Characterization of the Transmitted Virus in an Ongoing HIV-1 Epidemic Driven by Injecting Drug Use

Elena Dukhovlinova,[1] Alexey Masharsky,[2] Aleksandra Vasileva,[2] Alessandro Porrello,[1] Shuntai Zhou,[1] Olga Toussova,[3] Sergei Verevochkin,[2] Ekaterina Akulova,[4] Dmitrij Frishman,[2] David Montefiori,[5] Celia Labranche,[5] Irving Hoffman,[6] William Miller,[7] Myron S. Cohen,[6] Andrei P. Kozlov,[2,4] and Ronald Swanstrom[1,8]

Abstract

Understanding features of the HIV-1 transmission process has the potential to inform biological interventions for prevention. We have examined the transmitted virus in a cohort of people who inject drugs and who are at risk of HIV-1 infection through blood contamination when injecting in a group. This study focused on seven newly infected participants in St. Petersburg, Russia, who were in acute or early infection. We used end-point dilution polymerase chain reaction to amplify single viral genomes to assess the complexity of the transmitted virus. We also used deep sequencing to further assess the complexity of the virus. We interpret the results as indicating that a single viral variant was transmitted in each case, consistent with a model where the exposure to virus during transmission was limited. We also looked at phenotypic properties of the viral Env protein in isolates from acute and chronic infection. Although differences were noted, there was no consistent pattern that distinguished the transmitted variants. Similarly, despite the reduced genetic heterogeneity of the more recent subtype A HIV-1 epidemic in St. Petersburg, we did not see reduced variance in the neutralization properties compared to isolates from the more mature subtype C HIV-1 epidemic. Finally, in looking at members of injecting groups related to the acute HIV-1 infection/early subjects, we found examples of sequence linkage consistent with ongoing and rapid spread of HIV-1 in these groups. These studies emphasize the dynamic nature of this epidemic and reinforce the idea that improved prevention methods are needed.

Keywords: people who use drugs, genetic bottleneck, HIV in Russia, HIV envelope, HIV transmission, acute HIV infection

Introduction

The first cases of HIV-1 infection in USSR were detected in late 1980s,[1] but the expansion of HIV-1 epidemic started only in 1996 when a subtype A virus strain was introduced into the population of people who inject drugs (PWID).[2] Since then the HIV-1 epidemic in Russia has been growing rapidly, resulting in a total of 1,114,815 cases of infection, with over 100,000 new cases reported in 2016.[3]

The city of St. Petersburg is the second largest city in Russia and one of the regions with the highest prevalence of HIV-1 infection that approached 1% of the city population in 2016. PWIDs historically have been at the highest risk of HIV-1 acquisition in Russia and St. Petersburg[4,5]; however, the virus has been also spreading into the general population.[6] In 2016, 35% of new HIV-1 infections in St. Petersburg were attributed to injection drug use (data of St. Petersburg City AIDS Center, www.hiv-spb.ru, in Russian).

Acute HIV-1 infection (AHI) cases are characterized by extremely high viral loads and a proportionally high risk of further spread of infection, so monitoring and clinical management of such cases have become a standard practice in

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[2]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation.
[3]Pavlov State Medical University, St. Petersburg, Russian Federation.
[4]The Biomedical Center, St. Petersburg, Russian Federation.
[5]Laboratory for AIDS Vaccine Research and Development, Department of Surgery, Duke University, Durham, North Carolina.
[6]School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[7]College of Public Health, The Ohio State University, Columbus, Ohio.
[8]Department of Biochemistry and Biophysics, and the UNC Center for AIDS Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

many countries.[7] The studies of viral genetic and phenotypic diversity in AHI cases are aimed at finding the patterns in the transmitted isolates that can guide the design of a region-specific vaccine.[8,9] A stringent genetic bottleneck in heterosexual transmission has been observed and extensively studied for the subtype B and C viruses.[10–13]

We previously reported a cross-sectional study of active PWID who were in acute or early stages of HIV-1 subtype A infection, and found the HIV-1 genetic bottleneck present upon parenteral transmission, with the majority of infection cases attributed to a single viral variant.[14] In a similar study of PWID in Bangkok, the parenteral transmission was also associated with a similarly strong genetic bottleneck.[15] In contrast, the analysis of the transmitted HIV-1 in a PWID cohort of Montreal showed a less stringent bottleneck and a higher percentage of infections with a complex viral population.[16]

In this article, we analyzed longitudinal plasma samples collected through a cohort study that monitored HIV-1 acquisition in real time among PWID in St. Petersburg, Russia. The study design and cohort description are published elsewhere.[17] We used single genome amplification (SGA) and Illumina next-generation sequencing to detect minor variants present near the time of transmission among acutely or recently infected PWID. We confirmed the strong genetic bottleneck of HIV-1 transmission in a majority of studied participants. We also followed early evolution in longitudinal samples and observed a superinfection event during acute infection. Molecular epidemiology allowed the identification of transmission clusters within injecting subgroups. We also tested the idea that the overall less diverse isolates in St. Petersburg, due to the more recent introduction of HIV-1 into this population, might be more homogeneous in their neutralization properties than a more diverse and older epidemic.

Materials and Methods

Cohort description

A detailed description of the cohort is published elsewhere.[17] Briefly, we recruited and followed up individuals who were at high risk of HIV acquisition based on the criteria of frequent use of injectable drugs. The participants signed informed consent to donate the blood samples to be tested by real-time PCR, ELISA, and Western blot for HIV-1. They also agreed to answer a series of questions about their drug use behavior. Appropriate counseling was provided to each individual before and after the tests for HIV. The study was approved by Institutional Review Boards of UNC Chapel Hill and the Biomedical Center.

Single genome amplification

In SGA, the cDNA copies of the viral RNA are diluted to endpoint and nested polymerase chain reaction (PCR) amplifications are done at a dilution where a majority of the reactions are negative, ensuring that a majority of the positive amplifications were initiated from a single template; this approach ensures that each amplicon represents an independent observation, and the total number of amplicons obtained defines the sample size for querying the genetic diversity of the viral population.[18] Amplicons of individual copies of the HIV-1 *env* gene were obtained starting with viral RNA extracted from plasma and then amplifying cDNAs spanning the *env* gene using semi-nested PCR with the primers listed below. For the first round of SGA, we used cEnvA and EnvN, and for second round of SGA, we used cEnvA and EnvM. Primer sequences used for SGA sequencing are listed in Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com/aid).

The SGA sequences were deposited to the Genbank, ID: MH603967-MH604583. Sequence analysis was performed using BioEdit and Mega 6.0 software. Sequence alignment and Poisson Fitter analysis (after correction for APOBEC3G/F mutations) were performed using LANL database tool kit (https://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html). Calculations of mean pairwise nucleotide distances per sample and phylogenetic analysis of SGA-derived amplicons were performed using Mega 6.0 and plotted by Prism 6.0.

Illumina MiSeq sequencing

We used a previously published Primer ID approach to Illumina next-generation sequencing (NGS).[19] The NGS libraries were created with the *env*-specific primers **A3F** (*NGSAdapter1*-CAGTCATGACCTGGATGCAATGGGACA) and **A3R** (*NGSAdapter2*-CCCTATCTGTCCACCCAGCTACT) and sequenced using MiSeq at the UNC High-Throughput Sequencing Facility. We performed the bioinformatics data analysis as previously reported.[19]

For the NGS sequence analysis, we used two types of consensus sequences. First, the template consensus sequence (TCS) was generated from the sequence reads that contained the same Primer ID since each sequence was generated from a single cDNA molecule (and RNA template). Due to the high number of TCS sequences, we used randomly generated sets of 200 TCS sequences to compare the pairwise genetic diversity between the subjects.

