RESEARCH ARTICLE

Characterization of HIV diversity, phylodynamics and drug resistance in Washington, DC

Marcos Pérez-Losada[1,2,3]*, Amanda D. Castel[3], Brittany Lewis[3], Michael Kharfen[4], Charles P. Cartwright[5], Bruce Huang[1], Taylor Maxwell[1], Alan E. Greenberg[3], Keith A. Crandall[1], on behalf of the DC Cohort Executive Committee¶

1 Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Ashburn, VA, United States of America, **2** CIBIO-InBIO, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal, **3** Milken Institute School of Public Health, Department of Epidemiology and Biostatistics, The George Washington University, Washington, DC, United States of America, **4** District of Columbia Department of Health, Washington, DC, United States of America, **5** Laboratory Corporation of America, Burlington, NC, United States of America

¶ The complete membership of the author group can be found in the Acknowledgments.
* mlosada@gwu.edu

Citation: Pérez-Losada M, Castel AD, Lewis B, Kharfen M, Cartwright CP, Huang B, et al. (2017) Characterization of HIV diversity, phylodynamics and drug resistance in Washington, DC. PLoS ONE 12(9): e0185644. https://doi.org/10.1371/journal.pone.0185644

Editor: Dimitrios Paraskevis, National and Kapodistrian University of Athens, GREECE

Received: April 7, 2017

Accepted: September 16, 2017

Published: September 29, 2017

Data Availability Statement: All sequence files are available from the GenBank database (accession number(s) MF455515 – MF457397).

Funding: This study was partially supported by a DC D-CFAR pilot award, USA NIH CCNS214443F and NIH UL1TR000075 awards from the NIH National Center for Advancing Translational Sciences, the DC Cohort Study (UO1 AI69503-03S2), a supplement from the Women's Interagency Study for HIV (410722_GR410708), and a 2015 HIV Phylodynamics Supplement award

Abstract

Background

Washington DC has a high burden of HIV with a 2.0% HIV prevalence. The city is a national and international hub potentially containing a broad diversity of HIV variants; yet few sequences from DC are available on GenBank to assess the evolutionary history of HIV in the US capital. Towards this general goal, here we analyze extensive sequence data and investigate HIV diversity, phylodynamics, and drug resistant mutations (DRM) in DC.

Methods

Molecular HIV-1 sequences were collected from participants infected through 2015 as part of the DC Cohort, a longitudinal observational study of HIV+ patients receiving care at 13 DC clinics. Sequences were paired with Cohort demographic, risk, and clinical data and analyzed using maximum likelihood, Bayesian and coalescent approaches of phylogenetic, network and population genetic inference. We analyzed 601 sequences from 223 participants for *int* (~864 bp) and 2,810 sequences from 1,659 participants for *PR/RT* (~1497 bp).

Results

Ninety-nine and 94% of the *int* and *PR/RT* sequences, respectively, were identified as subtype B, with 14 non-B subtypes also detected. Phylodynamic analyses of US born infected individuals showed that HIV population size varied little over time with no significant decline in diversity. Phylogenetic analyses grouped 13.5% of the *int* sequences into 14 clusters of 2 or 3 sequences, and 39.0% of the *PR/RT* sequences into 203 clusters of 2–32 sequences. Network analyses grouped 3.6% of the *int* sequences into 4 clusters of 2 sequences, and 10.6% of the *PR/RT* sequences into 76 clusters of 2–7 sequences. All network clusters

from the District of Columbia Center for AIDS Research, an NIH funded program (P30AI117970), which is supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, NIGMS, NIDDK, and OAR. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health. CPC has a commercial affiliation with Laboratory Corporation of America. The Laboratory Corporation of America provided support in the form of salaries for author CPC, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section."

Competing interests: CPC has a commercial affiliation with Laboratory Corporation of America. The Laboratory Corporation of America provided support in the form of salaries for author CPC, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of this author are articulated in the 'author contributions' section." This does not alter our adherence to PLOS ONE policies on sharing data and materials. The other authors have declared that no competing interests exist.

were detected in our phylogenetic analyses. Higher proportions of clustered sequences were found in zip codes where HIV prevalence is highest (r = 0.607; P<0.00001). We detected a high prevalence of DRM for both *int* (17.1%) and *PR/RT* (39.1%), but only 8 *int* and 12 *PR/RT* amino acids were identified as under adaptive selection. We observed a significant (P<0.0001) association between main risk factors (men who have sex with men and heterosexuals) and genotypes in the five well-supported clusters with sufficient sample size for testing.

Discussion

Pairing molecular data with clinical and demographic data provided novel insights into HIV population dynamics in Washington, DC. Identification of populations and geographic locations where clustering occurs can inform and complement active surveillance efforts to interrupt HIV transmission.

Introduction

Washington, DC has consistently had one of the highest rates of HIV infection in the United States (US) with 371 new HIV cases reported in 2015 and a 2.0% HIV prevalence [1]. The HIV epidemic in DC is generalized: seven of the city's eight wards (geopolitical areas) have an HIV prevalence greater than one percent and men who have sex with men (MSM) and heterosexuals (HRH) account for 45% and 28% respectively of new diagnoses, with those of unknown risk accounting for 22%. The other 3% and 2% are IDU and sexual contact/IDU, respectively [1]. Measurement of the DC HIV continuum of care finds that approximately 11% of people living with HIV infection (PLWH) are estimated to be unaware of their infections [2] and only 73% of persons are continuously in care with 57% of all PLWH achieving viral suppression as of last report [1]. Given that a relatively high proportion of PLWH in care in DC have detectable viral loads (VL) [3] and that PLWH with VL of 1,500 copies/ml or higher are at increased risk for transmitting virus [4], identifying those individuals is critical to ensure they are receiving appropriate antiretroviral therapy and to curbing new infections. Furthermore, DC is a national and international city potentially containing a broad diversity of HIV variants with 11% of PLWH in DC being foreign-born [5]. Previous phylogenetic studies in the DC-Maryland region have found that 13% of people are infected with non-B-subtypes and of those, 81% were from the Maryland suburbs of DC [6]. Thus, further characterization of subtype distribution will assist in determining whether there are distinct HIV epidemics occurring in the region [6].

Started in 2011, the DC Cohort is a longitudinal observational cohort study that aims to characterize the quality of care being received among PLWH obtaining outpatient care at 13 clinic sites in DC. With over 8,000 participants enrolled as of December 2016, the Cohort provides a representative sample of the approximately 13,000 PLWH in DC who are in care [1]. The DC Cohort demographics are similar to those of all PLWH with respect to age, race/ethnicity, and sex; however, it is important to note that Cohort participants are those in care and may not reflect the care patterns of all PLWH in DC. All consented participants have de-identified data electronically exported monthly to a centralized database [7]. Data elements abstracted from sites' electronic medical records (EMRs) include demographics, HIV risk behaviors, diagnoses, laboratory tests, treatments, and procedures. An important element of the DC Cohort includes the periodic linkage of Cohort data to the DC Department of Health (DOH) HIV/AIDS Hepatitis, STD, TB Administration (HAHSTA) surveillance databases inclusive of molecular sequences from commercial laboratory reports, as well as location information such as zip code of residence. In this study, GW, DC DOH and LabCorp collaborated to further analyze these sequences beyond individual drug resistant variant calling–the only information being currently reported back to DOH from LabCorp. The objectives of these analyses were to: 1) characterize the diversity of HIV in the recent history of the DC epidemic, 2) identify drug resistant variants and sites under natural selection circulating in the DC population, 3) identify potential transmission networks, and 4) characterize associations of epidemiological factors with potential transmission networks, including geography.

Materials and methods

Ethics

Institutional Review Board (IRB071029) approval was obtained from The George Washington University IRB (which serves as the IRB of Record for eight of the participating sites), the DC DOH IRB, and the remaining site IRBs. Written informed consent was obtained and documented prior to conducting study procedures.

DC cohort

The DC DOH is one of 25 US jurisdictions funded by the Centers for Disease Control and Prevention to conduct Molecular HIV Surveillance [8]. Through this surveillance program, molecular sequence data generated by commercial laboratories is routinely sent to the DC DOH and incorporated into the surveillance database. For the purposes of this analysis, resistance data generated by LabCorp and sent to the DC DOH was incorporated into the DC Cohort database through the aforementioned linkage process. Between January 1, 2011 and March 31, 2015, 6,800 DC Cohort participants were enrolled in the study. The DC DOH received 3,411 sequences collected between 2011 and 2015 on DC Cohort participants, which were collected from 1,895 unique individuals. Using the most recently available post-DOH linked data for participants consented by March 31, 2015 and sequenced by June 15, 2015, 2,858 sequences were available representing 1,740 unique participants (i.e., sequence coverage is 25.6%). Descriptive and univariate analyses using Chi-square and Wilcoxon rank sum tests in SAS v9.3 were conducted to describe participants and examine differences between Cohort participants with and without sequence data (Table 1). The PROC FREQ function was used to perform the Chi-Square tests and PROC NPAR1WAY function was used to perform the Wilcoxon rank sum test. The sequence data were merged by participant ID with a limited set of DC Cohort demographic and clinical variables (e.g., age, race, sex, risk factor, CD4 and viral load count). The paired data were then analyzed in a de-identified manner using approaches described below.

