Nonstudent Young Men Put Students at High Risk of HIV Acquisition in Guangxi, China: A Phylogenetic Analysis of Surveillance Data

He Jiang,[1,2,3,a] Guanghua Lan,[1,a] Qiuying Zhu,[1,a] Shujia Liang,[1,a] Jianjun Li,[1,a] Yi Feng,[2,a] Mei Lin,[1,a] Hui Xing,[2,a] and Yiming Shao[2,3]

[1]Guangxi Key Laboratory of Major Infectious Disease Prevention Control and Biosafety Emergency Response, Guangxi Center for Disease Prevention and Control, Nanning, China, [2]State of Key Laboratory for Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, and [3]Guangxi Key Laboratory of AIDS Prevention and Treatment & Guangxi Universities Key Laboratory of Prevention and Control of Highly Prevalent Disease, School of Public Health, Guangxi Medical University, Nanning, Guangxi, China

Background. We sought to identify students and their sexual partners in a molecular transmission network.

Methods. We obtained 5996 HIV protease and reverse transcriptase gene sequences in Guangxi (165 from students and 5831 from the general populations) and the relevant demographic data. We constructed a molecular transmission network and introduced a permutation test to assess the robust genetic linkages. We calculated the centrality measures to describe the transmission patterns in clusters.

Results. At the network level, 68 (41.2%) students fell within the network across 43 (8.1%) clusters. Of 141 genetic linkages between students and their partners, only 25 (17.7%) occurred within students. Students were more likely than random permutations to link to other students (odds ratio [OR], 7.2; $P < .001$), private company employees aged 16–24 years (OR, 3.3; $P = .01$), private company or government employees aged 25–49 years (OR, 1.7; $P = .03$), and freelancers or unemployed individuals aged 16–24 years (OR, 5.0; $P < .001$). At the cluster level, the median age of nonstudents directly linked to students (interquartile range) was 25 (22–30) years, and 80.3% of them had a high school or higher education background. Compared with students, they showed a significantly higher median degree (4.0 vs 2.0; $P < .001$) but an equivalent median Eigenvector Centrality (0.83 vs 0.81; $P = .60$).

Conclusions. The tendency of genetic linkage between students and nonstudent young men and their important position in the HIV transmission network emphasizes the urgent need for 2-pronged public health interventions based on both school and society.

Keywords. HIV; network centrality; student; transmission network.

Students are the highest priority group for HIV prevention in China [1]. Between 2011 and 2015, the number of newly reported HIV cases in students aged 15–24 years increased, ranging from 15% to 50% annually, reaching 3236 in 2015 [2]. Although the increasing trend has slowed down since the Chinese government launched a series of policies to increase HIV health care services for students in 2015, students continue to be inordinately vulnerable to HIV in the existing social network [3]. In 2019, newly reported HIV cases in students accounted for 21.7% (3422/15 790) of the cases in the overall population of the same age [4].

Hence, some key populations may not have been covered by the existing policies. Although the 2-pronged interventions targeting both transmission source and susceptible populations are effective at reducing HIV incidence [5], the current research is limited to school-based prevention, such as conducting youth risk behavior surveillance [6], promoting HIV knowledge and sexual health education [7], and encouraging self-testing, preexposure prophylaxis, and postexposure prophylaxis [8–10]. The bottleneck of the current research is that the traditional questionnaire method cannot effectively uncover the overlapping transmission networks within students or between students and nonstudents. Fortunately, an increasing number of studies have supported that phylogenetic analysis can provide a method for determining the groups among whom HIV is spreading [11]. Here, we constructed a molecular transmission network based on 2 province-wide surveillance databases, namely the HIV sequences database and the National HIV/AIDS Comprehensive Response Information Management System for Guangxi. We introduced a permutation test to assess sampling bias. Moreover, we calculated the centrality measures and plotted target diagrams to illustrate the network position of individuals in clusters. We hypothesized that the network position would reflect the transmission patterns.

METHODS

Data Sources

For this phylogenetic analysis, we collected the sequence data from Guangxi Center for Diseases Prevention and Control (GXCDC). The HIV protease and reverse-transcriptase genetic sequences (*pol*, HXB2 coordinates: 2253–3554 nt, minimum length 1000 nt, proportion of ambiguous nucleotides <5%) for the province-wide cross-sectional HIV molecular surveys (between 1997 and 2020) were extracted, each representing a distinct group of people with HIV (PWH). The sequence data were linked to the National HIV/AIDS Comprehensive Response Information Management System for Guangxi. The process of selecting participants is described in a flow diagram (Supplementary Figure 1). Students were eligible if they met the following criteria: (1) aged 16 years and above; (2) self-reported as students and were confirmed by student ID cards; (3) attending high schools, colleges, or universities in Guangxi and residing locally when sampling; (4) diagnosed as HIV-positive in Guangxi. PWH who did not meet the above criteria were marked as nonstudents. Eventually, 165 *pol* sequences from 165 students were obtained.

