Emerging Microbes & Infections

ISSN: 2222-1751 (Online) Journal homepage: www.tandfonline.com/journals/temi20

Delineating the "Who-Where-How" of Persistent HIV Epidemics: A 23-Year Longitudinal Genetic Network, Phylodynamic, and Spatial Analysis for Precise Intervention in Rural China

Yi Chen, Xiaoshan Xu , Jianjun Li , Shujia Liang , Jinghua Huang , Kailing Tang , Hui Xing, Lingjie Liao , Yi Feng , Jingjun Huang , Wanling Chen , Yuhua Ruan & Ge Zhong

To cite this article: Yi Chen, Xiaoshan Xu , Jianjun Li , Shujia Liang , Jinghua Huang , Kailing Tang , Hui Xing, Lingjie Liao , Yi Feng , Jingjun Huang , Wanling Chen , Yuhua Ruan & Ge Zhong (13 May 2026): Delineating the "Who-Where-How" of Persistent HIV Epidemics: A 23-Year Longitudinal Genetic Network, Phylodynamic, and Spatial Analysis for Precise Intervention in Rural China, Emerging Microbes & Infections, DOI: 10.1080/22221751.2026.2671470

To link to this article: https://doi.org/10.1080/22221751.2026.2671470

Publisher: Taylor & Francis & The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd
Journal: *Emerging Microbes & Infections*
DOI: 10.1080/22221751.2026.2671470

Delineating the "Who-Where-How" of Persistent HIV Epidemics: A 23-Year Longitudinal Genetic Network, Phylodynamic, and Spatial Analysis for Precise Intervention in Rural China

Yi Chen[1,†], Xiaoshan Xu[2,†], Jianjun Li[3,†], Shujia Liang[3], Jinghua Huang[3], Kailing Tang[3], Hui Xing[2], Lingjie Liao[2], Yi Feng[2], Jingjun Huang[4], Wanling Chen[1], Yuhua Ruan[2,*], Ge Zhong[3,*]

1. The Guangxi Academy of Medical Sciences, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China.
2. State Key Laboratory of Infectious Disease Prevention and Control (SKLID), Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Center for AIDS/STD Control and Prevention (NCAIDS), Chinese Center for Disease Control and Prevention (China CDC), Beijing, China.
3. Guangxi Key Laboratory of Major Infectious Disease Prevention Control and Biosafety Emergency Response, Guangxi Center for Disease Control and Prevention, Nanning, China.
4. Guangxi University of Chinese Medicine, Nanning, China.

[†]These authors contributed equally to this work.

*These authors were corresponding authors. Corresponding authors' information: Yuhua Ruan, ruanyuhua92@chinaaids.cn, +86 13011098583. National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China. 102206.

Ge Zhong, gxzhongge@163.com, +86 13878166344. Guangxi Center for Disease Control and Prevention, Nanning, China. 530028.

Abstract

Despite "Treat All" strategies, HIV outbreaks persist among the elderly in rural Southwest China. To identify the core drivers ("Who"), geographic foci ("Where"), and socio-behavioral mechanism ("How") of sustained transmission in the post-"Treat All" era, we triangulated longitudinal genetic networks, phylodynamic and spatial analysis, and behavioral surveys using multi-source 1999-2021 data (N=5,094). Although antiretroviral therapy reduced community-level transmission risk by 21.0%, 107 molecular super-spreaders, predominantly males aged 50-69, were identified across four phases. These individuals exhibited poor treatment retention, with dropout rates reaching 28.2% and 38.5% having unsuppressed or missing viral loads by Phase 4. Notably, 94.4% of super-spreaders resided within 2 km of transportation arteries, exceeding 77.9% population baseline. Townships within this buffer exhibited significantly higher cumulative HIV cases independently of population metrics (adjusted Incidence Rate Ratio =1.43, 95% CI: 1.14-1.78). Characterized as infrastructure-anchored "risk hubs", these areas facilitate sustained viral dispersal through a "double-proximity" mechanism. Older men frequent nearby periodic markets along transit corridors for commercial sex, while female sex workers exhibit linear mobility across townships, effectively aligning their service cycles with the temporal rhythm of the periodic markets. Our methodology facilitates targeted node-

level testing and interventions, bridging clinical treatment and public health prevention to mitigate aging-driven epidemics in resource-limited regions.

Key words Longitudinal HIV genetic network; Phylogeographic analysis; Spatial epidemiology; "Treat All" policy; Precision public health

Background

The global scale-up of "Treat All" antiretroviral therapy (ART) policy represents a monumental achievement in the fight against HIV/AIDS, aiming to reduce mortality and interrupt transmission at the population level [1–3]. Yet, the persistence of localized HIV outbreaks poses a critical challenge to achieve universal epidemic control [4,5]. This implementation gap suggests that the public health impact of "Treat All" might be undermined not by drug efficacy, but by the existence of HIV super-spreaders and substantial silent reservoirs. These reservoirs, comprised of undiagnosed HIV-positive cases and high-risk individuals, remain beyond the reach of conventional screening [5].

This challenge was acutely evident in rural southwest China. Despite the nationwide rollout of "Treat All" in 2016 and significant progress under the Guangxi AIDS Conquering Project (GACP, 2010-2020) , featuring expanded ART coverage (scaling from 56.1% in 2016 to 86.7% by 2021), intensified education, and voluntary testing, which collectively decelerated epidemic growth [6,7], persistent high-risk behaviors, particularly unprotected sex, continued to fuel transmission, especially among the older population [8]. The number of newly diagnosed HIV/AIDS cases in Guangxi

remained at approximately 10,000 annually over the past five years, accounting for roughly 10% of the national total; of the newly diagnosed, heterosexuals continually accounted for over 90.0% [9], and the proportion of newly diagnosed aged ≥50 years increased sharply from 37.3% in 2010 to 67.2% in 2024, significantly exceeding the national average during the same period. Given that a small fraction of cases drives the majority of infections, precision-targeted interventions are essential to maximize public health returns [10].

However, precisely defining the "Who" and "Where" of these super-spreaders and "silent reservoirs", as well as "How" they make the epidemic persistent, remains a formidable challenge. Longitudinal viral phylogenetic data, which captures the underlying transmission network, offers a novel lens for such an audit [11,12]. When integrated with spatial analysis and behavioral survey, this approach moves beyond a simple census of diagnosed individuals to provide mechanistic insights into "Where" transmission is most active, "Who" is driving it, and "How" these dynamics sustain the epidemic. Yet, to date, most studies have relied on fragmented methodologies. For instance, spatial clustering based solely on reported cases [13,14] lacks molecular evidence to verify cross-regional linkages. Similarly, cross-sectional genetic "snapshots" [15–18] offer limited insight into the epidemic's long-term temporal evolution. Even recent advancements combining genetic networks with spatial analysis [19] remain largely descriptive, providing a static mapping of genetic linkages without capturing the longitudinal momentum of the epidemic's expansion. Furthermore, although macro-level phylogeographic studies [20] had traced CRF07_BC lineages across broad provincial

boundaries, such findings remain too coarse-grained to guide the localized, high-precision interventions necessary at the community level.

To bridge these gaps, we conducted a 23-year (1999-2021) longitudinal study in rural Southwest China, implementing a multidimensional framework that integrated Generalized Estimating Equations (GEE) modeling, Bayesian phylogeographic reconstruction and longitudinal genetic networks with spatiotemporal clustering and supplementary exploratory behavioral survey. Moving beyond static surveillance, this integrated approach enables us to evaluate the population-level impact of the "Treat All" policy while unmasking the "Who-Where-How" of the transmission engines and "silent reservoirs" that disproportionately sustain viral dispersal. By synthesizing these diverse data layers, we aimed to provide a robust evidence base for high-precision, context-specific interventions capable of neutralizing the hidden drivers and reversing the aging-associated trajectory of the HIV epidemic in rural Southwest China.

Materials and methods

Study setting and design

Guangxi is a high-prevalence region, with an HIV prevalence (0.25%) nearly triple the national average (0.09%) [21]. As one of China's top three provinces for new HIV/AIDS diagnoses, it epitomizes the unique epidemic trajectories observed in rural Southwest China that HIV epidemic transitioned from 1990s injecting drug user (IDU)-driven clusters to current elderly-dominated heterosexual transmission [22]. Within this context, Qinzhou, a strategic coastal port city and

transportation hub, was selected as the study site, as it ranks among the top three prefectures for HIV burden in the province. It serves as a representative model for regional HIV-1 dynamics [7,22]. This study retrospectively analyzed cases diagnosed from 1999 to 2021. Ethical approval was obtained from the Institutional Review Board of the China CDC (No. X221109713). The requirement for informed consent was waived by the ethics committee due to the study's retrospective nature and the use of de-identified data.

Study population and sample selection

The study population comprised all newly diagnosed HIV/AIDS cases in Qinzhou between 1999 and 2021. Residual serum specimens of these cases were obtained from routine HIV confirmatory tests (for ART-naïve cases) or CD4 count assessment (for ART-experienced cases). HIV-1 Ribonucleic Acid (RNA) was sequenced from pre-ART plasma samples cryopreserved at -80°C. For patients undergoing ART, pro-viral Deoxy ribonucleic Acid (DNA) was extracted from whole blood to ensure sequence recovery despite viral suppression. From this initial cohort, cases were excluded based on the following criteria: (i) age < 18 years; (ii) insufficient specimen volume; (iii) poor sequence quality (e.g., amplification failure, ≥ 5% mixed bases, pol sequences < 1,000 bp, or RT length < 238 codons); and (iv) incomplete epidemiological records. All nucleic acid extraction, amplification, and sequencing were performed at the Guangxi CDC HIV/AIDS Confirmatory Laboratory following established protocols [7,23,24].

Epidemiological data collection

The corresponding de-identified demographic and epidemiological data for the included cases were extracted from the Guangxi HIV/AIDS Comprehensive Response Information Management System of China's National System. The data encompassed age, gender, ethnicity, education, marital status, occupation, residence, HIV transmission route (risk), date of HIV diagnosis, baseline CD4 level (defined as the CD4 count at HIV diagnosis), HIV viral load, ART status (categorized as ART-naïve, active ART, or ART dropout). Dropout was assessed longitudinally and defined as treatment interruption or loss to follow-up for > 90 days since the last scheduled visit [25] . Patients who transferred to other ART authorities with confirmed referrals were continued as "active ART", while those with unverifiable status post-migration were classified as dropouts.

Epidemiological phasing

The study period was stratified into four distinct epidemiological phases: 1) Phase I (1999-2009): Pre-GACP phase, 2) Phase II (2010-2015): GACP intensive intervention phase, 3) Phase III (2016-2018): "Treat All" transition phase, and 4) Phase IV (2019-2021): "Treat All" full coverage phase. The rationale of the stratification is provided in Supplementary File 2 "1. Stratification of four epidemiological phases".

HIV genetic network reconstruction and topological importance quantified

HIV genetic networks were reconstructed using all available sequences (N=5,094) spanning four epidemiological phases (1999–2021). Pairwise genetic distances were calculated using HyPhy (version 2.2.4) under the Tamura-Nei 93 (TN93) model. Putative transmission links were identified

using a stringent threshold of 0.005 substitutions/site to capture recent (2-3 years) transmission events [26]. For each phase, township-level topological importance was quantified by calculating global parameters (nodes, edges, density, and number/size of clusters) and centrality indices (mean degree, betweenness, and eigenvector centrality). Individuals whose degree centrality exceeded the mean by two standard deviations ($\mu + 2\sigma$) within each respective phase were defined as molecular super-spreaders. This stringent statistical threshold ensures the identification of extreme outliers with the most significant impact on network expansion while accounting for temporal variations in network size and connection density across the four epidemiological phases. To protect participant privacy, specific geographic identifiers (e.g., township and street names) were redacted from the manuscript. However, these granular data were retained for internal geospatial analysis to identify high-risk clusters and inform local interventions. This approach balances epidemiological utility with strict confidentiality.

BEAST-based phylogeographic analysis

HIV-1 subtype identification was based on MAGE (v10.0). Bayesian phylogeographic analysis was performed based on RNA sequences of five major HIV subtypes/clusters using BEAST 1.8.4. The specific subtype distribution for these RNA sequences was illustrated in Supplementary file 1 Table S1. As detailed in Supplementary File 2 " 2. Phylodynamic and phylogeographic analysis", we processed sequences and configured BEAST models to investigate the epidemic's trajectory. Specifically, we employed a phylogenetic time-slicing approach to quantify temporal shifts in the demographic composition of each lineage across four epidemiological phases, visualized using

smoothed stacked area charts[27]. In parallel, Bayesian stochastic search variable selection (BSSVS) was applied to infer viral migration dynamics between subgroups [28].

Geospatial and genetic network integration analysis

Geospatial distribution of transmission hubs was analyzed by aggregating genetic network features at the township level (georeferenced by geometric centroids), visualized using ArcGIS 10.8. Intra-township links and inter-township connections were visualized to represent localized and regional connectivity, respectively. Specifically, mean degree centrality and Mean Molecular Cluster Growth (MMCG) were used to quantify connectivity and expansion intensity. MMCG's definition and calculation was described in Supplementary file 2 " 3. Mean Molecular Cluster Growth calculation". These metrics were further integrated with Geographic Information System (GIS) data, utilizing road proximity and spatial autocorrelation, to evaluate the impact of transportation infrastructure on viral dissemination. Negative Binomial regression analysis was further performed to isolate the impact of road proximity from the effects of general population aggregation. Details methods for spatial autocorrelation and road proximity analysis were provided in Supplementary file 2 " 4. Spatial autocorrelation analysis" and " 5. Road proximity analysis".

A supplementary exploratory behavioral survey

To explore the socio-behavioral mechanism of the observed molecular and spatiotemporal patterns, we conducted a supplementary exploratory behavioral survey targeted local men (N=84) across three townships along the major transportation corridors in June-July 2025. A hybrid sampling strategy

was employed. Initially, 22 diagnosed HIV cases were recruited as seeds for potential chain-referral; subsequently, to ensure broader community representation, we integrated venue-based sampling at traditional informal social gathering points (e.g., village entrances and community parks) frequented by older adults. Socio-demographic profiles, the frequency and spatial context of non-marital sexual encounters (including both commercial and casual sex), and condom use behaviors were collected from all participants. Additionally, data regarding the geographic location and mobility status of the encountered female sex workers (FSWs) were recorded. To ensure maximum participant protection and minimize social stigmatization, verbal informed consent was obtained in lieu of written documentation, and all data were de-identified and restricted to authorized public health personnel to maintain strict confidentiality. The interviews were approved by the IRB of the National Center for AIDS/STD Control and Prevention, Chinese CDC (IRB No. X221109713).

Statistical analysis

Analyses were performed in R (version 4.3.2) using sf for spatial processing, spdep for autocorrelation metrics and ggplot2 for visualization. Chi-square tests were used to compare categorical variables. A GEE model was used to assess the association between ART status and onward HIV transmission among all samples (N=5094), details in Supplementary file 2 "6. Generalized estimating equations analysis". A two-tailed $P < 0.05$ was considered significant.

Results

Epidemiological landscape and cohort characteristics (1999-2021)

From 1999 to 2021, a total of 7,381 HIV/AIDS cases were newly diagnosed in the study area. 69.0% (n = 5,094) met the inclusion criteria with high-quality sequence and complete epidemiological records. This sequencing cohort comprised 4,814 plasma RNA sequences and 280 pro-viral DNA sequences (Supplementary file 1 Figure S1). Demographic comparison confirmed that the 4,814-case RNA sub-cohort was statistically representative of the overall diagnosed population (all $P > 0.05$; Supplementary file 1 Table S4), thus validating its use for subsequent Bayesian phylodynamic (BEAST) analysis. As shown in Table 1, the overall study population were primarily composed of males (74.6%), cases aged over 50 years old (52.9%), Han ethnicity (89.6%), and farmers (75.0%), with heterosexual transmission (89.9%) being the predominant route. Over the 23-year span, the cohort exhibited a robust aging trend, with the proportion of cases aged 50-69 surging from 11.8% to 51.0% ($P_{trend} < 0.001$). The proportion of divorced/widowed cases rose from 11.5% to 25.4% ($P_{trend} < 0.001$). Further, a near-complete transition from injecting drug use to heterosexual transmission was observed ($P_{trend} < 0.001$). At the molecular level, these changes mirrored a transition toward a more diversified viral reservoir. While CRF01_AE and CRF08_BC remained the dominant strains in absolute numbers, their proportional contributions significantly contracted ($P_{trend} < 0.001$). Conversely, CRF07_BC experienced rapid expansion, increasing from 4.5% to 14.9% ($P_{trend} < 0.001$). Although CRF55_01B also showed a statistically significant upward trend ($P_{trend} < 0.001$), its overall prevalence remained below 2.0% throughout the study period (Details see Table 1 and Supplementary file 1 Figure S2).

Population-level protection of the "Treat All" policy

The implementation of the nationwide "Treat All" policy in 2016 provided valuable real-world evidence to evaluate the impact of immediate ART. Although its effective execution in rural areas exhibited a temporal lag, the population-level efficacy remains profound. GEE analysis based on 5094-sequence dataset with 19661 observations shows that: active ART was associated with a significant 21% reduction in HIV secondary transmission [Adjusted odds ratio (AOR) = 0.79, 95% CI: 0.73-0.87], while ART dropout showed no significant protective effect compared to being ART-naïve (See Table 2). However, GEE identified several factors independently associated with an increased risk of secondary transmission, most notably older age, which exhibited a graded increase in risk [35-49 years, AOR = 1.60 (95%CI: 1.20-2.14); 50-69 years, AOR = 3.59 (95%CI: 2.68-4.81); ≥70 years, AOR = 4.22 (95% CI: 3.09-5.77), all vs. < 25 years], male gender (AOR = 1.42, 95% CI: 1.29-1.56), infected with CRF07_BC (AOR = 1.94, 95% CI: 1.71-2.21) or CRF08_BC subtypes (AOR = 1.75, 95% CI: 1.59-1.92), residence in Qinbei (AOR = 1.17, 95% CI: 1.05-1.31), and higher baseline CD4 level [350-499 vs. < 350: AOR = 1.23 (95%CI: 1.10-1.37); ≥ 500 vs. < 350: AOR = 1.26 (95%CI: 1.10-1.45)], see Table 2.

The persistence of high-risk profiles suggests that specific subgroups may bypass macro-level policy gains. To determine how these risks translate into actual transmission, we performed longitudinal HIV genetic networks analysis to pinpoint the molecular super-spreaders ("Who") that function as the primary drivers of the epidemic.

Topological evolution of the longitudinal HIV genetic networks

Based on the 5094-sequence dataset, the topological characteristics of the HIV genetic networks exhibited significant expansion and structural shifts across the four epidemiological phases (See Table 3). The network expanded from 34 nodes and 51 edges (1999-2009) to 729 nodes and 1,408 edges (2019-2021). While the average degree peaked at 4.21 in 2016-2018, network density dropped from 0.0909 to 0.0053, signaling a transition from compact early clusters to a sparse, large-scale system. The number of clusters grew 43-fold (from 4 to 175). Notably, the maximum cluster size peaked at 51 in 2016-2018, suggesting this period was a critical window for the formation of major transmission chains. The number of molecular super-spreaders rose from 1 to a peak of 41 in 2016-2018. Although the proportion of molecular super-spreaders declined to 5.35% by the final phase, their edge proportion reached a record high of 61.65%. This indicated that a decreasing minority of individuals was responsible for an increasing majority of transmission links. The HIV genetic networks across the four epidemiological phases are shown in Supplementary file 1 Figure S4, and their geographic distributions are presented in Figure 2.

