



# Article Insights into HIV-1 Transmission Dynamics Using Routinely Collected Data in the Mid-Atlantic United States

Seble G. Kassaye <sup>1,\*</sup>, Zehava Grossman <sup>2,3,\*</sup>, Priyanka Vengurlekar <sup>1</sup>, William Chai <sup>4</sup>, Megan Wallace <sup>1</sup>, Soo-Yon Rhee <sup>5</sup>, William A. Meyer III <sup>6</sup>, Harvey W. Kaufman <sup>6</sup>, Amanda Castel <sup>7</sup>, Jeanne Jordan <sup>7</sup>, Keith A. Crandall <sup>8</sup>, Alisa Kang <sup>1</sup>, Princy Kumar <sup>1</sup>, David A. Katzenstein <sup>5</sup>, Robert W. Shafer <sup>5</sup>, and Frank Maldarelli <sup>2</sup>

- <sup>1</sup> Department of Medicine, Georgetown University, Washington, DC 20057, USA
- <sup>2</sup> HIV Dynamics and Replication Program, National Cancer Institute, Frederick, MD 21702, USA
- <sup>3</sup> School of Public Health, Tel Aviv University, Tel Aviv 69978, Israel
- <sup>4</sup> Warren Alpert Medical School, Brown University, Providence, RI 02912, USA
- <sup>5</sup> Department of Medicine, Stanford University, Stanford, CA 94305, USA
- <sup>6</sup> Quest Diagnostics, Secaucus, NJ 07094, USA
- Department of Epidemiology, George Washington University, Washington, DC 20052, USA
- <sup>8</sup> Computational Biology Institute, George Washington University, Ashburn, VA 20147, USA
- \* Correspondence: sgk23@georgetown.edu (S.G.K.); zehava.grossman@gmail.com (Z.G.)

Abstract: Background: Molecular epidemiological approaches provide opportunities to characterize HIV transmission dynamics. We analyzed HIV sequences and virus load (VL) results obtained during routine clinical care, and individual's zip-code location to determine utility of this approach. Methods: HIV-1 pol sequences aligned using ClustalW were subtyped using REGA. A maximum likelihood (ML) tree was generated using IQTree. Transmission clusters with  $\leq$ 3% genetic distance (GD) and  $\geq$ 90% bootstrap support were identified using ClusterPicker. We conducted Bayesian analysis using BEAST to confirm transmission clusters. The proportion of nucleotides with ambiguity  $\leq 0.5\%$  was considered indicative of early infection. Descriptive statistics were applied to characterize clusters and group comparisons were performed using chi-square or t-test. Results: Among 2775 adults with data from 2014–2015, 2589 (93%) had subtype B HIV-1, mean age was 44 years (SD 12.7), 66.4% were male, and 25% had nucleotide ambiguity  $\leq 0.5$ . There were 456 individuals in 193 clusters: 149 dyads, 32 triads, and 12 groups with  $\geq$  four individuals per cluster. More commonly in clusters were males than females, 349 (76.5%) vs. 107 (23.5%), *p* < 0.0001; younger individuals, 35.3 years (SD 12.1) vs. 44.7 (SD 12.3), p < 0.0001; and those with early HIV-1 infection by nucleotide ambiguity, 202/456 (44.3%) vs. 442/2133 (20.7%), p < 0.0001. Members of 43/193 (22.3%) of clusters included individuals in different jurisdictions. Clusters  $\geq$  four individuals were similarly found using BEAST. HIV-1 viral load (VL)  $\geq$  3.0 log<sub>10</sub> c/mL was most common among individuals in clusters  $\geq$  four, 18/21, (85.7%) compared to 137/208 (65.8%) in clusters sized 2-3, and 927/1169 (79.3%) who were not in a cluster (p < 0.0001). Discussion: HIV sequence data obtained for HIV clinical management provide insights into regional transmission dynamics. Our findings demonstrate the additional utility of HIV-1 VL data in combination with phylogenetic inferences as an enhanced contact tracing tool to direct HIV treatment and prevention services. Trans-jurisdictional approaches are needed to optimize efforts to end the HIV epidemic.

**Keywords:** molecular epidemiology; phylogenetic analysis; transmission networks; regional transmission dynamics; HIV drug resistance; clinical and phylogenetic data combined; contact tracing tool

# 1. Introduction

In February 2019, there was a call to end the HIV epidemic in the U.S. within a decade, using currently available treatment and prevention modalities [1]. Based on the



Citation: Kassaye, S.G.; Grossman, Z.; Vengurlekar, P.; Chai, W.; Wallace, M.; Rhee, S.-Y.; Meyer, W.A., III; Kaufman, H.W.; Castel, A.; Jordan, J.; et al. Insights into HIV-1 Transmission Dynamics Using Routinely Collected Data in the Mid-Atlantic United States. *Viruses* 2023, *15*, 68. https://doi.org/ 10.3390/v15010068

Academic Editor: Ester Ballana Guix

Received: 21 October 2022 Revised: 21 December 2022 Accepted: 23 December 2022 Published: 25 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). current epidemiology in 2016–2017 of new HIV cases, 57 jurisdictions (48 counties, 7 states, Washington, District of Columbia. (DC), and Puerto Rico) were selected as over 50% of HIV diagnoses were made in these locations [2]. An estimated 14% of people living with HIV (PWH) were not aware of their condition and were thought to contribute approximately 38% of new transmissions with the remaining new infections arising from individuals with known HIV infection who were either not in care or not virally suppressed despite being in care [3]. Continued progress towards elimination of the HIV epidemic requires diagnosis of a higher percentage of PWH. Consistent early detection and treatment with durable viral suppression would reduce transmission [4]. This goal may be enhanced by implementing strategies to detect transmission clusters. Those can be reconstructed using HIV-1 sequence data. Expanded molecular epidemiologic investigation that also incorporates HIV-1 viral load (VL) may (i) identify undiagnosed PWH or those with suboptimal treatment with continued viremia who need antiretroviral treatment for their own health, but also to limit transmission to others [5–7]; and (ii) identify contacts at risk of infection for targeted provision of prophylactic antiretrovirals including appropriate preor post-exposure prophylaxis to prevent HIV acquisition [8–11].

