

# An exploratory study on the social and genotypic clustering of HIV infection in men having sex with men

Shui Shan Lee<sup>a,b</sup>, Dennise K.P. Tam<sup>a</sup>, Yi Tan<sup>a</sup>, Wai Lai Mak<sup>c</sup>,  
Ka Hing Wong<sup>c</sup>, Jonathan H.K. Chen<sup>d</sup> and Wing Cheong Yam<sup>d</sup>

**Objective:** To explore the clustering of HIV infected men having sex with men (MSM) using social network approach in conjunction with the phylogenetic relationship of the virus strains.

**Design:** An exploratory study incorporating social network and phylogenetic analysis.

**Methods:** Recently diagnosed HIV-infected MSM attending one major HIV specialist clinic in Hong Kong were recruited in the study involving the administration of a self-administered questionnaire on behaviours and partnership patterns using a Likert Scale, the results of which were assessed using social network analysis and in context of the phylogenetic analysis from sequencing the HIV-1 *pol* gene, as part of the clinical investigation for genotypic resistance. Clusters were defined using social and molecular methods.

**Results:** An 'Internet-centred' cluster and 'Sauna-centred' cluster could be delineated using correspondence analysis and network diagrams. The main distinguishing features of MSM in the 'Internet-centred' social cluster were: younger age, higher education level, and multiple partner types. Three genetic clusters could be identified in the phylogenetic tree, two of which associated with Internet use and one with sauna for sex partnership. There were partial overlaps between social and genetic clusters. Characteristically, the virus strains in sauna users were more disperse compared with the closely knit configuration of those using Internet.

**Conclusion:** The principle of the duality of place and person can be strategically applied in epidemiologic investigation. The characterization of MSM cluster using anonymized network data provides a potentially powerful tool for informing public health intervention.

© 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins

*AIDS* 2009, **23**:000–000

**Keywords:** cluster, men having sex with men, phylogenetic, social network

## Introduction

Infectious diseases are characterized by their propensity to spread from person to person, through transmission routes unique to each pathogen. Individuals become infected largely because of common behavioural practice, which may involve sharing of food items, space (and therefore

airborne infections) or even sexual contacts, as in the case of sexually transmitted infections. Theoretically, all infection cases are linked, forming 'clusters' with shared characteristics. It is against this basic premise that outbreak investigation assumes a central position in the epidemiologic study of infections. Effectiveness of public health control is naturally dependent on the identification of

<sup>a</sup>Stanley Ho Centre for Emerging Infectious Diseases, <sup>b</sup>Department of Microbiology, The Chinese University of Hong Kong, <sup>c</sup>Special Preventive Programme, Centre for Health Protection, Department of Health, Hong Kong Special Administrative Region Government, and <sup>d</sup>Department of Microbiology, The University of Hong Kong.

Correspondence to Dr W.C. Yam, Department of Microbiology, The University of Hong Kong, Queen Mary Hospital Compound, Pokfulam Road, Hong Kong.

Tel: +852 28554821; fax: +852 28551241; e-mail: wcyam@hkucc.hku.hk

Received: 10 January 2009; revised: 25 April 2009; accepted: 5 May 2009.

DOI:10.1097/QAD.0b013e32832dc025

index cases, characterization of clusters, and the investigation of common source infections [1]. In recent years, advances in molecular epidemiology have sharpened the tool for defining infectious disease clusters [2,3], which could contribute to the development of interventions targeting individuals within clusters.

Unlike many pathogens, HIV (human immunodeficiency virus) poses a unique challenge to epidemiologic investigators. As a chronic infection with a long symptom-free period, identification of 'index' cases is almost always impossible. Driven by private behaviours, characterization of outbreaks of HIV is difficult compared to, say, food borne infections. Moreover, stigma and discrimination have been, and still are, causes for concern in targeting HIV patients in prevention programmes [4,5]. With the use of highly active antiretroviral therapy (HAART) in the last decade, HIV is becoming a manageable condition [6,7]. Early diagnosis and prompt treatment with HAART according to clinical protocol is the key to reduction of HIV associated morbidity and mortality. The idea of uncovering HIV clusters has emerged, alongside calls for normalizing HIV prevention [8]. Coincidentally the recent years have witnessed the reports of outbreaks of HIV in men having sex with men (MSM) in countries around the world [9–12]. Such occurrence underlines the pressing need for developing new means of effectively controlling HIV transmission, of which the targeting of people living with the virus is receiving more attention [13–15]. In the absence of any universal model, we undertook to explore the incorporation of partnership profile in the characterization of clusters in a cohort of HIV infected Chinese MSM in Hong Kong. In light of the nonrandom distribution of MSM in the society, principles of social network analysis (SNA) were used [16,17]. The clustering pattern was studied in conjunction with the phylogenetic makeup of the virus obtained from routine genotypic resistance testing (GRT), with a view of examining the applicability of such approach in enhancing epidemiologic surveillance.

## Methods

### Enrolment of patients

HIV positive MSM attending one major HIV specialist clinic (Integrated Treatment Centre) in Hong Kong were invited to join the study if they fulfilled the following inclusion criteria: age 18 years or above; Chinese ethnicity; diagnosis of HIV in the preceding 1 year; infection presumably occurred in Hong Kong. Patients assessed to be emotionally unstable by clinic staff (e.g. depression, with suicidal ideation) was excluded. With consent, each person was asked to complete a self-administered questionnaire following a thorough explanation of the objective of the study, and a description of the meaning of individual question. No personal

particulars were collected, but a code was assigned to each enrolled individual, which was used to link with results of GRT performed on all newly registered cases as part of surveillance of primary resistance. Approval was obtained from the Ethics Committee of the Chinese University of Hong Kong and the Department of Health.