Profiles of molecular super-spreaders

Based on HIV genetic networks constructed from all 5,094 available sequences, Table 4 summarizes the profiles of identified molecular super-spreaders across the four phases. The number of super-spreaders increased from 1 to 39, with their maximum degree centrality rising from 18 to 34. The transmission dynamics shifted from younger IDU populations in the early stages to an older, heterosexually active demographic in the final phase. Although men dominated majorly, the women composition increased from 0.0% in phase 1 to 20.5% in phase 4. By 2019-2021, 94.9% of

molecular super-spreaders were ≥50 years old. Within this group, farmers accounted for 82.1% and commercial heterosexual contact for 71.8% of cases. The subtype distribution shifted from an exclusive CRF08_BC presence (100.0%) in Phase 1 to a more diverse mixture in subsequent phases, eventually being overtaken by CRF07_BC (74.4%) in Phase 4. Geographically, molecular super-spreaders evolved from a localized focus in TX township to a multi-focal distribution, anchored in townships along major transportation hubs, such as townships of PJ, QT, and XX etc. While local ART coverage increased from 56.1% in 2016 to 86.7% by 2021, molecular super-spreaders exhibited a concurrent rise in dropout rates, escalating from 19.2% in Phase 2 to 28.2% in Phase 4. Furthermore, the proportion of these super-spreaders with unsuppressed (≥ 200 copies/mL) or missing viral loads rose from 0.0% in Phase 2 to 38.5% by Phase 4, indicating a marked decline in treatment adherence within the identified transmission networks.

To investigate how these individual-level transmission risks scale up to shape the broader regional epidemic, we further conducted phylodynamic and viral migration analyses across demographic subgroups.

Phylodynamic reconstruction and demographic transmission dynamics

Bayesian evolutionary reconstruction of HIV-1 subtypes/clusters

Bayesian phylodynamic analysis revealed that five HIV-1 subtypes independently converged into a shared heterosexual-driven transmission network in Qinzhou. CRF01_AE (Cluster 1 and 2) (Supplementary file 1 Figure S3-1 and S3-2) and CRF08_BC (Supplementary file 1 Figure S3-4)

originated from regional IDUs in the mid-1990s before transitioning to heterosexual transmission circa 1999. CRF07_BC (Supplementary file 1 Figure S3-3), initially circulating among men who have sex with men (MSM), later diverged into distinct heterosexual and MSM-hybrid clusters. While CRF55_01B (Supplementary file 1 Figure S3-5) predominantly affected younger men after 2005, other major lineages exhibited a progressive convergence toward older age groups, establishing a spatiotemporally overlapping transmission reservoir among the rural elderly. Time to most recent common ancestor (tMRCA) is shown in Supplementary file 1 Table S5, all major estimates had strong statistical support, posterior probability (PP) $\geqslant$0.90.

Temporal dynamics of lineage composition defined by age-gender state frequencies

The smoothed stacked area charts (Figure 1) delineate the temporal shifts in lineage composition across four epidemiological phases. By quantifying reconstructed ancestral states, these visualizations reveal a progressive demographic shift, as the 18-49 age group contracted, older cohorts (particularly those aged 50-69 and ≥70) expanded to become the predominant components across all major lineages. This transition was particularly aggressive in CRF08_BC, where the older cohort's contribution surged from negligible levels in the early 2000s to dominance by 2008. While CRF01_AE Cluster 1 maintained long-standing geriatric dominance, CRF01_AE Cluster 2 and CRF07_BC underwent a steady replacement of middle-aged males by older cohorts, a shift that accelerated after 2010. Male lineages consistently predominated the transmission pool. By 2015, these convergent trajectories had established the rural male elderly as the primary and expanding transmission reservoir for the regional HIV-1 flow.

Inference of asymmetric transmission dynamics between age-gender subgroups

Subsequently, BSSVS analysis (Supplementary Table S6, Figure S5) identified bidirectional and asymmetrical viral flows across age-gender cohorts. The 50-69M group exhibited the highest outward transition rates across all lineages, acting as the predominant source of viral dissemination. Specifically, in CRF01_AE Cluster 2, robust transmissions were observed from 50-69M to females to nearly all age groups (BF = 670,332), alongside significant internal transmission (50-69M → 50-69F). In contrast, the 35-49M group primarily facilitated both intra-group transmission and downward viral flow toward younger cohorts, notably via the 35-49M → 35-49M and 35-49M → 25-34F pathways, respectively, across all lineages except for CRF08_BC. While the lineage state frequency of the ≥70 group increased over time (Figure 1), BSSVS results characterized this cohort as a net recipient of infections. Transitions into the ≥70 group from the 50-69M reservoir were highly supported, yet outward transitions from the ≥70 group remained minimal across all lineages.

Coupled spatiotemporal evolution of HIV genetic networks with transportation infrastructure

Following the identification of "Who" drives the epidemic, we characterized the "Where" by integrating longitudinal genetic network metrics with spatial statistical analysis. The spatial distribution of network connectivity (represented by mean degree centrality) showed a progressive geographical dispersion over the study period. This trend was evidenced by the Global Moran's I, demonstrated in Supplementary file 1 Table S7, which declined from 0.298 (1999-2009, $P < 0.001$) to a non-significant 0.115 by 2019-2021 ($P = 0.057$). In contrast, recent network expansion,

measured by MMCG, exhibited an opposing trend of increasing spatial concentration. The clustering of MMCG remained consistently significant ($P < 0.01$) and peaked during the 2019-2021 period ($I = 0.373$, $Z = 4.765$), indicating that recent transmission hotspots have become increasingly localized.

Local indicators of spatial association (LISA) analysis demonstrated a significant geographic alignment between HIV expansion and transportation arteries. For historical connectivity (Figure 3-1), High-High hubs of degree centrality were initially concentrated at major highway arteries in the East-north region (1999-2009). Over the 23-year period, these connectivity hubs progressively migrated along the national and provincial highway arteries and established new transmission clusters in western areas by 2019-2021. Parallel LISA analysis of MMCG (Figure 3-2) confirmed that recent genetic growth followed a similar trajectory but with higher intensity. Recent hotspots were consistently anchored to and extended along the primary road networks, with genetic expansion mirroring the layout of national (blue lines) and provincial (orange lines) highway arteries.

Distance-base quantitative analysis further substantiated the role of transportation corridors as primary conduits for viral expansion. Correlation analysis identified a negative relationship between geographic proximity to roads and genetic network metrics (Figures 4-1 and 4-2). For historical connectivity, the correlation with road distance became significant after 2010 (2010-2015: $r = -0.361$, $P = 0.002$), while genetic growth peaked in its dependency on road proximity during 2016-2018 ($r = -0.323$, $P = 0.006$). Crucially, transportation buffer zone analysis (Figure 4-3) revealed a temporal emergence of stark divergence in transmission intensity based on road proximity. While no

significant difference was observed in the earliest phase (1999-2009, P = 0.808), townships located within 2-km buffer zones of major roads exhibited significantly higher molecular cluster growth compared to those in distal areas from 2016 onwards (2016-2018, P = 0.011; 2019-2021, P = 0.007). By the final period (2019-2021), the MMCG in near-road townships was more than double that of distal areas, confirming that the recent intensification of HIV transmission was predominantly concentrated along major transportation arteries. After adjusting for population size and density, the negative binomial regression model reveals a significant independent association between road proximity and the burden of cumulative reported HIV cases [adjusted Incidence Rate Ratio (IRR) = 1.431; 95% CI: 1.143-1.782, p = 0.0017]. Furthermore, while townships within the 2-km highway buffer account for 77.9% (2,710,918/3,479,690) of the total population in Qinzhou, they host a disproportionately high 94.4% (101/107) of the identified molecular super-spreaders ($X^2 = 17.65$, $P < 0.001$). These findings confirm that the transportation corridor effect persists independently of demographic aggregation.

To elucidate the social-behavioral mechanisms ("How") underlying the localization of the epidemic along transportation arteries, a supplementary exploratory behavioral survey targeting local men was conducted.

Behavioral drivers of corridor-based viral dispersal

A total of 84 men were recruited, among 22 seeds (diagnosed HIV cases), only two successfully initiated a single wave of recruitment, referring two and three followers, respectively. The remaining

participants (n=57) were recruited via venue-based sampling at informal gathering sites frequented by older rural men. The study participants were predominantly male (97.6%) and aged over 50 years (91.7%), with 48.8% of participants being widowed, divorced, or single. Their social and recreational activities primarily occurred in informal outdoor settings, including village entrances for card playing (75.0%), community parks (56.0%), public squares (54.8%) and periodic township markets (47.6%) etc. Regarding sexual risk, 48.8% of respondents reported engaging in casual or commercial heterosexual sex, with 70.7% practicing inconsistent condom use. Notably, 88.9% of commercial sexual encounters occurred in townships within a 2-km buffer of transportation arteries. The median number of commercial partners for those engaging in commercial sexual contacts was 2 (IQR: 2-4). Commercial sex encounters were characterized by low entry costs (low-tier). Specifically, over one-third of respondents (37.5%) reported prices below 50 CNY per transaction, while only 27.1% exceeded 100 CNY. Furthermore, 60.9% of respondents visited FSWs at least semi-annually. While 66.7% of encountered FSWs were from within Qinzhou, over 70.0% of respondents noted that these FSWs were highly mobile, rotating among townships along major transportation arteries to synchronize with local periodic market days. Detailed survey questions and results are demonstrated in Supplementary file 1 Table S8.

Discussion

The strength of this study lies in its multi-layered analytical framework, which moves beyond isolated methodologies toward a truly integrative molecular epidemiological approach. On a global scale, landmark research such as the Rakai Community Cohort Study in Uganda[29,30] has

demonstrated the power of coupling longitudinal cohorts with deep-sequencing phylogenetics to track transmission bottlenecks. However, our framework extends this paradigm by integrating fine-grained demographic stratification with transportation buffer zone analysis, which often overlooked in traditional phylodynamic studies. While recent advancements in China have begun to combine genetic networks with spatial mapping [19], they frequently remain descriptive and static. In contrast, our approach uniquely bridges the gap between molecular evidence and structural drivers. By triangulating longitudinal network evolution and phylogeographic reconstruction with transportation-buffer analysis, we identify that the "transportation corridor effect" represents an independent risk factor for the local HIV-1 burden. This finding was further substantiated by the "double-proximity" model identified through our supplementary behavioral survey. This model characterizes the localized, nearby mobility of older men (clients) to townships along the transportation arteries on market days, and the linear mobility of FSWs navigating these corridors to access client flow. Such a holistic "Who-Where-How" characterization provides the granular precision necessary to uncover undiagnosed cases and high-risk individuals within "silent reservoirs", a depth of insight that macro-level phylogeographic study [20] or purely genetic "snapshots" [15–18] cannot achieve. This study finding can further facilitate geographically targeted screening and enhanced case-finding along transportation arteries.

The GEE analysis confirmed the significant population-level success of the "Treat All" policy, with active ART reducing secondary transmission risk by 21%. This aligned with the "Treatment as Prevention" (TasP) paradigm, suggesting that early treatment had successfully blunted the overall

epidemic trajectory [31]. Regarding the global "95-95-95" targets, China reached an estimated "84-93-97" in 2022 [32]. Our findings suggest that a primary driver of this "first 95" shortfall in rural Southwest China is the under-detection and subsequent "invisibility" of high-risk marginalized groups in genetic networks. Specifically, while the sustained male dominance underscores older men as primary epidemic drivers, this pattern may be biased by the under-sampling of populations such as low-tier FSWs, as indicated in our behavioral survey and previous study[33]. Due to the clandestine nature of their work, high mobility, and significant healthcare barriers [34], these potential core transmitters often remain undiagnosed, leading to their underrepresentation in network analysis. Nevertheless, the rising female proportion among molecular super-spreaders (20.5% in Phase 4) signals an intensifying heterosexual transmission dynamic, further emphasizing the urgent need to bridge this detection gap through geographically and demographically targeted strategies. Furthermore, we identified a 28.2% ART dropout rate and 38.5% rate of unsuppressed or missing viral loads among Phase 4 molecular super-spreaders, reflecting a structural breakdown between the "second" and "third" 95-95-95 targets. The lack of protective effect in dropouts (AOR=1.01) further underscored that high nominal coverage often masks functional gaps in transmission hubs. In these settings, suboptimal adherence acts as a bottleneck, undermining ART's capacity to achieve the sustained viral suppression required to break the chain of transmission, a finding consistent with previous studies [35].

Our phylogeographic analysis indicated that the current HIV epidemic in Qinzhou has evolved from isolated outbreaks among IDUs into a mature epidemic fueled by external viral introductions and

subsequent sustained local circulation among older heterosexual populations. This shift is typified by a profound geriatric transition across all viral lineages. As shown in our temporal reconstructions, the contraction of younger cohorts (25-49 years) has led to the epidemic being increasingly sequestered within the oldest demographics, with males aged ≥ 50 emerging as the predominant state. This demographic "takeover", most aggressive in CRF08_BC and accelerating after 2010 in CRF07_BC, is further corroborated by our identification of molecular super-spreaders within these age groups. Such a transition aligns with the broader "graying" of China's HIV epidemic [36,37], a trend particularly pronounced in Qinzhou, which consistently ranked among the top five prefectures in China for new HIV diagnoses among elderly (≥ 50 years) individuals linked to commercial sex [38]. The BSSVS analysis elucidates the functional hierarchy driving this shift. Specifically, males aged 50-69 function as the primary hub of viral dissemination, driving infections both internally and upward to the oldest-old, a phenomenon likely facilitated by low-tier FSWs [39]. These results highlight a critical surveillance gap in rural commercial sex markets [40]. Conversely, despite their declining share in lineage composition, males aged 35-49 act as both intra-group transmission and downward bridges to younger females. This role mirrors patterns observed in other generalized epidemics and ensures the continuous re-seeding and long-term persistence of the regional viral pool [17]. Finally, although the ≥ 70 group is expanding rapidly, it functions predominantly as a transmission sink, absorbing infections from the male aged 50-69 reservoir with minimal outward spread [41]. Collectively, these findings pinpoint the "silent reservoirs" sustaining the regional viral flow, addressing the "Who" in the epidemic's transmission dynamics.

Our longitudinal genetic network and spatiotemporal analyses reveal a strategic transition of the epidemic from a localized cluster in Tanxu (Phase 1) to a broad geographic concentration anchored along major transportation arteries ("Where"). This spatial shift is underscored by the LISA analysis, which confirms genetic network growth along the transportation corridors, and the persistent impact of transit proximity, evidenced by a 43.1% increased risk of cumulative HIV cases in townships within 2 km buffer of highway arteries. Furthermore, the gap between the 94.4% molecular super-spreader concentration and the 77.9% population baseline confirms that transportation infrastructure acts as an independent functional driver. Rather than a mere proxy for population aggregation, these arteries serve as structural conduits that facilitate and anchor regional viral dispersal. This "corridor-driven" expansion mirrors the historical dissemination of the HIV-1 primordial ancestor from Kinshasa, which was similarly facilitated by colonial transportation networks[27].

To explain these dynamics, we propose a "double-proximity" model, which integrates supplementary exploratory behavioral survey data with molecular and spatial findings to characterize the spatiotemporal synchronization of high-risk behaviors ("How"). This model is defined by the convergence of two distinct mobility patterns: the localized mobility of clients and the linear mobility of providers. Specifically, older rural males ($\geqslant$ 50 years) exhibit restricted, routine-driven mobility, seeking low-tier commercial sex at township hubs along the transportation arteries during periodic market days. Conversely, low-tier FSWs exhibit linear, corridor-based mobility, rotating among venues along highway axes to access these periodic market-day client flows [42]. This dual alignment of geographic proximity and behavioral timing creates transient, high-intensity "risk hubs"

that facilitate the sustained dispersal of HIV-1 lineages across the regional network. This infrastructure-dependent pattern diverges from the long-distance migratory models seen in Sub-Saharan Africa [43] or urban MSM-driven epidemics in China [44]. Instead, it highlights how transportation corridors serve as the primary conduits for viral spread among an aging, non-migratory population.

Hence, future interventions should shift from passive surveillance to data-driven proactive engagement. Outreach must prioritize MMCG hotspots along transportation corridors to optimize resource allocation. Specifically, mobile outreach taskforces should provide point-of-care testing directly at clandestine venues, such as roadside inns, rental houses, and informal gathering spots, where elderly individuals and FSWs congregate. To normalize screening and reduce stigma, HIV testing should be integrated into routine geriatric physical exams at township hospitals or health centers under China's basic public health services. For identified cases, community-based follow-up models involving village doctors were essential to provide tailored adherence support and reduce high dropout rates among elderly cases especially for those molecular super-spreaders [45]. Finally, sustained prevention efforts, including peer education and venue-based outreach at village gathering spots for the elderly, remain crucial for controlling localized outbreaks.

This study has several limitations. First, as the sample is from only one prefecture in Guangxi, it may not fully represent the entire southwest China. However, since Qinzhou shares similar HIV/AIDS epidemic trends and characteristics with Guangxi, the findings still hold significance and can guide

precise intervention. Second, while the supplementary behavioral survey had a limited sample size and potential selection bias due to venue-based sampling, it was designed as a contextual probe rather than a large-scale representative study. These findings provide critical "ground-truth" for the socio-behavioral mechanisms driving transmission along transportation arteries. While not fully generalizable provincial-wide, they offer essential mechanistic insights into the localized dynamics of the epidemic. Third, the number of molecular super-spreaders may be underestimated due to our stringent identification threshold ($\mu + 2\sigma$) and the inherent exclusion of undiagnosed or un-sequenced individuals. While such sampling gaps, a common challenge in molecular epidemiology, could omit certain high-risk groups like low-tier FSWs, this longitudinal dataset remains the most representative empirical evidence currently available for characterizing regional transmission dynamics. Finally, while the link between road proximity and HIV burden was independent of population metrics, unmeasured socioeconomic confounders like gross domestic product or urbanization may persist. Future research incorporating multidimensional socioeconomic data is needed to isolate the specific effects of infrastructure from regional development.

Authors' contribution

YR and YC conceived the idea. YC, JL and XX wrote the study protocol and designed data analysis. KT, SL, JL, WC, JH and GZ contributed to data collection, management and provided laboratory support. JL, KT and YR verified the underlying data. YC and XX conducted genetic transmission network analysis. YC and JH conducted spatial analysis. HX, LL and YF supervised the study. YC

and XX drafted the manuscript. YR, GZ and SL revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The HIV sequences and related epidemiological data used in this study were collected by our team from Qinzhou, Guangxi, China. These data and related datasets are not publicly available but can be obtained upon reasonable request and approval from the Chinese Center for Disease Control and Prevention. Requests for accessing these data and datasets should be directed to YR via email: ruanyuhua92@chinaaids.cn.

Fundings

This work was supported by the Guangxi Natural Science Foundation Project (Grant no. 2025GXNSFAA069868), National Natural Science Foundation of China (Grant no. 82260670), National Natural Science Foundation of China (Grant no. 12571522), Ministry of Science and Technology of China (Grant no. 2022YFC2305205, 2022YFC2305201), Guangxi Key Laboratory of AIDS Prevention Control and Translation (Grant no. ZZH2020010), Guangxi Bagui Honor Scholarship, Guangxi Medical and Health Key Discipline Construction Project.