For the phylogenetic analysis, we used collapsed identical TCS to focus on the diversity within the viral population. Thus, we were able to assess the viral variants present in the subject's viral population and the representation of these variants, which was done using 200 randomly sampled sequences from the collapsed population. The calculations of pairwise distances and phylogenetic analysis were performed using Mega 6.0.

Pseudovirus generation

The SGA-derived *env* amplicons were reamplified using the Phusion Hot-start High-Fidelity DNA polymerase (Finnzymes), gel-purified using the QIAquick Gel Extraction Kit (QIAGEN), and cloned using the pcDNA3.1 Directional TOPO Expression Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Sequence-verified vectors were used for a transient transfection of 293T cells in equal ratio with the luciferase-containing backbone vector pNL4-3.LucR-E- plasmid (AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH) with the FuGENE 6 transfection reagent (Roche). The viral supernatants were collected after 48 h, filtered through 0.45 μm filters (Millipore), and stored in small aliquots at −80°C to minimize freeze–thaw cycles.

Neutralization assay

Assessment of neutralization properties of pseudoviruses was performed in the TZM-bl cell line at the Laboratory for AIDS Vaccine Research and Development, Duke University,

using a standard protocol.[20] The data are reported either as the serum dilution or the concentration of a monoclonal antibody needed to inhibit 50% of the virus infectivity (ID50 or IC50, respectively), as interpolated using 5-parameter curve-fitting.

Neutralization data analysis

Our analysis of antibody (Ab) concentrations obtained in a neutralization assay for subtype A pseudoviruses split into two groups (acute vs. chronic infection) was performed in the following manner: preliminarily, data were organized in a matrix with viruses on the rows and Abs on the columns. For achieving a balanced data analysis that accounts for the steps described below, when an Ab concentration was above the range of measurability, it was represented by a number that equals twice the maximum of that range. Vice versa, when an Ab concentration was below the range of measurability, it was represented by a number that equals half the minimum of that range. Then, concentration values were log2-transformed, and subsequently, medians were calculated for each column and subtracted from them. The log transformation exploits the fact that when, for instance, $x = 2 \times y$, $\log_2(x/y) = 1$; instead, if $x = 0.5 \times y$, $\log(x/y) = -1$. Consequently, proportional changes in either direction (increases and reductions) always become symmetrical with respect to the 0 (if there is no variation, i.e., when $x = y$: $\log_2(x/y) = \log_2(x/x) = 0$).

Thereafter, we performed a hierarchical clustering both of the rows and the columns using the uncentered correlation coefficient as our similarity metric and the average linkage clustering method. Clustering results were displayed through a heat map and two dendrograms (one for the columns, shown on the top, and one for the rows, shown on the left side). Due to the column-by-column median centering that we used, all log-transformed values are recalculated with respect to the medians (i.e., by subtracting the medians), which have the well-known property to have the same number of (column) values greater and lower than themselves. Consequently, in the heat map of log-transformed data, a general chromatic balance is achieved, in each column, between blue squares (log-transformed values that were above the median before being median centered) and yellow squares (log-transformed values that were below the median before being median centered), and differences among viruses are more easily detectable.

We compared subtype A pseudovirus neutralization data with a panel of subtype C viruses tested in a similar way and using shared Abs.[21] In this new dataset, five cases of subtype A viruses that were recombinants between founder variants of the virus of subject AHI_D and a complex recombinant virus SC1283 were removed from the analysis; also, seven missing values of subtype C could not be used for the PG9 Ab. As a preliminary step, we adjusted neutralization values that were above or below the detection threshold, separate and independent of our clustering analysis of subtype A, for the five Abs available. These extreme values, which were recorded in our spreadsheets as $<X$ or $>Y$, were changed into $X/2$ and $2Y$, respectively. Then, all these values were log10-transformed, keeping into account that log2- and log10-transformed values are directly related by the approximate formula: $3.3219 \cdot \log_{10}(x) = \log_2(x)$. These log transformations, different from what was previously described for the heat maps, were motivated by the need to shift the skewness values toward 0 and make the kurtosis values nearer to 3, thus obtaining probability distributions closer to the normal before performing statistical inference on them using parametric methods. In particular, (1) the absolute values of the skewness of all the probability distributions of the analyzed Abs were reduced after log transforming the data and (2) from a more qualitative perspective and more complexly, also, the absolute values of the difference between the kurtosis of these distributions (originally, all leptokurtic) and the number 3 were reduced after the log transformation. Additionally, the statistical tests described below were utilized relying on their robustness.

Then, we tested the differences of variance and mean concerning four groups of viruses (group 1: acute infection, subtype A; group 2: chronic infection, subtype A; group 3: acute infection, subtype C; and group 4: chronic infection, subtype C). In particular, we compared group 1 versus 3 and group 2 versus 4 for each Ab. First, *F*-tests (two sample tests for equal variance) were performed for these two comparisons. Later, for the similar two comparisons, we performed a *t*-test (of equality of the means) that either assumes that the samples come from independent random samples with the same variance (when the corresponding *F*-test *p* value was greater than .05) or that does not make this assumption (Welch's *t*-test, when the *F*-test *p* value was less than .05). Canonically, the *F*-tests were used to compare the spreading of the values around the means, that is, the variances, between groups, while the *t*-tests assessed the differences between the means themselves. Since there was no significant difference in the mean and variance within subtype A and C of samples of acute and chronic viruses for four Abs (PG9, 4E10, sCD4, and SA-C72), we tested for statistical differences between the subtype A (acute and chronic cases) and C (acute and chronic cases) using the *F*-test as well as the standard or Welch's *t*-test, based on the results of the corresponding *F*-test.

For the sake of completeness, also, the statistical results of the fifth Ab (HIVIG-C) were calculated and are shown with the others, despite the caveat that acute and chronic viruses of this Ab have a statistically significant difference in their variance for subtype C. All statistical tests were two tailed.

Tropism assay

We assessed the usage of CD4 and coreceptors in the inducible Affinofile cell line according to the protocol described elsewhere.[22] Pseudoviruses were incubated with Affinofile cells that were induced to either $CD4^{high}/CCR5^{high}$ or $CD4^{low}/CCR5^{high}$ expression profiles, which approximate the level of CD4 on T cells or macrophages, respectively. The pseudoviruses were incubated with cells in the presence or absence of AMD3100 and/or maraviroc (10 μM) to determine the ability to use CXCR4 or CCR5 as the coreceptor. Data were analyzed and plotted using Prism 6.0.

Results

Characterization of subjects with AHI with injection drug use as their major transmission risk factor

Briefly, we recruited a cohort of HIV-negative persons who were actively injecting drugs (PWID) in St. Petersburg, Russia, described in greater detail elsewhere.[17] The cohort participants were frequently monitored for plasma HIV-1

RNA to identify cases of AHI. Based on their drug use patterns and frequency of injection, these subjects were at high risk of HIV-1 acquisition through injection and this allowed analysis of the transmitted virus upon injecting drug use (IDU) transmission. A summary of the demographics of the seven subjects identified in acute infection, their risk factors, and the dates of plasma sampling after transmission are presented in Table 1.

Based on the provided behavioral data, that is, number of sex partners, number of sexual contacts before AHI detection, and frequency of drug injection, we assigned drug use (IDU), sexual transmission (Sex), or both as potential route of transmission for each AHI subject. Due to the low sample size and variations of responses, these assumptions cannot be verified statistically and should be treated as informational/qualitative data.

Also, included in the table are the days of sampling since the last HIV-negative test as well as the Fiebig stage based on the assessment of their antibody response. The day of detection of HIV-1 was used as their entry day (day 0) into the AHI cohort. In addition, eight members of their risk networks, defined as sharing needles and syringes, were recruited into this study.

