



OPEN

## HIV-1 subtype diversity and transmission strain source among men who have sex with men in Guangxi, China

Yi Chen<sup>1</sup>, Zhiyong Shen<sup>1</sup>, Yi Feng<sup>1,2</sup>, Yuhua Ruan<sup>1,2</sup>, Jianjun Li<sup>1</sup>, Shuai Tang<sup>1,4</sup>, Kailing Tang<sup>1,4</sup>, Shujia Liang<sup>1,4</sup>, Xianwu Pang<sup>1,4</sup>, Edward B. McNeil<sup>3</sup>, Hui Xing<sup>1,2</sup>, Virasakdi Chongsuvivatwong<sup>3</sup>✉, Mei Lin<sup>1</sup>✉ & Guanghua Lan<sup>1</sup>✉

With the rapid increase in HIV prevalence of men who have sex with men (MSM) in recent years and common human migration and travelling across different provinces in China, MSM are now finding it easier to meet each other, which might contribute to local HIV epidemics as well as fueling cross-province transmission. We performed a cross-sectional survey in 2018–2019 to investigate the current HIV subtype diversity and inferred HIV strain transmission origin among MSM in Guangxi province, China based on a phylogenetic analysis. Based on 238 samples, we found that the HIV-1 subtype diversity was more complicated than before, except for three major HIV subtypes/circulating recombinant forms (CRFs): CRF07\_BC, CRF01\_AE, CRF55\_01B, five other subtypes/CRFs (CRF59\_01B, B, CRF08\_BC, CRF67\_01B, CRF68\_01B) and five unique recombinant forms (URFs) were detected. In total, 76.8% (169/220) of samples were infected with HIV from local circulating strains, while others originated from other provinces, predominantly Guangdong and Shanghai. The high diversity of HIV recombinants and complicated HIV transmission sources in Guangxi MSM indicates that there has been an active sexual network between HIV positive MSM both within and outside Guangxi without any effective prevention. Inter-province collaboration must be enforced to provide tailored HIV prevention and control services to MSM in China.

Phylogenetics can provide an insight into the spread of infectious diseases such as HIV. It is known that even though HIV continually evolves and adapts in infected hosts, there are stringent bottlenecks leading to transmission of a single monophyletic variant to most (>95%) newly-infected persons<sup>1</sup>. Hence, information about the extent of genetic diversity at the population level can be attained by reconstructing HIV phylogenetic trees, including where different viral strains might have originated, and whether new strains are entering populations as an epidemic proceeds<sup>2</sup>. When combining HIV phylogenetic analysis with epidemiological information, HIV epidemiology can be classified into three general categories: (1) molecular Epidemiology, which allows an understanding of the risk factors for HIV transmission and epidemic spread; (2) phylodynamics, which reconstructs the epidemic history and quantifies epidemic growth or decline using viral genealogies and explicit population genetic models; and (3) phylogeography, which describes the distribution of subtype diversity, estimates the impact of human migration on viral spread, and places historical and risk factor data into geographic context in order to identify transmission hubs<sup>3</sup>.

It is well known that phylogenetic analyses have elucidated the origin and spread of the pandemic human immunodeficiency virus type (HIV-1) group M virus, both during the pre-epidemic period and the epidemic phase of spread among humans<sup>4</sup>. Additionally, it is also applied to evaluate small transmission chains between infected individuals, such as some forensic studies<sup>5–7</sup>, and to track the founder events that resulted in the spread of HIV-1 strains across vast geographic areas, specific countries, and within geographically restricted communities<sup>8</sup>.

<sup>1</sup>Institute of HIV/AIDS Prevention and Control, Guangxi Center of Disease Control and Prevention, Nanning 530028, China. <sup>2</sup>State Key Laboratory of Infectious Disease Prevention and Control (SKLID), Chinese Center for Disease Control and Prevention (China CDC), Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing 102206, China. <sup>3</sup>Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Hat Yai 90110, Thailand. <sup>4</sup>These authors contributed equally: Shuai Tang, Kailing Tang, Shujia Liang and Xianwu Pang. ✉email: cvirasak@medicine.psu.ac.th; gxlinmei@126.com; lgh605@163.com

Guangxi Zhuang Autonomous Region (Guangxi) is located in southwest China bordering Vietnam. The HIV epidemic in Guangxi has been driven by injecting drug users since 1996 and shifted to heterosexual transmission from 2006 onwards<sup>9</sup>. Meanwhile, transmission among men who have sex with men (MSM) has also increased and contributed to the HIV epidemic in Guangxi. The first MSM HIV case was reported in 2005 and since then, the number of reported HIV MSM cases has increased from 45 in 2008 to 868 in 2018. Annual surveillance data indicated that the HIV prevalence increased from 0.83% in 2008 to 5.0% in 2012, rose sharply to 11.2% in 2015<sup>10</sup>, and stabilized at around 10% from 2017 to 2018.

