Longitudinal typing of molecular HIV clusters in a statewide epidemic

Vlad Novitsky[a], Jon Steingrimsson[a], Mark Howison[b], Casey Dunn[c], Fizza S. Gillani[a], Akarsh Manne[a], Yuanning Li[c], Matthew Spence[d], Zoanne Parillo[d], John Fulton[a], Theodore Marak[d], Philip Chan[a,d], Thomas Bertrand[d], Utpala Bandy[d], Nicole Alexander-Scott[d], Joseph Hogan[a] and Rami Kantor[a]

Background: HIV molecular epidemiology is increasingly integrated into public health prevention. We conducted cluster typing to enhance characterization of a densely sampled statewide epidemic towards informing public health.

Methods: We identified HIV clusters, categorized them into types, and evaluated their dynamics between 2004 and 2019 in Rhode Island. We grouped sequences by diagnosis year, assessed cluster changes between paired phylogenies, t_0 and t_1, representing adjacent years and categorized clusters as stable (cluster in t_0 phylogeny = cluster in t_1 phylogeny) or unstable (cluster in $t_0 \neq$ cluster in t_1). Unstable clusters were further categorized as emerging (t_1 phylogeny only) or growing (larger in t_1 phylogeny). We determined proportions of each cluster type, of individuals in each cluster type, and of newly diagnosed individuals in each cluster type, and assessed trends over time.

Results: A total of 1727 individuals with available HIV-1 subtype B *pol* sequences were diagnosed in Rhode Island by 2019. Over time, stable clusters and individuals in them dominated the epidemic, increasing over time, with reciprocally decreasing unstable clusters and individuals in them. Conversely, proportions of newly diagnosed individuals in unstable clusters significantly increased. Within unstable clusters, proportions of emerging clusters and of individuals in them declined; whereas proportions of newly diagnosed individuals in growing clusters significantly increased over time.

Conclusion: Distinct molecular cluster types were identified in the Rhode Island epidemic. Cluster dynamics demonstrated increasing stable and decreasing unstable clusters driven by growing, rather than emerging clusters, suggesting consistent in-state transmission networks. Cluster typing could inform public health beyond conventional approaches and direct interventions.

AIDS 2021, **35**:1711–1722

Keywords: emerging clusters, growing clusters, HIV cluster types, HIV transmission networks, molecular HIV clusters, stable clusters

[a]Brown University, [b]Research Improving People's Life, Providence, RI, [c]Yale University, New Haven, CT, and [d]Rhode Island Department of Health, Providence, RI, USA.

Correspondence to Vlad Novitsky, MD, PhD, Research Scientist, Division of Infectious Diseases, Department of Medicine, Brown University, Providence, RI, 02906, USA.

E-mail: VNovitsky@lifespan.org

Received: 5 November 2020; revised: 19 April 2021; accepted: 11 May 2021.

DOI:10.1097/QAD.0000000000002953

Introduction

Identification and characterization of molecular HIV clusters is increasingly used to understand structure and dynamics of transmission networks, identify transmission outbreaks in real-time, estimate transmission rates, project averted infections and inform epidemiology and prevention [1–12]. Utilizing HIV-1 *pol* sequences generated through routine drug resistance screening, outcomes of such studies help characterize HIV clustering and reveal dynamics of HIV transmission networks.

Transmission networks research has been informative in characterizing HIV epidemics. For example, early-stage HIV infections are more likely to be found in clusters [13], epidemiological and molecular surveillance could be synergistic [14], high density of sampling can improve cluster detection [15–18] and novel parameters, such as a network-based risk score, could help better characterize transmission networks [19]. Notably, analysis and prediction of growing clusters over time are of particular interest [20] as they could help identify high-priority groups, enhance prevention efforts, and enable prediction of increased transmission in a population [2,15,20–25]. Moreover, such clusters have a mean transmission rate eight times the national average in the United States [23] and their identification could provide further opportunity for public health interventions to interrupt new HIV transmissions [15,20,21,23–27].

Despite recent advances, it remains elusive whether and how HIV-1 clustering results could be translated into actionable public health interventions, beyond outbreak investigations, to prevent new HIV infections in real time. Molecular HIV clusters are typically characterized by cluster size or by traits of cluster members, such as risk factors, sociodemographics features, drug resistance, geography, and recency of infection. Common outcomes of cluster analyses [1,2,5,10], including HIV outbreaks [8,9], are usually binary, that is 'clustered' or 'not clustered', while potential qualitative and/or quantitative differences between clusters or their stability over time are not typically considered, particularly in real-life epidemics. Other than growing clusters, types of molecular HIV clusters are understudied, and their role in characterizing an epidemic remains unclear. Comprehensive longitudinal categorization of cluster types, or cluster typing, could increase the granularity of understanding transmission networks and viral spread mechanisms, and enhance the use of molecular HIV clusters to inform and impact public health activities and targeted deployment of resources.

According to the Rhode Island Department of Health [28], 2674 Rhode Islanders were diagnosed and living with HIV through the end of 2018; 91% were infected via sexual contact and 55% of newly diagnosed cases are gay, bisexual, or other MSM; and 89% of those with HIV know their status, 74% are engaged in care, and 69% are virally suppressed. In previous studies, we and others addressed patterns and trends of the Rhode Island HIV-1 epidemic including epidemiology [29–33], prevention among high-risk groups [34–44] and adolescents [45–47], co-infections [48–50], drug resistance [29,30,51–53] and transmission networks [21].

In this study, we conduct cluster typing to characterize a real-life, densely sampled statewide HIV-1 epidemic, assess trends of different cluster types over time and discuss the potential impact of cluster typing, as an extended tool of molecular epidemiology. We hypothesize that cluster typing could assist in better understanding the local epidemic and its dynamics, which could inform public health to design targeted interventions.

Methods

HIV-1 sequences and associated data

Analyses included all available HIV-1 sequences that were ever obtained for drug resistance testing as part of routine clinical care of individuals with HIV residing in Rhode Island and in care at the adult Immunology Center or pediatric Hasbro Clinic, the two largest HIV Centers in Rhode Island, which serve ~80% of the state's HIV population [21,51,54]. Demographic, clinical, and laboratory de-identified data were obtained from individuals' clinical charts, collected through routine clinical care, and included age, gender, race, ethnicity, behavior risk factors, and country of birth. Sequencing of partial HIV-1 *pol* was performed at certified commercial laboratories by Sanger sequencing. Additional sequence quality assessment was performed using SQUAT [55] and Stanford Database tools [56,57]. Multiple sequence alignment was generated by *mafft* v7.450 [58]. HIV-1 subtyping was performed by using REGA [59], COMET [60], and RIP [61,62] with minor discrepancies resolved on a case-by-case basis. HIV-1 subtype B *pol* sequences (HXB2 nucleotide positions 2253–3554; earliest single sequence per person) from individuals with a known year of HIV-1 diagnosis were included in the analyses. The study was approved by, and a consent waiver was obtained from, the Institutional Review Board at The Miriam Hospital, Providence, Rhode Island.

Identification of molecular HIV clusters

Viral sequences were grouped cumulatively by year of HIV-1 diagnosis. Molecular HIV clusters were identified in each annual sequence dataset by combining Maximum Likelihood phylogeny (RAxML v.8.10.2 [63,64] using GTRCAT model and fast bootstrap) with mean pairwise distance thresholds. To assess whether stringency of thresholds for inferring clusters affects cluster typing outcomes, we performed a sensitivity analysis by using the following four criteria, covering a broad range of

commonly used thresholds [65]: bootstrap support of at least 0.95 with mean pairwise TN93 distance threshold of 0.015 substitutions per site or less; bootstrap support of at least 0.90 with mean pairwise distance threshold of 0.030 substitutions/site or less; bootstrap support of at least 0.85 with mean pairwise distance threshold of 0.030 substitutions/site or less; and bootstrap support of at least 0.80 with mean pairwise distance threshold of 0.045 substitutions/site or less.

