

Preplanned Studies

Identifying the Key Nodes of HIV Molecular Transmission Network Among Men Who Have Sex with Men — Guangzhou, Guangdong Province, China, 2015–2017

Juan Yang¹; Zhigang Han²; Huifang Xu²; Hui Xing¹; Peng Xu¹; Weibin Cheng²; Yuzhou Gu²; Fan Lyu^{1,†}

Summary

What is already known about this topic?

Identifying the most influential spreaders in human immunodeficiency virus (HIV) transmission networks is crucial for developing effective prevention strategies.

What is added by this report?

This study identified key nodes of the HIV molecular transmission network among men who have sex with men (MSM) by utilizing linkages between sequences to reconstruct the transmission network at the molecular level.

What are the implications for public health practice?

This study could act as an important supplement of laboratory results to epidemiological studies and suggests that interdisciplinary research could inspire new ideas for finding breakthroughs on HIV/acquired immunodeficiency syndrome (AIDS) prevention and control.

Based on reports of the acquired immunodeficiency syndrome (AIDS) epidemic in China in December 2017, sexual transmission accounted for more than 90% of total infections, and 26.86% of the sexual transmission infections were men who have sex with men (MSM) (1). According to the research conducted by Ethan Morgan and his colleagues, it is necessary to conduct investigations that focus on networks of target populations rather than traditional epidemiological factors such as geographic areas of high incidence (2). Identifying the most influential spreaders of the human immunodeficiency virus (HIV) transmission networks is crucial to develop effective prevention strategies.

Analyzing the structure of networks provides an optimal way to confirm the location and the role of key nodes that play key roles in accelerating HIV transmission in the network. Given the hidden nature of the MSM population, it is difficult to confirm the relationship ties between any two members of the

community, which is the first step to analyze network structure in traditional epidemiologic field investigations. Phylogenetics provides probabilities for network structure analysis in HIV research. Inferring putative transmission is the process of utilizing molecular phylogenetics analyzed by using HIV sequences to identify transmission events in groups of individuals (3). This study identified key nodes of the HIV molecular transmission network among MSM by utilizing linkages between sequences to reconstruct the HIV transmission network at the level of molecular genetics.

A total of 184 sequences of the HIV-1 pol full-length gene were assessed and stratified over 2 periods based on the year of sample collection (2015–2017). All 184 sequences were aligned with all known sequences in the HIV database (<http://hiv-web.lanl.gov/content/index>, operated by Triad National Security, LLC for the U.S. Department of Energy's National Nuclear Security Administration) using the Basic Local Alignment Search Tool (BLAST) before analysis. The length of the HIV-1 pol gene was 3,045 base pairs (bp) and the nucleotide positions of pol were 2,147–5,192 according to HXB2 subtype B reference strain (GenBank accession number K03455). The sequences were edited with the software Sequencher (version 5.0, Gene Codes Corporation, Ann Arbor, MI, USA). The reference sequences that were available on HIV Database) covered the major HIV-1 subtypes/CRFs. Among the 184 successfully amplified pol full-length sequences, 44.02% (81/184) were CRF07_BC, 33.15% (61/184) were CRF01_AE, 13.04% (24/184) were 01_B, 3.80% (7/184) were B, and 5.98% (11/184) were others.

HIV molecular transmission network was based on genetic distance (4). Putative transmission links in the network were identified with dichotomized data, which was determined by whether the pairwise genetic distance was less than 0.015 substitutions per site within all sequences (5). In our study, the Tamura-Nei 93 pairwise genetic distances were calculated by Mega

[Mega 7.0: Molecular Evolutionary Genetics Analysis across computing platforms (Kumar S, Stecher G, and Tamura K 2016)] (6).

All social network analyses was conducted by UCINET 6.0 (version 6.05; Borgatti, Everett, and Freeman, 2002). The methods were described in the Supplementary Materials (available in <http://weekly.chinacdc.cn/>). All statistical analyses were performed with SAS (version 9.4, SAS Institute Inc., Cary, NC, USA). Multivariate logistic regression model was used to analyze the demographic characteristics of the key nodes.