To build the largest background sequences database, we collected all available HIV *pol* sequences from 2 sources: (1) the GXCDC HIV sequences database mentioned above; (2) all published HIV *pol* sequences (with the same quality control requirement as sequences from GXCDC) from Los Alamos National Laboratories HIV Sequence Database (https://www.hiv.lanl.gov/). When 2 or more sequences from an individual were available, the earliest sequence was retained. We retrieved 171 365 background sequences.

Patient Consent

The Ethics Review Board of Guangxi Center for Diseases Prevention and Control approved the study (Certificate No. GXIRB2016-0047-1). Each PWH signed written consent before blood sampling, allowing the use of the blood samples and anonymized data in future epidemiological studies. We signed a confidentiality agreement and were authorized to use these databases for this study.

Transmission Network Construction

The HIV *pol* sequences from students and the background population were merged into a sequence database. All *pol* sequences were automatically aligned by MAFFT (version 7) [12] and visually inspected by BioEidit (version 7.1). The HIV subtypes were determined based on an approximate maximum likelihood phylogenetic tree analysis, which was constructed by IQ-TREE (version 1.6.12) [13] with a general time-reversible model and was visualized by FigTree (version 1.4.3). Using the Tamura-Nei model, 93 pairwise nucleotide genetic distances were calculated between each pair of sequences in the merged database. A genetic distance threshold (GD) of 0.5% was used to infer the rapid transmission clusters [14]. As no background sequence obtained outside Guangxi directly linked to student sequences at this threshold, we considered that recent HIV transmissions among students in Guangxi are predominantly derived from circulating local strains. To improve the specificity in identifying high-risk transmission subpopulations, we finally included 165 HIV *pol* sequences from students and 5831 Guangxi background sequences. The number of newly diagnosed PWH and the sampling rates of HIV *pol* sequences by year are shown in Supplementary Table 1. The HIV molecular transmission network was constructed using MicrobeTrace (version 0.7.0) [15].

Statistical Analysis

At the network level, we compared the predictive factors of falling within the Guangxi network and falling within the clusters containing at least 1 student (student cluster) using univariable and multivariable logistic regression models. Age was classified into 3 categories: (1) age 16–24 years, representing young adults; (2) age 25–49 years, representing adults, and (3) age ≥50 years, representing the elderly. The occupation was classified into 4 categories: (1) student; (2) farm or factory worker, representing the relatively low-income subpopulations living in rural areas or temporarily migrating to urban areas [16]; (3) freelancer, unemployed, or retired, representing the relatively low-income subpopulations living in urban areas; and (4) private company or government employees, representing the relatively high-income subpopulations living in urban areas. To minimize the impact of missing data (Supplementary Table 2) on the coefficient estimates of predictive factors, we applied multiple imputations by chained equations [17] to obtain suitable imputed data before fitting the logistic regression model. A sensitivity analysis was also performed to inspect the distribution of original and imputed data (Supplementary Table 3). To address potential sampling bias and test whether the network contained more robust linkages between students and their partners than randomized linkages, we introduced a permutation test, which has been detailed previously [18]. Briefly, randomized pairwise genetic linkages were generated by randomly permuting subpopulation labels on the network 1000 times. Statistical significance was calculated by comparing the observed proportion of pairwise genetic linkages with the distribution for the proportion of randomized pairwise genetic linkages.

At the cluster level, the individuals in the student clusters were classified into 3 subgroups: (1) students; (2) nonstudents directly linked to students (DLS); and (3) nonstudents indirectly linked to students (ILS). We compared the centrality measures between the 3 subgroups using Dunn's Kruskal-Wallis test for multiple comparisons. In a cluster, a node represents an individual, and an edge represents a link between 2 nodes. The centrality measures in this study included [19]: (1)

Degree, the sum of edges connected to a node, represents the gregariousness of a node. (2) Betweenness Centrality is the number of shortest paths between 2 nodes through a given node. A node with a high Betweenness Centrality tends to provide a bridge between two large clusters. (3) Eigenvector Centrality, the sum of the weighted Degree of a node and its neighbors, represents the potential importance of a node in the cluster.

All statistical tests were 2-sided, and $P < .05$ was considered statistically significant. Adjusted P for multiple comparisons was calculated by Dunn's Kruskal-Wallis test. Statistical analyses were performed with R (version 4.1.0). The R packages "mice" (version 3.7.0), "sna" (version 2.6), "igraph" (version 1.2.6), and "FSA" (version 0.9.1) were used to conduct multiple imputations, generate random permutations, calculate centrality measures, and conduct multiple comparisons, respectively.