Acknowledgments

We extend our gratitude to the Qinzhou Center for Disease Control and Prevention for their invaluable contribution to data collection. Special thanks to Yaomin Qin for his technical support for the visualization of lineage composition shifts presented in Figure 1.

Disclosure statement

No potential conflict of interest was reported by the author(s).

List of abbreviations:

ART : antiretroviral therapy ; GACP : Guangxi AIDS Conquering Project;

IDU : injection drug use; RNA: Ribonucleic Acid ; DNA: Deoxy ribonucleic Acid;

MCMC: Markov chain Monte Carlo; GTR: general time-reversible;

ESS: Effective Sample Size; UCLN: uncorrelated lognormal;

MCC: Maximum clade credibility; BSSVS: Bayesian stochastic search variable selection;

BF: Bayes Factors; TN93: Tamura-Nei; LISA: Local Indicators of Spatial Association;

GEE: Generalized Estimating Equations; tMRCA: time to the most recent common ancestor;

HPD: Highest posterior density; AOR: Adjusted odds ratio;

MSM: men who have sex with men; VCT: voluntary counseling and testing.

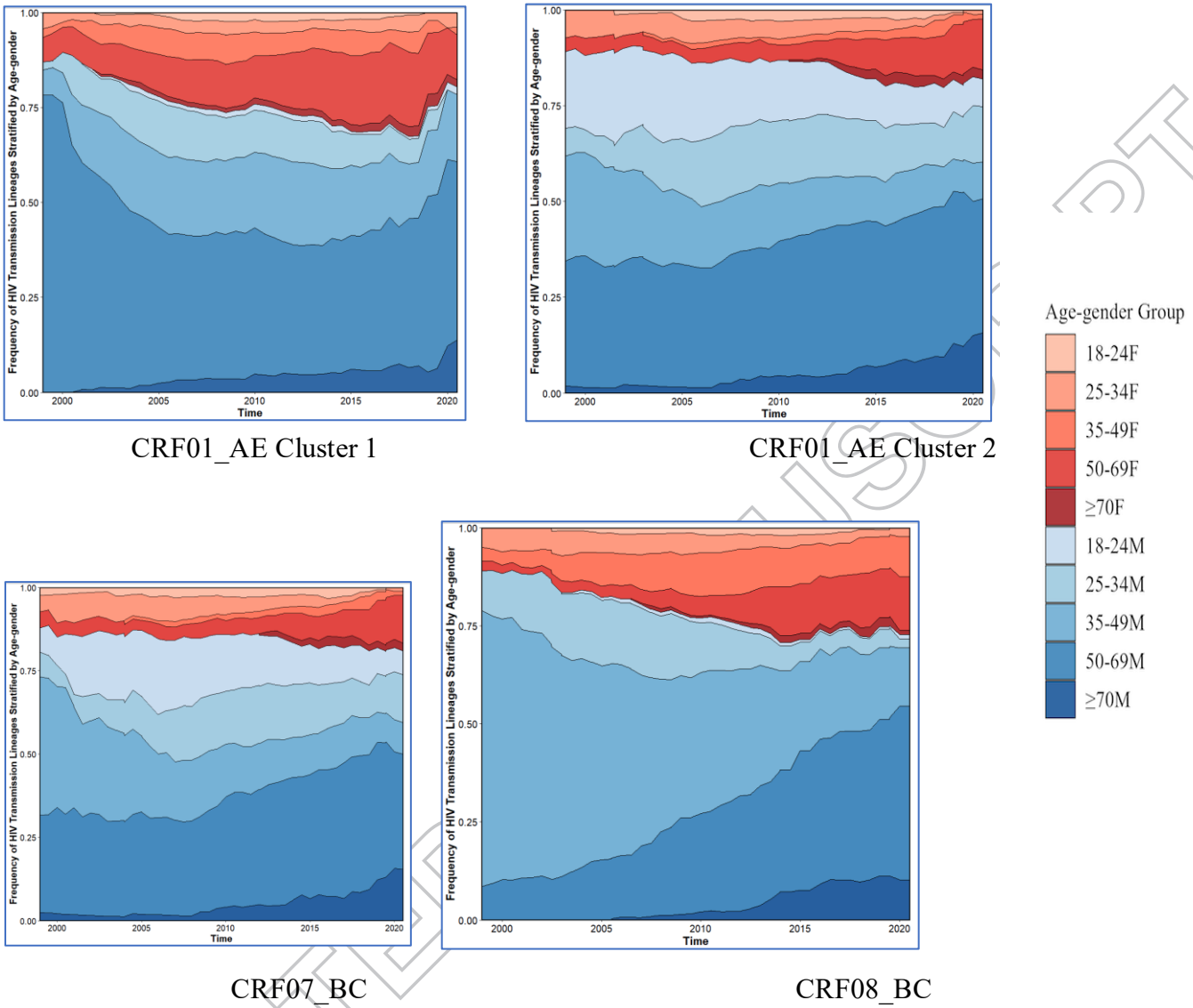
Reference

1. Mugenyi L, Nanfuka M, Byawaka J, Agaba C, Mijumbi A, Kagimu D, *et al.* Effect of universal test and treat on retention and mortality among people living with HIV-infection in Uganda: An interrupted time series analysis. PLoS ONE 2022;17:e0268226.

2. Rosen JG, Muraleetharan O, Walker A, Srivastava M. Pediatric Antiretroviral Therapy Coverage and AIDS Deaths in the "Treat All" Era. Pediatrics 2023;151:e2022059013.
3. Chao A, Spiegelman D, Khan S, Walsh F, Mazibuko S, Pasipamire M, *et al.* Mortality under early access to antiretroviral therapy vs. Eswatini's national standard of care: the MaxART clustered randomized stepped-wedge trial. HIV Med 2020;21:429–40.
4. Pandey A, Galvani AP. The global burden of HIV and prospects for control. Lancet HIV 2019;6:e809–11.
5. Zhao B, Qiu Y, Song W, Kang M, Dong X, Li X, *et al.* Undiagnosed HIV Infections May Drive HIV Transmission in the Era of "Treat All": A Deep-Sampling Molecular Network Study in Northeast China during 2016 to 2019. Viruses 2022;14:1895.
6. Ma S, Chen Y, Lai X, Lan G, Ruan Y, Shen Z, *et al.* Predicting the HIV/AIDS epidemic and measuring the effect of AIDS Conquering Project in Guangxi Zhuang Autonomous Region. PloS One 2022;17:e0270525.
7. Chen Y, Cao Z, Li J, Chen J, Zhu Q, Liang S, *et al.* HIV transmission and associated factors under the scale-up of HIV antiretroviral therapy: a population-based longitudinal molecular network study. Virol J 2023;20:289.
8. He J, Zhong S, Qin C, Nong A, Lin Z, Liang H, *et al.* The trend, prevalence and potential risk factors of secondary HIV transmission among HIV/AIDS individuals receiving ART in Guangxi, China: a longitudinal cross-sectional study. Emerg Microbes Infect 2024;13:2429622.
9. Ge X, Yang W, Zhu Q, Wu X, Shen Z, Zhu J, et al. Epidemiological characteristics of HIV/AIDS in Guangxi Zhuang Autonomous Region,2010-2017.Chin J Epidemiol, 2019 March;40(3):315-321. (In Chinese)
10. He S, Dong W, Fairley CK, Li Z, Wei Y, Lai H, *et al.* Optimizing health resource allocation for improving timely HIV diagnosis in China. J Int AIDS Soc 2024;27:e26221.
11. Han X, Zhao B, An M, Zhong P, Shang H. Molecular network-based intervention brings us closer to ending the HIV pandemic. Front Med 2020;14:136–48.
12. Oster AM, France AM, Panneer N, Bañez 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 1999 2018;79:543–50.
13. Qin Q, Guo W, Tang W, Mahapatra T, Wang L, Zhang N, *et al.* Spatial Analysis of the Human Immunodeficiency Virus Epidemic among Men Who Have Sex with Men in China, 2006-2015. Clin Infect Dis Off Publ Infect Dis Soc Am 2017;64:956–63.
14. Huang G, Cheng W, Xu Y, Yang J, Jiang J, Pan X, *et al.* Spatiotemporal Pattern and Its Determinants for Newly Reported HIV/AIDS Among Older Adults in Eastern China From 2004 to 2021: Retrospective Analysis Study. JMIR Public Health Surveill 2024;10:e51172.
15. Pang X, Ma J, He Q, Tang K, Huang J, Fang N, *et al.* HIV molecular transmission networks among students in Guangxi: unraveling the dynamics of student-driven HIV epidemic. Emerg Microbes Infect 14:2459142.
16. Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO. HIV Transmission Networks Among Transgender Women in Los Angeles County: network analysis of surveillance data. Lancet HIV 2019;6:e164–72.

17. de Oliveira T, Kharsany ABM, Gräf T, Cawood C, Khanyile D, Grobler A, *et al.* Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. Lancet HIV 2017;4:e41–50.
18. Dávila-Conn V, García-Morales C, Matías-Florentino M, López-Ortiz E, Paz-Juárez HE, Beristain-Barreda Á, *et al.* Characteristics and growth of the genetic HIV transmission network of Mexico City during 2020. J Int AIDS Soc 2021;24:e25836.
19. Yuan D, Yu B, Liang S, Fei T, Tang H, Kang R, *et al.* HIV-1 genetic transmission networks among people living with HIV/AIDS in Sichuan, China: a genomic and spatial epidemiological analysis. Lancet Reg Health - West Pac 2022;18:100318.
20. Gan M, Zheng S, Hao J, Ruan Y, Liao L, Shao Y, *et al.* Spatiotemporal Patterns of CRF07_BC in China: A Population-Based Study of the HIV Strain With the Highest Infection Rates. Front Immunol 2022;13:824178.
21. Cai C, Tang H, Li D, Qin Q, Chen F, Jin Y, *et al.* Evolution of HIV Epidemic and Emerging Challenges — China, 1989–2023. China CDC Wkly 2024;6:1251–6.
22. Chen H, Luo L, Pan SW, Lan G, Zhu Q, Li J, *et al.* HIV Epidemiology and Prevention in Southwestern China: Trends from 1996-2017. Curr HIV Res 2019;17:85–93.
23. Xing H, Ruan Y, Hsi JH, Kan W, Liao L, Leng X, *et al.* Reductions in virological failure and drug resistance in Chinese antiretroviral-treated patients due to lamivudine-based regimens, 2003-12. J Antimicrob Chemother 2015;70:2097–103.
24. Liao L, Xing H, Shang H, Li J, Zhong P, Kang L, *et al.* The Prevalence of Transmitted Antiretroviral Drug Resistance in Treatment-Naive HIV-Infected Individuals in China. JAIDS J Acquir Immune Defic Syndr 2010;53:S10–4.
25. Cai WP, Chen XJ, Li HQ, Li TS, Liu ZF, Lu HZ, et al. Manual of the National Free Antiretroviral Treatment. 4rd ed. Beijing: People's Medical Publishing House; 2016. (May 22, 2023). Retrieved from https://cread.jd.com/read/startRead.action?bookId=30533861&readType=1.
26. USCDC. Detecting and Responding to HIV Transmission Clusters - a Guide for Health Departments. US: CDC; 2018:132.
27. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, *et al.* The early spread and epidemic ignition of HIV-1 in human populations. Science 2014;346:56–61.
28. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol 2009;5:e1000520.
29. Ratmann O. Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: a population-based study in Rakai, Uganda. 2020;7.
30. Monod M, Brizzi A, Galiwango RM, Ssekubugu R, Chen Y, Xi X, *et al.* Longitudinal population-level HIV epidemiologic and genomic surveillance highlights growing gender disparity of HIV transmission in Uganda. Nat Microbiol 2024;9:35–54.
31. Sun X, Chen H, Liang S, Xiao T, Zeng Y, Liu H, *et al.* The Effectiveness of Antiretroviral Therapy in Mitigating New HIV Infections in Southwest China: An Ecological Study. Infect Drug Resist 2025;18:2943–50.

32. Ye J. Ye J, Lan Y, Wang J, Feng Y, Lin Y, Zhou Y, et al. Improvement in the 95-95-95 Targets Is Accompanied by a Reduction in Both the Human Immunodeficiency Virus Transmission Rate and Incidence in China. J Infect Dis. 2024 Nov 15;230(5):1202-1214.
33. Chen X, Qin C, Chen R, Huang Y, Xu Y, Tang Q, *et al.* Epidemiological profile and molecular genetic characterization of HIV-1 among female sex workers and elderly male clients in Guangxi, China. Emerg Microbes Infect 10:384–95.
34. Wulifan JK. Female transactional sex workers' experiences and health-seeking behaviour in low-middle income countries: a scoping review. BMC Public Health 2024;24:2749.
35. Safari F, Hashempour A, Khodadad N, Alinazari M matin K, Akbarinia S, Tabatabaee SA. What leads to treatment failure among people living with HIV: first systematic review from the middle east and North Africa. Virol J 2025;22:344.
36. Zhang C. HIV vulnerability among older Chinese men: new perspectives. Lancet Glob Health 2024;12:e742.
37. Chen J, Chang Y, Wu Y, Tang H, Wu G, Sun J, *et al.* Trends, Age-Period-Cohort Effects, and Projections in the Incidence and Mortality of Human Immunodeficiency Virus/AIDS Among the Elderly in China. J Infect Dis 2025;231:414–23.
38. Qin QQ, Cai C, Jin YC, Chen FF, Li DM, Tang HL. Analysis of reported HIV/AIDS patients aged 50 years old and above infected by commercial heterosexual sex in China from 2015 to 2022. Chin J AIDS STD, 2023,29(11):1197-1201.(In Chinese)
39. Chen L, His JH, Wu X, Shen Z, Lu H, Chen H, *et al.* Disparities in HIV and syphilis prevalence and risk factors between older male clients with and without steady sex partners in southwestern rural China. BMC Infect Dis 2017;17:269.
40. Arias Garcia S, Chen J, Calleja JG, Sabin K, Ogbuanu C, Lowrance D, *et al.* Availability and Quality of Surveillance and Survey Data on HIV Prevalence Among Sex Workers, Men Who Have Sex With Men, People Who Inject Drugs, and Transgender Women in Low- and Middle-Income Countries: Review of Available Data (2001-2017). JMIR Public Health Surveill 2020;6:e21688.
41. Sokurenko EV, Gomulkiewicz R, Dykhuizen DE. Source-sink dynamics of virulence evolution. Nat Rev Microbiol 2006;4:548–55.
42. Hendrickson ZM, De Jesus M, Barrington C, Cole SW, Kennedy CE, Sisson LN, *et al.* "You know that we travel a lot": Mobility narratives among female sex workers living with HIV in Tanzania and the Dominican Republic. Robinson J, ed. PLOS Glob Public Health 2024;4:e0003355.
43. Okano JT, Busang L, Seipone K, Valdano E, Blower S. The potential impact of country-level migration networks on HIV epidemics in sub-Saharan Africa: the case of Botswana. Lancet HIV 2021;8:e787–92.
44. Zhang J, Yao J, Jiang J, Pan X, Luo M, Xia Y, *et al.* Migration interacts with the local transmission of HIV in developed trade areas: A molecular transmission network analysis in China. Infect Genet Evol 2020;84:104376.
45. Adherence to the Continuum of Care | NIH. 2021. (January 8, 2026). Retrieved from https://clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-arv/adherence-continuum-care.



CRF01_AE Cluster 1

CRF01_AE Cluster 2

CRF07_BC

CRF08_BC

Figure 1. Temporal dynamics of lineage composition based on age-gender state frequencies for major HIV-1 lineages. F, female; M, male. Continuous temporal mapping of ancestral age-gender traits was performed using a Bayesian phylogeographic framework. State frequencies were estimated using discrete trait ancestral reconstruction, representing the posterior probability distribution of age-gender attributes across the reconstructed phylogenetic trees. The Y-axis represents the relative contribution of each age-gender group to the total Markov jump events at a given time point. Each panel represents a specific HIV-1 lineage (subtype/cluster), showing the proportional contribution of ten demographic groups (stratified by age and gender) to the regional epidemic pool over time. A notable expansion in the proportion of older males (50-69M) is observed in later years within CRF01_AE Cluster 2, CRF07_BC, and CRF08_BC.

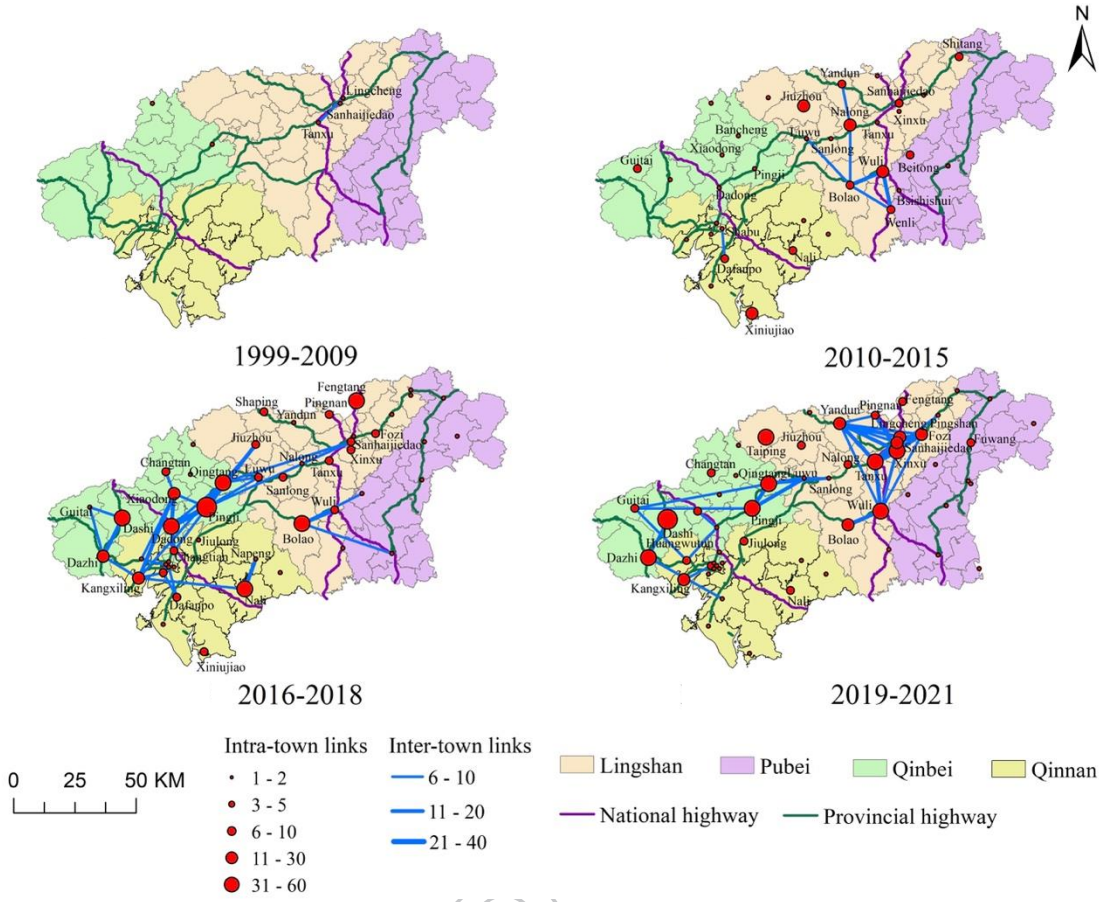


Figure 2 Spatiotemporal visualization of HIV transmission networks in Qinzhou (1999-2021). Township nodes (red circles) are sized by intra-link volume; connecting lines (blue) represent inter-link strength, weighted by thickness. Transportation routes shown as green (provincial highways) and dark purple (national highways) lines. Administrative regions (districts or counties) color-coded: Lingshan (beige), Pubei (lavender), Qinbei (light green), Qinnan (mustard).