Sequencing

Sequencing-based analyses of regions of the HIV-1 polymerase (*pol*) gene were performed by LabCorp for the detection of anti-retroviral resistance polymorphisms in regions encoding the protease (PR; codons 1–99), reverse transcriptase (RT; codons 1–400), and integrase (INT; codons 1–288) coding regions. In brief, HIV-1 RNA was recovered from plasma samples and reverse transcription followed by nested polymerase-chain-reaction (RT-PCR) performed using HIV-1 specific primer sets. Upon successful amplification, products were subjected to Sanger sequencing (BigDye® Terminator v3.1, ThermoFisher) and sequence reads analyzed

Table 1. Demographic and clinical characteristics of DC Cohort participants stratified by availability of sequence data.

	Total N[1] = 6,800	Participants Sequenced N[1] = 1,740	Participants not Sequences N[1] = 5,060	p-value[2]
Median Age (IQR)[3]	47 (36.4,54.6)	43 (31.6,50.7)	48.3 (38.5,55.7)	< .0001
Race/ethnicity[3]				
Non-Hispanic Black	5,317 (78.3)	1,486 (85.4)	3,831 (75.8)	
Non-Hispanic White	982 (14.5)	148 (8.5)	834 (16.5)	
Hispanic	334 (4.9)	77 (4.4)	257 (5.1)	< .0001
Other	146 (2.1)	26 (1.5)	120 (2.4)	
Unknown	12 (0.2)	3 (0.2)	9 (0.2)	
Sex at Birth[3]				
Male	4,938 (72.6)	1,162 (66.8)	3,776 (74.6)	
Female	1,862 (27.4)	578 (33.2)	1,284 (25.4)	< .0001
Country of Birth[3]				
US	1,117 (82.6)	128 (75.3)	989 (83.7)	
Non-US	235 (17.4)	42 (24.7)	193 (16.3)	.0070
State of Residence[3]				
DC	4,994 (73.4)	1,553 (89.2)	3,441 (68.0)	
MD	1,298 (19.1)	155 (8.9)	1,143 (22.6)	< .0001
VA	406 (6.0)	29 (1.7)	377 (7.4)	
Other	102 (1.5)	3 (0.2)	99 (2.0)	
HIV Risk Factor[3]				
MSM	3,272 (48.1)	844 (48.5)	2,428 (48.0)	
MSM/IDU	110 (1.6)	26 (1.5)	84 (1.6)	
IDU	959 (14.1)	234 (13.4)	725 (14.3)	< .0001
Heterosexual	1,961 (28.8)	555 (31.9)	1,406 (27.8)	
Other[4]	262 (3.8)	39 (2.2)	223 (4.1)	
Unknown	236 (3.5)	42 (2.4)	194 (3.8)	
Co-morbidities[3]				
Hepatitis C	820 (12.1)	115 (6.6)	705 (13.9)	< .0001
Hepatitis B	235 (3.5)	41 (2.4)	194 (3.8)	0.0036
Syphilis	260 (3.8)	84 (4.8)	176 (3.5)	0.0113
Gonorrhea	40 (0.6)	25 (1.4)	15 (0.3)	< .0001
Chlamydia	40 (0.6)	16 (0.9)	24 (0.5)	0.0362
Median Duration of Infection (yrs)(IQR)[3]	10 (5,17)	8 (2,15)	11 (6,17)	< .0001
Median CD4 count (cells/μl)(IQR)[3]	513 (323,723)	348.5 (171,534)	566 (391.5,772)	< .0001
Viral Load (copies/ml) (IQR)[3]				
<200	3109 (59.9)	88 (5.9)	3,021 (81.9)	
200–399	159 (3.1)	46 (3.1)	113 (3.1)	
400–9999	733 (14.1)	473 (31.6)	260 (7.1)	< .0001
≥10,000	1188 (22.8)	892 (59.5)	296 (8.0)	
ARV Exposure[3]				
Experienced	5,905 (86.8)	1,171 (67.3)	4,734 (93.6)	
Naïve	345 (5.1)	117 (6.7)	228 (4.5)	< .0001
Unknown	550 (8.1)	452 (26)	98 (1.9)	
ARV Regimen Type[3]				
PI-based	145 (2.4)	80 (1.8)	65 (4.2)	
NRTI-based	234 (3.8)	152 (3.3)	82 (5.3)	
NNRTI—based	13 (4.3)	7 (0.1)	6 (0.4)	< .0001
INSTI-based	56 (0.9)	42 (0.9)	14 (0.9)	

(*Continued*)

Table 1. (*Continued*)

	Total N^1 = 6,800	Participants Sequenced N^1 = 1,740	Participants not Sequences N^1 = 5,060	p-value^2
Dual-Class	3,069 (50.4)	2,270 (50.1)	799 (51.3)	
ARV Resistance^3				
At least one PI	874 (13.9)	210 (7.9)	664 (18.3)	< .0001
At least one NRTI	1,687 (26.8)	537 (20.2)	1,150 (31.6)	< .0001
At least one NNRTI	1,645 (26.1)	660 (24.8)	985 (27.1)	0.0446
Other	24 (0.4)	19 (0.7)	5 (0.1)	0.0002

[1] Represents total number of Cohort participants enrolled through March 31, 2015. Totals may not sum to *N* due to missing data

[2] Chi-square or Wilcoxon test

[3] Data at the time of first sequence

[4] Perinatal, blood transfusion, hemophilia/coagulation disorder

https://doi.org/10.1371/journal.pone.0185644.t001

using Sequencher DNA Sequence Analysis Software (Gene Codes Corp.). This process resulted in the generation of two independent contiguous sequences of 1497bp of *PR/RT* (corresponding to nucleotide positions 2253–3749 of HBX2CG [GenBank accession K03455]) and 864bp *int* (corresponding to nucleotide positions 4230–5093 of HBX2CG).

Analyses

Sequence data were collected from HIV positive plasma samples from DC Cohort individuals targeting either a portion of the *PR/RT* and/or *int*. We also included 170 subtype reference sequences from the Los Alamos HIV database (http://www.hiv.lanl.gov/) to assign sequences to particular subtype clades. Sequence data were aligned using MAFFT [9]. The best-fit model of molecular evolution [10] was estimated from the data using jModelTest [11]. A maximum likelihood phylogenetic estimate [12] was made using RAxML and 3 codon-position partitions [13] with the best-fit model for each partition. Nodal support was estimated using the bootstrap approach with 1,000 replicates [14]. Bayesian trees were also inferred using MrBayes [15] and 3 codon-position partitions. We ran four chains (one cold and three heated) for $4\text{x}10^7$ generations sampling every 2,000 steps for the *int* region and for 10^8 generations sampling every 4,000 steps for the *PR/RT* region. Each run was repeated twice. Convergence and mixing of the Markov chains were assessed in Tracer [16]. Phylogenetic transmission (infection) clusters [17] were defined as those clades with bootstrap proportions $>70\%$ and posterior probabilities >0.95. Transmission chains were also assessed using a recently described network approach [18, 19] implemented in HIV-Trace (http://test.datamonkey.org/hivtrace). We used genetic distance thresholds of 0.01 substitutions/site for identifying potential transmission partners (see [18]).

HIV subtype B relative genetic diversity (i.e., population size over time) was inferred in BEAST [20] using the *PR/RT* sequence data for all infected individuals born in the US. We assume that these patients were also infected in the US. No sequence data were available for *int* from US born patients. We used the GMRF Bayesian Skyride model [21], the HKY substitution model with gamma-distributed among-site rate heterogeneity, and a relaxed clock (log-normal) model of rate of substitution [22]. We used the date of HIV-1 diagnosis to calibrate the analysis and a normal prior with mean = 0.001 and SD = 0.0005 for ucld.mean. We performed two runs $2\text{x}10^7$ generations long sampling every 1,000 generations. Parameter uncertainty was summarized in the 95% highest posterior density (HPD) intervals. All output generated by BEAST was analyzed in Tracer [23].

HIV-1 subtype identification was done using the REGA subtyping tool [24, 25] and validated via phylogenetic analyses above. Haplotype diversity (h), the number of segregating sites (S), nucleotide diversity (π), Watterson genetic diversity (θ) and recombination rate (r) were estimated using DnaSP v. 5.10.1 [26]. Nucleotide ambiguities in the alignment were arbitrarily resolved for these analyses. We identified drug resistant mutations by BLASTing [27] nucleotide sequences against the Stanford HIV Drug Resistance Database (https://hivdb.stanford.edu) using the HIVdb Program. We then identified nucleotide positions under positive selection using Fast Unconstrained Bayesian AppRoximation (FUBAR) [28], while accounting for recombination GARD [29, 30]. These analyses were carried out in HyPhy [31].

