

Exploring and Predicting the Drivers of Ongoing HIV-I Transmission in Guangyuan, Sichuan

Yan Zhang¹, Haolin Jiang², Wenkai Xiang³, Jun Zhu³, Xueqin Hou³, Shu Liang¹, Dan Yuan¹, Chang Zhou¹, Ling Su¹

¹Center for AIDS/STD Control and Prevention, Sichuan Center for Disease Control and Prevention, Chengdu, People's Republic of China;

²Department of Epidemiology and Statistics, Chengdu Medical College, Chengdu, People's Republic of China; ³Center for AIDS/STD Control and Prevention, Guangyuan Center for Disease Control and Prevention, Guangyuan, People's Republic of China

Correspondence: Ling Su, Center for AIDS/STD Control and Prevention, Sichuan Center for Disease Control and Prevention, No. 6, Middle School Road, Wuhou District, Chengdu, Sichuan Province, 610000, People's Republic of China, Tel +8618982092195, Email sl7512@163.com

Purpose: Guangyuan was selected as the first pilot city of molecular transmission network in Sichuan Province to implement dynamic monitoring. This study aim to insight the characteristics of HIV-1 molecular epidemiology and explore the influencing factors of transmission dynamics. Furthermore, it predict the driving factors of network expansion by established a transmission risk prediction model.

Patients and Methods: A longitudinal cohort study was conducted to obtain a total of 1434 plasma samples from newly diagnosed HIV-infected patients from 2010 to June 2022. Phylogenetic relationship and cluster analysis were performed using HIV-1 polymerase (*pol*) gene sequences to study the risk factors of clustering. We applied Logistic ML algorithms to establish a transmission risk prediction model, and model performance was checked using 10-fold cross-validation in the training set and receiver operating characteristic (ROC) curve analysis.

Results: A total of 1360 *pol* sequences linked demographics obtained in this study cover approximately 94.8% of newly notified infections from 2010 to June 2022. The major epidemic genotypes were CRF07_BC, CRF01_AE, CRF08_BC and B subtypes, accounting for 93.82% of all. The differences of some clinical and demographic factors (eg, age, marital status) were statistically significant ($P<0.05$). We identified 136 clusters containing 654 HIV-1 *pol* sequences and observed that some characteristics (eg, over 50 years, married) were more likely to associated to the clusters ($P<0.05$). The predictive model showed excellent predictive ability to forecast cluster growth.

Conclusion: The epidemic genotypes were relatively complex and diverse in Guangyuan. There was a potential transmission association caused widely spread in local area after the new strains entering. The transmission risk prediction model showed excellent predictive ability to forecast cluster growth which can predict the risk factors causing clusters expansion and provide a guidance for precise intervention strategies.

Keywords: HIV-1, genotype, cluster analysis, transmission risk prediction model

Introduction

In recent years, the number of newly reported HIV/AIDS cases per year has been increasing in Sichuan province.¹ Although it has achieved remarkable results by constantly exploring and innovating prevention and control strategies, the number of surviving HIV/AIDS cases in Sichuan province has exceeded 210,000 as of December 31, 2022 according to the statistics of the National AIDS Comprehensive Prevention and Control Information System, the proportion of newly reported HIV/AIDS cases over 50 years old has been increasing in recent years. The number of newly reported HIV/AIDS cases in 2022 was 19,018, of which 60.47% were infected individuals aged 50 and above, 89.33% were infected by heterosexual transmission. There was still a long way to go in HIV prevention and control. Guangyuan was a low-epidemic city that located in the northeast of Sichuan Province,² but the ratio of homosexual transmission was far higher than the average level of the Sichuan province.³ In combination with the goals of “three 90%” proposed by UNAIDS⁴ and China's 13th Five-Year Plan for prevention and control of AIDS, China regards expanding detection coverage and finding out the HIV-infected individuals maximally as one of the main measures of AIDS prevention and control, in order to continue to control the epidemic of AIDS

at a low level. In 2018, Sichuan province launched the prevention and control project in tackling HIV/AIDS, taking detection and discovery as the key strategy of HIV/AIDS prevention and control. Various measures and approaches were adopted to expand the coverage of HIV/AIDS detection, such as incorporating HIV detection into pre-marital check-up and health check-up of service personnel in key public places, promoting voluntary detection of healthy people in combination with free health check-up for all, conducting Provider-initiated HIV Testing and Counseling (PITC). These measures can help to treat and manage patients as soon as possible to reduce transmission. Diagnosis and detection continues to improve with the comprehensive expansion of screening coverage.⁵ Within the outbreak of COVID-19 in China, Sichuan Province adopted a series of prevention and control measures at the end of December 2019, such as initiating the first-level emergency response to public health emergencies, banning mass gathering activities, and implementing the closed management of residential communities,^{6,7} which had a great impact on the production and live of residents. As the staff of grassroots CDC focused on the prevention and control of COVID-19 and the decreased of patient visits, the number of HIV/AIDS cases reported in the first quarter of 2020 in Sichuan province was reduced significantly.⁸ It was particularly important to update prevention and control strategies in a timely manner. As we all know, the spread of AIDS depends on the contact of individuals in social networks, and the flow of HIV-infected-patients in social networks affects the direction of HIV infection.⁹ It was time-consuming to conduct field epidemiological investigation; therefore, molecular epidemiology has played an increasingly important role in prevention and research of AIDS in recent years. In combination with molecular transmission network and epidemiology, we use the theories related to evolutionary dynamics and molecular biology to construct a transmission network with potential correlation through the genetic information of HIV-infected patients. We analysed the transmission mechanism of HIV from the perspective of groups and find more hidden high-risk groups to provides new ideas and methods for precise prevention and control of AIDS. As a pilot city of molecular transmission network in Sichuan Province, continuous dynamic surveillance has been carried out since 2016. The study aim to insight the characteristics of HIV-1 molecular epidemiology and explore the influencing factors of transmission dynamics by combining molecular transmission network with epidemiological investigation, and followed the transmission dynamics in these clusters across a 12 years time period. Additionally, we established a transmission risk model to predict the driving factors of network expansion, which can identify high-risk groups related to molecular transmission networks.¹⁰ It provides an important basis for discovering the law of HIV transmission and formulating precise prevention and control strategies in the further.

Materials and Methods

Study Participants

A total of 1434 blood samples of the newly diagnosed HIV-infected patients, which were the residual plasma from HIV-1 diagnostic tests between 2010 and June 2022 in Guangyuan. The epidemiological information was obtained from the China Information System for Disease Prevention and Control of AIDS, such as gender, age, marital status, education level, route of transmission, contact history and other basic demographic characteristics.

The study was reviewed and approved by the Ethics Committee of the Sichuan Center for Disease Control and Prevention and conducted following the Helsinki Declaration of 1964.

Methods

RNA Extraction, Amplification and Sequencing

Five millilitres of venous blood was extracted with EDTA-K2 anticoagulant vessel, and the plasma was separated and stored at -80°C then be transported to the Sichuan Center for Disease Control and Prevention by cold chain. The Nucleic Acid of the virus was obtained from 200 μL plasma by using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche, Branchburg, NJ). The reverse transcription polymerase chain reaction (RT-PCR) was performed using QIAGEN one-step RT-PCR kit and Taq PCR Master Mix kit to amplify the full-length protease gene in the *pol* region and the first 300 codons of the reverse transcriptase gene. The amplification primers and reaction conditions were reported in literature.¹¹ The PCR products were dealt with 1% agarose-gel electrophoresis, and the amplified positive products were purified and sequenced by Beijing Human Genome Research Center. The *pol* region sequence of HIV (1060 base pairs, HXB2:2254–3313) was obtained.

Sequence Analysis and Genotype Determination

Sequencher 5.1 was used to splice the sequence fragments, the BioEdit 7.1 was used to edit and correct the sequence.¹² Clustal W program in MEGA7.0 was used to perform multiple comparison between the assemble sequences and international reference sequences¹³ to establish sample sequence data sets. Reference sequences were downloaded from the Los Alamos National Laboratory HIV Sequence Database (<https://www.hiv.lanl.gov>), including major HIV-1 subtypes and regroups. Fasttree software was used to construct an approximate maximum likelihood (ML) phylogenetic tree for *pol* sequences using the GTR+G+I nucleotide substitution model.¹⁴ The Shimodaira–Hasegawa (SH) test was used to calculate the node (branch point) value of evolutionary tree, and the SH-like support $\geq 85\%$ was considered as the criterion for determine genotype.¹⁵ The online analysis tool of HIV Databases BLAST (<https://www.hiv.lanl.gov/content/index>) was used to check the results. If the genotype of sequence could not be determined by phylogenetic tree and HIV Databases BLAST, then was considered as URFs.

