS pairs compared with if the same individuals were randomly assorted into the same number of pairs (see Figure 1, Supplemental Digital Content, <http://links.lww.com/QAI/B328>). HLA-B subanalyses were performed with pairs being stratified by risk groups: men-who-have-sex-with-men (MSM), heterosexuals (HET), and intravenous drug users (IDUs). All the above analyses were then repeated while restricting for individuals of white ethnicity. To correct for overtesting, the Benjamini–Hochberg procedure was performed for the pair/clusters analyses, with a false discovery rate set at 0.05 and the number of tests equal to 64, the total number of pair/clusters analyses performed.

Clusters Analysis

For the clusters analyses of HLA homogeneity, clusters were identified from the phylogeny as follows: (1) forming a clade of at least 2 individuals’ sequences (thus including pairs as well), (2) having a maximum cophenetic distance between any of the individuals’ sequences of 0.03, (3)

comprising individuals who are over 80% from the SHCS, and (4) having data on at least a single HLA allele for a minimum of a third of the cluster individuals (0.333 is the median proportion of HLA coverage in the clusters), disregarding perinatally infected patients. An individual can only belong to a single cluster (ie, smaller-nested clusters are disregarded).

For each clusters analysis (for HLA-A, -B, and -C), the proportion of all possible individual-to-individual comparisons that had an allelic match was calculated (see Figure 2, Supplemental Digital Content, <http://links.lww.com/QAI/B328>). The average intracluster proportion of allelic matching in the SHCS clusters was compared with a distribution of the average of 10,000 randomizations where the individuals were assorted randomly into the same number of clusters of the same sizes. This yielded a *P* value determining whether there is significantly greater HLA homogeneity in the SHCS clusters compared with randomly assorted clusters. An HLA-B subanalysis was performed with clusters being stratified by the risk group as in the pairs analysis (at least two-thirds of a cluster’s individuals must belong to the risk group to be included).

Protective Allele Analysis

To better understand the nature of the HLA alleles found in the transmission phylogeny, the distribution of 2 HLA-B alleles, B27 and B57, both protective in terms of HIV-disease progression, was examined for patients within the transmission pairs and clusters for the HLA-B 2-digit analysis and those outside of any pair/cluster.³⁵ We used a *Z* test for 2 proportions to test whether the protective alleles are over- or underrepresented in the pairs and clusters.

RESULTS

Demographic Overview

The SHCS consists of 20,089 patients at the time of the analysis (January 2018), of which about half (*n* = 11,922) had at least 1 sequence that was used to construct the transmission phylogeny (Table 1). Considering the limited coverage of HLA data on the patients, 336 transmission pairs (phylogenetic cherries) were identified as having HLA allelic data for both individuals and fitting other criteria as explained in the Methods section, as well as 1319 individuals identified in valid clusters (see Figure 3, Supplemental Digital Content, <http://links.lww.com/QAI/B328>). The individuals in pairs/clusters are more likely men (78.42% and 81.27% in pairs and clusters, respectively vs. 72.42% for the whole cohort) and more likely to be MSM (49.70% and 50.42% vs. 38.65%) and less likely to be individuals who contracted HIV through heterosexual contact or intravenous drug usage (Table 1). This can be explained by the dominance of IDUs in the first decade of the Swiss epidemic, before the introduction of highly active antiretroviral treatment and resistance testing and sequencing, during which most passed away.²⁰ The predominance of MSM in the pairs/clusters is expected—most recent HIV infections in Switzerland occurred among

TABLE 1. General Characteristics of SHCS Patients and Those Analyzed in the Pairs and Clusters

	Swiss HIV Cohort Study (SHCS)	SHCS Patients with Sequences in the Transmission Phylogeny	SHCS Patients in Transmission Pairs Fitting Criteria	SHCS Patients in Transmission Clusters Fitting Criteria
Total number	20,089	11,922	672	1319
Median age (IQR)	32.69 (26.78–41.00)	27.00 (33.04–41.23)	33.60 (28.04–39.68)	33.88 (27.85–41.00)
Male (%)	14,548 (72.42)	8554 (71.75)	527 (78.42)	1072 (81.27)
Risk group				
MSM (%)	7764 (38.65)	4738 (39.74)	334 (49.70)	665 (50.42)
Heterosexual (%)	6624 (32.97)	4245 (35.61)	205 (30.51)	340 (25.78)
Intravenous drug user (%)	4824 (24.01)	2430 (20.38)	116 (17.26)	276 (20.92)
Others (%)	877 (4.37)	509 (4.27)	17 (2.53)	38 (2.88)
White (%)	13,611 (67.75)	9199 (77.16)	640 (95.24)	1276 (96.74)

Overall demographic information of patients in the SHCS, as well as those represented in the transmission phylogeny, and those with HLA data in the analyzed pairs and clusters as fitting the criteria (described in the Materials and Methods section).
IQR, interquartile range.