We were interested in assessing the genetic complexity of the virus transmitted by IDU and in characterizing the phenotypic features of the viral Env protein encoded in these transmitted viruses. To accomplish these goals, we performed SGA of viral RNA isolated from longitudinal plasma samples of seven AHI subjects and the eight members of their networks. The number of *env* gene amplicons obtained for each plasma sample is presented in Table 1. All the viral strains obtained in this study belonged to the subtype A virus lineage specific to the Former Soviet Union countries.

A phylogenetic analysis of a sampling of all of the *env* gene sequences is shown in Figure 1. This analysis revealed three transmission clusters: (1) AHI subject AHI_D with partners CHR_F and CHR_G, although the reported infected partners CHR_E, AHI_H, CHR_I, and CHR_J had viruses unrelated to this cluster (by phylogenetic analysis); (2) AHI subjects AHI_A and AHI_B with partner CHR_C; and (3) AHI

Table 1. Study Subjects and Single Genome Amplification Data Summary

								SGA	
Subject ID	*Risk*	*Status*	*Time point*	*Days after neg visit*	*Days post enrollment*	*Fiebig stage*	*Amplicons (N)*	*Poisson fitter (CI)*	*Star phylogeny*
AHI_A	IDU	AHI	1	32	0	III/IV	29	26 (11, 40)	y
			2	39	7	III/IV	n/a	n/a	n/a
			3	46	14	III/IV	16	36 (27, 46)	y
			4	53	21	V	16	34 (21, 46)	y
			5	63	31	V	29	46 (39, 54)	y
AHI_B	IDU>sex	AHI	1	18	0	I/II	20	15 (9, 21)	y
			2	28	10	III/IV	22	25 (16, 35)	y
			3	67	49	V	16	39 (29, 49)	y
			4	128	110	V	8	n/a	n/a
			5	246	228	Est	18	n/a	n/a
CHR_C	n/a	P	1	n/a	n/a	Est	5	n/a	n/a
AHI_D	IDU>sex	AHI	1	0[a]	0	I/II	11	17 (9, 25)	y
			2	7	7	III/IV	17	24 (15, 33)	y
			3	14	14	V	17	n/a	n/a
			4	21	21	V	19	n/a	n/a
			5	28	28	V	12	n/a	n/a
			6	279	279	Est	25	n/a	n/a
CHR_E	n/a	P	1	n/a	n/a	Est	17	n/a	n/a
CHR_F	n/a	P	1	n/a	n/a	Est	8	n/a	n/a
CHR_G	n/a	P	1	n/a	n/a	Est	15	n/a	n/a
CHR_H	n/a	P	1	n/a	n/a	Est	15	n/a	n/a
CHR_I	n/a	P	1	n/a	n/a	Est	13	n/a	n/a
CHR_J	n/a	P	1	n/a	n/a	Est	11	n/a	n/a
AHI_K	IDU or sex	AHI	1	29	0	III/IV	28	26 (11, 42)	y
			2	36	7	III/IV	39	36 (29, 44)	y
			3	43	14	V	17	39 (29, 48)	y
AHI_L	n/a	P	1	n/a	n/a	Est	16	n/a	n/a
AHI_M	IDU	AHI	1	68	0	III/IV	19	44 (29, 59)	n
			2	75	7	III/IV	18	40 (29, 52)	n
			3	82	14	V	8	65 (36, 94)	n
AHI_N	IDU or sex	AHI	1	35	0	III/IV	41	45 (39, 52)	n
			2	43	8	III/IV	17	52 (41, 64)	n
			3	50	15	V	19	61 (49, 73)	n
AHI_O	IDU>sex	AHI	1	0[a]	0	I/II	15	13 (4, 21)	y
			2	7	7	I/II	23	15 (7, 22)	y

[a]Identified as AHI at enrollment.

AHI, acute HIV-1 infection; CI, 95% confidence interval; IDU, injecting drug use; n/a, not available; SGA, single genome amplification.

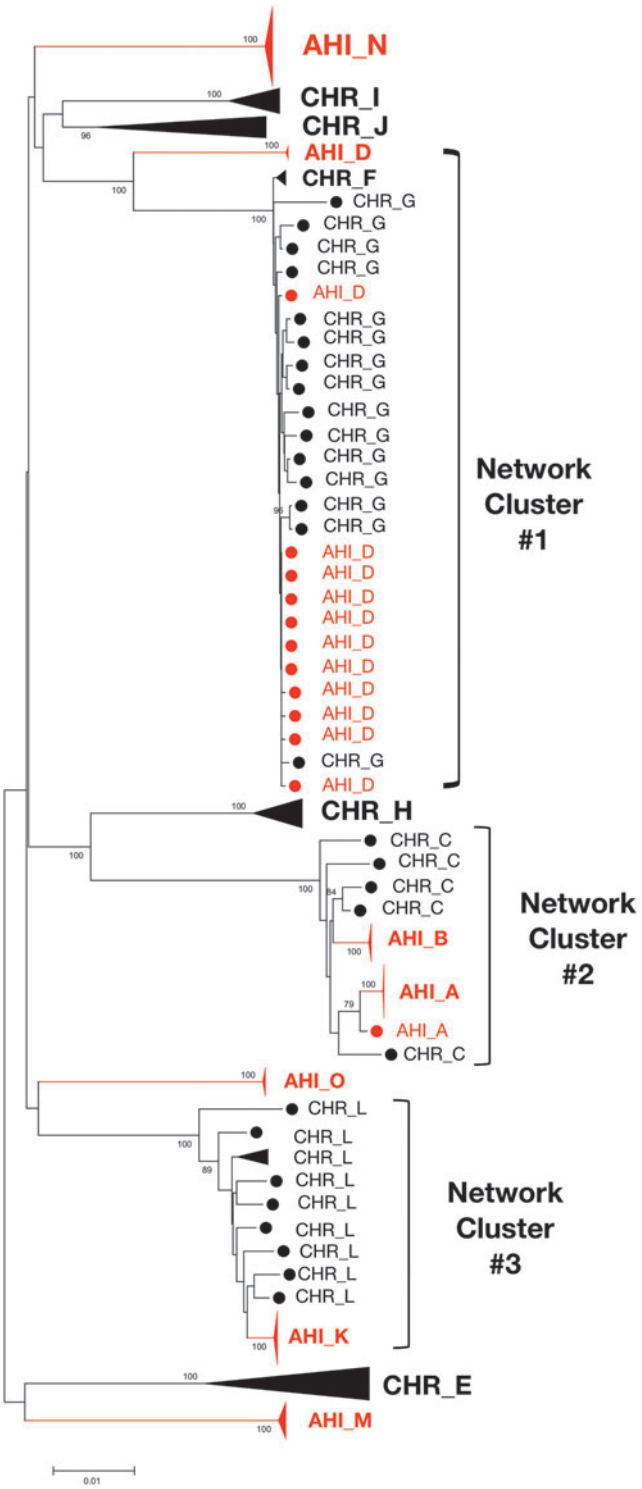

FIG. 1. Phylogenetic analysis of full-length env genes obtained by SGA from the transmitted (*red*) and chronic (*black*) virus shows the transmission clusters and a genetic bottleneck among AHI subjects. *Brackets* indicate the transmission clusters. *Triangles* indicate the phylogenetic clusters of diverse yet related sequences collapsed for visual convenience; the length of the *triangles* is proportional to the genetic diversity within the cluster. AHI, acute HIV-1 infection; SGA, single genome amplification.

subject AHI_K and partner CHR_L. AHI subjects AHI_M, AHI_N, and AHI_O each appeared as a distinct phylogenetic lineage. The features of the viral sequences in the three transmission clusters are described below.