### Questionnaire on partnership

The self-administered questionnaire was modified from that of an earlier study on MSM in Hong Kong [18]. In the first section a participant was asked to provide sociodemographic information including age, education, employment status, and with whom they were living with, followed by information on the setting of infection—speculated year of HIV infection, the place where they tested HIV positive, and the reasons for HIV testing.

The core questions explored respondents' partnership practice in the 1-year period prior to their HIV infection. Partnership practice refers to the type and number of partners (lover, regular partner, and casual partner), and locations that they frequented for seeking sex partnership. The latter was referred as 'partner sourcing' in the study. Five categories of venues for partner sourcing were defined, that is, public toilets, bar, sauna, sex party, gatherings. The use of Internet, though not a physical location, was assessed as a 'virtual' venue for partner sourcing. Respondents were asked if they had used any of the partnership venues for sourcing sex partners. For those indicating a positive response to a specified venue, their use was ranked using a standardized 6-points Likert Scale (ranging from 1 – 'Very Rarely' to 6 – 'Always'), again referring to one's practice in the 1-year period prior to the speculated time of infection. This was followed by questions on the respondent's sexual behaviours, (mutual masturbation, oral sex, anal sex) and condom use (oral sex, anal sex). Similarly, a 6-points Likert Scale was created to capture responses to these questions.

### Correspondence and social network analysis

We made an assumption that MSM who used similar venues for sourcing sex partners shared common partnership characteristics. A matrix was constructed by tabulating the Likert scores for each category of partnership venues (column) against an individual MSM (row). Correspondence analysis was performed using UCINET version 6.174 [19]. The Likert scores for each person were dichotomized (nil versus any Likert score) to allow two-mode networks to be constructed, the latter for illustrating linkage of people with venues. The 2-mode data were converted into one mode by firstly dichotomizing Likert scores (nil and Likert 1–3 versus 4–6) followed by the determination of cross products and minimums, and secondly, separating '0' linkages from other scores. The link relationship was measured in term of density and centrality – density is the ratio of actual to possible linkages, whereas centrality is a measure of the importance of each person as regards his linkage to others.

NetDraw 2.083 and Pajek 1.01 were used in generating the diagrams.

### Phylogenetic analysis

GRT sequences from the enrolled patients were obtained. Briefly, total RNA was extracted from patient plasma using the QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany). The entire PR and upto 400 codons of the HIV-1 *pol* gene was reverse transcribed and amplified by an in-house genotyping system or ViroSeq HIV-1 Genotyping System version 2.0 as described previously [20]. The amplified fragments were then directly sequenced in both directions with BigDye Terminator Cycle Sequencing Ready Reaction kit version 1.1 (Applied Biosystems, California, USA) with five specific primers. Cycle sequencing products from both assays were purified using the AutoScreen96 (GE Healthcare, USA) before loading into the Prism 3130 Genetic Analyzer (Applied Biosystems, California, USA).

Sequences were aligned by the use of CLUSTAL\_X and edited by BioEdit program. Phylogenetic tree was constructed using the Maximum Likelihood approach implemented in PhyML 3.412 [21] with 1000 bootstrap replications. A general-time-reversal (GTR) substitution model with gamma distributed rate heterogeneity of four rate categories ( $\Gamma_4$ ) and a proportion of invariable sites suggested by Modeltest were used. Clustering was considered when there were three or more closely related cases in conjunction with bootstrap value exceeding 900/1000.

### Statistical analysis

The Data were analysed using SPSS version 13.0. Univariate analyses including Chi-square test and independent sample *t*-test was used for categorical and continuous variables respectively. Differences were classified as significant when *p*-value was less than 0.05.

## Results

### Characteristics of respondents

Between July 2007 and March 2008, 84 recently diagnosed HIV MSM were recruited. Data from four

of the enrolled participants were discarded: one denying any form of sexual activities, two did not have sex in Hong Kong, and one was unable to read the questionnaire. Seven respondents with missing data on partner sourcing were also excluded. The mean age of the 73 respondents was 33.5 years (standard deviation = 8.9 years). A majority were working full time, whereas less than 10% were unemployed. About half were living with family, 16% with sex partners, and 22% living alone. Some 40% of the respondents tested HIV positive through attending a voluntary counselling and testing (VCT) service. Over half were tested HIV positive after presenting to a clinic or hospital with some forms of complications. There were, however, no demographic differences between patients who presented with a complication compared with those who did not. All individuals were treatment naïve at the time of blood sampling. Samples from 49 out of 73 cases (67.1%) were amplified, revealing no major resistance mutation in the protease or reverse transcriptase genes.

### Partnership and behavioural pattern

Three main types of partners were reported by the respondents: lovers (70.8%), regular partners (42.5%), casual (82.2%) partners. Lovers were sex partners that respondents considered to be emotionally attached to and in stable relationship; regular partners were those with sexual relationship alone over extended periods; and casual partners were typically one-night stand partners without emotional attachment nor extended relationship. Only two respondents admitted having a 'Money Boy' (sex in exchange for money or kind), which were excluded in the analysis because of the small number involved. The distribution of the practice of masturbation, oral and anal sex is at Table 1.

Table 2 shows the distribution of Likert scores of respondents as regards their venues for sex partner sourcing in the 1-year period prior to their HIV infection. Sex partners here were not differentiated between the three main types above (lover, regular and casual partner). 'Internet' was treated as one form of 'venue', which was further separated in three formats – chatroom, instant messaging programme (IMP) and bulletin board system (BBS). Overall, above 60% of the respondents had used sauna for partner sourcing, followed

**Table 1. Distribution of behavioural pattern and partner-sourcing venues.**

	Types of partners and sexual behaviours					
	Masturbation		Oral sex		Anal sex	
	Reported <sup>a</sup> practice	Frequent <sup>b</sup> practice	Reported <sup>a</sup> practice	Frequent <sup>b</sup> practice	Reported <sup>a</sup> practice	Frequent <sup>b</sup> practice
Lover	47 (92.2%)	37 (72.5%)	47 (92.2%)	37 (72.5%)	45 (88.2%)	28 (54.9%)
Regular	28 (90.3%)	19 (61.3%)	30 (96.8%)	24 (77.4%)	31 (100%)	25 (80.6%)
Casual	58 (95.1%)	35 (57.4%)	59 (96.7%)	38 (62.3%)	59 (96.7%)	33 (54.1%)

<sup>a</sup>Reported practice – admitted having had such practice.