A study conducted in Guangxi in 2013 targeting newly diagnosed HIV positive cases aged 18 years or older identified four circulating recombinant forms (CRFs) among MSM cases comprising CRF07\_BC (44.4%), CRF01\_AE (38.9%), CRF55\_01B (11.1%) and CRF08\_BC (5.6%)<sup>11</sup>. However, in recent years, with the development of modern transportation, especially the establishment of high speed rail, MSM traveling to seek sexual partners outside of their local residence has become more common. Gay websites and Gay mobile phone applications are popular mediums for Chinese MSM seeking male sexual partners<sup>12,13</sup>. In order to understand the changing situation of the HIV epidemic in China, information of HIV cross-province transmission among MSM should be documented. The objective of our study is to identify the current diversity of HIV subtypes in Guangxi MSM, and further analyze the temporal-geographical origins of HIV transmission sources.

## Methods

**Study participants and blood collection.** Our study participants were recruited from an antiviral therapy (ART) clinic and a MSM Voluntary Counseling and Testing (VCT) clinic in Nanning from September 2018 to November 2019. These were subjects who had started attending the clinics from 2007 to 2019. In 2018, we recruited 371 subjects, 328 of whom were already on ART, of which HIV DNA was successfully amplified in 192. In 2019, we screened 52 treatment-naïve subjects from a VCT clinic, of which 46 gave HIV RNA amplification. Finally, 238 HIV nucleotide sequences were obtained for genotyping and phylogenetic analysis (see Supplementary Fig. S1 for details). Socio-demographic data comprising age, marital status, area of residence, ethnicity, sexual identity and education level were collected based on face-to-face structured interviews by trained research staff.

**Ethics statement.** The study was approved by the Guangxi institutional review board (approval number GXIRB 2019-0010-1). Informed consent from all participants was obtained after full explanation of the contents was given. All participants in this study were de-identified to maintain their anonymity. All research methods in this study were performed in accordance with the approved guidelines.

**HIV DNA/RNA extraction, amplification, and sequencing.** Whole blood specimens were collected and plasma was isolated and sent under cold chain to the laboratory in our study. Since the patients on antiretroviral treatment were virally suppressed, HIV DNA from subjects on ART were extracted from 200 µl of whole blood using the QIAamp DNA Blood Mini Kit (250) according to the manufacturer's handbook. HIV RNA from ART-naïve subjects was extracted from 200 µl of plasma by the NucliSENS easyMAG platform (BioMérieux, Boxtel, Netherlands) as per the manufacturer's instructions. Then, nested polymerase chain reaction (PCR) was used for viral DNA /RNA to obtain fragments of the pol gene (HXB2, positions 2147–3462 for a total of 1315 bp)<sup>14</sup>. Primers MAW26 (5'-TGGAAATGTGGAAAGGAAGGAC-3' and RT21 (5'-CTGTATTTCTGCTATTAAGTCTTTTGGATGGG-3') in 25 µl reaction volumes was used for pol fragment first amplified with One Step Reverse Transcription PCR (Takara, Dalian, China) and Cycling conditions were as follows: 50 °C for 30 min; 94 °C for 2 min; 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min 30 s; and 72 °C for 10 min. Then, the second PCR was conducted using 2 × Taq PCR MasterMix (Tiangen, Beijing, China) with primers Pro-1 (5'-CAGAGCCAACAGCCCCACCA-3') and RT-20 (5'-CTGCCAGTTCTAGCTCTGCTTC-3') in 50 µl reaction volumes. The following cycling conditions were applied: 94 °C for 5 min; 30 cycles of 94 °C for 30 s, 63 °C for 30 s, and 72 °C for 2 min 30 s; and 72 °C for 10 min. Each step was operated with appropriate negative controls to detect possible contamination throughout the experiments. PCR products were analyzed with 1% agarose gel electrophoresis. Positive PCR products were purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA) and sequenced directly on an ABI 3730XL automated sequencer with BigDye terminators (Applied Biosystems, Foster City, CA, USA) by Beijing Biomed Technology Development Co., Ltd (Beijing, China).

**HIV subtyping.** HIV subtypes were determined based on Maximum-likelihood tree analysis, which was generated using IQ TREE 2.0.6<sup>15</sup> with the best-fitting model of TVM + F + R5. This software can determine best-fitting model for the aligned study sequences automatically<sup>16,17</sup>. The topology of trees was tested by ultrafast bootstrap (UFBoot) approximation with 1000 bootstrap replicates to reduce computing time while achieving more unbiased branch supports<sup>18</sup>. Clusters with a bootstrap value ≥ 0.95 (95%) were defined as a phylogenetic cluster<sup>17</sup>.

248 reference sequences were used for HIV subtype determination. 170 HIV-1 reference sequences of HIV-1 M,N,O,P group were downloaded from Los Alamos National Laboratory (LANL) HIV Sequence Database (<https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>, accessed in July, 2020) by BLAST (see Supplementary S1). 59 CRF reference sequences were obtained by BLAST from China local HIV sequence dataset from the Division of Virology and Immunology, National Center for AIDS/STD Control and Prevention (NCAIDS), China CDC representing the common circulating strains in China, including subtypes of CRF01\_AE, CRF07\_BC, CRF08\_BC, CRF55\_01B and CRF59\_01B. Alignment was done using the "HIV ALIGN" software on the LANL HIV Sequence Database, and manual adjustments were followed using BioEdit software<sup>19</sup>, considering protein coding sequences. Based on the initial Neighbor-joining tree (Reconstructed by MEGA 7.0), most of

the subtypes of sample sequences were found to be close to CRF01\_AE, CRF07\_BC, CRF55\_01B, CRF59\_01B. Therefore, HIV BLAST was performed on HIV sequence database for some sample sequences close to these four subtypes, and the reference sequences achieved were added into the previous reference sequences dataset. Finally, after removal of duplicated sequences, 248 reference sequences for our 238 study sample sequences were obtained, containing 81 CRF01\_AE, 70 CRF07\_BC, 7 CRF55\_01B, 7 CRF5901\_B and 83 other subtype reference sequences.