RESULTS

Demographic Characteristics

We obtained 5996 HIV *pol* sequences, each representing a unique PWH, of which 165 were sampled from students between 2010 and 2020 and 5831 were sampled from nonstudents between 1997 and 2020. Compared with nonstudents, students showed the following distinct characteristics: age 16–24 years (91.5% vs 11.5%), male (97.6% vs 71.3%), acquired HIV through men who have sex with men (MSM; 80.0% vs 18.7%), CRF01_AE Cluster 4 (13.9% vs 5.6%), CRF01_AE Cluster 5 (7.3% vs 2.0%), CRF07_BC Cluster 1 (54.5% vs 10.8%), and unique recombinant form (10.9% vs 7.3%) (Table 1).

Network Inference

At the network level, 1886 (31.5%) of 5996 PWH were clustered at the GD of 0.5%. The Guangxi network is composed of 531 clusters ranging in size from 2 to 76 nodes (Supplementary Figure 2). Among 5831 nonstudents, 1818 (31.2%) fell within the network across 529 (99.6%) clusters. Compared with nonstudents, students were more likely to fall within the network (n = 68, 41.2% vs 31.2%; $P = .006$), but they were only distributed in 43 (8.1%) clusters. In the univariable and multivariable analyses, male, single, high school or higher educational background, and MSM were the positive predictive factors of clustering (Supplementary Tables 4 and 5). However, age and occupation had different effects on falling within the Guangxi network and falling within student clusters. In age categories, age ≥50 years showed a higher adjusted odds ratio (aOR) than others of falling within the Guangxi network but showed a lower aOR than others of falling within the student cluster. In the occupation categories, all occupations had a similar aOR of falling within the Guangxi network, but farming or factory work showed a lower aOR than others of falling within the student clusters (Table 2). In the

Table 1. Characteristics of Studied Students and Nonstudents

Characteristics	Student, No. (%)	Nonstudent, No. (%)
Total No.	165	5831
Age		
16–24 y	151 (91.5)	669 (11.5)
25–49 y	14 (8.5)[a]	3003 (51.5)
≥50 y	0 (0)	2159 (37.0)
Median (IQR), y	21 (19–22)	43 (31–56)
Sex		
Male	161 (97.6)	4159 (71.3)
Female	4 (2.4)	1672 (28.7)
Marital status		
Single	165 (100)	2089 (35.8)
Married[b]	0 (0)	2758 (47.3)
Divorced or widowed	0 (0)	984 (16.9)
Education background		
Junior high school and below	0 (0)	4233 (72.6)
High school and above	165 (100)[c]	1598 (27.4)
Occupation		
Student	165 (100)	0 (0)
Farming or factory worker	0 (0)	3051 (52.3)
Freelancer, unemployed, or retired	0 (0)	1586 (27.2)
Private company or government employee	0 (0)	1194 (20.5)
Infectious route		
Heterosexual contact	33 (20.0)	4174 (71.6)
Men who have sex with men	132 (80.0)	1092 (18.7)
Intravenous drug use	0 (0)	565 (9.7)
HIV strain		
CRF01_AE Cluster 1	1 (0.6)	1633 (28.0)
CRF01_AE Cluster 2	1 (0.6)	929 (15.9)
CRF01_AE Cluster 3	0 (0)	63 (1.1)
CRF01_AE Cluster 4	23 (13.9)	328 (5.6)
CRF01_AE Cluster 5	12 (7.3)	114 (2.0)
CRF07_BC Cluster 1	90 (54.5)	632 (10.8)
CRF07_BC Cluster 2	5 (3.0)	494 (8.5)
CRF08_BC	2 (1.2)	977 (16.8)
CRF55_01B	13 (7.9)	208 (3.6)
URF	18 (10.9)	424 (7.3)
B and C	0 (0)	29 (0.5)

Abbreviations: IQR, interquartile range; URF, unique recombinant form.

[a]Fourteen students in this category were aged 25–29 years.

[b]Married does not include same-sex marriage.