Cluster typing

We defined cluster types by tracking the evolution of molecular HIV clusters over time, between pairs of adjacent annual sequence datasets, and over time. After grouping sequences by year of HIV-1 diagnosis and identifying clusters in each annual dataset (by each of the four cluster definition criteria outlined above), we compared clusters in each of the 16 pairs of phylogenies, each pair representing adjacent years between 2003 and 2019. The earlier phylogeny in each pair was designated t_0 and the later phylogeny t_1. Changes in the composition and structure of clusters between each t_0 and t_1 phylogenies were identified (e.g. clusters identified in the 2004 phylogeny vs. clusters in the 2003 phylogeny; 2005 vs. 2004; 2006 vs. 2005, etc.). Differences between the t_0 and t_1 phylogenies in each pair resulted from newly diagnosed cases in the t_1 phylogeny (Fig. 1). Derived cluster types were defined as either stable, that is, the same cluster detected in both the t_0 and t_1 phylogenies; or unstable, that is, the cluster in the t_1 phylogeny was not the same as in the t_0 phylogeny. Unstable clusters were further categorized as emerging, that is, identified in the t_1 phylogeny only; growing, that is, increased in size in the t_1 phylogeny as compared with the t_0 phylogeny; merging, that is, multiple clusters detected in the t_0 phylogeny that merged in the t_1 phylogeny), growing-merging, lost, and reduced clusters (schematically shown in Fig. 1). Cluster types were summarized per each pair of phylogenies, t_0 and t_1. For each cluster type, three major outcome measures were assessed: proportion of each cluster type of the total identified clusters in the t_1 phylogeny; proportion of individuals in each cluster type of the total individuals in clusters in the t_1 phylogeny; and proportion of newly diagnosed individuals in each cluster type of the total newly diagnosed individuals with sequences in the t_1 phylogeny. As per definition, stable clusters did not include newly diagnosed individuals, proportions of newly diagnosed individuals in stable clusters were not assessed. Lastly, we determined trends over time for all outcome measures.

Inference about time trends

To quantify time-trends in cluster types, we calculated the Mann Kendall statistic. This statistic uses all pairs of distinct years and calculates the difference between the proportion of pairs with an increase in the cluster type (i.e. the proportion of the cluster type is higher in the more recent year) and the proportion of pairs with a decrease in the cluster type (i.e. the proportion of the cluster type is lower in the more recent year). A Mann Kendall statistic of 1 indicates an increase in the proportion of the cluster type for the whole time period and a Mann Kendall statistic of −1 indicates a decrease in the proportion of the clustering type for the whole time period. For confidence interval construction and hypothesis testing, the variance of the Mann Kendall statistic was estimated using a block bootstrap with block size of four and 10 000 bootstrap replicates [66]. The block bootstrap was used to account for potential autocorrelation. To ensure that the confidence intervals fall between −1 and 1, the confidence intervals are based on the $\log(-\log((x+1)/2))$ transformation. Confidence intervals cannot be calculated when the Mann Kendall statistic is equal to 1 or −1, and are therefore not reported. We used local polynomial regression to plot trends over time in the proportions of different cluster types and to calculate associated pointwise confidence intervals. Confidence intervals cannot be calculated when the observed proportion is equal to 0 or 1, and are therefore, not reported. To ensure that the confidence intervals for the proportions fall between 0 and 1 a $\log(-\log(x))$ transformation was used. To account for that different number of sequences were used to build the different phylogenies, the analysis was weighted by the inverse of the standard deviation of the proportion estimator. As we do not account for multiple comparisons, the P values reported here can be viewed as descriptive statistics quantifying degree of statistical association and inferences about time trends should be considered exploratory.

Results

HIV-1 diagnoses in Rhode Island and study dataset

According to the Rhode Island Department of Health, the annual number of new HIV-1 diagnoses in the state ranged from 24 in 1984, increased sharply during the late 1980s, peaked to 307 in 1990, and had a steep decline by 1995 with a more steady decline over the last 25 years to 73 in 2019 (Supplementary Figure S1; green line, http://links.lww.com/QAD/C150). The annual proportions of individuals with available sequences out of all individuals with HIV increased throughout the years from a median of 30% (IQR 19–40%) during 1984–2003 to 73% (IQR 69–81%) during 2004–2019 because of increased routine pre-antiretroviral therapy (ART) resistance testing. To minimize sampling bias, all cluster typing analyses in this study were, therefore, performed for the 2003/2004–2018/2019 data sets (Supplementary Figure S1; to the right of the vertical dashed line, http://links.lww.com/QAD/C150), during which the gap between newly HIV-diagnosed individuals and those with no sequence data was smaller and relatively stable.