The Definition of Cluster

The pairwise distances between sequences were based on the fact that the variation degree of *pol* gene region in single infected individuals is less than 1% within 3 years after infection.^{16,17} The pairwise distances less than 1% were selected to perform network visualization by Cytoscape v3.7.1 according to the previous study of our laboratory.¹⁸ A node represents an HIV-infected individuals.¹⁹ Two or more nodes can form a cluster.

Statistical Analysis

In continuous parameters following normal distribution were described by mean SD. Categorical parameters were presented as numerical values and proportions. Chi-square tests, Fisher's exact tests, *t*-tests and logistic regression analysis were performed using SAS 9.4 (Version 9.4, SAS Institute, Inc., Cary, NC, USA). The difference was statistically significant when $P < 0.05$ with bilateral test. ML algorithms was also performed with the help of SAS 9.4 R (Version 4.1.1). The pairwise distances between sequences were calculated by using HyPhy2.2.1 in Tamura-NEI93 model.

“Whether clustering or not” was taken as the dependent variable in univariate and multivariate analysis, while age, gender, ethnicity, marital status, education level, the first CD4+T cells, current address, genotype, the time of notified infection and transmission route were taken as covariates.

The 07BC1 and AE4 clusters were used as dependent variables for transmission risk prediction to investigate the transmission risk factors of 07BC1 and AE4 clusters in Guangyuan. Variables were selected by the Enter method of Logistic regression model and the inclusion criterion was set as 0.05 while the exclusion criterion was 0.15. Hosmer-Lemeshow (H-L test) was used to evaluate the fitting effect of the model, and model performance was checked using 10-fold cross-validation in the training set and test set ROC curve analysis. The significance level was set at 0.05.

Results

Demographic Characteristics and Distribution of HIV-I Genetic Forms

A total of 1360 *pol* sequences with linked demographics obtained in this study cover approximately 94.8% of newly notified infections in Guangyuan from 2010 to June 2022. The major epidemic genotypes were CRF07_BC, CRF01_AE, CRF08_BC and B subtypes, accounting for 93.82% of all.

The proportion of CRF07_BC in HIV-infected individuals under 25 years old was 56.50% (113/200), which was significantly higher than those aged 26–49 years old and over 50 years old. As age increases, the proportion of CRF07_BC decreases gradually. Among HIV-infected individuals over 50 years old, the proportions of CRF01_AE and CRF08_BC were 38.25% (197/515) and 14.37% (74/515), respectively, which showing a significantly higher prevalence compared to those under 25 years old. Additionally, the proportion of CRF07_BC was significantly higher in male patients (60.58%; 356/587) compared to female patients (43.90%; 111/253). The proportion of CRF07_BC in unmarried patients was 54.31% (208/383), which was significantly higher than those in married (42.37%; 322/760). While the CRF01_AE and CRF08_BC in married accounted for 35.39% (269/760) and 12.63% (96/760) respectively, which showing a significantly higher prevalence compared to unmarried. The proportion of CRF07_BC was 78.85% (41/52) in other nationalities except Han, which was more than 3/4 of all subtypes. Besides, from the perspective of transmission route, the dominant epidemic subtype was CRF07_BC in intravenous drug users (IDUs) and MSM, accounting for 78.95% (30/38) and 58.09% (140/241), respectively. However, the proportion of CRF01_AE and CRF08_BC

were 35.07% (377/1075) and 11.53% (124/1075) in the population infected by heterosexual transmission, which were significantly higher than the proportion of these two subtypes in the IDUs and MSM. The main subtype of infection caused by mother-to-child transmission (MCT) was CRF07_BC in Guangyuan, accounting for 83.33% (140/241). The proportion of CRF07_BC and CRF08_BC in downtown of Guangyuan were significantly higher than that in suburbans, which were 51.74% (208/402) and 13.93% (56/402), respectively. CRF07_BC accounted for more than half of the epidemic strains in downtown. The proportion of CRF01_AE in suburbans was 37.35% (347/929), which was significantly higher than those in downtown. In addition, the proportion of CRF07_BC in detention house was 82.76% (24/29), which was more than 3/4 of all subtypes in detention house. The CRF07_BC was 60.56% (109/180) in proportion of HIV-infected patients with the first CD4+T cells ≥ 500 cells/ul, and the proportion gradually decreased with the decrease of the first CD4+T cells. The CRF01_AE and CRF08_BC in HIV-infected patients with the first CD4+T cells ≤ 200 cells/ul accounted for 40.33% (173/429) and 13.05% (56/429), respectively, which was significantly higher than those with the first CD4+T cells ≥ 500 cells/ul, and the proportion gradually decreased with the first CD4+T cells increasing. There were 60.29% (820/1360) of the HIV-infected patients were found in medical institutions, 31.03% (422/1360) of them were from the surveillance of key populations, and about 4% were from custodial population and other sources respectively. Among them, the detection ratio of CRF07_BC was higher in high-risk population (52.13%; 220/422) and custodial population (78.57%; 44/56). The detection rates of CRF01_AE and CRF08_BC in medical institutions were higher than those from other sources. According to occupational classification, the highest detection rate was 55.88% (760/1360) in farmers/migrant workers, and the proportion of CRF01_AE was higher than that in housework/unemployed and other occupations, while the proportion of CRF07_BC was relatively higher in housework/unemployed (52.27%; 138/264) and other occupations (55.06%; 185/336), which was significantly higher than those in farmers/migrant workers, as shown in Table 1.

Risk Factors of Clustering Analysis

We inferred a partial HIV transmission network from HIV-infected individuals in Guangyuan between 2010 and June 2022. Molecular transmission cluster analysis identified 136 clusters containing 654 HIV-infected individuals (654/1360, 48.1%), the cluster sizes ranged from 2 to 130. There were 7 clusters with more than 10 cases, 5 clusters with size ≥ 5 cases, 34 clusters with 3–4 cases, and 90 pair-to-pair transmission clusters. Comparing the characteristics of individuals involved in clusters with those out of clusters, gender, age, marital status, ethnicity, route of transmission, the time of notified infection, current address, subtype, and sample source were influential factors for clustering by univariate analysis. Female (OR: 1.19, 95% CI: 1.00–1.42), over 50 years old (OR: 1.54, 95% CI: 1.20–1.98), married (OR: 1.54, 95% CI: 1.20–1.98), current address in downtown (OR: 1.27, 95% CI: 1.08–1.49), CRF08_BC subtype (OR: 1.36, 95% CI: 1.06–1.74), and high-risk population (OR: 1.21, 95% CI: 1.03–1.43) were significantly associated to clusters ($P < 0.05$) (Table 2). The females accounted for only 23.60% of the total HIV-infected patients, but 54.83% of them were entered clusters, and the risk of clustering was 1.19 times that of males. Patients aged under 25 years old and over 50 years accounted for 14.71% and 37.87% of the total HIV-infected patients, but the proportion of clustering was 38.5% and 59.22%, respectively. The patients over 50 years old had a 1.54 times greater risk of clustering than under 25 years old. Compared with unmarried, cluster risk of married was 1.38 times. In addition, a majority of the infected patients in Guangyuan were Han nationality and the cluster risk was higher than that of others nationality. The majority of the others nationality were Yi nationality. Cluster risk of patients in downtown was 1.27 times that of suburbans. Cluster risk of patients notified infection between 2020 to June 2022 was 0.73 times that of notified infection before 2017. In terms of subtypes, CRF08_BC accounted for 9.41% of the total cases only, but 61.72% of the cases entered clusters, and the cluster risk was 1.36 times that of CRF07_BC. Secondly, there were 60.29% cases detected from medical institutions, of which only 45.49% entered clusters, while the high-risk population accounted for 55.21% of the transmission cluster but 31.03% of the total cases. However, the cluster risk of the high-risk population was 1.21 times that of the population detected from medical institutions.