Molecular epidemiologic approaches to characterize HIV transmission have been used in research and public health settings to describe local transmission dynamics [12–14]. Such studies have augmented our understanding of transmission dynamics previously based primarily on epidemiologic data alone. In 2014–2015, molecular epidemiologic approaches were critical to detect and monitor an HIV outbreak leading to 181 new HIV infections in southeastern Indiana related to drug injection [15]. The U.S. Centers for Disease Control and Prevention have since added a molecular surveillance component to routine public health activities to identify HIV outbreaks. Since then, this approach has been used to investigate outbreaks across the U.S. [16]. The approach focuses on identifying rapidly growing transmission clusters, with selection of a very low genetic distance threshold of 0.5% defining closely related viral sequences among individuals with recently diagnosed HIV. Only a small fraction of the new diagnoses invokes a public health investigation, with most investigations conducted by local health jurisdictions and limited to intra-state analyses.

DC has the highest HIV prevalence in the U.S. with an estimated 1.8% of the population living with HIV in 2019 [17]. DC residents have the highest estimated lifetime risk of acquiring HIV infection (1:39) followed in sixth place by neighboring Maryland (1:85) [18]. The adjacent geographic location of DC to Maryland and Virginia and the sharing of work force result in social and cultural interactions across borders, impacting the HIV epidemic. Using public health HIV surveillance data from these three contiguous jurisdictions, we found that approximately 21% of the PWH in the DC metropolitan area received care in more than one jurisdiction demonstrating the converging and overlapping epidemics in this high HIV transmission region [19]. Active movement of PWH in the metropolitan DC area is demonstrated with routine surveillance data that demonstrate outmigration of approximately 42% of PWH originally diagnosed in DC in 2019, and in-migration contributing 17% of the 12,408 PWH actively receiving care in DC [17]. Investigating HIV transmission dynamics in this region, we previously analyzed HIV pol sequence data from individuals who had enrolled in clinical or observational studies from early in the HIV pandemic [20–22]. Transmission clusters were common among men who have sex with men (MSM) and younger individuals. Individuals with evidence of recent infection based on low proportion of nucleotide ambiguity were more likely to belong in a transmission cluster [23,24].

In this study, we used HIV-1 sequences generated for clinical purposes along with limited demographic data, to characterize HIV transmission patterns in this high prevalence region. Our supposition is this approach would identify transmission clusters that traverse state borders, across this porous interconnected region, highlighting the importance of trans-jurisdictional surveillance to maximize its benefit in reducing HIV infection. We further postulate that laboratory surrogate markers for HIV care continuum metrics such as HIV-1 VL will provide an important discriminating factor to guide clinical and public health interventions.

#### 2. Materials and Methods

#### 2.1. Study Population

This is a retrospective molecular epidemiology study, using HIV sequencing data generated as part of routine clinical care from PWH receiving medical treatment in the mid-Atlantic region including DC, Maryland (MD) and Virginia (VA). HIV pol data sequenced using the Sanger method from 2008–2015 (protease, reverse transcriptase, and, when available, integrase) were obtained under a data use agreement with one national clinical laboratory. Additional information including age, sex, state, zip code, and year of service was available to the research team. A subset of individuals had quantitative HIV-1 RNA VL testing performed for routine clinical care at the Clinical Laboratory Improvement Amendments of 1988 (CLIA) certified national reference laboratory using a U.S. Food and Drug Administration cleared test method from Roche Molecular Systems (Branchburg, NJ, USA) with a lower limit of detection of 1.3 log<sub>10</sub> copies/mL (c/mL). Data from samples obtained within a year of the sequencing date were included in this analysis. When zip code or state of residence were not provided, we used where medical service was provided as a location marker.

### 2.2. Sequence Alignment and Characterization

The earliest sequence available for each individual age  $\geq$ 18 years, obtained during 2014–2015 were used. We aligned protease (Pr) and reverse transcriptase (RT) regions of HIV-1 pol sequences in Clustal-Omega version clustal-omega-1.2.1using HXB2 Pr-RT as the reference sequence spanning nucleotides 2252–3443 of the HXB2 reference sequence [25]. The Sierra pipeline to the Stanford HIV Database was used in sequence evaluation (subtype, gene coverage, and quality) and assessment of HIV-1 drug resistance [26]. The proportion of nucleotides with ambiguity  $\leq$ 0.5% was used to indicate early or recent HIV infection within one year of infection [23,24].

## 2.3. Cluster Analyses

# 2.3.1. Maximum Likelihood

A maximum-likelihood (ML) phylogenetic tree was constructed using IQ-TREE multicore [27] v1.5.0-beta with the general time-reversible (GTR) substitution model with invariable sites (I) and gamma (G) distribution (GTR+I+G) with 1000 bootstraps [28]. The FASTA sequences and the IQtree tree files were uploaded into ClusterPicker (v1.2.3) to identify sequences that were in clusters with  $\geq$ 95% bootstrap node support for pre-specified genetic distance (GD) thresholds of 0.5%, 1.5%, 2.0%, 3.0% and 4.5% [27].

#### 2.3.2. Bayesian Analysis

To probe the likelihood that clusters actually represented local transmission networks, the specificity of cluster designation was assessed using a probability-based approach, the Bayesian evolutionary analysis sampling trees (BEAST) [29]. We identified sequences that were in groups of four or more at a GD  $\leq$ 3% using the ML method. In order to verify the veracity and robustness of the transmission clusters we performed a blast search in the GenBank HIV sequence database for each sequence in this subset and selected the ten sequences most similar to each sequence in our data set [30]. Excluding duplicate sequences, sequences without dates, and those of significantly shorter length, we were left with 120 GenBank sequences. A total of 183 sequences comprising the study and GenBank sequences, with an additional subtype D reference sequence, were used for the BEAST analysis. The sequences were aligned as above. BEAST analysis was performed using the GTR+I+G substitution model, uncorrelated log-normal relaxed molecular clock [31], the parametric tree prior coalescent assumption of the Gaussian Markov random field (GMRF) skyride [32,33], and Markov chain Monte Carlo (MCMC) of iterations for 135 million states

to achieve a posterior effective sample size of 302. Clusters of sequences that grouped together with a posterior probability nodal support of 95% or greater were identified and compared to the original clusters identified using the ML method to confirm the robustness of grouping.