[c]High school and college students.

observed HIV molecular transmission network, 25 (17.7%) of 141 observed genetic linkages between students and their partners were student–student linkages, and 116 (82.2) of 141 were student–nonstudent linkages, which indicated that students and nonstudents might have strong linkages in the HIV transmission network. Further, we conducted permutation tests to assess the robust genetic linkages and find the key subpopulation in nonstudents. Four groups had significantly higher proportions of observed pairwise genetic linkages to

Table 2. Comparison of the Predictive Factors Associated With Falling Within the Guangxi Molecular Transmission Network and Falling Within the Student Clusters

Characteristics	Falling Within Guangxi Network		Falling Within Student Clusters	
	aOR (95% CI)	P	aOR (95% CI)	P
Age				
25–49 y	1		1	
16–24 y	1.08 (0.89–1.31)	.43	1.47 (1.08–1.98)	.01
≥50 y	1.74 (1.51–2.01)	<.001	0.30 (0.13–0.61)	.002
Sex				
Female	1		1	
Male	1.24 (1.07–1.43)	.003	2.03 (1.13–3.88)	.02
Marital status				
Married[a]	1		1	
Single	1.18 (1.01–1.39)	.04	2.19 (1.41–3.50)	.001
Divorced or widowed	0.92 (0.78–1.08)	.31	0.88 (0.39–1.82)	.75
Education background				
Junior high school and below	1		1	
High school and above	1.25 (1.08–1.46)	.004	2.86 (1.97–4.22)	<.001
Occupation				
Farming or factory worker	1		1	
Freelancer, unemployed, or retired	1.13 (0.98–1.31)	.09	2.78 (1.80–4.41)	<.001
Private company or government employee	1.07 (0.90–1.26)	.44	2.47 (1.59–3.93)	<.001
Student	0.89 (0.61–1.28)	.52	5.44 (3.16–9.55)	<.001
Infectious route				
Heterosexual contact	1		1	
Men who have sex with men	1.83 (1.54–2.19)	<.001	2.14 (1.53–3.02)	<.001
Intravenous drug use	0.51 (0.39–0.65)	<.001	0.03 (0.00–0.16)	.001

Student clusters = clusters containing at least 1 student.

Abbreviation: aOR, adjusted odds ratio.

[a]Married does not include same-sex marriage.

students than random links: (1) other students (aOR, 7.2; $P < .001$); (2) private company employees aged 16–24 years (aOR, 3.3; $P = .01$); (3) private company or government employees aged 25–49 years (aOR, 1.7; $P = .03$); and (4) freelancers or unemployed individuals aged 16–24 years (aOR, 5.0; $P < .001$), respectively (Figure 1). These results showed that the four groups mentioned above were the key subpopulation that had robust genetic linkages with students.

At the cluster level, 43 student clusters were composed of 68 students, 132 DLS, and 109 ILS. The majority of these individuals were male (students, 98.5%; DLS, 92.4%; ILS, 97.2%). More than 80% of nonstudents (including DLS and ILS) had a high school or higher educational background, and <15% of nonstudents were farming or factory workers (Supplementary Table 6). Among the 3 subgroups, students were the youngest (median [interquartile range {IQR}], 20 [19–22] years; adjusted $P_{\text{student vs DLS}} < .001$; adjusted $P_{\text{student vs ILS}} < .001$), followed by DLS (median [IQR], 25 [22–30] years; adjusted $P_{\text{DLS vs ILS}} < .001$) and ILS (median [IQR], 28 [24–35] years). Additionally, the student subgroup had the highest proportion of PWH age 16–24 years (91.2%), followed by the DLS subgroup (46.2%) and the ILS subgroup (29.4%).

The target diagrams (Figure 2) illustrate how the Degree indeed captures the notions of centrality: the nearer to the center,

the higher the Degree. However, the small number of individuals in a single cluster made it difficult to detect statistical differences between the 3 subgroups. For example, the largest student cluster was composed of 6 students, 21 DLS, and 49 ILS. For the Degree in this cluster, there was no significant difference between students (median [IQR], 4.0 [1.3–6.8]) and DLS (median [IQR], 11.0 [4.0–12.0]; adjusted $P_{\text{student vs DLS}} = .12$). Similarly, there was no significant difference between students (median [IQR], 4.0 [1.3–6.8]) and ILS (median [IQR], 2 [1–3]; adjusted $P_{\text{student vs ILS}} = .20$). But the Degree for DLS (median [IQR], 11.0 [4.0–12.0]) was significantly higher than that for ILS (median, 2 [1–3]; adjusted $P_{\text{DLS vs ILS}} < .001$) in the cluster. Therefore, we compared the centrality measures for 309 individuals in all 43 student clusters.