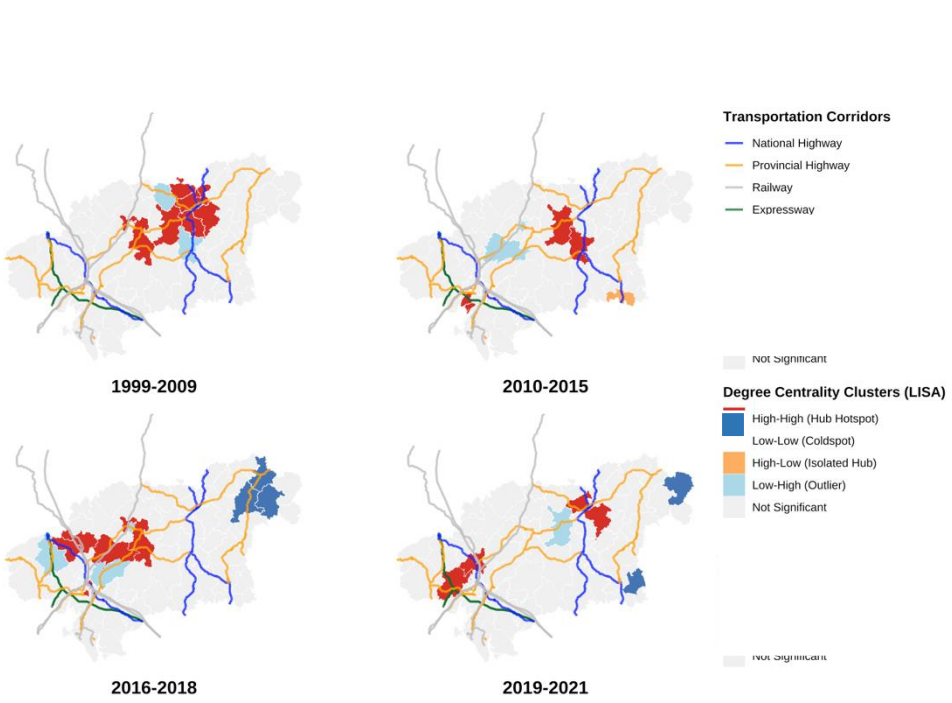


Figure 3-1 Spatiotemporal local indicators of spatial association analysis of association between HIV genetic network degree centrality and transportation corridors (1999-2021)

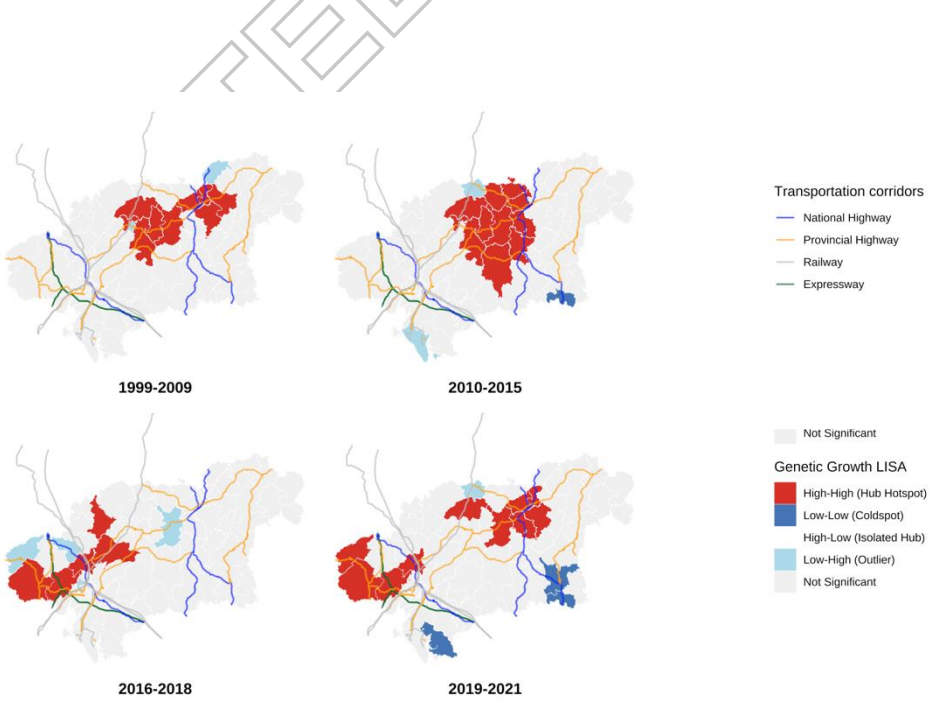


Figure 3-2 Spatiotemporal local indicators of spatial association analysis of association between HIV genetic cluster growth and transportation corridors (1999-2021)

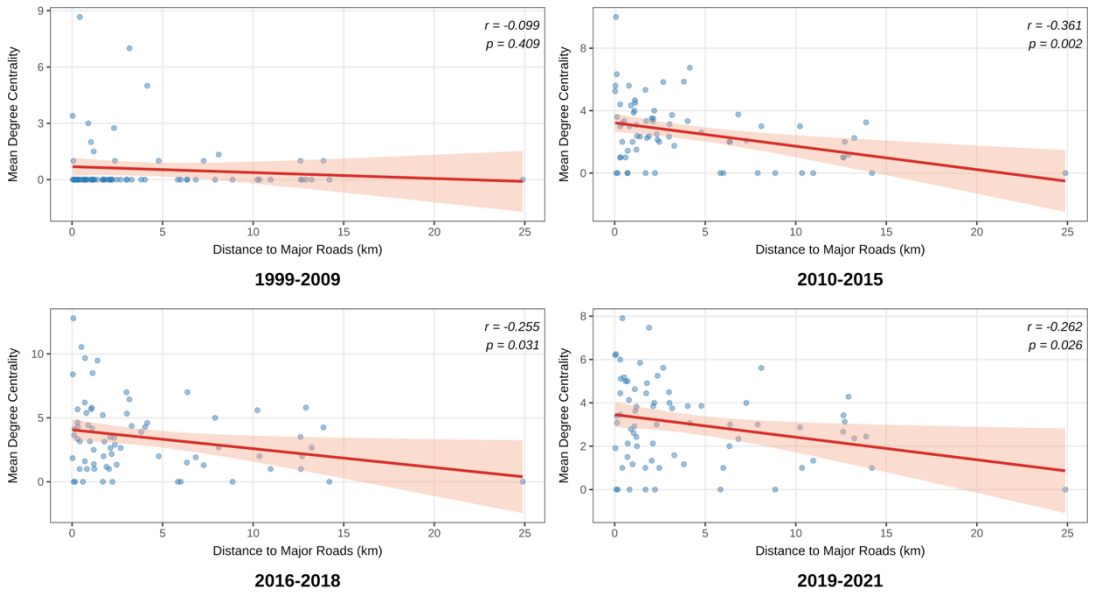


Figure 4-1 Correlation analysis of HIV genetic network degree centrality and geographic proximity to transportation network

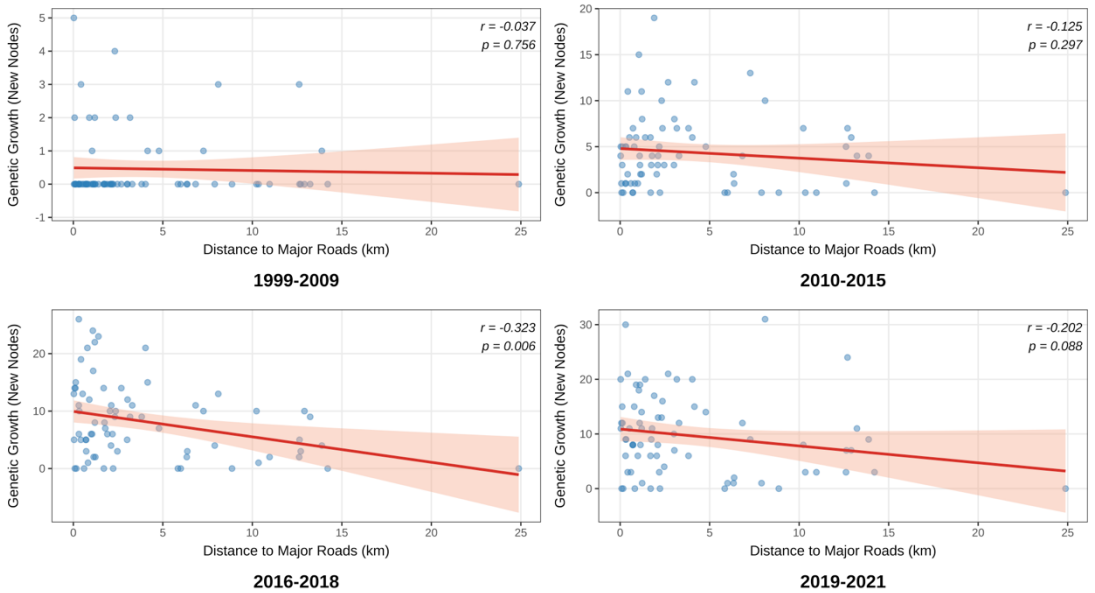


Figure 4-2 Correlation analysis of HIV genetic network growth and geographic proximity to transportation network

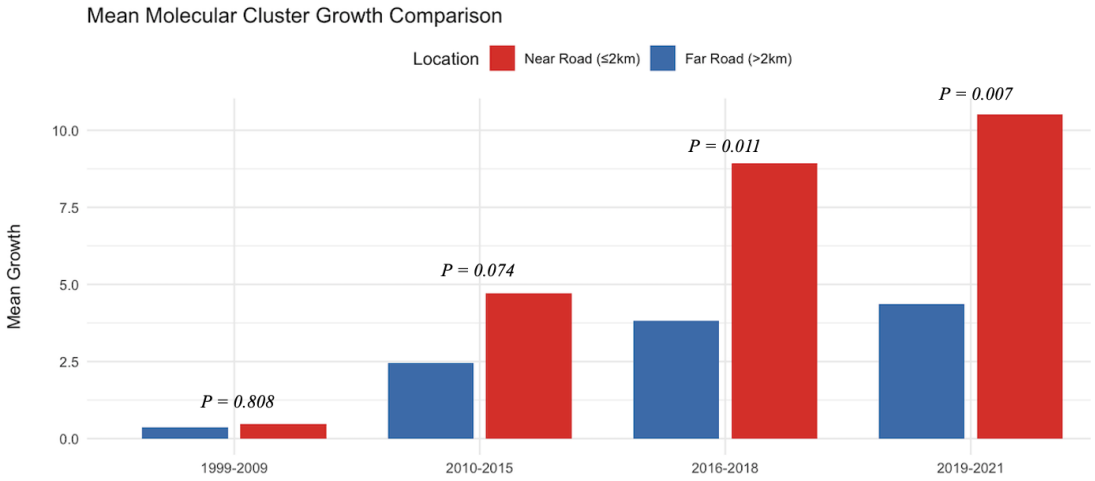


Figure 4-3 Comparison of mean HIV molecular cluster growth between townships within and outside the 2-km transportation buffer zones

Table 1 Basic social-demographic characteristics of all participants (N=5,094)

	Total	1999-2009	2010-2015	2016	2017	2018	2019	2020	2021	X^2 (Overall P)	P_{trend}
Total	5094 (100.0)	288 (100.0)	1139 (100.0)	698 (100.0)	597 (100.0)	602 (100.0)	674 (100.0)	525 (100.0)	571 (100.0)		
Age (Years)										401.16 (< 0.001)	
18-24	184 (3.6)	2 (0.7)	52 (4.6)	23 (3.3)	31 (5.2)	24 (4.0)	20 (3.0)	14 (2.7)	18 (3.2)		0.816
25-34	725 (14.2)	95 (33.0)	230 (20.2)	105 (15.0)	75 (12.6)	72 (12.0)	68 (10.1)	43 (8.2)	37 (6.5)		**< 0.001**
35-49	1491 (29.3)	151 (52.4)	366 (32.1)	206 (29.5)	176 (29.5)	147 (24.4)	179 (26.6)	131 (25.0)	135 (23.6)		**< 0.001**

50-69	2192 (43.0)	34 (11.8)	386 (33.9)	303 (43.4)	262 (43.9)	300 (49.8)	340 (50.4)	276 (52.6)	291 (51.0)		**< 0.001**
≥70	502 (9.9)	6 (2.1)	105 (9.2)	61 (8.7)	53 (8.9)	59 (9.8)	67 (9.9)	61 (11.6)	90 (15.8)		**< 0.001**
Gender										24.69 (< 0.001)	
Female	1296 (25.4)	47 (16.3)	269 (23.6)	198 (28.4)	154 (25.8)	153 (25.4)	200 (29.7)	127 (24.2)	148 (25.9)		**< 0.001**
Male	3798 (74.6)	241 (83.7)	870 (76.4)	500 (71.6)	443 (74.2)	449 (74.6)	474 (70.3)	398 (75.8)	423 (74.1)		**< 0.001**
Ethnicity										32.14 (< 0.001)	
Han	4564 (89.6)	269 (93.4)	1026 (90.1)	626 (89.7)	537 (89.9)	533 (88.5)	597 (88.6)	493 (93.9)	483 (84.6)		0.009
Other	530 (10.4)	19 (6.6)	113 (9.9)	72 (10.3)	60 (10.1)	69 (11.5)	77 (11.4)	32 (6.1)	88 (15.4)		0.009
Education										54.31 (< 0.001)	
Illiteracy	236 (4.6)	8 (2.8)	43 (3.8)	36 (5.2)	25 (4.2)	26 (4.3)	42 (6.2)	21 (4.0)	35 (6.1)		0.042
Elementary school	2602 (51.1)	115 (39.9)	574 (50.4)	351 (50.3)	311 (52.1)	300 (49.8)	350 (51.9)	299 (57.0)	302 (52.9)		**< 0.001**
Junior school	1830 (35.9)	147 (51.0)	428 (37.6)	248 (35.5)	206 (34.5)	226 (37.5)	228 (33.8)	168 (32.0)	179 (31.3)		**< 0.001**

High school and above	426 (8.4)	18 (6.3)	94 (8.3)	63 (9.0)	55 (9.2)	50 (8.3)	54 (8.0)	37 (7.0)	55 (9.6)		0.161
Marital status										138.83 (< 0.001)	
Single	1178 (23.1)	130 (45.1)	314 (27.6)	125 (17.9)	123 (20.6)	120 (19.9)	133 (19.7)	108 (20.6)	125 (21.9)		**< 0.001**
Married	2903 (57.0)	125 (43.4)	620 (54.4)	448 (64.2)	369 (61.8)	352 (58.5)	393 (58.3)	295 (56.2)	301 (52.7)		**< 0.001**
Divorced/Widowed	1013 (19.9)	33 (11.5)	205 (18.0)	125 (17.9)	105 (17.6)	130 (21.6)	148 (22.0)	122 (23.2)	145 (25.4)		**< 0.001**
Occupation										263.20 (< 0.001)	
Farmer	3822 (75.0)	119 (41.3)	827 (72.6)	554 (79.4)	445 (74.5)	476 (79.1)	532 (78.9)	418 (79.6)	451 (79.0)		**< 0.001**
House worker	653 (12.8)	76 (26.4)	158 (13.9)	67 (9.6)	71 (11.9)	68 (11.3)	76 (11.3)	72 (13.7)	65 (11.4)		**< 0.001**
Retired	91 (1.8)	2 (0.7)	26 (2.3)	17 (2.4)	15 (2.5)	10 (1.7)	12 (1.8)	4 (0.8)	5 (0.9)		0.887
Other	528 (10.4)	91 (31.6)	128 (11.2)	60 (8.6)	66 (11.1)	48 (8.0)	54 (8.0)	31 (5.9)	50 (8.8)		**< 0.001**
Risk										1288.24 (< 0.001)	
Injecting drug	419 (8.2)	181 (62.8)	123 (10.8)	26 (3.7)	18 (3.0)	29 (4.8)	16 (2.4)	14 (2.7)	12 (2.1)		**< 0.001**

Heterosexual	4581 (89.9)	107 (37.2)	1002 (88)	658 (94.3)	560 (93.8)	562 (93.4)	643 (95.4)	504 (96.0)	545 (95.4)		**< 0.001**
Homosexual	94 (1.8)	0 (0.0)	14 (1.2)	14 (2.0)	19 (3.2)	11 (1.8)	15 (2.2)	7 (1.3)	14 (2.5)		**< 0.001**
Residence										107.02 (< 0.001)	
Qinnan	1056 (20.7)	39 (13.5)	261 (22.9)	145 (20.8)	128 (21.4)	123 (20.4)	156 (23.1)	79 (15.0)	125 1.9)		> 0.050
Qinbei	1083 (21.3)	61 (21.2)	227 (19.9)	163 (23.4)	143 (24.0)	123 (20.4)	152 (22.6)	81 (15.4)	133 (23.3)		> 0.050
Lingshan	2283 (44.8)	168 (58.3)	512 (45.0)	273 (39.1)	234 (39.2)	284 (47.2)	279 (41.4)	300 (57.1)	233 (40.8)		0.020
Pubei	635 (12.5)	20 (6.9)	126 (11.1)	111 (15.9)	87 (14.6)	71 (11.8)	82 (12.2)	64 (12.2)	74 (13.0)		**< 0.001**
Kaifaqu	37 (0.7)	0 (0.0)	13 (1.1)	6 (0.9)	5 (0.8)	1 (0.2)	5 (0.7)	1 (0.2)	6 (1.1)		> 0.050
Subtype/Cluster										450.48 (< 0.001)	
CRF01_AE Cluster 1	549 (10.8)	23 (8.0)	154 (13.5)	81 (11.6)	73 (12.2)	53 (8.8)	66 (9.8)	46 (8.8)	53 (9.3)		**0.001**
CRF01_AE Cluster 2	1858 (36.5)	70 (24.3)	511 (44.9)	296 (42.4)	239 (40.0)	227 (37.7)	241 (35.8)	147 (28.0)	127 (22.2)		**< 0.001**
CRF07_BC	516 (10.1)	13 (4.5)	71 (6.2)	57 (8.2)	73 (12.2)	52 (8.6)	86 (12.8)	79 (15.0)	85 (14.9)		**< 0.001**

CRF08_BC	1619 (31.8)	148 (51.4)	305 (26.8)	187 (26.8)	149 (25.0)	197 (32.7)	217 (32.2)	192 (36.6)	224 (39.2)		**< 0.001**
CRF55_01B	55 (1.1)	0 (0.0)	4 (0.4)	6 (0.9)	10 (1.7)	16 (2.7)	4 (0.6)	4 (0.8)	11 (1.9)		**< 0.001**
Others	497 (9.8)	34 (11.8)	94 (8.3)	71 (10.2)	53 (8.9)	57 (9.5)	60 (8.9)	57 (10.9)	71 (12.4)		0.040

Note: X^2 (Overall P) represents the results of Pearson's chi-square test for overall distribution across study periods.