Because the HIV sequences are related through an hierarchical evolutionary history, they are not independent and typical genotype to phenotype associations cannot be performed without taking the dependence structure into account [32]. We intended to use treescanning [33] for the analyses, but found that there was insufficient resolvable structure for *int* data set and although there were several large clades for *PR/RT*, there was no resolution between them at the base of the tree (i.e., low bootstrap values and posterior probabilities for clade structure among these large clades). Therefore, among the few clades large enough for statistical inference, we treated each clade as independent and conducted simple contingency table tests for association with sex, race and risk behavior where each clade was treated as a separate factor and the remaining sequences were grouped into one factor using a permutation chi-square test [34]. We used the chisq.test function in the core "stats" package of R [35] with the simulate.p.value option. We generated 100 million Monte Carlo permutations [36] for each test to obtain empirical p-values. To increase the number of cases per cell in our chi-square tests, we limited the risk category to the two main risk types (HRH, IDU and MSM). Because MSM and HRH are highly correlated with sex, we also analyzed gender excluding MSM.

Results

Samples

We paired sequence data with the most recently available post-DOH linked demographic and clinical data for participants consented by March 31, 2015 and sequenced by June 15, 2015. We collected 601 partial *int* sequences and 2,810 partial *PR/RT* sequences from 223 participants and 1,659 participants, respectively, with a few participants being sequenced for both genes. All sequences have been deposited in GenBank under accession numbers MF455515 – MF457397. Aggregated demographic and clinical information (race, sex, age, risk factor, viral load, CD4+ count, sequence date, HIV diagnosis date, and zip code) for participants are shown in Table 1 along with a comparison between DC Cohort participants who had sequences available to those without sequences. Our data came from participants who resided predominantly in Washington DC (73.4%) and had dominant HIV risk factors of MSM (48.1%) and heterosexual (28.8%). Participants with sequence availability were significantly more likely to be younger, black, female, non-US born, and DC residents ($p<0.0001$ for all) compared to those without sequence data. Slightly more participants with sequences were infected through heterosexual sexual contact, and a higher proportion had a history of hepatitis C, syphilis, gonorrhea, and chlamydia co-infections at the time of sequencing ($p<0.05$ for all). Participants with sequences had a shorter duration of infection, lower CD4 counts, higher viral loads, and were less likely to be treatment experienced ($p<0.0001$ for all). Among those on ARVs, a higher proportion of those sequenced were on dual-class regimens and had a lower prevalence of resistance to PIs, NRTIs, and NNRTIs ($p<0.001$ for all).

HIV-1 diversity

Our subtyping analysis of *int* found that 220 of the participants were infected with subtype B virus and identified three additional subtypes: A (1 individual), C (1), and G (1). The *PR/RT* sequences were also heavily subtype B (1,557 of 1,659 participants), but with a greater diversity of other subtypes including: AG (9 individuals), CD (1), BF (2), A (7), C (32), D (4), F (1), G (3), J (1), AK (1), BA (1), BD (29), BF (7), DB (1) and unknown (3) (see also Fig 1).

All the DNA sequences from the same individual comprised identical replicates, hence we only used one representative per individual for our analyses. The subtype B *PR/RT* gene showed higher diversity rates, substitution rates (estimated as in [37]) and minimum number of recombinant events than the subtype B *int* gene, but lower recombination/gen rates (Table 2). Watterson genetic diversity (θ) for subtype B *int* was lower in the IDU group compared to subtype B HRH and MSM likely due to the smaller sample size of the IDU (Table 2). For the subtype B *PR/RT* sequence data, we also included perinatally infected participants and that group had the highest nucleotide diversity, θ and non-synonymous substitution rates despite its smaller sample size.

No noticeable differences in diversity and substitution rates were observed among subtypes B, BD, and C for all the indices compared except for recombination per gene, which was about three times higher for subtype BD compared to the other two subtypes.

Drug resistance

A total of 38 subtype B *int* sequences (2 HRH, 3 IDU and 20 MSM, and 13 “other risk”) showed at least one DRM with an overall prevalence (HIV sequences including at least one DRM/total number of sequences) of 17.3%. Thirty-five IN Major mutations and 26 IN Accessory mutations were detected (Table 3). A total of 584 subtype B *PR/RT* sequences (215 HRH, 28 IDU, 232 MSM, 26 PER and 83 “other risk”) showed at least one DRM with a prevalence of 39.1%. NRTI, NNRTI and RT SDRMs included the highest proportions of resistant individuals (301 to 461) and total DRM (557 to 931) and unique DRM (59 to 80) (Table 3). A total of 10 subtype BD *PR/RT* sequences (1 HRH, 8 MSM and 1 “other risk”) showed at least one DRM with a prevalence of 34.5%. NRTI, NNRTI and RT SDRMs also included the highest proportions of resistant individuals (6 to 10) and total DRM (10 to 22) and unique DRM (7 to 14) (Table 3). Finally, a total of 8 subtype C *PR/RT* sequences (3 HRH and 5 “other risk”) showed at least one DRM with a prevalence of 25.0%. NNRTI and RT SDRMs included the highest proportions of resistant individuals (7) and total DRM (12 to 13) and unique DRM (9 to 12) (Table 3). It is important to highlight that since only 6.7% of the patients in our cohort are treatment naïve, we cannot confirm that these DRM are actually being transmitted. Amino acid mutations counted as a DRM for each subtype and gene region are also presented in S1 Table.

Subtype B *int* DRM caused amino acid changes in 11 different codons while Subtype B *PR/RT* DRM caused amino acid changes in 72 different codons (Table 3). Similarly, subtype BD and C *PR/RT* DRM caused amino acid changes in 13 and 19 different codons, respectively. These codons did not correspond to those inferred by FUBAR as being positively selected (8 and 12 amino acids in the *int* and *PR/RT* genes, respectively), except codon 140 in *PR/RT*, which matched in both analyses. Positively selected sites included 8 amino acids for Subtype B *int* (72, 201, 206, 218, 227, 230, 265 and 283), 14 for Subtype B *PR/RT* (12, 13, 19, 35, 37, 59, 79, 95, 136, 140, 203, 236, 301 and 312), 2 for Subtype BD *PR/RT* (136 and 312) and 5 for Subtype C *PR/RT* (19, 65, 236, 274 and 275).

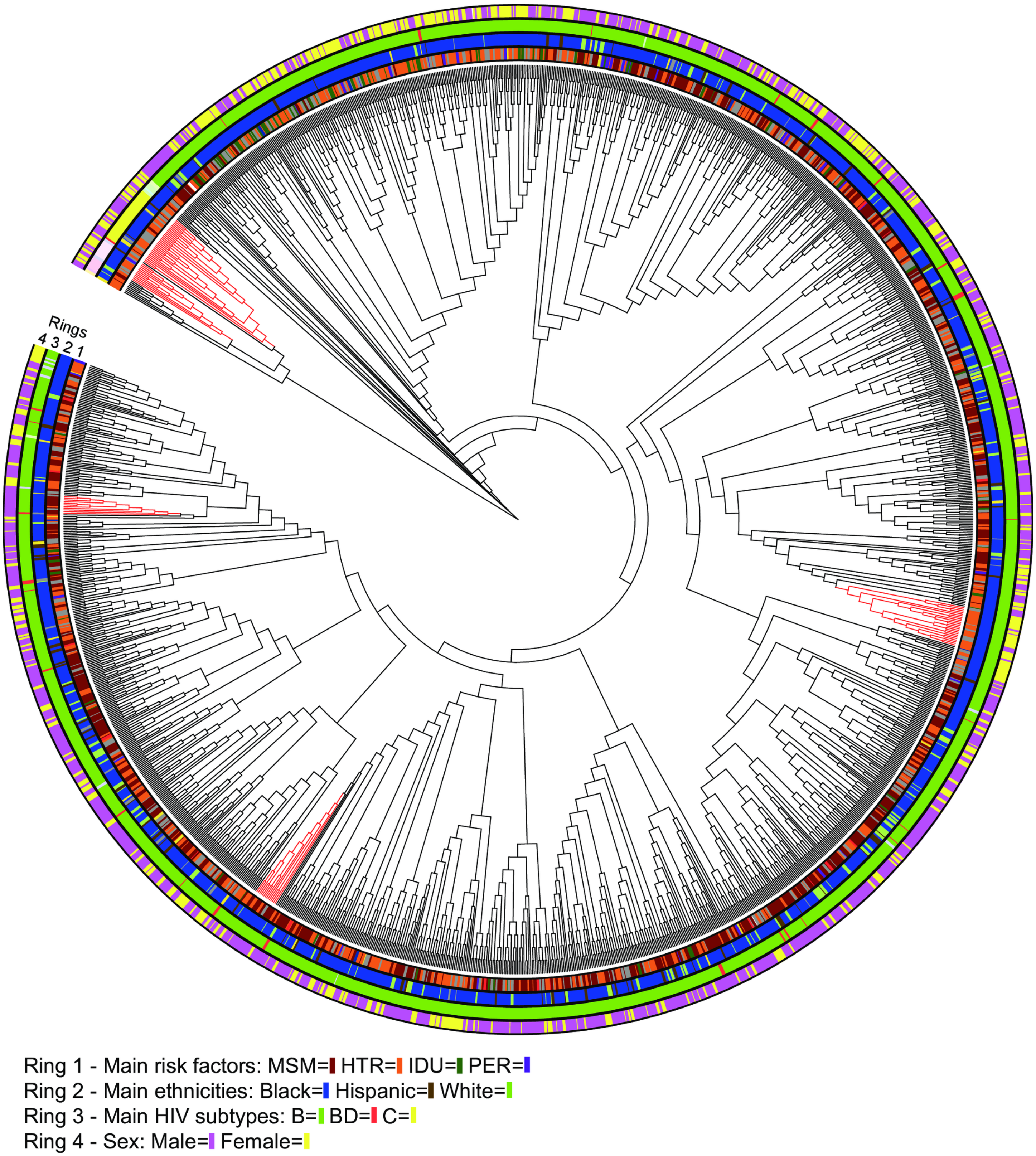
Rings
4 3 2 1
Ring 1 - Main risk factors: MSM=▮ HTR=▮ IDU=▮ PER=▮
Ring 2 - Main ethnicities: Black=▮ Hispanic=▮ White=▮
Ring 3 - Main HIV subtypes: B=▮ BD=▮ C=▮
Ring 4 - Sex: Male=▮ Female=▮