The Degree for students (median [IQR], 2.0 [1.0–5.0]; adjusted $P_{\text{student vs DLS}} < .001$) and ILS (median [IQR], 2.0 [1.0–4.0]; adjusted $P_{\text{DLS vs ILS}} < .001$) was significantly lower than that for DLS (median [IQR], 4.0 [2.0–8.0], respectively). Similarly, the Betweenness Centrality for students (median [IQR], 0 [0–2.0]; adjusted $P_{\text{student vs DLS}} = .01$) and ILS (median [IQR], 0 [0–3.3]; adjusted $P_{\text{DLS vs ILS}} = .02$) was significantly lower than that for DLS (median [IQR], 0.6 [0–7.1]). However, the Eigenvector Centrality for students (median
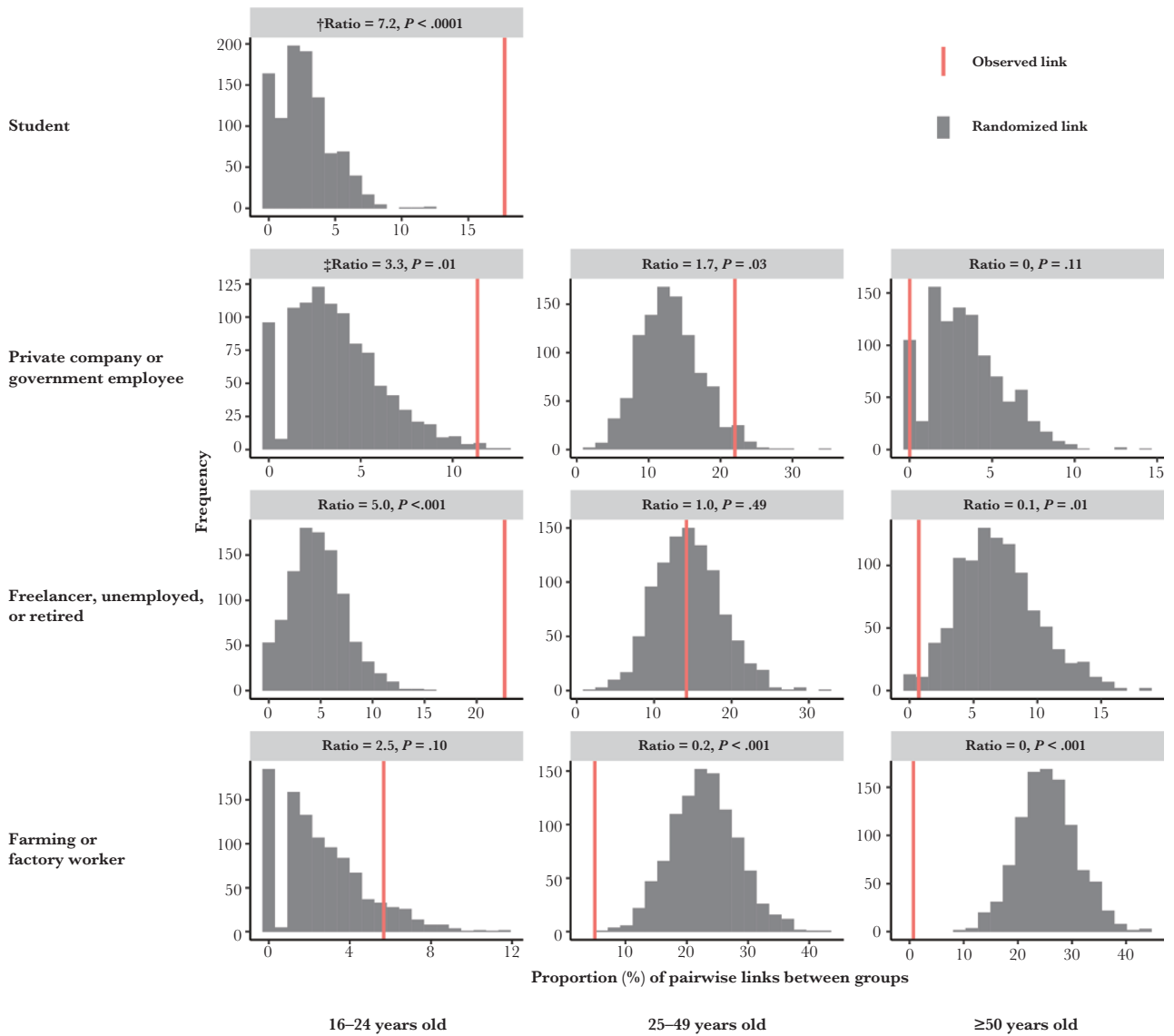
Figure 1. Comparison of the observed proportion of pairwise genetic linkages and the distribution for the proportion of randomized pairwise genetic linkages. [a]165 students aged 16–29 years. [b]226 private company employees aged 16–24 years. Ratio equals the proportion of the observed proportion of pairwise genetic linkages divided by the median of the proportion of randomized pairwise genetic linkages. Ratios >1 and $P < .05$ indicated a significantly stronger link in the observed network than random links.
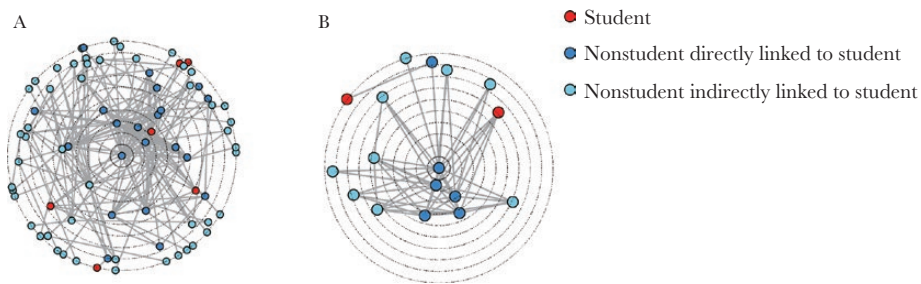