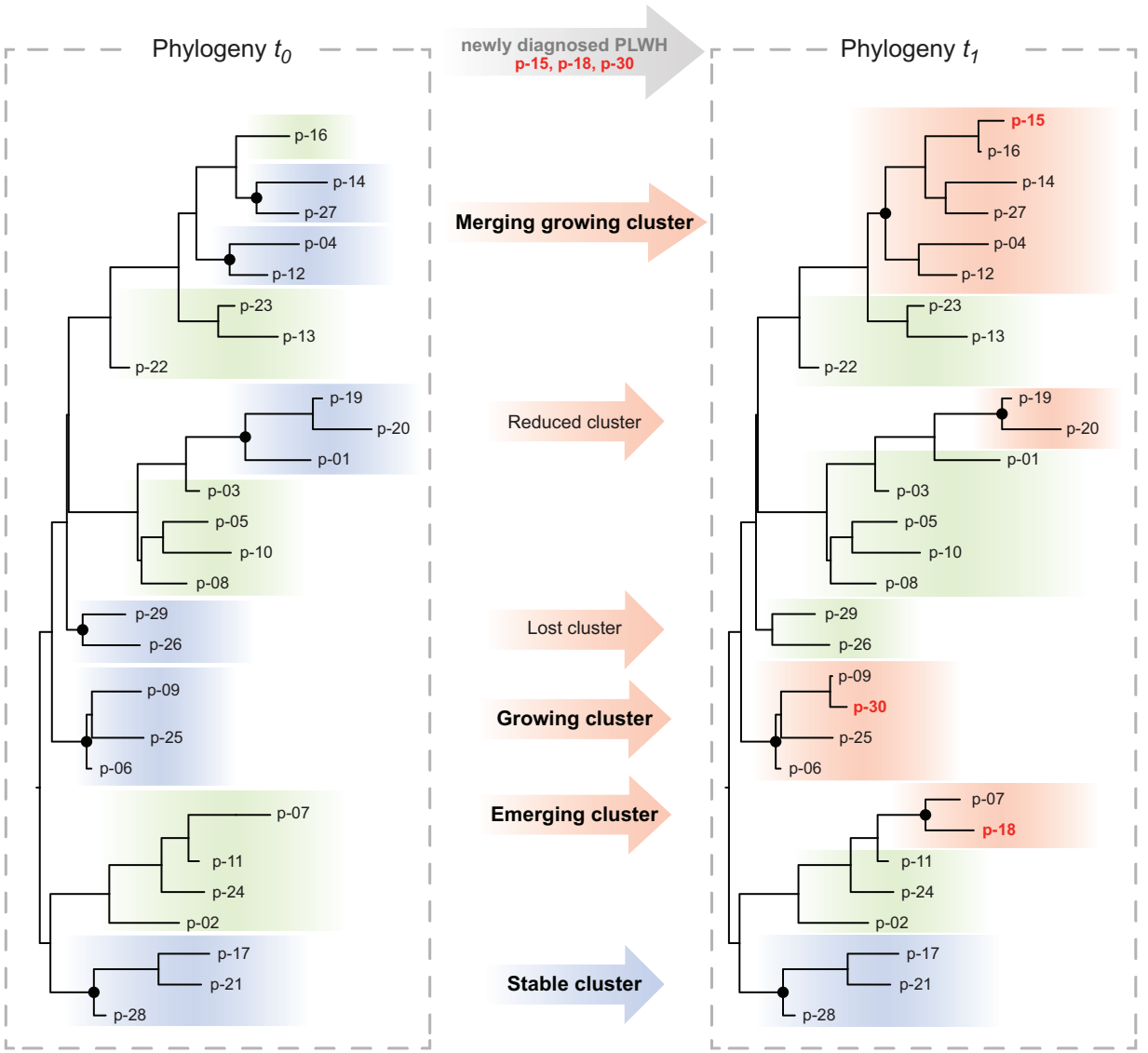


Fig. 1. Cluster typing concept. The cluster typing concept and the relevant clusters that are being considered in analyses are presented. Two schematic phylogenies, at t_0 (left) and at t_1 (right) are shown, representing two annual sequence datasets of any adjacent years (e.g. 2003 and 2004, 2004 and 2005, etc.). The t_1 phylogeny includes the same sequences as the t_0 phylogeny, plus three cases of newly diagnosed individuals in t_1 (p-15, p-18, and p-30, shown in red). Identified clusters (as defined in Methods) are outlined by black circles at the relevant ancestral nodes, also highlighted by blue (stable clusters) or red (unstable clusters) clouds in the t_1 phylogeny. Nodes without circles represent nonclustered sequences according to the cluster definition criteria, and are outlined by green clouds. Different cluster types resulting from changes between the t_0 and t_1 phylogenies are shown by arrows between the two trees.

Of a total of 2149 available sequences by the end of 2019 (single earliest sequence per person), 90% ($n = 1936$) were HIV-1 subtype B and 10% ($n = 213$) non-B subtypes (Supplementary Figure S1; gray and pink lines, http://links.lww.com/QAD/C150). As no clustering is expected between HIV-1 subtypes, and the non-B sequence data were relatively few, only HIV-1 subtype B sequences with documented year of HIV-1 diagnosis were included in this study (693 sequences from people diagnosed by 2003 and 1034 sequences from people diagnosed between 2004 and 2019). Among the 1727 individuals with HIV-1 subtype B, median age at diagnosis was 33 years (IQR 26–42 years), 75% were men (68% of those MSM), 29% were Hispanics, 27% black or African-American, and 66% white, 21% used illegal substances or injected nonprescription drugs, 10% had a history of being incarcerated, 43% had mental illnesses, and 71% were born in the USA.

Molecular HIV clusters in Rhode Island

The annual numbers of identified clusters, proportion of individuals in clusters out of total individuals, and

proportion of newly diagnosed individuals in clusters out of total newly diagnosed individuals had all significantly increased from 2004 to 2019 by all four cluster definition criteria, though with occasional plateauing and fluctuations (Supplementary Figure S2, http://links.lww.com/QAD/C151). Depending on the criteria, of 693 sequences available by 2003, 4–15% were in 15–43 clusters of size 2–9; and of 1727 sequences available by 2019, 26–45% were in 147–195 clusters of size 2–41 (Supplementary Figure S2, http://links.lww.com/QAD/C151).

Cluster typing: stable clusters

Cluster types were identified in 16 pairs of adjacent year phylogenies between 2003/2004 and 2018/2019. Stable clusters between adjacent year pairs dominated in the HIV-1 Rhode Island epidemic. Proportions of stable clusters (out of the total identified clusters) demonstrated a statistically significant increase over time, irrespective of cluster definition criteria (Fig. 2a; Mann Kendall statistic 0.79–0.87; *P* values 0.003–0.005), ranging from 54 to 63% in 2003/2004 to 91 to 92% in 2018/2019.

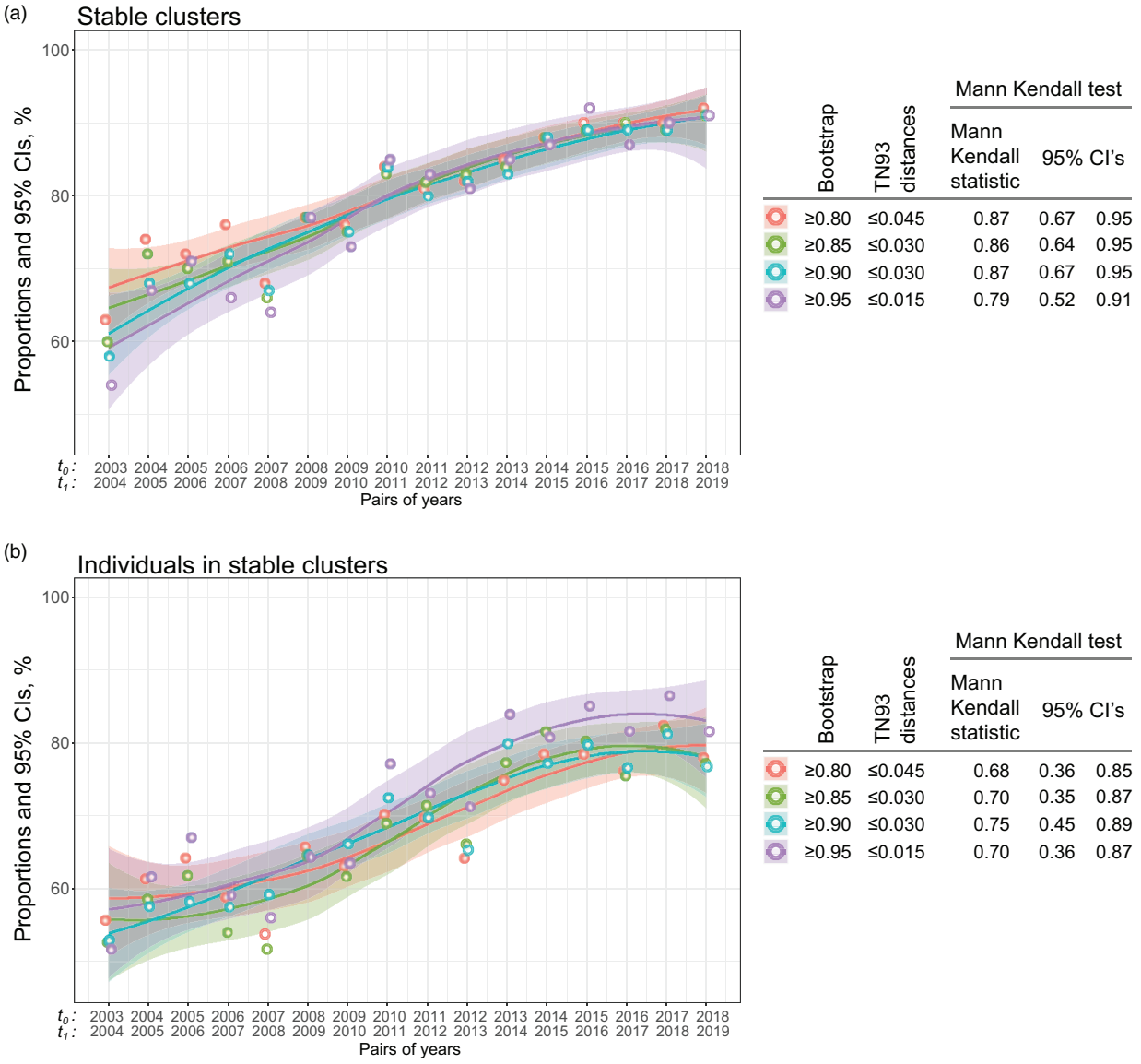


Fig. 2. Cluster typing: trends in proportion of stable clusters over time. Proportions (*y*-axis) and 95% CIs (shaded areas) of stable clusters in phylogenies of pairs of adjacent years (t_0 and t_1) are shown from 2003/2004 to 2018/2019 (*x*-axis). The four cluster definition criteria (colors, bootstrap supports and mean pairwise distances) used for sensitivity analyses are outlined in panels to the right of the graphs. Estimated proportions for each pair of years are shown as dots in the graphs, while estimated Mann Kendall statistic and 95% CIs are shown as lines and shades in the graphs; and are presented in panels to the right of the graphs. (a) Proportions and 95% CIs of stable clusters out of total clusters. (b) Proportions and 95% CIs of individuals in stable clusters out of total individuals with sequences. CI, confidence interval.