P_{trend} denotes the significance of linear trends for specific categories, calculated via the Cochran-Armitage trend test.

Table 2 Factors associated with HIV molecular linkage between baseline cases (1999–2016) and newly diagnosed cases (2017–2021): A Generalized Estimating Equation analysis (N=5,094; Observations=19,661)

Variable	Observations	≥ 1 (%)	1 (%)	≥ 2 (%)	OR (95%CI)	P value	AOR (95%CI)	P value
Total	19661	3463 (17.6)	1567 (8.0)	1896 (9.6)				
ART status								
ART-naive	6292	1297 (20.6)	577 (9.2)	720 (11.4)	Ref.		Ref.	
Active ART	12377	1984 (16.0)	921 (7.4)	1063 (8.6)	0.74(0.68-0.80)	< 0.001	0.79(0.73-0.87)	< 0.001
ART dropout	992	182 (18.3)	69 (7.0)	113 (11.4)	0.87(0.73-1.03)	0.132	0.90(0.75-1.08)	0.274
Age (years)								
< 25	742	63 (8.5)	39 (5.3)	24 (3.2)	Ref.		Ref.	
25-34	3140	220 (7.0)	134 (4.3)	86 (2.7)	0.81 (0.61-1.09)	0.163	0.89(0.66-1.21)	0.465
35-49	6017	694 (11.5)	340 (5.7)	354 (5.9)	1.41 (1.07-1.84)	0.014	1.60(1.20-2.14)	0.001
50-69	7988	1930 (24.2)	822 (10.3)	1108 (13.9)	3.43 (2.64-4.47)	< 0.001	3.59(2.68-4.81)	< 0.001
⩾ 70	1774	556 (31.3)	232 (13.1)	324 (18.3)	4.92 (3.73-6.49)	< 0.001	4.22(3.09-5.77)	< 0.001
Gender								
Female	4954	730 (14.7)	356 (7.2)	374 (7.5)	Ref.		Ref.	
Male	14707	2733 (18.6)	1211 (8.2)	1522 (10.3)	1.32 (1.21-1.44)	< 0.001	1.42(1.29-1.56)	< 0.001

Education								
Illiteracy	867	229 (26.4)	89 (10.3)	140 (16.1)	Ref.		Ref.	
Elementary school	7248	1099 (15.2)	508 (7.0)	591 (8.2)	0.66 (0.56-0.77)	< 0.001	0.85(0.72-1.01)	0.072
Junior school	9905	1897 (19.2)	859 (8.7)	1038 (10.5)	0.50 (0.42-0.59)	< 0.001	0.81(0.68-0.98)	0.027
High school and above	1641	238 (14.5)	111 (6.8)	127 (7.7)	0.47 (0.39-0.58)	< 0.001	0.67(0.53-0.84)	0.001
Marital status								
Single	4680	578 (12.4)	302 (6.5)	276 (5.9)	Ref.		Ref.	
Married	11288	2145 (19.0)	951 (8.4)	1194 (10.6)	1.66 (1.51-1.84)	< 0.001	0.97(0.86-1.09)	0.614
Divorced/Widowed	3693	740 (20.0)	314 (8.5)	426 (11.5)	1.78 (1.58-2.00)	< 0.001	0.90(0.78-1.03)	0.131
Ethnicity								
Han	17682	3028 (17.1)	1397 (7.9)	1631 (9.2)	Ref.		Ref.	
Other	1979	435 (22.0)	170 (8.6)	265 (13.4)	1.36(1.22-1.53)	< 0.001	1.09(0.96-1.25)	0.179
Occupation								
Farmer	14512	2671 (18.4)	1212 (8.4)	1459 (10.1)	Ref.		Ref.	
House worker	2569	419 (16.3)	202 (7.9)	217 (8.4)	0.86 (0.77-0.97)	< 0.001	1.10(0.97-1.25)	0.141
Retired	389	118 (30.3)	43 (11.1)	75 (19.3)	1.93 (1.55-2.41)	< 0.001	1.18(0.93-1.51)	0.172
Other	2191	255 (11.6)	110 (5)	145 (6.6)	0.58 (0.51-0.67)	< 0.001	0.83(0.71-0.98)	0.023
Risk								
Injection drug use	1944	122 (6.3)	78 (4)	44 (2.3)	Ref.		Ref.	
Heterosexual	17365	3307 (19.0)	1469 (8.5)	1838 (10.6)	3.51 (2.91-4.24)	< 0.001	2.64(2.13-3.27)	< 0.001
Homosexual	352	34 (9.7)	20 (5.7)	14 (4.0)	1.60 (1.07-2.38)	0.021	2.25(1.45-3.48)	< 0.001
Location								
Qinnan	4108	864 (21.0)	370 (9.0)	494 (12.0)	Ref.		Ref.	

Qinbei	4213	969 (23.0)	368 (8.7)	601 (14.3)	1.12 (1.01-1.24)	0.030	1.17(1.05-1.31)	0.006
Lingshan	8741	1243 (14.2)	651 (7.4)	592 (6.8)	0.62 (0.57-0.69)	< 0.001	0.67(0.60-0.75)	< 0.001
Pubei	2452	353 (14.4)	162 (6.6)	191 (7.8)	0.63 (0.55-0.72)	< 0.001	0.65(0.57-0.75)	< 0.001
Kaifaqu	147	34 (23.1)	16 (10.9)	18 (12.2)	1.13 (0.76-1.67)	0.540	1.22(0.80-1.85)	0.355
Subtype								
CRF01_AE	9942	1575 (15.8)	790 (7.9)	785 (7.9)			Ref.	
CRF07_BC	1779	439 (24.7)	152 (8.5)	287 (16.1)	1.74 (1.54-1.96)	< 0.001	1.94(1.71-2.21)	< 0.001
CRF08_BC	5992	1142 (19.1)	512 (8.5)	630 (10.5)	1.25 (1.15-1.36)	< 0.001	1.75(1.59-1.92)	< 0.001
Other	1948	307 (15.8)	113 (5.8)	194 (10.0)	0.99 (0.87-1.14)	0.928	1.42(1.23-1.63)	< 0.001
Baseline CD4 level								
< 350	14632	12088 (61.5)	1170 (6.0)	1374 (7.0)				
350-499	2931	2378 (12.1)	232 (1.2)	321 (1.6)	1.10 (1.00-1.22)	0.055	1.23(1.10-1.37)	< 0.001
≥ 500	1900	1581 (8.0)	141 (0.7)	178 (0.9)	0.96 (0.84-1.09)	0.520	1.26(1.10-1.45)	0.001
Missing	198	151 (0.8)	24 (0.1)	23 (0.1)	1.48 (1.06-2.06)	0.020	1.15(0.79-1.68)	0.468

Note: * ART refers to Antiretroviral Therapy.

Table 3 HIV genetic network metrics across four distinct epidemiological phases (N=5,094)

Phase	No. of nodes	No. of edges	Network density	Average degree	Average degree centrality	Average betweenness	Average eigenvector centrality	Maximum degree centrality
1999-2009	34	51	0.0909	3.00	0.0909	0.032475	0.2303	0.5455
2010-2015	328	548	0.0102	3.34	0.0102	0.000152	0.0452	0.0550
2016-2018	627	1321	0.0067	4.21	0.0067	0.000033	0.0300	0.0431

2019-2021	729	1408	0.0053	3.86	0.0053	0.000017	0.0347	0.0467

Phase	Maximum betweenness	Maximum eigenvector centrality	No. of super spreaders	Total Edge Count of Super spreaders	Proportion of Super spreaders	Number of clusters	Maximum cluster size	Average Cluster Size	Edge Proportion of Super spreaders
1999-2009	0.4054	1	1	18	0.0294	4	28	8.5	0.35294118
2010-2015	0.0070	1	26	319	0.0793	69	37	4.75	0.58211679
2016-2018	0.0023	1	41	764	0.0654	148	51	4.24	0.57834974
2019-2021	0.0017	1	39	868	0.0535	175	36	4.17	0.61647727

Table 4 Spatiotemporal profiles and epidemiological characteristics of molecular super-spreaders across four distinct phases (1999–2021) (N=5,094)

Network Indicators	Phase 1	Phase 2	Phase 3	Phase 4
	1999-2009	2010-2015	2016-2018	2019-2021
Number of super spreaders	1	26	41	39
Male (n, %)	1 (100.0)	23 (88.5)	36 (87.8)	31 (79.5)
Mean Degree Centrality	18.0	12.3	18.6	22.3
Maximum degree Centrality	18	18	27	34
Demographics				
Age	< 25 (100.0%)	50–69 (50.0%); 35-49 (26.9%); ≥70: (19.2%); 25-34: (3.8%)	50-69 (65.9%); ≥70 (17.1%); 35-49 (17.1%)	50-69 (59.0%); ≥70 (35.9%); 35-49 (5.1%)

Occupation	Farmer (100.0%)	Farmer (80.8%); House worker or business (15.4%); Other (3.8%)	Farmer (85.4%); Retired (9.8%); House worker or business (4.9%)	Farmer (82.1%); House worker or business (15.4%); Retired (2.6%)
Marital status	Single (100.0%)	Married (53.8%); Divorced or widowed (26.9%); Single (19.2%)	Married (73.2%); Divorced or widowed (19.5%); Single (7.3%)	Married (59%); Divorced or widowed (30.8%); Single (10.3%)
Risk Factors				
Transmission route	IDU (100%)	Commercial (42.3%); Non-marital heterosexual (23.1%); IDU: 15.4%; Non-marital non-commercial (15.4%); Spouse positive (3.8%)	Commercial (82.9%); Spouse positive (9.8%); Non-marital non-commercia (7.3%)	Commercial (71.8%); Non-marital non-commercia (23.1%); Spouse positive (5.1%)
ART* and Virological Profile				
Active ART	0 (0.0%)	12 (46.2%)	6 (14.6%)	7 (17.9%)
ART dropout	1 (100.0%)	5 (19.2%)	10 (24.4%)	11 (28.2%)
Active ART but				
1. Viral load ≥ 200 copies/mL	0 (0.0%)	0 (0.0%)	4 (9.8%)	1 (2.6%)
2. Viral load missing	1 (100.0%)	0 (0.0%)	6 (14.6%)	14 (35.9%)
Dominant HIV subtype	CRF08_BC (100%)	Other (69.2%); CRF01_AE_C2 (26.9%); CRF08_BC (3.8%)	CRF08_BC (43.9%); CRF07_BC (19.5%); Other (19.5%); CRF01_AE_C2 (17.1%)	CRF07_BC (74.4%); CRF08_BC (23.1%); CRF01_AE_C2 (2.6%)
Geographic Focus				

Key Townships	TX: 100%	BL: 12%; JS: 8%; KXL: 8%; WL1: 8%; JZ: 8%; WL2: 8%; NL: 8%; SH: 4%; DZ: 4%; WF: 4%; XX: 4%; TX: 4%; LC: 4%; CT1: 4%; LW: 4%; QT: 4%; HWT: 4%	PJ: 19.5%; QT: 17.1%; NP: 4.9%; LW: 4.9%; SH: 2.4%; SL: 2.4%; DC: 2.4%; BT: 2.4%; XY: 2.4%; DD: 2.4%; DS: 2.4%; DZ: 2.4%; ZC: 2.4%; AS: 2.4%; XD: 2.4%; JS: 2.4%; KXL: 2.4%; WL1: 2.4%; JZ: 2.4%; TX: 2.4%; SB: 2.4%; GT: 2.4%; NL: 2.4%; HT: 2.4%; HWT: 2.4%; LM: 2.4%	TX: 12.8%; XX: 10.3%; LC: 10.3%; YD: 10.3%; FZ: 7.7%; SH: 5.1%; PJ: 5.1%; WL2: 5.1%; CT2: 5.1%; DD: 2.6%; DS: 2.6%; ZC: 2.6%; XJ: 2.6%; PN: 2.6%; PS: 2.6%; XNJ 2.6%; FW: 2.6%; NL: 2.6%; QT: 2.6%; HWT: 2.6%

Note: Super spreader was defined as its degree centrality exceeded the threshold of two standard deviations above the mean ($\mu + 2\sigma$) for each respective phase.

*ART denotes antiretroviral therapy.

Supplementary file 1

Table S1 Distribution of HIV subtypes/clusters among study samples with RNA sequence in Qinzhou (1999-2021) (N=4,814)

Items	Numbers	%
Total	4814	100.0
HIV subtypes		
CRF01_AE	2505	52.1
cluster 1	*508*	*10.6*
cluster 2	*1770*	*36.7*
cluster 3	*14*	*0.3*
cluster 4	*13*	*0.3*
cluster 5	*7*	*0.2*
Other CRF01_AE clusters	*193*	*4.0*
CRF07_BC	491	10.2
CRF08_BC	1527	31.7
CRF55_01B	49	1.0
B	18	0.4
Other CRFs or URFs	224	4.6

Note: CRF refers to circulating recombinant form, URF refers to Unique recombinant form.

Table S2 Comparison of baseline characteristics between sampled and overall study participants for CRF01_AE cluster 2 (N=1,770) and CRF08_BC (N=1,527) Subtypes in Qinzhou

	CRF01_AE cluster 2				CRF_08BC		
	All sequences (N, %)	Samples (N, %)	*P* value		All sequences (N, %)	Samples (N, %)	*P* value
Total	1770 (100.0)	500 (100.0)			1527 (100.0)	500 (100.0)	
Age (years)							
18-24	29 (1.6)	25 (5.0)	1.000		38 (2.5)	29 (5.8)	1.000
25-34	213 (12.0)	75 (15.0)	1.000		208 (13.6)	69 (13.8)	0.998
35-49	447 (25.3)	128 (25.6)	0.992		549 (35.9)	185 (37)	0.992
50-69	856 (48.4)	203 (40.6)	0.134		594 (38.9)	169 (33.8)	0.477
≥70	225 (12.7)	69 (13.8)	0.999		138 (9)	48 (9.6)	0.999
Gender							
Male	1329 (75.1)	369 (73.8)	0.953		1121 (73.4)	368 (73.6)	0.989
Female	441 (24.9)	131 (26.2)	0.997		406 (26.6)	132 (26.4)	0.982
Education							
Illiteracy	100 (5.6)	22 (4.4)	0.999		78 (5.1)	25 (5.0)	1.000
Elementary school	930 (52.5)	247 (49.4)	0.768		812 (53.2)	269 (53.8)	0.985
Junior school	617 (34.9)	203 (40.6)	1.000		548 (35.9)	182 (36.4)	0.987
High school	113 (6.9)	28 (5.6)	0.999		89 (5.8)	24 (4.8)	0.999
Marital status							
Single	319 (18.0)	110 (22.0)	1.000		392 (25.7)	121 (24.2)	0.945
Married	1062 (60)	297 (59.4)	0.961		834 (54.6)	285 (57.0)	0.998
Divorced/Widowed	389 (22)	93 (18.6)	0.791		301 (19.7)	94 (18.8)	0.978
Occupation							
Farmer	1329 (75.1)	355 (71)	0.656		1174 (76.9)	376 (75.2)	0.932
Other	441 (24.9)	145 (29)	1.000		353 (23.1)	124 (24.8)	0.998
Risk							
Injecting drug use	39 (2.2)	39 (7.8)	1.000		265 (17.3)	96 (19.2)	0.999
Heterosexual	1727 (97.6)	457 (91.4)	0.185		1262 (82.7)	404 (80.8)	0.941
Homosexual	4 (0.2)	4 (0.8)	1.000		0 (0.0)	0 (0.0)	-
Residence							
Qinnan	602 (34)	160 (32)	0.897		114 (7.5)	38 (7.6)	0.999
Qinbei	476 (26.9)	146 (29.2)	0.999		238 (15.6)	74 (14.8)	0.989
Lingshan	529 (29.9)	158 (31.6)	0.998		1032 (67.6)	348 (69.6)	0.998
Pubei	145 (8.2)	34 (6.8)	0.997		82 (5.4)	25 (5.0)	1.0000
Other	18 (1.0)	2 (0.4)	1.000		3 (0.2)	0 (0.0)	-

Note: Sequences from CRF01_AE cluster 2 and CRF08_BC subtypes were derived from the 4814 - RNA sequence dataset.

Table S3 Distribution of HIV-1 sequences used in the final Bayesian analysis for five HIV-1 subtypes/clusters

HIV-1 subtype	No. of original study samples*	No. of study samples for Bayesian analysis (Post-subsampling)*	No. of historical reference sequences from Guangxi # (Year of detection)	Total sample size for Bayesian analysis
Total	4345	2048	61	2109
CRF01_AE cluster 1	508	508	10 (2005-2007)	518
CRF01_AE cluster 2	1770	500 @	27 (1997-2009)	527
CRF07_BC	491	491	12 (2009-2010)	503
CRF08_BC	1527	500 @	8 (1997-2009)	508
CRF55_01B	49	49	4 (2009-2013)	53

Note: * The study samples were derived from the 4814-RNA sequence dataset.

Historical reference sequences from Guangxi (n=61, sampled 1997-2013) were retrieved from the provincial HIV surveillance database of Guangxi CDC as temporal anchors for MRCA (most recent common ancestor) calibration.

@ For subtype/clusters with more than 1000 original sequences, a stratified random subsampling strategy was applied to select 500 representative sequences to ensure Bayesian computational feasibility while preserving demographic representativeness.

Table S4 Comparison of demographic characteristics between all newly diagnosed HIV/AIDS cases and samples with RNA sequences in Qinzhou (1999-2021)

Variables	Newly diagnosed cases N (%)	Samples with RNA sequence N (%)	*P* -value
Total	7381 (100.0)	4814 (100.0)	
Age (years)			
18-24	408 (5.5)	178 (3.7)	1.000
25-34	1056 (14.3)	697 (14.5)	1.000
35-49	2112 (28.6)	1389 (28.9)	1.000
50-69	3133 (42.4)	2059(42.8)	1.000
≥70	672 (9.1)	491 (10.2)	1.000
Gender			
Male	5542 (75.1)	3588 (74.5)	1.000
Female	1839 (24.9)	1226 (25.5)	1.000

Education			
Illiteracy	360 (4.9)	232 (4.8)	1.000
Elementary school	3791 (51.4)	2451 (50.9)	1.000
Junior school	2534 (34.3)	1722 (35.8)	1.000
High school and above	696 (9.4)	409 (8.5)	1.000
Marital status			
Single	1831 (24.8)	1126 (23.4)	1.000
Married	4087 (55.4)	2727 (56.7)	1.000
Divorced/Widowed	1463 (19.8)	961 (20.0)	1.000
Occupation			
Farmer	5521 (74.8)	3602 (74.8)	1.000
Other	1860 (25.2)	1212 (25.2)	1.000
Risk			
Injecting drug	352 (4.8)	407 (8.5)	0.997
Heterosexual	6675 (90.4)	4320 (89.7)	1.000
Homosexual	222 (3.0)	87 (1.8)	1.000
Unkown	132 (1.8)	0 (0.0)	-
Residence			
Qinnan	1156 (15.7)	1037 (21.5)	0.117
Qinbei	1402 (19.0)	1060 (22.0)	0.996
Lingshan	3531 (47.8)	2129 (44.2)	1.000
Pubei	1238 (16.8)	551 (11.5)	1.000
Other	54 (0.7)	37 (0.8)	1.000

Table S5 The time to most recent common ancestor for main HIV subtypes and clusters in Qinzhou

HIV subtypes/clusters	Guangxi historical reference sequences @			Study samples+		tMRCA*- 95%HPD#
	Year of detection	No. of samples		Year of detection	No. of samples	
Total	-	61		-	2048	-
CRF01_AE cluster 1	2005-2007	10		2010-2020	508	1997.9 (1996.1- 1999.9)
CRF01_AE cluster 2	1997-2009	27		2010-2020	500	1993.6 (1991.9- 1997.2)
CRF07_BC	2009-2010	12		2011-2020	491	2000.4 (1993.2- 2002.7)
CRF08_BC	1997-2009	8		2011-2020	500	1995.4 (1989.4- 1996.7)
CRF55_01B	2009-2013	4		2014-2020	49	2001.1 (1995.4- 2002.8)

Note: * tMRCA denotes time to most recent common ancestor.

