HIV-1 genetic transmission networks among people living with HIV/AIDS in Sichuan, China: a genomic and spatial epidemiological analysis

*Dan Yuan,[a,†] Bin Yu,[b,†] Shu Liang,[a,†] Teng Fei,[d,***] Houlin Tang,[e] Rui Kang,[d] Yiping Li,[a] Li Ye,[a] Peng Jia,[d,f,**] and Shujuan Yang,[c,f,*]*

[a]Center for AIDS/STD Control and Prevention, Sichuan Center for Disease Control and Prevention, Chengdu, China
[b]West China Second University Hospital, Sichuan University, Chengdu, China
[c]West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China
[d]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[e]National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
[f]International Institute of Spatial Lifecourse Epidemiology (ISLE), Wuhan University, Wuhan, China

Summary

Background Spatialized HIV genetic transmission networks can help understand dynamic changes of HIV-1 at the regional level. This study aimed to combine genomic, epidemiological, and spatial data to investigate the patterns of the HIV-1 epidemic at both individual and regional levels among people living with HIV (PLWH) with virological failure of antiretroviral therapy (ART).

Methods We evaluated the transmission patterns of 5,790 PLWH with identified pol sequences of the five main HIV-1 subtypes (B, CRF08_BC, CRF85_BC, CRF07_BC, and CRF01_AE) in Sichuan Province, China. A phylogenetic cluster was defined as a group of sequences with genetically similar HIV strains, with all phylogenetic clusters forming an HIV-1 genetic transmission network for each subtype. Logistic regression was used to identify the potential risk factors for phylogenetic clustering. Spatial analysis was applied to demonstrate the geographic patterns of phylogenetic clustering rates; intensity matrices and flow maps were made to demonstrate the intensity of transmission within and between cities.

Findings There were 2,159 (37.3%) of 5,790 PLWH, distributed in 452 phylogenetic clusters. Some individual clinical and behavioral factors were associated with phylogenetic clustering, including a viral load of >50,000 copies/ml (OR=1.16, 95%CI=1.02-1.33), infection of other sexually transmitted diseases (OR=1.38, 95%CI=1.12-1.69), and ≥5 non-marital sexual partners (OR=1.25, 95%CI=1.03-1.51), while >3 years of treatment since the initial ART was associated with less likelihood of phylogenetic clustering (OR=0.82, 95%CI=0.70-0.97). The phylogenetic clustering rates varied regionally and were highest in the central region of Sichuan, especially for subtype CRF08_BC. The significant spatial clusters of high and low phylogenetic clustering rates were detected in the east (Dazhou for B; Zigong and Luzhou for CFR08_BC) and west (Yaan and Ganzi for CRF07_BC), respectively. The proportion of intercity transmission varied across cities from 0.14 (Yibin) and 1.00 (Ganzi). Stronger intercity transmission than average existed between some cities, e.g., between Deyang and Neijiang. CRF07_BC was the most widespread subtype between cities, and CRF85_BC (a novel HIV-1 subtype) showed strong intercity transmission (e.g., between Yibin and Guangan).

Interpretation The phylogenetic clustering rates and intercity connections of HIV-1 have varied geographically, possibly due to varying human mobility, traffic convenience, and economic activities. Our findings enhanced the understanding of the dynamics of HIV-1 transmission from individual to city level, and demonstrated a novel cross-disciplinary (epidemiological, genetic, and spatial) approach by which we identified high-risk populations and areas. Our approach could be adapted to other regions for precision public health interventions.

---

*Correspondence: Prof Shujuan Yang, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China.
**Correspondence: Prof Peng Jia, School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.
***Correspondence: Prof Teng Fei, School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
  E-mail addresses: feiteng@whu.edu.cn (T. Fei), jiapengff@hotmail.com (P. Jia), rekiny@126.com (S. Yang).

[†] Joint first authors.

Research in Context

Evidence before this study

HIV genetic transmission network is a group of HIV-infected individuals having genetically similar HIV strains. Genetic transmission network analysis has been combined with demographic and epidemiological data to identify factors underlying HIV transmission. However, HIV transmission networks have varied across regions. Spatial analysis may be combined with genetic transmission networks and traditional epidemiological analyses to reveal spatial heterogeneities of HIV transmission networks. We searched PubMed for all article types published from database inception to 21 March 2021, with the terms "HIV", "spatial", "genetic transmission network", "clustering", and "link". We found that few studies have revealed spatial heterogeneities of HIV transmission networks. A recent study (2020) has reconstructed the directed HIV transmission networks of a population-based sample on the basis of HIV deep sequence data, and examined HIV transmission between spatial clusters of HIV cases and surrounding inland communities in Uganda. In addition, there have been some phylogeographic analyses that integrate phylogenetics and geography to reveal the evolution of HIV strains and the transmission of HIV, but none of them have revealed the transmission between regions.

Added value of this study

This study fills an important knowledge gap about the spatial characteristics of HIV-1 genetic transmission network, and revealed the transmission links between different regions in Sichuan Province, China. We used cross-disciplinary approaches (i.e., epidemiological, spatial, genomic approaches) to analyze the HIV-1 transmission network among people living with HIV/AIDS with virologic failure of antiretroviral therapy (ART). From the traditional perspective, we sequenced and constructed HIV-1 genetic transmission networks, calculated the phylogenetic clustering rates, and subsequently analyzed the factors of phylogenetic clustering rates. These analyses help to find out high-risk populations which should be monitored at an individual level. Furthermore, we spatialized HIV-1 genetic transmission network, with explanation of spatial distribution and spatial clustering patterns of phylogenetic clustering rates, and spatial links of HIV-1 between cities. These analyses contributed to better understandings of the dynamic characteristics of all HIV-1 subtypes from the regional level. Briefly, we found phylogenetic clustering rate was highest in the central part of Sichuan. The cluster map help to

better understand the spatial interaction of adjacent regions. Additionally, the intensity matrixes revealed most cities with strong intracity links, but some cities with strong transmission links with another city (even some geographically distant regions); the flow map suggested the subtype CRF07_BC a highest intensity of regional connections, and novel subtype CRF85_BC showed spread trend.

Implication of all the available evidence

Integrating genomic, epidemiological, and spatial methods is useful in revealing spatial heterogeneities of HIV-1 subtypes, characteristics of infected individuals (e.g., transmission routes), as well as HIV transmission between regions. Such cross-disciplinary approaches could be adapted to other regions for precision public health interventions (e.g., spatially explicit interventions).

Introduction

Human immunodeficiency virus (HIV) infection is considered one of the most devastating infectious diseases in human history, and it has been responsible for nearly 76 million infections globally.[1] The prevention of HIV spread is a key priority on the global health agenda, with the necessary step of identifying the patterns of HIV transmission. Geographic differences (e.g., economic, cultural, and demographic differences across geographic locations) may to some extent determine the patterns of HIV transmission.[2] Since the increasing spatial analyses of HIV have exhibited considerable spatial heterogeneities in the prevalence and incidence of HIV,[3,4] region-specific interventions have attracted much attention. A recent spatial epidemiological study of HIV in Malawi[5] has shown that region-specific HIV control, such as geographically focused and age-specific approaches to confidential HIV testing for men who have sex with men (MSM), was more effective than universal approaches to reduce broad HIV transmission, owing to a better understanding of inter-regional epidemic dynamics.[6,7] In China, a meta-analysis revealed the significant spatial heterogeneity of different HIV subtypes (also referred to as HIV-1 subtypes) among different populations,[8] which helped us understand the dynamic changes of HIV-1 subtypes from a spatial perspective. In view of the severe epidemic of HIV-1 in China,[9] the spatial dynamics of different subtypes are worthy of an in-depth study.