Fig 1. Cladogram of the *PR/RT* gene showing risk factors, ethnicities, subtypes and sex in four concentric rings. Main phenotypes within each ring are represented with different colors. Well-supported clades comprised of >10 HIV sequences are also indicated in red.

https://doi.org/10.1371/journal.pone.0185644.g001

Table 2. HIV DNA polymorphism and drug resistant mutations (DRM).

	Diversity					Substitutions		Recombination		DRM
	N	S	h	Pi	θ (W)	Pi(s)	Pi(ns)	R/gene	Rm	Total (relative %)
SubB										
Int										
ALL	220	434	220	0.054	0.084	0.188	0.024	1133	100	38 (17.3)
HRH	78	345	78	0.053	0.081	0.185	0.023	831	83	2 (2.6)
IDU	13	180	13	0.051	0.067	0.174	0.024	3134	37	3 (23.1)
MSM	71	350	71	0.054	0.084	0.190	0.024	1178	85	20 (28.2)
PR/RT										
ALL	1557	720	1556	0.064	0.090	0.222	0.028	975	149	584 (39.1)
HRH	512	614	512	0.063	0.088	0.220	0.028	1210	141	215 (38.5)
IDU	85	456	85	0.055	0.089	0.187	0.024	348	111	28 (32.6)
MSM	605	636	604	0.065	0.090	0.228	0.028	1304	143	232 (36.4)
PER	37	424	37	0.070	0.099	0.215	0.039	385	101	26 (68.4)
SubBD										
PR/RT										
ALL	29	353	29	0.067	0.088	0.239	0.029	2872	89	10 (34.5)
HRH	8	199	8	0.066	0.075	0.231	0.030	4363	37	1 (3.4)
MSM	18	294	18	0.067	0.083	0.243	0.028	1293	61	8 (27.6)
SubC										
PR/RT										
ALL	32	354	32	0.066	0.086	0.243	0.027	957	91	8(25.0)
HRH	16	273	16	0.066	0.080	0.236	0.029	1895	69	3 (9.4)

Diversity (N = sequences, S = segregating sites, h = haplotypes, Pi = nucleotide diversity, θ = Watterson genetic diversity), substitutions (Pi(s) = synonymous, non-synonymous Pi(ns)), recombination (R) (R/gen = R per gen, and Rm = minimum number of recombinant events) rates. Total and relative (total/N) proportion (%) of HIV strains including DRM. MSM = Men who have sex with men, HRH = heterosexuals, IDU = intravenous drug users, PER = perinatal

https://doi.org/10.1371/journal.pone.0185644.t002

Phylodynamics and transmission networks

The past demographic analysis of DC subtype B *PR/RT* sequences from US born patients in BEAST (Fig 2) indicates that HIV relative genetic diversity has not decreased significantly over the last 25 years, despite an increasing prevalence of HIV infection earlier on in the epidemic followed by more recent decreases in HIV incidence rates [1, 38]. Additionally BEAST analyses aggregating HIV sequences by the two main sexual orientations of their hosts (MSM and HTR) generated similar Skyride plots to that presented in Fig 2. Our ML phylogenies and bootstrap analysis (70% support) grouped 13.5% of the *int* sequences into 14 phylogenetic clusters comprised of 2 to 3 sequences, while 39.0% of the *PR/RT* sequences grouped into 203 phylogenetic clusters comprised of 2 to 32 sequences (S1 Fig, S2 Fig and S3 Fig). These clusters were subsequently confirmed by our Bayesian analyses at a P≥0.95. Higher proportions of clustered sequences were found in zip codes where HIV prevalence is highest (r = 0.607; P<0.00001) with southeastern Washington, DC zip codes, being the highest (Fig 3). A network approach (HIV-Trace) grouped 3.6% of the *int* sequences into 4 clusters of 2 sequences and 10.6% of the *PR/RT* sequences into 76 clusters of 2–7 sequences (Fig 4). Hence, this approach identified fewer and smaller clusters compared to the ML-bootstrap approach.

Table 3. Drug resistant mutations.

	Sub B – *int*			Sub B – *PR/RT*			Sub BD – *PR/RT*			Sub C – *PR/RT*		
	S	**TM**	**UM**	**S**	**TM**	**UM**	**S**	**TM**	**UM**	**S**	**TM**	**UM**
IN Major	24	35	17	-	-	-	-	-	-	-	-	-
IN Accessory	25	26	6	-	-	-	-	-	-	-	-	-
PR Major	-	-	-	87	154	31	0	0	0	1	1	1
PR Accessory	-	-	-	72	107	22	0	0	0	2	2	2
NRTI	-	-	-	301	557	80	6	13	9	3	4	4
NNRTI	-	-	-	410	663	69	6	10	7	7	13	12
PR SDRMs	-	-	-	90	186	26	0	0	0	1	1	1
RT SDRMs	-	-	-	461	931	59	10	22	14	7	12	9
PI TSMs	-	-	-	33	34	11	0	0	0	1	1	1
NRTI TSMs	-	-	-	52	59	13	1	1	1	1	1	1
NNRTI TSMs				14	15	12	1	1	1	0	0	0
DRM Codons		11			72			21			22	
FUBAR Codons		8			14			2			5	

Sequences (S), Total Mutations (TM) and Unique Mutations (UM) conferring resistance to antiretroviral drugs (IN Major to NNRTI TSMs) for genes int and PR/RT in HIV subtypes (Sub) B, BD and C. DRM amino acid codons and codons under adaptive selection (FUBAR) are also listed

https://doi.org/10.1371/journal.pone.0185644.t003

Phenotypic associations

We found five well-supported (≥70% bootstrap proportion and ≥0.95 posterior probability) *PR/RT* clades of 11 to 32 HIV sequences. None of the clades showed exclusive mapping of a particular risk group (Fig 1). However, our chi-square analyses detected a strong association ($P<0.00001$) between those five clades and risk (MSM, IDU and HRH) and sex (Table 4), still highly significant after a Bonferroni threshold for multiple tests of $P<0.0167$. Because MSM and HRH are so highly correlated with sex, the sex significant results may be due to correlation with risk factor. Therefore, we re-ran the test for sex excluding MSM (Table 4) and the evidence for a sex effect dropped dramatically ($P = 0.095$), suggesting that the sex association was likely due to its correlation with the most common risk categories. The other clinical variables were almost constant across clades (e.g., race/ethnicity) or showed low numbers of cases per cell in our chi-square test (e.g., IDU).

Previous HIV studies in British Columbia (Canada) [39] and Switzerland [40] using intra-patient or inter-patient sampling detected higher numbers of viral sequences falling within clusters (3120–4431, (55–57% of all sequences)) than our study (647 (39.0%) of the *PR/RT* sequences). Similarly, HIV epidemics in those two studies were dominated by 3–5 clusters of 1051–107 sequences and 14–17 clusters of 29–88 sequences. The HIV DC population was grouped into 203 phylogenetic clusters of 2 to 32 *PR/RT* sequences, of which only 5 clusters comprised 11–32 individuals. Ongoing work by our group including intra-patient sequences from multiple patients generated via high-throughput (MiSeq) technology and covering a larger geographic area (DC and neighboring states) will allow us to apply some on the methodologies (e.g., patristic distances) in [39, 40] to identify transmission clusters and consider multiple clinical co-variables simultaneously.