**HIV genetic network construction.** The genetic network was constructed using HIV-TRACE (Transmission Cluster Engine, <http://demo.hivtrace.org/network.html>)<sup>20</sup>. Sample HIV pol sequences were aligned to HXB2 reference sequence and the pairwise genetic distances were calculated under the Tamura-Nei 93 model<sup>21</sup>. 8 sequences with ambiguous nucleotides were removed. The genetic distance between nodes (each node represented a sequence or an individual) up to 0.15% substitutions per site were linked to each other. The genetic distance threshold of 0.15% was recommended by the Centers for Disease Control and Prevention (CDC) in the United States (US)<sup>22</sup>.

**HIV transmission strain source inference.** HIV transmission strain sources were analyzed using a Bayesian temporal-geographical origin analysis. As there were only four sequences with CRF59\_01B, three sequences with subtype B, two sequences for CRF\_08BC and one for CRF67\_01B, CRF68\_01B, CRF01\_BC respectively, no phylogenetic trees were needed to determine their transmission source. We determined the transmission source of the strains among HIV positive MSM individuals with CRF01\_AE, CRF07\_BC, and CRF55\_01B, respectively.

Regarding the pol gene (PR-RT region) reference sequences selection, which was the second reference dataset in our study, the five most similar sequences to each sample sequence from the Los Alamos HIV Sequence Database were selected using the “HIV BLAST” tool (accessed June, 2020). 1,100 reference sequences were obtained and duplicated sequences and those without sampling year or geographical information were removed. Finally, 374 unique pol sequences with sampling time and geographical information were obtained and used as reference sequences for determining HIV transmission strain sources for our 220 pol gene sample sequences. Among the 374 reference sequences, one was from Vietnam and the rest were from China. The 594 (374 + 220) pol sequences were separated into three datasets by different subtypes, CRF01\_AE (n = 299), CRF07\_BC (n = 243), and CRF55\_01B (n = 52). These three sequence datasets were aligned by “HIV ALIGN” on the Los Alamos HIV Sequence Database (accessed June, 2020).

To reconstruct the spatial dynamics and estimate the sources of the strains of our study participants, a Bayesian discrete phylogeographic approach was performed using Markov chain Monte Carlo (MCMC) runs of 300 million generations with BEAST v.1.8.4 under a Bayesian skygrid demographic model. The first 10–30% of the states from each run were discarded as the burn-in<sup>23,24</sup>. All three data sets were analyzed using a general time-reversible (GTR) model specifying a gamma distribution based on a Neighbor-Joining tree as a prior on each relative substitution rate and a relaxed uncorrelated lognormal (UCLN) molecular clock model to infer the timescale of HIV evolution with a gamma distribution prior on the mean clock rate (shape = 0.001, scale = 1000)<sup>25</sup>. The Bayesian MCMC output was analyzed using Tracer v1.6 (<http://beast.bio.ed.ac.uk/Tracer>). The Maximum clade credibility (MCC) trees were generated in Tree Annotator v1.8.4. Finally, the most probable temporal and geographical origin of HIV transmission strains for the study participants were inferred according to the output of the posterior of Bayesian estimation and visualized on MCC trees using the FigTree software v1.4.3 (<http://beast.bio.ed.ac.uk>). The tree figure editing was using Adobe Illustrator 2020(V 24.0.1.341).

Sequences with posterior probability over 0.7 were defined as Guangxi local circulating strains, and the posterior probability refers to the support of locations at the nodes. The HIV transmission strain source provinces of our study samples were then tabulated by different HIV subtypes.

**Statistical analysis.** Statistical analyses for this study were performed using the R language and environment (R version 4.0.0). The distribution of HIV subtypes with epidemiological characteristics were described. Chi-squared tests were used to test the associations between HIV subtypes and epidemiological characteristics, as well as the association between HIV subtype and HIV transmission source.

**Study nucleotide sequence.** Considering the risk of breaching patient confidentiality, we submitted a random sample of 25% of all HIV pol gene nucleotide sequences obtained from our study to GenBank under accession numbers MW573881- MW573938. (see Supplementary S2 for details).

## Results

**Demographic characteristics of study samples.** Table 1 summarizes demographic information of the 238 study participants. The sample was predominated by young adults, single, rural residents, Han ethnicity, homosexual sexual identity, and a college or higher education level. Three major HIV subtypes were identified; CRF01\_AE, CRF07\_BC, and CRF55\_01B with a small proportion of others. The only variable associated with HIV subtypes was area of residence; except for the CRF01\_AE subtype, a higher proportion of participants with other subtypes lived in a rural area.