#### 2.3.3. Network Analysis

Network analysis was performed to visually display the effect of different GD thresholds on the network composition. The aligned HIV sequence data were uploaded into HIV-TRACE [34] to implement the Tamura Nei 93 (TN93) substitution model that determines the pairwise distances between sequences [35]. The ambiguities between nucleotide pairs were resolved at an ambiguity fraction of 0.05 and the overlap was set at 500. The program generated a list of the nodes (represents sequences) and edges (connections between the nodes) used to derive network figures. The node and link attribute files were integrated with the remaining metadata and uploaded in the web-based HIV-TRACE visualization tool. Network figures were generated for GD thresholds of 0.5%, 1.5%, 2.0%, and 3.0%.

# 2.4. Statistical Analysis

We used R script to generate descriptive statistics to characterize the frequency and distribution of demographic and cluster data within the study population [36]. We conducted comparisons between groups using t-test and chi square tests for continuous and categorical data, respectively.

#### 2.5. Ethical Considerations and Disclosures

This study was reviewed and approved by the Georgetown University Institutional Review Board under the terms of the data use agreement allowing the use of limited personal identifiers including zip code, with appropriate data security safeguards instituted by the Georgetown University Information Systems. Data from zip codes with fewer than 20 individuals were omitted to avoid unintended loss of privacy for these individuals. The terms of use limit access to these data to the study team and their delegates that include members of the national reference laboratory that generated these data. Publications resulting from this data set required review and approval by the national reference laboratory.

## 3. Results

The analysis included 2775 individuals aged  $\geq$ 18 years with genotype sequences available from 2014–2015 (Table 1). The mean age was 42.9 years (SD: 12.7, range: 18–79). The majority were males, (N = 1843, 66.4%). Residence was in MD for 2124 participants (76.5%); VA for 375 (13.5%); and DC for 242 (8.7%). Subtype B HIV-1 was the predominant subtype, 2589 (93.3%), of which 643/2589 (24.8%) sequences had the proportion of nucleotide ambiguity  $\leq$ 0.5 suggestive of recent infection.

Table 1. Cohort Socio-demographic and HIV-1 Characteristics, 2014–2015.

		Subtype B HIV-1 N = 2589	Non-B Subtypes N = 186	Total N = 2775
Gender	Female	823 (31.8%)	105 (56.5%)	928 (33.4%)
	Male	1762 (68.1%)	81 (43.5%)	1843 (66.4%)
	Not Reported	4 (0.15%)	0	4 (0.14%)
State	DC	232 (9%)	10 (5.4%)	242 (8.7%)
	MD	1988 (76.8%)	136 (73.1%)	2124 (76.5%)

		Subtype B HIV-1 N = 2589	Non-B Subtypes N = 186	Total N = 2775
	VA	338 (13.1%)	37 (19.9%)	375 (13.5%)
	Not Reported	31 (1.2%)	3 (1.6%)	34 (1.2%)
Age	18–37 years	902 (34.8%)	64 (34.4%)	966 (34.8%)
	38–52 years	1014 (39.1%)	91 (48.9%)	1105 (39.8%)
	$\geq$ 53 years	673 (26.0%)	31 (16.6%)	704 (25.3%)
Ambiguity (Amb) *	$Amb \le 0.5$	643 (24.8%)	29 (15.6%)	672 (24.2%)
	Amb 0.5–0.99	443 (17.1%)	26 (13.9%)	469 (16.9%)
	$Amb \geq 1$	1503 (58.1%)	130 (69.8%)	1633 (58.8%)
	Female	107 (4.1%)	1 (0.5%)	108 (3.9%)
Clustered Genetic Distance 3.0%	Male	349 (13.5%)	3 (1.6%)	352 (12.7%)
	Total	456 (17.6%)	4 (2.2%)	460 (16.6%)

Table 1. Cont.

\* Ambiguity (Amb)—proportion of nucleotides within the sequence with ambiguity. Percentages may not add up to 100% due to rounding.

#### 3.1. Cluster Data

Among 2589 individuals with subtype B HIV-1, 456 (17.6%) aligned within 193 clusters using the ML method (GD 3%, 90% bootstrap support) (Table 2). The majority of clusters were dyads (149) and triads (32). There were 12 clusters of  $\geq$ 4 individuals: 4 clusters of 4; 5 clusters of 5; and 1 cluster each with 6, 7, and 8 individuals. Proportionally, males were more likely to be in clusters than females, 349/1762 (19.8%) vs. 107/823 (13.0%), p < 0.0001. The age distribution of individuals based on cluster designation is demonstrated in Figure 1. Individuals in clusters were younger, with a mean age of 35.3 years (SD 12.1), compared to 44.7 years (SD 12.3) among those not in clusters (p < 0.0001). Individuals in clusters were more likely to have low proportion of nucleotides with ambiguity,  $\leq 0.5\%$ , suggestive of recent infection, with 201/456 (44.1%) having the low ambiguity score among those in clusters compared with 442/2133 (20.7%) among those not in clusters, p < 0.0001. The majority of individuals in clusters with low sequence ambiguity were males, 81% (163/201), and only 19% (38/201) were female. There was no statistically significant difference in the proportion of individuals in a cluster by state of residence. Of 193 clusters, 43 (22.3%) included individuals residing in different states: DC-MD in 20 clusters; MD-VA in 18 clusters; VA-DC in 4 clusters; and DC-MD-VA in 1 cluster.

When using a more stringent genetic distance threshold of 2%, the number of clusters decreased to a total of 109:88 dyads, 14 triads, four clusters of four, and three clusters of five individuals. Even fewer clusters (72) were identified using a genetic distance cutoff of 1.5%, with 59 dyads, 10 triads, and 3 clusters of 4 individuals. At the genetic distance threshold of 0.5%, only 13 dyads and 1 triad were identified.