Multivariate analysis showed that age, the time of notified infection, current address and sample source all affected the formation of clusters. The cluster risk of patients aged over 50 years old was 1.47 times that of under 25 years old, and high-risk populations were 1.27 times higher than that of those detected from medical institutions. Compared with

Table 1 General Distribution Characteristics of the Epidemic Strains in Guangyuan

Characteristics	Total(%) N is Equal to 1360	CRF07_BC	CRF01_AE	CRF08_BC	B	Others	χ^2	P
Gender								
Male	1039(76.40)	495(47.64)	354(34.07)	84(8.08)	42(4.04)	64(6.16)	1.819	0.177
Female	321(23.60)	145(45.17)	98(30.53)	44(13.71)	14(4.36)	20(6.23)		
Age, years								
≤25	200(14.71)	113(56.50)	55(27.50)	6(3.00)	8(4.00)	18(9.00)	28.824	<0.001***
26–49	645(47.43)	335(51.94)	200(31.01)	48(7.44)	29(4.50)	33(5.12)		
≥50	515(37.87)	192(37.28)	197(38.25)	74(14.37)	19(3.69)	33(6.41)		
Marital Status								
Unmarried	383(28.16)	208(54.31)	114(29.77)	13(3.39)	18(4.70)	30(7.83)	13.666	0.001**
Divorced or widowed	217(15.96)	110(50.69)	69(31.80)	19(8.76)	8(3.69)	11(5.07)		
Married	760(55.88)	322(42.37)	269(35.39)	96(12.63)	30(3.95)	43(5.66)		
Ethnicity								
Han	1308(96.18)	599(45.80)	447(34.17)	126(9.63)	54(4.13)	82(6.27)	—	<0.001***
Others	52(3.82)	41(78.85)	5(9.62)	2(3.85)	2(3.85)	2(3.85)		
Route of Transmission								
Hetero	1075(79.04)	465(43.26)	377(35.07)	124(11.53)	43(4.00)	66(6.14)	—	<0.001***
MSM	241(17.72)	140(58.09)	71(29.46)	2(0.83)	12(4.98)	16(6.64)		
IDUs	38(2.79)	30(78.95)	4(10.53)	2(5.26)	0(0.00)	2(5.26)		
MCT	6(0.44)	5(83.33)	0(0.00)	0(0.00)	1(16.67)	0(0.00)		
Education Level								
College degree or above	164(12.06)	93(56.71)	46(28.05)	4(2.44)	7(4.27)	14(8.54)	7.218	0.027*
High school or technical secondary school	284(20.88)	140(49.30)	95(33.45)	16(5.63)	12(4.23)	21(7.39)		
Middle school or below	912(67.06)	407(44.63)	311(34.10)	108(11.84)	37(4.06)	49(5.37)		
Time of notified infection								
Before 2017	661(48.60)	313(47.35)	221(33.43)	70(10.59)	25(3.78)	32(4.84)	0.313	0.855
2018–2019	356(26.18)	161(45.22)	129(36.24)	30(8.43)	11(3.09)	25(7.02)		
2020–2022	343(25.22)	166(48.40)	102(29.74)	28(8.16)	20(5.83)	27(7.87)		
Current address								
Suburban	929(68.31)	408(43.92)	347(37.35)	72(7.75)	40(4.31)	62(6.67)	16.278	<0.001***
Downtown	402(29.56)	208(51.74)	101(25.12)	56(13.93)	16(3.98)	21(5.22)		
Detention house	29(2.13)	24(82.76)	4(13.79)	0(0.00)	0(0.00)	1(3.45)		

(Continued)

Table I (Continued).

Characteristics	Total(%) N is Equal to 1360	CRF07_BC	CRF01_AE	CRF08_BC	B	Others	χ^2	P
The first CD4⁺T cells (cells/ul)								
<200	429(31.54)	150(34.97)	173(40.33)	56(13.05)	20(4.66)	30(6.99)	36.246	<0.001***
200–499	656(48.24)	336(51.22)	211(32.16)	51(7.77)	25(3.81)	33(5.03)		
≥500	180(13.24)	109(60.56)	40(22.22)	11(6.11)	6(3.33)	14(7.78)		
Unknown	95(6.99)	45(47.37)	28(29.47)	10(10.53)	5(5.26)	7(7.37)		
Sample source								
Medical Institutions	820(60.29)	348(42.44)	294(35.85)	90(10.98)	36(4.39)	52(6.34)	30.710	<0.001***
High-risk population	422(31.03)	220(52.13)	129(30.57)	33(7.82)	15(3.55)	25(5.92)		
Custodial population	56(4.12)	44(78.57)	9(16.07)	0(0.00)	0(0.00)	3(5.36)		
Others/unknown	62(4.56)	28(45.16)	20(32.26)	5(8.06)	5(8.06)	4(6.45)		
Occupation								
Farmers/migrant workers	760(55.88)	317(41.71)	271(35.66)	82(10.79)	37(4.87)	53(6.97)	21.985	<0.001***
Domestic worker and unemployed	264(19.41)	138(52.27)	75(28.41)	27(10.23)	4(1.52)	20(7.58)		
Others/unknown	336(24.71)	185(55.06)	106(31.55)	19(5.65)	15(4.46)	11(3.27)		

Notes: represents the exact fisher probability; *P<0.05; **P<0.01; ***P<0.001.

Abbreviations: Hetero, heterosexual transmission; MSM, men who have sex with men; IDUs, intravenous drug users; MCT, Mother to child transmission.

Table 2 The Association Between Cluster Formation and Covariates Analysis by Logistic Regression

Variables	Total (%) N is Equal to 1360	Nonclustering (%) N is Equal to 706	Clustering (%) N is Equal to 654	Univariate Analysis		Multivariate Analysis	
				OR (95% CI)	P	OR (95% CI)	P
Gender							
Male	1039(76.40)	561(53.99)	478(46.01)	Reference			
Female	321(23.60)	145(45.17)	176(54.83)	1.19(1.00–1.42)*	0.047	1.18(0.98–1.42)	0.084
Age, years							
≤25	200(14.71)	123(61.50)	77(38.50)	Reference			
26–49	645(47.43)	373(57.83)	272(42.17)	1.10(0.85–1.41)	0.481	1.00(0.74–1.35)	0.989
≥50	515(37.87)	210(40.78)	305(59.22)	1.54(1.20–1.98)*	0.001	1.47(1.04–2.08)*	0.031
Marital Status							
Unmarried	383(28.16)	236(61.62)	147(38.38)	Reference			
Divorced or widowed	217(15.96)	112(51.61)	105(48.39)	1.26(0.98–1.62)	0.070	1.11(0.81–1.52)	0.503
Married	760(55.88)	358(47.11)	402(52.89)	1.38(1.14–1.66)*	0.001	1.20(0.91–1.57)	0.191
Ethnicity							
Han	1308(96.18)	664(50.76)	644(49.24)	Reference			
Others	52(3.82)	42(80.77)	10(19.23)	0.39(0.21–0.73)*	0.003	0.50(0.25–1.01)	0.052
Route of transmission							
MSM	241(17.72)	126(52.28)	115(47.72)	Reference			
Hetero	1075(79.04)	546(50.79)	529(49.21)	1.03(0.84–1.26)	0.765	0.93(0.71–1.21)	0.585
IDUs	38(2.79)	29(76.32)	9(23.68)	0.50(0.25–0.98)*	0.043	0.71(0.32–1.58)	0.404
MCT	6(0.44)	5(83.33)	1(16.67)	0.35(0.05–2.50)	0.295	0.50(0.07–3.65)	0.492
Education Level							
College degree or above	164(12.06)	88(53.66)	76(46.34)	Reference			
High school or technical secondary school	284(20.88)	153(53.87)	131(46.13)	1.00(0.75–1.32)	0.974	0.88(0.66–1.18)	0.402
Middle school or below	912(67.06)	465(50.99)	447(49.01)	1.06(0.83–1.35)	0.651	0.88(0.66–1.18)	0.396
Time of notified infection							
Before 2017	661(48.60)	309(46.75)	352(53.25)	Reference			
2018–2019	356(26.18)	188(52.81)	168(47.19)	0.89(0.74–1.06)	0.198	0.83(0.69–1.01)	0.058
2020–2022	343(25.22)	209(60.93)	134(39.07)	0.73(0.60–0.90)*	0.002	0.69(0.56–0.85)*	0.001
Current address							
Suburban	929(68.31)	511(55.01)	418(44.99)	Reference			
Downtown	402(29.56)	172(42.79)	230(57.21)	1.27(1.08–1.49)*	0.003	1.27(1.05–1.53)*	0.013
Detention house	29(2.13)	23(79.31)	6(20.69)	0.46(0.21–1.03)	0.059	0.61(0.23–1.62)	0.318
The first CD4 +T cells (cells/ul)							
<200	429(31.54)	219(51.05)	210(48.95)	Reference			
200–499	656(48.24)	334(50.91)	322(49.09)	1.00(0.84–1.19)	0.975	1.01(0.85–1.21)	0.884
≥500	180(13.24)	95(52.78)	85(47.22)	0.96(0.75–1.24)	0.780	1.02(0.78–1.32)	0.904
Unknown	95(6.99)	58(61.05)	37(38.95)	0.80(0.56–1.13)	0.200	0.90(0.63–1.30)	0.581

(Continued)

Table 2 (Continued).