HIV-1 genetic transmission network analysis is an effective strategy for identifying individuals at high risk.[10,11] Specifically, the HIV-1 genetic transmission networks ascribe a putative transmission link to any pair of viral sequences with a high degree of genetic similarity.[12] HIV-1 genetic transmission network analysis is a tool for revealing obscured transmission patterns which would not have been found by traditional epidemiological approaches (e.g., self-reported heterosexual males whose viruses cluster only with those from MSM).[13] Also, HIV-1 genetic transmission network analysis has become an increasingly important component in population-based surveillance of infectious diseases.[11] For instance, the HIV-1 genetic transmission networks among people living with HIV/AIDS (PLWH) in San Mateo County (California, United States) have identified a striking role of intracommunity transmission dynamics among African Americans.[14] In another study, HIV-1 genetic transmission network analysis helped identify a central role of non-disclosed MSM in HIV-1 spread in China.[15] However, the HIV-1 genetic transmission networks alone cannot capture the spatial characteristics of HIV transmission. Revealing the spatial variation of genetic transmission networks could provide more detailed information for designing region-specific interventions.[4,7]

Sichuan Province, with the highest prevalence of HIV-1 in China (161,456 PLWH in 2018), has a unique and heterogeneous population of PLWH with multiple ethnicities, language communities, living environments, and social norms, overlapping high-risk HIV-related groups of people who inject drugs, sex workers, and men who have sex with men (MSM).[16−18] This study adopted cross-disciplinary approaches, combining spatial, epidemiological, and genetic transmission network analyses, to better understand the patterns of the HIV-1 epidemic at both individual and regional levels in Sichuan, China. The findings are important evidence for developing precise strategies to reduce HIV-1 epidemic.

Methods

Study design and participants
Sichuan is located in the southwest of China. PLWH were screened and selected from the basic information system for AIDS Prevention and Control of the Sichuan Provincial Center for Disease Control and Prevention (CDC), if meeting the three inclusion criteria: (1) PLWH in Sichuan Province between January 2018 and December 2018, (2) receiving antiretroviral therapy (ART) for at least 6 months, (3) being diagnosed as the virological failure in ART (i.e., HIV RNA level>1000 copies/ml, according to the 2020 National Guideline for Detection of HIV/AIDS in China).[19]

The data of 7,011 PLWH with virological failure of ART in total were extracted from the information

system. Among these PLWH, pol sequence information was successfully exported from 5,926 (84.5%) samples. After eliminating duplicate samples and other HIV-1 subtype sequences, 5,790 PLWH with pol sequences of the five main HIV-1 subtypes (i.e., subtype B, CRF08_BC, CRF85_BC, CRF07_BC, and CRF01_AE) were finally included in this study. For these PLWH, their information was exported from their medical records in the system, including sociodemographic (e.g., age, sex, ethnicity, marital status), HIV-related behavioral (e.g., syringe sharing, non-marital sex, homosexual sex), and HIV/AIDS-related clinical characteristics (e.g., viral load, infection of other sexually transmitted diseases [STDs]), as well as the possible infection routes of HIV (heterosexual, homosexual, intravenous drug use, sexual contact and intravenous drug use, and mother-to-child) recalled by the PLWH at the time of initial diagnosis.

The study protocol was approved by the Ethics Committee of the Sichuan Center for Disease Control and Prevention (SCCDCIRB2017002). The study was following the Helsinki Declaration of 1964.

Laboratory tests and genetic analysis
About 5 ml of venous blood was extracted from each PLWH to test the CD4+ T cell counts and the viral load of HIV-1. Flow cytometry was used to quantify the CD4+ T cells in the local CDCs. The plasma samples were isolated from each participant and preserved in a -86C freezer before sending the cold chain to the Sichuan CDC for viral load measurement.

To infer transmission networks by phylogenetic analysis, viral nucleic acid was obtained from 200 ml plasma of PLWH by extraction machines (MagNA Pure LC 2.0 system, Roche, Branchburg, NJ). Similar to the previous report,[17] sequences were generated from the HIV-1 pol. Briefly, the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) was used to amplify the full-length protease gene in the pol region and the first 300 codons of the reverse transcriptase gene. The PCR products were dealt with electrophoresis with 1% agarose gel, and the amplified positive products were purified and sequenced by Beijing Genomics Research Center Ltd. The current antiretroviral drugs are usually reverse transcriptase inhibitors and could cause pol gene mutations. Therefore, we excluded 62 codon positions (33 in protease and 29 in reverse transcriptase) associated with drug resistance mutations on the basis of the HIV drug resistance database (https://hivdb.stanford.edu), to reduce the potential confounding effects of convergent evolution on sequence analysis.[20−24]

The phylogenetic tree and genetic distance-based methods were used to analyze the HIV-1 genetic transmission networks.[17] First, phylogenetic trees were constructed in FastTree 3.0 by using a maximum likelihood method to compare sequences, to identify HIV-1

subtypes. For each subtype, the network was formed based on all phylogenetic clusters, where a phylogenetic cluster is defined as a group of sequences with genetically similar HIV strains. The pairwise Tamura-Nei 93 (TN93) genetic distance was calculated among all sequences in HyPhy 2.2.4. When a given distance threshold was assigned, any two sequences with a closer distance than the threshold distance formed an initial cluster; each other sequence joined a specific cluster if closer to any sequence in that cluster than the threshold distance.[11] This step iterated until all separate sequences had a farther distance from any sequence in the existing clusters than the threshold distance, and hence cannot join any cluster. According to the National Technical Guideline for HIV Transmission Network Monitoring and Intervention in China, the distance threshold resulting in the maximum number of clusters among the five subtypes was used as the final genetic distance threshold, and all clusters formed under this threshold were considered final and visualized as the HIV-1 genetic transmission networks of each subtype. The visual editing of the genetic transmission networks were completed in Cytoscape 3.5 (https://cytoscape.org), which has been widely used to construct molecular networks,[25] also suggested by the National Technical Guideline for HIV Transmission Network Monitoring and Intervention in China.

The phylogenetic clustering rate was defined as the proportion of the PLWH within the phylogenetic clusters over all PLWH in a given area. A higher phylogenetic clustering rate indicated more genetic relatedness of HIV-1 among PLWH in the networks, which also reflected the more putative transmission link between infected individuals and severe HIV-1 epidemic dynamics.

Statistical analysis

The multivariate logistic regression model was used to examine the association between each of the exposure variables of interest (i.e., HIV/AIDS-related clinical characteristics and HIV-related risk behaviors) and phylogenetic clustering (1 = individuals within the phylogenetic clusters, and 0 = individuals outside the phylogenetic clusters). Being within the phylogenetic clusters implies increased transmission, so the six available HIV/AIDS-related clinical characteristics and HIV-related risk behaviors in our dataset, which have been associated with increased transmission, were used as the categorical exposure variables of interest in this study: the duration of treatment since initial ART (<1, 1-3, >3 years), viral load (1,000-10,000, >10,000-50,000, >50,000 copies/ml), infection of other STDs (no, yes), and the numbers of syringe sharers (0, 1-4, ≥5), non-marital sexual partners (0, 1-4, ≥5), and homosexual partners (0, 1-4, ≥5); the six common sociodemographic characteristics were adjusted for: age (≤14, 15-49, ≥50 years), sex (male, female), ethnicity (Han, minority),

marital status (married and living with spouse, unmarried, divorced/widowed), educational level (illiteracy, primary school, middle school or higher), and occupation (farmer, employed, retired, student/child).[26−28] The statistical analysis was conducted in SPSS (version 21.0, Inc, Chicago, IL, USA).

