Infer HIV transmission dynamics from gene sequences among young men who have sex with men in China

Xin Jin, Jinjin Cheng, Jie Lou[*]

Department of Mathematics, Shanghai University, 99 Shangda Road, Shanghai, 200444, PR China

ABSTRACT

To investigate the transmission dynamics and temporal and spatial migration characteristics of HIV spread among men who have sex with men (MSM) in China, a total of 1012 HIV-1 partial pol sequences, including five subtypes, were studied. Bayesian analysis were applied for each subtype to infer its dynamic characters including the effective reproductive number (R_e) and migration process. The mean curve of each R_e was almost always greater than 1 (even the 95% highest posterior density (HPD) lower value) along with time, which supports the necessity for a comprehensive study about risk behaviors among young MSM group in China. We also should reappraise the free treatment strategy, especially the therapeutic effect during the free treatment policy.

1. Introduction

More than 30 years have passed since the first HIV case was reported among men who have sex with men (MSM) in the USA(Centers for Disease, 1981). MSM continues to be the major cause of new infection in the globe, including China (Beyrer et al., 2012). Even though China has implemented a number of effective public health measures to prevent and treat HIV/AIDS, which effectively reduced the spread of the disease (Lu et al., 2008; Qian et al., 2006; Wu et al., 2007; Zhang et al., 2011), the rapid increase of HIV infections among MSM population runs counter to the overall HIV epidemic (Li et al., 2011; Ye et al., 2012). Today, China has approximately 18 million MSM, making them one of the most important target risk populations for HIV prevention(Shang et al., 2012). Better understanding the dynamics of HIV transmission could assist in the design and evaluation of preventive HIV/AIDS interventions.

A Bayesian approach about viral gene sequences can help to integrate various sources of information and uncertainties in the analyses. Phylogenetic analysis of viral sequences provides an alternative and independent route for reconstructing transmission networks of some RNA virus such as HIV/AIDS, Ebola virus, Dengue virus and even COVID-19 virus (ALIZON et al., 2014; GRENFELL et al., 2004; RAGONNET-CRONIN et al., 2018; VILLABONA-aRENAS et al., 2016). The core algorithm of Bayesian phylogenetic methods is Metropolis-Hastings Markov Chain Monte Carlo (MCMC) (HASTINGS, 1970; METROPOLIS et al., 1953). They are commonly used for viruses to infer epidemiological processes from genetic data (Pomeroy et al., 2008; Pybus et al., 2001). But only a few similar work can be found for HIV spread in MSM population in China. In

---

* Corresponding author.
   E-mail address: jie.lou@126.com (J. Lou).
   Peer review under responsibility of KeAi Communications Co., Ltd.

literature (Li et al., 2015), a total of 1205 HIV-1 nucleotide sequences of the 1.0-kb pol gene (HXB2: 2253−3278 nts) from newly diagnosed MSM between 2009 and 2014 from 13 Chinese provinces were determined. The sequences can be downloaded from GenBank with accession numbers KR822836-KR824040. In this study, total five HIV-1 subtypes were identified based on phylogenetic analyses of partial pol sequences. To estimate the evolutionary rate and the time of the most recent common ancestor (tMRCA) for the lineages, authors used Bayesian skyline plot demographic model under BEAST v.1.8.0.

In this study, we will further explore the transmission characteristics of HIV-1, such as its transmission dynamics and temporal and spatial migration characteristics among different areas, through the same 1205 gene sequences as that in (Li et al., 2015). We hope this can provide deeper insight into the evolutionary dynamics of HIV epidemics for future prevention.

2. Method

2.1. The BDSKY model

The Birth-Death skyline (BDSKY) model was used to estimate the effective reproductive numbers R_e for the five HIV lineages in this manuscript (Stadler et al., 2013). As one of Bayesian phylogenetic methods, BDSKY model is a random description of the population changing dynamics. It includes mainly three parameters. The transmission rate λ tells us the rate at which infected individuals infect susceptibles. The sampling rate ψ describe how likely it is for an infected individual to be sampled. The death rate μ is the rate at which lineages disappear (go extinct) from a population without being sampled, including individual's death, successful treatment or behavior changing. BDSKY can model piecewise constant transmission, death and sampling rates which change along with time. Define the effective reproductive number $Re = \lambda/(\psi + \mu)$ at each period, making it feasible to estimate the spread dynamics.

Sequence alignment was done using Mega, resulting in a total length of 1201bp for the partial pol sequences. Akaike Information Criterion (AIC) method (Akaike, 1974) was used to choose appropriate model among different substitution models (HKY+Γ+I, GTR+Γ+I) and clock models (strict, uncorrelated lognormal, uncorrelated exponential). Finally, the strict clock model and the GTR+Γ+I substitution model were chosen. BEAST analysis was performed using Markov Chain Monte Carlo (MCMC) runs of 50 million generations, discard the first 20% samples and sampled every 1000 steps. The Bayesian MCMC output was analyzed using Tracer v1.6.