Variables	Total (%) N is Equal to 1360	Nonclustering (%) N is Equal to 706	Clustering (%) N is Equal to 654	Univariate Analysis		Multivariate Analysis	
				OR (95% CI)	P	OR (95% CI)	P
Subtype							
CRF07_BC	640(47.06)	349(54.53)	291(45.47)	Reference			
CRF01_AE	452(33.24)	221(48.89)	231(51.11)	1.12(0.95–1.34)	0.185	1.06(0.89–1.27)	0.510
CRF08_BC	128(9.41)	49(38.28)	79(61.72)	1.36(1.06–1.74)*	0.016	1.12(0.86–1.46)	0.391
B	56(4.12)	37(66.07)	19(33.93)	0.75(0.47–1.19)	0.216	0.76(0.48–1.22)	0.255
Others	84(6.18)	50(59.52)	34(40.48)	0.89(0.62–1.27)	0.521	0.88(0.61–1.26)	0.484
Sample source							
Medical institutions	820(60.29)	447(54.51)	373(45.49)	Reference			
High-risk population	422(31.03)	189(44.79)	233(55.21)	1.21(1.03–1.43)*	0.02	1.27(1.06–1.51)*	0.009
Custodial population	56(4.12)	38(67.86)	18(32.14)	0.71(0.44–1.13)	0.15	1.58(0.85–2.94)	0.148
Others/unknown	336(24.71)	183(54.46)	153(45.54)	1.06(0.73–1.54)	0.745	1.24(0.85–1.82)	0.262
Occupation							
Farmers/migrant workers	760(55.88)	374(49.21)	386(50.79)	Reference			
Domestic workers/unemployed	264(19.41)	149(56.44)	115(43.56)	0.86(0.70–1.06)	0.148	0.89(0.71–1.12)	0.31
Others/unknown	62(4.56)	32(51.61)	30(48.39)	0.90(0.74–1.08)	0.253	0.90(0.73–1.10)	0.299

Notes: *P<0.05 were considered statistically significant.

Abbreviations: Hetero, heterosexual transmission; MSM, men who have sex with men; IDUs, intravenous drug users; MCT, Mother to child transmission.

patients in suburbans, those in downtown had a higher risk, which was 1.27 times that of patients in suburbans. The risk of patients notified infection between 2020 to June 2022 was 0.69 times that of notified infection before 2017.

In this study, clusters with more than 10 cases were selected as key clusters to be monitored, and the characteristics of each cluster are shown in Table 3. We observed that the 07BC1 and AE15 clusters were associated with individuals reporting homosexual transmission. There were mainly young males in downtown with a median age of 32 and 27, respectively. 07BC1 was the largest cluster involve 130 cases in Guangyuan. After the first case was found in 2011, it expanded year by year and became the largest transmission cluster in this area, with virological failure ratio of 70.77%. The other large clusters were mainly spread by heterosexual, with a median age of 50–59 years old. The number of male cases more than 68% in these clusters except AE2, indicating that the spread of the majority clusters which transmitted by heterosexual behavior were caused by elderly men over 50 years old. While the majority of the patients in AE2 were females over 50 years old which were distributed in suburbans mainly. AE4 was the cluster with the second largest number of cases after 07BC1. 92.31% of those cases were distributed by a county, 99.04% of the cases were acquired through heterosexual transmission but only one male case was infected by homosexual transmission. The expansion of each cluster over years is shown in Figure 1. With the exception of 07BC1, 08BC1 and AE2, the number of new cases entering the transmission network increased significantly in 2019 and decreased after 2020.

Univariate Analysis

Univariate analysis was performed between 07BC1, AE4 and other clusters with more than 10 cases, indicating significant differences in gender, marital status, ethnicity, transmission route, contact history, education level, the time of notified infection, current address, the first CD4+T cells, genotype, antiretroviral therapy (ART) or not, sample source, and occupation (Table 4). Transmission risk prediction models were established for 07BC1 and AE4 clusters respectively, and the final results of multivariate Logistic regression analysis showed that the risk factors of 07BC1 cluster was as follows: the time of notified infection before 2017 (OR: 4.717, 95% CI: 1.308–17.008), IDUs (OR: 10.371, 95% CI: 3.151–30.630), MSM (OR: 11.918, 95% CI: 3.290–43.173), aged 26–49 years old (OR: 14.856, 95% CI: 5.495–40.165) and over 50 years old (OR: 20.755, 95% CI: 4.244–101.508), male (OR: 4.238, 95% CI: 1.397–12.851), others or unknown occupation (OR: 5.032, 95% CI: 1.456–17.384).

Establishment and Evaluation of Transmission Risk Prediction Model

First we check the performance of Logistic regression algorithms through 10-fold cross validation, it showed a good performance in ROC curve analysis among 10-fold cross validation (average AUC were at least 0.85, the specific details see Figure 2). Furthermore, we established the risk prediction model, the result of transmission risk expression of 07BC1 cluster in this study was as follows:

Table 3 General Distribution Characteristics of the Clusters (Nodes ≥ 10) in Guangyuan

Cluster Number	Subtypes	Time of Cluster Formation	Nodes	Characteristics of Clusters				
				Numbers of Male (%)	Transmission Route (%)	Current Address (%)	Median Age	Viral Load $\leq 1000\text{cps/mL}$ After ART (%)
07BC1	CRF07_BC	2011	130	123(94.62)	63(48.46) ^a	76(58.46) ^c	32	92(70.77)
AE4	CRF01_AE	2011	104	77(74.04)	103(99.04) ^b	95(92.31) ^d	54	76(73.08)
08BC2	CRF08_BC	2011	35	24(68.57)	35(100.00) ^b	25(71.43) ^c	59	24(68.57)
AE14	CRF01_AE	2012	20	14(70.00)	20(100.00) ^b	11(55.00) ^c	55	16(80.00)
08BC1	CRF08_BC	2012	18	13(72.22)	18(100.00) ^b	13(72.22) ^c	51	15(83.33)
AE2	CRF01_AE	2015	15	5(33.33)	15(100.00) ^b	13(86.67) ^d	59	12(80.00)
AE15	CRF01_AE	2016	10	9(90.00)	6(60.00) ^a	8(80.00) ^c	27	9(90.00)

Notes: ^arepresents the transmission route is MSM; ^brepresents the transmission route is Hetero; ^crepresents the current address is downtown; ^drepresents the current address is suburbans.

Abbreviation: ART, antiretroviral therapy.

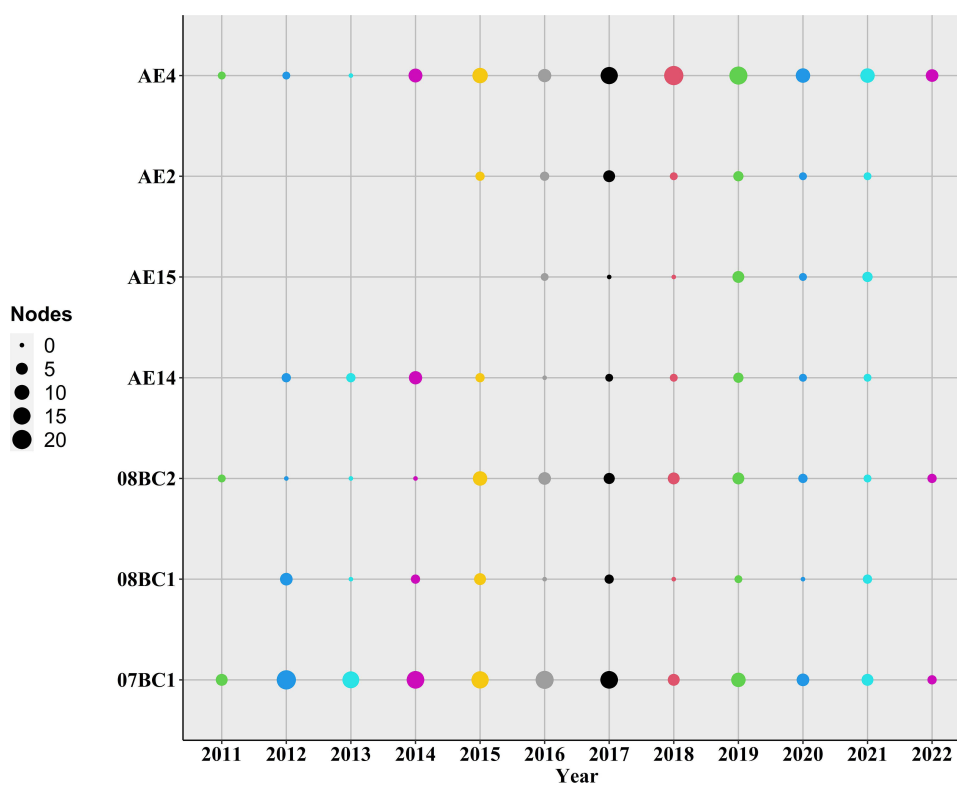


Figure 1 The expansion of each cluster over years (bubble chart).

Notes: The nodes in AE4 was 104; the nodes in AE2 was 15; the nodes in AE15 was 10; the nodes in AE14 was 20; the nodes in 08BC2 was 35; the nodes in 08BC1 was 18; the nodes in 07BC1 was 130.

$$\text{Logit}(\pi) = 1.551 * (\text{Before2017}) + 2.339 * (\text{IDUs}) + 2.478 * (\text{MSM}) + 2.698 * (26 - 49) + 3.033 * (350) \\ + 1.444 * (\text{male}) + 1.616 * (\text{others or unknown occupation}) - 0.829$$

π was the probability of 07BC1, the specific information of the model see Table 5.