		Clustered N = 456	Non-Clustered N = 2133	Total N = 2589	Chi-Square
Gender	Female	107 (23.5%)	716 (33.6%)	823 (31.8%)	p < 0.0001
	Male	349 (76.5%)	1413 (66.2%)	1762 (68.1%)	
	Mean (SD)	35.3 (12.1)	44.7 (12.3)	43 (12.8)	p < 0.0001
Age	18–37 years	289 (63.4%)	613 (28.7%)	902 (34.8%)	
	38–52 years	114 (25.0%)	900 (42.2%)	1014 (39.2%)	<i>p</i> < 0.00001
	$\geq$ 53 years	53 (11.6%)	620 (29.1%)	673 (26.0%)	
Ambiguity (Amb) Score	Amb $\leq 0.5$	202 (44.3%)	442 (20.7%)	644 (24.9%)	
	Amb 0.5–0.99	99 (21.7%)	343 (16.1%)	442 (17.1%)	<i>p</i> < 0.00001
	$Amb \ge 1$	155 (34.0%)	1348 (63.2%)	1503 (58.1%)	-
State	DC	36 (7.9%)	196 (9.2%)	232 (9%)	
	MD	370 (81.1%)	1618 (75.9%)	1988 (76.8%)	- 
	VA	48 (10.5%)	290 (13.6%)	338 (13.1%)	p = 0.09 (105)
	NR	2 (0.4%)	29 (1.4%)	31 (1.2%)	-

**Table 2.** Comparative characteristics of individuals with subtype B HIV-1 identified within and outside clusters with a genetic distance threshold of 3%.

Percentages may not add up to 100% due to rounding.



**Figure 1.** Distribution of individuals over clustered, non-clustered and sex by age group, genetic distance 3%. The bar graph demonstrates the frequency and age distribution of individuals in transmission clusters. The proportion of those in clusters (shown in gray) compared with those not in clusters (shown in yellow) is higher among younger individuals.

# 3.2. BEAST Analysis

All the clustered sequences with  $\geq$ 4 members found using the maximal likelihood approach with a genetic distance threshold of 3% remained in clusters, with posterior probability support  $\geq$ 0.95 (Figure 2a,b) when analyzed using BEAST. Twenty-one out of the sixty-two individuals had HIV-1 VL available, and 18/21 (85.7%) had HIV-1 VL  $\geq$  3.0 log<sub>10</sub> c/mL. There was evidence of transmitted K103N/K103S drug resistance mutations within a large transmission cluster as shown in Figure 2b, with the resistance mutation noted among multiple members of the cluster. The range and frequency of major HIV-1 resistance mutations are shown in Figure S1.



Figure 2. Phylogenetic trees of sequences in clusters with  $\geq 4$  members. (a). Study population sequences and closest related sequences from GenBank. Bayesian phylogenetic tree of 62 subtype B HIV-1 sequences that were identified in clusters size four or more based on GD 3% using the maximal likelihood method were included alongside the most closely related sequences from the GenBank HIV Sequence database at Los Alamos (labeled "LA"). Bayesian evolutionary analysis sampling trees [29] analysis was performed using the GTR+I+G substitution model, uncorrelated log-normal relaxed molecular clock, tree coalescent assumption of the Gaussian Markov random field (GMRF) skyride [31–33], and run of 135 million states to achieve an effective sample size of 302. The cluster relationships were maintained with high posterior probability of  $\geq$ 95% when analyzed using this BEAST approach. Male participants are indicated in purple and females indicated in green, and sequences from Los Alamos indicated in black. (b). Cluster characteristics: HIV-1 drug resistance mutations and viral load. HIV-1 viral load data are written in black with the corresponding sequence when available. Among the 21 individuals with HIV-1 VL data available, 18 (85.7%) had levels >3 log10 c/mL. Four HIV-1 drug resistance associated mutations were identified in two clusters including the resistance mutations Y188H, K101E, K103N/S, and G190S. One cluster had multiple cluster members with drug resistance mutations.

# 3.3. Network Analysis

Using HIV-TRACE we identified 519 nodes linked to at least one other node at GD 3.0% (Figure 3a–d). A total of 189 clusters were formed, of which the largest had 33 nodes with 46 connected links at a mean genetic distance of 2.8% and 1.39 links per node. The second-largest dense cluster consisted of 15 nodes making 38 links with a higher transmission speed of 2.53 links per node. The male transmission clusters comprised 66% of identified networks and the female clusters were 10.5%. Recent infection based on the proportion of nucleotide ambiguity  $\leq$ 0.5% identified a higher number of nodes from 2014 than from 2015. Networks spanning more than one jurisdiction were identified in 75 nodes comprising 14% of individuals identified to be in a transmission cluster.



**Figure 3.** HIV-TRACE Transmission Networks by Genetic Distance. Cluster networks configured in HIV-TRACE from HIV pol sequences are inferred using genetic distances between sequences. Cluster networks demonstrate decreasing linkages with lower genetic distance cutoffs ((**a**)  $\leq$ 3%; (**b**)  $\leq$ 2%; (**c**)  $\leq$ 1.5%; and (**d**)  $\leq$ 0.5%), although multijurisdictional networks are identified at all genetic distance cutoffs. Using the less restrictive genetic distance cutoff of 3% provides greater detail and the resulting network visualization provides context and highlights connections that more completely represent potentially important HIV-1 transmission dynamics.

#### 3.4. Transmission Clusters and Viremia

HIV-1 VL was available for 54% (N = 1394) of the study population (Table 3). The majority, 78% (1091/1394) had VL  $\geq$ 3.0 log<sub>10</sub> copies/milliliter (c/mL); 3.5% (49/1394) had VL 2.7–3.0 log<sub>10</sub> c/mL; 17.4% (242/1394) had VL 1.3–2.7 log<sub>10</sub> c/mL, and 0.01% (12/1394) individuals had VL <1.3 log<sub>10</sub> c/mL. The median VL was 4.4 log<sub>10</sub> c/mL (IQR 25–75%: 3.2–4.99). The quantitative VL differed by cluster size with the highest VL among those in clusters of four or greater, median VL 4.6 log<sub>10</sub> c/mL, compared with 3.9 c/mL log<sub>10</sub> for those in clusters of 2 and 3, and 4.4 log<sub>10</sub> c/mL varied among different-sized cluster groups and was highest in the group with cluster size  $\geq$ 4, 85.7% (18/21); 65.8% (137/208) among those in cluster size 2–3; and 79.3% (927/1169) among those individuals not in a cluster (*p* < 0.0001).