<sup>b</sup>Frequent practice – a score of 4 or above on Likert Scale.

**Table 2. Distribution of Likert scale scores for venues used by respondents for sourcing sex partners in the one year period prior to HIV infection.**

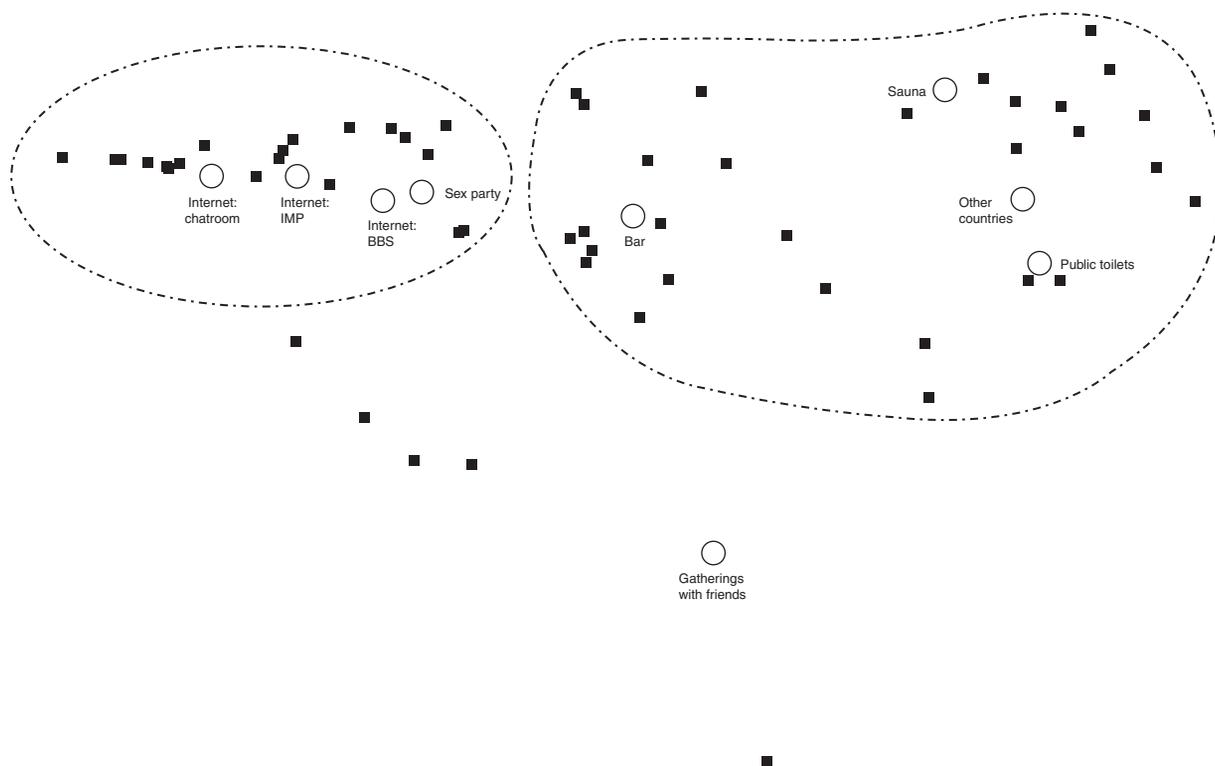
	Public Toilet	Bar	Sauna	Sex Party	Friends Gathering	Internet
Nil response	61 (83.6%)	49 (67.1%)	26 (35.6%)	60 (82.2%)	52 (71.2%)	30 (41.1%)
Likert Scale						
1	5 (6.8%)	7 (9.6%)	6 (8.2%)	7 (9.6%)	3 (4.1%)	6 (8.2%)
2	3 (4.1%)	4 (5.5%)	7 (9.6%)	2 (2.7%)	1 (1.4%)	1 (1.4%)
3	1 (1.4%)	5 (6.8%)	10 (13.7%)	1 (1.4%)	12 (16.4%)	7 (9.6%)
4	2 (2.7%)	3 (4.1%)	13 (17.8%)	2 (2.7%)	3 (4.1%)	11 (15.1%)
5	0 (0.0%)	3 (4.1%)	4 (5.5%)	0 (0.0%)	1 (1.4%)	10 (13.7%)
6	1 (1.4%)	2 (2.7%)	7 (9.6%)	1 (1.4%)	1 (1.4%)	8 (11.0%)

by the use of Internet (chatroom 50.6%; IMP 48.1%; BBS 14.3%). Bars (34.6%) and gatherings with friends (30.8%) were other channels for sourcing sex partners. Fifteen respondents participated in sex parties (19.0%) or visited public toilets (16.5%) for sex partner sourcing.