Network cluster 1 contained sequences from AHI subject AHI_D and the sequences related to two of his PWID network members CHR_F and CHR_G, who had established HIV-1 infection as assessed by Western blot results. The viral population of AHI_D was homogeneous at the time of initial entry screening and 7 days after entry into the study, in each case giving rise to a star-like phylogeny. The presence of a homogeneous population is consistent with the transmission of a single viral genome in the transmission event. The presence of a largely homogeneous population early after transmission allowed the use of Poisson Fitter to estimate the days postinfection, which, after elimination of sequences, likely to have been hypermutated by APOBEC3G/F, was estimated to be 17 days, consistent with the Fiebig Stage I/II designation based on antibody assessment (Table 1).

The phylogenetic analysis and highlighter plot of full-length *env* sequences from the entry time point of AHI_D indicated that the founder virus of AHI_D was most closely related to the viruses in CHR_G. Even though CHR_G was WB positive at enrollment and was classified as chronically infected, the viral population of CHR_G, as seen in the highlighter plot (Supplementary Fig. S1A), was relatively homogeneous. The viral population of CHR_F was also related to AHI_D, although not as close as to CHR_G, and homogeneous, indicating these were likely recent infections within the AHI_D cluster from a common source.

Surprisingly, at 14 days after study entry, the viral population of AHI_D revealed a second largely homogeneous variant that was distinct from the entry founder virus and could not be attributed to contamination from any sample tested in this study. We interpret this pattern as a superinfection event during the acute infection period. The plasma samples from day 21 after study entry and later revealed extensive recombination between the initial founder virus and the superinfecting variant; however, we cannot exclude the possibility that a third variant was introduced during these later time periods, which appeared as recombinants with the initial two viruses.

Network cluster 2 included 2 AHI subjects, AHI_A and his network and sexual partner AHI_B, as well as their PWID network member CHR_C, who had an established infection. Subject AHI_A was considered AHI, given a visit 30 days previously where he was assessed as HIV negative. We detected a largely homogeneous viral population (Supplementary Fig. S1B), and analysis of the diversity with the Poisson Fitter program gave an estimate of 26 days postinfection for this subject, consistent with his Fiebig III/IV designation. The homogeneity of the population is consistent with the transmission of a single variant.

Subject AHI_B tested positive for HIV-1 at screening, and her virus was highly homogeneous, consistent with a recent transmission event of a single virus variant. Poisson Fitter analysis suggested a population age of just 15 days, consistent with the Fiebig I/II designation.

The viral population of network partner CHR_C was diverse and related to both AHI_B and AHI_A, suggesting CHR_C as a possible source of infection for both AHI_B and AHI_A (or an equivalent partner with a similar virus with

similar diversity). The placement of these sequences within the phylogenetic tree (Fig. 1) indicates that the virus was not passed directly between AHI_A and AHI_B.

Network cluster 3 included AHI subject AHI_K and the chronically infected network partner CHR_L (Fig. 1). The viral population of AHI_K was homogeneous at entry (Supplementary Fig. S2A) with the Poisson Fitter analysis suggesting days postinfection at 26, consistent with the Fiebig III/IV designation from antibody analysis (Table 1). Again, the homogeneity of this population was consistent with a single variant being transmitted. In contrast, the viral population in CHR_L was much more diverse, indicating that this person had been infected for a much longer period of time.

We also had three individuals with AHI, which did not have any risk network members enrolled in the study: AHI_M, AHI_N, and AHI_O. The lineages for these viruses were distinct from the three network clusters and from each other (Fig. 1). Subject AHI_O was identified as being in acute infection at screening with a Fiebig I/II designation. The viral population was very homogeneous (Supplementary Fig. S2B), with the Poisson Fitter analysis suggesting an infection 15 days before screening, and the homogeneity of the viral population is again consistent with transmission of a single variant. The viral populations of AHI_M and AHI_N that appeared to be older in those single nucleotide variants (Supplementary Fig. S2C, D) were becoming fixed in the population, making the Poisson Fitter estimates of transmissions at 44 and 45 days before sampling, respectively, less accurate, but still consistent with the Fiebig stage designation of III/IV for each. However, the lack of distinct lineages within either of these entry samples is still consistent with the transmission of a single variant in each case.

In this cohort of seven individuals with AHI, all seven subjects displayed a severe transmission bottleneck, which we interpret as a single variant being transmitted in each case. In addition, it was relatively easy to identify linked transmissions within these self-identified partner networks. In one network, two of six partners had viruses that were phylogenetically linked to the AHI subject (AHI_D), and all three of the viral populations within this network were relatively young, suggesting recent linked transmission events within the network. Similarly, the network of AHI subject AHI_A included a linked chronically infected partner and a partner in acute infection, which was even more recent. The other three AHI subjects were not linked to the other members of the cohort. However, the ability to find linked infections within these networks without extensive sampling indicates they represent foci of active HIV-1 transmission.

The low variability on pairwise nucleotide diversity among transmitted isolates, as analyzed using the full-length *env* gene sequences (SGA), was consistent with the bottleneck hypothesis (Fig. 2A). The nucleotide diversity mean values with range for transmitted isolates were 0.0007 (0.0004–0.0013), while for chronic isolates, they were much higher—0.01345 (0.0023–0.0326), $p = .0003$ by Mann–Whitney test.

Nucleotide substitution rate varies among transmitted isolates

We analyzed the nucleotide substitution rate of the transmitted virus using the BEAST software package, employing tip-dating by incorporating into the presets the days postinfection according to the Fiebig stage, or alternatively using a consensus sequence of the HIV-1 strain created for each subject based on their viral population present at the earliest available time point of infection. The viral population of AHI_D was excluded from the analysis due to the multiple variants present.

The results for the AHI subjects show some differences in the substitution rates (Fig. 2B). The estimated substitution rate values fluctuate between 2×10^{-5} to 7×10^{-5}. There are several reasons why there may be such a range in apparent substitution rates. First, the different founder viruses may vary in the rate of misincorporation by reverse transcriptase. Second, very early immune responses may provide diversifying selective pressure; these early responses could be due to a higher state of immune activation in some individuals, reactivity to fortuitous cross-reactive antigens, or the presence of low-level cytotoxic T lymphocyte (CTL) activity due to repeated exposure to HIV-1 before establishing a systemic infection (these drivers of early diversity are not mutually exclusive). Finally, repeated superinfection from the same donor who is in an acute or early stage of HIV infection is another possibility for adding diversity to a population that would mimic an apparent high substitution rate.

Primer ID deep sequencing results reveal few minor variants in the pool of transmitted viruses

The samples previously analyzed by SGA were sequenced using the Primer ID NGS approach that corrects the error rate and quantifies the sampling depth for the detection of minor variants. We chose to sequence a fragment of the *env* gene that spans the region from C1-V1-V2-C2, although the paired-end reads do not quite meet. The use of Primer ID allows each RNA template to be evaluated independently as a distinct sequence even after the PCR step, avoiding the problem of PCR resampling. The goal was to determine if we could identify any minor viral variants not sampled by SGA, which might be present at a lower level than the founder virus.

For the subject AHI_D, we were able to detect the second transmitted variant within 2 weeks postscreening, compared to 3 weeks by SGA. We also performed NGS analysis of his PWID network members CHR_E, CHR_G, CHR_I, and CHR_J (Fig. 3A). No sample was available for analysis from subject CHR_F. We detected the same founder variant that belonged to CHR_G group, and a superinfection with the variant that belonged to an unidentified subject, consistent with what was observed by SGA.