Similar trends were observed for the proportion of individuals in stable clusters (out of the number of all individuals in clusters; Fig. 2b; Mann Kendall statistic 0.68–0.75; *P* values 0.008–0.012). Proportions increased by more than 20% at all cluster definition criteria, ranging from 52 to 56% in 2003/2004 to 78 to 82% in 2018/2019.

Cluster typing: unstable clusters

Given that the proportion of stable clusters increased through time, the proportion of unstable clusters (out of the total identified clusters; reciprocal to the proportion of stable clusters) showed a statistically significant decrease over time across all analyzed cluster definition criteria (Fig. 3a; Mann Kendall statistic from −0.87 to −0.79; *P* values 0.003–0.006). Between 2003/2004 and 2018/2019, the range of proportions of unstable clusters decreased from 37–46% to 8–9%.

Trends in proportions of individuals in unstable clusters (out of the total number of clustered individuals; reciprocal to the proportion of individuals in stable clusters) similarly decreased over time (Fig. 3b; Mann Kendall statistic from −0.75 to −0.68; *P* values 0.008–0.014). Between 2003/2004 and 2018/2019, the range of proportions of individuals in unstable clusters decreased from 44–48% to 18–22%.

In contrast to these declining trends, the proportions of newly diagnosed individuals in unstable clusters (out of the total number of newly diagnosed individuals with sequences) demonstrated significant increases at all analyzed cluster definition criteria (Fig. 3c; Mann Kendall statistic 0.52–0.58; *P* values 0.005–0.008). Three of four criteria produced similar proportions, increasing from 29–33% in 2003/2004 to 75–81% in 2018/2019, respectively. The most strict cluster definition criteria produced a curve that was similar in shape (Mann Kendall statistic 0.52; *P* value 0.006) but had lower values and increased from 20% in 2003/2004 to 58% in 2018/2019.

Cluster typing: emerging clusters

To better understand the composition of the unstable clusters, we further categorized them to different cluster types. Between 2003/2004 and 2018/2019, the proportions of emerging clusters (out of the total number of clusters) decreased from 27–46% to 2–4%, demonstrating a significant decline (Fig. 4a; Mann Kendall statistic −0.80 to −0.75; *P* values 0.006–0.010).

The proportion of individuals in emerging clusters (out of all individuals in clusters) showed similar trends and decreased from 23–48 to 2–3% (Fig. 4b; Mann Kendall statistic −0.77 to −0.75; *P* values 0.006–0.011). However, the proportions of newly diagnosed individuals in emerging clusters (out of the total number of newly diagnosed individuals with sequences) demonstrated fluctuating, wave-like patterns with overall stability (∼20%) and nonsignificant trends over time (Fig. 4c; Mann Kendall statistic −0.28 to −0.03; *P* values 0.073–0.855).

Cluster typing: growing clusters

Growing clusters, another category of unstable clusters, demonstrated slight increase in their proportions out of the number of identified clusters between 2003/2004 and 2008/2009 followed by slow gradual decline over the last decade (Fig. 5a; Mann Kendall statistic −0.53 to −0.21; *P* values 0.053–0.332). Proportions of growing clusters increased from 0–14% in 2003/2004 to 12–15% in 2009/2010 followed by decline to 5% by 2018/2019.

The proportions of individuals in growing clusters (out of the total number of clustered individuals) demonstrated similarly shaped and nonsignificant trends (Fig. 5b; Mann Kendall statistic −0.40 to −0.05; *P* values 0.114–0.805). Proportions increased from 0–21% in 2003/2004 to 25–31% in 2007–2009 and then dropped to 15–20% by 2019.

In contrast, the proportions of newly diagnosed individuals in growing clusters (out of the total number of newly diagnosed individuals with sequences) showed steady increases between 2003/2004 and 2018/2019 (Fig. 5c; Mann Kendall statistic 0.68–0.77; *P* values 0.003–0.006). Despite some year-to-year fluctuations, the proportion of newly diagnosed individuals in growing clusters increased from 0–10 to 33–55% at the different cluster definition criteria.

We did not see statistical evidence to support existence of any other unstable cluster types, such as, merging, growing-merging, lost, and reduced clusters.

Discussion

We identified distinct types of molecular HIV-1 clusters and used them to longitudinally characterize a real life, densely sampled statewide HIV epidemic during 2004–2019. The steady decline of newly diagnosed individuals in Rhode Island during that time was accompanied by increasing number of identified clusters and proportions of individuals in clusters and of newly diagnosed individuals in clusters. The proportion of stable clusters of all clusters increased over time and dominated the epidemic, suggesting good overall control. In contrast, proportion of newly diagnosed individuals in growing, rather than other types of unstable clusters steadily increased over the same time period. Cluster typing could, therefore, help to more comprehensively characterize transmission network dynamics in a statewide epidemic and provide more granular understanding of viral spread mechanisms by categorization of clusters and individuals in them. These data can be incorporated in and guide routine public health activities and be