### Defining clusters

Using Likert scores to denote the strength of linkage between an individual and a venue, correspondence analysis was performed, (Fig. 1) on the assumption that the chi-square distance between points reflects correlation among variables [22]. In this study we described closely related individuals in the correspondence analysis as social clusters. The results demonstrated clustering of MSM around 'Internet' on the left, in close proximity

with 'sex party'. A less well-defined cluster associated with sauna can be delineated on the right of the diagram. Discriminant analysis was then performed to compare between the 'Internet-centred' and 'Sauna-centred' social clusters (Table 3). The main distinguishing features of MSM in the 'Internet-centred' social cluster were: younger age, higher education level, going for regular HIV testing, having all three types of partners (lover, regular and casual) simultaneously. There was no association between either of the two social clusters and the individuals' initial presentation with complications. Among Internet users, chatroom was the most popular channel for partner sourcing. Comparing between MSM of 'Internet-centred' social cluster with noncluster Internet users, there was a higher proportion



**Fig. 1. Two-mode correspondence analysis on a cohort of HIV infected MSM.** Seventy-three MSM (open circles, with serial no.) are plotted using data on their use of venues (solid square) for sourcing sex partners in the 1-year period prior to their infection, as derived from a six-point Likert Scale. The circle with dotted line on the right represents the 'Sauna-centred' social cluster, whereas the left refers to the 'Internet-centred' social cluster.

**Table 3. Discriminant analysis between 'Internet-centred' and 'Sauna-centred' social cluster from two-mode correspondence analysis.**

	Internet (N=25)	Sauna (N=38)	ORs/95% CI
Age at infection ( <i>t</i> -test)*	27.96 (7.2)	31.14 (9.0)	<i>t</i> = -3.47, <i>P</i> < 0.01
Education level*			
Tertiary or above	16 (64.0%)	11 (29.7%)	4.20/1.43, 12.36
Secondary or below	9 (36.0%)	26 (70.3%)	
Residency			
Living with others	18 (72.0%)	28 (75.7%)	0.83/0.26, 2.62
Living alone	7 (28.0%)	9 (24.3%)	
HIV testing			
VCT	11 (45.8%)	12 (34.3%)	1.62/0.56, 4.70
Non-VCT	13 (54.2%)	23 (65.7%)	
Reasons for HIV testing			
Perceived chance of infection	8 (32.0%)	7 (18.4%)	2.08/0.64, 6.74
Partner was infected	1 (4.0%)	4 (10.5%)	0.35/0.04, 3.37
Peer encouragement	1 (4.0%)	5 (13.2%)	0.28/0.03, 2.51
Regular HIV testing*	7 (28.0%)	3 (7.9%)	4.54/1.05, 19.68
Hospitalization	4 (16.0%)	13 (34.2%)	0.37/0.10, 1.29
STD infection	2 (8.0%)	4 (10.5%)	0.74/0.13, 4.37
Other complications	5 (20.0%)	2 (5.3%)	4.50/0.80, 25.35
Partner types			
Lover*	22 (88.0%)	21 (56.8%)	5.59/1.42, 22.00
Regular partner*	18 (72.0%)	10 (26.3%)	7.20/2.32, 22.35
Casual partner	21 (84.0%)	34 (89.5%)	0.62/0.14, 2.74
All three partner types*	13 (52.0%)	7 (18.4%)	4.81/1.54, 14.93

\**P* < 0.05.

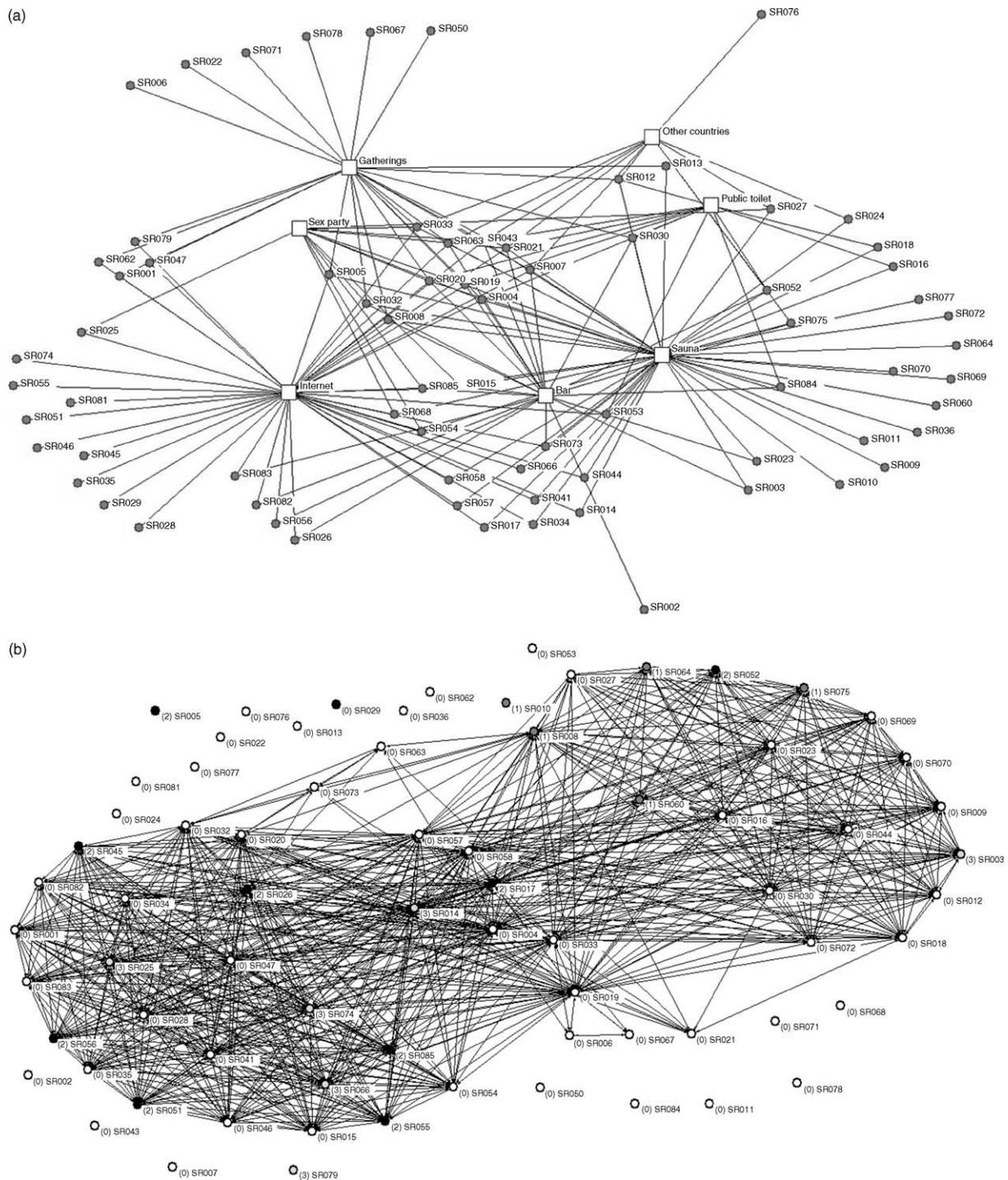
of frequent chatroom users in the former (84% vs. 21.4%  $\chi^2 = 14.85$ , *P* < 0.01).