MSM.³⁶ The patients in the pairs/clusters are more likely of white ethnicity than in the general SHCS (95.24% and 96.74% vs. 67.75%). This is due to HLA imputation techniques being validated for those of white ethnicity; hence, white patients are more likely to have available HLA data. In addition, white patients generally had greater follow-up within the cohort, thus being more likely to having had viral sequencing performed.³⁷ The predominance of white–nonwhite couples within heterosexual couples is another factor in the relative underrepresentation of patients from the heterosexual risk group having HLA data within the pairs and clusters.³⁶

HLA Homogeneity Analyses

For the pairs analyses, we found greater-than-random homogeneity for HLA-B. Three hundred one, 325, and 310 pairs were identified, respectively, for the HLA-A, -B, and -C analyses. When analyzing all pairs unrestricted by the ethnicity or risk group, there was only a significant difference in HLA homogeneity observed with the HLA-B supertype analysis, where 65.5% of the pairs had an HLA-B allele supertype match compared with an average of 60.8% in the randomly assorted pairs (P value = 0.014) (Table 2). When restricting to the 59 pairs where both individuals were infected through heterosexual transmission, there was significantly greater than random HLA-B concordance at the 4-digit and 2-digit levels (4-digit: 25.4% vs. 15.9%, P value = 0.017; 2-digit: 35.6% vs. 23.1%, P value = 0.006) than compared with the randomly assorted pairs, although the 4-digit level analysis significant result is potentially a result of overtesting as suggested after applying the Benjamini–Hochberg procedure. As for the 34 pairs where both individuals were IDUs, the HLA-B 2-digit and supertype analyses demonstrated more HLA-concordant pairs compared with the randomized pairs [2-digit: 38.2% vs. 24.9%, P value = 0.018 (a potential significant value due to overtesting); supertype: 82.4% vs. 67.8%, P value = 0.007]. The results were largely unchanged when restricting to pairs with only white

individuals (see Table 1, Supplemental Digital Content, <http://links.lww.com/QAI/B328>).

A similar result was found for the clusters analyses. Four hundred sixty-nine, 496, and 475 clusters fitting the defined criteria were identified for the HLA-A, -B, and -C analyses, respectively (Table 3). The average cluster size ranged between 3.911 and 3.981 depending on the HLA type, with about two-thirds of the cluster individuals having HLA

TABLE 2. Statistical Analyses of HLA Homogeneity in Transmission Pairs

HLA-Type (Risk Group)	HLA Analysis Level	No. of Pairs	Mean Proportion of Pairs With a Match	Average Value in Randomizations	<i>P</i>
A (all)	4-digit	301	0.405	0.380	0.131
	2-digit		0.442	0.415	0.120
B (all)	4-digit	325	0.200	0.182	0.168
	2-digit		0.265	0.241	0.136
C (all)	Supertype	310	0.655	0.608	0.014
	4-digit		0.271	0.287	0.718
B (MSM)	2-digit	142	0.406	0.420	0.691
	4-digit		0.190	0.192	0.468
B (HET)	2-digit	59	0.239	0.249	0.560
	Supertype		0.648	0.606	0.087
B (IDU)	4-digit	34	0.254	0.159	0.017*
	2-digit		0.356	0.231	0.006
	Supertype		0.695	0.636	0.093
	4-digit		0.265	0.184	0.064
	2-digit		0.382	0.249	0.018*
	Supertype		0.824	0.678	0.007

Overview of statistical analyses measuring HLA allelic homogeneity in transmission pairs (proportion of pairs with at least 1 matching HLA allele). Different analyses were performed analyzing varying HLA types (A, B, and C), levels of HLA classification specificity (4-digit, 2-digit, and supertype), and risk groups (all individuals, MSM, HET, and heterosexual). Significant P values of tests (<0.05) are denoted in bold.