2.2. The BDMM model

The multitype birth-death model (BDMM) was used to construct two subtypes' temporal and spatial migration characteristics (DeniseTanja et al., 2016). As an extension of BDSKY model, population structure is added in BDMM model. This model assumes that samples are taken from d discrete subgroups. Therefore, the parameters of BDMM model include dn transmission rates λ, dn death rates μ and dn sampling rates ψ. We use m_{ij} to represent the migration rate of individuals from subgroup i to subgroup j, so there are $(d^2 - d)$ migration rates. The migration time when subtypes were introduced into different areas were calculated.

We set the MCMC step to 80 million and discard the first 20% samples. Sample every 1000 steps. Finally the effective sample size of phylogenetic tree parameter is more than 200. The Bayesian MCMC output was analyzed using Tracer v1.6.

3. Results

We extracted the effective reproductive number R_e over time from gene sequences of each HIV-1 subtype through BDSKY model (Stadler et al., 2013) (Fig. 1). It shows the median line of Re and its 95% HPD interval for each of the five subtypes (CRF07BC, CRF01_AE_4, CRF01_AE_5, CRF 5501_B and CRFB_EU). There are big differences in the transmission dynamics among different subtypes. The Re of CRF07_BC and CRF01_AE_4 showed a downward trend. On the contrary, CRF01_AE_5 and CRF55_01B showed an upward trend. CRFB_EU was special, showing a wobbling pattern and finally reaching a peak of 3.05 (95% HPD: 1.71−5.00). Subtypes CRF01_AE_5 and CRFB_EU seemed more competitive than other subtypes as time goes on.

In addition, we can infer that the epidemic histories of the five subtypes were quite different. B_EU first entered MSM population very early, in about 1989, and was followed by CRF01_AE_5 (1997), CRF01_AE_4 (1998), CRF55_01B (2004), and CRF07_BC (2004) one by one (Fig. 1).

HIV outbreaks in each area of China have its own epidemic dynamic. But these epidemic dynamics are intertwined with the ancestor migration process of HIV. It is necessary to take into account the temporal and spatial structure of the sequences. Here, as the subtype having the largest R_e in the early stage of HIV transmission in MSM, total 48 CRF07_BC sequences sampled from three representative areas (Henan province in central China, Beijing in northern China and Shenzhen in southern China) were studied through BDMM model. The result was shown in Fig. 2.

In Fig. 2, the end of the Maximal Clade Credibility (MCC) tree is the annotated information of the sequence, which includes subtype type, sequence number, sampling area, and sampling time of sequence. Different colors on the MCC tree represent different area of the sequences. Different branching lengths of the tree represent evolutionary time. Color changes indicate the area changes. Each node is marked with a type probability to indicate the possibility of this type for this node.

We can see that subtype CRF07_BC sequences in the three areas can be traced back to Shenzhen, which can infer that this subtype in MSM population of Beijing and Henan might be introduced from Shenzhen. That is, we can understand as
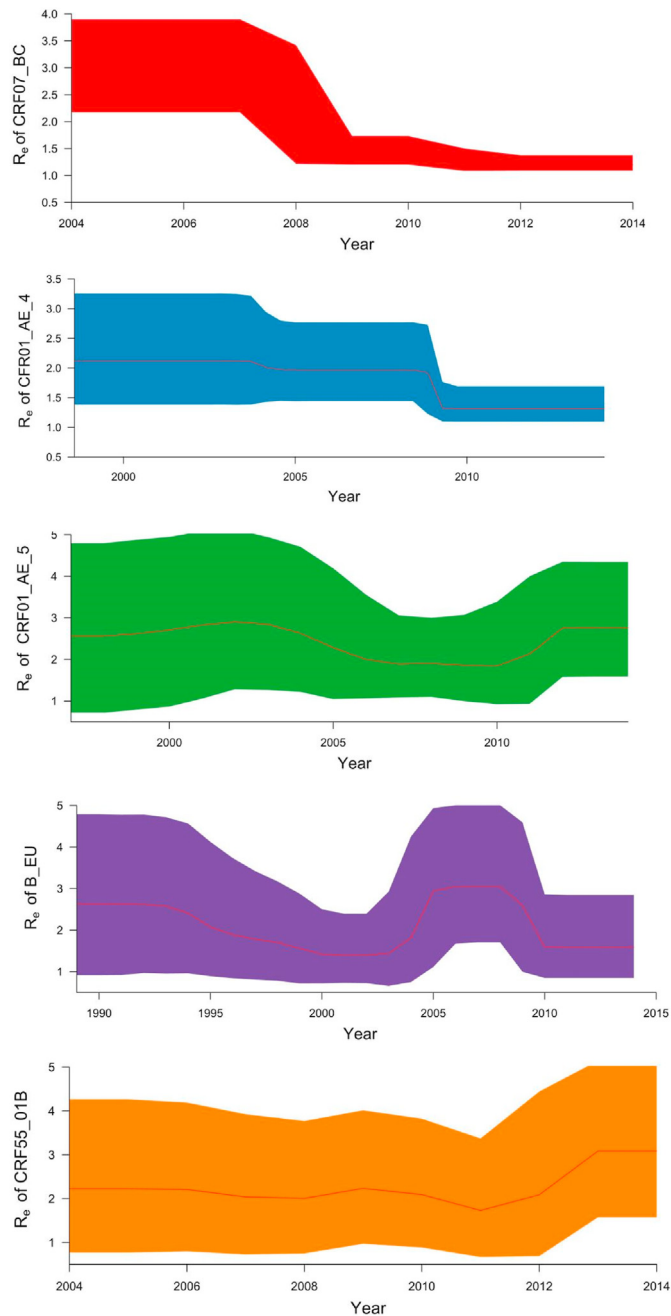
Fig. 1. Median of the effective reproductive number Re of different subtypes and their HPD intervals.