**HIV subtypes among MSM in Guangxi.** Figure 1 illustrates the maximum-likelihood phylogenetic tree based on sample and reference sequences. References with a red clade represents those commonly identified around the world. The samples were predominated by three major HIV CRFs: CRF07\_BC, CRF01\_AE,

Variable	CRF01_AE	CRF07_BC	CRF55_01B	Others	Total	P value ( $\chi^2$ test)
Total	89	107	24	18	238 (100.0)	
<b>Age group (years)</b>						
≤ 25	34 (40.0)	43 (43.0)	8 (36.4)	5 (29.4)	90 (40.2)	0.878
26–35	34 (40.0)	41 (41.0)	9 (40.9)	7 (41.2)	91 (40.6)	
> 36	17 (20.0)	16 (16.0)	5 (22.7)	5 (29.4)	43 (19.2)	
<b>Marital status</b>						
Married	9 (10.3)	11 (10.4)	3 (12.5)	5 (27.8)	28 (11.9)	0.214
Single	78 (89.7)	95 (89.6)	21 (87.5)	13 (72.2)	207 (88.1)	
<b>Residence</b>						
Rural	38 (43.7)	69 (65.7)	15 (62.5)	10 (55.6)	132 (56.4)	0.020
Urban	49 (56.3)	36 (34.3)	9 (37.5)	8 (44.4)	102 (43.6)	
<b>Ethnicity</b>						
Han	50 (61.7)	62 (62.0)	12 (54.5)	11 (61.1)	135 (61.1)	0.931
Zhuang	31 (38.3)	38 (38.0)	10 (45.5)	7 (38.9)	86 (38.9)	
<b>Sexual identity</b>						
Bisexual	16 (19.5)	19 (19.6)	4 (19.0)	4 (22.2)	43 (19.7)	0.984
Homosexual	66 (80.5)	78 (80.4)	17 (81.0)	14 (77.8)	175 (80.3)	
<b>Education</b>						
Junior school or less	9 (10.3)	13 (12.4)	4 (17.4)	6 (33.3)	32 (13.7)	0.178
High school	32 (36.8)	38 (36.2)	10 (43.5)	7 (38.9)	87 (37.3)	
College or higher	46 (52.9)	54 (51.4)	9 (39.1)	5 (27.8)	114 (48.9)	
<b>HIV diagnosed year</b>						
2007–2016	40 (44.9)	48 (44.9)	7 (29.2)	5 (27.8)	100 (42.0)	0.285
2017–2019	49 (55.1)	59 (55.1)	17 (70.8)	13 (72.2)	138 (58.0)	

**Table 1.** Demographic characteristics for different HIV-1 genotypes among MSM in Nanning. Column totals may not necessarily add up due to missing values.

CRF55\_01B. Five other subtypes/CRFs were CRF59\_01B, B, CRF08\_BC, CRF67\_01B, and CRF68\_01B. Five unique recombinant forms (URFs) were discovered (shown with a blue clade in Fig. 1).