Of the 184 HIV-1 sequences that were of patients diagnosed between 2015–2017, 75 sequences had at least one relationship tie with another patient (Figure 1). The characteristics of the participants are presented in Table 1. Social network analysis demonstrated that there were 14 cliques that included at least 3 nodes (Supplementary Materials, Supplementary Table S1, available in <http://weekly.chinacdc.cn/>). The biggest clique includes 24 members, and there were some cliques sharing the same members. Cliques 1–8 shared a lot of same members, and clique 9 only included 4 members that did not share any member with others.

The clique co-membership method yields a large

subgroup consisting of cliques 1–8 with a median subgroup of cliques 10 and 12, 4 smaller groups including cliques 9, 11, 13, and 14, and the outsiders. We denoted the 6 subgroups as A, B, C, D, E, and F. M026 acted as a broker between Subgroup B and F, E and F, D and F, as well as D and E. Subgroup B and D shared 2 actors {30, M026} acting as brokers between them. There were 3 shared members between groups B and E, respectively: {R12, M026, M056}.

From the result of lambda analysis (Supplementary Table S2, available in <http://weekly.chinacdc.cn/>), there were 17 lambda sets with $\lambda = 1$ that have a minimum of 1 independent path linking for any two actors. The largest λ was 19; it include 2 actors {27, M057}. A little bit smaller λ were 15 and 10, the actors in the lambda sets were {4, 27, M057} and {26, M050, 4, 27, M057} respectively. All of the above 5 nodes were nested hierarchically in the set with $\lambda = 1$, which has the largest number of members. These five nodes have the most relationship ties in the set and were in the most active central position.

Finally, we identified 9 key nodes by using cohesive subgroup analysis in the HIV molecular transmission network; {30, M026, R12, M056} acted as brokers between subgroups, and {26, M050, 4, 27, M057} were confirmed as the most active nodes in one

TABLE 1. Characteristics of the study population according to categories of number of connections in Guangzhou, Guangdong Province, China, 2015–2017.

Characteristics	Number of respondents N (%)	Number of connections			P value*
		0	1	≥2	
Total	184 (100.00)	109 (59.24)	36 (19.57)	39 (21.20)	
Age (years)					0.21
18–25	69 (37.50)	47 (68.12)	10 (14.49)	12 (17.39)	
26–35	71 (38.59)	37 (52.11)	19 (26.76)	15 (21.13)	
≥36	44 (23.91)	25 (56.82)	7 (15.91)	12 (27.27)	
Educational level					0.53
Primary school	44 (23.91)	27 (61.36)	6 (13.64)	11 (25.00)	
Junior and senior high school	44 (23.91)	29 (65.91)	7 (15.91)	8 (18.18)	
College and above	96 (52.17)	53 (55.21)	23 (23.96)	20 (20.83)	
Marital status					0.26
Married	29 (15.76)	13 (44.83)	7 (24.14)	9 (31.03)	
Unmarried	143 (77.72)	89 (62.24)	28 (19.58)	26 (18.18)	
Divorced	12 (6.52)	7 (58.33)	1 (8.33)	4 (33.33)	
Time of diagnosis					0.01
2015–2016	100	53 (53.00)	28 (28.00)	19 (19.00)	
2016–2017	84	56 (66.67)	8 (9.52)	20 (23.81)	

Note: All percentages are line percentages.

* P value for chi-square test for categorical variables.

subgroup. We analyzed the demographic characteristics of these key nodes. From the results of multivariate logistic regression model, young MSM born in the 1990s (aged 18–25) and 1980s (aged 26–35) was 0.06 and 0.12 times, respectively, likely to be a key node than older MSM born in the 1970s (aged 36 and older) or before (Table 2).

DISCUSSION

Of 184 newly-HIV diagnosed MSM, 40.76% were linked to other MSM. Social network analysis demonstrated that 9 key nodes were detected.