Spatial analysis

Spatial analysis was conducted to examine the spatial distribution and spatial clustering patterns of phylogenetic clustering rates of the five HIV-1 subtypes and HIV transmission routes. To map the phylogenetic clustering rates of the HIV-1 subtypes, we counted the number of PLWH infected with each of the five subtypes of HIV-1 in each city of Sichuan, among which the proportion of PLWH in the genetic transmission networks was calculated as the phylogenetic clustering rate of each subtype.

The Anselin local Moran's I statistic was employed to the spatial unit of city to detect spatial clusters of phylogenetic clustering rates, resulting in five categorical values[29]: "high-high" means that a given city with a high phylogenetic clustering rate is surrounded by the cities also with a high phylogenetic clustering rate; "low-low" means that a city with a low phylogenetic clustering rate is surrounded by the cities also with a low phylogenetic clustering rate; "low-high" means that a city with a low phylogenetic clustering rate is surrounded by the cities with a relatively high phylogenetic clustering rate; "high-low" means that a city with a high phylogenetic clustering rate is surrounded by the cities with a relatively low phylogenetic clustering rate; a value of zero means that the phylogenetic clustering rates are distributed randomly among cities (i.e., no spatial clustering patterns). A z-test was conducted to determine if each spatial clustering pattern of phylogenetic clustering rates was significantly different from a random distribution.

To reflect HIV transmission types geographically, for each HIV-1 subtype, the proportion of intracity transmission in each city was calculated by dividing the number of links between PLWH in the city by the total number of links with any PLWH in the city (each of the PLWH was assigned to the cities s/he lived); the remaining proportion was referred to as the proportion of intercity transmission (i.e., the proportion of links between PLWH in and outside the city). To demonstrate finer-scale HIV transmission status among cities, the transmission network for each HIV-1 subtype was visualized and colored differently in intensity matrices and flow maps. The color of the grid cell at the intersection of two cities in an intensity matrix represented the number of linkages between the PLWH in two cities. Flow maps visualized the intensity of intercity HIV transmission by coloring the links of each HIV-1 subtype differently and scaling the line width by the number of

linkages, which were consistent with values in the intensity matrices. The flow maps were created in QGIS (version 3.10).[30]

In addition, we examined the variations of the compositions of age, sex, and transmission routes across cities. The Pielou index, an index of diversity across all categories in a given area, was further employed to demonstrate the within-city heterogeneity in the five possible HIV transmission routes considered in this study. It was calculated as $(-\sum_{i=1}^{s} p_i \times \ln p_i / \ln S)$, where p_i is the proportion of the ith type of transmission route in each city, and s=5 in this study (the number of types of transmission route). The value of the Pielou index falls between 0 (complete homogeneity) and 1 (complete heterogeneity).

Role of the funding source
The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.

Results

Analysis of HIV-1 genetic transmission networks
There were 2,159 of 5,790 sequences of PLWH detected within the genetic transmission networks with the genetic distance threshold of 0.01, distributed in 452 phylogenetic clusters (Figure 1), which were more than the numbers of clusters formed with any other genetic distance thresholds. The 452 phylogenetic clusters included 223 (49.3%), 120 (26.6%), 80 (17.7%), 18 (4.0%), and 11 (2.4%) clusters for subtypes CRF_07 BC, CRF_01 AE, CRF_08 BC, CRF_85 BC, and B, respectively. Among the pol sequences within the networks, 601 were only linked to one pol sequence, and 1,558 were linked to more than 2 pol sequences.

According to the epidemiological survey, most (73.6%) of the 5,790 PLWH were infected through heterosexual transmission, followed by intravenous drug use (14.4%), homosexual (6.5%), mother-to-child (2.8%), and both sexual contact and intravenous drug use transmission (2.7%). This order remained among the 2,159 PLWH within the phylogenetic clusters, where the major route of transmission was heterosexual
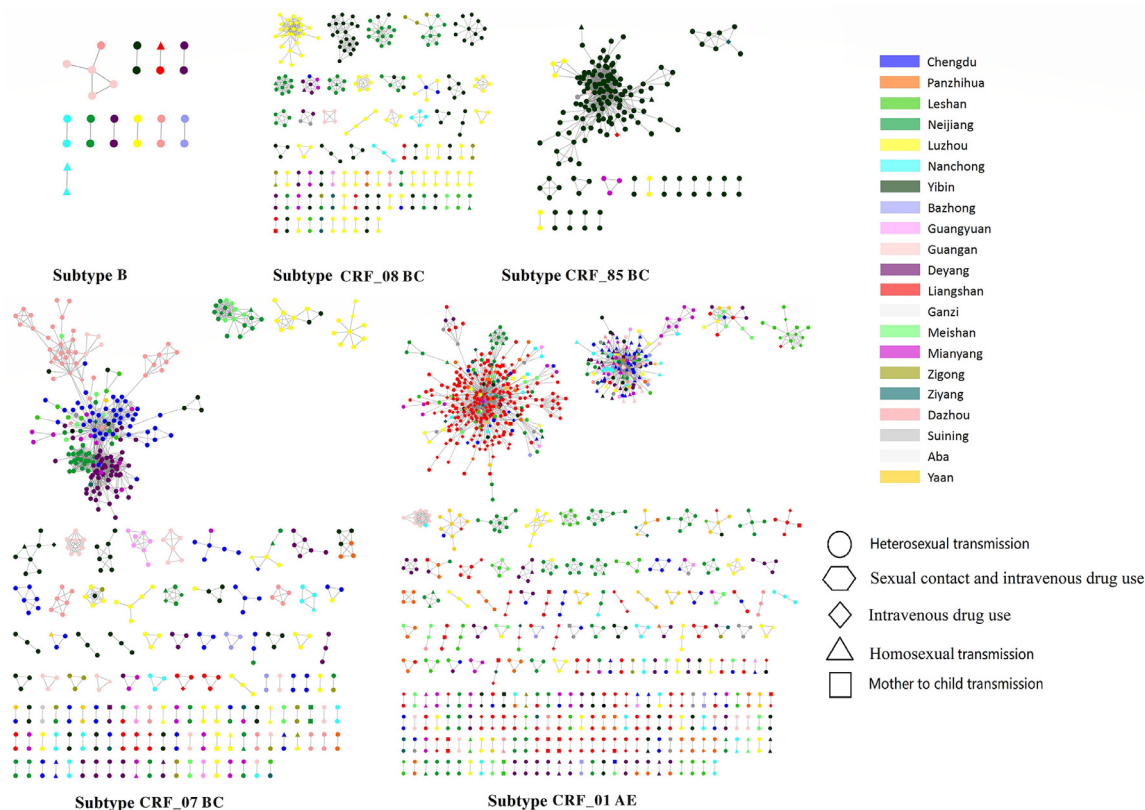

Figure 1. Genetic transmission networks of the HIV-1 subtypes. There were 11 clusters (2.4%) for subtype B, with sizes from 2 to 5 pol sequences; 80 clusters (17.7%) for subtype CRF_08 BC, with sizes from 2 to 26 pol sequences; 18 clusters (4.0%) for subtype CRF_85 BC, with sizes from 2 to 127 pol sequences; 120 clusters (26.6%) for subtype CRF_01 AE, with sizes from 2 to 206 pol sequences; and 223 clusters (49.3%) for subtype CRF_07 BC, with sizes from 2 to 342 pol sequences.