Discussion

Molecular surveillance and subtype diversity

The predominant HIV-1 subtype circulating among a subgroup of DC Cohort participants comprised of mainly Non-Hispanic Blacks from 2011–2015 was subtype B, which accounted

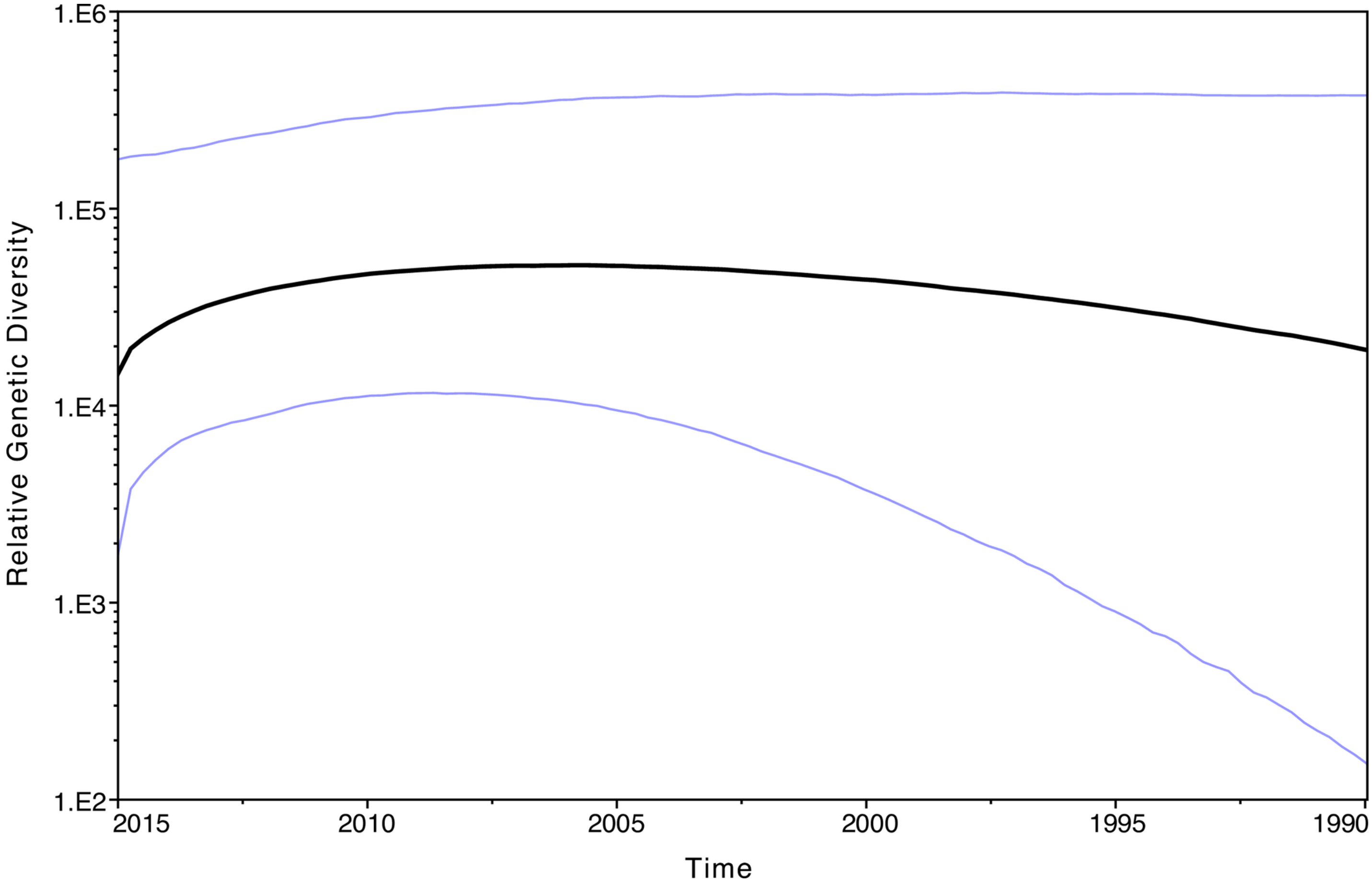

Fig 2. Bayesian skyride plot of HIV-1 subtype B *PR/RT* past population dynamics in US born patients. Black lines show the median estimate and blue lines the 95% high posterior density limits of the relative genetic diversity over time.

https://doi.org/10.1371/journal.pone.0185644.g002

for 98.7% of the infections for *int* and 93.9% for *PR/RT*. Subtype B is also the predominant subtype in the US and Western Europe. We also detected another 15 subtypes that accounted for the remaining sequences. A study by Kassaye et al. [41] also focused on HIV diversity in DC used *pol* sequences collected from 641 individuals enrolled between 1994–2013 and showed a higher frequency of non-subtype B sequences of HIV (88.5% subtype B and 11.4% non-B HIV-1 subtypes). These differences could be due to demographic and clinical characteristics of the participants in each cohort, changes in the dynamics of HIV in DC over the different time frames of the two studies and/or the fact that the Kassaye et al. study surveyed over a period of time that was double the timeframe of this study. HIV prevalence among ethnic minorities is very high in DC, much less among the heterosexual white population, indicating that the HIV epidemic is concentrated in key populations defined by sexual orientation and ethnicity. Yet, nearly all patients harbored subtype B viruses, which suggests that the local epidemic amongst ethnic minorities was seeded in the US and is at present not primarily driven by importation from sub-Saharan Africa or other parts of the world where non-B subtypes circulate. The finding of a small proportion of patients with non-subtype B virus, nonetheless, highlights the cultural diversity of DC and the immigration of persons from other areas of the world. Higher diversity of non-B HIV-1 subtypes has been reported in other immigrant-rich North American cohorts on the East Coast such as Rhode Island (8.3%) [42], Maryland (12.9%) [6], and

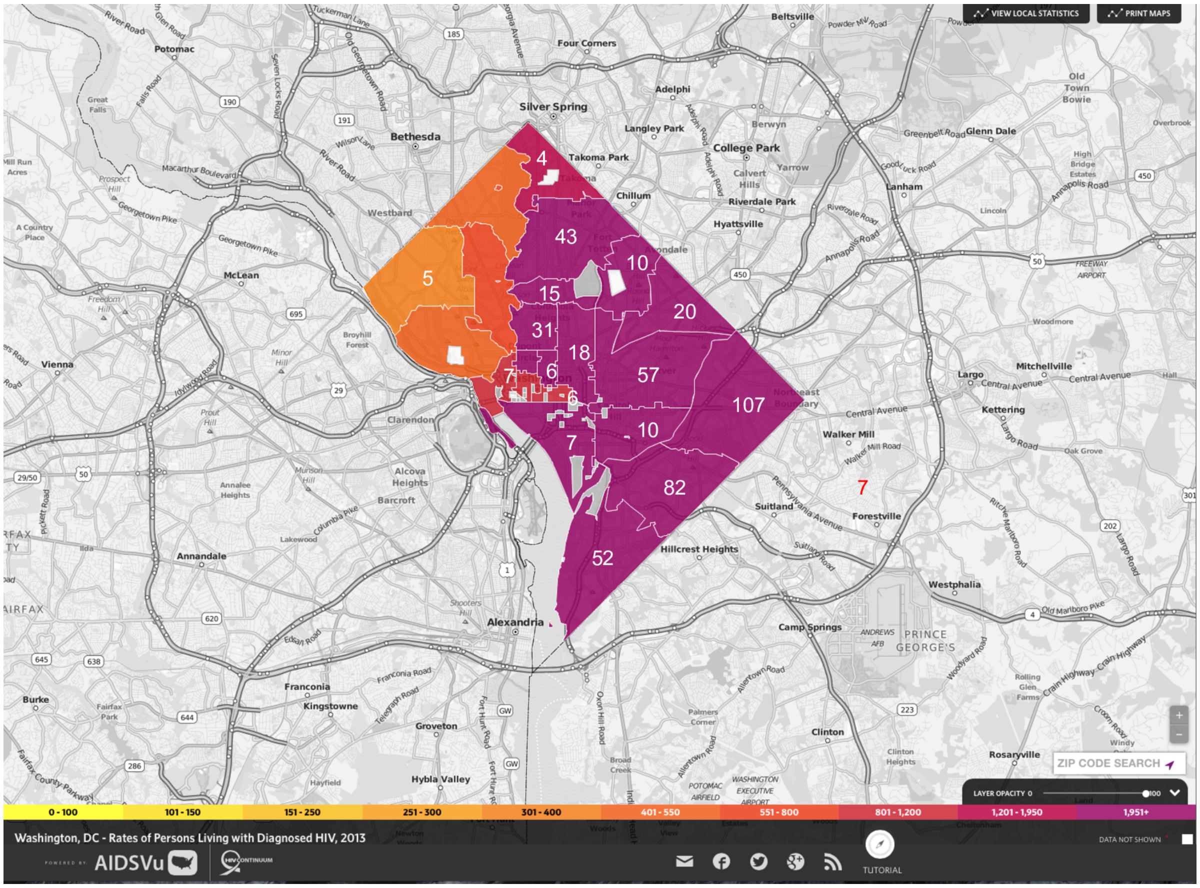

Fig 3. Geographic distribution of transmission networks by zip code overlaid on Washington DC map with rates of persons living with diagnosed HIV in DC. Only zip codes containing ≥4 HIV sequences that fall into a ML cluster (70% bootstrap support) are shown.

https://doi.org/10.1371/journal.pone.0185644.g003

New York City (43.4%) [43]. Similarly, the prevalence of non-B infections has increased from 5.9% in 2006 to 8.5% in 2013 in 7 US states (Colorado, Connecticut, Michigan, New York,

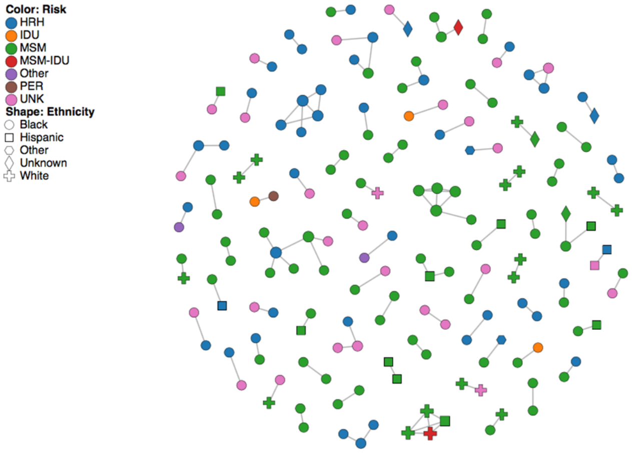

Fig 4. *PR/RT* network (HIV-Trace) of Washington DC HIV-1 isolates showing risk factors and ethnicities.

https://doi.org/10.1371/journal.pone.0185644.g004

Table 4. Chi-square tests of association between phylogenetic clades and demographic variables.