Figure 2. Target diagrams showing individuals' positions in the 2 largest student clusters. A, The largest cluster is composed of 6 students, 21 nonstudents directly linked to students (DLS), and 49 nonstudents indirectly linked to students (ILS). B, The second-largest cluster is composed of 2 students, 6 DLS, and 9 ILS. The target diagrams illustrate that the closer a node is to the center, the higher its degree. Because the small number of individuals in a single cluster made it difficult to detect statistical differences, we compared centrality measures for 309 individuals in all 43 student clusters.

Table 3. Centrality Measures for Students and Nonstudents in the Student Clusters

Centrality	Student	DLS	ILS	Adjusted $P_{student\ vs\ DLS}$	Adjusted $P_{student\ vs\ ILS}$	Adjusted $P_{DLS\ vs\ ILS}$
Degree	2.0 (1.0–5.0)	4.0 (2.0–8.0)	2.0 (1.0–4.0)	<.001	.66	<.001
Betweenness Centrality	0 (0–2.0)	0.6 (0–7.1)	0 (0–3.3)	.01	.81	.02
Eigenvector Centrality	0.81 (0.62–1.00)	0.83 (0.61–1.00)	0.18 (0.01–0.64)	.60	<.001	<.001

Data are presented as median (IQR). Statistical significance was calculated using Dunn's Kruskal-Wallis test for multiple comparisons.

Abbreviations: DLS, nonstudents directly linked to students; ILS, nonstudents indirectly linked to students; IQR, interquartile range.

[IQR], 0.81 [0.62–1.00]; adjusted $P_{student\ vs\ ILS}$ < .001) and DLS (median [IQR], 0.83 [0.61–1.00]; adjusted $P_{DLS\ vs\ ILS}$ < .001) was significantly higher than that for ILS (median [IQR], 0.18 [0.01–0.64]) (Table 3).

DISCUSSION

This is the first study using large-scale province-wide surveillance databases to infer HIV transmission networks in students and characterize their network position. Although the small number of students was an inherent limitation, the largest number of background sequence data and the most detailed demographic data in Guangxi provided the greatest possibility to identify potential sources of HIV acquisition in students.

As the HIV transmission network is constructed based on genetic similarity, the genetic linkages may not fully represent actual transmission linkages. Although potential transmission clusters and risk populations may be found under a lenient GD of 1.5% [20], it is difficult to detect the high-priority clusters and identify the high-risk subgroups in a larger cluster [21]. In this study, 18.7% of nonstudent PWH and 80.0% of student PWH acquired HIV from MSM based on self-report. Therefore, the transmission cluster inferred by a lenient GD may include many PWH with a weak epidemiological linkage. Additionally, China is one of the countries with the largest number of HIV-1 subtypes in the world [22], and a lenient GD may generate false-positive genetic linkages between some HIV recombinant forms. Therefore, we chose a conservative genetic threshold of 0.5 to infer the rapid transmission network [14].

In the rapid transmission network, a larger number of students gathered in a small number of clusters, which were composed of both students and nonstudents. This phenomenon suggested that the primary sources of rapid transmission may be concentrated in certain subpopulations. In this study, some demographic and risk characteristics were demonstrated to be robust predictive factors of falling within the Guangxi network and student clusters, such as male gender, single marital status, high school or higher educational background, and being an MSM. However, age 25–49 years performed a weak predictive effect on falling within the Guangxi network but a strong effect on falling within student clusters. Also, in the occupation categories, freelancer, unemployed or retired, private company or government employee, and student showed similar predictive effects. As the high proportion of individuals aged ≥50 years

and farming or factory workers in the Guangxi network may lead to more false-positive genetic linkages, we further conducted permutation tests to assess the robust genetic linkages. The consistent results of multivariable logistic regression analyses and permutation tests indicated that young MSM living in urban areas had robust genetic linkages with students in the complex network. In previous interviews, the median age at first anal intercourse among MSM has continuously decreased from 33 years to 18 years over the past few decades [23], but <42% of MSM aged <25 years have undergone first-time HIV testing in China [24]. We have demonstrated that insufficient HIV testing in young MSM accelerates HIV transmission between students and nonstudents. However, we need to emphasize that the generalizability of our results may be limited. A previous nationwide spatial analysis suggested that sociodemographic indicators such as population density and per capita gross domestic product had statistically significant positive correlations with spatial–temporal clusters of MSM [25]. Therefore, the risk of young MSM may be underestimated when our results are externally validated in the developed region.