(80.9%), and then intravenous drug use (9.4%), homosexual (6.5%), mother-to-child (1.7%), and both sexual contact and intravenous drug use transmission (1.5%). However, the routes of HIV transmission were usually cross-linked. For example, of the 140 within-cluster PLWH reporting infection through homosexual contact, 114 (81.4%) were genetically linked to other 239 PLWH infected by heterosexual contact.

Determinants and spatial distribution of HIV-1 genetic transmission networks

There were 2,159 (37.3%) of 5,790 PLWH appearing within the genetic transmission networks (Table 1). Among them, most of the HIV-1 subtypes were CRF07_BC (51.3%) and CRF01_AE (26.2%). In general, PLWH were more likely to be within the phylogenetic clusters if aged ≥50 years (compared to those aged <15), divorced/widowed (compared to those married and living with spouse), educated (compared to illiteracy), and retired (compared to farmers); they were less likely to be within the phylogenetic clusters if being minorities (compared to the Hans), unmarried, and students or children. After controlling for these sociodemographic characteristics, some clinical and behavioral factors were associated with phylogenetic clustering, including a viral load of >50,000 copies/ml (OR=1.16, 95% CI=1.02-1.33), infection of other STDs (OR=1.38, 95%CI=1.12-1.69), and ≥5 non-marital sexual partners (OR=1.25, 95% CI=1.03-1.51), while >3 years of treatment since the initial ART was associated with less likelihood of phylogenetic clustering (OR=0.82, 95% CI=0.70-0.97).

For the routes of HIV transmission, significant heterogeneities were observed in different cities, especially Liangshan (located in the southwest of Sichuan Province, and being a gathering area of ethnic minorities) with the most diversity of transmission routes (Pielou index > 0.66 for PLWH within the genetic transmission networks) (Figure 2).

Spatial patterns of phylogenetic clustering rates

The phylogenetic clustering rates varied across cities, with the highest rates observed in the central regions (Deyang and Neijiang) and with rates in the east generally higher than in the west. All these subtypes showed relatively high phylogenetic clustering rates in the eastern part of Sichuan (Figure 3). The high phylogenetic clustering rates of CRF08_BC were mainly distributed in the central region, and of subtypes B and CRF85_BC were not as widely distributed as other subtypes.

Significant spatial clusters of high phylogenetic clustering rates were detected in the east (Dazhou for subtype B; Zigong and Luzhou for subtype CFR08_BC), and of low phylogenetic clustering rates were detected in the west (Yaan and Ganzi for subtype CRF07_BC) (Figure 3). Some cities with high phylogenetic clustering rates were surrounded by cities with low rates, such as Deyang and Neijiang for subtype B, Deyang for subtype CRF01_AE, Nanchong for subtype CRF08_BC. Also, some cities with low phylogenetic clustering rates were surrounded by cities with high rates, such as Nanchong for subtype CRF85_BC.

Transmission links within and between cities

The intensity of transmission links within and between cities was generally diverse for different subtypes. The intensity matrices showed strong intracity links in most cities (in darker blue along the diagonals of the intensity matrices), and strong intercity links in some regions (e.g., between Deyang and Neijiang) (Figure 4). Most HIV-1 subtypes were mainly transmitted within the cities, with some apparent exceptions, such as between Yibin and Guangan for subtype CRF85_BC. The proportion of intercity transmission varied across cities from 0.14 (Yibin) and 1.00 (Ganzi), being over 0.5 in 14 of the 21 cities (Table S1). The age, sex, and transmission route compositions of the PLWH also varied across cities (Tables S2-S4). The proportion of transmission through homosexual sex in intercity links (13.99%) was higher than that in all links (6.34%), while the proportion of transmission through heterosexual sex was lower in intercity links (73.72%) than in all links (82.38%) (Figure S1).

The spatial patterns of intercity transmission were different across subtypes. The subtype CRF07_BC was the most widespread subtype (Figure 5), with the strongest link observed between Neijiang and Liangshan, and between Guangan and Chengdu; the strongest intercity transmissions for CRF01_AE and CRF85_BC were found in Meishan-Neijiang-Deyang and Yibin-Guangan, respectively. For subtypes B and CRF85_BC, the intercity transmissions were generally weak, with slightly high transmission observed in Dazhou-Guangan and Guangan-Yibin, respectively. The subtype CRF85_BC, a novel HIV-1 subtype, showed strong transmission between Yibin and Guangan, which do not border each other (Figures 4 and 5). Transmission networks with strong links were also identified, such as among Deyang, Neijiang, Liangshan, and Mianyang.

Discussion

In this study, we found that, at the individual level, the higher viral load, infection of other STDs, more non-marital sexual partners, and the short period since the initial ART of HIV were associated with phylogenetic clustering in the genetic transmission networks. Furthermore, at the regional level, we found that the spatial distribution of phylogenetic clustering rates was clustered in the central part of Sichuan, especially for subtype CRF08_BC. The significant spatial clusters of high and low phylogenetic clustering rates

Variables	Number (%)		OR (95% CI)	
	Within networks (n=2,159)	**Total (n=5,790)**	**Unadjusted**	**Adjusted**[†]
Age (years)				
≤14	31 (1.4)	142 (2.5)	1.00 (Ref.)	-
15-49	993 (46.0)	3,406 (58.8)	1.47 (0.98-2.21)	-
≥50	1,135 (52.6)	2,242 (38.7)	**2.45 (2.55-5.51)**[**]	-
Sex				
Male	1,547 (71.7)	4,153 (71.7)	1.00 (Ref.)	-
Female	612 (28.3)	1,637 (28.3)	1.01 (0.89-1.13)	-
Ethnicity				
Han	1,785 (82.7)	4,172 (72.1)	1.00 (Ref.)	-
Minority	362 (16.8)	1,618 (27.9)	**0.38 (0.33-0.43)**[**]	-
Marital status				
Married (living with spouse)	1,176 (54.5)	3,115 (53.8)	1.00 (Ref.)	-
Unmarried	402 (18.6)	1,306 (22.6)	**0.73 (0.64-0.84)**[*]	-
Divorced/Widowed	581 (26.9)	1,369 (23.6)	**1.22 (1.07-1.38)**[**]	-
Educational level				
Illiteracy	443 (20.5)	1,395 (24.1)	1.00 (Ref.)	-
Primary school	935 (43.3)	2,353 (40.6)	**1.42 (1.23-1.63)**[**]	-
Middle school or higher	781 (36.1)	2,042 (35.3)	**1.33 (1.15-1.54)**[**]	-
Occupation				
Farmer	1,539 (69.2)	3,959 (68.4)	1.00 (Ref.)	-
Employed	512 (23.7)	1,526 (26.4)	0.98 (0.86-1.11)	-
Retired	59 (2.7)	123 (2.1)	**1.54 (1.07-2.20)**[*]	-
Student/Child	49 (2.3)	182 (3.1)	**0.61 (0.44-0.86)**[*]	-
Duration of treatment since initial ART (years)				
<1	410 (19.0)	1,000 (17.3)	1.00 (Ref.)	1.00 (Ref.)
1-3	993 (46.0)	2,459 (42.4)	0.95 (0.82-1.11)	1.01 (0.87-1.18)
>3	756 (35.0)	2,331 (40.3)	**0.69 (0.59-0.81)**[**]	**0.82 (0.70-0.97)**[*]
Viral load (copies/ml)				
1,000-10,000	701 (32.5)	2,084 (36.0)	1.00 (Ref.)	1.00 (Ref.)
>10,000-50,000	682 (31.6)	1,781 (30.8)	**1.22 (1.07-1.40)**[*]	1.11 (0.97-1.27)
>50,000	776 (35.9)	1,925 (33.2)	**1.33 (1.17-1.52)**[**]	**1.16 (1.02-1.33)**[*]
Infection of other STDs				
No	1,965 (91.0)	5,358 (92.5)	1.00 (Ref.)	1.00 (Ref.)
Yes	194 (9.0)	432 (7.5)	**1.41 (1.16-1.72)**[**]	**1.38 (1.12-1.69)**[**]
Number of syringe sharers				
0	1,970 (91.2)	5,055 (87.3)	1.00 (Ref.)	1.00 (Ref.)
1-4	121 (5.6)	488 (8.4)	**0.51 (0.42-0.64)**[**]	1.10 (0.86-1.41)
≥5	68 (3.1)	247 (4.3)	**0.60 (0.45-0.79)**[**]	1.36 (0.99-1.87)
Number of non-marital sex partners				
0	730 (39.1)	2,261 (39.1)	1.00 (Ref.)	1.00 (Ref.)
1-4	1,157 (53.6)	3,529 (60.9)	1.39 (1.24-1.56)	0.95 (0.84-1.08)
≥5	272 (12.6)	623 (10.8)	1.63 (1.36-1.95)	**1.25 (1.03-1.51)**[*]
Number of homosexual partners				
0	2,028 (93.9)	5,434 (93.9)	1.00 (Ref.)	1.00 (Ref.)
1-4	94 (4.4)	250 (4.3)	1.01 (0.78-1.32)	1.07 (0.81-1.42)
≥5	37 (1.7)	106 (1.8)	0.90 (0.60-1.35)	0.96 (0.64-1.46)