Figure 3 is a two-mode network showing the relationship between MSM and partner-sourcing venues, with lines indicating their connection with the respective venue. The diagram shows that most MSM sourced sex partners either through Internet or Sauna. The interpersonal linkages were assessed after transformation into one-mode network of MSM in binary relationships (Fig. 2(b)). The density of the latter network, after excluding the standalone cases, was 0.51. Three forms of centrality were calculated: degree centrality is the number of links incident on one person; closeness is a measure of the geodesic distance (therefore closeness) between one person and all other reachable persons; and betweenness centrality refers to the fraction of the shortest links passing through a person [19]. In this reconstructed network, the median degree centrality of 'Internet-centred' social cluster MSMs and 'Sauna-centred' MSMs was 28 (range 0 to 45) and 16 (range 0 to 47) respectively. A higher proportion of 'Sauna-centred' social cluster MSMs had a degree centrality of zero compared with 'Internet-centred' social cluster cases (34.2 vs. 12%). There was little betweenness in the network, with a mean Freeman betweenness of 12.06 and a standard deviation of 25.37. The normalized closeness measure, using Freeman geodesic paths, was 74.12, with a standard deviation of 9.46. The cases with highest closeness and betweenness centrality were those with a preference for both sauna and Internet for sourcing partners, with a varying Likert score for other venues.

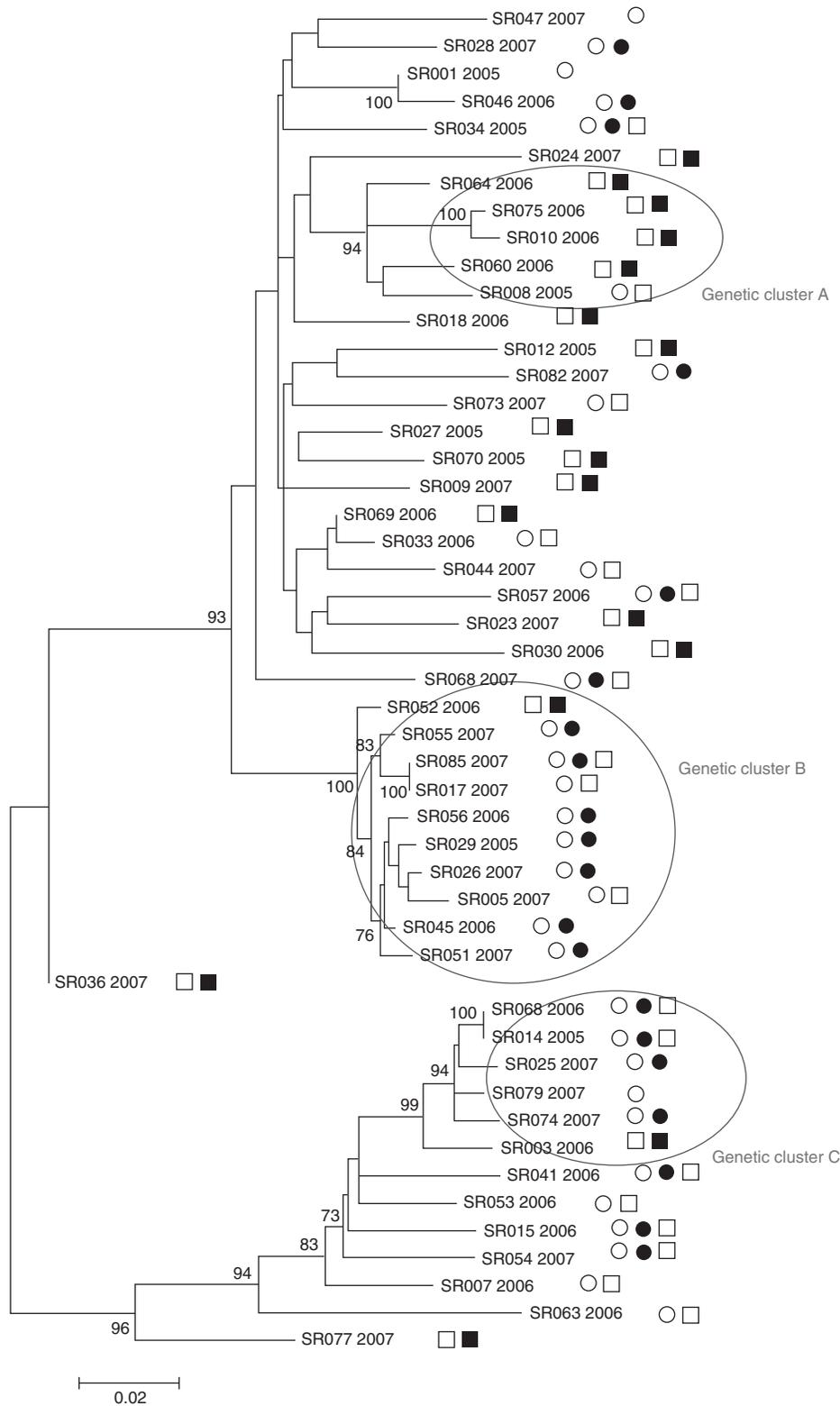
A phylogenetic tree was constructed (Fig. 3). There was no statistically significant difference in the proportion of Internet users between sequenced and nonsequenced samples. Three genetic clusters could be visualized. MSMs in Genetic Cluster A had used sauna for sourcing sex partner, whereas a majority of the individuals in Genetic Cluster B (9 out of 10) and C (5 out of 6) had used Internet for partner sourcing. MSM from the 'Internet-centred' social cluster accounted for seven out of 10 members in genetic cluster B, four out of six in cluster C, and none in cluster A. The speculated year of infection was 2006 in all except one person in genetic cluster A, whereas that for B and C could be dated to range from 2005 to 2007, with over half (nine out of 16) in 2007. The Internet associated genetic cluster B was well defined, whereas the sauna associated genetic cluster A was smaller and had longer branch lengths. The small genetic cluster C appeared to be just beginning to emerge. The number of MSM is however too small for comparison to be made between the two Internet-associated genetic clusters (B and C). Overall, the virus strains of MSM from 'Sauna-centred' social clusters were genetically less closely related to one another than those from 'Internet-centred' social cluster.