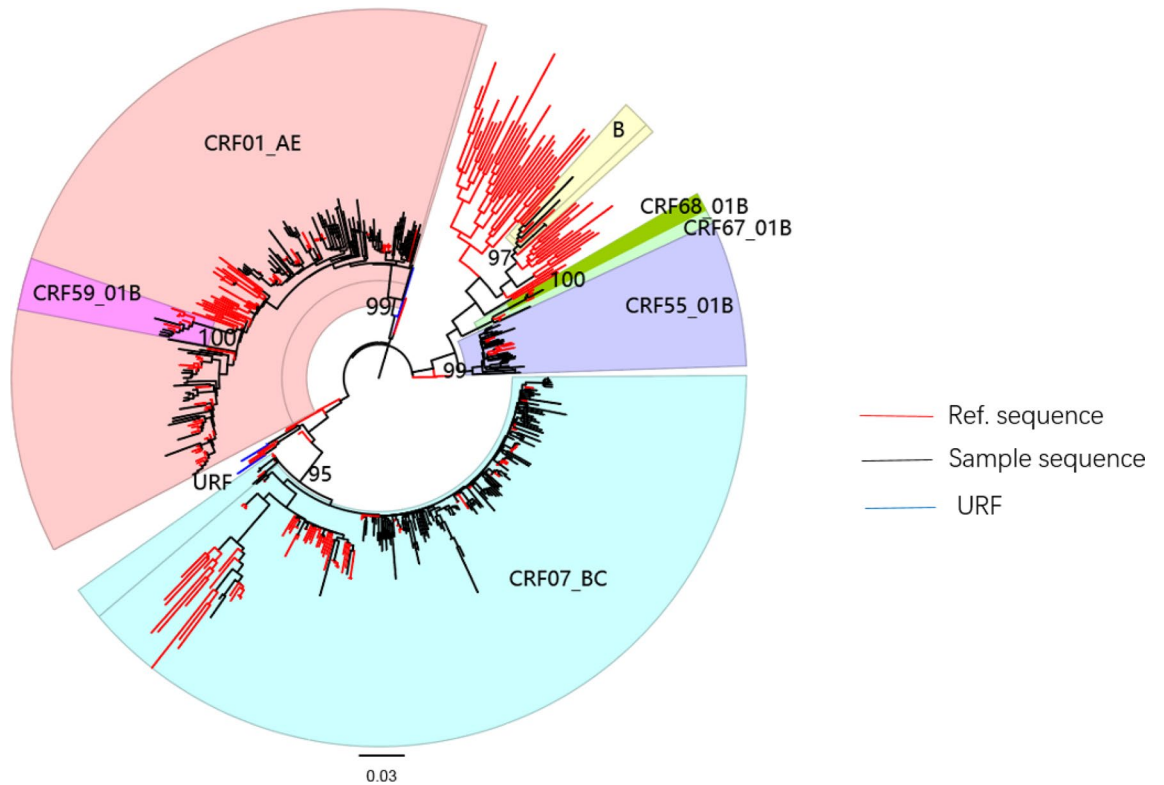
### HIV genetic network construction

Under the genetic distance threshold of 0.15%, 52.6% (121/230) of nodes (sequences) were included into the genetic network, 25 clusters were identified, and the largest cluster consisted of 38.8% (47/121) of nodes, which had subtype CRF07\_BC. The sizes of the remaining clusters ranged from 2 to 7. This indicates that sequences with CRF07\_BC tended to gather as one big cluster than other subtypes. The result is shown in Supplementary Figure S2.

**HIV transmission strain sources.** Subtypes/CRF B, CRF59\_01B, CRF08\_BC, CRF67\_01B and CRF68\_01B were excluded from the analysis due to small sample sizes ( $n < 5$ ).

The HIV transmission strain sources for 220 sample sequences with CRF01\_AE, CRF07\_BC, CRF55\_01B were visualized on MCC trees (see Supplementary Fig. S3 for details). Based on the MCC trees, most Guangxi local circulating strains were spread directly from Guangdong province. There were two clusters of HIV CRF01\_AE strains originating from Guangdong province and then spreading directly into Guangxi, or via Shanghai/Liaoning province into Guangxi. HIV CRF07\_BC strains were spread into Guangxi by three clusters, and most Guangxi local strains were spread directly from Guangdong, followed by originating in Guangdong province and then spreading through Shanghai into Guangxi. While there were two clusters spreading HIV CRF55\_01B strains from Guangdong to Guangxi directly, these developed into Guangxi local circulating strains. The temporal signal was indicated by the Bayesian phylogeographic analysis results. The probable time of introduction for CRF01\_AE strains with effective population size circulating in Guangxi MSM was probably in 2006.2 (2004.3–2008.4) from Liaoning province but originated from Guangdong, for CRF07\_BC strains was probably in 2008.6 (2006.9–2010.5) from Guangdong, and CRF55\_01B was probably in 2008.02 (2007.5–2008.6) from Guangdong. The temporal signal was indicated by a red dot on the Bayesian MCC trees for three subtypes and shown in Supplementary Fig. S3.

The source provinces were summarized by three common subtypes and groups of participants in Tables 2 and 3. Based on the criteria of posterior probability over 0.7, transmission source of 196 samples (89.1%) could be determined. The HIV transmission source of 169 samples (76.8%) were Guangxi local circulating strains. The remaining 27 strains were from outside Guangxi. 101 (96.2%) of 105 HIV strains with CRF07\_BC were primary local strains in Guangxi MSM community, compared to 50 (74.6%) of 67 CRF01\_AE and 18 (75.0%) of 24 CRF55\_01B (Fisher's exact test,  $p < 0.001$ ). The subtype with the most diversified source was CRF01\_AE.



**Figure 1.** Maximum-likelihood phylogenetic trees of HIV-1 pol sequences from 238 MSM in Nanning, Guangxi, China. The topology of trees was tested by IQ Tree (V.2.0.6<sup>15</sup>)ultrafast bootstrap (UFBoot) approximation with 1000 bootstrap replicates, the bootstrap values were showed on the corresponding nodes, Clusters with bootstrap value  $\geq 0.95$  (95%) were defined as phylogenetic cluster. Red clade denoted reference sequence, black clade denoted samples sequences and blue clade represented URF.

Posterior probability	Province	Total	Subtype			Sample source	
			CRF01_AE	CRF07_BC	CRF55_01B	ART clinic	VCT clinic
> 0.7	Guangxi	169	50	101	18	140	29
	Guangdong	17	10	1	6	16	1
	Shanghai	6	3	3	0	1	5
	Beijing	1	1	0	0	1	0
	Shanxi	2	2	0	0	1	1
	Liaoning	1	1	0	0	1	0
< 0.7		24	22	2	0	20	4
Total		220	89	107	24	180	40

**Table 2.** Distribution of HIV transmission strains circulating provinces.

Province	Total	Subtype			<i>P</i> (Fisher's Exact test)	Sample source		<i>P</i> (Fisher's Exact test)
		CRF01_AE	CRF07_BC	CRF55_01B		ART clinic	VCT clinic	
Guangxi	169	50	101	18	< 0.001	140	29	0.288
Non-Guangxi	27	17	4	6		20	7	

**Table 3.** Comparison of HIV transmission source by different subtype and sample group.

Subjects from ART clinic and those from VCT clinic did not differ in source of their HIV strains (Fisher's exact test,  $p = 0.288$ ).

## Discussion

In this study we found that the majority of HIV positive MSM in Guangxi were young, single, identified themselves as homosexual, and were well educated. The proportion aged 25 years or less was higher than those in studies from Shenzhen, Beijing and Shanghai<sup>26–28</sup>, as well as the United States<sup>29</sup>. Our sample was similar to those in Kunming<sup>30</sup> and Thailand<sup>31</sup>.