***Table 1*: Association of each predictor variable of interest with being within the genetic transmission networks**
[†] Adjusted for sex, age, ethnicity, marital status, and occupation and education levels. ART: antiretroviral therapy; STDs: sexually transmitted diseases.
* p<0.05;
** p<0.001

(a) For all participants
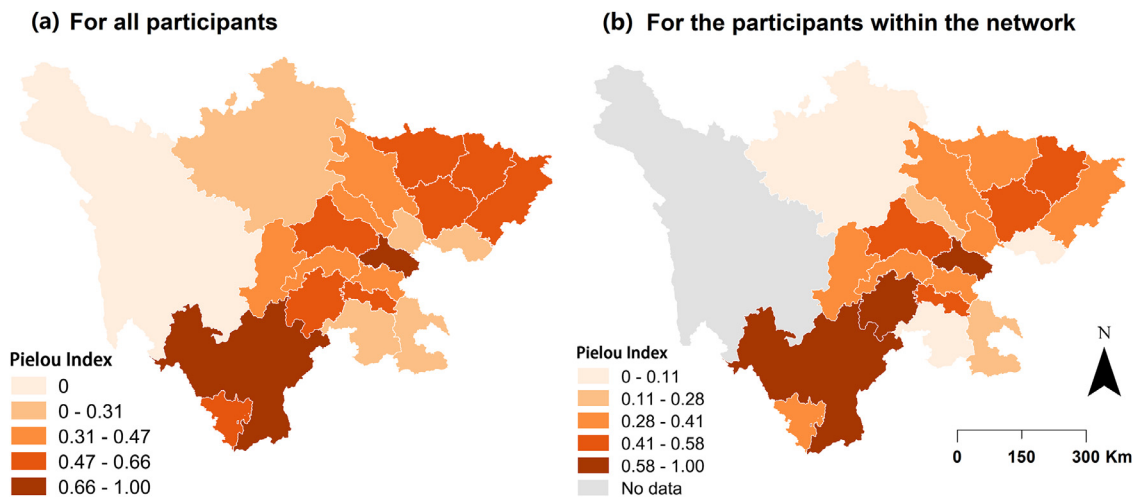
(b) For the participants within the network

Figure 2. Map of heterogeneities in HIV transmission routes, represented by the Pielou Index. A smaller value of the Pielou index represents more homogeneous routes of HIV transmission within a given city, while a larger value of the Pielou index represents more heterogeneous routes of HIV transmission.
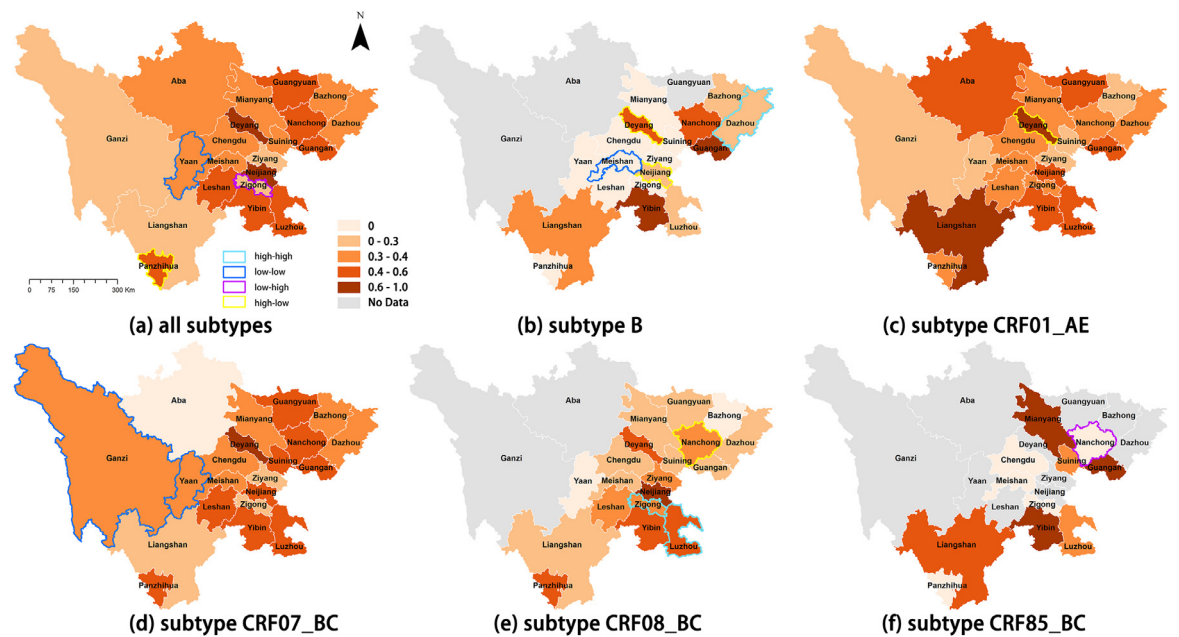

Figure 3. Spatial distribution and significant spatial clusters of the phylogenetic clustering rates of HIV-1 subtypes. A smaller (larger) value represents the lower (higher) phylogenetic clustering rate, while a larger value represents the higher phylogenetic clustering rate.

were detected in the east and west, respectively. The proportions of intercity transmission among all intracity and intercity transmission links varied across cities substantially. The subtype CRF07_BC was the most widespread subtype between cities, and subtype CRF85_BC showed strong intercity, and even multi-city, transmission.