	Risk	Sex	Sex w/o MSM
Clade	**Total (% HRH)**	**Total (% Female)**	**Total (% Female)**
1	11 (18.2%)	11 (0%)	2 (0%)
2	20 (70%)	25 (48%)	16 (68.8%)
3	20 (100%)	22 (81.8%)	21 (85.7%)
4	10 (100%)	11 (63.6%)	10 (70%)
5	12 (0%)	13 (7.7%)	1 (100%)
P-value	<1.00E-08	9.50E-07	0.095

https://doi.org/10.1371/journal.pone.0185644.t004

South Carolina, Texas, and Washington) [44]. Lower rates, however, of non-B subtypes have been reported for other Southeastern US states like North Carolina [45], which is consistent with the national average prevalence of non-B subtypes (3.27%) estimated in 2011 [46].

Our estimates of *int* and *PR/RT* genetic diversity in subtype B viruses from DC were relatively high (e.g., θ = 0.08–0.09, for all main subtypes) compared to other US subtype B strains available in Los Alamos HIV database for *int* (θ = 0.075) and *PR/RT* (θ = 0.067). This suggests that the HIV epidemic in DC is likely to be mature and that extensive exchange between risk groups has been ongoing for years, which is also supported by our phylodynamic results as discussed below. High HIV genetic and subtype diversity is of concern as it complicates vaccine development by increasing the chances for the evolution of vaccine resistance or the failure for specific epitopes to work against broad diversity [47].

Our population estimators of genetic diversity, recombination, and selection showed differences between patients grouped by risk factor for some estimators. This may suggest differences in transmission and HIV-1 dynamics among risk groups. HIV Subtype B in perinatally infected participants showed higher levels of diversity compared to other risk groups likely due to the uniqueness of mother-infant transmission in perinatal infections (i.e., potentially high HIV effective population size) and exposure to many different ARV regimens.

Our phylodynamic analyses of the US born subtype B sequences showed that the relative genetic diversity of the DC viral population has not significantly decreased over the last 25 years. This is consistent with previous phylodynamic analyses of US subtype B sequences collected between 1981 and 2006 [47]. More importantly, the epidemic persists among certain groups like MSM, where infection rates are still very high and continue to be a source of continued transmission [1, 3, 47]. Therefore, department of health prevention efforts should continue to focus on high-risk groups as well as the general population [1, 48].

Phylogenetic structure of HIV-1 in Washington DC

Our phylogenetic analyses of subtype B sequences did not reveal clear evidence that HIV-1 populations in DC are structured by any of the epidemiological and clinical factors studied. These results agree with previous subtype B star-like phylogenies reported for DC [41]. Geographically broader phylogenetic studies across the US also showed lack of phylogenetic structuring based on transmission type, sociodemographic factors, and geographic location [47]. Keele et al. [49], for example, showed that viral *env* genes evolving from individual transmitted or founder HIV-1 subtype B viruses generally exhibited a star-like phylogeny, such as the one observed in North American viruses [47]. Given the maturity of the HIV-1 epidemic in DC and the fact that the virus is thought to mutate at a rate of 1% per year [50, 51], the possibility exists that different clades could have emerged in different wards or high-risk groups in the city. Indeed, phylogenetic structuring based on these factors has been observed before between subtypes in, for example, Africa [52] and Asia [53], and within subtypes in, for example, Vietnam [54] and China [55]. But contrary to what happened in those HIV/AIDS epidemics, our star-like gene genealogies of HIV-1 in Washington DC suggest the DC epidemic expanded uniformly (i.e., no phylogenetic structure) across the metropolitan area and across the epidemiological risk types [56, 57].

Transmission networks

The extent to which transmission of HIV-1 is clustered is not clear. Some studies [58–69] report high clustering (24 to 65%) levels, while others [18, 47, 67] show much lower values (7 to 17%) for the same subtypes and transmission routes. Our comprehensive phylogenetic analyses of HIV-1 from DC show moderate proportions of subtype B infections (13.5% to 39.0% depending on the gene) falling into clusters, confirming that transmission chains play a role in HIV-1 transmission and spread of HIV in DC [41]. Previous studies of HIV-1 clustering in DC and North America reported values of ~17% [41, 47], but a recent study of 86 patients in Chicago showed levels of 36% [69]. Moreover, differences in clustering have also been observed between subtypes, transmission routes, mental health and geographic regions [47, 62, 63, 69]. Our phenotypic association testing (chi-square) showed significant ($P<0.00001$) association between *PR/RT* clusters and risk factor (transmission group), suggesting that MSM and HRH may comprise largely independent transmission networks in the DC area, as seen in other US states [44, 70].

Evolution of drug resistance

We detected a high prevalence of Drug Resistance Mutations (DRM) for both subtype B *int* (17.1%) and *PR/RT* (39.1%) with similar prevalence across risk groups except for among heterosexuals when looking at *int*, (2.6%) and perinatally infected participants when looking at *PR/RT* (68.4%). High rates of DRM were also detected in subtype BD (34.5%) and C (25.0%) for *PR/RT*. *PR/RT* prevalence rates reported in this study are higher than those previously reported for the DC area in 1994–2013 using *pol* sequences [41]. DRM rates higher than 50% are frequently observed in large sequence databases in Europe (e.g., UK, http://www.hivrdb.org.uk and Switzerland, www.shcs.ch) when similar patient groups are considered (i.e., ART experienced). This finding may also reflect patient selection. Future analyses should determine if DRMs of current treatment regimens cluster, and if there is evidence for increasing prevalence of DRMs based on current treatment regimens in the DC area. Similarly, studies including only naïve patients are needed to determine to what extent these DRM are actually being transmitted in the DC area.

Finally, we also detected 8 *int* and 12 *PR/RT* amino acids under adaptive selection. These correspond to sites evolving in the DC HIV-1 subtype B population that have not been fixed in HIV-1 as those conferring drug resistance. These amino acid sites involved positions that, in the future, may confer resistance to antiretrovirals if they become fixed in the population and could be helpful in informing ARV regimen choices.

Conclusions

Routinely collected commercial sequences are useful for examining transmission dynamics and can provide an historical context for the HIV epidemic. Our ability to combine them with surveillance, clinical, and demographic indicators enhances their utility in understanding transmission patterns and geospatial distribution. Given the large number of sequences analyzed, the data presented in this study inform our understanding of the molecular epidemiology of HIV infection in Washington DC. Moreover, they help lay the foundation for future work in which molecular and epidemiologic data could be used synergistically to target public health interventions to interrupt HIV transmission networks.

Supporting information

S1 Table. Amino acid mutations counted as drug resistant mutations for each main subtype and gene region analyzed in this study.
(XLSX)

S1 Fig. Maximum likelihood phylogenetic tree of Washington DC HIV-1 *int* isolates. Clades supported by bootstrap proportions ≥70% are indicated with an asterisk. These clades were also supported by Bayesian posterior probabilities ≥0.95.
(PDF)

S2 Fig. Maximum likelihood phylogenetic tree of Washington DC HIV-1 *PR/RT* isolates.
(PDF)

S3 Fig. Maximum likelihood consensus phylogenetic tree of Washington DC HIV-1 *PR/RT* isolates showing clades supported by bootstrap proportions ≥70%. These clades were also supported by Bayesian posterior probabilities ≥0.95.
(PDF)

Acknowledgments

The authors would also like to acknowledge the DC Cohort Executive Committee Members. DC Cohort data in this analysis were collected by the DC Cohort investigators and research staff located at: Cerner Corporation (Jeffrey Binkley, Harlen Hays, Thilakavathy Subramanian, Robert Taylor, Dana Franklin, Rachel Hart; Nabil Rayeed); Children's National Medical Center Adolescent (Lawrence D'Angelo) and Pediatric (Natella Rakhmanina) clinics; Family and Medical Counseling Service (Angela Wood); Georgetown University (Princy Kumar); George Washington Medical Faculty Associates (David Parenti); George Washington University Department of Epidemiology and Biostatistics (James Peterson, Naji Younes, Lindsey Powers Happ, Maria Jaurretche, Arpi Terzian); Howard University (Saumil Doshi); La Clinica Del Pueblo (Ricardo Fernandez); MetroHealth (Annick Hebou); National Institutes of Health (Carl Dieffenbach, Henry Masur); Unity Health Care (Gebeyehu Teferi); Veterans Affairs Medical Center (Debra Benator); Washington Hospital Center (Maria Elena Ruiz); Whitman-Walker Health (David Hardy, Deborah Goldstein). We would also like to acknowledge the Research Assistants at all of the participating sites, the DC Cohort Community Advisory Board, and staff of the DC Department of Health HIV/AIDS, Hepatitis, STD, TB Administration Strategic Information Division. We would also like to thank the site PIs, RAs, the participants themselves, the DC Department of Health, the DC CFAR, and the National Institutes of Health for their numerous contributions to the DC Cohort. Finally, we thank Gabriel Leventhal and two anonymous reviewers for helpful comments/suggestions on previous versions of this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Contributions

Conceptualization: Marcos Pérez-Losada, Amanda D. Castel, Alan E. Greenberg, Keith A. Crandall.