Table 4 Univariate Analysis Between 07BC1, AE4 and Other Clusters

Characteristics	N(%)	07BC1 (n=130)	AE4 (n=104)	Others (n=98)	χ^2	P
Gender						
Male	265(79.82)	123(94.62)	77(74.04)	65(66.33)	30.809	<0.001***
Female	67(20.18)	7(5.38)	27(25.96)	33(33.67)		
Age, years						
≤25	37(11.14)	31(23.85)	0(0.00)	6(6.12)	135.366	<0.001***
26–49	144(43.37)	90(69.23)	28(26.92)	26(26.53)		
≥50	151(45.48)	9(5.96)	76(50.23)	66(43.71)		
Marital status						
Divorced or widowed	63(18.98)	18(13.85)	24(23.08)	21(21.43)	77.075	<0.001***
Unmarried	81(24.40)	65(50.00)	5(4.81)	11(11.22)		
Married	188(56.63)	47(36.15)	75(72.12)	66(67.35)		
Ethnicity						
Han	329(99.10)	128(98.46)	104(100.0)	97(98.98)	—	0.045*
Hui	1(0.30)	0(0.00)	0(0.00)	1(1.02)		
Yi	2(0.60)	2(1.54)	0(0.00)	0(0.00)		

(Continued)

Table 4 (Continued).

Characteristics	N(%)	07BC1 (n=130)	AE4 (n=104)	Others (n=98)	χ^2	P
Route of transmission						
Hetero	257(77.41)	60(46.15)	103(99.04)	94(95.92)	—	<0.001***
MSM	72(21.69)	67(51.54)	1(0.96)	4(4.08)		
IDUs	3(0.90)	3(2.31)	0(0.00)	0(0.00)		
Contact history						
Homosexual intercourse among MSM	70(21.08)	65(50.00)	1(0.96)	4(4.08)	117.412	<0.001***
Hetero-nonmarital and noncommercial	123(37.05)	29(22.31)	43(41.35)	51(52.04)		
Hetero-nonmarital	93(28.01)	24(18.46)	41(39.42)	28(28.57)		
Spouse or permanent partner	23(6.93)	2(1.54)	11(10.58)	10(10.20)		
Others	23(6.93)	10(7.69)	8(7.69)	5(5.10)		
Education Level						
Middle school or below	217(65.36)	51(39.23)	87(83.65)	79(80.61)	69.630	<0.001***
College degree or above	40(12.05)	33(25.38)	2(1.92)	5(5.10)		
High school or technical secondary school	75(22.59)	46(35.38)	15(14.42)	14(14.29)		
Time of notified infection						
2018–2019	67(20.18)	6(4.62)	37(35.58)	24(24.49)	52.689	<0.001***
2020–2022	52(15.66)	13(10.00)	24(23.08)	15(15.31)		
Before 2017	213(64.16)	111(85.38)	43(41.35)	59(60.20)		
Current address						
C county	21(6.33)	9(6.92)	2(1.92)	10(10.20)	254.082	<0.001***
J county	40(12.05)	21(16.15)	0(0.00)	19(19.39)		
L area	138(41.57)	76(58.46)	5(4.81)	57(58.16)		
Q county	9(2.71)	7(5.38)	0(0.00)	2(2.04)		
W county	107(32.23)	8(6.15)	96(92.31)	3(3.06)		
Unknown	17(5.12)	9(6.92)	1(0.96)	7(7.14)		
The first CD4+T cells (cells/ul)						
<200	110(33.13)	27(20.77)	45(43.27)	38(38.78)	26.318	<0.001***
200–499	171(51.51)	77(59.23)	52(50.00)	42(42.86)		
≥500	36(10.84)	22(16.92)	4(3.85)	10(10.20)		
Unknown	15(4.52)	4(3.08)	3(2.88)	8(8.16)		
ART or not						
No	9(2.71)	6(4.62)	0(0.00)	3(3.06)	—	<0.001***
Yes	289(87.05)	112(86.15)	92(88.46)	85(86.73)		
Others	34(10.24)	12(9.23)	12(11.54)	10(10.20)		
Viral load (cps/mL)						
≤1000	240(72.29)	92(70.77)	76(73.08)	72(73.47)	5.156	0.272
>1000	28(8.43)	7(5.38)	12(11.54)	9(9.18)		
Untreated	64(19.28)	31(23.85)	16(15.38)	17(17.35)		
Subtypes						
CRF01_AE	149(44.88)	0(0.00)	104(100.0)	45(45.92)	455.947	<0.001***
CRF07_BC	130(39.16)	130(100.0)	0(0.00)	0(0.00)		
CRF08_BC	53(15.96)	0(0.00)	0(0.00)	53(54.08)		
Sample source						
Medical Institutions	196(59.04)	55(42.31)	68(65.38)	73(74.49)	—	<0.001***
High-risk population	111(33.43)	62(47.69)	27(25.96)	22(22.45)		
Custodial population	8(2.41)	7(5.38)	0(0.00)	1(1.02)		
Others/unknown	17(5.12)	6(4.62)	9(8.65)	2(2.04)		
Occupation						
Domestic worker/unemployed	58(17.47)	28(21.54)	9(8.65)	21(21.43)	53.063	<0.001***
Farmers/migrant workers	171(51.51)	38(29.23)	77(74.04)	56(57.14)		
Others/unknown	103(31.02)	64(49.23)	18(17.31)	21(21.43)		

Notes: —represents the exact fisher probability; *P<0.05; ***P<0.001.

Abbreviations: Hetero, heterosexual transmission; MSM, men who have sex with men; IDUs; intravenous drug users; ART, antiretroviral therapy.

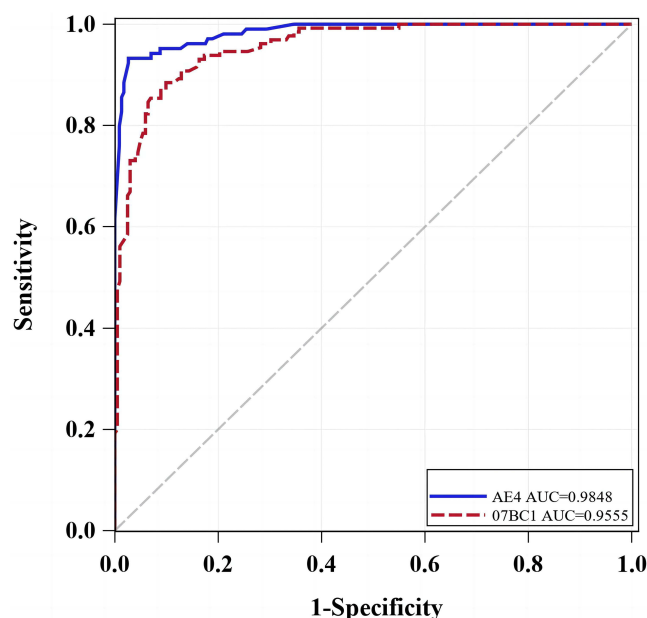


Figure 2 Evaluation of sensitivity, specificity of prediction models.

Notes: The solid blue line indicates the sensitivity and specificity of AE4. The dashed red line indicates the sensitivity and specificity of 07BC1.

Abbreviation: AUC, area under the curve.

The result of transmission risk expression of AE4 cluster was as follows:

$$\text{Logit}(\pi) = 1.572 * (\text{Viralload1000cps/ml}) + 1.959 * (2018 - 2019) + 3.019 * (\text{WCounty}) \\ + 3.050 * (\text{Hetero}) + 6.969 * (\text{others or unknown sample source}) - 5.832$$

π was the probability of AE4 clusters, the specific information of the model see [Table 6](#).

The chi-square value of H-L test was 0.95 and $P=0.329$, It indicates that the model has a good fit, the ROC-AUC of 07BC1 and AE4 were 0.9848 and 0.9555 respectively ([Figure 2](#)), indicating that the models have a good ability of differentiation.

Table 5 Risk Prediction Results of 07BC1 Cluster

Variables	OR (95% CI)	Estimate	SE	Wald χ^2	P
Gender					
Female	Reference				
Male	4.238(1.397,12.851)	1.444	0.566	6.509	0.011
Age, years					
≤25	Reference				
26–49	14.856(5.495,40.165)	2.698	0.507	28.279	<0.001
≥50	20.755(4.244,101.508)	3.033	0.810	14.024	<0.001
Route of transmission					
Hetero	Reference				
IDUs	10.371(3.151,30.630)	2.339	1.083	9.991	<0.001
MSM	11.918(3.290,43.173)	2.478	0.657	14.239	<0.001
Time of notified infection					
2018–2019	Reference				
2020–2022	1.532(0.311,7.544)	0.427	0.813	0.275	0.600
Before 2017	4.717(1.308,17.008)	1.551	0.654	5.619	0.018

(Continued)

Table 5 (Continued).