By using the clique co-membership method, there were four key nodes acting as brokers between subgroups. It could be inferred that there were a lot of subgroups connected by sharing co-members in the HIV molecular transmission network among MSM in Guangzhou City, Guangdong Province, China.

The four nodes that occupied important bridge locations were critical in controlling and understanding the spread processes as well as for developing effective prevention strategies.

Selecting candidates who connect across groups of otherwise disconnected individuals (such individuals are known as “bridging actors”) based on their network positions was shown to be more likely to enhance the

diffusion of innovative HIV prevention interventions when compared to other centrally located popular opinion leaders (7). Some HIV-infected MSM called as key nodes mediated the transmission of HIV among different subpopulations. Young MSM were less likely to promote HIV transmission than older MSM.

Based on connectivity cohesive subgroup analysis, known as the lambda sets method, we detected 5 key nodes. They were possibly taking on some kind of leadership role. In fact, they were active only in several subgroups of the transmission network in this study, rather than participating in the whole network of HIV transmission. In our study, there were at least three independent subgroups with members closely connected to each other within them. Therefore, it is immensely vital for HIV prevention and control to determine subgroups with different characteristics in HIV transmission network among MSM.

In recent years, HIV incidence in young Chinese MSM was significantly higher than that of older MSM (8). However, based on our results, MSM who were younger than 25 years old were less likely to promote the wide spread of HIV than older MSM. The results of our survey on the social interaction patterns of this group also confirmed this point: MSM aged about 30 and above were more likely to have condomless anal intercourse (CAI) with those of different ages (9). The

TABLE 2. The demographic characteristics of key nodes in the HIV transmission network in Guangzhou, Guangdong Province, China, 2015–2017.

Characteristics	Key nodes N (%)	Others N (%)	Adjusted OR 95% CI	P value
Total	9 (4.89)	175 (95.11)		
Age (years)				
18–25	1 (1.45)	68 (98.55)	0.06 (0.01–0.74)	0.03
26–35	2 (2.82)	69 (97.18)	0.12 (0.02–0.84)	0.03
≥36	6 (13.64)	38 (86.36)	1.00	
Educational level				
Primary school	3 (6.82)	41 (93.19)	0.42 (0.06–2.95)	0.38
Junior and senior high school	1 (2.27)	43 (97.73)	0.28 (0.03–3.07)	0.29
College and above	5 (5.21)	91 (94.79)	1.00	
Marital status				
Married	3 (10.34)	26 (89.66)	1.35 (0.11–16.12)	0.81
Unmarried	5 (3.50)	138 (96.50)	1.40 (0.09–22.93)	0.81
Divorced	1 (8.33)	11 (91.67)	1.00	
Sample resource				
NHS	4 (4.00)	96 (96.00)	0.61 (0.14–2.75)	0.52
VCT	5 (5.95)	79 (94.05)	1.00	

Abbreviations: HIV=human immunodeficiency virus; OR=odds ratio; CI=confidence interval; NHS= the National HIV sentinel Surveillance; VCT=HIV voluntary counseling and testing clinics.