Data curation: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Michael Kharfen, Charles P. Cartwright, Keith A. Crandall.

Formal analysis: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Bruce Huang, Taylor Maxwell, Keith A. Crandall.

Funding acquisition: Amanda D. Castel, Alan E. Greenberg.

Investigation: Marcos Pérez-Losada, Amanda D. Castel, Taylor Maxwell, Alan E. Greenberg, Keith A. Crandall.

Methodology: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Michael Kharfen, Charles P. Cartwright, Bruce Huang, Taylor Maxwell, Keith A. Crandall.

Project administration: Amanda D. Castel, Brittany Lewis, Alan E. Greenberg, Keith A. Crandall.

Resources: Amanda D. Castel, Michael Kharfen, Charles P. Cartwright, Alan E. Greenberg, Keith A. Crandall.

Software: Marcos Pérez-Losada, Taylor Maxwell, Keith A. Crandall.

Supervision: Marcos Pérez-Losada, Amanda D. Castel, Alan E. Greenberg, Keith A. Crandall.

Validation: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Michael Kharfen, Charles P. Cartwright, Keith A. Crandall.

Visualization: Marcos Pérez-Losada, Amanda D. Castel.

Writing – original draft: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Michael Kharfen, Charles P. Cartwright, Bruce Huang, Taylor Maxwell, Alan E. Greenberg, Keith A. Crandall.

Writing – review & editing: Marcos Pérez-Losada, Amanda D. Castel, Brittany Lewis, Taylor Maxwell, Alan E. Greenberg, Keith A. Crandall.

References

1. District of Columbia Department of Health HIV/AIDS H, STD, and TB Administration (HAHSTA). Annual Epidemiology & Surveillance Report: Surveillance Data Through December 2016. Washington, DC: DC Department of Health, 2016.
2. Kang HIHQATTRSMCTGJ. Prevalence of Diagnosed and Undiagnosed HIV Infection—United States, 2008–2012. Morbidity Mortality Weekly Report. 2015; 64(24):657–62. PMID: 26110835
3. Castel AD, Kalmin MM, Hart RL, Young HA, Hays H, Benator D, et al. Disparities in achieving and sustaining viral suppression among a large cohort of HIV-infected persons in care—Washington, DC. AIDS Care. 2016; 28(11):1355–64. https://doi.org/10.1080/09540121.2016.1189496 PMID: 27297952; PubMed Central PMCID: PMCPMC5084086.
4. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. The New England journal of medicine. 2000; 342(13):921–9. https://doi.org/10.1056/NEJM200003303421303 PMID: 10738050.
5. Counsil AI. New Americans in Washington, D.C.: The Political and Economic Power of Immigrants, Latinos, and Asians in Our Nation's Capital 2015 [cited 2016 May 6]. Available from: http://www.immigrationpolicy.org/sites/default/files/docs/new_americans_washington_dc_2015.pdf).
6. Carr JK, Osinusi A, Flynn CP, Gilliam BL, Maheshwari V, Zhao RY. Two Independent Epidemics of HIV in Maryland. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2010; 54(3):297–303. https://doi.org/10.1097/QAI.0b013e3181e0c3b3 PMID: 20505517
7. Greenberg AE, Hays H, Castel AD, Subramanian T, Happ LP, Jaurretche M, et al. Development of a large urban longitudinal HIV clinical cohort using a web-based platform to merge electronically and manually abstracted data from disparate medical record systems: technical challenges and innovative solutions. J Am Med Inform Assoc. 2015. https://doi.org/10.1093/jamia/ocv176 PMID: 26721732.

8. Oster AM, Wertheim JO, Hernandez AL, Ocfemia MCB, Saduvala N, Hall HI. Using Molecular HIV Surveillance Data to Understand Transmission Between Subpopulations in the United States. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2015; 70(4):444–51. https://doi.org/10.1097/QAI.0000000000000809 PMID: 26302431

9. Kotah S. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution. 2013; 30:772–80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

10. Posada D, Crandall KA. Selecting models of nucleotide substitution: An application to Human Immunodeficiency Virus 1 (HIV-1). Molecular biology and evolution. 2001; 18(6):897–906. PMID: 11371577

11. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 2012; 9(8):772. https://doi.org/10.1038/nmeth.2109 PMID: 22847109

12. Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol. 1981; 17:368–76. PMID: 7288891

13. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90. https://doi.org/10.1093/bioinformatics/btl446 PMID: 16928733

14. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985; 39:783–91. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x PMID: 28561359

15. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology. 2012; 61(3):539–42. https://doi.org/10.1093/sysbio/sys029 PMID: 22357727; PubMed Central PMCID: PMC3329765.

16. Rambaut A, Drummond AJ. Tracer: MCMC trace analysis tool. 1.5 ed. Edinburgh: Institute of Evolutionary Biology; 2009. p. http://tree.bio.ed.ac.uk/software/tracer/.

17. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nature reviews Genetics. 2009; 10(8):540–50. Epub 2009/07/01. nrg2583 [pii] https://doi.org/10.1038/nrg2583 PMID: 19564871.

18. Wertheim JO, Kosakovsky Pond SL, Forgione LA, Mehta SR, Murrell B, Shah S, et al. Social and Genetic Networks of HIV-1 Transmission in New York City. PLoS pathogens. 2017; 13(1):e1006000. https://doi.org/10.1371/journal.ppat.1006000 PMID: 28068413; PubMed Central PMCID: PMCPMC5221827 following competing interests: JOW is a paid consultant for the Centers for Disease Control and Prevention.

19. Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, et al. The global transmission network of HIV-1. The Journal of infectious diseases. 2014; 209(2):304–13. https://doi.org/10.1093/infdis/jit524 PMID: 24151309; PubMed Central PMCID: PMCPMC3873788.

20. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012; 29(8):1969–73. https://doi.org/10.1093/molbev/mss075 PMID: 22367748; PubMed Central PMCID: PMC3408070.

21. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Molecular biology and evolution. 2008; 25(7):1459–71. Epub 2008/04/15. msn090 [pii] https://doi.org/10.1093/molbev/msn090 PMID: 18408232; PubMed Central PMCID: PMC3302198.

22. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS biology. 2006; 4(5):e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

23. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer2014.

24. Alcantara LCJ, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic acids research. 2009; 37:W634–W42. https://doi.org/10.1093/nar/gkp455 PMID: 19483099

25. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics. 2005; 21 (19):3797–800. https://doi.org/10.1093/bioinformatics/bti607 PMID: 16076886

26. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–2. https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325

27. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997; 25:3389–402. PMID: 9254694

28. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, et al. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Molecular biology and evolution. 2013; 30(5):1196–205. https://doi.org/10.1093/molbev/mst030 PMID: 23420840

29. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. Bioinformatics. 2006; 22(24):3096–8. https://doi.org/10.1093/bioinformatics/btl474 PMID: 17110367.

30. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Molecular biology and evolution. 2006; 23(10):1891–901. https://doi.org/10.1093/molbev/msl051 PMID: 16818476.

31. Pond SLK, Muse SV. HyPhy: Hypothesis Testing Using Phylogeneties. Statistical Methods in Molecular Evolution. Statistics for Biology and Health. New York: Springer; 2005. p. 125–81.

32. Felsenstein J. Phylogenies and the comparative method. American Naturalist. 1985; 125:1–15.

33. Templeton AR, Maxwell T, Posada D, Stengard JH, Boerwinkle E, Sing CF. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics. 2005; 169(1):441–53. https://doi.org/10.1534/genetics.104.030080 PMID: 15371364

34. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: Chi-square and the problem of small samples. Molecular biology and evolution. 1989; 6:539–45. PMID: 2677600

35. RDevelopmentCoreTeam. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2008.

36. Hope ACA. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society Series B (Methodological). 1968; 30(3):582–98.

37. Hudson RR. Estimating the recombination parameter of a finite population model without selection. Genetical Research, Cambridge. 1987; 50:245–50.