Cluster Data		HIV Viral Load (VL)			
	Number of HIV VL Tests N = 1394	Median in log <sub>10</sub> Copies/mL (IQR 25–75%)	$\geq$ 3.0 log $_{10}$ c/mL	<3.0 log <sub>10</sub> c/mL	
Overall HIV VL	1394	4.4 (3.2–4.99)	1091 (78.3%)	303 (21.7%)	
Cluster size $\geq$ 4 (N = 62)	21	4.6 (3.8–5.2)	18 (85.7%)	3 (14.3%)	
Cluster size 3 (N = 96)	47	3.1 (2.3–4.6)	26 (55.3%)	21 (44.7%)	
Cluster size 2 (N = 298)	161	4.0 (2.1–4.8)	111 (68.9%)	50 (31.1%)	
Cluster size 2 and 3 (N = 394)	208	3.9 (2.2–4.8)	137 (65.9%)	71 (34.1%)	
Not in clusters (N = 2133)	1169	4.4 (3.3–5.0)	927 (79.3%)	242 (20.7%)	

**Table 3.** Relationship between HIV-1 viral load (VL) and transmission cluster size (genetic distance 3%).

# 4. Discussion

This collaborative project was designed to determine the utility of regional sequence and HIV-1 VL data generated as part of clinical care for the characterization of HIV transmission patterns, given the highly interconnected mid-Atlantic metropolitan DC area. An important element of our research is the public-private-academic partnership that allowed us to access this representative regional dataset generated in the course of routine clinical care with multi-jurisdictional representation under a legally binding data sharing agreement. Enrolling such a comprehensive cohort to obtain sequence and laboratory data for research purposes only would have been challenging and costly. With the safeguards that we have placed both in the design and implementation of the study to protect individual's privacy, we demonstrate the feasibility and utility of such collaborations. Our findings of the relative frequency of trans-jurisdictional transmission clusters in this highly interconnected region highlights the need to overcome restrictions in data sharing based on traditional jurisdictional borders, regional or national. Such restrictions limit collaborations and innovative strategies that span jurisdictions are essential to adequately respond to the call to end the HIV epidemic.

We demonstrate the high fidelity of rapid maximal likelihood phylogenetic approaches for identifying transmission clusters, further validated with the more time intensive probability-based sequence and network analyses. Overall, we rarely identified large transmission clusters, with most clusters containing two or three sequences in this convenience sample of sequences. There were few transmission clusters with size 4 or greater. This is likely the result of the sampling framework that was not focused exclusively on newly diagnosed PWH, but rather included all individuals for whom HIV sequence data were generated for clinical management, including those with chronic HIV-1 infection. In addition to MSM and younger individuals, we also identified a subset of individuals in their fourth or higher decades of life, including women, in transmission clusters. Interventions to end the HIV epidemic in the region will need targeting also these important and often missed sub-populations that remain at risk for HIV-1 infection.

Evidence supporting the validity of rapid maximal likelihood based phylogenetic methods shows that large data sets can be analyzed to support contact tracing within relatively short periods of time, with computational capacity that is generally available in many public health settings. In this context, more recent uses of HIV sequence data have emerged that focus on identifying new HIV outbreaks. These projects have primarily used very narrow genetic distance thresholds of 0.5–1.5% [13,14]. While this may be appropriate for identifying emerging large outbreaks, these cutoffs yield o reduced case numbers for investigation when applied to our regional dataset, with only fourteen clusters (thirteen dyads and one triad) fitting the most stringent metric. Most individuals found in larger clusters, identified using the genetic distance threshold of 3% and verified using a network and probability based phylogenetic analysis with BEAST, would have been missed. A significant proportion of individuals with evidence of linked transmission in our analyses demonstrated very low proportion of nucleotide ambiguity, suggesting that some of these infections indeed could have been more recently acquired. The relatedness of individuals within clusters using a less stringent genetic distance cutoff could be epidemiologically important in a mature epidemic such as in the mid-Atlantic region, where linked transmissions may span decades. The consistent findings despite using different methods of inference to determine sequence relatedness demonstrate justifies using a 3% genetic distance threshold to robustly identify relevant transmission clusters. The persistence of the identified clusters after the addition of the most closely related sequences from GenBank suggests supports likely epidemiologic linkages among the persons from whom sequences were obtained. While use of currently accepted genetic distance thresholds in public health molecular surveillance programs are set to identify emerging outbreaks, we demonstrated how these thresholds would miss identifying epidemiologically important targets for intervention in the context of an established generalized epidemic. Relaxing the stringency beyond 3%, though, appeared counter-productive in our case as it may result in artifactual linkages, a finding that has previously been reported by others [13].

The addition of selected laboratory testing such as HIV-1 VL further increases the potential utility of molecular sequence data. HIV-1 VL is an important marker of linkage, engagement and retention in care in the context of the HIV care continuum [37]. The most recent report indicates that 57% of the estimated 1.2 million individuals with HIV in the U.S. had viral suppression [38]. Ideally, all individuals with detectable viral load would benefit from interventions to support linkage to care. We demonstrate how overlaying molecular epidemiology with HIV-1 VL data allows for further risk stratification such that individuals with evidence for recent infection based on sequence ambiguity, cluster membership, and high HIV-1 VL would be preferentially targeted for support services if resources are constrained. Natural history studies have demonstrated HIV-1 VL of approximately 3  $\log_{10}$  c/mL as the threshold for transmission risk, with escalating VL associated with increasing transmission risk [39]. More recent data have identified that ART use with durable viral suppression decreases risk of transmission [5–7]. The relatively high proportion of individuals within clusters with HIV-1 VL  $\geq 3\log_{10}$  copies/mL highlights the need for targeted and effective ART programs. While all individuals should receive ART for their own health, those who are in transmission clusters with high VL could be prioritized for adherence support and interventions to halt further transmission. Individuals who are in dyads and triads are often not considered high value targets for intervention and contact tracing but are more frequently identified than the larger transmission networks that are currently the target of public health interventions [40]. The addition of HIV-1 VL data to transmission cluster data allows one to discern which individuals are biologically more likely to transmit and hence should be prioritized for supportive intervention. Such an approach may allow us to break through the current plateau in new HIV-1 diagnoses towards the goal of ending the HIV epidemic.

Earlier molecular epidemiology studies in clinical settings have demonstrated the decreased transmission risk with antiretroviral treatment and viral suppression [41]. We demonstrate how molecular epidemiologic elements can bolster efforts to identify individuals with potential for transmission at the local population level, using routinely collected clinical management laboratory data. Ideally, all individuals with viremia should receive interventions to support engagement in care. In the absence of unlimited resources, however, priority could be directed to those with elevated viremia for whom there is molecular epidemiologic support for risk for transmission [42], thereby providing an additional incen-

tive for intervening both for the sake of the individual's health and to reduce transmission to others. Such active monitoring with expanded use of molecular data in conjunction with clinical metadata such as HIV-1 VL would contextualize and refine assessments of transmission risk [42]. Individuals in larger cluster sizes who remain viremic have greater potential to transmit HIV, risk that could be effectively mitigated with appropriate outreach to ensure engagement in care and effective suppressive antiretroviral therapy. Discovery of risk before transmission and successful mitigation represents the holy grail of infectious disease control, and successful implementation would serve to accelerate the progress towards ending the HIV epidemic.