*Significant P values that become nonsignificant after the Benjamini–Hochberg procedure, taking into account 64 total tests, with a false discovery rate of 0.05.

TABLE 3. Statistical Analyses of HLA Homogeneity in Transmission Clusters

HLA-Type (Risk Group)	HLA Analysis Level	No. of Clusters	Average Cluster Size	Average Percentage of Individuals in Clusters with HLA Data	Mean Proportion of Intracluster Matches	Average Value in Randomizations in Clusters	P
A (all)	4-digit	469	3.981	65.82	0.379	0.376	0.447
	2-digit				0.413	0.411	0.456
B (all)	4-digit	496	3.911	66.47	0.191	0.185	0.329
	2-digit				0.291	0.251	0.009
	Supertype				0.659	0.611	0.002
C (all)	4-digit	475	3.979	66.35	0.294	0.293	0.478
	2-digit				0.410	0.425	0.810
	4-digit				0.171	0.193	0.695
B (MSM)	2-digit	214	4.000	69.51	0.245	0.253	0.616
	Supertype				0.642	0.615	0.131
	4-digit				0.249	0.162	0.010
B (HET)	2-digit	81	3.037	76.02	0.371	0.244	0.002
	Supertype				0.714	0.637	0.027*
	4-digit				0.222	0.183	0.137
B (IDU)	2-digit	78	4.641	57.73	0.341	0.246	0.010
	Supertype				0.699	0.602	0.007

Overview of statistical analyses measuring HLA allelic homogeneity in transmission clusters (average proportion of all comparisons within clusters where at least 1 match exists). Different analyses were performed analyzing varying HLA types (A, B, and C), levels of HLA classification specificity (4-digit, 2-digit, and supertype), and risk groups (all individuals, MSM, HET, and heterosexual). Significant *P* values of tests (<0.05) are denoted in bold.

*Significant *P* values that become nonsignificant after the Benjamini–Hochberg procedure, taking into account 64 total tests, with a false discovery rate of 0.05.

data (Table 3). HLA-B 2-digit and supertype analyses demonstrated a higher average proportion of intracluster comparisons with an allelic match compared with randomized clusters (2-digit: 0.291 vs. 0.251, *P* value = 0.009; supertype: 0.659 vs. 0.611, *P* value = 0.002). A similar pattern was found when restricted to clusters where at least two-thirds of the individuals are HETs [4-digit: 0.249 vs. 0.162, *P* value = 0.010; 2-digit: 0.371 vs. 0.244, *P* value = 0.002; supertype:

0.714 vs. 0.637, *P* value = 0.027 (potential significant value due to over-testing)] and IDUs (2-digit: 0.341 vs. 0.246, *P* value = 0.010; supertype: 0.699 vs. 0.602, *P* value = 0.007). The results were essentially unchanged in the white ethnicity subanalysis (see Table 2, Supplemental Digital Content, <http://links.lww.com/QAI/B328>).

The 3 levels of analyses across the 3 HLA types are summarized in Figure 1, showing greater HLA homogeneity