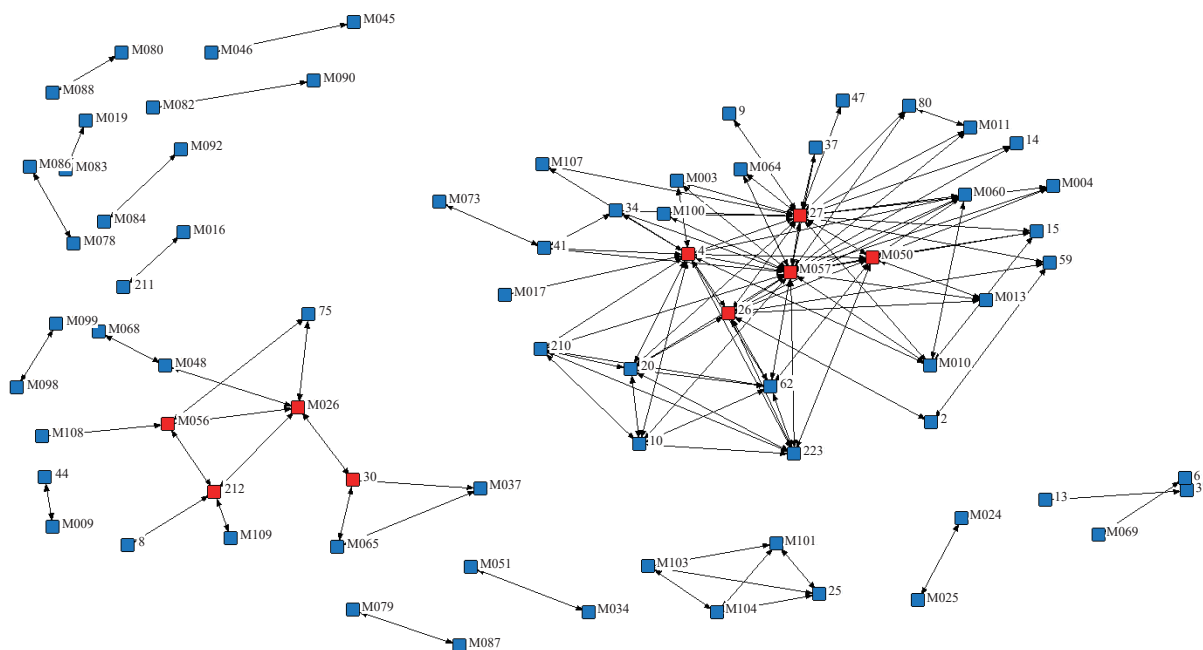


FIGURE 1. Network diagram of 75 nodes who had at least 1 relationship tie with another node among 184 sequences of men who have sex with men in Guangzhou, Guangdong Province, China, 2015–2017.

Note: Genetic distance: the pairwise genetic distance is equal or less than 0.015 substitutions per site within all sequences. Red represents key nodes: The name of the nodes is the laboratory code, and the sample name beginning with “M” came from 2015–2016. The line between any two nodes displayed the propagation relationship; however, the lines do not denote directionality. Cliques and lambda sets were obtained by analysis and cannot be seen directly from the picture. See Supplementary Table S1 and Table S2.

point of intervention activities should be to improve awareness of self-protective measures in young MSM and to promote HIV testing and antiretroviral therapy in older MSM.

This study was subject to some limitations. Without a universally accepted standard, we used genetic distance less than 0.015 as the criterion when inferring putative transmission ties of the sequences. Some sequences with propagative relationships may be misclassified as false negatives. Furthermore, the network used to analyze structure characteristics in this paper was a partial network, so the number and scale of the subgroups may be underestimated, and some key nodes were not successfully identified. Large sample size research is needed to explore the demographic and behavioral characteristics of key nodes. Moreover, sequences were obtained from newly-HIV-diagnosed MSM during 2015–2017. We did not include the cases of patients who were infected through heterosexual and drug injection, and our conclusions did not apply to other populations.

There were a lot of subgroups connected by sharing co-members in HIV molecular transmission network among MSM in Guangzhou. Some HIV-infected

MSM, known as key nodes, mediated the transmission of HIV among different subpopulations. Young MSM under 25 were less likely to promote HIV transmission than older MSM. This study reflected the important supplement of laboratory results to epidemiological studies and provided new ideas for finding breakthroughs in HIV prevention and control.

Conflicts of interest: No conflicts of interest declared.

Acknowledgements: Yali Zhang and Qiuyan Yu.

Funding: Supported by the National Natural Science Foundation of China (NSFC) (No. 71473234 and 71573239).

doi: 10.46234/ccdcw2021.198

Corresponding author: Fan Lyu, fanlv0925@163.com.

¹ The National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China;

² Department of HIV/AIDS Control and Prevention, Center for Disease Control and Prevention, Guangzhou, Guangdong, China.