Not only students but also their sexual partners should be considered high-priority groups for interventions. In student clusters, DLS occupied a central position because they had the highest median Degree among the 3 subgroups. Additionally, DLS had the highest median Betweenness Centrality among the 3 subgroups, which indicated that DLS provided a bridge between students and ILS. Compared with DLS, students were currently located at a relatively peripheral position because they had a lower median Degree than DLS. However, students had a high potentiality to become the next superspreaders because they had a high median Eigenvector Centrality, similar to DLS, whereas ILS were located at peripheral positions and had a low risk of transmission. Notably, ILS were the oldest in the student cluster, followed by DLS and students. Therefore, we propose a hypothesis that there is a cycle of HIV transmission from men with a median age of 28 years to men with a median age of 25 years to students (Figure 3). This hypothesis is reasonable because that age-disparate sexual relationship is associated with a high frequency of unprotected sex behaviors [26, 27]. Although genetic linkages in the molecular transmission network do not suggest the direction or order of transmission, the network positions of individuals can be used to infer the transmission patterns and provide the necessary evidence to prioritize targeted interventions [20]. The most effective methods at present
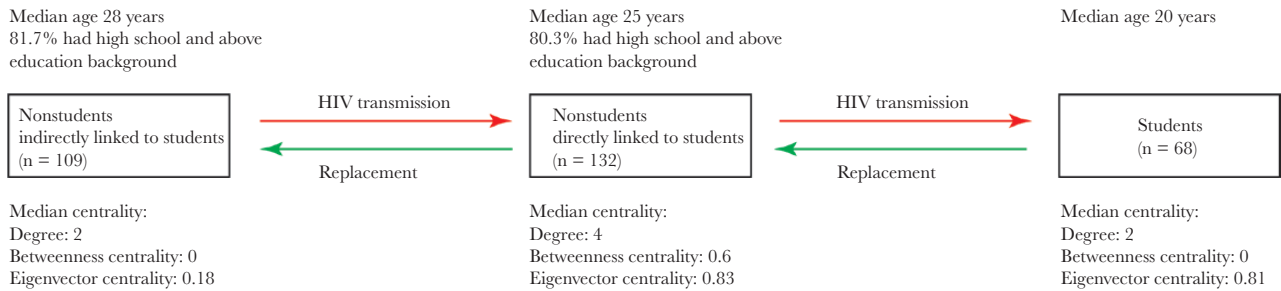
Figure 3. Schematic presentation of putative HIV transmission cycle in students and nonstudents.

to block this cycle are conducting detailed behavioral surveys on DLS and providing them with targeted HIV health care services. In parallel, providing services for students will reduce future HIV prevalence. Additionally, a small number of females were found in the student clusters. Why do these females appear in the clusters dominated by MSM? A possible explanation is that they were the unmarried female sexual partners or wives of MSM. A meta-analysis showed that 17.0% of MSM in China were currently married to a female. Also, 26.3% of MSM had had at least 1 heterosexual contact in the last 6 months, but the rate of condom use was only 25.6% [28]. Our study suggested that MSM could be the critical bridge of HIV transmission to students as well as the general female population.

The study has some limitations. First, a sampling bias was inevitable because some students might conceal their student identity. Second, some nonstudent young men who were diagnosed as HIV-positive might have acquired HIV when they were students. A BED-capture enzyme immunoassay for identifying recent infections and a detailed questionnaire for understanding when students graduated were not available in this study. Therefore, these limitations open avenues for further research. The inclusion of such data in future studies can improve the accuracy of targeted interventions to reduce secondary transmission.

CONCLUSIONS

In conclusion, this study revealed that students tended to cluster not only with other students but also with nonstudent young men. Both students and nonstudents occupied an important position in the HIV transmission network. These data emphasize the urgent need for 2-pronged public health interventions based on both school and society.