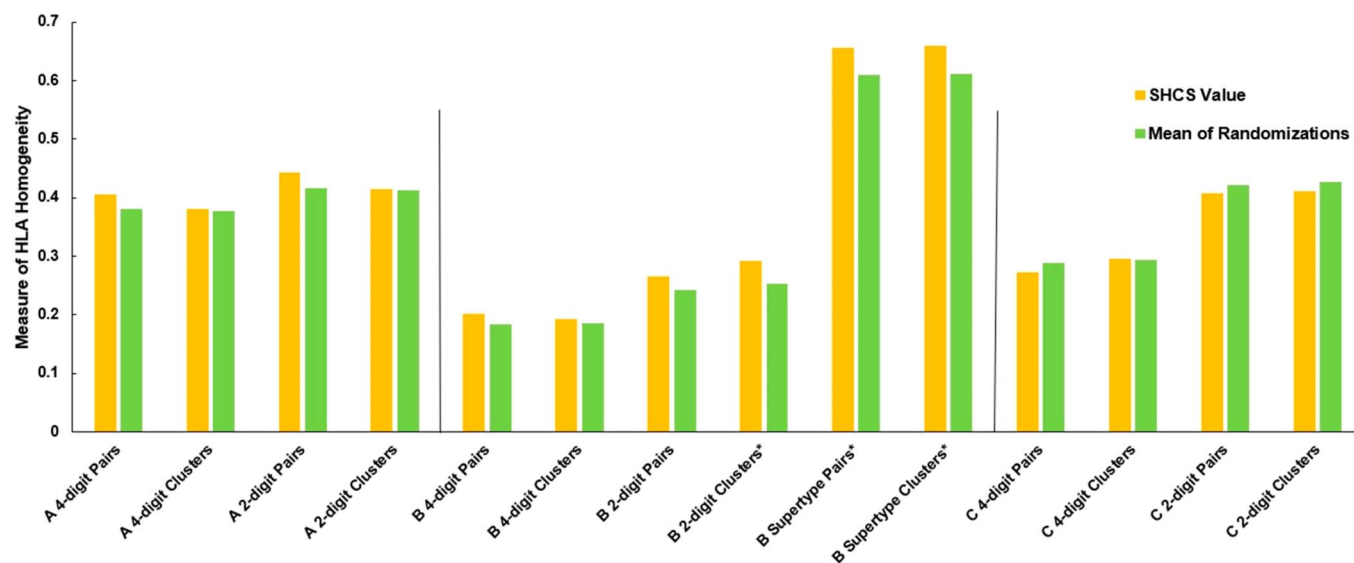


FIGURE 1. Overview of pairs and clusters analyses and HLA homogeneity. Visual overview of statistical analyses of HLA allelic homogeneity (average proportion of pairs with at least 1 matching HLA allele or average proportion of all comparisons within clusters with at least 1 matching allele). Different analyses were performed of the varying HLA class-I types (A, B, and C) and levels of HLA classification specificity (4-digit, 2-digit, and supertype).

in SHCS pairs/clusters compared with the randomized pairs/clusters when analyzing HLA-A and HLA-B, although significantly greater homogeneity was only observed with HLA-B 2-digit clusters, supertype pairs, and supertype clusters. This pattern was not observed for HLA-C. The highest homogeneity measures were found in the supertype analyses, followed by the 2-digit and 4-digit analyses, which is expected considering the greater number of HLA categorizations as specificity increases.

Protective Allele Analysis

As summarized in Table 4, HLA-B27 was shown to not be under- or overrepresented within the 1297 individuals of the HLA-B 2-digit analyses compared with the 3412 individuals outside of the pairs/clusters (proportions: 0.076 vs. 0.080, P value = 0.651). HLA-B57 was somewhat underrepresented for the individuals within the pairs/clusters with nonsignificant differences (0.067 vs. 0.084, P value = 0.057). The distribution of both protective alleles overall was not found to be concentrated within or without the pairs/clusters (0.141 vs. 0.161, P value = 0.094).

DISCUSSIONS

The HIV transmission pairs and clusters in the SHCS demonstrate significantly greater HLA-B homogeneity than if the individuals were randomly assorted. These findings were further strengthened when we restricted to individuals with white ethnicity to control for race. This observed effect with HLA-B is in line with other research that suggests the dominant role of HLA-B in determining HIV pathogenesis.^{14,38,39} This greater role for HLA-B is likely explained by the HLA-B locus' greater observed capacity for recombination and mutation compared with other loci. Considering HLA-B's greater peptide-binding versatility, it is thus no surprise that the virus' ability to escape HLA-B binding is critical to escape.⁴⁰

Several processes could explain our findings. If the virus is transmitted within a cluster between individuals with similar HLA alleles, then the developed escape mutations that are transmitted do not require further mutations or reversions

to the wild type as the virus is already primed to the respective HLA context of the newly infected individuals. Goulder et al⁴¹ suggested in their article demonstrating an association between mother-child HLA concordance and transmission that cytotoxic T lymphocyte (CTL) escape pressure could increase viremia and transmissibility. Such a mechanism would be supported by the literature suggesting faster disease progression in the recipient where there is an HLA match between a transmission pair.⁴² In such a scenario, larger clusters would demonstrate greater HLA homogeneity. Another potential mechanism behind the greater homogeneity may involve superinfection. If an individual was superinfected by a second strain of HIV, the superinfecting strain may have a greater probability to supplant the resident strain if it is preadapted to the HLA alleles of the recipient, which is more likely if recipient and new transmitter are HLA concordant. In such a scenario, the transmission phylogeny would show a phylogenetic affinity between 2 HLA-concordant individuals even if the identified infecting individual was not the primary transmitter. Previous research estimates the annual incidence of HIV superinfection in already infected individuals to be as high as 7.7%, which could explain the appearance of people with similar HLA-B types clustering together in our analysis, particularly among IDUs who may experience more superinfection events.^{43,44} This being said, we emphasize that our study and findings are exploratory in nature. We are thus unable to assert a mechanism, novel or not, behind these observations.