Submitted: May 10, 2021; Accepted: September 15, 2021

REFERENCES

1. NCAIDS, NCSTD, China CDC. Update on the AIDS/STD epidemic

- in China in December 2017. *Chin J AIDS STD* 2018;24(2):111. <http://dx.doi.org/10.13419/j.cnki.aids.2018.02.01>. (In Chinese).
2. Morgan E, Oster AM, Townsell S, Peace D, Benbow N, Schneider JA. HIV-1 infection and transmission networks of younger people in Chicago, Illinois, 2005–2011. *Public Health Rep* 2017;132(1):48 – 55. <http://dx.doi.org/10.1177/0033354916679988>.
 3. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. *Curr Opin HIV AIDS* 2014;9(2):126 – 33. <http://dx.doi.org/10.1097/COH.0000000000000040>.
 4. Oster AM, Wertheim JO, Hernandez AL, Ocfemia MCB, Saduvala N, Hall HI. Using molecular HIV surveillance data to understand transmission between subpopulations in the United states. *J Acquir Immune Defic Syndr* 2015;70(4):444 – 51. <http://dx.doi.org/10.1097/QAI.0000000000000809>.
 5. Pines HA, Wertheim JO, Liu L, Garfein RS, Little SJ, Karris MY. Concurrency and HIV transmission network characteristics among MSM with recent HIV infection. *AIDS* 2016;30(18):2875 – 83. <http://dx.doi.org/10.1097/QAD.0000000000001256>.
 6. Li XS, Gao R, Zhu KX, Wei FR, Fang K, Li W, et al. Genetic transmission networks reveal the transmission patterns of HIV-1 CRF01_AE in China. *Sex Transm Infect* 2018;94(2):111 – 6. <http://dx.doi.org/10.1136/sextrans-2016-053085>.
 7. Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. *Soc Sci Med* 2015; 125:192 – 202. <http://dx.doi.org/10.1016/j.socscimed.2013.12.034>.
 8. Mao X, Wang Z, Hu Q, Huang C, Yan H, Wang Z, et al. HIV incidence is rapidly increasing with age among young men who have sex with men in China: a multicentre cross-sectional survey. *HIV Med* 2018;19(8):513 – 22. <http://dx.doi.org/10.1111/hiv.12623>.
 9. Yang J, Xu HF, Li S, Cheng WB, Gu YZ, Xu P, et al. The characteristics of mixing patterns of sexual dyads and factors correlated with condomless anal intercourse among men who have sex with men in Guangzhou, China. *BMC Public Health* 2019;19(1):722. <http://dx.doi.org/10.1186/s12889-019-7082-9>.

Supplementary Materials

Cohesive subgroups analysis is a powerful and mathematically rigorous method to characterize network robustness. The strength lies in the capacity to detect strong connections among nodes that not only have no neighbors in common, but that may be distantly separated in the graph (1).

Cliques

A clique is a subgroup of actors in which each actor is adjacent to any other actors in it, and it is impossible to add any other actors to the clique without violation of this condition (2). In our study, we constrain the minimum size of any clique to three.

When there are many cliques, it is difficult to interpret the result of cohesive subgroups for the overlap between cliques, which can result in hidden features of the structure. A method to solve this issue would be to try to remove or reduce the overlap by performing additional analysis such as clique co-membership(2). The first step is to combine cliques who have more than 2/3 of all actors being shared. After the first step, if there are still too many cliques, those that share more than 1/3 of the same members can be merged (3). From a small number of cliques, we can detect a set of key nodes acting as the bridge between subgroups.

Lambda Sets

Lambda sets, based on the property that members of the set have greater edge connectivity with other members than with non-members, have been shown to correspond to a particular hierarchical clustering of the nodes in a network (4). It is a maximal subset of actors who have more edge-independent paths connecting them to each other than to outsiders since actors in lambda sets with connectivity λ have a minimum of λ independent paths linking any one to any other. When λ is large, a lambda set describes a subset that is relatively difficult to disconnect by means of edge removals (4). In infectious disease research, we can detect those who are the most active in the subgroup, which is the most important for disease control.

SUPPLEMENTARY TABLE S1. Clique analysis in HIV transmission network with 184 nodes in Guangzhou, Guangdong Province, China, 2015–2017.