One of the limitations of our study is the sampling frame only covers an estimated quarter of data that are generated from individuals in the region. While the lower sampling density may decrease the number of transmission clusters that are identified due to missing data, our findings suggest that the identified clusters are robust. The persistence of the identified clusters after the addition of the most closely related sequences from GenBank suggests that the inferred relationships were not simply due to sampling bias. Our data are limited to PWH receiving care and does not capture the important subset of PWH who are not in care. Although having near complete sampling from the affected community would result in finding more transmission clusters, even this subset yields important information with actionable findings, of the kind pivotal to ending the HIV epidemic efforts. While individual data related to social contacts was not available in our data set, partner information is often under-reported [43]. The dataset does not include ART histories and thus limits full characterization of the HIV care-continuum for the studied population.

Lastly, there has been a mixed response to using molecular epidemiology to characterize transmission dynamics both in the scientific community and among advocacy groups [44–47]. It remains critical to assess any presumed negative consequences. Our analysis validates the utility of genotypic and viral load data even in the absence of detailed epidemiologic data and may be an important adjunctive public health tool in areas where concerns related to stigma and marginalization may limit the engagement of people living with HIV [48].

## 5. Summary

Optimal use of molecular genotypic data should include more relaxed genetic distance cutoffs to identify putative transmission clusters. These data should be analyzed in the context of additional routine clinical and laboratory data that are collected during the implementation and monitoring of standard of care treatment and prevention programs. Such an approach would provide context and identify opportunities for intervention to guide care teams to provide differentiated care services and optimize resource allocation and utilization. Only through bold and integrated interventions will we reach the next milestone towards ending the HIV epidemic.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v15010068/s1, Figure S1: Frequency of major reverse transcriptase and protease mutations in B and Non-B Subtype HIV-1.

**Author Contributions:** Conceptualization, S.G.K. and D.A.K.; methodology, S.-Y.R. and R.W.S.; software, P.V., S.-Y.R. and A.K.; formal analysis, Z.G. and P.V.; investigation, S.G.K. and Z.G.; resources, A.K.; data curation, P.V., W.A.M.III, H.W.K. and A.K.; writing—original draft preparation, S.G.K.; writing—review and editing, S.G.K., Z.G., P.V., W.C., M.W., S.-Y.R., W.A.M.III, H.W.K., A.C., J.J., K.A.C., A.K., P.K., R.W.S. and F.M.; supervision, S.G.K., D.A.K.; acquisition, S.G.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has received funding from the Washington, DC, Metropolitan Site of the Women's Interagency Human immunodeficiency Virus (HIV) Study (WIHS) (multi-principal investigators, M. Y. and S. G. K.) is supported by the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID; U01AI034994) and co-funded by the National Cancer Institute, the National Institute on Drug Abuse, and Eunice Kennedy Shriver National Institute

of Child Health and Human Development. This project was funded in part with federal funds (grant KL2TR000102, previously KL2RR031974) from the National Center for Research Resources and the National Center for Advancing Translational Sciences, a trademark of DHHS, part of the Re-Engineering the Clinical Research Enterprise. This research has also been facilitated by the services and resources provided by 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, NIDDK, NIMHD, NIDCR, NINR, FIC and OAR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**Informed Consent Statement:** Patient consent was waived as the study made use of existing data that were used for research purposes as a limited data set with limited identifiers: ZIP codes, birth dates, or other dates only. Consent was not sought as it was not thought to be feasible given the retrospective design of the study, and the research could not be practicably conducted without the waiver of consent.

**Data Availability Statement:** The data were made available to the study team through a Data Licensing and Services agreement between QUEST Diagnostics and Georgetown University and does not permit transfer of the data. The national reference laboratory has rights to the data.

Acknowledgments: In memory of our dear friend, mentor, and colleague David Katzenstein, who died from COVID-19 in January 2021 prior to the publication of this manuscript. He was deeply engaged in discussions from study inception to review of the findings and was included on an earlier abstract and poster presentation at the 24th Workshop on Virus Evolution and Molecular Epidemiology Workshop, August 2019, Hong Kong. We acknowledge his significant contributions to the work presented in this paper and recognize his deep and lasting contributions to the field.

**Conflicts of Interest:** SK: Development of HIV educational materials with Integritas Communications, LLC; WAM and HWK are employees of and own stock in Quest Diagnostics. The other co-authors declare no conflict of interest.