Mate choice and geographic clustering are often identified as confounders in HLA-related studies. Some studies have suggested that partner choice in humans may preferentially occur between HLA-discordant individuals as to increase offspring HLA diversity.⁴⁵ However, the evidence for this is inconclusive.⁴⁶ Regardless, our interpretation of the results would be unaffected—any mate selection effect would mean we would expect the null hypothesis to be disproportionately greater heterogeneity. Such a scenario would only strengthen the significance of the observed homogeneity in our data. In addition, only about a quarter of the SHCS individuals analyzed in our study are classified as having heterosexual contact as the transmission route, whereas about half are MSM (Table 1). No research to date examines how MSM partner selection may be related to HLA concordance.

Another potential confounder is geographic clustering, where HLA types tend to appear at different frequencies in different regions due to migration patterns and selection pressures that geographically vary; thus, people who live near each other are more likely to be HLA concordant.⁴⁷ We believe that such confounding can be mostly disregarded in our setting for the following reason: despite the HLA variation within Switzerland, the large majority of the individuals in our study are MSM or IDUs.⁴⁸ These 2 high-risk populations cluster in urban areas, forming subpopulations with people from diverse regions, largely eliminating geographic confounding for our study context.^{49,50} If geography or mate selection were to play a role in our observed results, then we would have observed HLA-A and -C homogeneity as well, which we do not.

TABLE 4. Distribution of Protective HLA Alleles B27 and B57 Within and Without the HLA-B Pairs/Clusters Analyses

HLA Allele	Proportion Having Protective Allele within Pairs/Clusters Analyzed (n = 1297)	Proportion Having Protective Allele outside Pairs/Clusters Analyzed (n = 3412)	P of Z-test of 2 Proportions
B27	0.076 (99)	0.080 (274)	0.651
B57	0.067 (87)	0.084 (286)	0.057
B27 or B57	0.141 (183)	0.161 (549)	0.094

Distribution of protective HLA-B alleles 27 and 57 inside and outside the pairs and clusters analyzed for the HLA-B 2-digit analyses, and the Z-test P values indicating if protective alleles are significantly over- or underrepresented in patients in the pairs/clusters.

In the subanalyses of IDUs, the subpopulation initially hardest hit by the Swiss HIV epidemic, we found even more significant HLA-B homogeneity at the 2-digit and supertype levels of analysis compared with the main analyses. Frequent needle sharing combined with the higher chance of HIV transmission by intravenous needle usage compared with sexual transmission routes (except receptive anal intercourse) may lead to more superinfection events among IDUs, leading to greater observed HLA homogeneity.⁵¹

Another potential explanation for the greater effect among IDUs may be that the IDUs in the Swiss cohort were primarily infected in the early phases of the epidemic. As research suggests greater CTL escape mutations with the progression of the epidemic, it is possible that the pattern described in our study attenuates with the passage of time and is thus more pronounced among IDUs.^{16,19}

Because roughly half of HIV+ people in Switzerland are in the SHCS, the transmission clusters are missing individuals that would otherwise be included, meaning that some transmission pairs are not true pairs but are instead part of a transmission network with the intermediary individuals missing. Furthermore, only a fraction of the individuals in the SHCS had HLA data, which could be used in our analysis. However, such an effect would just dilute our findings, as opposed to artificially increasing observed homogeneity.⁵² This combined with the high coverage of the SHCS compared with other HIV cohort studies strengthens our confidence in our findings.

The novel population-level approach used in this article provides new insights into HIV and HLA's coevolutionary dynamics, posing new questions about the mechanisms that HIV uses to adapt to the host's HLA type. Our study provides the first indication that HLA diversity is a selective factor for HIV-1 at the population level and not just at the individual level.