In the network cluster 2, the virus from both AHI subjects AHI_A and AHI_B cluster with different variants presents in CHR_C more than with each other, indicating that CHR_C, either directly or indirectly, was a donor for both of these subjects (Fig. 3B).

In the network cluster 3, deep sequencing allowed us to find a variant in the donor CHR_L that was identical to the founder virus of AHI_K over the shorter NGS amplicon (Fig. 3C). The phylogenetic tree also revealed a very minor group of AHI_K sequences that mixed with another cluster of CHR_L, indicating a possible minor variant in the viral population of AHI_K due either to infection with two viruses or, given its low abundance, introduction of a second variant from the same donor.

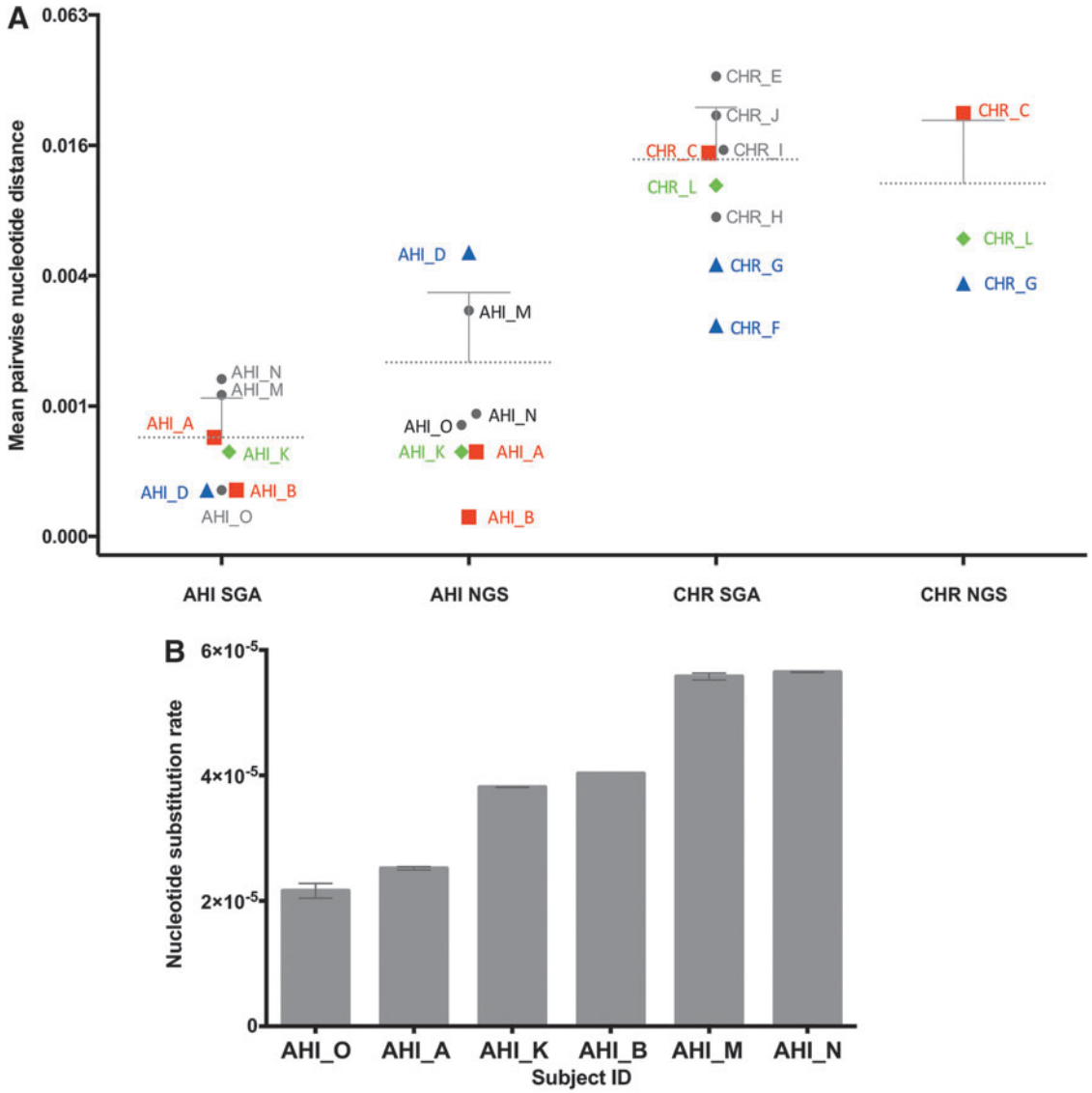

FIG. 2. Assessment of nucleotide diversity and viral evolution among AHI and chronic (CHR) subjects. **(A)** Mean pairwise nucleotide distance with standard deviation for either full-length *env* gene (AHI) or its fragment (NGS). The values are shown for members of the network cluster 1 (*blue*), network cluster 2 (*red*), and network cluster 3 (*green*), as well as for the subjects with unrelated virus (gray). **(B)** Estimation of nucleotide substitution rates of AHI-derived full-length *env* amplicons by BEAST. We used lognormal type of molecular clock for tip-dating of the viral variants by setting a consensus sequence set as point 0. The order of the subjects on *X*-axis is determined by the increase in nucleotide substitution rate. NGS, next-generation sequencing.

The NGS analysis of the other study subjects revealed that AHI_M had closely related viral variants that could be due to multiple variants or multiple events of superinfection, but from a person in early infection, or, more likely, the early diversification of a single transmitted variant in this somewhat later AHI case (Fig. 3D). Similarly, the viral populations of AHI_O and AHI_N showed the single viral variant for each (Fig. 3D).

Using the Primer ID NGS approach, we characterized viral populations of AHI subjects and their partners. We were able to detect an additional donor for AHI_D not sampled at an earlier time point by SGA. However, overall, the conclusions about population diversity drawn from the SGA analysis were confirmed using the deep sequencing approach, specifically that transmission in these IDU cases is dominated by a severe genetic bottleneck that most often results in the infection being established by a single variant.

Phenotypic analysis of transmitted strains shows predominantly R5 T cell-tropic strains among AHI subjects

We used the full-length *env* genes of the isolates from the subjects of this and our previously published study of PWID in St. Petersburg[14] to create a panel of *env* genes from selected SGA amplicons cloned into an expression vector representing 11 different subjects in acute infection. The expression vectors were used to create pseudotyped viruses encoding a reporter gene (luciferase) that was used to test for infection of Affinofile cells. We first assessed coreceptor specificity by carrying out