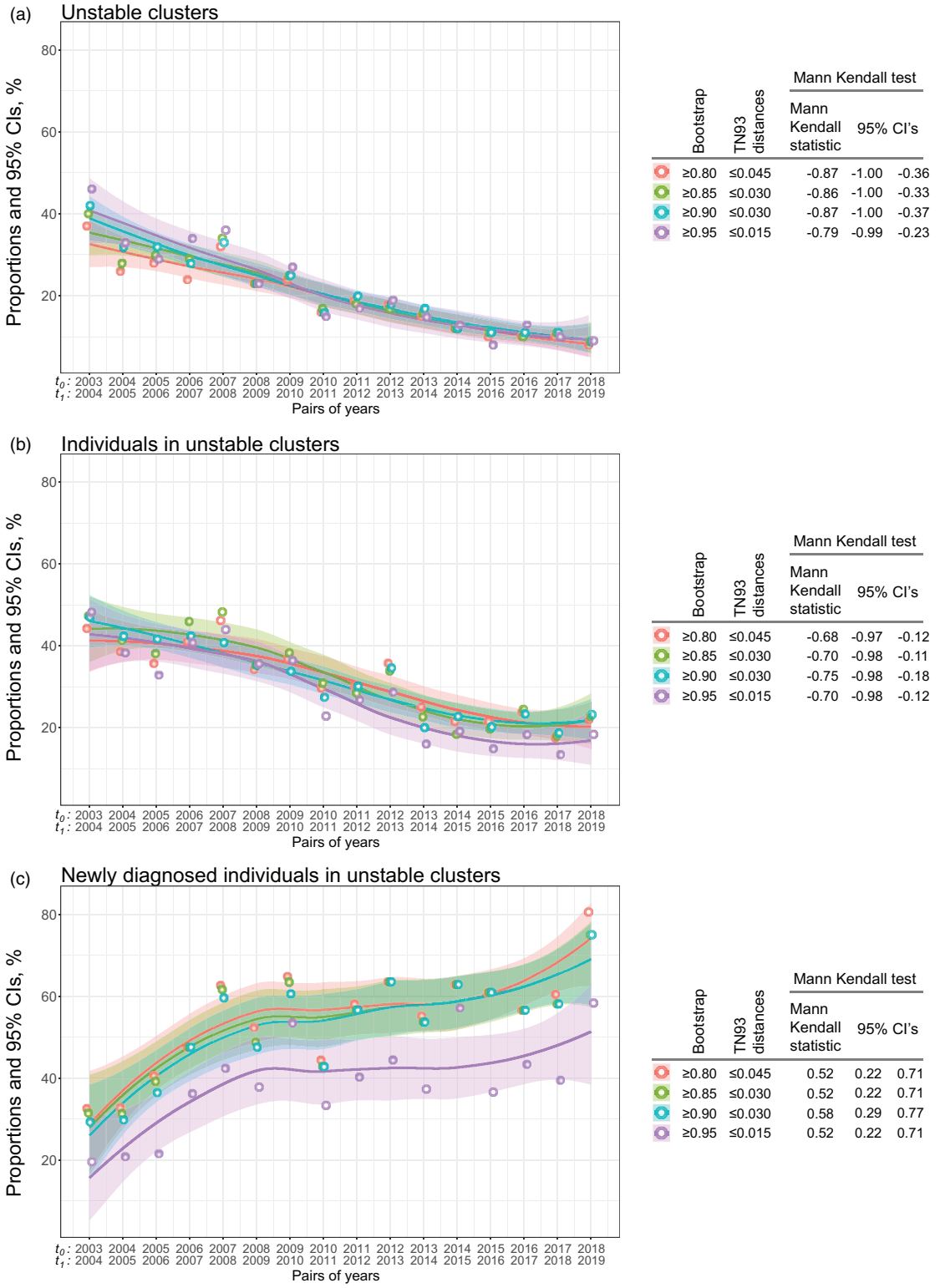
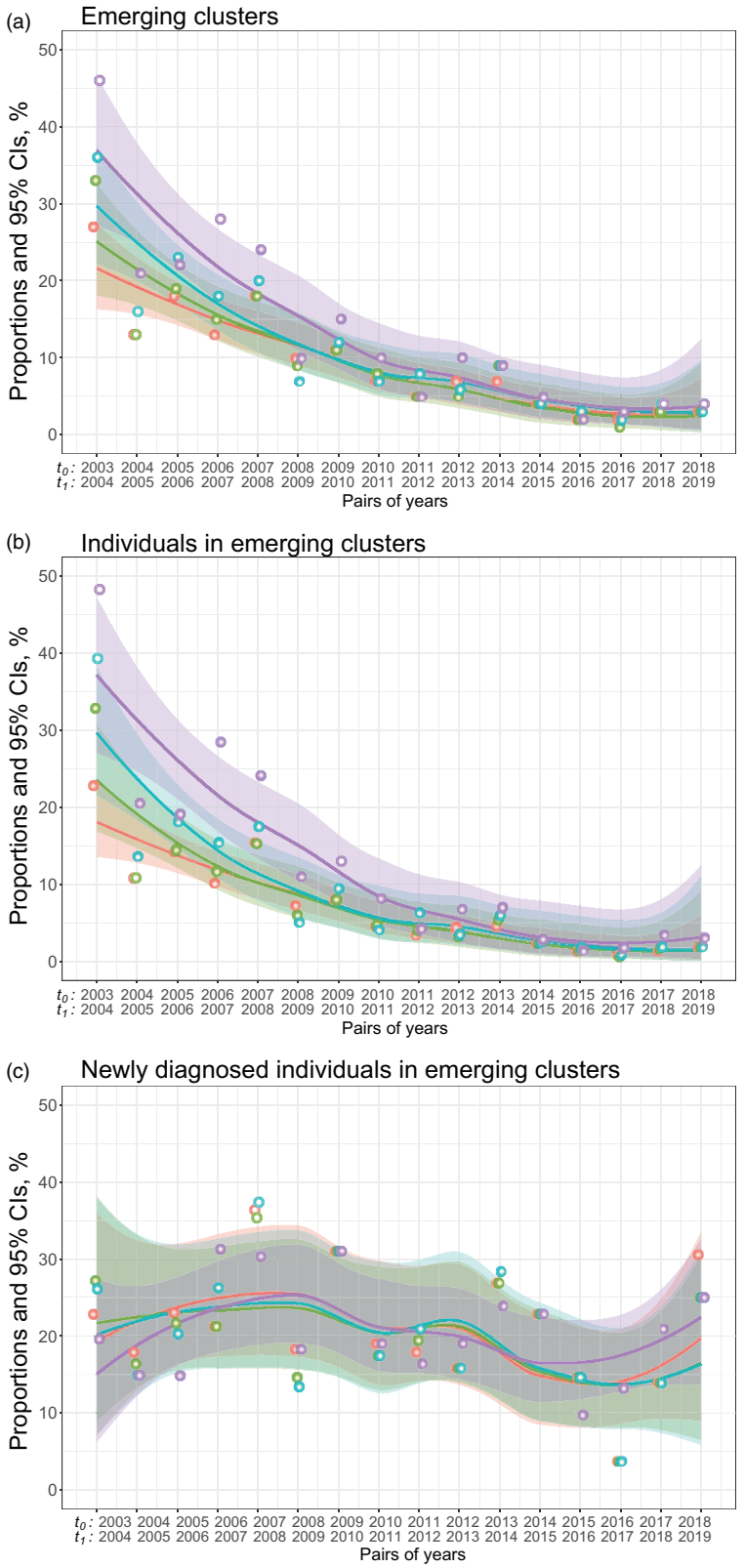


Fig. 3. Cluster typing: trends in proportion of unstable clusters over time. For details see Fig. 2 legend. (a) Proportions and 95% CIs of unstable clusters out of total clusters. (b) Proportions and 95% CIs of individuals in unstable clusters out of total individuals with sequences. (c) Proportions and 95% CIs of newly diagnosed individuals in unstable clusters out of total number of newly diagnosed individuals with sequences. CI, confidence interval.



Bootstrap	TN93 distances	Mann Kendall test: Mann Kendall statistic	95% CI's	
≥0.80	≤0.045	-0.80	-0.99	-0.21
≥0.85	≤0.030	-0.79	-0.99	-0.22
≥0.90	≤0.030	-0.75	-0.98	-0.20
≥0.95	≤0.015	-0.76	-0.99	-0.16

Bootstrap	TN93 distances	Mann Kendall test: Mann Kendall statistic	95% CI's	
≥0.80	≤0.045	-0.77	-0.99	-0.18
≥0.85	≤0.030	-0.77	-0.99	-0.18
≥0.90	≤0.030	-0.77	-0.99	-0.22
≥0.95	≤0.015	-0.75	-0.99	-0.15

Bootstrap	TN93 distances	Mann Kendall test: Mann Kendall statistic	95% CI's	
≥0.80	≤0.045	-0.26	-0.61	0.09
≥0.85	≤0.030	-0.28	-0.59	0.04
≥0.90	≤0.030	-0.20	-0.53	0.12
≥0.95	≤0.015	-0.03	-0.31	0.23

Fig. 4. Cluster typing: trends in proportion of emerging clusters over time. For details see Fig. 2 legend. (a) Proportions and 95% CIs of emerging clusters out of total clusters. (b) Proportions and 95% CIs of individuals in emerging clusters out of total individuals with sequences. (c) Proportions and 95% CIs of newly diagnosed individuals in emerging clusters out of total number of newly diagnosed individuals with sequences. CI, confidence interval.