## Discussions

On the basis of an integration of social network properties and the phylogenetic relationships of the implicating virus, we developed an exploratory study to characterize



**Fig. 2. Network diagrams of MSM in the study ( $n = 73$ ).** (a) A 2-mode network of MSM (dark circle with serial number) and their venues (open square with label) for partner-sourcing, with each link representing a positive response of the MSM to the question on the respective venue; (b) A one-mode network created from data transformed from two-mode network after dichotomization and cross-products determination (see Methods). Most of the nodes on the left side of the network represent MSM on the 'Internet-centred' social cluster, whereas those on the right are largely MSM in 'Sauna-centred' social cluster. The freestanding circles are those without linkage after transformation. Each case is denoted by serial number (SR) with prefix representing cases from genetic cluster A (dark grey circles), B (black circles), C (light grey), and (0) from nongenetic cluster cases (open circle).



**Fig. 3. Phylogenetic tree of HIV positive MSM.** Open circle: Internet user; solid circle: 'Internet-centred' social cluster; open square: sauna user; solid square: 'Sauna-centred' social cluster; year indicating that speculated for HIV infection.

clustering of HIV infected MSM in Hong Kong. Though these were all recently diagnosed HIV infection, interpretation is cautioned, as they did not necessarily

indicate recent infection. To date relatively few studies have used social network methodologies to study HIV epidemiology [23–25], a phenomenon that has stemmed

from a number of reasons. Firstly, conventional full network analysis requires the collection of complete linkage data from all individuals, a method that has proven to be impossible in HIV epidemiology, both because of the technical difficulties as well as the problem with protecting the confidentiality of respondents. Such a massive exercise, even if conducted, would unlikely be repeatable. Secondly, there is the lack of a robust measure for quantifying linkage between HIV infected individuals. In the investigation of acute infection – for example, influenza, gonorrhoea or severe acute respiratory syndrome (SARS) – the characteristics of connections between individuals over a short period of time can be mapped directly. The same approach cannot be effectively applied to HIV, which is characterized by a long asymptomatic period and chronic disease.

To explore common source infection, we turned our attention to the duality of place and people in defining partnership, an approach that has been described in the investigation of sexual structure in the population [26]. We reckon that the location for partner sourcing is more important than location for sex, as two persons might choose to have sex in a place (for example, one's own home) that bears no relationship with how they are connected in the first place. We consider that without partner sourcing, there would have been no sexual activity between a respondent and his partner(s). Our study does not have the power to determine if an MSM actually knew another MSM in a specific venue. As MSM using the same partner-sourcing venue are more likely to have behaved in a similar way, the potential for virus transmission is theoretically higher than those who have used a different venue. By ranking the people–place relationship with Likert scale, a common tool in behavioural studies, we were able to infer the strength of relationship between the MSM. Unlike approaches founded on a presumed random distribution of MSM in the society, we attempted to map their social linkage and thereby define clusters that might have underlined transmission of the virus. The characterization of clusters has in fact been pursued in the study of other infection like acute sexually transmitted infection (STI) [27–29] and tuberculosis [30–33], the contexts of which are different from HIV, a chronic infection that is acquired sexually. In this study we revealed two social clusters by analysing data collected on people–place affinity. ‘Internet-centred’ social cluster members had a higher degree centrality, meaning that there is more linkages incident on each MSM in the cluster, compared to ‘Sauna-based’ social cluster. This pattern confirmed our belief that the use of Internet might have predisposed to the recent increase in HIV transmission in MSM in Hong Kong [18]. Compared with a recent study of genetic clusters of subtype B strains in MSM [34], our study further suggested the emergence of multiple clusters, which is also associated with Internet use for sex partner sourcing.