The three major CRFs identified were CRF07\_BC, CRF01\_AE and CRF55\_01B. Additionally, five CRFs (four of which were newly detected in Guangxi in recent five years) and five URFs were detected. We created a table (see Supplementary Table S1) to compare the HIV subtype/CRF diversity in western and Asian countries, as well as provinces or large cities in China. The HIV subtype/CRF diversity in Guangxi MSM in the current study was more complicated than that in 2013<sup>11</sup> and those in southeast countries such as Thailand<sup>31</sup>, Malaysia<sup>32</sup> and Singapore<sup>33</sup>. Degrees of diversity were high in Shanghai<sup>28</sup>, Guangdong<sup>26</sup>, Beijing<sup>34</sup> and Sichuan<sup>35</sup>. The predominant CRFs in our study were similar to other provinces in China<sup>26, 35, 36</sup>, but different from those in southeast Asian countries<sup>31–33</sup>, United States<sup>29</sup>, Canada and European countries<sup>37–41</sup>. The newly detected CRFs in our study had been circulating in the northeast and middle and southern parts of China since 2013<sup>42–46</sup>, which concurred with the results from our temporal geographical analysis.

In our study, genetic network analysis identified one large cluster with subtype CRF07\_BC, while the Bayesian phylogeographic analysis confirmed that the majority of Guangxi local HIV circulating strains were CRF07\_BC. This subtype increased rapidly in recent years and has been seen among MSM in many cities of China. Among Shenzhen MSM, CRF07\_BC increased from 12.5% in 2006 to 43.2% in 2012<sup>26</sup>. In Fujian, the proportion of CRF07\_BC infections among recently infected MSM expanded rapidly from 19.0% in 2012 to 41.9% in 2013<sup>47</sup>. A phylodynamic analysis also revealed that the CRF07\_BC strain had driven two rapid HIV spreading waves with effective population size in China, the second wave coincided with the expanding MSM cluster<sup>48</sup>. HIV CRF01\_AE, the second-most common subtype in our study, originated in Thailand in the early 1990s<sup>49, 50</sup>. This subtype had been circulating among IDUs and sexually active populations in China for over 30 years<sup>51, 52</sup>. It is relatively virulent in that it causes rapid CD4 T-cell decline and rapid disease progression<sup>53</sup>. This may explain its ineffective transmission and reported decline since 2010.

More than 86% of HIV infections from our sample originated from Guangxi local circulating strains. For strains circulating outside Guangxi, all except four were probably transmitted from economic developed cities of China, such as Guangdong and Shanghai. The most probable time of the three common HIV CRF strains were effectively introduced into Guangxi MSM populations between 2006 and 2008.

The source provinces of HIV strains were not neighboring provinces of Guangxi with low economic level such as Yunnan and Guizhou. In fact, most were from large cities of China including Guangdong and Shanghai. They contributed to a new emergence of CRFs and URFs of HIV and increased the complexity of the HIV epidemic among MSM populations in Guangxi. Previous studies reported that MSM living in south-western rural China tended to migrate to the eastern urban areas of China<sup>54, 55</sup> where the HIV prevalence is high<sup>56–58</sup>. This migration might have facilitated HIV transmission back to the provinces where these MSM came from. Although a large proportion of MSM tended to be infected from local circulating strains, the proportion of MSM infected with HIV from non-local circulating strains (13.7%) is still higher than that of Guangxi heterosexuals (5.9%)<sup>11</sup> which suggests that the MSM sexual network is broader and more complicated than the heterosexual one.

Our study has certain limitations. First, the samples were collected from an ART clinic and a VCT clinic. Although we could get DNA sequences from ART clinic subjects, we could not get RNA sequences because their virus had been well suppressed. However, this has little effect on the HIV strain transmission source inference or genetic network study<sup>59</sup>. On the other hand, the subjects from the VCT clinic might not well represent recently infected MSM, most of whom might not have access to this service.

In summary, a high diversity of HIV recombinants and complicated HIV transmission sources were found in Guangxi MSM. This suggests that there has been an active sexual network between HIV positive MSM both within and outside Guangxi without effective prevention. Inter-province collaboration must be enforced to provide tailored HIV prevention and control services to MSM in China.

Received: 23 September 2020; Accepted: 30 March 2021

Published online: 15 April 2021

## References

- Brenner, B. G., Ibanescu, R.-I., Hardy, I. & Roger, M. Genotypic and phylogenetic insights on prevention of the spread of HIV-1 and drug resistance in 'real-world' settings. *Viruses* **10**, (2017).
- Holmes, E. C. & Garnett, G. P. Genes, trees and infections: Molecular evidence in epidemiology. *Trends Ecol. Evol.* **9**, 256–260 (1994).
- Dennis, A. M. *et al.* Phylogenetic studies of transmission dynamics in generalized HIV epidemics: An essential tool where the burden is greatest?. *J. Acquir. Immune Defic. Syndr.* **1999**(67), 181–195 (2014).
- Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* **18**, 182–192 (2012).
- Ou, C. Y. *et al.* Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**, 1165–1171 (1992).
- Lemey, P. *et al.* Molecular testing of multiple HIV-1 transmissions in a criminal case. *AIDS Lond. Engl.* **19**, 1649–1658 (2005).
- Abecasis, A. B., Pingarilho, M. & Vandamme, A.-M. Phylogenetic analysis as a forensic tool in HIV transmission investigations. *AIDS* **32**, 543–554 (2018).
- Gräf, T., Delatorre, E. & Bello, G. Phylogenetics applied to the human immunodeficiency virus type 1 (HIV-1): from the cross-species transmissions to the contact network inferences. *Mem. Inst. Oswaldo Cruz* **115**, e190461 (2020).