ACKNOWLEDGMENTS

The authors thank the patients who participated in the Swiss HIV Cohort Study; the physicians and study nurses, for excellent patient care provided to participants; the resistance laboratories for high-quality genotyping drug resistance testing; SmartGene (Zug, Switzerland), for technical support; Alexandra Scherrer, Susanne Wild, and Anna Traytel from the SHCS data center for data management, and Marianne Amstutz, Danièle Perraudin, and Mirjam Minichiello for administration.

The members of the Swiss HIV Cohort Study include the following: A. Anagnostopoulos, M.B., E.B., J.B., D. L. Braun, H. C. Bucher, A. Calmy, M.C., A. Ciuffi, G. Dollenmaier, M. Egger, L. Elzi, J. Fehr, J.F., H. Furrer (Chairman of the Clinical and Laboratory Committee), C. A. Fux, H. F. Günthard (President of the SHCS), D. Haerry (deputy of "Positive Council"), B. Hasse, H.H. Hirsch, M. Hoffmann, I. Hösli, M. Huber, C. Kahlert, L. Kaiser, O. Keiser, T.K., R.D. Kouyos, H. Kovari, B. Ledergerber, G. Martinetti, B. Martinez de Tejada, C. Marzolini, K.J. Metzner, N. Müller, D. Nicca, P.P., G. Pantaleo, M.P., A. Rauch (Chairman of the Scientific Board), C. Rudin (Chairman of the Mother & Child

Substudy), A.U. Scherrer (Head of Data Center), P. Schmid, R. Speck, M. Stöckle, P. Tarr, A. Trkola, P.V., G. Wandeler, R. Weber, and S.Y.

REFERENCES

1. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: heterozygote advantage and B* 35-Cw* 04 disadvantage. *Science*. 1999;283:1748–1752.
2. McLaren PJ, Coulonges C, Bartha I, et al. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proc Natl Acad Sci U S A*. 2015;112:14658–14663.
3. Leslie AJ, Pfaffertott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med*. 2004;10:282–289.
4. Markov PV, Pybus OG. Evolution and diversity of the human leukocyte antigen (HLA). *Evol Med Public Health*. 2015;2015:1.
5. Zinkernagel RM, Doherty PC. MHC-restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. *Adv Immunol*. 1979;27:51–177.
6. Borghans JA, Beltman JB, De Boer RJ. MHC polymorphism under host-pathogen coevolution. *Immunogenetics*. 2004;55:732–739.
7. Chang CH, Kist NC, Chester TL, et al. HIV-infected sex workers with beneficial HLA-variants are potential hubs for selection of HIV-1 recombinants that may affect disease progression. *Sci Rep*. 2015;5:11253.
8. Sun X, Shi Y, Akahoshi T, et al. Effects of a single escape mutation on T cell and HIV-1 co-adaptation. *Cell Rep*. 2016;15:2279–2291.
9. Deng K, Peretea M, Rongvaux A, et al. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature*. 2015;517:381–385.
10. Frater AJ, Brown H, Oxenius A, et al. Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. *J Virol*. 2007;81:6742–6751.
11. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med*. 2007;13:46.
12. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science*. 2007;317:944–947.
13. Kaslow RA, Carrington M, Apple R, et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med*. 1996;2:405.
14. Goepfert PA, Lum W, Farmer P, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *J Exp Med*. 2008;205:1009–1017.
15. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol*. 2004;4:630–640.
16. Carlson JM, Listgarten J, Pfeifer N, et al. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. *J Virol*. 2012;86:5230–5243.
17. MacDonald KS, Embree J, Njenga S, et al. Mother-child class I HLA concordance increases perinatal human immunodeficiency virus type 1 transmission. *J Infect Dis*. 1998;177:551–556.
18. Polycarpou A, Ntais C, Korber BT, et al. Association between maternal and infant class I and II HLA alleles and of their concordance with the risk of perinatal HIV type 1 transmission. *AIDS Res Hum Retroviruses*. 2002;18:741–746.
19. Payne R, Muenchhoff M, Mann J, et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proc Natl Acad Sci*. 2014;111:E5393–E5400.
20. Kouyos RD, Von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis*. 2010;201:1488–1497.
21. Davenport MP, Loh L, Petravic J, et al. Rates of HIV immune escape and reversion: implications for vaccination. *Trends Microbiol*. 2008;16:561–566.
22. Swiss HIV Cohort Study, Schoeni-Affolter F, Ledergerber B, et al. Cohort profile: the Swiss HIV Cohort study. *Int J Epidemiol*. 2009;39:1179–1189.