High phylogenetic clustering rates of all HIV-1 subtypes were mainly distributed in the central and eastern regions of Sichuan Province. This might be due to human mobility, relatively developed traffic, and strong economic activity.[31] Note that regions with high phylogenetic clustering rates may not match regions with severe epidemics due to different percentages of PLWH
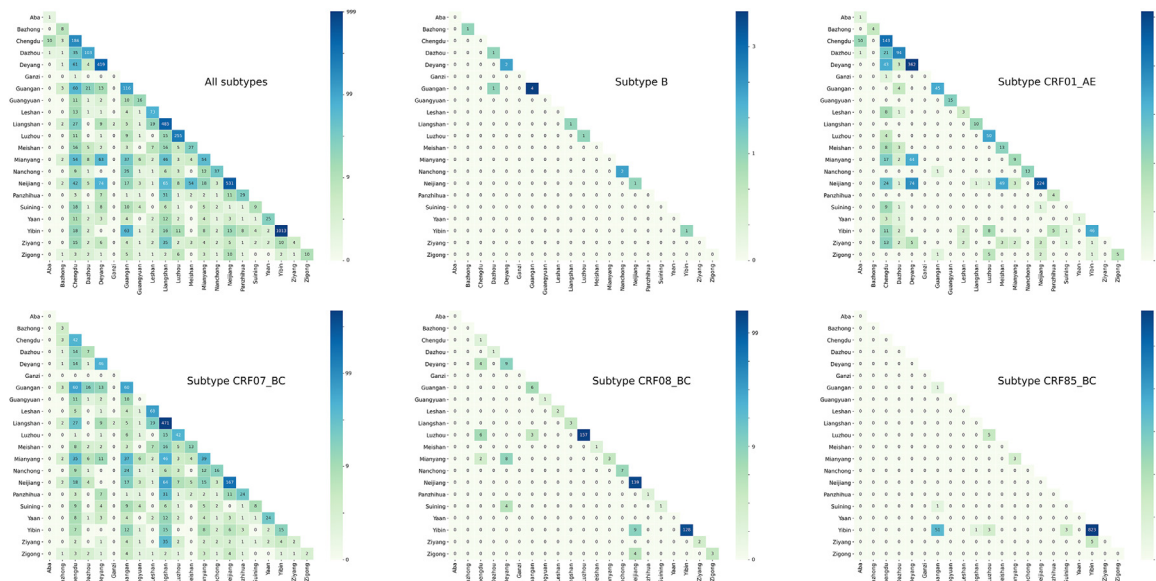
Figure 4. Intensity matrices of HIV-1 transmission links between cities in Sichuan (intensity matrices). The color of the grid cell at the intersection of two cities represents the number of linkages between the people living with HIV (PLWH) in two cities. A base-10 logarithmic color bar was used to enhance contrast.

with virological failure, which determined the size of the samples included in our study. For example, as the worst-hit city by HIV epidemic in Sichuan,[32] Liangshan (the largest traditional settlement of Yi minority in China) did not show the highest phylogenetic clustering rate. China's 13th five-year plan has strengthened the coverage of ART and increased the successful HIV suppression in Liangshan,[33] curbing the spread of HIV and thus may affect the phylogenetic clustering rate of virological failure in ART to some extent. Some cities with relatively less severe HIV epidemics showed higher phylogenetic clustering rates, such as Nanchong, located in the northeastern part of Sichuan Province. This highlights that regions without severe HIV epidemic may also deserve attention for possible high phylogenetic clustering rates, which could be revealed by spatial analyses. Furthermore, our results, along with previous findings, have demonstrated that spatial clustering can be used to suggest the relevance of intercity transmission.[4] For example, the Global Fund, the World Health Organization (WHO), and the Joint United Nations Program on HIV/AIDS (UNAIDS) have advocated geo-targeted HIV control intervention.[4] The cross-disciplinary methods of detecting regions with high phylogenetic clustering rates in our study undoubtedly provide new thinking on the design and optimization of geo-targeted interventions for lowering phylogenetic clustering rates.

There are at least three highlights in the intercity links of HIV infection, which would benefit from our combined analyses and deserve to be discussed for a

better understanding of the local situation of HIV infection. First, in addition to the high incidence rate, the cities with strong transmission (e.g., between Deyang and Neijiang) should be concerned. Especially, the transmission linkage between some cities geographically distant from each other was even stronger than between cities close to each other, which implies that local health sectors in those distant cities may increase information exchange for better launching joint interventions. Second, focused interventions should be taken based on the components of intercity transmission. For example, cities with the high proportion of intercity transmission (e.g., Ganzi) should focus on the prevention of virus inflow. Notably, we found Chengdu, the capital of Sichuan, showed similar proportions of intracity and intercity transmission, which suggests that HIV prevention efforts in provincial capitals or metropolises may need joint interventions both inside and outside (e.g., collaborating with other cities). Third, the epidemic dynamics of HIV-1 subtypes should be noticed. The flow maps in this study suggest that some small cities (e.g., Guangan) may be key sites of HIV transmission (especially for CRF07_BC, which was the most widespread subtype) and thus the frontline areas in need of urgent interventions. Further analyses should be conducted at finer scales by considering the transmission intensity among PLWH, to prioritize high-risk individuals for tailored interventions on the basis of individuals' characteristics, behaviors, and mobility patterns.

Some relatively novel subtypes of HIV-1 spread between cities in recent years, such as subtype
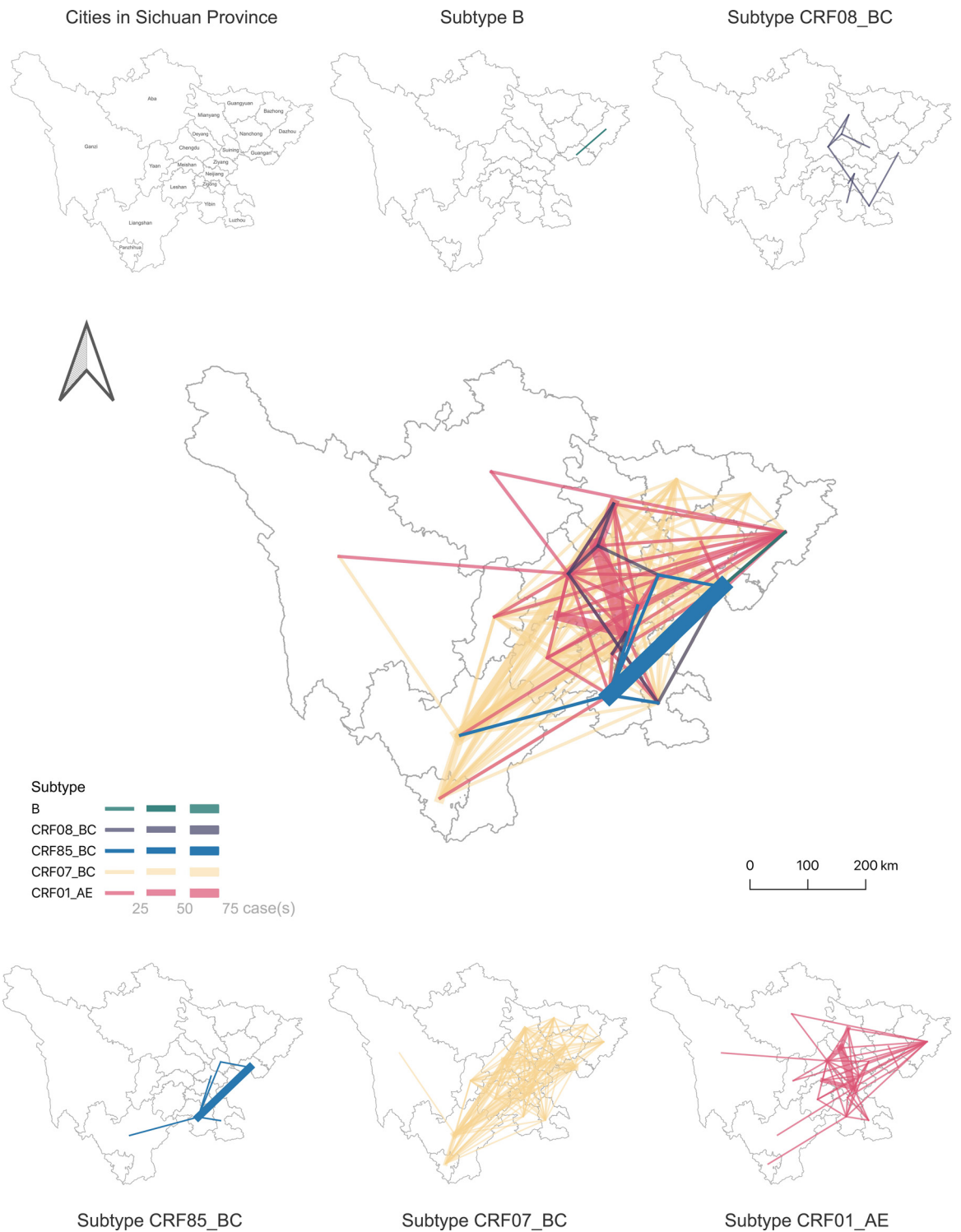
Figure 5. Flow map of intensity of HIV transmission linkages observed in the HIV-1 genetic transmission networks between cities in Sichuan. Each line represents the HIV transmission linkage between two cities, with each color of the line representing a different HIV-1 subtype and with the thickness of the line representing the number of linkages between the PLWH in two cities.