Variables	OR (95% CI)	Estimate	SE	Wald χ^2	P
Current address					
C county	Reference				
W county	0.283(0.054,1.478)	-1.261	0.843	2.239	0.135
J county	3.162(0.634,15.782)	1.151	0.820	1.970	0.160
L area	1.193(0.290,4.911)	0.177	0.722	0.060	0.807
Q county	2.183(0.219,21.741)	0.781	1.173	0.443	0.505
Unknown	1.321(0.177,9.836)	0.278	1.024	0.074	0.786
Occupation					
Domestic workers and unemployed	Reference				
Farmers/migrant workers	1.707(0.504,5.784)	0.535	0.623	0.737	0.391
Others/unknown	5.032(1.456,17.384)	1.616	0.633	6.525	0.011

Abbreviations: SE, standard error; Hetero, heterosexual transmission; MSM, men who have sex with men; IDUs, intravenous drug users.

Table 6 Risk Prediction Results of AE4 Cluster

Variables	OR (95% CI)	Estimate	SE	Wald χ^2	P
Route of transmission					
Hetero	Reference				
IDUs	0.356(0.116,1.087)	-1.033	0.570	<0.001	0.989
MSM	21.125(7.956,56.041)	3.050	0.498	4.149	0.042
Time of notified infection					
Before 2017	Reference				
2018–2019	7.093(1.590,31.647)	1.959	0.763	6.592	0.010
2020–2022	4.199(0.749,23.530)	1.435	0.879	2.663	0.103
Current address					
C county	Reference				
J county	0.076(0.515,3.233)	-2.576	1.913	0.004	0.952
L area	0.669(0.101,4.430)	-0.403	0.965	0.174	0.676
Q county	0.054(0.013,0.224)	-2.913	0.725	<0.001	0.985
W county	55.645(2.347,178.538)	3.019	1.105	29.667	<0.001
Unknown	4.778(0.285,80.143)	1.564	1.439	1.182	0.277
Viral load (cps/mL)					
≤1000	Reference				
>1000	4.818(1.295,17.907)	1.572	0.670	4.409	0.036
untreated	0.182(0.037,0.893)	-1.702	0.810	2.161	0.142
Sample resource					
Medical Institutions	Reference				
High-risk population	2.303(0.584,9.088)	0.834	0.700	1.419	0.234
Custodial population	0.027(0.002,2.525)	-5.603	2.311	0.001	0.980
Others/unknown	7.163(1.143,44.860)	6.969	1.936	12.963	<0.001
Occupation					
Domestic workers/unemployed	Reference				
Farmers/migrant workers	3.502(0.591,20.743)	1.253	0.908	2.907	0.067
Others/unknown	0.861(0.114,6.521)	-0.149	1.033	0.021	0.885

Abbreviations: SE, standard error; Hetero, heterosexual transmission; MSM, men who have sex with men; IDUs, intravenous drug users.

Discussion

Guangyuan was one of the first pilot city of molecular transmission network in Sichuan Province. Continuous and dynamic surveillance was carried out nearly 10 years. Based on molecular epidemiology, the study deeply understand the characteristics of HIV-1 molecular epidemiology and reveal the risk factors of transmission dynamics by combining longitudinal cohort data with statistical methods for the first time. Besides, it established a transmission risk prediction model to predict the driving factors of network expansion.

The epidemic situation in Guangyuan was still dominated by heterosexual transmission, and most of the infected individuals were farmers or migrant workers. The genotypes of HIV-1 were relatively complex and diverse, with the main epidemic subtypes being CRF01_AE and CRF07_BC, and others such as CRF08_BC and B, which were consistent with the results of Su et al.²⁰ With the genetic variation and recombination of genes between different genotypes, numerous new circulating recombinant forms (CRFs) or unique recombinant forms (URFs) such as CRF55_01B, CRF85_BC, CRF64_BC, CRF59_01B, CRF67_01B and CRF80_0107 had appeared successively since 2015. The results may be related to the phenomenon that Guangyuan was a major city of labor export in Sichuan province, and some farmers and migrant workers were infected during the other cities. They resulting in a large number of imported cases when return to their hometown.²¹ In addition, Sichuan province was located in the southwest of China, and the AIDS epidemic was affected to varying degrees by Yunnan and other neighboring provinces, which caused the genotypes complexly and diversely.^{22,23} With the increasingly convenient transportation and personnel flow between the city in the province, the genotypes of HIV-1 tend to be more and more complex in Guangyuan. CRF85_BC was a recombinant strain of subtype B and C identified for the first time in Sichuan province, which was considered to have been introduced from Yibin²⁴ and gradually spread in other cities of Sichuan province in recent years.²³ CRF55_01B was generated by the recombination of CRF01_AE and subtype B.²⁵ It first appeared in the newly diagnosed patients in 2013 in Guangyuan and epidemic among males mainly, which was consistent with previous studies that CRF55_01B originated in Shenzhen and prevailed in MSM mainly.^{26,27} It suggests that attention should be paid to imported HIV-infected individuals and surveillance of local subtypes should be strengthened to reduce transregional transmission and the generation of recombination genotypes.

In this study, the patients with characteristics of aged under 25 years old, unmarried and MSM were more likely to be infected by CRF07_BC. The proportion of CRF07_BC decreases with age increasing, which indicates that CRF07_BC was mainly epidemic among MSM in Guangyuan. While CRF01_AE and CRF08_BC were mainly popular in the population that over 50 years old, married and route of heterosexual transmission which was consistent with Su et al.²⁰ The proportion of CRF01_AE and CRF08_BC increases with aging. CRF08_BC and CRF07_BC were prevalent among IDUs in Yunnan early²⁸ and then introduced into Liangshan, Sichuan.²⁹ Besides, the patients whose first CD4+T cells ≥ 500 cells/ul were infected CRF07_BC mostly and the proportion of CRF07_BC decreased with the decrease of the first CD4+T cells. While the patients whose first CD4+T cells ≤ 200 cells/ul were more likely to be infected CRF01_AE and the proportion of CRF01_AE decreased with the increase of the first CD4+T cells. It maybe indicated that the ratio of late detection was increases with aging. While the ratio of late detection was low in MSM because of the higher risk awareness with more educated and younger.

It was particularly important to adjust prevention and control strategies timely to control the epidemic. We constructed a molecular transmission network of newly diagnosed patients to deeply investigate the transmission characteristics of each genotypes in Guangyuan in the past 12 years. We explored the risk factors of transmission dynamics based on epidemiology and analyzed the characteristics of clusters with more than 10 cases, which providing a basis for the formulation of precise prevention and control strategies. The results showed that 48.1% of the HIV-infected patients entered the molecular transmission network, suggesting that there was a potential transmission association, which caused widely spread in local area after the newly strains entering, and the main transmission route were MSM and heterosexual transmission. According to the fitting analysis of Poisson regression model and Logistic regression model, we observed that females, aged over 50 years old, married, current address in downtown, CRF08_BC and high-risk population are regarded as risk factors for clustering, which suggests that we should pay more attention to the elderly over 50 years old and continue to strengthen the surveillance of high-risk population. Although the proportion of patients that under 25 years old is relatively low, most of them were infected by homosexual transmission and have more possibilities to entered clusters, which also deserves attention. Besides, the proportion of CRF08_BC among HIV-infected patients was not very high but the ratio of clustering was high, which was

inconsistent with the results of Zhou et al.³⁰ It indicating that the clustering rate of different subtypes varies among regions, and timely molecular network surveillance is conducive to more efficient prevention and control of HIV transmission.

In addition, CRF07_BC was epidemic in downtown. CRF07_BC was more likely to access to the network and forming the largest transmission cluster 07BC1 in Guangyuan, in which number of males accounted for 94.62%, but homosexual transmission accounted for only 50%. Previous studies on MSM in Guangyuan showed that MSM were more inclined to conceal his homosexual behavior in the face of medical institutions.³¹ In addition, Partial MSM in the molecular transmission networks were also connected to heterosexual male and female according the study of Yuan et al.³² Most MSM choose to marry with women to hide their identity due to Chinese family values and fear of prejudice, which increasing the risk of transmission to their spouses.³³ Therefore, it can be speculated that the 07BC1 cluster was more likely to be from infections among individuals reporting homosexual transmission, which suggests that CRF07_BC have a high risk for transmission in Guangyuan. CRF01_AE was epidemic in suburbans mainly. AE4 was the largest heterosexual transmission cluster in W County of Guangyuan, which has been expanding year by year since the first HIV-infected patient was discovered in 2011. Preliminary studies show that the sex trade of elderly whoring clients in low-grade entertainment place was the main reason for the network transmission.³⁴ In recent years, the reported number and composition ratio of elderly male HIV-infected patients in Sichuan province have increased year by year,³⁵ suggesting that more attention should be paid to the surveillance of elderly men and sex workers in this area in the future, and the intervention of spouses of HIV-infected patients should be strengthened to reduce transmission within the family.