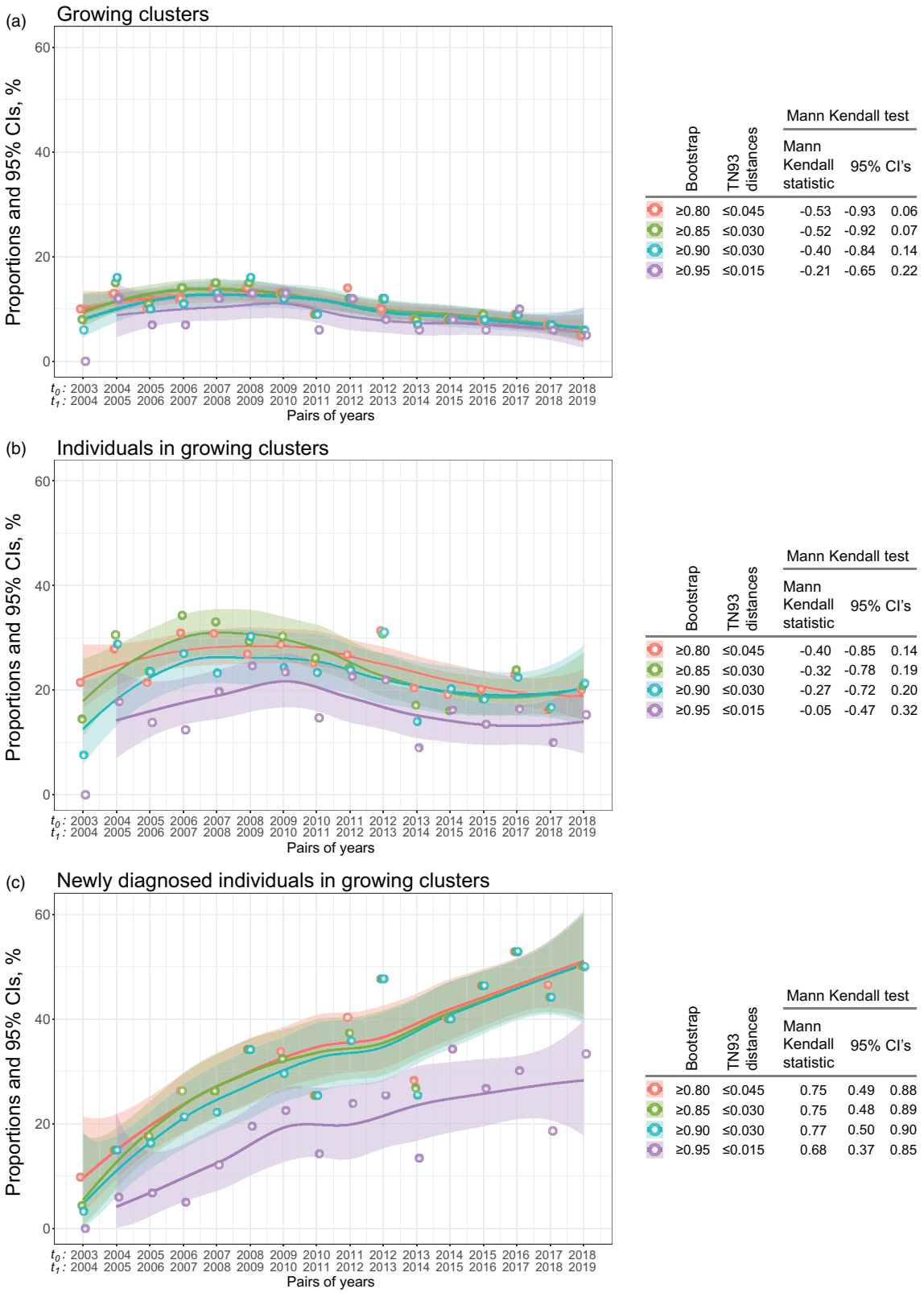


Fig. 5. Cluster typing: Trends in proportion of growing clusters over time. For details see Fig. 2 legend. When proportions were equal to zero, 95% CIs cannot be calculated. (a) Proportions and 95% CIs of growing clusters out of total clusters. (b) Proportions and 95% CIs of individuals in growing clusters out of total individuals with sequences. (c) Proportions and 95% CIs of newly diagnosed individuals in growing clusters out of total number of newly diagnosed individuals with sequences. CI, confidence interval.

integrated into partner services to prevent HIV transmissions.

Though the significance and benefit of molecular cluster typing needs to be demonstrated as well as validated in other populations, we demonstrate that it might identify new and important epidemiological dynamics. HIV-1 molecular clusters are not all equal, yet are commonly treated as such, partly because of cross-sectional rather than longitudinal analyses. Some clusters are stable over time, which might indicate low likelihood of future activity, represent an extinct viral lineage or transmission chain, and therefore, require less public health attention. Other clusters are changing over time, which might require better consideration for intervention. For example, detection of an unstable cluster might trigger targeted contact tracing services towards cluster members or their contacts, who might be out of care, not on ART, not virologically suppressed, undiagnosed, or at high risk for HIV acquisition.

The dominance and extent of stable clusters detected in Rhode Island throughout the observed time period, together with the persistent increase in overall proportions of individuals in stable clusters and decreasing proportions of unstable clusters might represent a favorable trend, suggesting saturation and control of a local epidemic. Statewide transitions from unstable to stable clusters over time, enabled by longitudinal sequence aggregation and cluster typing, could be, therefore, considered an ultimate goal in HIV epidemic control, providing insight and guiding public health interventions.

On the other hand, further categorization of unstable clusters in the Rhode Island HIV epidemic demonstrated trends that can be interpreted as concerning, because of stability of newly diagnosed individuals in emerging clusters, despite decreasing overall emerging clusters and individuals in these clusters. Even more disturbing is a significant increase in newly diagnosed individuals in growing clusters despite little changes in overall growing clusters and individuals in these clusters over time. These trends, that there are more newly diagnosed individuals in unstable clusters with time, driven by growing rather than emerging clusters, may indicate more persistent, rather than short-lived, growth over time, as well as consistent in-state HIV transmission chains, requiring better focus. Such observations create opportunities for stronger local efforts and interventions, such as more rigorous and upscaled HIV diagnostics and testing and enhanced partner services to identify HIV-infected-unaware individuals, intensified preexposure prophylaxis for high-risk individuals, and boosted community engagement.

For optimized results, the data required for the proposed cluster typing would be obtained, analyzed and disseminated state-wide, with potential for cross state collaborations. These data could help prioritize clusters and facilitate focused partner services in a timely manner to target the interventions directly to prioritized clusters, their members, and their partners. Such interventions could be particularly helpful for large jurisdictions and/or those with limited partner services capacity, within or across state-lines. Incorporating cluster typing into longitudinal analyses could provide information beyond conventional molecular epidemiological approaches, including investigations of transmission outbreaks. The scope and magnitude of these speculations and the benefit of cluster typing for public health still needs to be demonstrated, and studies are needed to evaluate its generalizability and applicability to statewide HIV epidemics and beyond.

The performed sensitivity analysis by using a broad range of cluster definition criteria demonstrated robustness and validity of the cluster typing approach. Almost all trends and proportions identified during cluster typing in this study showed high similarities across the range of criteria. In rare cases, when discordance was identified, shapes of analyzed trend curves were still similar irrespective of the criteria. Analytical approaches in molecular epidemiology (e.g. software tools and thresholds) as well as scenarios in which these approaches can and should be used (e.g. outbreak investigations vs. routine use to guide partner services) are heterogenous with limited justification of their specific uses. There is no consensus on optimal analytical approaches, and many phylogenetic and distance cluster definition criteria are being used to determine and characterize HIV molecular clusters [7–17,65]. Such approaches might need to be further refined towards specific purposes, such as the use of more strict cluster definition criteria in transmission outbreak investigations vs. the use of more relaxed criteria in historical analyses and routine public health activities. Although the relative stability of longitudinal cluster typing based on cluster definition criteria demonstrated here is reassuring, cluster typing will require exploration in other HIV epidemics and populations.

A limitation of this study, beyond inevitable incompleteness of sequence data because of individual migration, commuting, and transience, is incomplete sampling during early stages of the HIV-1 epidemic in Rhode Island, which could negatively affect cluster typing analyses. To minimize sampling bias, we restricted the analysis to 2004–2019, a time period in which more than 70% of diagnosed individuals in Rhode Island had available viral sequences. On the other hand, our dataset is highly representative of the Rhode Island epidemic as we had access to ~80% of individuals with HIV in the state. Though possibly limiting generalizability, this further emphasizes potential implications of cluster typing in larger jurisdictions and the need for its evaluation in other settings. Additionally, making social assumptions based on phylogenetic analysis is inferential and likely inaccurate, and is an inherent limitation of molecular epidemiology approaches. Lastly, we categorized datasets based on

timing of HIV diagnoses rather than HIV infections, which would have been more accurate, but is information that is conventionally hard to obtain.

In conclusion, though the existing literature at times considers growing clusters in HIV epidemic descriptions, we propose that cluster typing can and should be a potential addition to cluster analyses (e.g. those promoted by CDC [8,9]) and evaluations of local HIV-1 epidemics. This approach, when applied to the densely sampled Rhode Island epidemic, demonstrated intervention potential, beyond an acute outbreak mediation. Robustness and benefits of this approach for public health need to be determined and demonstrated in these and other settings. However, such a panoramic longitudinal perspective, beyond individual and cross-sectional cluster detection and routine contact tracing, could advance understanding of HIV-1 epidemics, and lead to more precise public health interventions.