HPD denotes highest posterior density.

@ The Guangxi historical reference sequences (n=61; sampled 1997-2013) were retrieved from Guangxi CDC institutional database to serve as tMRCA calibration.

+ The study samples were derived from 4,814-RNA sequence dataset.

Table S6 Bayesian Stochastic Search Variable Selection (BSSVS) results of main HIV subtypes/clusters transmission relationships between different age-gender groups in Qinzhou

		CRF01_AE cluster 1		CRF01_AE cluster 2		CRF07_BC		CRF08_BC	
From	To	Mean counts	BF	Mean counts	BF	Mean counts	BF	Mean counts	BF
≥70M	18-24F	0	0.8	0	0	0	0.5	0	1.4
≥70M	25-34F	0	0.6	0	0	0	0.4	0	1.0
≥70M	35-49F	0	0.8	0	0	0	0.4	0	0.4
≥70M	50-69F	0	0.8	0	0	3	12.3	0	1.1
≥70M	≥70F	0	3.0	0	0	2	13.9	1	7.9
50-69M	18-24F	0	0.2	11	670332.1	0	0.2	0	0.3
50-69M	25-34F	0	0.3	0	0	5	9.8	0	0.9
50-69M	35-49F	11	1620	15	670332.1	14	536630.2	0	0.1
50-69M	50-69F	72	573158.4	52	670332.1	59	536630.2	39	282614.9
50-69M	≥70F	6	255.0	8	670332.1	11	336.6	3	13.9
35-49M	18-24F	0	0.4	0	0	1	29.8	25	282614.9
35-49M	25-34F	19	114625.0	26	670332.1	7	48.4	22	282614.9
35-49M	35-49F	34	573158.4	0	0	2	24.8	42	282614.9

35-49M	50-69F	2	1.5	0	0	0	0.3	1	1.4
35-49M	≥70F	0	0.2	0	0	0	0.2	0	0.6
25-34M	18-24F	7	4434.9	4	670332.1	0	1.1	1	4.7
25-34M	25-34F	4	440.2	0	0	0	0.3	0	0.3
25-34M	35-49F	0	1.3	0	0	0	0.2	0	0.2
25-34M	50-69F	0	0.3	0	0	0	0.7	0	0.3
25-34M	≥70F	0	0.2	0	0	0	0.1	0	0.4
18-24M	18-24F	0	0.6	0	0	6	90.6	0	0.8
18-24M	25-34F	0	0.5	0	0	0	0.1	0	0.6

		CRF01_AE cluster 1		CRF01_AE cluster 2		CRF07_BC		CRF08_BC	
From	To	Mean counts	BF	Mean counts	BF	Mean counts	BF	Mean counts	BF
18-24M	35- 49F	0	0.5	0	0	0	0.4	0	0.6
18-24M	50-69F	0	0.5	0	0	0	0.6	0	0.5
18-24M	≥70F	0	0.5	0	0	0	0.1	0	0.5
≥70F	18-24M	0	1.1	0	0	0	1.2	0	1.1
≥70F	25-34M	0	1.3	0	0	0	1.2	0	1.0
≥70F	35-49M	0	1.4	0	0	0	1.4	0	1.5
≥70F	50-69M	0	1.1	0	0	0	1.3	0	1.3
≥70F	≥70M	0	1.3	0	0	0	1.3	0	0.8
50-69F	18-24M	0	1.9	0	0	0	0.5	0	0.7
50-69F	25-34M	0	1.1	0	0	0	0.7	0	0.8
50-	35-49M	0	0.8	0	0	0	1.1	0	1.4

69F									
50-69F	50-69M	1	2.4	0	0	0	2.2	0	1.0
50-69F	≥70M	0	0.6	3	670332.1	2	4.5	0	1.3
35-49F	18-24M	0	3.4	1	670332.1	0	0.6	0	0.9
35-49F	25-34M	0	1.7	0	0	0	0.7	0	0.6
35-49F	35-49M	0	0.8	0	0	0	1.7	0	1.2
35-49F	50-69M	0	1.0	0	0	0	0.7	0	1.1
35-49F	≥70M	0	0.6	0	0	0	0.5	0	0.5
25-34F	18-24M	0	0.7	1	670332.1	0	0.6	0	0.6
25-34F	25-34M	0	0.9	0	0	0	0.6	0	0.6
25-34F	35-49M	0	1.1	0	0	0	1.1	0	0.6
25-34F	50-69M	0	0.6	0	0	0	0.6	0	0.6
25-34F	≥70M	0	0.4	0	0	0	0.4	0	0.5

		CRF01_AE cluster 1		CRF01_AE cluster 2		CRF07_BC		CRF08_BC	
From	To	Mean counts	BF	Mean counts	BF	Mean counts	BF	Mean counts	BF
18-24F	18-24M	0	0.8	2	670332.1	0	0.5	0	4.1
18-24F	25-34M	0	0.5	0	0	0	0.5	0	2.1
18-24F	35-49M	0	0.4	0	0	0	0.5	0	0.4
18-24F	50-69M	0	0.4	0	0	0	0.5	0	0.3

18-24F	≥70M	0	0.3	0	0	0	0.4	0	0.4

Note: BF refers to Bayes Factor; All transmission events with decisive support (BF > 100) were associated with a posterior probability ≥ 0.9. F, female; M, Male;
This table presents the complete output of the BSSVS analysis; results with low BF values should be interpreted with caution as they indicate weaker statistical evidence for viral migration. All samples were derived from the 4814 RNA dataset.

Table S7 Global Moran's I analysis of Mean Degree Centrality and Mean Genetic Cluster Growth across four distinct epidemiological phases

Phase	Mean Degree Centrality			Mean Genetic Cluster Growth		
	Global Moran I	z value	p value	Global Moran I	z value	p value
1999-2009	0.298	4.316	0.000	0.314	4.235	0.000
2010-2015	0.184	2.460	0.007	0.223	2.950	0.002
2016-2018	0.208	2.753	0.003	0.205	2.699	0.003
2019-2021	0.115	1.584	0.057	0.373	4.765	0.000

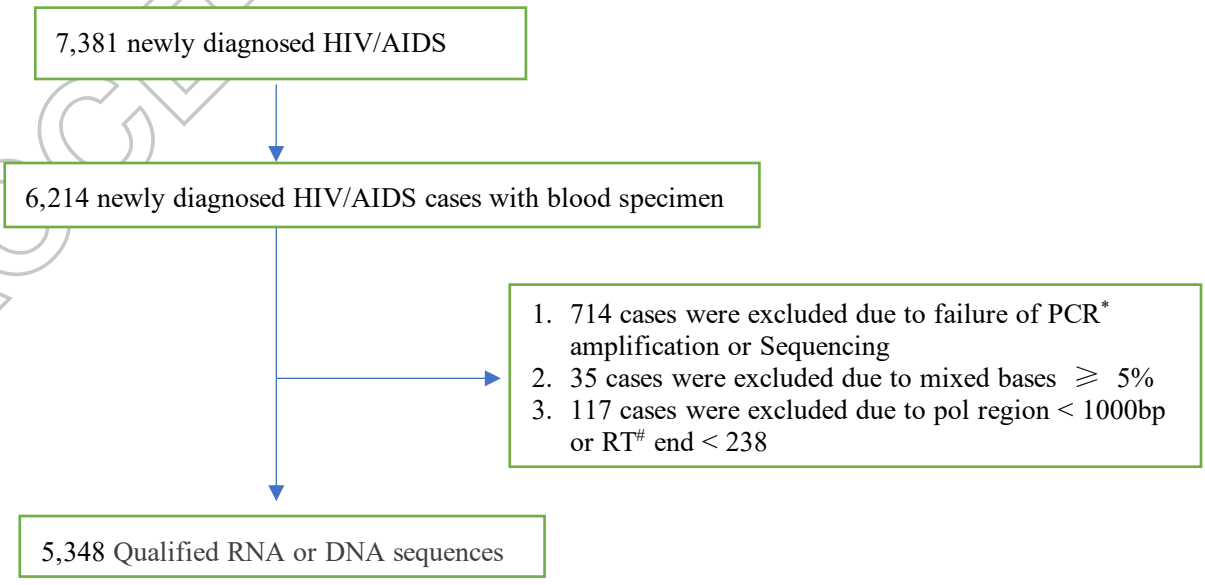
Note: Samples were derived from the overall study samples (N=5,094)

Table S8 Questions and analysis of respondent driven sampling based behavioral survey

Variable	Number	%
Total	84	100.0
Q1. Age (years)		-
< 50	7	8.3
50-69	24	28.6
⩾70	53	63.1
Q2. Gender		
Male	82	97.6
Q3. Marital status		
Single	15	17.9
Married	43	51.2
Widow/divorced	26	30.9
Q4. Where are your social & recreational venues ?(Multiple-choice question)		
Informal outdoor social gatherings (e.g., card playing at village entrances)	63	75.0
Community Park	47	56.0

Public little square	46	54.8
Traditional periodic markets	40	47.6
Street-side vending areas	29	34.5
Private residences of friends	13	15.5
Indoor recreational centers (Mahjong)	11	13.1
Indoor recreational centers (Chess parlors)	7	8.3
Q5. What is your primary source of income ? (Multiple-choice question)		
Pension	36	42.9
Labor earnings	43	51.2
Children's support	41	48.8
Minimum living guarantee subsidies	8	9.5
Q6. Have you ever engaged in casual or commercial heterosexual sex ?	41	48.8
Q7. Have you ever engaged in casual heterosexual sex ?		
Yes	20	23.8
No	64	76.2
Q8. What is your number of casual heterosexual partners? [Median (IQR)]	1.0 (1.0, 2.0)	-
Q9. Have you ever visited commercial sex venues ?		
Yes	38	45.2
No	46	54.8
Q10. Have you ever engaged in commercial heterosexual sex ?		
Yes	36	42.9
No	48	57.1
Q11. What about the locations of your commercial sex encounter ? (n=36)		
Within township/street 2-km buffer of transportation arteries	32	88.9
Beyond township/street 2-km buffer of transportation arteries	4	11.1
Q12. What is the number of your commercial heterosexual partners? [Median (IQR)]	2.0 (2.0, 4.0)	-
Q13. What is your frequency of visiting commercial or casual heterosexual partners ? (n=41)		
≥ Once weekly	1	2.4
> Once monthly to ≤ once weekly	10	24.4

> Once quarterly to ≤ once monthly	7	17.1
> Semi-annually to ≤ once quarterly	7	17.1
> Annually to ≤ semi-annually	12	29.3
Once only	4	9.7
Q14. How about your condom use with commercial or casual heterosexual partners ? (n=41)		
Never	13	31.7
Occasionally	13	31.7
Frequently	3	7.3
Consistently	12	29.3
Q15. Are you aware of the typical price of commercial sex in your local venues? (CNY/transaction) (n=48)		
< 50	18	37.5
50-100	17	35.4
> 100	13	27.1
Q16. What is the geographic origin of commercial sex workers you encountered (n=36)		
Local (Within Qinzhou city)	24	66.7
Non-local (Within Guangxi, but outside Qinzhou)	9	25.0
Out-of-province (Outside Guangxi)	3	8.3
Q17. Were the commercial sex workers you encountered based permanently in a specific township's venues? (n=36)		
No. Non-fixed (Rotating among townships along transportation arteries, synchronized with market days)	26	72.2
Yes. Fixed (Permanently based within a single township).	10	27.8



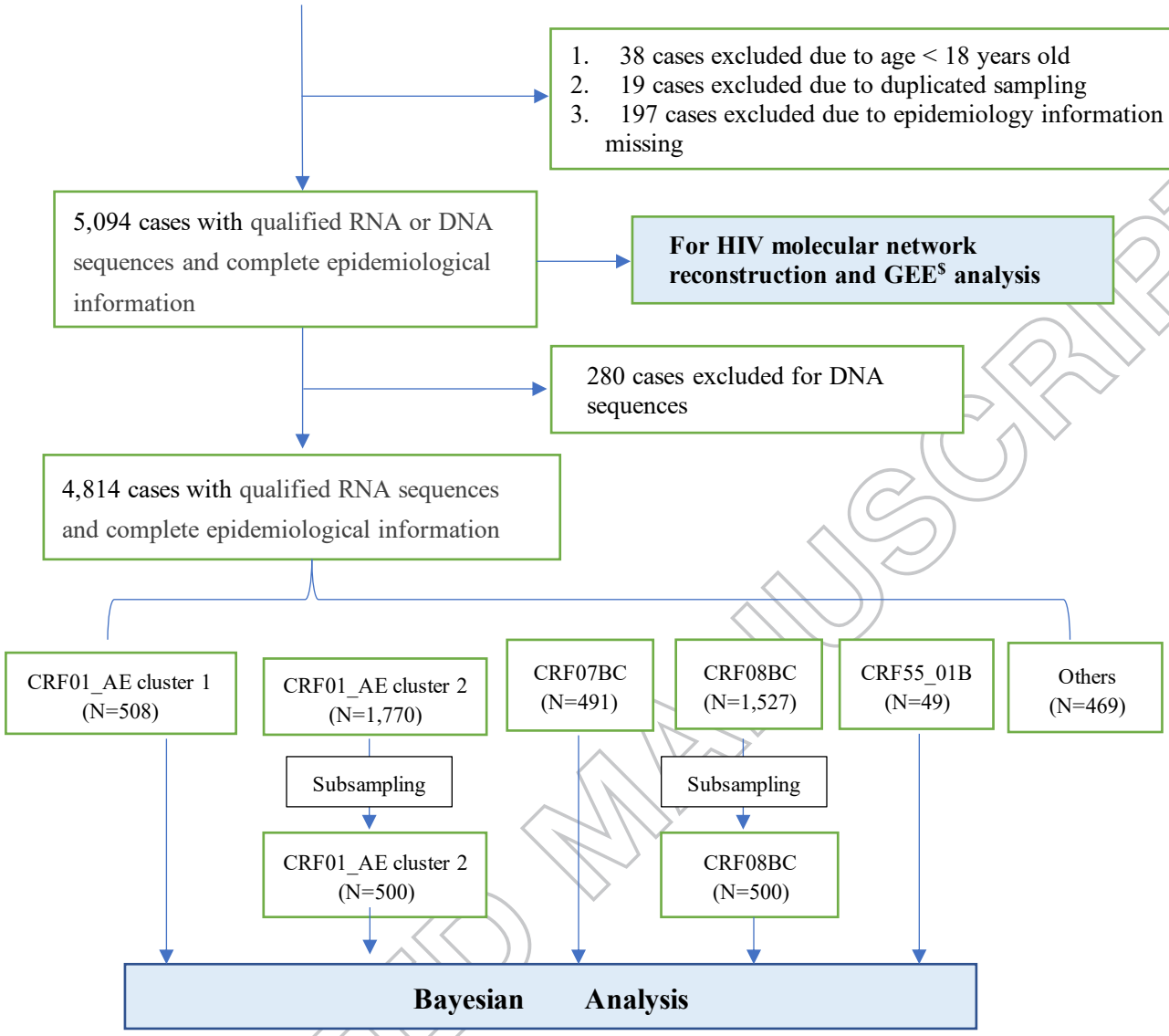


Figure S1 Flowchart of study samples inclusion and exclusion

Note: * PCR denotes Polymerase Chain Reaction. # RT denotes Reverse Transcriptase. $ GEE refers to Generalized Estimating Equations.

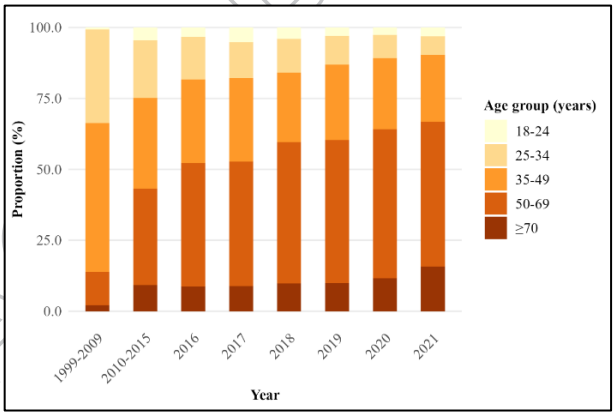


Figure S 2-1

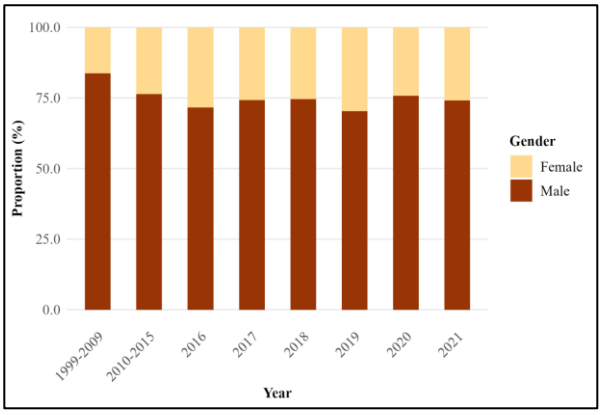


Figure S 2-2

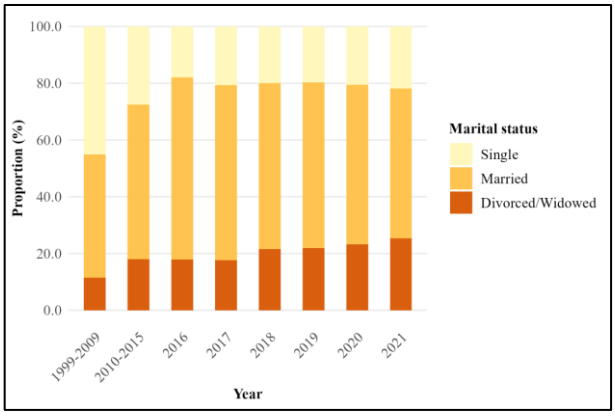


Figure S 2-3

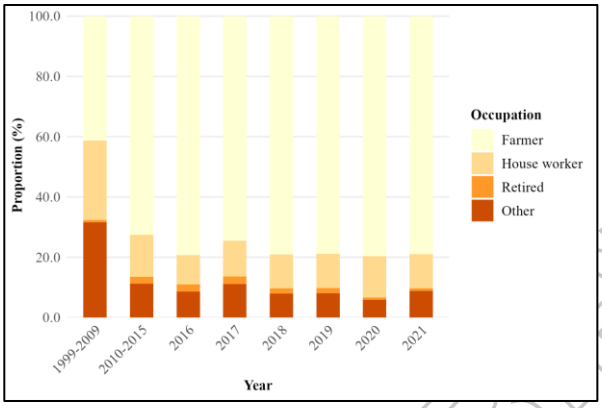


Figure S 2-4

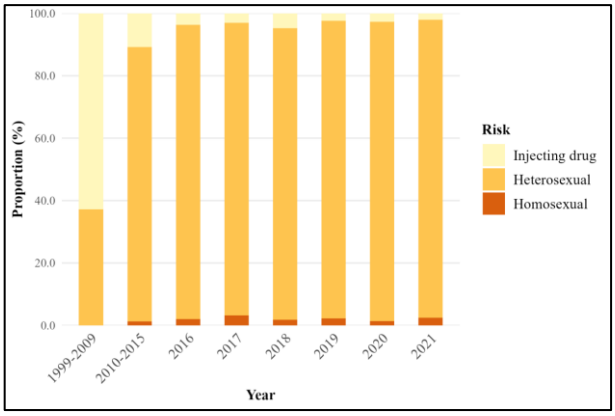


Figure S 2-5

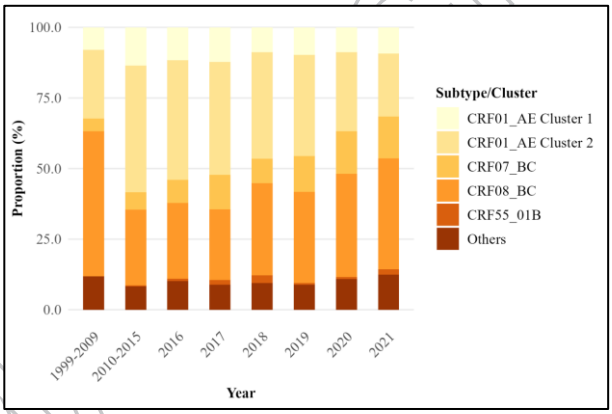


Figure S 2-6

Figure S2 Temporal shifts in the proportional distribution of age, gender, marital status, occupation, risk and HIV-1 subtypes/clusters (1999–2021), visualized via 100% stacked column graphs. (N=5,094)