The HIV-infected patients in Guangyuan were mainly the Han nationality, and there were also a few other ethnic populations, mainly the Yi nationality. The patients of Yi nationality were infected by CRF07_BC mainly, and more than half of them were infected through injecting drug, which was similar to the epidemic situation in Liangshan in the early study.³⁶ However, most of the patients of Yi nationality did not enter the local transmission clusters, and the majority of them came from detention house. Therefore, it was inferred that the formation of transmission clusters of CRF07_BC was not caused by Yi patients in Guangyuan. We considering that MSM were prone to cause the clusters rapidly expansion as the characteristics of multiple sexual partners, unprotected high-risk sexual behavior, secluded place and strong mobility.^{37,38} It was inferred that MSM were the driving force of the ongoing epidemic of CRF07_BC in Guangyuan and the continuous expansion of 07BC1 cluster. An early study also reported that MSM transmission was be the main risk factor for transmission clusters.³⁹ Besides, as the characteristics of convenient transportation and mobility of downtown, which can lead to the further rapid spread of CRF07_BC in this area, further confirming that attention should be paid to the MSM, especially those live in downtown. We can continue to use the internet to carry out continuous online health education, condom promotion, AIDS counseling, mobilization testing and other services for MSM in Guangyuan, and even consider promoting pre-exposure prophylactics and other measures, so as to reduce the HIV transmission among MSM. Secondly, male patients accounted for a high proportion in the majority of clusters in this study, but the majority of patients were female in AE2 only. Preliminary investigation showed that most women had heterosexual non-marital and non-commercial sex with a male in AE2 cluster.¹⁸ By June 2022, 5 cases were added in AE2 cluster, all of which were from J County, suggesting that more attention should be paid to the detection and intervention of left-behind women over 50 years old in this county in the future. Under the influence of the expanded testing policy, the number of new cases in most transmission clusters was increased significantly in 2019, indicating that expanded testing played an important role in the timely detection of patients. While the decline in the number of new cases after 2020 was attribute to the local dynamic molecular network surveillance and targeted intervention of key groups and regions to contain network expansion. On the one hand, the impact of COVID-19 has reduced the flow of people and slowed down network expansion. In the future, we will continue to perform the dynamic surveillance of molecular network in Guangyuan to take precise intervention timely and effective.

The predictive of transmission risk models were established for 07BC1 and AE4 clusters respectively. The predictive model developed and validated existing HIV surveillance data and showed excellent predictive ability to forecast cluster growth in Guangyuan. 07BC1 and AE4 were the largest clusters in Guangyuan, among which 07BC1 was mainly transmitted by homosexual transmission. The median age of 07BC1 was 32 years old, and the current address was mainly in downtown. However, AE4 was mainly transmitted by heterosexual transmission, and the cases were mainly elderly men over 50 years old in suburbans.

The results showed that the time of notified infection before 2017, IDUs and MSM, aged 26–49 and over 50 years old, male, others or unknown occupation were the risk factors for 07BC1 clusters. The ratio of viral load $\leq 1000\text{cps/mL}$ after one year of ART was only 70.77% which may be one of the reasons for the network expansion. The other occupations were mainly commercial services, retirees and students and so on. This result suggests that we should pay more attention to intravenous drug users and MSM in Guangyuan in the future, especially aged 26–49 years. To prevent the continuous expansion of 07BC1, we should strengthen publicity and intervention for those engaged in business services, students and retirees. The characteristics of viral load $> 1000\text{cps/mL}$, the time of notified infection between 2018 and 2019, live in W County, heterosexual transmission, others or unknown sample resource were risk factor for AE4 clusters. The ratio of viral load $\leq 1000\text{cps/mL}$ after one year of ART was also just 73.08%, which may be due to the fact that patients of this cluster were mainly elderly patients with low knowledge of AIDS prevention and treatment, low risk awareness and low medication compliance.⁴⁰ From 2018 to 2019, the number of new cases in the network increased significantly, and then trend to steady, mainly because with the implementation of the comprehensive AIDS prevention and control policy to expand testing, the detection rate increased significantly, a large number of patients were discovered, and the newly diagnosed patients showed an upward trend. The detection rate in 2018–2019 was significantly higher than that in other years, which also reflected the remarkable effect of expanded testing. However, the detection rate declined in recent two years.⁴¹ Combined with the implementation of “discovery is treatment” and other policies, the effective detection rate of HIV in Sichuan province was only 0.717% in 2020,⁴² which also suggests that it was crucial to adjust prevention and control strategies in timely. In addition to expanding detection scientifically and orderly, it was necessary to strengthen the surveillance of key populations and take precise intervention in low-epidemic areas. Health education and high-risk sexual behavior intervention should be strengthened for elderly men to raise their awareness of disease risk and urge them to detection timely.

However, the number of new cases of cluster 07BC1 decreased from 2018 to 2019. It indicating that the detection of MSM was still insufficient and may caused by the hidden places of MSM and strong mobility which make it difficult to manage. On the other hand, it also reflects the initial effectiveness of intervention measures for MSM through dynamic surveillance of molecular network.⁴³ In addition, we should strengthen publicity and education, promote condom use, strengthen peer education and Internet intervention for MSM. Besides, dynamic monitoring through molecular network can help us find hot spots and carry out precise intervention due to their strong mobility.

Limitations

This study also has some limitations. The network was not complete due to the lack of some sequences of newly diagnosed HIV-infection patients and undiscovered patients. In addition, the sequences obtained by Sanger sequencing which cannot represent all quasispecies in patients. In the future, more efforts will be strengthened to discover the potential patients and improved the sample quality to carry out dynamic surveillance of molecular transmission network continuously. Besides, high-throughput sequencing will be actively applied to obtain more abundant and real gene information of patients. All the measures will provide scientific support for epidemic prevention and control in Guangyuan.

Conclusion

The expanding detection was low efficiency with high cost, while the traditional epidemiological investigation was time-consuming and labor-intensive in a low-epidemic area. So the identification of high-risk groups and fast-growing clusters by molecular transmission network was crucial for epidemic prevention and control. According to the surveillance data of infectious disease from National AIDS Comprehensive Prevention and Control Information System, the number of newly diagnosed patients in Sichuan province continued to decline in 2022, but the decline range was reduce. Only the newly diagnosed HIV-infected patients in Guangyuan declined the most, which was closely related to the implementation of dynamic surveillance of molecular transmission network and precise prevention and control since 2016. Based on the construction of molecular transmission network and the support of field epidemiological data, this study analyzed the epidemic characteristics of HIV-1 strains in Guangyuan from 2010 to June 2022, explore the risk factors of transmission dynamics of clusters by followed these clusters across a 12 years time period, and established a transmission risk model

for the two largest transmission clusters to predict the risk factors causing expansion, which provide a guidance for effective and precise intervention of crucial populations and regions to reduce transmission in the future.

Acknowledgments

The authors would like to thank participants from Guangyuan Center for Disease Control and Prevention for their hard work on sample collection. Additionally, the authors would like to sincerely appreciate all investigators for their help and support.