References

1. Li G, Jiang Y, Zhang L. HIV upsurge in China's students. Science **2019**; 364:711.
2. Zou H, Tucker JD, Fan S, et al. Learning about HIV the hard way: HIV among Chinese MSM attending university. Lancet Infect Dis **2018**; 18:16–8.
3. Ge L, Li DM, Tang HL, et al. Tends on HIV and syphilis prevalence and sexual behaviors among young students in China, 2015-2019. Zhonghua Liu Xing Bing Xue Za Zhi **2021**; 42:602–7.
4. Cai C, Tang HL, Chen FF, et al. Characteristics and trends of newly reported HIV infection in young students in China, 2010-2019. Zhonghua Liu Xing Bing Xue Za Zhi **2020**; 41:1455–9.
5. Bbosa N, Ssemwanga D, Ssekagiri A, et al. Phylogenetic and demographic characterization of directed HIV-1 transmission using deep sequences from high-risk and general population cohorts/groups in Uganda. Viruses **2020**; 12:331.
6. Kann L, McManus T, Harris WA, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ **2018**; 67:1–114.
7. Kusuma M, Kidd T, Muturi N, et al. HIV knowledge and stigma among dietetic students in Indonesia: implications for the nutrition education system. BMC Infect Dis **2020**; 20:663.
8. Bunting SR, Garber SS, Goldstein RH, et al. Student education about pre-exposure prophylaxis (PREP) varies between regions of the United States. J Gen Intern Med **2020**; 35:2873–81.
9. Koris AL, Stewart KA, Ritchwood TD, et al. Youth-friendly HIV self-testing: acceptability of campus-based oral HIV self-testing among young adult students in Zimbabwe. PLoS One **2021**; 16:e0253745.
10. Han J, Li J, Wang KR, et al. Status and influencing factors of knowledge awareness and service acceptance of HIV non-occupational post-exposure prophylaxis of men who have sex with men among college students among three cities of China, 2019. Zhonghua Yu Fang Yi Xue Za Zhi **2020**; 54:1220–6.
11. Han X, Zhao B, An M, et al. Molecular network-based intervention brings us closer to ending the HIV pandemic. Front Med **2020**; 14:136–48.
12. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform **2019**; 20:1160–6.
13. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol **2016**; 65:997–1008.
14. Wertheim JO, Panneer N, France AM, et al. Incident infection in high-priority HIV molecular transmission clusters in the United States. AIDS **2020**; 34:1187–93.
15. Campbell EM, Patala A, Shankar A, et al. Phylodynamic analysis complements partner services by identifying acute and unreported HIV transmission. Viruses **2020**; 12:145.
16. Qin F, Liang B, Liang H, et al. High HIV prevalence, low awareness of and willingness to use HIV pre-exposure prophylaxis among male migrant workers in Southwestern, China: a short report. AIDS Care **2021**; 33:760–6.
17. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol **2020**; 37:1322–31.
18. Ragonnet-Cronin M, Hu YW, Morris SR, et al. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. Lancet HIV **2019**; 6:e164–72.

19. Farine DR, Whitehead H. Constructing, conducting and interpreting animal social network analysis. J Anim Ecol **2015**; 84:1144–63.

20. Rhee SY, Magalis BR, Hurley L, et al. National and international dimensions of human immunodeficiency virus-1 sequence clusters in a Northern California clinical cohort. Open Forum Infect Dis **2019**; 6:XXX–XX.

21. Oster AM, France AM, Panneer N, et al. Identifying clusters of recent and rapid HIV transmission through analysis of molecular surveillance data. J Acquir Immune Defic Syndr **2018**; 79:543–50.

22. Hao J, Zheng S, Gan M, et al. Changing proportions of HIV-1 subtypes and transmitted drug resistance among newly diagnosed HIV/AIDS individuals — China, 2015 and 2018. China CDC Wkly **2021**; 3:1133–8.

23. Zou H, Xu J, Hu Q, et al. Decreasing age at first anal intercourse among men who have sex with men in China: a multicentre cross-sectional survey. J Int AIDS Soc **2016**; 19:20792.

24. Zhang J, Xu JJ, Song W, et al. HIV incidence and care linkage among MSM first-time-testers in Shenyang, China 2012-2014. AIDS Behav **2018**; 22:711–21.

25. Qin Q, Guo W, Tang W, et al. Spatial analysis of the human immunodeficiency virus epidemic among men who have sex with men in China, 2006-2015. Clin Infect Dis **2017**; 64:956–63.

26. Hurt CB, Matthews DD, Calabria MS, et al. Sex with older partners is associated with primary HIV infection among men who have sex with men in North Carolina. J Acquir Immune Defic Syndr **2010**; 54:185–90.

27. Bajunirwe F, Semakula D, Izudi J. Risk of HIV infection among adolescent girls and young women in age-disparate relationships in Sub-Saharan Africa. AIDS **2020**; 34:1539–48.

28. Chow EP, Wilson DP, Zhang L. What is the potential for bisexual men in China to act as a bridge of HIV transmission to the female population? behavioural evidence from a systematic review and meta-analysis. BMC Infect Dis **2011**; 11:242.