# References

- 1. Fauci, A.S.; Redfield, R.R.; Sigounas, G.; Weahkee, M.D.; Giroir, B.P. Ending the HIV Epidemic: A Plan for the United States. *JAMA* 2019, *321*, 844. [CrossRef] [PubMed]
- Office of Infectious Disease and HIV/AIDS Policy, HHS. About Ending the HIV Epidemic in the U.S.: Priority Jurisdictions: Phase 1; Date Last Updated: 3 November 2020. Available online: https://www.hiv.gov/federal-response/ending-the-hiv-epidemic/ jurisdictions/phase-one (accessed on 24 November 2021).
- 3. Centers for Disease Control and Prevention. CDC Vital Signs. Ending the HIV Epidemic: HIV Treatment Is Prevention. 2019. Available online: https://www.cdc.gov/vitalsigns/end-hiv/pdf/vs-0318-end-hiv-H.pdf (accessed on 10 December 2021).
- 4. Jenness, S.M.; Johnson, J.A.; Hoover, K.W.; Smith, D.K.; Delaney, K.P. Modeling an integrated HIV prevention and care continuum to achieve the Ending the HIV Epidemic goals. *AIDS* **2020**, *34*, 2103–2113. [CrossRef] [PubMed]
- Donnell, D.; Baeten, J.M.; Kiarie, J.; Thomas, K.K.; Stevens, W.; Cohen, C.R.; McIntyre, J.; Lingappa, J.R.; Celum, C. Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: A prospective cohort analysis. *Lancet Lond Engl.* 2010, 375, 2092–2098. [CrossRef] [PubMed]
- Cohen, M.S.; Chen, Y.Q.; McCauley, M.; Gamble, T.; Hosseinipour, M.C.; Kumarasamy, N.; Hakim, J.G.; Kumwenda, J.; Grinsztejn, B.; Pilotto, J.H.; et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. N. Engl. J. Med. 2011, 365, 493–505. [CrossRef]
- Cohen, M.S.; Chen, Y.Q.; McCauley, M.; Gamble, T.; Hosseinipour, M.C.; Kumarasamy, N.; Hakim, J.G.; Kumwenda, J.; Grinsztejn, B.; Pilotto, J.H.; et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N. Engl. J. Med.* 2016, 375, 830–839. [CrossRef]
- Centers for Disease Control and Prevention. US Public Health Service: Preexposure Prophylaxis for the Prevention of HIV Infection in the United States—2021 Update: A Clinical Practice Guideline. Available online: https://www.cdc.gov/hiv/pdf/ risk/prep/cdc-hiv-prep-guidelines-2021.pdf (accessed on 6 December 2022).
- Grant, R.M.; Lama, J.R.; Anderson, P.L.; McMahan, V.; Liu, A.Y.; Vargas, L.; Goicochea, P.; Casapía, M.; Guanira-Carranza, J.V.; Ramirez-Cardich, M.E.; et al. Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. N. Engl. J. Med. 2010, 363, 2587–2599. [CrossRef]

- Baeten, J.M.; Donnell, D.; Ndase, P.; Mugo, N.R.; Campbell, J.D.; Wangisi, J.; Tappero, J.W.; Bukusi, E.A.; Cohen, C.R.; Katabira, E.; et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *N. Engl. J. Med.* 2012, 367, 399–410. [CrossRef]
- 11. Abbasi, J. Long-acting Cabotegravir Shot Prevents HIV Among Women. JAMA 2020, 324, 2247. [CrossRef] [PubMed]
- 12. Little, S.J.; Pond, S.L.K.; Anderson, C.M.; Young, J.A.; Wertheim, J.O.; Mehta, S.R.; May, S.; Smith, D. Using HIV Networks to Inform Real Time Prevention Interventions. Harrigan PR, ed. *PLoS ONE* **2014**, *9*, e98443. [CrossRef]
- 13. Wertheim, J.O.; Kosakovsky Pond, S.L.; Forgione, L.A.; Mehta, S.R.; Murrell, B.; Shah, S.; Smith, D.M.; Scheffler, K.; Torian, L.V. Social and Genetic Networks of HIV-1 Transmission in New York City. *PLoS Pathog* **2017**, *13*, e1006000. [CrossRef]
- Lubelchek, R.J.; Hoehnen, S.C.; Hotton, A.L.; Kincaid, S.L.; Barker, D.E.; French, A.L. Transmission Clustering Among Newly Diagnosed HIV Patients in Chicago, 2008 to 2011: Using Phylogenetics to Expand Knowledge of Regional HIV Transmission Patterns. JAIDS J. Acquir. Immune Defic. Syndr. 2015, 68, 46–54. [CrossRef] [PubMed]
- 15. Peters, P.J.; Pontones, P.; Hoover, K.W.; Patel, M.R.; Galang, R.R.; Shields, J.; Blosser, S.J.; Spiller, M.W.; Combs, B.; Switzer, W.M.; et al. HIV Infection Linked to Injection Use of Oxymorphone in Indiana, 2014–2015. *N. Engl. J. Med.* **2016**, 375, 229–239. [CrossRef]
- Oster, A.M.; Lyss, S.B.; McClung, R.P.; Watson, M.; Panneer, N.; Hernandez, A.L.; Buchacz, K.; Robilotto, S.E.; Curran, K.G.; Hassan, R.; et al. HIV Cluster and Outbreak Detection and Response: The Science and Experience. *Am. J. Prev. Med.* 2021, *61*, S130–S142. [CrossRef]
- District of Columbia Department of Health HIV/AIDS, Hepatitis, STD, and TB Administration. Annual Epidemiology and Surveillance Report; Data Through December 2019. 2020. Available online: https://dchealth.dc.gov/sites/default/files/dc/ sites/doh/publication/attachments/2020-HAHSTA-Annual-Surveillance-Report.pdf (accessed on 26 July 2022).
- Singh, S.; Hu, X.; Hess, K. Estimating the Lifetime Risk of a Diagnosis of HIV Infection in the United States. 2022. Available online: https://www.croiconference.org/abstract/estimating-the-lifetime-risk-of-a-diagnosis-of-hiv-infection-in-the-unitedstates/ (accessed on 19 December 2022).
- Ocampo, J.M.F.; Smart, J.C.; Allston, A.; Bhattacharjee, R.; Boggavarapu, S.; Carter, S.; Castel, A.D.; Collmann, J.; Flynn, C.; Hamp, A.; et al. Improving HIV Surveillance Data for Public Health Action in Washington, DC: A Novel Multiorganizational Data-Sharing Method. *JMIR Public Health Surveill.* 2016, 2, e3. [CrossRef]
- Kassaye, S.G.; Grossman, Z.; Balamane, M.; Johnston-White, B.; Liu, C.; Kumar, P.; Young, M.; Sneller, M.C.; Sereti, I.; Dewar, R.; et al. Transmitted HIV Drug Resistance Is High and Longstanding in Metropolitan Washington, DC. *Clin. Infect. Dis.* 2016, 63, 836–843. [CrossRef] [PubMed]
- 21. Gibson, K.M.; Steiner, M.; Kassaye, S.; Maldarelli, F.; Grossman, Z.; Perez-Losada, M.; Crandall, K.A. A 28-Year History of HIV-1 Drug Resistance and Transmission in Washington, DC. *Front. Microbiol.* **2019**, *10*, 369. [CrossRef] [PubMed]
- Gibson, K.M.; Jair, K.; Castel, A.D.; Bendall, M.L.; Wilbourn, B.; Jordan, J.A.; Crandall, K.A.; Pérez-Losada, M.; the DC Cohort Executive Committee. A cross-sectional study to characterize local HIV-1 dynamics in Washington, DC using next-generation sequencing. *Sci. Rep.* 2020, *10*, 1989. [CrossRef]
- 23. Kouyos, R.D.; von Wyl, V.; Yerly, S.; Böni, J.; Rieder, P.; Joos, B.; Taffé, P.; Shah, C.; Bürgisser, P.; Klimkait, T.; et al. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2011**, *52*, 532–539. [CrossRef]
- 24. Andersson, E.; Shao, W.; Bontell, I.; Cham, F.; Cuong, D.D.; Wondwossen, A.; Morris, L.; Hunt, G.; Sönnerborg, A.; Bertagnolio, S.; et al. Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet Infect. Dis.* **2013**, *18*, 125–131. [CrossRef]
- Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 2011, 7, 539. [CrossRef]
- 26. Rhee, S.Y.; Gonzales, M.J.; Kantor, R.; Betts, B.J.; Ravela, J.; Shafer, R.W. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 2003, *31*, 298–303. [CrossRef]
- 27. Ragonnet-Cronin, M.; Hodcroft, E.; Hué, S.; Fearnhill, E.; Delpech, V.; Brown, A.J.L.; Lycett, S. Automated analysis of phylogenetic clusters. *BMC Bioinform.* 2013, 14, 317. [CrossRef]
- 28. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
- 29. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 2007, 7, 214. [CrossRef] [PubMed]
- 30. Sayers, E.W.; Cavanaugh, M.; Clark, K.; Ostell, J.; Pruitt, K.D.; Karsch-Mizrachi, I. GenBank. *Nucleic Acids Res.* 2019, 48, D84–D86. [CrossRef]
- 31. Drummond, A.J.; Ho, S.Y.W.; Phillips, M.J.; Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* 2006, *4*, e88. [CrossRef]
- 32. Drummond, A.J.; Nicholls, G.K.; Rodrigo, A.G.; Solomon, W. Estimating Mutation Parameters, Population History and Genealogy Simultaneously from Temporally Spaced Sequence Data. *Genetics* **2002**, *161*, 1307–1320. [CrossRef]
- Minin, V.N.; Bloomquist, E.W.; Suchard, M.A. Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. *Mol. Biol. Evol.* 2008, 25, 1459–1471. [CrossRef]