In this study, we did not set out to validate social clusters with genetic results, as the methodologies were totally different. Instead we are proposing to interpret SNA results in context of the phylogenetic analysis, so as to provide insights in the transmission dynamics of HIV in MSM. By resorting to sequences initially collected for resistance testing, we were able to simplify the process of sample collection. The same approach has been adopted by researchers exploring molecular phylodynamics for reconstructing HIV transmission network [35]. Three genetic clusters were uncovered in our study, two of which composed of members using Internet as the predominant channel for partner-sourcing and one using largely sauna instead. There was no perfect matching between members of genetic and social clusters. Interestingly, one member in each of the two Internet-based genetic clusters had not used Internet, whereas a member in the sauna based genetic cluster actually used Internet. These MSM probably formed the bridge between the two social networks, allowing the virus to spread through the wider MSM community in Hong Kong. These ‘bridges’ were all MSM using multiple channels for partner sourcing. Previous studies have suggested that the characterization of genetic clusters is a powerful tool to evaluate transmission chains, especially if primary HIV infection could be determined and tracked [36]. In field epidemiology this is not always possible. As an alternative, we used the year of infection suggested by interviewees to label the cases in our phylogenetic tree (Fig. 3). Although such information is subject to recall bias, we can see that the Internet-based genetic cluster appeared at a later date than the sauna-based genetic cluster, an observation that ties in with the general pattern of an increasing popularity of Internet as a means of sex partner sourcing. The temporal pattern cannot be validated because the exact time of infection is not ascertained in the study population. Behavioural questionnaires can never reflect the whole truth, and it is possible that a fraction of the ‘recent’ diagnoses may have resulted from old rather than incident infections. Interpretation is also limited by the small number of patients, though they represent a sizable proportion of all newly diagnosed cases seen in the largest clinic in the territory.

In conclusion, we have explored an interdisciplinary approach that incorporates behavioural science, SNA and molecular epidemiology for enhancing HIV surveillance, using MSM as an example. The application of a Likert Scale for characterizing partner-sourcing behaviours in conjunction with routinely collected genotyping data had provided us with a manageable procedure in busy clinics for generating potentially useful epidemiological information. By triangulating results obtained from the genetic and social network studies, intervention can be targeted at social clusters bearing significant genetic identity or vice versa. Further automation of phylogenetic and network analysis may enable clusters to be detected efficiently,

paving the way for an effective integration of HIV prevention and treatment, an intervention made possible through expanded access to antiretroviral therapy at population level [37]. Our new approach is, however, not without limitations. The ranking of people–place linkage with a Likert Scale and subsequent data transformation may have distorted the meaning of cluster. Likewise social network measures like centrality and density may not be interpretable in the same way as conventional SNA after multiple steps of data transformation. On the contrary, the broad categorization of ‘venues’ for partner sourcing may need to be further refined if a standard investigation tool is needed for advancing HIV surveillance. Collecting specific details of the people–place network data may, however, expose the identity of individuals, thereby infringing on ethical issues which demand attention of epidemiologists undertaking case investigations. On the phylogenetic analysis, the use of sequences for GRT may also have restricted the power of the study to define genetic clusters. Moreover, we were unable to exclude inclusion bias of regular partners of MSM, who may self-select to enroll in the clinical service. Finally, the assumptions embodied in the integrative approach should continue to be debated, such that more robust investigative tools could be developed in strengthening epidemiologic surveillance of HIV infection.

## Acknowledgement

The authors thank all patients at the Integrated Treatment Centre for consenting to participate in the study, and all nursing and medical staff for their assistance in the conduction of the study. The social network based studies (SPF-004) and the genotype resistance testing studies (MSS-099R) were supported by the AIDS Trust Fund of the Hong Kong Special Administrative Region Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Contributions of authors: S.S.L. conceptualized the study and prepared the manuscript drafts; D.T. conducted questionnaire survey and performed the social network analyses; Y.T. worked on the phylogenetic analysis and defined genetic cluster; W.L.M. and K.H.W. collected the clinical samples and provided support in clinical evaluation of patients involved in the study; J.C. performed genotype resistance testing and coordinated sample collection; W.C.Y. designed the genotype resistance testing study and contributed to the editing; all read and approved the final version of the manuscript.

Conflicts of interest: The authors declare that there are no conflicts of interests. The opinions and assertions contained herein are private views of the authors and

do not reflect those of the Centre for Health Protection, Department of Health, Hong Kong Special Administrative Region Government.