9. Chinese Center for Disease Control and Prevention. Analysis of HIV/STD epidemic in 2007. Beijing, China. [http://www.china.cdc.cn/jkzt/crb/zl/azb/jszl\\_2219/201103/t20110311\\_28108.html](http://www.china.cdc.cn/jkzt/crb/zl/azb/jszl_2219/201103/t20110311_28108.html).
10. Guanghua, L. *et al.* HIV, syphilis and behavioral risk factors among men who have sex with men in a drug-using area of south-western China: Results of 3 cross-sectional surveys from 2013 to 2015. *Medicine (Baltimore)* **97**, e0404 (2018).
11. Li, J. *et al.* HIV-1 transmissions among recently infected individuals in Southwest China are predominantly derived from circulating local strains. *Sci. Rep.* **8**, 12831 (2018).
12. Dong, M.-J. *et al.* The prevalence of HIV among MSM in China: A large-scale systematic analysis. *BMC Infect. Dis.* **19**, 1000 (2019).
13. Wei, L. *et al.* Use of gay app and the associated HIV/syphilis risk among non-commercial men who have sex with men in Shenzhen, China: a serial cross-sectional study. *Sex. Transm. Infect.* **95**, 496–504 (2019).
14. Liao, L. *et al.* Genotypic analysis of the protease and reverse transcriptase of HIV type 1 isolates from recently infected injecting drug users in western China. *AIDS Res. Hum. Retroviruses* **23**, 1062–1065 (2007).
15. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
16. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
17. Minh, B. Q., Trifinopoulos, J., Schrempf, D. & Schmidt, H. A. IQ-TREE version 2.0: Tutorials and Manual Phylogenomic software by maximum likelihood. 140.
18. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
19. Hall T A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT[C]//Nucleic acids symposium series. [London]: Information Retrieval Ltd., c1979-c2000., 1999, 41(41): 95–98.
20. Kosakovsky Pond, S. L., Weaver, S., Leigh Brown, A. J. & Wertheim, J. O. HIV-TRACE (TRANsmiSSion Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol. Biol. Evol.* **35**, 1812–1819 (2018).
21. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
22. Detecting and Responding to HIV Transmission Clusters - a Guide for Health Departments. June 2018. Draft Version 2.0. 132.
23. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
24. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
25. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
26. Zhao, J. *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.* **6**, 28703 (2016).
27. Song, Y. X. *et al.* Prevalence of transmitted drug resistance among HIV-1 treatment-naïve patients in Beijing. *Epidemiol. Infect.* **146**, 339–344 (2018).
28. Li, X. *et al.* HIV-1 genetic diversity and its impact on baseline CD4+T cells and viral loads among recently infected men who have sex with men in Shanghai China. *PLoS ONE* **10**, e0129559 (2015).
29. Oster, A. M. *et al.* Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013. *Ann. Epidemiol.* **27**, 244–251.e1 (2017).
30. Chen, M. *et al.* HIV-1 genetic transmission networks among men who have sex with men in Kunming, China. *PLoS ONE* **13**, (2018).
31. Chang, D. *et al.* Molecular epidemiology of a primarily MSM acute HIV-1 cohort in Bangkok, Thailand and connections within networks of transmission in Asia. *J. Int. AIDS Soc.* **21**, e25204 (2018).
32. Ng, K. T. *et al.* Evolutionary history of HIV-1 subtype B and CRF01\_AE transmission clusters among men who have sex with men (MSM) in Kuala Lumpur Malaysia. *PLoS ONE* **8**, e67286 (2013).
33. Ng, K. T. *et al.* Phylodynamic profile of HIV-1 subtype B, CRF01\_AE and the recently emerging CRF51\_01B among men who have sex with men (MSM) in Singapore. *PLoS ONE* **8**, e80884 (2013).
34. Sun, L. *et al.* Multiple HIV-1 subtypes were found circulating in Shijingshan District of Beijing China. *AIDS Res. Hum. Retroviruses* **35**, 494–499 (2019).
35. Yuan, D. *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)* **98**, e17585 (2019).
36. Hao, M. *et al.* Identification of a Novel HIV-1 Second-Generation Recombinant Form (CRF01\_AE/07\_BC) in Men Who Have Sex with Men in Beijing China. *AIDS Res. Hum. Retroviruses* **35**, 500–504 (2018).
37. Paraskevis, D. *et al.* HIV-1 molecular transmission clusters in nine European countries and Canada: Association with demographic and clinical factors. *BMC Med.* **17**, 4 (2019).
38. Frange, P. *et al.* HIV-1 subtype B-infected MSM may have driven the spread of transmitted resistant strains in France in 2007–12: impact on susceptibility to first-line strategies. *J. Antimicrob. Chemother.* **70**, 2084–2089 (2015).
39. Pérez-Parra, S. *et al.* Phylodynamic and phylogeographic profiles of subtype B HIV-1 epidemics in South Spain. *PLoS ONE* **11**, e0168099 (2016).
40. Parczewski, M. *et al.* Expanding HIV-1 subtype B transmission networks among men who have sex with men in Poland. *PLoS ONE* **12**, e0172473 (2017).
41. Kazennova, E. *et al.* Genetic variants of HIV Type 1 in men who have sex with men in Russia. *AIDS Res. Hum. Retroviruses* **33**, 1061–1064 (2017).
42. Han, X. *et al.* Genome sequences of a novel HIV-1 circulating recombinant form (CRF59\_01B) identified among men who have sex with men in Northeastern China. *Genome Announc.* **1**, (2013).
43. Jia, D. *et al.* Two-year cross-sectional studies reveal that single, young MSMs in Shenzhen, China are at high risk for HIV infection. *Viol. J.* **16**, 83 (2019).
44. Zhang, W. *et al.* Trends of HIV-1 subtypes among young people in Hangzhou China. *AIDS Res. Hum. Retroviruses* **33**, 219–227 (2017).
45. Yang, Y. *et al.* Phylogenetic and temporal dynamics of human immunodeficiency virus type 1 CRF01\_AE and CRF07\_BC among recently infected antiretroviral therapy-naïve men who have sex with men in Jiangsu province, China, 2012 to 2015: A molecular epidemiology-based study. *Medicine (Baltimore)* **97**, e9826 (2018).
46. Zhang, D. *et al.* Genetic characterization of HIV-1 epidemic in Anhui Province China. *Viol. J.* **17**, 17 (2020).
47. Chen, Z.-W. *et al.* Surging HIV-1 CRF07\_BC epidemic among recently infected men who have sex with men in Fujian China. *J. Med. Virol.* **90**, 1210–1221 (2018).
48. Zhang, M. *et al.* Phylodynamic analysis revealed that epidemic of CRF07\_BC strain in men who have sex with men drove its second spreading wave in China. *AIDS Res. Hum. Retroviruses* **33**, 1065–1069 (2017).
49. Pau, C. P. *et al.* Highly specific V3 peptide enzyme immunoassay for serotyping HIV-1 specimens from Thailand. *AIDS Lond. Engl.* **7**, 337–340 (1993).
50. Wirachsilp, P. *et al.* Surveillance of subtype and genetic variation of the circulating strains of HIV-1 in Thailand. *Southeast Asian J. Trop. Med. Public Health* **38**, 814–827 (2007).