- Kosakovsky Pond, S.L.; Weaver, S.; Leigh Brown, A.J.; Wertheim, J.O. HIV-TRACE (TRAnsmission Cluster Engine): A Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol. Biol. Evol.* 2018, 35, 1812–1819. [CrossRef]
- 35. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526. [CrossRef]
- R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.R-project.org/ (accessed on 19 December 2022).
- Phillips, H.J. AHEAD: Monitoring Progress toward Achieving the Nation's Viral Suppression Goals. Available online: https:// www.hiv.gov/blog/ahead-monitoring-progress-toward-achieving-nation-s-viral-suppression-goals (accessed on 26 July 2022).
- 38. Centers for Disease Control and Prevention. Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data—United States and 6 Dependent Areas, 2019; HIV Surveillance Supplemental Report 2021, 26. Available online: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html (accessed on 21 December 2022).
- Quinn, T.C.; Wawer, M.J.; Sewankambo, N.; Serwadda, D.; Li, C.; Wabwire-Mangen, F.; Meehan, M.O.; Lutalo, T.; Gray, R.H. Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. N. Engl. J. Med. 2000, 342, 921–929. [CrossRef] [PubMed]
- Oster, A.M.; France, A.M.; Panneer, N.; Ocfemia, M.C.B.; Campbell, E.; Dasgupta, S.; Switzer, W.M.; Wertheim, J.O.; Hernandez, A.L. Identifying Clusters of Recent and Rapid HIV Transmission Through Analysis of Molecular Surveillance Data. *J. Acquir. Immune Defic. Syndr.* 2018, 79, 543–550. [CrossRef]
- Fisher, M.; Pao, D.; Brown, A.E.; Sudarshi, D.; Gill, O.N.; Cane, P.; Buckton, A.J.; Parry, J.V.; Johnson, A.M.; Sabin, C.; et al. Determinants of HIV-1 transmission in men who have sex with men: A combined clinical, epidemiological and phylogenetic approach. *AIDS* 2010, 24, 1739–1747. [CrossRef]
- Poon, A.F.Y.; Joy, J.B.; Woods, C.K.; Shurgold, S.; Colley, G.; Brumme, C.J.; Hogg, R.S.; Montaner, J.S.G.; Harrigan, P.R. The Impact of Clinical, Demographic and Risk Factors on Rates of HIV Transmission: A Population-based Phylogenetic Analysis in British Columbia, Canada. J. Infect. Dis. 2015, 211, 926–935. [CrossRef] [PubMed]
- Edelman, E.J.; Gordon, K.S.; Hogben, M.; Crystal, S.; Bryant, K.; Justice, A.C.; D. A. Fiellin for the VACS Project Team. Sexual Partner Notification of HIV Infection Among a National United States-Based Sample of HIV-Infected Men. *AIDS Behav.* 2014, 18, 1898–1903. [CrossRef]
- 44. McClelland, A.; Guta, A.; Gagnon, M. The rise of molecular HIV surveillance: Implications on consent and criminalization. *Crit. Public Health.* **2020**, *30*, 487–493. [CrossRef]
- 45. Dawson, L.; Benbow, N.; Fletcher, F.E.; Kassaye, S.; Killelea, A.; Latham, S.R.; Lee, L.M.; Leitner, T.; Little, S.J.; Mehta, S.R.; et al. Addressing Ethical Challenges in US-Based HIV Phylogenetic Research. *J. Infect. Dis.* **2020**, 222, 1997–2006. [CrossRef]
- 46. Mehta, S.R.; Schairer, C.; Little, S. Ethical issues in HIV phylogenetics and molecular epidemiology. *Curr. Opin. HIV AIDS.* 2019, 14, 221–226. [CrossRef]
- 47. Sandset, T. The ethical and epistemological pitfalls of translating phylogenetic HIV testing: From patient-centered care to surveillance. *Humanit Soc. Sci. Commun.* **2020**, *7*, 19. [CrossRef]
- 48. Ayala, G.; Bahati, M.; Balan, E.; Chang, J.; Do, T.D.; Fokeerbux, N.A.; Hassan, A.; Kerboghossian, J.; Poonkasetwatana, M.; Saavedra, J.; et al. Partner Notification: A Community Viewpoint. *J. Int. AIDS Soc.* **2019**, 22, e25291. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.