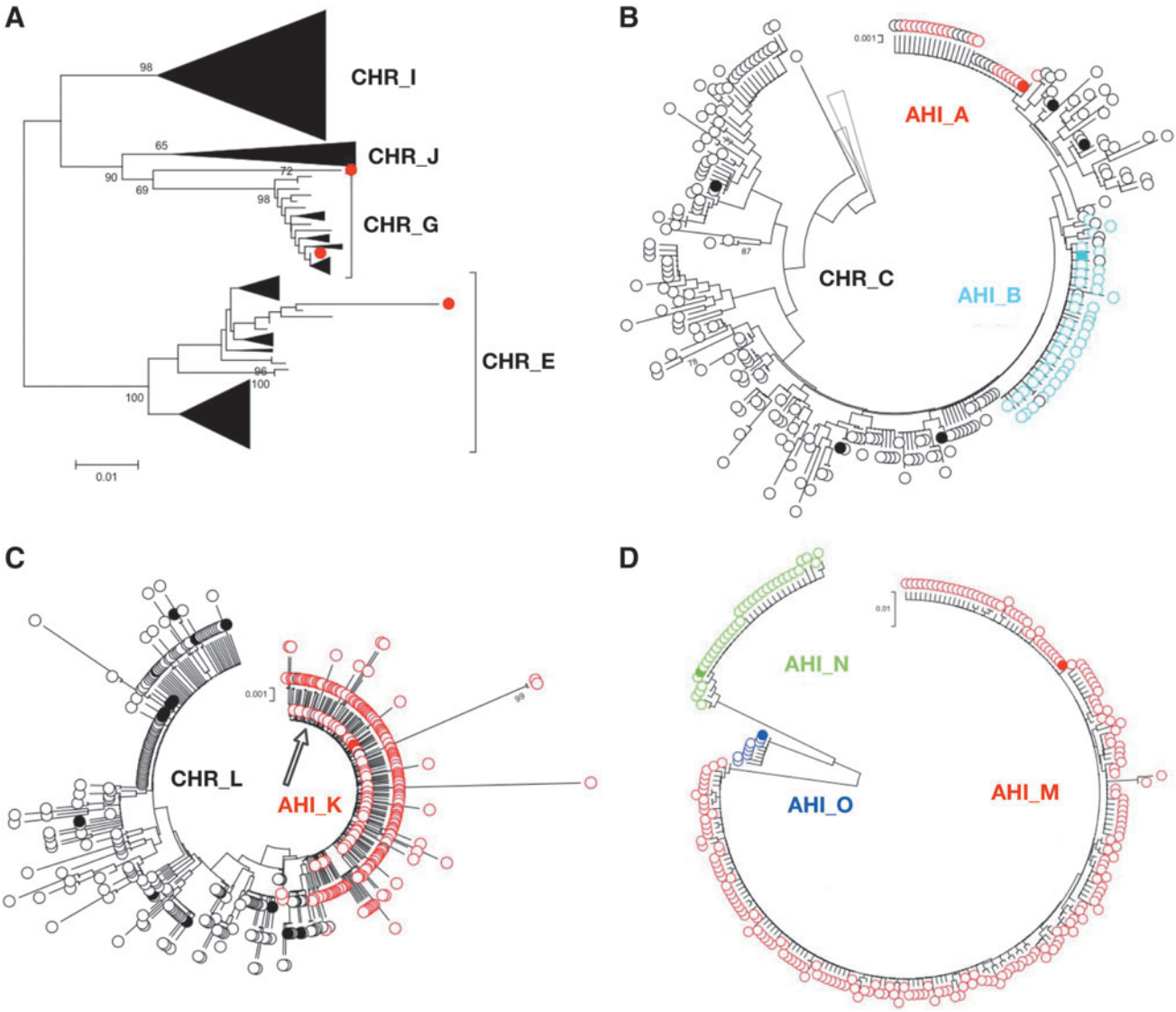

FIG. 3. Phylogenetic analysis of the deep sequencing data (TCS) obtained for the clusters and unrelated AHI cases reaffirm a single-variant transmission in all the AHI cases. Collapsed TCS are labeled with *open circles* and SGA-derived variants placed on the same tree for the reference are labeled with *closed circles*. *Different colors* indicate the AHI subjects, *black* indicates chronic subjects. **(A)** Network cluster 1 and other PWID related to AHI_D by risk behavior, with AHI_D viral variants *highlighted* in *red closed circles*. **(B)** Network Cluster 2. **(C)** Network Cluster 3, with the *arrow* pointing to the identical sequences of AHI_K (*red*) and CHR_L (*black*). **(D)** Three unrelated AHI subjects that were placed on the same tree for visual convenience. PWID, people who inject drugs; TCS, template consensus sequence.

the infections with or without a CCR5 antagonist and/or a CXCR4 antagonist. As can be seen in Figure 4A, all of the *env* genes from viruses in acute infection encoded proteins that used CCR5 for entry and not CXCR4, as is typically seen in sexual transmission. One of the *env* genes cloned from a participant in chronic infection (SC1457) used CXCR4 exclusively for entry.

We next tested the *env* clones for entry phenotype with respect to whether they required a high density of CD4 for efficient entry (T cell tropic) or if they were variants that could use a low density of CD4 on the cell surface for efficient entry (macrophage tropic). Shown in Figure 4B are the results indicating that all the *env* genes from both the acute and chronic infection samples encoded proteins that failed to mediate efficient entry at a low density of CD4, thus making all the *env* genes tested T cell tropic, again similar to that seen in sexual transmission.

Neutralization sensitivity assays of chronic and acute subtype A isolates, and comparing subtype A and subtype C isolates

We tested the neutralization sensitivity of the pseudotyped viruses derived from transmitted and chronic isolates to a panel of HIV-specific antibodies. The panel included 4 polyclonal serum samples from HIV-infected individuals of subtype C, HIVIG-C from pooled sera, soluble CD4, and the monoclonal antibodies 4E10, PG9, and CH01-31.