Acknowledgements

Funding: National Institute of Allergy and Infectious Diseases: R01AI136058; K24AI134359; P30AI42853 (Providence/Boston Center for AIDS Research).

Partial data were presented previously at the 27th International Dynamics & Evolution of HIV and Other Human Viruses Meeting (virtual), 10–13 May 2020.

Authors' contributions: V.N. and R.K. conceived the study. V.N., J.S., and R.K. designed the study. V.N., M.H., F.G., A.M., and R.K. collected data. V.N., J.S,. M.H., C.W.D., J.H., and R.K. analyzed data. V.N. and R.K. wrote the first version of the manuscript; V.N., J.S., M.H., F.G., Y.L., J.F., Z.P., M.S., T.M., P.C., T.B., U.B., N.A.-S., C.W.D., J.H., and R.K. critically reviewed and finalized the manuscript.

Conflicts of interest

M.H. reports fees from Competition Economics and The Miriam Hospital for consulting, outside the submitted work. All other authors declare that they have no competing interests.

References

1. Grabowski MK, Herbeck JT, Poon AFY. **Genetic cluster analysis for HIV prevention.** *Curr HIV/AIDS Rep* 2018; **15**:182–189.
2. Wertheim JO, Chato C, Poon AFY. **Comparative analysis of HIV sequences in real time for public health.** *Curr Opin HIV AIDS* 2019; **14**:213–220.
3. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjornsson J. **Defining HIV-1 transmission clusters based on sequence data: a systematic review and perspectives.** *AIDS* 2017; **31**:1211–1222.
4. Dennis AM, Herbeck JT, Brown AL, Kellam P, de Oliveira T, Pillay D, *et al*. **Phylogenetic studies of transmission dynamics in generalized HIV epidemics: an essential tool where the burden is greatest?** *J Acquir Immune Defic Syndr* 2014; **67**:181–195.
5. Grabowski MK, Redd AD. **Molecular tools for studying HIV transmission in sexual networks.** *Curr Opin HIV AIDS* 2014; **9**:126–133.
6. Poon AF. **Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks.** *Virus Evol* 2016; **2**:vew031.
7. Poon AF, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, *et al*. **Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study.** *Lancet HIV* 2016; **3**:e231–e238.
8. CDC. Detecting and responding to HIV transmission clusters. A guide for health departments. 2018. Available at: https://www.cdc.gov/hiv/pdf/funding/announcements/ps18-1802/CDC-HIV-PS18-1802-AttachmentE-Detecting-Investigating-and-Responding-to-HIV-Transmission-Clusters.pdf [Accessed 31 May 2020]
9. CDC. HIV cluster and outbreak detection and response. 2018. Available at: https://www.cdc.gov/hiv/programresources/guidance/molecular-cluster-identification/index.html [Accessed 31 May 2020]
10. Gonsalves GS, Crawford FW. **Dynamics of the HIV outbreak and response in Scott County, IN, USA, 2011-15: a modelling study.** *Lancet HIV* 2018; **5**:e569–e577.
11. Vrancken B, Rambaut A, Suchard MA, Drummond A, Baele G, Derdelinckx I, *et al*. **The genealogical population dynamics of HIV-1 in a large transmission chain: bridging within and among host evolutionary rates.** *PLoS Comput Biol* 2014; **10**:e1003505.
12. Hall HI, Holtgrave DR, Maulsby C. **HIV transmission rates from persons living with HIV who are aware and unaware of their infection.** *AIDS* 2012; **26**:893–896.
13. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SD. **Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection.** *PLoS Comput Biol* 2012; **8**:e1002552.
14. Mehta SR, Murrell B, Anderson CM, Kosakovsky Pond SL, Wertheim JO, Young JA, *et al*. **Using HIV sequence and epidemiologic data to assess the effect of self-referral testing for acute HIV infection on incident diagnoses in San Diego, California.** *Clin Infect Dis* 2016; **63**:101–107.
15. Dasgupta S, France AM, Brandt MG, Reuer J, Zhang T, Panneer N, *et al*. **Estimating effects of HIV sequencing data completeness on transmission network patterns and detection of growing HIV transmission clusters.** *AIDS Res Hum Retroviruses* 2019; **35**:368–375.
16. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. **Impact of sampling density on the extent of HIV clustering.** *AIDS Res Hum Retroviruses* 2014; **30**:1226–1235.
17. Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJ, PANGEA_HIV Consortium, ICONIC Project. **Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic.** *Sci Rep* 2016; **6**:39489.
18. Chaillon A, Delaugerre C, Brenner B, Armero A, Capitant C, Nere ML, *et al*. **In-depth sampling of high-risk populations to characterize HIV transmission epidemics among young MSM using PrEP in France and Quebec.** *Open Forum Infect Dis* 2019; **6**:ofz080.
19. Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, *et al*. **Using HIV networks to inform real time prevention interventions.** *PLoS One* 2014; **9**:e98443.
20. Wertheim JO, Murrell B, Mehta SR, Forgione LA, Kosakovsky Pond SL, Smith DM, Torian LV. **Growth of HIV-1 molecular transmission clusters in New York City.** *J Infect Dis* 2018; **218**:1943–1953.
21. Chan PA, Hogan JW, Huang A, DeLong A, Salemi M, Mayer KH, Kantor R. **Phylogenetic investigation of a statewide HIV-1 epidemic reveals ongoing and active transmission networks among men who have sex with men.** *J Acquir Immune Defic Syndr* 2015; **70**:428–435.
22. Avila D, Keiser O, Egger M, Kouyos R, Boni J, Yerly S, *et al*., Swiss HIV Cohort Study. **Social meets molecular: combining phylogenetic and latent class analyses to understand HIV-1 transmission in Switzerland.** *Am J Epidemiol* 2014; **179**:1514–1525.