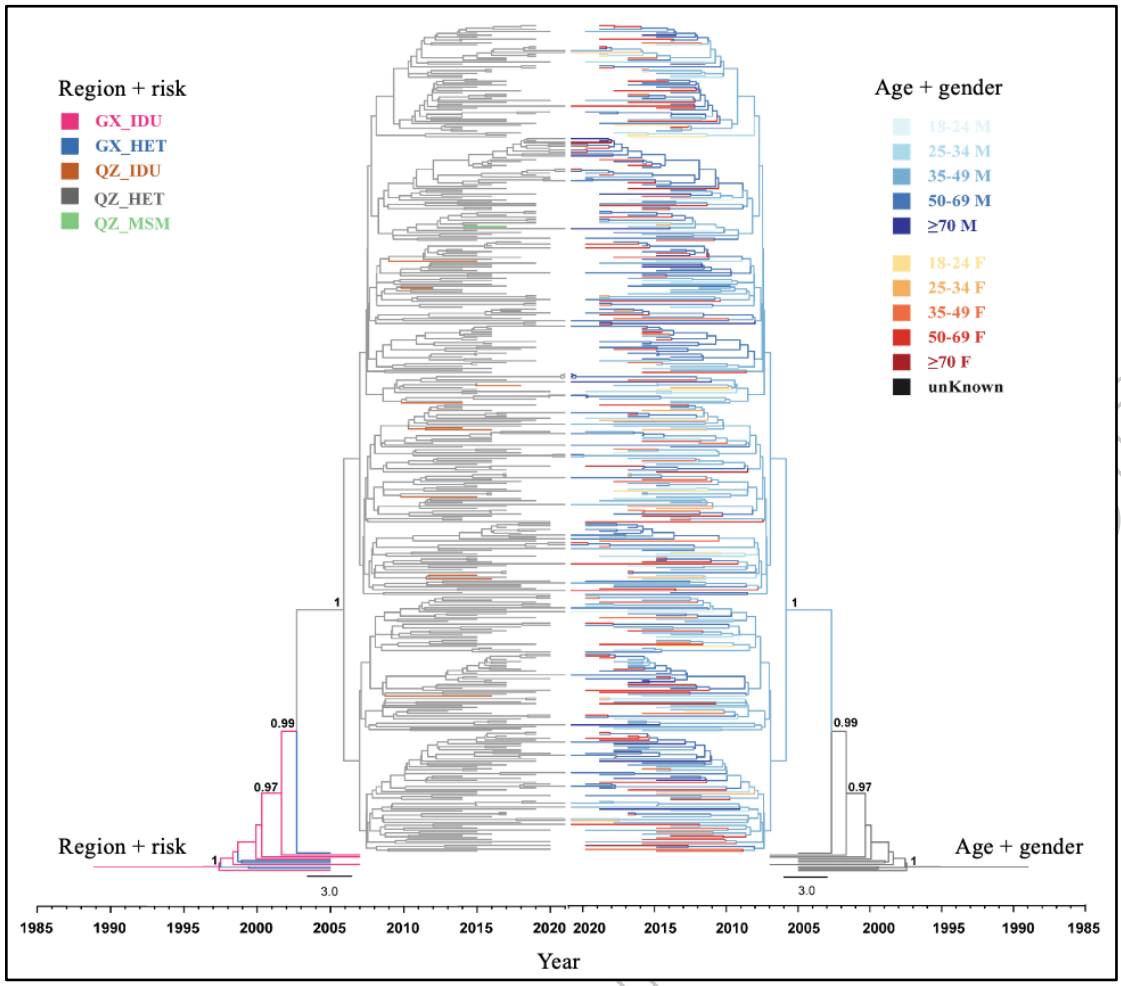


Figure S3-1 Maximum clade credibility tree of HIV-1 CRF01_AE cluster 1 transmission patterns. (A) Left panel: Phylogeographic reconstruction showing transmission sources by region and risk group (GX: Guangxi; QZ: Qinzhou). (B) Right panel: Demographic transmission patterns by age and gender (M: male; F: female). Branches display posterior probabilities and nodes indicate divergence times (year-scale). Trees were time-scaled using Bayesian evolutionary analysis and visualized with FigTree v1.4.4. Scale bars denote evolutionary time in years. This dataset included 518 RNA sequences.

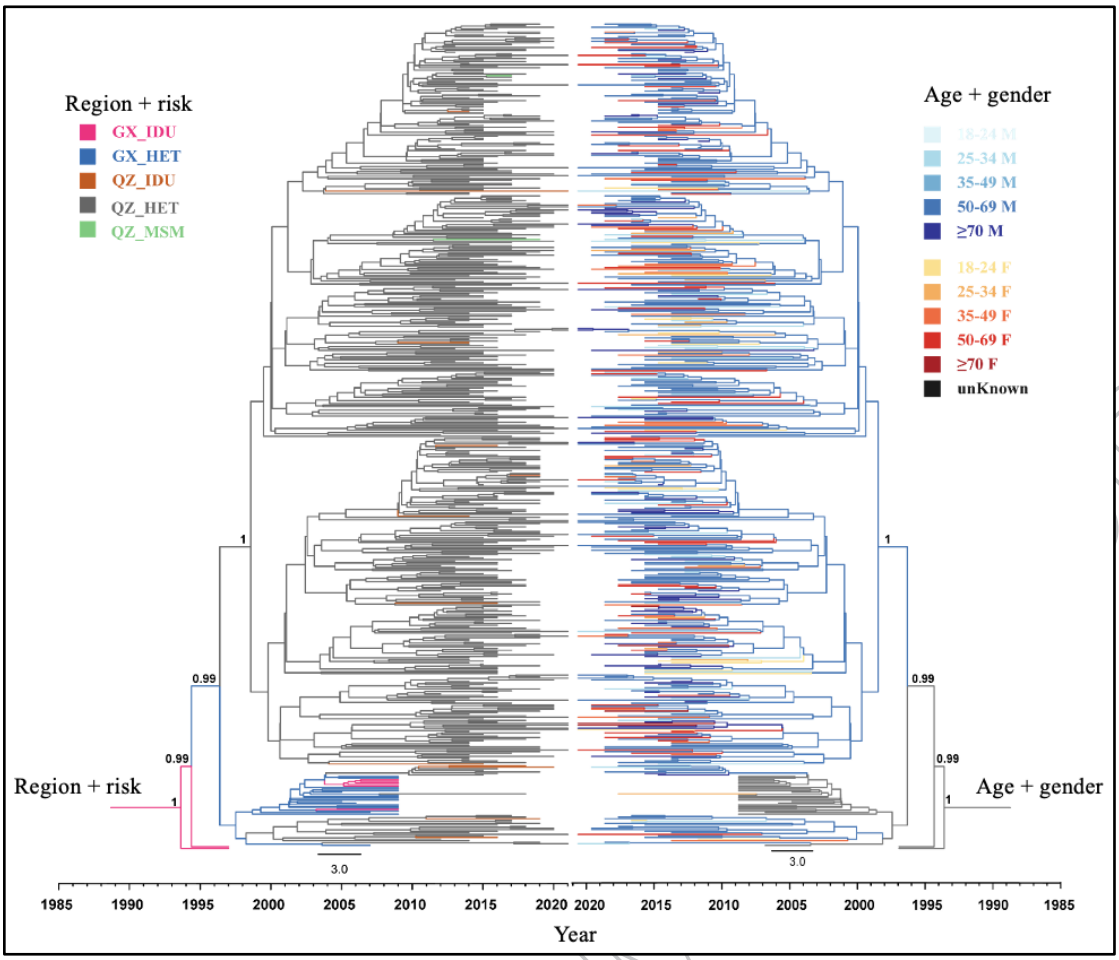


Figure S3-2 Maximum clade credibility tree of HIV-1 CRF01_AE cluster 2 transmission patterns. (A) Left panel: Transmission sources stratified by region and risk factors (GX: Guangxi; QZ: Qinzhou). (B) Right panel: Transmission patterns by age group and gender (M: male; F: female). Branches display posterior probabilities and nodes indicate divergence times (year-scale). Trees were reconstructed using Bayesian evolutionary analysis and visualized with FigTree (v1.4.4). Scale bars denote evolutionary time in years. This dataset included 527 RNA sequences.

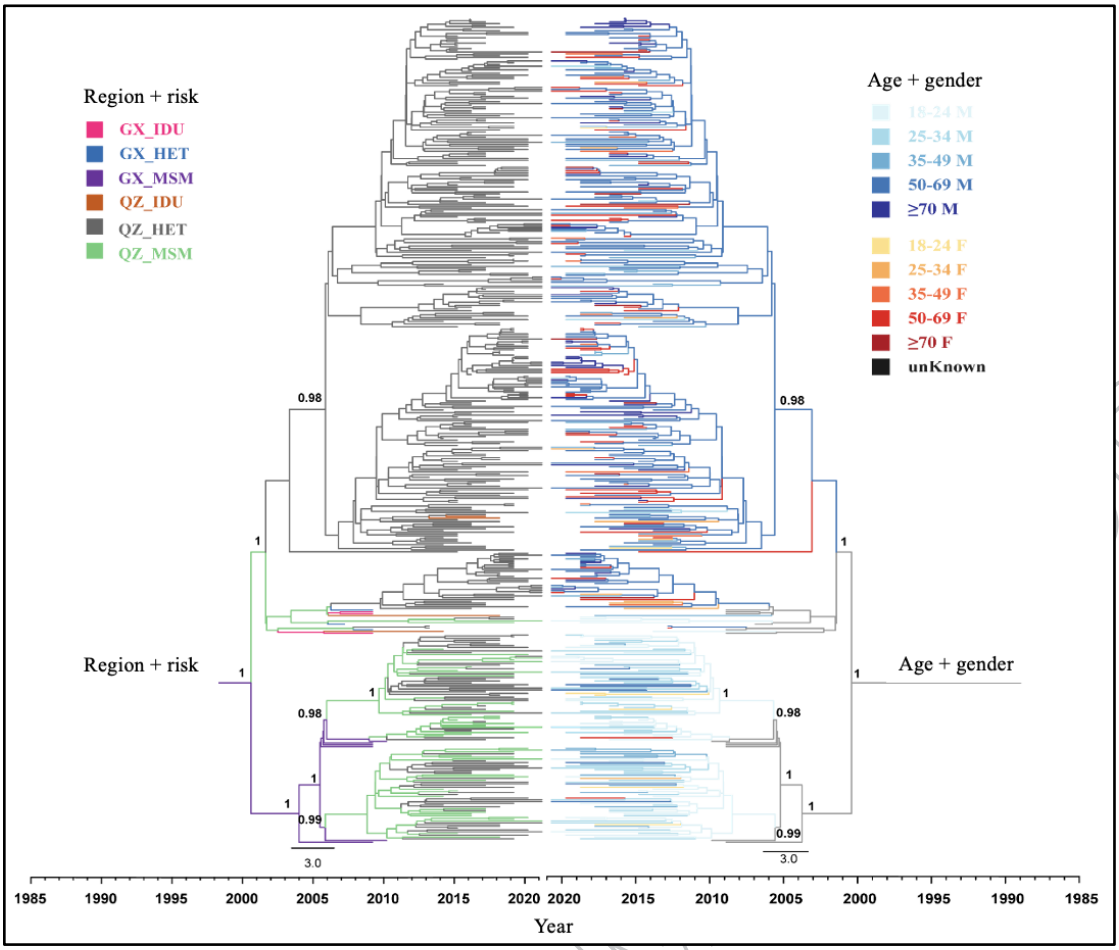


Figure S3-3 Maximum clade credibility tree of HIV-1 CRF07_BC transmission patterns. (A) Left panel: Transmission sources stratified by region and risk factors (GX: Guangxi; QZ: Qinzhou). (B) Right panel: Transmission patterns by age group and gender (M: male; F: female). Branches display posterior probabilities and nodes indicate divergence times (year-scale). Trees were reconstructed using Bayesian evolutionary analysis and visualized with FigTree (v1.4.4). Scale bars denote evolutionary time in years. This dataset included 503 RNA sequences.

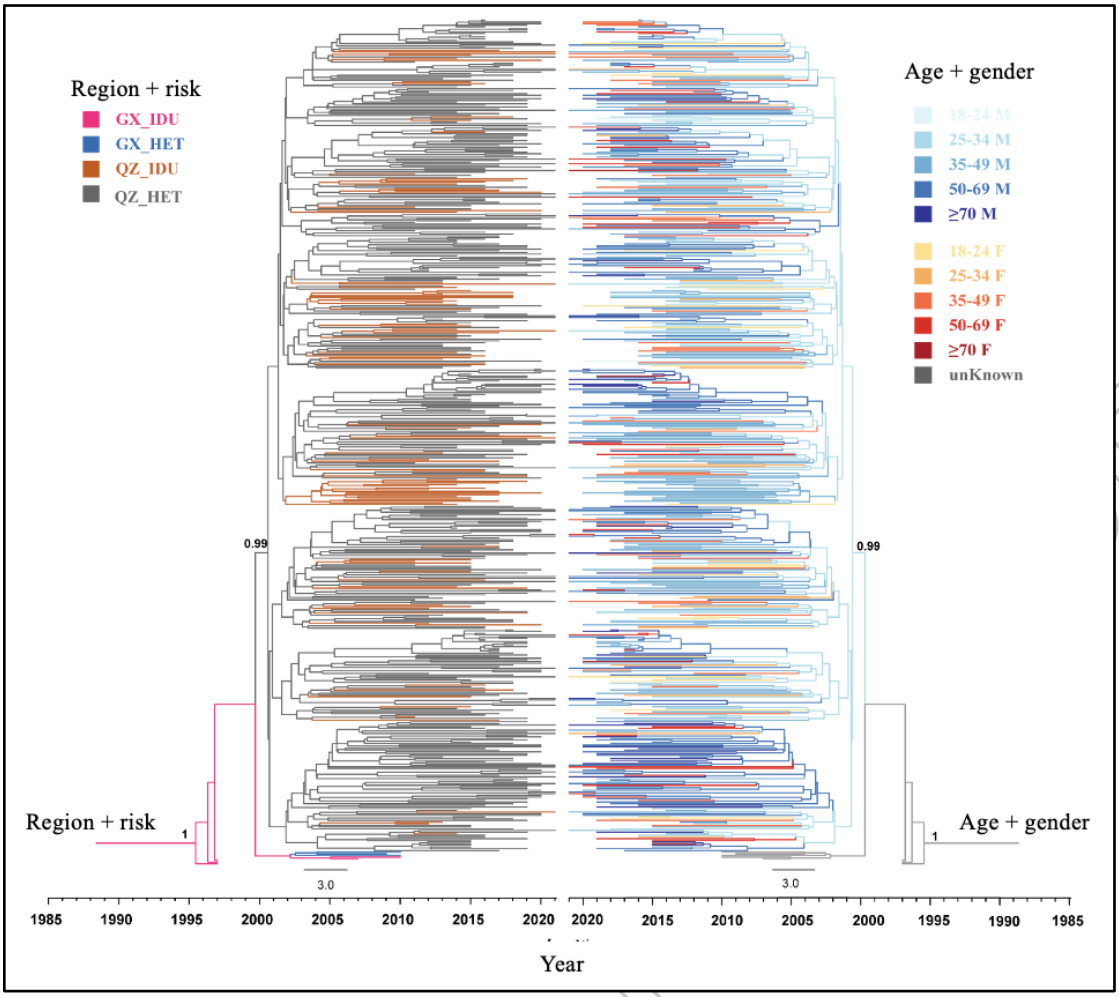


Figure S3-4 Maximum clade credibility tree of HIV-1 CRF08_BC transmission patterns. (A) Left panel: Transmission sources stratified by region and risk factors (GX: Guangxi; QZ: Qinzhou). (B) Right panel: Transmission patterns by age group and gender (M: male; F: female). Branches display posterior probabilities and nodes indicate divergence times (year-scale). Trees were reconstructed using Bayesian evolutionary analysis and visualized with FigTree (v1.4.4). Scale bars denote evolutionary time in years. This dataset included 508 RNA sequences.