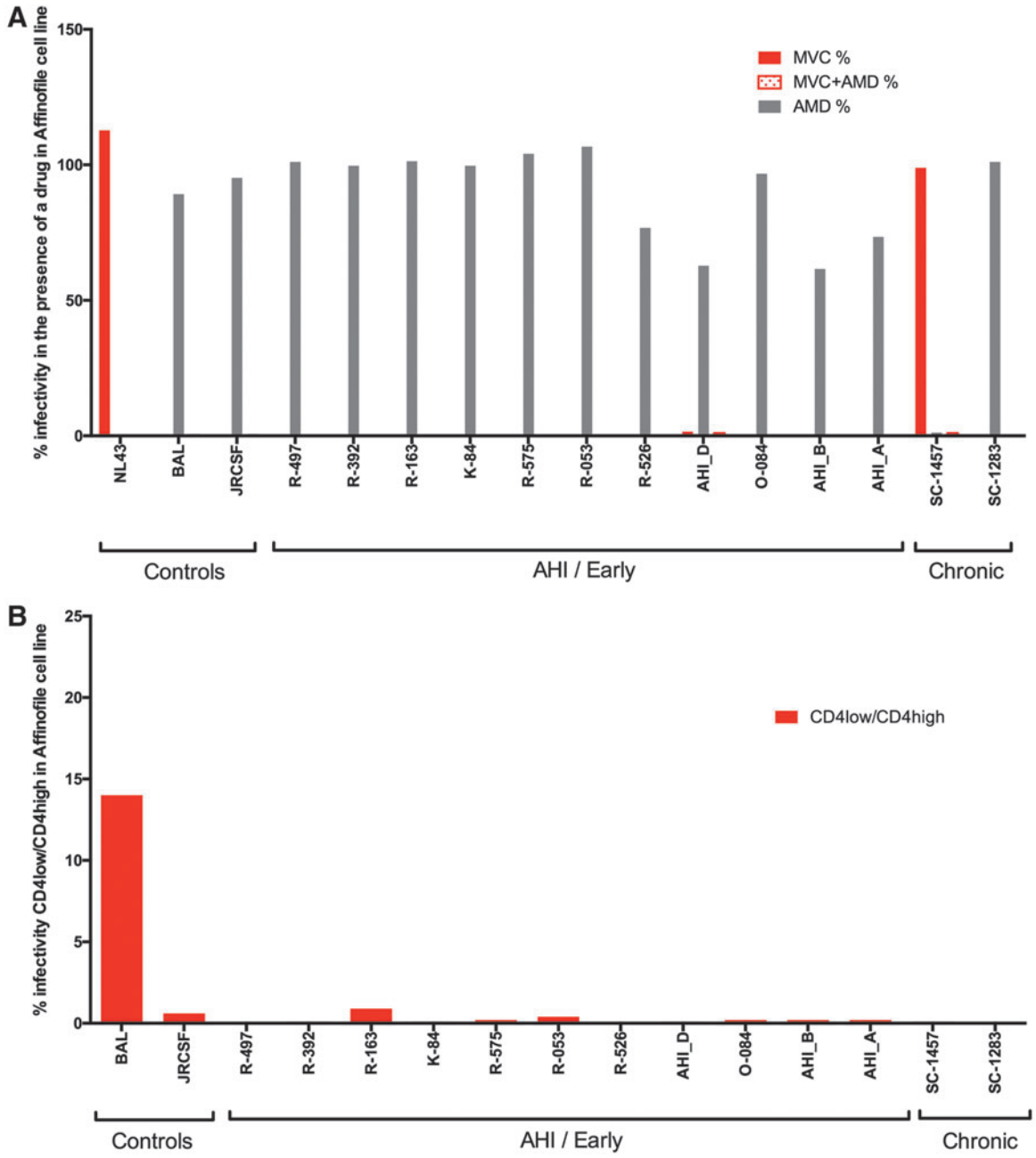

FIG. 4. Affinofile cell culture assay of HIV-1 subtype A pseudoviruses shows little variability between acute and chronic isolates. **(A)** CXCR4 virus detected in a chronically infected subject. **(B)** All the viruses tested in the assay had a T cell-tropic phenotype. MVC indicates maraviroc, a blocking agent for CCR5 coreceptor, while AMD stands for AMD3100, a blocking agent for CXCR4 coreceptor.

A hierarchical clustering analysis revealed three groups of subtype A viruses that differed in their neutralization pattern (Supplementary Fig. S3). Group 1 consisted of seven viruses predominantly derived from acutely or recently infected subjects (top of the heat map); Group 3 contained three isolates from chronic and two from acute subjects (bottom of the heat map, five viruses); and finally, Group 2 was equally split between both chronic and acute isolates (central portion of the heat map, six viruses). At present, we do not know the basis for this clustering, but it is clear that the AHI viruses are distributed throughout the three groups, arguing there is no specific phenotype for the transmitted variants, at least as assessed with these reagents.

Because the introduction of a subtype A virus into the IDU populations of the former Soviet Union countries happened more recently than the establishment of the subtype C epidemic in southern Africa[11] and the subtype B epidemic in the United States,[10] the viral population in the subtype A epidemic is less diverse. This is apparent comparing the sequences of the full-length *env* genes collected from subjects in acute infection of these three geographical regions approximately in the same period (Supplementary Fig. S4). We were interested in testing the hypothesis that the neutralization properties would be more homogeneous in the more genetically homogeneous viral population, that is, subtype A.

We compared previously published data[21] that focused on subtype C isolates from individuals with a predominantly heterosexual risk of acquisition versus our data set of subtype A isolates.

The two main variables of these probability distributions (mean and variance) were statistically compared and the corresponding *p* values are listed in Supplementary Table S2. Specifically, we were interested in comparing isolates from acute infection between the two groups, and isolates from chronic infection between the two groups. Chronic subjects of subtype A and C differed in both variance and mean of the neutralization sensitivity to 4E10, with the subtype A probability distribution being less compact and with a lower mean. For HIVIG-C, subtype A pseudoviruses had a lower mean compared with subtype C for both transmitted and chronic isolates; for sCD4, the mean value was higher for acute virus of subtype A versus subtype C pseudoviruses. However, there was no consistent pattern of difference in variance between subtype A and C viruses, despite the difference in genetic heterogeneity, and the differences that were observed would not be significant after accounting for multiple comparisons.

To increase the power of our analyses, we compared subtype A and subtype C clones, pooling AHI and chronic isolates, assuming that the similarities within subtypes would be larger than the putative differences between subtypes. Most of the trends (of the means) in the AHI and chronic comparisons remained, but we did not see less variance for the subtype A isolates compared to the genetically more diverse subtype C isolates. We conclude that the bottleneck of introduction of the subtype A virus into eastern Europe, and specifically into St. Petersburg, happened sufficiently long ago to obscure phenotypic similarities in neutralization properties.

Discussion

We analyzed the cell-free HIV-1 subtype A viral populations of PWID that had a very high risk of injection-related HIV-1 infection. A viral population that originates from a single viral variant, as an extreme genetic bottleneck, in the absence of selective pressure should evolve in a so-called ''star phylogeny'' pattern. The viral sequence data obtained by SGA and analyzed with the Poisson Fitter model confirmed a star phylogeny pattern for five out of seven AHI subjects and indicated the possibility of multiple consecutive infections within a PWID network, specifically in the case of the subject AHI_D. In the failed Poisson Fitter model (subjects AHI_M and AHI_N), the viral population contained minor variants that were genetically very similar to the founder virus, likely coming from the same donor (if they were in acute infection) or representing the initial fixation of a few mutations from a single transmission event; given this interpretation, this would indicate that all seven AHI subjects identified in this study were each infected with a single variant.

There are several scenarios that could increase early diversity in a newly infected person. In the population with high HIV-1 incidence like in St. Petersburg, and PWID networks where multiple members are infected, frequent exposure to the virus before establishing a systemic infection could result to early CTL escape mutations in the recipients. Alternatively, superinfection events from the same donor due to frequent contacts through drug use or sex may also lead to additional early diversity, but given the high similarity of the variants, this would only be the case if the donor was also in a very early stage of infection; multiple variants from a donor in chronic infection would appear much more distinct.

We performed Primer ID deep sequencing to test the hypothesis that at the earliest stages of AHI, there are multiple viruses that later are outgrown by the most fit variant. Our data reject this hypothesis and confirm the presence of a very strong bottleneck in AHI subjects. In a previously published cross-sectional study using SGA, we observed a genetic bottleneck in 9 out of 13 subjects with acute and recent HIV-1 infection.[14] In this longitudinal study combining the data obtained by SGA and NGS, we interpret the sum of the data as being most consistent with transmission of a single variant in each case. Combining these data with the previously described findings, we see that 16 out of 20 (80%) PWID individuals infected with HIV-1 subtype A had a severe genetic bottleneck during the HIV-1 transmission event.

This phenomenon is similar to that seen in heterosexual transmission and provides a framework for considering transmission under these circumstances. It appears to distinguish a model where there is a high probability of transmission because of virus-containing blood where multiple variants would be expected to be transmitted from a model where the amount of exposure to infectious material is much less, such that when infection does occur, it does so with a lower probability as evidenced by the presence, most often, of a single variant.

If the genetic bottleneck is due to biological selection of the fittest virus, we might expect to see differences in the phenotypic properties of Env protein derived from AHI compared to chronically infected subjects. The results did not reveal phenotypic differences in transmitted versus chronically circulating viruses of IDU subtype A, nor any macrophage-tropic virus present in our study group, so all the transmitted isolates and all but one chronic isolate belonged to CCR5-dependent T cell-tropic class of viruses. Thus parenteral transmission does not select strongly for a specific phenotype, at least by the measures used, and the transmitted variants appear similar to those that are typically found in the blood of chronically infected people. However, the sample size in this study was relatively small and therefore not able to detect differences of smaller magnitude, if they exist.

We assessed the neutralization sensitivity of the pseudoviruses to a panel of anti-HIV-1 sera and antibodies and compared the data within the panel and with the published HIV-1 subtype C data.[21] The hierarchical clustering analysis identified three clusters of the clones: one derived from mostly acute/recent infection cases, and two others derived from a mix of both chronic and acute cases. Although the nature of the higher sensitivity of viruses in the group (1) is unclear, they may be useful to further study as the candidate strains for the vaccine design against the transmitted isolates of subtype A. However, we were unable to generate evidence that the more recent nature of this subtype A epidemic retains less diversity in its neutralization properties, despite the fact that there is less diversity in the genetic population relative to older portions of the HIV-1 epidemic.