Funding

This study was supported by Natural Science Foundation of Sichuan Province (grant number 2022NSFSC1547) and Scientific research project of Sichuan Center for Disease Control and Prevention (grant number ZX202018).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Yuan FS, Liu L, Liu LH, et al. Epidemiological and spatiotemporal analyses of HIV/AIDS prevalence among older adults in Sichuan, China between 2008 and 2019: a population-based study. *Int J Infect Dis.* 2021;105:769–775. doi:10.1016/j.ijid.2021.02.077
2. Huang WL, Li T, Gong Y, et al. Analysis of social organizations' participation in AIDS prevention and control fund projects in Sichuan Province in 2016. *J Prev Med Inf.* 2018;34(12):1478–1481, 1486.
3. Hou XQ, Zhu J, Xiong L, et al. Epidemiological characteristics and trends of AIDS in Guangyuan from 1996 to 2018. *Mod Prev Med.* 2020;47(03):388–391+418.
4. UNAIDS. 90-90-90-An ambitious treatment target to help end the AIDS epidemic [R/OL]; 2017.
5. Feng L, Xie RX, Zhang LL, et al. Win the battle against AIDS and set sail: measures and results of AIDS control in Sichuan Province. *Mod Prev Med.* 2022;49(20):3649–3654. doi:10.20043/j.cnki.MPM.202209900
6. General Office of sichuan Provincial People's Government. Notice of the general office of the Sichuan Provincial People's Government on the issuance of first-level response measures to the outbreak of pneumonia caused by COVID-19 [EB/OL]; 2020.
7. Sichuan Provincial Health Commission. Announcement of Sichuan emergency response headquarters for pneumonia caused by COVID-19 (No. 2). [EB/OL]. 2020.
8. Yuan FS, Liu L, Liu LH, et al. Analysis of the correlation between HIV/AIDS cases that newly reported reports and COVID-19 in Sichuan province in the first quarter of 2020. *Chin J AIDS STD.* 2020;206(11):1178–1181. doi:10.13419/j.cnki.aids.2020.11.09
9. Wertheim JO, Murrell B, Mehta SR, et al. Growth of HIV-1 molecular transmission clusters in New York City. *J Infect Dis.* 2018;218(12):1943–1953. doi:10.1093/infdis/jiy431
10. Feng Y, Wang D, Xing H. HIV molecular transmission network helps precise prevention and control of AIDS in China *Chin J AIDS STD.* 2023;232(01):1–8. doi:10.13419/j.cnki.aids.2023.01.01
11. Yang JZ, Chen WJ, Zhang WJ, He L, Zhang JF, Pan XH. Molecular epidemiology and transmission of HIV-1 infection in Zhejiang province, 2015. *Chin J Epidemiol.* 2017;38(11):1551–1556. doi:10.3760/cma.j.issn.0254-6450.2017.11.023
12. Wang FX, Zhou H, Ling H, et al. Subtype and sequence analysis of HIV-1 strains in Heilongjiang Province. *Chin Med J.* 2007;120(22):2006–2010. doi:10.1097/00029330-200711020-00013
13. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–1874. doi:10.1093/molbev/msw054
14. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. *Curr Opin HIV AIDS.* 2014;9(2):126–133. doi:10.1097/COH.0000000000000040
15. Liu K, Linder CR, Warnow T. RAXML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *PLoS One.* 2011;6(11):e27731. doi:10.1371/journal.pone.0027731
16. CDC Guidance: detecting, investigating, and responding to HIV transmission clusters. Prevention TRCF; 2018.
17. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol.* 1999;73(12):10489–10502. doi:10.1128/JVI.73.12.10489-10502.1999
18. Yuan D, Yu B, Liang S, 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.* 2021;18:100318. doi:10.1016/j.lanwpc.2021.100318
19. Chinese Center for Disease Control and Prevention. *Guideline for the Analysis of HIV-1 Molecular Transmission Networks*. Beijing, China: Chinese Center for Disease Control and Prevention; 2013.
20. Su L, Zhu J, Zeng YL, et al. Epidemiological characteristics of HIV-1 strains in Guangyuan from 2011 to 2018. *Chin J AIDS STD.* 2020;203(08):831–835. doi:10.13419/j.cnki.aids.2020.08.09
21. Su L, Liang S, Hou X, et al. Impact of worker emigration on HIV epidemics in labour export areas: a molecular epidemiology investigation in Guangyuan, China. *Sci Rep.* 2018;8(1):16046. doi:10.1038/s41598-018-33996-6
22. Su L, Yang H, Hu Y, et al. Molecular epidemiological investigation of HIV-1 in Sichuan Province in 2014. *Chin J AIDS STD.* 2016;159(12):946–949. doi:10.13419/j.cnki.aids
23. Zhang Y, Dai J, Li Z, et al. Using molecular network analysis to explore the characteristics of HIV-1 transmission in a China-Myanmar border area. *PLoS One.* 2022;17(5):e0268143. doi:10.1371/journal.pone.0268143

24. Su L, Wei D, Yang H, et al. Identification of a Novel HIV-1 Circulating Recombinant Form (CRF85_BC) in Sichuan, China. *AIDS Res Hum Retroviruses*. 2016;32(9):895–899. doi:10.1089/AID.2016.0053
25. Han X, An M, Zhang W, et al. Genome Sequences of a novel HIV-1 circulating recombinant form, CRF55_01B, identified in China. *Genom Announc*. 2013;1(1):e00050–12. doi:10.1128/genomeA.00050-12
26. Zhao J, Chen L, Chaillon A, et al. The dynamics of the HIV epidemic among men who have sex with men (MSM) from 2005 to 2012 in Shenzhen, China. *Sci Rep*. 2016;6(1):28703. doi:10.1038/srep28703
27. Zhang J, Guo Z, Pan X, et al. Highlighting the crucial role of Hangzhou in HIV-1 transmission among men who have sex with men in Zhejiang, China. *Sci Rep*. 2017;7(1):13892. doi:10.1038/s41598-017-14108-2
28. Yang R, Kusagawa S, Zhang C, Xia X, Ben K, Takebe Y. Identification and characterization of a new class of human immunodeficiency virus type 1 recombinants comprised of two circulating recombinant forms, CRF07_BC and CRF08_BC, in China. *J Virol*. 2003;77(1):685–695. doi:10.1128/jvi.77.1.685-695.2003
29. Li L, Wei D, Hsu WL, et al. CRF07_BC strain dominates the HIV-1 epidemic in injection drug users in Liangshan Prefecture of Sichuan, China. *AIDS Res Hum Retroviruses*. 2015;31(5):479–487. doi:10.1089/AID.2014.0120
30. Zhou C, Liang S, Li Y, et al. Characterization of HIV-1 molecular epidemiology and transmitted drug-resistance in newly diagnosed HIV-infected patients in Sichuan, China. *BMC Infect Dis*. 2022;22(1):602. doi:10.1186/s12879-022-07576-z
31. Zhuoma L, Zhang Y, Yan T, et al. Non-disclosed men who have sex with men within local MSM HIV-1 genetic transmission networks in Guangyuan, China. *Front Public Health*. 2022;10:956217. doi:10.3389/fpubh.2022.956217
32. Yuan D, Zhong X, Li Y, et al. Molecular transmission network of newly reported HIV infections in Pengzhou, Sichuan Province: a study based on genomics and spatial epidemiology. *Int J Environ Res Public Health*. 2023;20(3):2523. doi:10.3390/ijerph20032523
33. Ling Q, Li PL, Tang L, et al. Research progress on susceptibility of heterosexual partners with HIV in men who have sex with men. *Chin J Epidemiol*. 2020;10(10):1731–1734.
34. Zeng YL, Mi JJ, Lv D, et al. Analysis of transmission characteristics of a large HIV-1 molecular cluster in a low-epidemic county. *Chin J Dis Control*. 2020;24(09):1096–1100. doi:10.16462/j.cnki.zhjbkz.2020.09.021
35. Yuan FS, Liu L, Su L, Zeng YL, Liang S. The related factors of new HIV infection among older men in Sichuan, China: a case-control study. *Epidemiol Infect*. 2022;150:e156. doi:10.1017/S0950268822001352
36. Feng Y, Ling S, Xiao L, et al. Molecular epidemiology of HIV-1 in Liangshan Prefecture in 2009. Chinese. *J AIDS STD*. 2014;129(06):403–405. doi:10.13419/j.cnki.aids.2014.06.001
37. 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):2875–2883. doi:10.1097/QAD.0000000000001256
38. Dong Y, Liu S, Xia D, et al. Prediction Model for the risk of HIV Infection among MSM in China: validation and stability. *Int J Environ Res Public Health*. 2022;19(2):1010. doi:10.3390/ijerph19021010
39. Di Giallonardo F, Pinto AN, Keen P, et al. Subtype-specific differences in transmission cluster dynamics of HIV-1 B and CRF01_AE in New South Wales, Australia. *J Int AIDS Soc*. 2021;24(1):e25655. doi:10.1002/jia2.25655
40. Yuan FS, Liang S, Zeng YL, et al. HIV/AIDS knowledge and commercial sexual behavior among middle-aged and elderly men in Sichuan Province. *Chin J AIDS STD*. 2022;223(04):459–463. doi:10.13419/j.cnki.aids.2022.04.17
41. Huang CL, Yang L. Source analysis of HIV testing samples in Suining from 2013 to 2018. *J Prev Med Inf*. 2019;35(10):1143–1147.
42. Wu XM, Yuan FS, Yuan D, et al. Analysis of HIV detection and related factors in Sichuan province in 2020. *Occupat Health Damage*. 2023;38(01):1–6.
43. Huang M, Wang DH, Hou XQ, et al. Analysis of HIV transmission among MSM in Guangyuan from 2007 to 2018. *J Prev Med Inf*. 2019;35(12):1431–1434.

Infection and Drug Resistance

Dovepress

Publish your work in this journal

Infection and Drug Resistance is an international, peer-reviewed open-access journal that focuses on the optimal treatment of infection (bacterial, fungal and viral) and the development and institution of preventive strategies to minimize the development and spread of resistance. The journal is specifically concerned with the epidemiology of antibiotic resistance and the mechanisms of resistance development and diffusion in both hospitals and the community. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/infection-and-drug-resistance-journal>