51. Su, Y., Liu, H., Wu, J., Zhu, L. & Wang, N. Distribution of HIV-1 genotypes in China: A systematic review. *Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi* **35**, 1164–1168 (2014).
52. He, X. *et al.* A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. *PLoS ONE* **7**, e47289 (2012).
53. Cui, H. *et al.* Rapid CD4+ T-cell decline is associated with coreceptor switch among MSM primarily infected with HIV-1 CRF01\_AE in Northeast China: *AIDS* **33**, 13–22 (2019).
54. Qin, Q. *et al.* The characteristics of HIV-positive men who have sex with men in China and predictors of their migration, 2008–2015. *Zhonghua Yu Fang Yi Xue Za Zhi* **50**, 938–942 (2016).
55. Mi, G. *et al.* Hidden and mobile: A web-based study of migration patterns of men who have sex with men in China. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **62**, 1443–1447 (2016).
56. Weng, R.-X., Hong, F.-C., Yu, W.-Y. & Cai, Y.-M. Compare HIV/syphilis infections between age groups and explore associated factors of HIV/syphilis co-infections among men who have sex with men in Shenzhen, China, from 2009 to 2017. *PLoS ONE* **14**, e0223377 (2019).
57. Sun, Y. M. *et al.* Surveillance for HIV infection in MSM selected through respondent driven sampling in Beijing, 2005–2012. *Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi* **37**, 1383–1391 (2016).
58. Lin, H. *et al.* High prevalence of HIV infection and bisexual networks among a sample of men who have sex with men in Eastern China. *PLoS ONE* **10**, e0129300 (2015).
59. Kang, R. *et al.* Using longitudinal genetic-network study to understand HIV treatment-as-prevention: a population-based observational study. *AIDS* **Published Ahead of Print**, (2021).

## Acknowledgements

This work was supported by grants of Guangxi Key Laboratory of AIDS Prevention Control and Translation, Guangxi Bagui Honor Scholar, National Natural Science Foundation of China (11971479), Guangxi Natural Science Foundation Project (grants 2020GXNSFAA159020), and National Natural Science Foundation of China (81502862). The authors would like to thank Dr. Hongyan Lu and Dr. Zhi Gan for samples collection in local fields, and Dr. Lei Liu, Dr. Aibo Dong, Dr. Mengze Gan, Dr. Kang Li for technical support.

## Author contributions

Y.F., Y.C., Z.Y.S., H.X. conceived and designed the study, J.J.L., K.L.T., S.J.L., X.W.P., S.T. collected the samples and performed the experiments, Y.C. analyzed the data and drafted the manuscript, V.C., G.H.L., Y.H.R., M.L. and E.M. provided critical comments. All authors reviewed and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87745-3>.

**Correspondence** and requests for materials should be addressed to V.C., M.L. or G.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021