Cliques	Number of nodes	ID
1	24	10 14 15 20 26 27 34 37 4 41 62 80 R10 R3 M003 M004 M010 M011 M050 M057 M060 M064 M100 M107
2	22	14 15 20 26 27 34 37 4 59 62 80 R3 M003 M004 M010 M011 M050 M057 M060 M064M100M107
3	22	14 15 20 26 27 34 37 4 47 59 80 9 M003 M004 M010 M011 M050 M057 M060 M064M100M107
4	24	10 14 15 20 26 27 34 37 4 41 62 80 R10 R3 M003 M004 M010 M011 M013 M050 M057 M060 M064 M100
5	22	14 15 20 26 27 34 37 4 59 62 80 R3 M003 M004 M010 M011 M013 M050 M057 M060 M064 M100
6	17	10 20 26 27 34 4 41 62 R10 R3 M003 M010 M017M050 M057 M060 M107
7	12	2 20 26 27 4 59 62 R3 M013 M050 M057 M060
8	5	34 4 41 M057 M073
9	4	25 M101 M103 M104
10	6	30 75 R12 M026 M048 M056
11	4	30 M026 M037 M065
12	5	75 R12 M026 M056 M108
13	5	8 R12 M026 M056 M109
14	3	M026 M048 M068

SUPPLEMENTARY TABLE S2. Lambda sets in HIV transmission network with 184 nodes in Guangzhou, Guangdong Province, China, 2015–2017.

λ	The number of sets	Actors
1	17	1: (3, 13); 2: (44, M009); 3: (R11, M016); 4: (M024, M025); 5: (M045, M046); 6: (M034, M051); 7: (6, M069); 8: (M019, M083); 9: (M078, M086); 10: (M079, M087); 11: (M080, M088); 12: (M028, M090); 13: (M084, M092); 14: (M098, M099); 15: (25, M101, M103, M104); 16: (47, 9, M017, M073, 14, 2, 37, 41, 59, 80, M003, M004, M011, M013, 15, 34, M010, 10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060, M064, M100, M107); 17: (8, M048, 75, R12, M026, M056, 30, M037, M065, M068, M108, M109, M110, M111)
2	4	1: (25, M101, M103, M104); 2: (14, 2, 37, 41, 59, 80, M003, M004, M011, M013, 15, 34, M010, 10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060, M064, M100, M107); 3: (75, R12, M026, M056); 4: (30, M037, M065)
3	3	1: (25, M101, M103, M104); 2: (41, 59, 80, M003, M004, M011, M013, 15, 34, M010, 10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060); 3: (M026, M056)
4	1	1: (15, 34, M010, 10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060)
5	1	1: (M010, 10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060)
6	1	1: (10, R10, 20, 62, R3, 26, M050, 4, 27, M057, M060)
8	1	1: (20, 62, R3, 26, M050, 4, 27, M057)
10	1	1: (26, M050, 4, 27, M057)
15	1	1: (4, 27, M057)
19	1	1: (27, M057)

REFERENCES

1. Sinkovits RS, Moody J, Oztan BT, White DR. Fast determination of structurally cohesive subgroups in large networks. *J Comput Sci* 2016;17(Pt1):62 – 72. <http://dx.doi.org/10.1016/j.jocs.2016.10.005>.
2. Borgatti SP, Everett MG, Johnson JC. *Analyzing social networks*. London: SAGE Publications. 2013. <http://find.nlc.cn/search/showDocDetails?docId=6305356386126200219&dataSource=ucs09&query=Analyzing%20Social%20Networks>.
3. Liu J. *Lecture on Whole Network Approach: A Practical Guide to UCINET (2nd)*. Shanghai. Gezhi press and Shanghai People's Publishing press. 2014. (In Chinese).
4. Borgatti SP, Everett MG, Shirey PR. LS sets, lambda sets and other cohesive subsets. *Soc Netw* 1990;12(4):337 – 57. [http://dx.doi.org/10.1016/0378-8733\(90\)90014-Z](http://dx.doi.org/10.1016/0378-8733(90)90014-Z).