When we compare the results of AHI studies in the world, it should be noted that the members of the cohorts in St. Petersburg, Bangkok, and Montreal used different types of

drugs and manifested different risk behavior. The members of St. Petersburg and Bangkok cohort reported heroin use, whereas the participants of Montreal cohort reported injection of cocaine. Injection of cocaine may imply more frequent drug administration, thus increasing the concomitant infection risks of needle sharing and sexual activity.[23] So the high multiplicity of the infection reported in the Montreal cohort may be associated with the multiple viral transmission events from the same or different donors.

An epidemiological study of Russian PWID networks published earlier suggests that the spread of HIV-1 infection among PWID in Russia happens most often within the small clusters or networks of drug users who share unsafe injection practices.[24] In our study, we detected three transmission clusters, and our analysis of the viral diversity in PWID networks along with the epidemiological data gathered in the previous studies suggests that the HIV-1 prevention measures focused on PWID networks would be one of the most efficient ways to control the epidemic.

Russia has seen an unprecedented growth of HIV-1 transmission among PWID since the onset of the epidemic, with extremely high incidence rates in St. Petersburg and other cities that have a large population of PWID. It is known that primary HIV-1 infection is characterized by an extremely high blood viral load and proportionally high probability of transmission. The genetic bottleneck and an overall low heterogeneity of subtype A HIV-1 in Russia may be a direct consequence of a scenario when a low-complexity viral population is getting transmitted, and the viral spread in a limited space of a drug use cluster with frequent injections is fast and efficient. In this case, the multiple-variant transmission could happen before selective pressure in the donor, which could lead to an underestimate of the average complexity of the transmitted virus.

So far, there are little to no preventive measure in Russia that focus on reducing the risk of spread from acutely infected individuals, and we show that such individuals are widespread within the networks of active drug users. The clinical testing requirements in the high-risk groups, hospitals, and blood banks should be expanded to include HIV-1 RNA or p24 assessment that would allow a more accurate and timely diagnosis of HIV-1 infection and subsequent preventive measures for the at-risk network members.

Acknowledgments

This study was supported by NIH grant P30 A150410 UNC Center for AIDS Research, and agreement between the UNC and the Biomedical Center; E.D. was partially supported through IAS-NIDA Postdoctoral Research Fellowships in HIV and Drug Use; the data analysis and publication were partially supported by the Grant of the Russian Science Foundation #15-14-00026.

Author Disclosure Statement

UNC is pursuing IP protection for Primer ID, and R.S. is listed as a co-inventor and has received nominal royalties. Presently, E.D. is employed with Janssen Pharmaceutical Companies of Johnson and Johnson. No competing financial interests exist for the other authors.

References

1. Kozlov AP, Volkova GV, Malykh AG, Stepanova GS, Glebov AV: Epidemiology of HIV infection in St. Petersburg, Russia. J Acquir Immune Defic Syndr 1993;6:208–212.
2. Bobkov A, Cheingsong-Popov R, Selimova L, *et al.*: An HIV type 1 epidemic among injecting drug users in the former Soviet Union caused by a homogeneous subtype A strain. AIDS Res Hum Retroviruses 1997;13:1195–1201.
3. Beyrer C, Wirtz AL, O'Hara G, Léon N, Kazatchkine M: The expanding epidemic of HIV-1 in the Russian Federation. PLoS Med 2017;14:e1002462.
4. Kozlov AP, Shaboltas AV, Toussova OV, *et al.*: HIV incidence and factors associated with HIV acquisition among injection drug users in St Petersburg, Russia. AIDS 2006; 20:901–906.
5. Dukhovlinova EN, Masharsky A, Toussova O, *et al.*: Two independent HIV epidemics in Saint Petersburg, Russia revealed by molecular epidemiology. AIDS Res Hum Retroviruses 2014:150127063121001.
6. Toussova O, Shcherbakova I, Volkova G, Niccolai L, Heimer R, Kozlov A: Potential bridges of heterosexual HIV transmission from drug users to the general population in St. Petersburg, Russia: Is it easy to be a young female? J Urban Health 2009;86:121–130.
7. Rutstein SE, Sellers CJ, Ananworanich J, Cohen MS: The HIV treatment cascade in acutely infected people: Informing global guidelines. Curr Opin HIV AIDS 2015;10: 395–402.
8. Ronen K, Sharma A, Overbaugh J: HIV transmission biology: Translation for HIV prevention. AIDS 2015;29:2219–2227.
9. Rademeyer C, Korber B, Seaman MS, *et al.*: Features of recently transmitted HIV-1 Clade C viruses that impact antibody recognition: Implications for active and passive immunization. PLOS Pathog 2016;12:e1005742.
10. Keele BF, Giorgi EE, Salazar-Gonzalez JF, *et al.*: Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 2008;105:7552–7557.
11. Abrahams MR, Anderson JA, Giorgi EE, *et al.*: Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J Virol 2009;83:3556–3567.
12. Carlson JM, Schaefer M, Monaco DC, *et al.* Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 2014;345:1254031–1254031.
13. Carlson JM, Du VY, Pfeifer N, *et al.*: Impact of preadapted HIV transmission. Nat Med 2016;22:606–613.
14. Masharsky AE, Dukhovlinova EN, Verevochkin SV, *et al.*: A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia. J Infect Dis 2010;201:1697–1702.
15. Sterrett S, Learn GH, Edlefsen PT, *et al.*: Low multiplicity of HIV-1 infection and no vaccine enhancement in VAX003 injection drug users. Open Forum Infect Dis 2014;1:ofu056.
16. Bar KJ, Li H, Chamberland A, *et al.*: Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol 2010;84:6241–6247.
17. Toussova O, Kozlov A, Verevochkin S, *et al.*: A Cohort approach to real time detection of acute HIV infections among people who inject drugs in St. Petersburg, Russia. AIDS Res Human Retroviruses 2017;AID.2017.0076.
18. Salazar-Gonzalez JF, Bailes E, Pham KT, *et al.*: Deciphering human immunodeficiency virus type 1 transmission

and early envelope diversification by single-genome amplification and sequencing. J Virol 2008;82:3952–3970.
19. Zhou S, Jones C, Mieczkowski P, Swanstrom R: Primer ID validates template sampling depth and greatly reduces the error rate of next generation sequencing of HIV-1 genomic RNA populations. J Virol 2015;JVI.00522-15-52.
20. Montefiori DC: Measuring HIV neutralization in a luciferase reporter gene assay. Methods Mol Biol 2009;485: 395–405.
21. Ping L-H, Joseph SB, Anderson JA, *et al.*: Comparison of viral Env proteins from acute and chronic infections with subtype C human immunodeficiency virus type 1 identifies differences in glycosylation and CCR5 utilization and suggests a new strategy for immunogen design. J Virol 2013;87: 7218–7233.
22. Joseph SB, Lee B, Swanstrom R: Affinofile Assay for Identifying Macrophage-Tropic HIV-1—BIO-PROTOCOL. Available at bio-protocolorg, accessed July 10, 2018.
23. Bruneau J, Daniel M, Abrahamowicz M, *et al.*: Trends in human immunodeficiency virus incidence and risk behavior among injection drug users in Montreal, Canada: A 16-year longitudinal study. Am J Epidemiol 2011;173: 1049–1058.
24. Hoffman IF, Latkin CA, Kukhareva PV, *et al.*: A peer-educator network HIV prevention intervention among injection drug users: Results of a randomized controlled trial in St. Petersburg, Russia. AIDS Behav 2013;17:2510–2520.

Address correspondence to:
Ronald Swanstrom
Lineberger Comprehensive Cancer Center
University of North Carolina at Chapel Hill
450 West Drive, Room 22-006
Chapel Hill, NC 27599-7260

E-mail: risunc@med.unc.edu