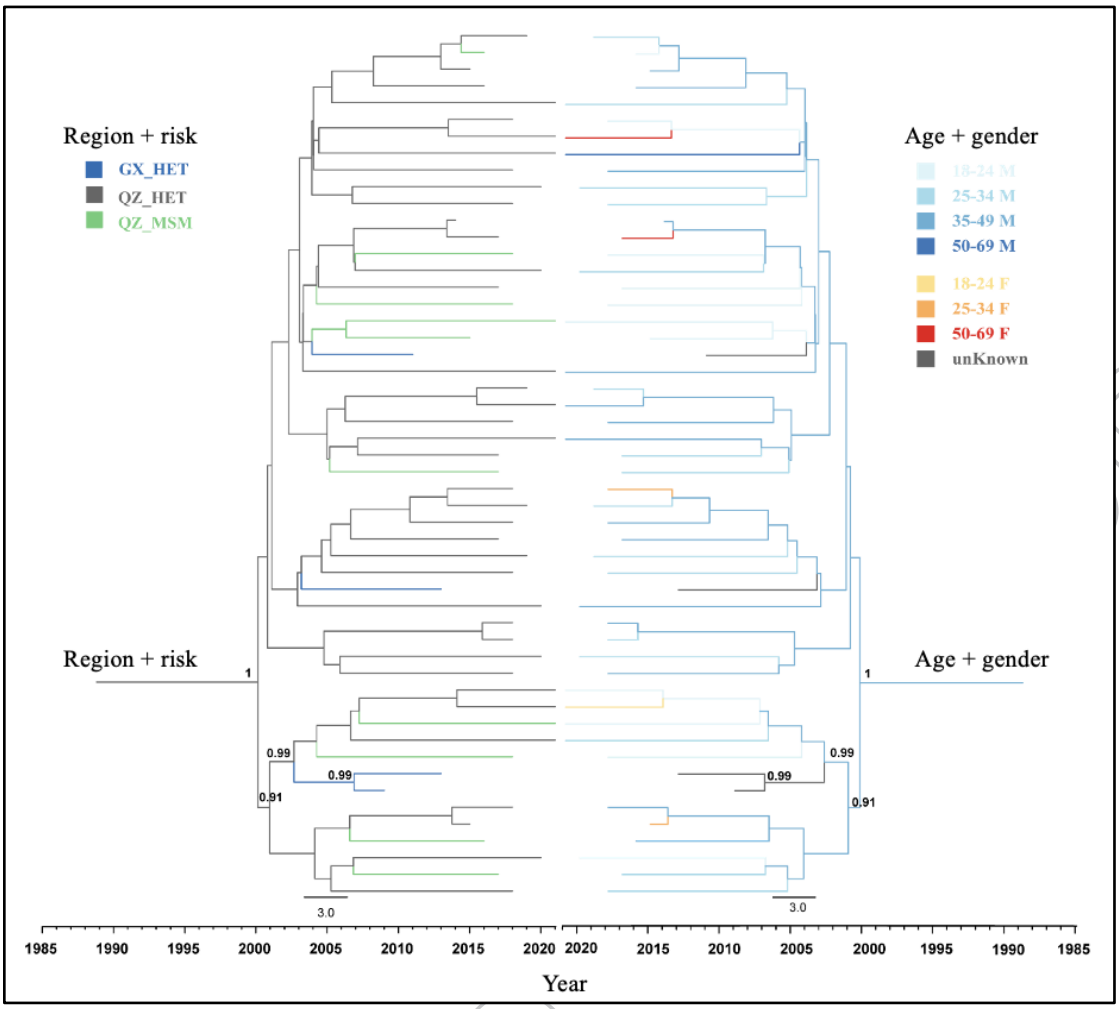


Figure S3-5 Maximum clade credibility tree of HIV-1 CRF55_01B transmission patterns. (A) Left panel: Transmission sources stratified by region and risk factors (GX: Guangxi; QZ: Qinzhou). (B) Right panel: Transmission patterns by age group and gender (M: male; F: female). Branches display posterior probabilities and nodes indicate divergence times (year-scale). Trees were reconstructed using Bayesian evolutionary analysis and visualized with FigTree (v1.4.4). Scale bars denote evolutionary time in years. This dataset included 53 RNA sequences.

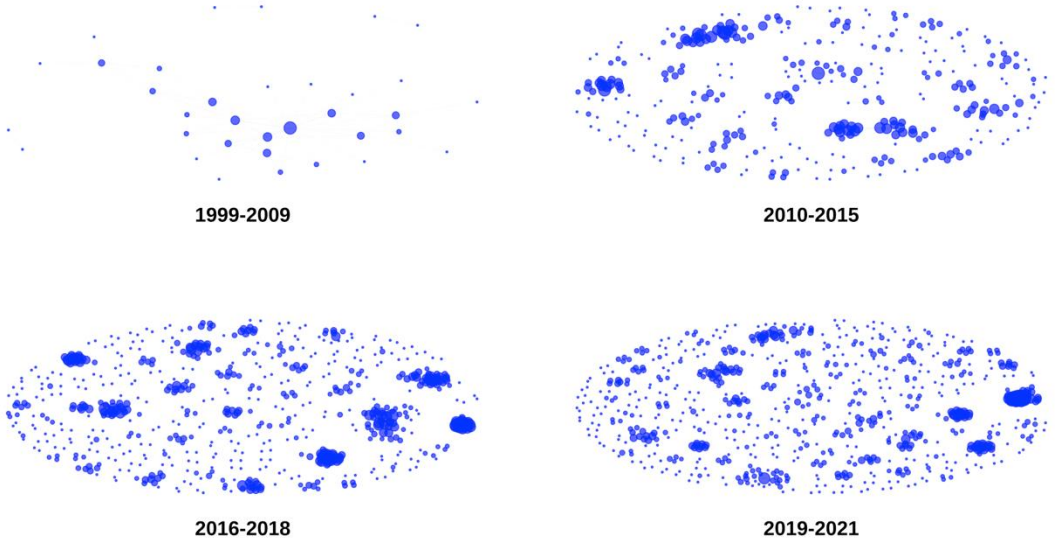


Figure S4 Phase-specific HIV-1 genetic networks (1999-2021). Nodes (blue dots) represent individual cases diagnosed in each phase, linked by a genetic distance threshold of 0.005 substitutions/site. The visualization highlights the evolving complexity of transmission clusters over time.

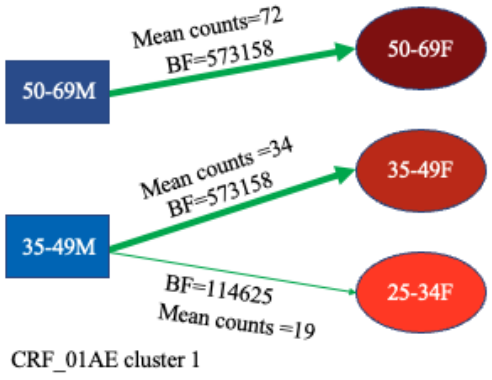


Figure S5-1

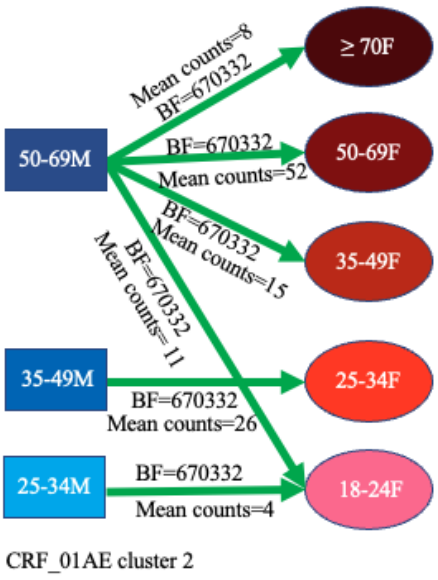


Figure S5-2

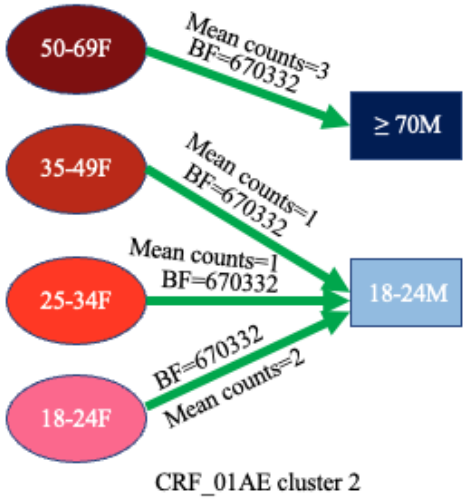


Figure S5-3

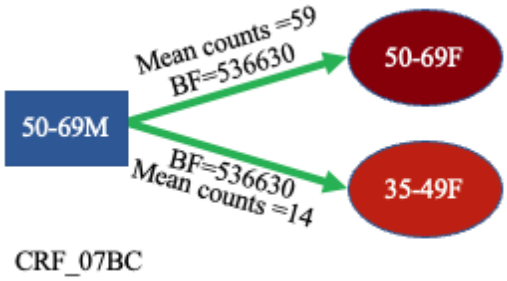


Figure S5-4

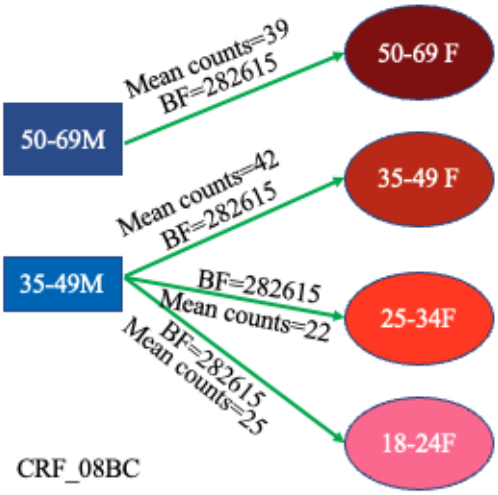


Figure S5-5

Figure S5 The transmission events between age-gender groups across HIV-1 subtypes. ⩾70M denotes men at least 70 years old, 50-69M denotes men in 50-69 years old, 35-49M denotes men in 35-49 years old, 18-24F denotes women in 18-25 years old, 25-34F denotes women in 25-34 years old, 35-49F denotes women in 35-49 years old, 50-69F denotes women in 50-69 years old, ⩾70F denotes women at least 70 years old. Only statistically robust transmission events with Bayes Factor (BF) ⩾100,000 and posterior probability ⩾ 0.9 were shown.

Supplementary file 2

Methods for details

1. Stratification of four epidemiological phases

The 23-year study period (1999-2021) was stratified into four distinct epidemiological phases based on key public health milestones and policy shifts.

Phase I [1999-2009, Pre-Guangxi AIDS Conquering Project (GACP) phase]: This period represents a stage of natural epidemic growth characterized by the implementation of routine prevention and control measures. During this phase, the epidemic expanded without the intensive, large-scale interventions that were later introduced under the GACP, providing a baseline of viral transmission dynamics in the region.

Phase II (2010-2015, GACP intensive intervention phase): This phase is marked by the large-scale implementation of regional prevention and control measures under the GACP, aiming to curb the rapidly rising epidemic.

Phase III (2016-2018, "Treat All" transition phase): Although the national "Treat All" policy was officially initiated in 2016, its comprehensive coverage in rural Southwest China was progressively phased in. This period represents the transition from targeted clinical eligibility to universal access in resource-limited settings.

Phase IV (2019-2021, "Treat All" full coverage phase): This phase is defined by the stabilized and comprehensive implementation of the "Treat All" strategy across the rural study areas, representing the contemporary management era.

2. Phylodynamic and phylogeographic analysis

2.1 Sequence preparation and subsampling

HIV-1 sequences were edited and aligned using BioEdit (v7.0.9.0) and the HIV Align tool. Subtypes and clusters were identified via phylogenetic reconstruction in MEGA (v10.0). For the two major subtype/clusters exceeding 1,000 sequences (CRF01_AE Cluster 2 and CRF08_BC), we performed stratified random subsampling (n=500 each) to ensure Bayesian computational feasibility while

preserving demographic representativeness (Supplementary file 1 Table S2). To enhance molecular clock calibration, 61 high-quality historical reference sequences (sampled 1997-2013) from the provincial HIV surveillance database of Guangxi Center for Disease Control and Prevention were integrated to corresponding subtype/cluster dataset as temporal anchors to estimate the Time of the Most Recent Common Ancestor (tMRCA) for each subtype/cluster (Supplementary file 1 Table S3).

2.2 Bayesian evolutionary reconstruction of HIV-1 subtypes/clusters

Phylodynamic analyses were conducted in BEAST (v1.8.4) for five lineages: CRF01_AE (Clusters 1 & 2), CRF07_BC, CRF08_BC, and CRF55_01B. Other lineages were excluded from these analyses due to insufficient sample sizes for robust inference. We employed a General Time-Reversible (GTR) substitution model, a relaxed uncorrelated lognormal (UCLN) molecular clock, and a Skygrid demographic prior [1–4]. Markov Chain Monte Carlo (MCMC) chains were run for 1 billion iterations, with the initial 10-30% discarded as burn-in. Convergence was confirmed using Tracer (v1.6), ensuring all parameters achieved an Effective Sample Size (ESS) >200 [5]. Maximum Clade Credibility (MCC) trees were generated via TreeAnnotator and visualized in FigTree (v1.4.4).

a) Temporal composition and migration dynamics of viral lineages

Due to sample size constraints, this analysis focused on four predominant HIV-1 lineages (CRF01_AE clusters 1 and 2, CRF07_BC, and CRF08_BC), excluding the less prevalent CRF55_01B.

(1) Temporal dynamics of lineage composition defined by age-gender state frequencies: To characterize the temporal evolution of viral lineage distributions across different demographic (age-gender) subgroups, we extracted lineage state frequencies from the annotated MCC trees. Following

established phylodynamic visualization protocols [1], we performed time-slicing across the phylogeny at 0.5-year intervals from the root to the most recent sampling date. This approach allowed us to quantify the temporal shifts in lineage composition and capture how the relative proportions of each age-gender group within the regional viral pool evolved over the 23-year period. At each discrete time point, we identified all co-existing viral lineages and extracted their most probable reconstructed age-gender ancestral states. The relative proportions of lineages belonging to each demographic group were then visualized using smoothed stacked area charts, implemented programmatically via the treeio [10] and ggplot2 R packages.

(2) Inference of asymmetric transmission dynamics between age-gender subgroups: The migration dynamics and transmission pathways between age-gender subgroups were reconstructed using the Bayesian Stochastic Search Variable Selection (BSSVS) framework [6], complemented by Markov jump analysis to quantify the directionality and intensity of viral spread throughout the study period [7]. The statistical robustness of the directional transmission pathways between age-gender subgroups was evaluated over the entire 1999-2021 period using Bayes Factors (BF) and posterior probabilities calculated via SpreaD3 v0.9.7.1 [8]. Following the standard interpretation scales, BF values are categorized as follows: $10 \leqslant BF < 100$ indicates strong evidence and $BF \geqslant 100$ indicates decisive evidence [9]. To ensure the highest level of inferential stringency and minimize spurious linkages within our complex transmission network, we adopted a highly conservative threshold of $BF \geqslant 100{,}000$ and posterior probability $\geqslant 0.9$.

3. Mean Molecular Cluster Growth calculation

The localized expansion intensity was quantified as Mean Molecular Cluster Growth (MMCG). For a given township *j* at stage *t*, MMCG is defined as:

$$\text{MMCG}_{j,t} = \frac{\sum_{vi \in V_t} 1(v_i \in \text{Township}_j \ \wedge \ vi \notin V_{t-1})}{N_j}.$$

Parameter Annotation: V_t, V_{t-1}: Sets of nodes in the genetic network at stage *t* and *t-1*, respectively.

1(·): An indicator function identifying "growth nodes" (incident cases) geographically located within township *j*.

N_j: The normalization factor. In this study, $N_j = 1$ was applied to quantify the absolute expansion magnitude as a proxy for transmission intensity, this approach emphasizes total viral lineage influx along key transportation corridors and eliminates inaccuracies associated with transient population denominators in rural areas.

Zero-filling: Townships with no incident cases were assigned a value of 0 to ensure a continuous surface for subsequent spatial modeling.

4. Spatial autocorrelation analysis

Temporal shifts in spatial clustering of mean degree centrality and MMCG were monitored using the Global Moran's I index at each phase. To identify geographic hotspots, Local Indicators of Spatial Association (LISA) analysis [12] based on Local Moran's I was performed for both HIV genetic network mean degree centrality and MMCG. A Queen's contiguity weight matrix was used to categorize townships into High-High (hotspots), Low-Low (cold spots), and spatial outliers (High-Low and Low-High). Statistical significance was verified using a Z-test based on asymptotic normal distribution.

5. Road proximity analysis

We calculated the minimum geodesic distance from each township's geometric centroid to the nearest segment of the regional transportation network (comprising expressways, national and provincial highways) using the sf package in R. Pearson correlation coefficients (r) were employed to quantify the association of both township-level mean degree centrality and MMCG with road proximity. For categorical comparison, townships were stratified into "Near-road" (≤2 km) and "Far-road" (> 2 km) groups using a spatial buffer. Disparities in MMCG between townships with "Near-road" (≤ 2 km) and "Far-road" (> 2 km) were validated using the Wilcoxon rank-sum test. To isolate the impact of road proximity from the effects of general population aggregation, we performed a Negative Binomial regression analysis. While the model used cumulative reported HIV case counts as the outcome, the inclusion of a log(Total_Pop) offset term effectively modeled the cumulative incidence across townships. This framework was specifically chosen to address the overdispersion inherent in the count data and the potential confounding by demographic clustering. The formula was specified as: $\log(\mu) = \beta_0 + \beta_1 *$ Road_Proximity $+ \beta_2 *$Pop_Density + log(Total_Pop), where μ represents the number of cumulative reported HIV cases.

6. Generalized estimating equations analysis

A Generalized Estimating Equations (GEE) model was used to assess the association between ART status and onward HIV transmission among 5,094 HIV/AIDS cases diagnosed between 1999 and 2021. To evaluate the "Treat All" policy, we constructed a dynamic transmission reservoir that

yielded 19,661 observations. These observations capture the ART status of potential sources at the end of each year *t* (2016-2020), with each status linked to the occurrence of new molecular linkages in the subsequent year *t+1* (2017–2021) [13]. This lagged design ensures that the identified drivers (ART status) precede the transmission events in time. The reservoir was maintained as follows: (1) individuals diagnosed between 1999 and 2016 (excluding those deceased by the end of 2016) formed the initial baseline pool; (2) cases newly diagnosed between 2017 and 2020 were sequentially incorporated into the pool in the year following their diagnosis; and (3) individuals were removed annually at the end of the year of their death to maintain an active reservoir. Cases diagnosed in 2021 were used to ascertain transmission outcomes but did not contribute to the observations as potential sources. The ART status was categorized as ART-naïve, active ART, or ART dropout. Within-subject correlations were addressed using an exchangeable correlation matrix [14]. The model was adjusted for socio-demographics, risk, residence, HIV subtype, and baseline CD4 level. A two-tailed $P < 0.05$ was considered significant.

List of abbreviations:

GACP: Guangxi AIDS Conquering Project;

tMRCA: time to the most recent common ancestor;

GTR: general time-reversible;

UCLN: uncorrelated lognormal;

MCMC: Markov chain Monte Carlo;

ESS: Effective Sample Size;

MCC: Maximum clade credibility;

BSSVS: Bayesian stochastic search variable selection;

BF: Bayes Factors;

MMCG: Mean Molecular Cluster Growth;

LISA: Local Indicators of Spatial Association;

GEE: Generalized Estimating Equations;

ART: antiretroviral therapy.

Reference

1. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, *et al.* The early spread and epidemic ignition of HIV-1 in human populations. Science 2014;346:56–61.
2. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 2012;29:1969–73.
3. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol 2013;30:713–24.
4. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol 2006;4:e88.
5. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 2018;67:901–4.
6. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol 2009;5:e1000520.
7. Minin VN, Suchard MA. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 2008;56:391–412.
8. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol 2016;33:2167–9.
9. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773–795.
10. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, *et al.* Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol Biol Evol 2020;37:599–603.
11. Minin VN, Suchard MA. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 2007;56:391–412.
12. Moran P a. P. Notes on continuous stochastic phenomena. Biometrika 1950;37:17–23.

13. Chen Y, Cao Z, Li J, Chen J, Zhu Q, Liang S, *et al.* HIV transmission and associated factors under the scale-up of HIV antiretroviral therapy: a population-based longitudinal molecular network study. Virol J 2023;20:289.
14. Ballinger GA. Using Generalized Estimating Equations for Longitudinal Data Analysis. Organ Res Methods 2004;7:127–50.