CRF85_BC.[34] Specifically, although severe epidemic dynamics of subtype CRF85_BC was mainly observed in Yibin, it also spread in some other cities (e.g., Guangan and Ziyang). Also, some cities (e.g., Zigong) in Sichuan showed an increased number of elderly PLWH mainly infected by non-married heterosexual transmission in recent years,[35] which might be attributed to the migration of high-risk populations (e.g., sex workers). Since subtype CRF85_BC is mainly found in elderly people infected by heterosexual transmission as a previous report,[34] elderly PLWH infected by non-marital heterosexual transmission within regions with strong intercity links (e.g., Yibin) might play a central role in the spatial flow of subtype CRF85_BC. Certainly, there may be more fundamental social factors associated with the spatial flow of the virus. Just as a previous study revealed, the heterogeneity in the health of urban dwellers, increased rates of contact, and mobility of people results in a high risk of disease transmission in large urban populations.[36] Cities become incubators where all conditions are met for outbreaks to occur, and migration and urbanization may lead to strengthened disease transmission links between cities in the future.[36] In sum, our findings highlight the necessity of integrating epidemiological, genetic, and spatial analyses, so that to identify the targeted hotspots and spatial flow of the virus, with consideration of the population characteristics for accurate identification of the population in key regions needs to be intervened. Especially for novel subtypes that are about to spread across cities, the transmission patterns could be analyzed by cross-disciplinary approaches.

This study used cross-disciplinary (epidemiological, genetic, and spatial) approaches to depict the HIV-1 genetic transmission and HIV intercity transmission networks. The strengths of this study include the identification of 1) the high-risk populations, especially transmission from virally unsuppressed patients, which is different from the majority of clustering analysis usually based on all PLWH and particularly useful for HIV prevention among disproportionately higher-risk populations; 2) some linkages of patients unrevealed by traditional epidemiological surveys, which could guide more targeted and customized retrospective surveys to find out forgotten facts due to recall bias; 3) intracity and intercity links of HIV transmission; and 4) the flow trend of HIV-1 subtypes. There are important public health implications from this study. We found that the proportion of homosexual sex in intercity transmission links was higher than that in all transmission links; in contrast, the proportion of heterosexual sex in intercity transmission links was lower than that in all transmission links. The more intercity transmission links associated with homosexual sex may reflect the frequent mobility and possibly multiple sexual partners in MSM, and the wide application of Gay mobile apps may allow

convenient seeking of sexual partners from different cities.[17,37] These detected patterns suggest that the population with specific transmission route may have specific spatial behavioral patterns, which may provide new clues for interventions for higher-risk populations.

Our study still has several limitations. First, some spatial characteristics of HIV-1 subtypes dynamically changed, with the HIV-1 subtypes spreading from the original regions to other regions. However, our study cannot capture the mobility patterns. Also, drug-induced HIV mutations may cause bias in our sequence analysis, which would be overcome by collecting sequences before ART to improve the surveillance of HIV transmission. Second, although important for designing precise HIV intervention methods and strategies, HIV genetic transmission networks have some limitations, such as the large number of HIV sequences and the high phylogenetic clustering rates required for constructing an effective transmission network, and the difficulty in clarifying the order of transmission in the networks. Also, the structure of the transmission network and spatial patterns of phylogenetic clustering rates may vary by the genetic distance threshold, which is usually selected for specific purposes of different studies. Third, the transmission network constructed from phylogenetic data alone, due to uncertain sampling density, may not reflect the actual intercity or intracity links of HIV transmission, which would be better identified by supplemental mobility information of PLWH from traditional epidemiological surveys. Lastly, we used province-wide data collected from various health facilities and all participants were PLWH with virological failure of ART, so the results may not be generalized to the whole population without caution. Moreover, the sample size in this study is limited and the situation in Sichuan Province could not represent the overall situation in China. In future studies, higher phylogenetic clustering rates and more advanced methods to construct such an HIV genetic transmission network are expected to improve the accuracy of the transmission links among regions.

Conclusions

Understanding the pathways via which HIV-1 subtypes have been transmitted is crucial for precisely targeting limited health resources to the regions of greatest need and guiding relevant policy-making. This study demonstrates the use of a novel combination of cross-disciplinary (epidemiological, genetic, and spatial) approaches to assist in achieving precision public health: at the individual level, genetic transmission network analysis was used to infer the dynamics of HIV transmission and identify the high-risk populations; at the regional level, the spatial distribution and clustering patterns of the phylogenetic clustering rates and spatial connections of

HIV-1 between cities revealed the geographic variations in HIV-1 transmission and high-risk areas, strengthening the individual-level evidence for targeting interventions. Future efforts could consider more spatial big data, such as mobile phone signaling data, social media and internet-based data, migration data and traffic flow data, to deepen our understanding of the interaction among host, agent, and environment in the HIV context,[38] as well as to better predict the individual's risk of infection at any time over the life course in the context of spatial lifecourse epidemiology.[39,40] Such efforts in regions where data are available would be able to accurately estimate the dynamic patterns of HIV-1 transmission in those regions, and also be extrapolated to other regions where relevant data are not available, in order to provide better solutions to precise HIV prevention and control in those regions.

Authors' contribution
SY and PJ conceived the idea. SY, PJ and TF wrote the study protocol and designed data analysis. SL, DY, HT, LY and YL contributed to data collection, management and provided laboratory support. DY, BY, RK, TF and SY verified the underlying data. SL, DY, LY and YL conducted genetic transmission network analysis. TF, RK and PJ conducted spatial analysis. SY, PJ and TF supervised the study. BY, TF, SY and DY drafted the manuscript. PJ, SY, TF, BY and DY revised the manuscript. All authors read and approved the final manuscript.

Data Availability
The data was collected by experienced scientists and health personnel. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Editor note: The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of Competing Interest
We declare no competing interests.