23. Oster AM, France AM, Panneer N, Banez Ocfemia MC, Campbell E, Dasgupta S, *et al*. **Identifying Clusters of Recent and Rapid HIV Transmission Through Analysis of Molecular Surveillance Data.** *J Acquir Immune Defic Syndr* 2018; **79**:543–550.
24. Wertheim JO, Panneer N, France AM, Saduvala N, Oster AM. **Incident infection in high-priority HIV molecular transmission clusters in the United States.** *AIDS* 2020; **34**:1187–1193.
25. Billock RM, Powers KA, Pasquale DK, Samoff E, Mobley VL, Miller WC, *et al*. **Prediction of HIV transmission cluster growth with statewide surveillance data.** *J Acquir Immune Defic Syndr* 2019; **80**:152–159.
26. Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. **Ending the HIV epidemic: a plan for the United States.** *JAMA* 2019; **321**:844–845.
27. Wertheim JO, Oster AM, Switzer WM, Zhang C, Panneer N, Campbell E, *et al*. **Natural selection favoring more transmissible HIV detected in United States molecular transmission network.** *Nature Commun* 2019; **10**:5788.
28. Rhode Island Department of Health. RHODE ISLAND HIV, Sexually Transmitted Diseases, Viral Hepatitis, and Tuberculosis Surveillance Report. Available at: https://health.ri.gov/publications/surveillance/2019/HIVSTD.pdf [Accessed 12 April 2021]
29. Alexander NE, Chan PA, Rogo TO, Parillo Z, Browning CA, Isabell BS, *et al*. **Interrupting transmission of HIV and other sexually transmitted infections in Rhode Island.** *Med Health R I* 2012; **95**:241–244.
30. Chan PA, Montgomery MC, Rose J, Tributino A, Crowley C, Medina MM, *et al*. **Statewide evaluation of new HIV diagnoses in rhode island: implications for prevention.** *Public Health Rep* 2018; **133**:489–496.
31. Flanigan TP, Cu-Uvin S, Fiore T, Vigilante K, Kizirian J, Carpenter CC. **HIV infection in Rhode Island women.** *R I Med* 1993; **76**:459–460465-456.
32. Keenlyside RA, Dondero L, Corkery M, Johnson A. **The epidemiology and control of AIDS and HIV infection in Rhode Island.** *R I Med J* 1987; **70**:35–42.
33. Leeper S, Fillion K, Gillani FS, Rana A, Tashima K, Ross H. **Decade of HIV in Rhode Island: demographic and clinical characteristics of patients diagnosed in 2001 and 2010.** *R I Med J (2013)* 2013; **96**:35–40.
34. Beckwith CG, Bazerman L, Cornwall AH, Patry E, Poshkus M, Fu J, Nunn A. **An evaluation of a routine opt-out rapid HIV testing program in a Rhode Island jail.** *AIDS Educ Prev* 2011; **23 (3 Suppl)**:96–109.
35. Beckwith CG, Cohen J, Shannon C, Raz L, Rich JD, Lally MA. **HIV testing experiences among male and female inmates in Rhode Island.** *AIDS Read* 2007; **17**:459–464.
36. Beckwith CG, Moreira CC, Aboshady HM, Zaller N, Rich JD, Flanigan TP. **A success story: HIV prevention for injection drug users in Rhode Island.** *Subst Abuse Treat Prev Policy* 2006; **1**:34.
37. Mitty JA, Bazerman LB, Selwyn K, Beckwith CG, Zaller ND, Rich JD. **Decrease in the proportion of injecting drug use-related HIV/AIDS in Massachusetts, New York, Connecticut, and Rhode Island.** *AIDS Read* 2008; **18**:596–600.
38. Desai AA, Latta ET, Spaulding A, Rich JD, Flanigan TP. **The importance of routine HIV testing in the incarcerated population: the Rhode Island experience.** *AIDS Educ Prev* 2002; **14 (5 Suppl B)**:45–52.
39. Farley JL, Mitty JA, Lally MA, Burzynski JN, Tashima K, Rich JD, *et al*. **Comprehensive medical care among HIV-positive incarcerated women: the Rhode Island experience.** *J Womens Health Gend Based Med* 2000; **9**:51–56.
40. Rich JD, Dickinson BP, Liu KL, Case P, Jesdale B, Ingegneri RM, *et al*. **Strict syringe laws in Rhode Island are associated with high rates of reusing syringes and HIV risks among injection drug users.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1998; **18 Suppl 1**:S140–S141.
41. Rich JD, Dickinson BP, Macalino G, Flanigan TP, Towe CW, Spaulding A, Vlahov D. **Prevalence and incidence of HIV among incarcerated and reincarcerated women in Rhode Island.** *J Acquir Immune Defic Syndr* 1999; **22**:161–166.
42. Rich JD, Macalino G, Merchant RC, Salas C, Marcussen P, Grundy M, Spaulding A. **HIV seroprevalence of adult males incarcerated for a sexual offense in Rhode Island, 1994-1999.** *JAMA* 2002; **288**:164–165.
43. Wakeman SE, Zaller ND, Flanigan TP, Pinkston M, Montague BT, Rich JD. **HIV among marginalized populations in Rhode Island.** *Med Health R I* 2009; **92**:244–246.
44. Chan PA, Towey C, Poceta J, Rose J, Bertrand T, Kantor R, *et al*. **Online hookup sites for meeting sexual partners among men who have sex with men in Rhode Island, 2013: a call for public health action.** *Public Health Rep* 2016; **131**:264–271.
45. Pugatch D, Anderson BJ, O'Connell JV, Elson LC, Stein MD. **HIV and HCV testing for young drug users in Rhode Island.** *J Adolesc Health* 2006; **38**:302–304.
46. Pugatch D, Rich JD, Flanigan TP. **HIV infection in Rhode Island adolescents.** *Med Health R I* 1997; **80**:331–333.
47. Ryoo HJ, Nazareth K, Chan PA, Reinert SE, Koster M. **HIV testing for at-risk adolescents at Rhode Island Hospital.** *R I Med J* 2013; **98**:25–28.
48. Gampper SN, George JA, Carter EJ, Jesdale BM, Flanigan TP, Mayer KH, De Groot AS. **Co-infection with Mycobacterium tuberculosis and HIV in high risk clinical care setting in Rhode Island.** *AIDS Care* 1998; **10**:221–229.
49. Macalino GE, Vlahov D, Sanford-Colby S, Patel S, Sabin K, Salas C, Rich JD. **Prevalence and incidence of HIV, hepatitis B virus, and hepatitis C virus infections among males in Rhode Island prisons.** *Am J Public Health* 2004; **94**:1218–1223.
50. Zaller N, Noak A, Poshkus M, McDonald L, Rich JD. **The overlapping epidemics of viral hepatitis/HIV, addiction and incarceration: the situation in Rhode Island.** *Med Health R I* 2005; **88**:424–427.
51. Rogo T, DeLong AK, Chan P, Kantor R. **Antiretroviral treatment failure, drug resistance, and subtype diversity in the only pediatric HIV clinic in Rhode Island.** *Clin Infect Dis* 2015; **60**:1426–1435.
52. Chan PA, Reitsma MB, DeLong A, Boucek B, Nunn A, Salemi M, Kantor R. **Phylogenetic and geospatial evaluation of HIV-1 subtype diversity at the largest HIV center in Rhode Island.** *Infect Genet Evol* 2014; **28**:358–366.
53. Chan PA, Tashima K, Cartwright CP, Gillani FS, Mintz O, Zeller K, Kantor R. **Short communication: transmitted drug resistance and molecular epidemiology in antiretroviral naive HIV type 1-infected patients in Rhode Island.** *AIDS Res Hum Retroviruses* 2011; **27**:275–281.
54. Reece R, Delong A, Matthew D, Tashima K, Kantor R. **Accumulated preswitch resistance to more recently introduced one-pill-once-a-day antiretroviral regimens impacts HIV-1 virologic outcome.** *J Clin Virol* 2018; **105**:11–17.
55. Delong AK, Wu M, Bennett D, Parkin N, Wu Z, Hogan JW, Kantor R. **Sequence quality analysis tool for HIV type 1 protease and reverse transcriptase.** *AIDS Res Hum Retroviruses* 2012; **28**:894–901.
56. Shafer RW. **Rationale and uses of a public HIV drug-resistance database.** *J Infect Dis* 2006; **194 Suppl 1**:S51–S58.
57. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. **Human immunodeficiency virus reverse transcriptase and protease sequence database.** *Nucleic Acids Res* 2003; **31**:298–303.
58. Katoh K, Standley DM. **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol* 2013; **30**:772–780.
59. Pineda-Pena AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, *et al*. **Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools.** *Infect Genet Evol* 2013; **19**:337–348.
60. Struck D, Lawyer G, Ternes AM, Schmit JC, Bercoff BD. **COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification.** *Nucleic Acids Res* 2014; **42**:e144.
61. Siepel AC, Halpern AL, Macken C, Korber BT. **A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences.** *AIDS Res Hum Retrovirus* 1995; **11**:1413–1416.
62. Myers G, Korber BT, Hahn BH, Jeang K-T, *et al*. editors. *Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos, NM: Los Alamos National Laboratory; 1995. pp. III-35–III-60.
63. Stamatakis A. **RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies.** *Bioinformatics* 2014; **30**:1312–1313.
64. Stamatakis A. **Using RAxML to infer phylogenies.** *Curr Protoc Bioinformatics* 2015; **51**:11–14.
65. Novitsky V, Steingrimsson J, Howison M, Gillani F, Li Y, Manne A, *et al*. **Empirical comparison of analytical approaches for identifying molecular HIV-1 clusters.** *Sci Rep* 2020; **10**:18547.
66. Davison AC, Hinkley DV. Bootstrap methods and their application. No. 1. Cambridge University Press. 1997. Available at: http://statwww.epfl.ch/davison/BMA/CUPsample.pdf.