23. Los Alamos HIV Sequence Database [database online]. Los Alamos, NM: U.S. Department of Energy: Los Alamos National Security. Los Alamos HIV Sequence Database; 2017.
24. Rhee SY, Gonzales MJ, Kantor R, et al. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 2003;31:298–303.
25. International Antiviral Society-USA. *HIV Drug Resistance Mutations Figures and User Notes*. 2018. Available at: <https://www.iasusa.org/content/drug-resistance-mutations-in-HIV>. Accessed June 25 2018.
26. Capella-Gutiérrez S, Silla-Martínez JM, et al. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–1973.
27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–1797.
28. Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5:e9490.
29. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013;8:e64683.
30. Szolek A, Schubert B, Mohr C, et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* 2014;30:3310–3316.
31. Dillthey AT, Gourraud PA, Mentzer AJ, et al. High-accuracy HLA type inference from whole-genome sequencing data using population reference graphs. *PLoS Comput Biol.* 2016;12:e1005151.
32. Leslie A, Matthews PC, Listgarten J, et al. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *J Virol.* 2010;84:9879–9888.
33. Study TI. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* 2010;330:1551.
34. Sidney J, Peters B, Frahm N, et al. HLA class I supertypes: a revised and updated classification. *BMC Immunol.* 2008;9:1.
35. Streeck H, Lichterfeld M, Alter G, et al. Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *J Virol.* 2007;81:7725–7731.
36. Turk T, Bachmann N, Kadelka C, et al. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. *Elife.* 2017;6:e28721.
37. Shilahi M, Marzel A, Yang WL, et al. Genotypic resistance tests sequences reveal the role of marginalized populations in HIV-1 transmission in Switzerland. *Sci Rep.* 2016;6:27580.
38. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med.* 2003;54:535–551.
39. Kiepiela P, Leslie AJ, Honeyborne I, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature.* 2004;432:769–775.
40. McAdam SN, Boyson JE, Liu X, et al. A uniquely high level of recombination at the HLA-B locus. *Proc Natl Acad Sci USA.* 1994;91:5893–5897.
41. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature.* 2001;412:334.
42. Crawford H, Lum W, Leslie A, et al. Evolution of HLA-B* 5703 HIV-1 escape mutations in HLA-B* 5703-positive individuals and their transmission recipients. *J Exp Med.* 2009;206:909–921.
43. Redd AD, Quinn TC, Tobian AA. Frequency and implications of HIV superinfection. *Lancet Infect Dis.* 2013;13:622–628.
44. Yerly S, Jost S, Monnat M, et al. HIV-1 co/super-infection in intravenous drug users. *AIDS.* 2004;18:1413–1421.
45. Ober C, Weitkamp LR, Cox N, et al. HLA and mate choice in humans. *Am J Hum Genet.* 1997;61:497–504.
46. Tregenza T, Wedell N. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol Ecol.* 2000;9:1013–1027.
47. Piazza A, Menozzi P, Cavalli-Sforza LL. The HLA-A, B gene frequencies in the world: migration or selection? *Hum Immunol.* 1980;1:297–304.
48. Buhler S, Nunes JM, Nicoloso G, et al. The heterogeneous HLA genetic makeup of the Swiss population. *PLoS One.* 2012;7:e41400.
49. Wandeler G, Gsponer T, Bregenzer A, et al; Swiss HIV Cohort Study. Hepatitis C virus infections in the Swiss HIV Cohort Study: a rapidly evolving epidemic. *Clin Infect Dis.* 2012;55:1408–1416.
50. Grob PJ. *Zuercher <<Needle-Park>>*. 2nd ed. Zurich, Switzerland: Chronos Verlag; 2009.
51. Patel P, Borkowf CB, Brooks JT, et al. Estimating per-act HIV transmission risk: a systematic review. *AIDS.* 2014;28:1509–1519.
52. Kusejko K, Kadelka C, Marzel A, et al. Inferring the age difference in HIV transmission pairs by applying phylogenetic methods on the HIV transmission network of the Swiss HIV Cohort Study. *Virus Evol.* 2018;4:vey024.