Supplementary materials
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.lanwpc.2021.100318.

References

1 Joint United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS Data 2017.

2 Gibney KB, Cheng AC, Hall R, Leder K. Sociodemographic and geographical inequalities in notifiable infectious diseases in Australia: a retrospective analysis of 21 years of national disease surveillance data. *The Lancet Infectious Diseases* 2017;**17**(1):86–97.

3 Dowdy DW, Golub Je Fau - Chaisson RE, Chaisson Re Fau - Saraceni V, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *(1091-6490 (Electronic))* 2021.

4 Ratmann O, Kagaayi J, Hall M, et al. Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: a population-based study in Rakai. *Uganda. The Lancet HIV* 2020;**7**(3):e173–e83.

5 Wirtz AL, Trapence G, Kamba D, et al. Geographical disparities in HIV prevalence and care among men who have sex with men in Malawi: results from a multisite cross-sectional survey. *The Lancet HIV* 2017;**4**(6):e260–e9.

6 Anderson SJ, Cherutich P, Kilonzo N, et al. Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study. *(1474-547X (Electronic))* 2021.

7 Jia P, Yang S. Time to spatialise epidemiology in China. *The Lancet Global Health* 2020;**8**(6):e764–e5.

8 Xiao P, Li J, Fu G, Zhou Y, Huan X, Yang H. Geographic Distribution and Temporal Trends of HIV-1 Subtypes through Heterosexual Transmission in China: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health* 2017;**14**(7):830.

9 Lu L, Jia M, Ma Y, et al. The changing face of HIV in China. *Nature* 2008;**455**(7213):609–11.

10 Brenner B, Wainberg MA, Roger M. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. *AIDS* 2013;**27**(7):1045–57.

11 Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjornsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS* 2017;**31**(9):1211–22.

12 Wertheim JO, Kosakovsky Pond SL, Forgione LA, et al. Social and Genetic Networks of HIV-1 Transmission in New York City. *PLoS Pathog* 2017;**13**(1):e1006000.

13 Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. *The Lancet HIV* 2019;**6**(3):e164–e72.

14 Dalai SC, Junqueira DM, Wilkinson E, et al. Combining Phylogenetic and Network Approaches to Identify HIV-1 Transmission Links in San Mateo County. *California. Front Microbiol* 2018;**9**:2799.

15 Yan H, He W, Huang L, et al. The Central Role of Nondisclosed Men Who Have Sex With Men in Human Immunodeficiency Virus-1 Transmission Networks in Guangzhou. *China. Open Forum Infect Dis* 2020;**7**(5):ofaa154.

16 Yu B, Zhou J, Gong Y, et al. Self-efficacy Mediates Perceived Benefits and Barriers of Adherence of Heroin-dependent Patients to Methadone for Addiction Treatment: A Health Belief Model Study. *J Addict Med* 2020;**14**(4):e110–e7.

17 Yuan D, Du Z, Zhou J, et al. HIV-1 subtype diversity, drug resistance, and genetic transmission networks in men who have sex with men with virologic failure in antiretroviral therapy in Sichuan, China, 2011 to 2017. *Medicine (Baltimore)* 2019;**98**(43):e17585.

18 Yang C, Latkin C, Luan R, Nelson K. Peer norms and consistent condom use with female sex workers among male clients in Sichuan province. *China. Soc Sci Med* 2010;**71**(4):832–9.

19 National Center for AIDS/STD Control & Prevention. National Guideline for Detection of HIV/AIDS. http://ncaids.chinacdc.cn/tzgggd/202005/t20200518_216798.htm. Accessed 18 May 2020.

20 Wheeler WH, Ziebell RA, Zabina H, et al. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. *Aids* 2010;**24**(8):1203–12.

21 Pines HA, Wertheim JO, Liu L, Garfein RS, Little SJ, Karris MY. Concurrency and HIV transmission network characteristics among MSM with recent HIV infection. *AIDS* 2016;**30**(18):2875–83.

22 Fabeni L, Rozera G, Berno G, et al. Molecular Transmission Dynamics of Primary HIV Infections in Lazio Region, Years 2013-2020. *Viruses* 2021;**13**(2):176.

23 Shao Y. HIV drug resistance monitoring strategies and detection techniques. Beijing: People's Medical Publishing House; 2010.

24 US-CDC. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. *Detecting and Responding to HIV Transmission Clusters: A Guide for Health Departments* 2018. Avaliable at https://www.cdc.gov/hiv/pdf/funding/announcements/ps18-1802/CDC-HIV-PS18-1802-AttachmentE-Detecting-Investigating-and-Responding-to-HIV-Transmission-Clusters.pdf. Accessed 17 October 2018.

25 Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**(11):2498–504.

26 Chow EPF, Grulich AE, Fairley CK. Epidemiology and prevention of sexually transmitted infections in men who have sex with men at risk of HIV. *Lancet HIV* 2019;**6**(6):e396–405.

27 Poon AF, Gustafson R, Daly P, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV* 2016;**3**(5):e231–8.

28 Zhang L, Chow EP, Jing J, et al. HIV prevalence in China: integration of surveillance data and a systematic review. *Lancet Infect Dis* 2013;**13**(11):955–63.

29 Moran PA. Notes on continuous stochastic phenomena. *Biometrika* 1950;**37**(1-2):17–23.

30 QGIS Development Team. QGIS Geographic Information System, Open Source Geospatial Foundation, 2009. Avaliable at: http://qgis.org. Accessed 15 October 2021.

31 Eshraghian EA, Ferdos SN, Mehta SR. The Impact of Human Mobility on Regional and Global Efforts to Control HIV Transmission. *Viruses* 2020;**12**(1):67.

32 Yuan D, Liu M, Jia P, et al. Prevalence and determinants of virological failure, genetic diversity and drug resistance among people living with HIV in a minority area in China: a population-based study. *BMC Infect Dis* 2020;**20**(1):443.

33 National Development and Reform Commission (NDRC), People's Republic of China. The 13th five-year plan for economic and social development of the people's republic of China. https://en.ndrc.gov.cn/policyrelease_8233/201612/P020191101482242850325.pdf. Accessed 16 Mar 2016.

34 Su L, Feng Y, Liang S, et al. The origin and spread of CRF85_BC, driven by heterosexual transmission among older people in Sichuan, China. *BMC Infect Dis* 2020;**20**(1):772.

35 He X, CY Wang D, Zhang Y. Analysis of HIV/AIDS Surveillance Data in Zigong, 2008-2013. *Journal of Occupational Health and Damage* 2016.

36 Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect Dis* 2011;**11**(2):131–41.

37 Cao B, Liu C, Stein G, et al. Faster and Riskier? Online Context of Sex Seeking Among Men Who Have Sex With Men in China. *Sex Transm Dis* 2017;**44**(4):239–44.

38 Jia P, Dong W, Yang S, Zhan Z, Tu L, Lai S. Spatial Lifecourse Epidemiology and Infectious Disease Research. *Trends Parasitol* 2020;**36**(3):235–8. https://doi.org/10.1016/j.pt.2019.12.012.

39 Jia P. Spatial lifecourse epidemiology. *The Lancet Planetary Health* 2019;**3**(2):e57–9. https://doi.org/10.1016/S2542-5196(18)30245-6.

40 Jia P, Lakerveld J, Wu J. Top 10 Research Priorities in Spatial Lifecourse Epidemiology. *Environ Health Persp* 2019;**127**(7):74501. https://doi.org/10.1289/EHP4868.