38. Murray CJL, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2014; 384(9947):1005–70. https://doi.org/https://doi.org/10.1016/S0140-6736(14)60844-8

39. Poon AF, Joy JB, Woods CK, Shurgold S, Colley G, Brumme CJ, et al. The impact of clinical, demographic and risk factors on rates of HIV transmission: a population-based phylogenetic analysis in British Columbia, Canada. The Journal of infectious diseases. 2015; 211(6):926–35. https://doi.org/10.1093/infdis/jiu560 PMID: 25312037; PubMed Central PMCID: PMC4351365.

40. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. The Journal of infectious diseases. 2010; 201(10):1488–97. https://doi.org/10.1086/651951 PMID: 20384495.

41. Kassaye SG, Grossman Z, Balamane M, Johnston-White B, Liu C, Kumar P, et al. Transmitted HIV Drug Resistance Is High and Longstanding in Metropolitan Washington, DC. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2016; 63(6):836–43. https://doi.org/10.1093/cid/ciw382 PMID: 27307507; PubMed Central PMCID: PMC4996138.

42. Chan PA, Reitsma MB, DeLong A, Boucek B, Nunn A, Salemi M, et al. Phylogenetic and geospatial evaluation of HIV-1 subtype diversity at the largest HIV center in Rhode Island. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2014; 28:358–66. https://doi.org/10.1016/j.meegid.2014.03.027 PMID: 24721515; PubMed Central PMCID: PMCPMC4190103.

43. Lin HH, Gaschen BK, Collie M, El-Fishaway M, Chen Z, Korber BT, et al. Genetic characterization of diverse HIV-1 strains in an immigrant population living in New York City. Journal of acquired immune deficiency syndromes. 2006; 41(4):399–404. https://doi.org/10.1097/01.qai.0000200663.47838.f1 PMID: 16652046.

44. Oster AM, Switzer WM, Hernandez AL, Saduvala N, Wertheim JO, Nwangwu-Ike N, et al. Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013. Ann Epidemiol. 2017; 27(4):244–51 e1. Epub 2017/03/21. https://doi.org/10.1016/j.annepidem.2017.02.002 PMID: 28318764.

45. Dennis AM, Hue S, Learner E, Sebastian J, Miller WC, Eron JJ. Rising prevalence of non-B HIV-1 subtypes in North Carolina and evidence for local onward transmission. Virus Evol. 2017; 3(1):vex013. https://doi.org/10.1093/ve/vex013 PMID: 28567304; PubMed Central PMCID: PMCPMC5442504.

46. Pyne MT, Hackett J Jr., Holzmayer V, Hillyard DR. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. Journal of clinical microbiology. 2013; 51(8):2662–9. https://doi.org/10.1128/JCM.00880-13 PMID: 23761148; PubMed Central PMCID: PMCPMC3719628.

47. Pérez-Losada M, Jobes DV, Sinangil F, Crandall KA, Posada D, Berman PW. Phylodynamics of gp120 sequences from a Phase 3 HIV-1 vaccine trial in North America. Molecular biology and evolution. 2010; 27:417–25.

48. DC Department of Health. DC DOH 90/90/90/50 Plan: Ending the HIV Epidemic in the District of Columbia by 2020. 2017.

49. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, 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(21):7552–7. https://doi.org/10.1073/pnas.0802203105 PMID: 18490657.

50. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. Journal of virology. 1999; 73(12):10489–502. PMID: 10559367.

51. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. Science. 2000; 288:1789–96. PMID: 10846155

52. Papathanasopoulos MA, Hunt GM, Tiemessen CT. Evolution and diversity of HIV-1 in Africa—a review. Virus Genes. 2003; 26(2):151–63. Epub 2003/06/14. PMID: 12803467.

53. Oelrichs RB, Crowe SM. The molecular epidemiology of HIV-1 in South and East Asia. Curr HIV Res. 2003; 1(2):239–48. Epub 2004/03/27. PMID: 15043206.

54. Liao H, Tee KK, Hase S, Uenishi R, Li XJ, Kusagawa S, et al. Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. Virology. 2009; 391(1):51–6. Epub 2009/06/23. S0042-6822(09)00313-4 [pii] https://doi.org/10.1016/j.virol.2009.05.023 PMID: 19540543.

55. Cheng CL, Feng Y, He X, Lin P, Liang SJ, Yi ZQ, et al. Genetic characteristics of HIV-1 CRF01_AE strains in four provinces, southern China. Zhonghua Liu Xing Bing Xue Za Zhi. 2009; 30(7):720–5. Epub 2009/12/05. PMID: 19957600.

56. Marjoram P, Donnelly P. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics. 1994; 136(2):673–83. Epub 1994/02/01. PMID: 8150290; PubMed Central PMCID: PMC1205816.

57. Rosenberg NA, Hirsh AE. On the use of star-shaped genealogies in inference of coalescence times. Genetics. 2003; 164(4):1677–82. Epub 2003/08/22. PMID: 12930771; PubMed Central PMCID: PMC1462671.

58. Brenner BG, Roger M, Moisi DD, Oliveira M, Hardy I, Turgel R, et al. Transmission networks of drug resistance acquired in primary/early stage HIV infection. AIDS. 2008; 22(18):2509–15. Epub 2008/11/14. https://doi.org/10.1097/QAD.0b013e3283121c90 PMID: 19005274; PubMed Central PMCID: PMC2650396.

59. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. The Journal of infectious diseases. 2007; 195(7):951–9. Epub 2007/03/03. JID37441 [pii] https://doi.org/10.1086/512088 PMID: 17330784.

60. Ahumada-Ruiz S, Flores-Figueroa D, Toala-Gonzalez I, Thomson MM. Analysis of HIV-1 pol sequences from Panama: identification of phylogenetic clusters within subtype B and detection of antiretroviral drug resistance mutations. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2009; 9(5):933–40. Epub 2009/06/30. S1567-1348(09)00144-0 [pii] https://doi.org/10.1016/j.meegid.2009.06.013 PMID: 19559103.

61. Bezemer D, van Sighem A, Lukashov VV, van der Hoek L, Back N, Schuurman R, et al. Transmission networks of HIV-1 among men having sex with men in the Netherlands. AIDS. 2010; 24(2):271–82. Epub 2009/12/17. https://doi.org/10.1097/QAD.0b013e328333ddee PMID: 20010072.

62. Cuevas MT, Muñoz-Nieto M, Thomson MM, Delgado E, Iribarren JA, Cilla G, et al. HIV-1 transmission cluster with T215D revertant mutation among newly diagnosed patients from the Basque Country, Spain. Journal of acquired immune deficiency syndromes. 2009; 51(1):99–103. Epub 2009/03/14. https://doi.org/10.1097/QAI.0b013e318199063e PMID: 19282784.

63. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. BMC infectious diseases. 2010; 10:262. Epub 2010/09/09. 1471-2334-10-262 [pii] https://doi.org/10.1186/1471-2334-10-262 PMID: 20822507; PubMed Central PMCID: PMC2940905.

64. Pao D, Fisher M, Hue S, Dean G, Murphy G, Cane PA, et al. Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. AIDS. 2005; 19(1):85–90. Epub 2005/01/01. 00002030-200501030-00010 [pii]. PMID: 15627037.

65. Thomson MM, Vinogradova A, Delgado E, Rakhmanova A, Yakovlev A, Cuevas MT, et al. Molecular epidemiology of HIV-1 in St Petersburg, Russia: predominance of subtype A, former Soviet Union variant, and identification of intrasubtype subclusters. Journal of acquired immune deficiency syndromes. 2009; 51(3):332–9. Epub 2009/04/14. https://doi.org/10.1097/QAI.0b013e31819c1757 PMID: 19363451.

66. Yerly S, Junier T, Gayet-Ageron A, Amari EB, von Wyl V, Gunthard HF, et al. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS. 2009; 23(11):1415–23. Epub 2009/06/03. https://doi.org/10.1097/QAD.0b013e32832d40ad PMID: 19487906.

67. Nguyen L, Hu DJ, Choopanya K, Vanichseni S, Kitayaporn D, van Griensven F, et al. Genetic analysis of incident HIV-1 strains among injection drug users in Bangkok: evidence for multiple transmission clusters during a period of high incidence. Journal of acquired immune deficiency syndromes. 2002; 30 (2):248–56. Epub 2002/06/05. PMID: 12045688.

68. Perez-Losada M, Jobes DV, Sinangil F, Crandall KA, Arenas M, Posada D, et al. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PloS one. 2011; 6(3):e16902. Epub 2011/03/23. https://doi.org/10.1371/journal.pone.0016902 PMID: 21423744; PubMed Central PMCID: PMC3053363.

69. Morgan E, Nyaku AN, D'Aquila RT, Schneider JA. Determinants of HIV phylogenetic clustering in Chicago among young black men who have sex with men from the uConnect cohort. Journal of acquired immune deficiency syndromes. 2017. https://doi.org/10.1097/QAI.0000000000001379 PMID: 28328553.

70. Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, et al. Characterizing HIV transmission networks across the United States. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2012; 55(8):1135–43. Epub 2012/07/13. https://doi.org/10.1093/cid/cis612 PMID: 22784872; PubMed Central PMCID: PMCPMC3529609.