## References

- Centers for Disease Control and Prevention. **Guidelines for investigating clusters of health events.** *MMWR Recomm Rep* 1990; **39(RR-11)**:1–23.
- Foxman B, Riley L. **Molecular epidemiology: focus on infection.** *Am J Epidemiol* 2001; **153**:1135–1141.
- Shpilberg O, Dorman JS, Ferrell RE, Trucco M, Shahar A, Kuller LH. **The next stage: molecular epidemiology.** *J Clin Epidemiol* 1997; **50**:633–638.
- Simbayi LC, Kalichman SC, Strebel A, Cloete A, Henda N, Mqeketo A. **Disclosure of HIV status to sex partners and sexual risk behaviours among HIV-positive men and women, Cape Town, South Africa.** *Sex Transm Infect* 2007; **83**:29–34.
- Woodhouse DE, Potterat JJ, Rothenberg RB, Darrow WW, Klovdahl AS, Muth SQ. **Ethical and legal issues in social network research: the real and the ideal.** *NIDA Res Monogr* 1995; **151**:131–143.
- Hariri S, McKenna MT. **Epidemiology of human immunodeficiency virus in the United States.** *Clin Microbiol Rev* 2007; **20**:478–488.
- Kallings LO. **The first postmodern pandemic: 25 years of HIV/AIDS.** *J Intern Med* 2008; **263**:218–243.
- Rutland E, Foley E, O’Mahony C, Miller RF, Maw R, Kell P, *et al.* **How normalised is HIV care in the UK? A survey of current practice and opinion.** *Sex Transm Infect* 2007; **83**:151–154.
- Buchacz K, Klausner JD, Kerndt PR, Shouse RL, Onorato I, McElroy PD, *et al.* **HIV incidence among men diagnosed with early syphilis in Atlanta, San Francisco, and Los Angeles, 2004 to 2005.** *J Acquir Immune Defic Syndr* 2008; **47**:234–240.
- LikataVICIUS G, Klavs I, Devaux I, Alix J, Nardone A. **An increase in newly diagnosed HIV cases reported among men who have sex with men in Europe, 2000–6: implications for a European public health strategy.** *Sex Transm Infect* 2008; **84**:499–505.
- The Commission on AIDS in Asia. *Redefining AIDS in Asia, Crafting an effective response.* India: Oxford University Press 2008. [Epub ahead of print]
- Wolitski RJ, Valdiserri RO, Denning PH, Levine WC. **Are we headed for a resurgence of the HIV epidemic among men who have sex with men?** *Am J Public Health* 2001; **91**:883–888.
- Crepaz N, Lyles CM, Wolitski RJ, Passin WF, Rama SM, Herbst JH, *et al.* **Do prevention interventions reduce HIV risk behaviours among people living with HIV? A meta-analytic review of controlled trials.** *AIDS* 2006; **20**:143–157.
- Centers for Disease Control and Prevention. **Incorporating HIV prevention into the medical care of persons living with HIV. Recommendations of CDC, the Health Resources and Services Administration, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America.** *MMWR Recomm Rep* 2003; **52(RR-12)**:1–24.
- Gordon CM, Forsyth AD, Stall R, Cheever LW. **Prevention interventions with persons living with HIV/AIDS: state of the science and future directions.** *AIDS Educ Prev* 2005; **17 (1 Suppl A)**:6–20.
- Morris M. *Network epidemiology: a handbook for survey design and data collection.* Oxford: Oxford University Press; 2004.
- Rothenberg RB, Woodhouse DE, Potterat JJ, Muth SQ, Darrow WW, Klovdahl AS. **Social networks in disease transmission: the Colorado Springs Study.** *NIDA Res Monogr* 1995; **151**:3–19.
- Lee SS, Tam DKP, Ho RLM, Wong KH. **Shift of preference of location for sexual partnership in men having sex with men in an Asian population.** *Sexual Health* 2008; **5**:373–374.
- Hanneman RA, Riddle M. *Introduction to social network methods.* <http://faculty.ucr.edu/~hanneman/>; 2005. [Accessed 23 December 2008]
- Chen JH, Wong KH, Chan K, Lam HY, Lee SS, Li P, *et al.* **Evaluation of an in-house genotyping resistance test for HIV-1 drug resistance interpretation and genotyping.** *J Clin Virol* 2007; **39**:125–131.

21. Guindon S, Gascuel O. **A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003; **52**:696–704.
22. Greenacre M, Pardo R. **Subset correspondence analysis: visualizing relationships among a selected set of response categories from a questionnaire survey.** *Sociol Methods Res* 2006; **35**:193–218.
23. Helleringer S, Kohler HP. **Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi.** *AIDS* 2007; **21**:2323–2332.
24. Rothenberg R, Muth SQ, Malone S, Potterat JJ, Woodhouse DE. **Social and geographic distance in HIV risk.** *Sex Transm Dis* 2005; **32**:506–512.
25. Knapper CM, Roderick J, Smith J, Temple M, Birley HD. **Investigation of an HIV transmission cluster centred in South Wales.** *Sex Transm Infect* 2008; **84**:377–380.
26. Frost SD. **Using sexual affiliation networks to describe the sexual structure of a population.** *Sex Transm Infect* 2007; **83** (Suppl 1):i37–i42.
27. De P, Singh AE, Wong T, Yacoub W, Jolly AM. **Sexual network analysis of a gonorrhoea outbreak.** *Sex Transm Infect* 2004; **80**:280–285.
28. Rothenberg RB, Sterk C, Toomey KE, Potterat JJ, Johnson D, Schrader M, et al. **Using social network and ethnographic tools to evaluate syphilis transmission.** *Sex Transm Dis* 1998; **25**:154–160.
29. Cunningham SD, Michaud JM, Johnson SM, Rompalo A, Ellen JM. **Phase-specific network differences associated with the syphilis epidemic in Baltimore city, 1996–2000.** *Sex Transm Dis* 2004; **31**:611–615.
30. Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. **A preventable outbreak of tuberculosis investigated through an intricate social network.** *Clin Infect Dis* 2001; **33**:1801–1806.
31. Andre M, Ijaz K, Tillinghast JD, Krebs VE, Diem LA, Metchock B, et al. **Transmission network analysis to complement routine tuberculosis contact investigations.** *Am J Public Health* 2007; **97**:470–477.
32. Cook VJ, Sun SJ, Tapia J, Muth SQ, Arguello DF, Lewis BL, et al. **Transmission network analysis in tuberculosis contact investigations.** *J Infect Dis* 2007; **196**:1517–1527.
33. Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, Musser JM. **Networks and tuberculosis: an undetected community outbreak involving public places.** *Soc Sci Med* 2001; **52**:681–694.
34. Leung TWC, Mak D, Wong KH, Wang Y, Song YH, Tsang DNC, et al. **Molecular epidemiology demonstrated three emerging clusters of HIV type 1 subtype B infection in Hong Kong.** *AIDS Res Hum Retroviruses* 2008; **24**:903–910.
35. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Brown AJL. **Episodic sexual transmission of HIV revealed by molecular phylodynamics.** *PLoS Med* 2008; **5**:e50; doi:10.1371/journal.pmed.0050050.
36. Brenner B, Roger M, Routy J-P, Moisi D, Ntemgwa M, Matte C, et al., and the Quebec Primary HIV Infection Study Group. **High rates of forward transmission events after acute/early HIV-1 infection.** *J Infect Dis* 2007; **195**:951–959.
37. Salomon JA, Hogan DR, Stover J, Staneck KA, Walker N, Ghys PD, et al. **Integrating HIV prevention and treatment: from slogans to impact.** *PLoS Med* 2005; **2**:e16; doi:10.1371/journal.pmed.0020016.