Shenzhen is the source of transmission of this HIV subtype in these three cities (type probability is 64%). Between 2007 and 2008, this subtype was mainly transmitted among MSM population within Shenzhen. Since 2008, it was gradually introduced into Beijing and Henan due to the ancestors' migration among different areas. It can be seen that the strain of Shenzhen was first introduced into Beijing and began to spread stably among MSM population in Beijing after 2009. This subtype was introduced into Henan from Shenzhen in early 2008 (which was relatively later than into Beijing), but it was steadily transmitted among MSM population in Henan in the first half of 2008, which was much earlier than in Beijing. The possible reason could be that the MSM population's mobility in Henan was much smaller than that in Beijing. On the whole, subtype CRF07_BC of Henan is mostly from Shenzhen and a small amount from Beijing. However, CRF07_BC of Beijing was all originated from the migration transmission of Shenzhen ancestors.
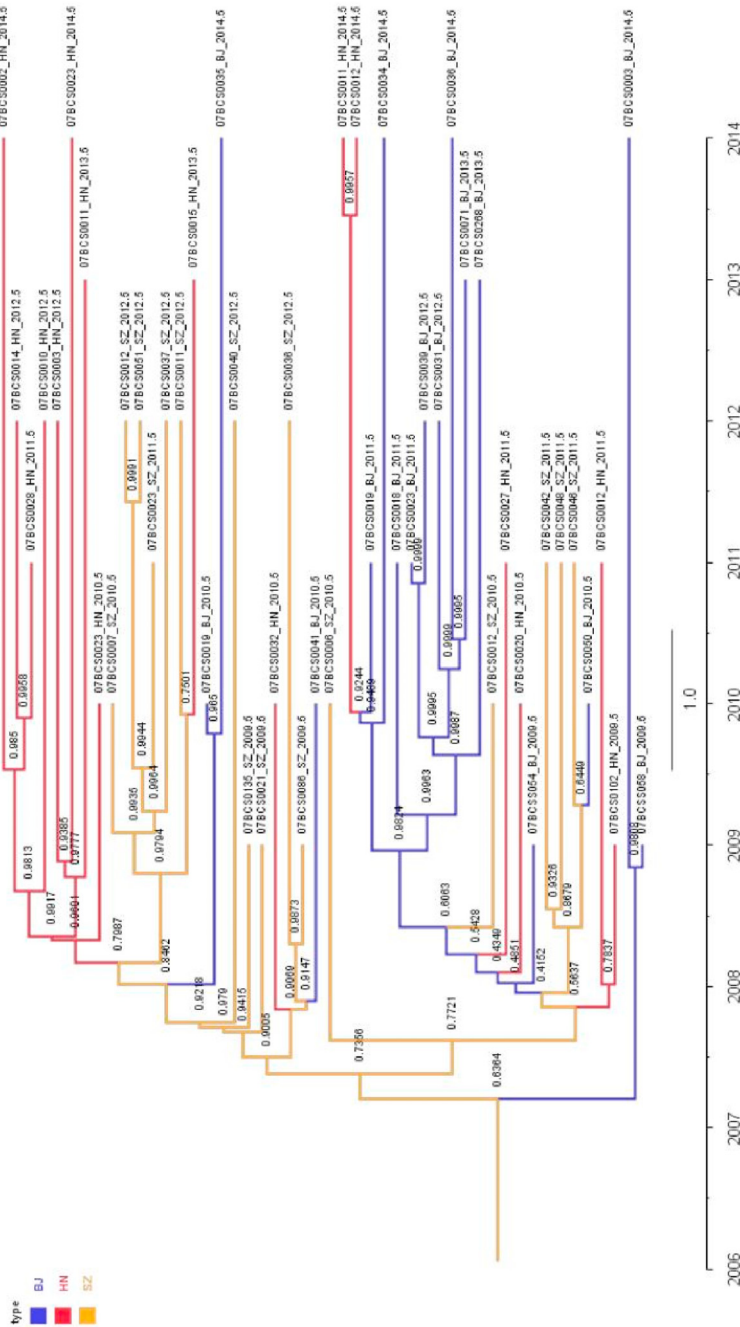
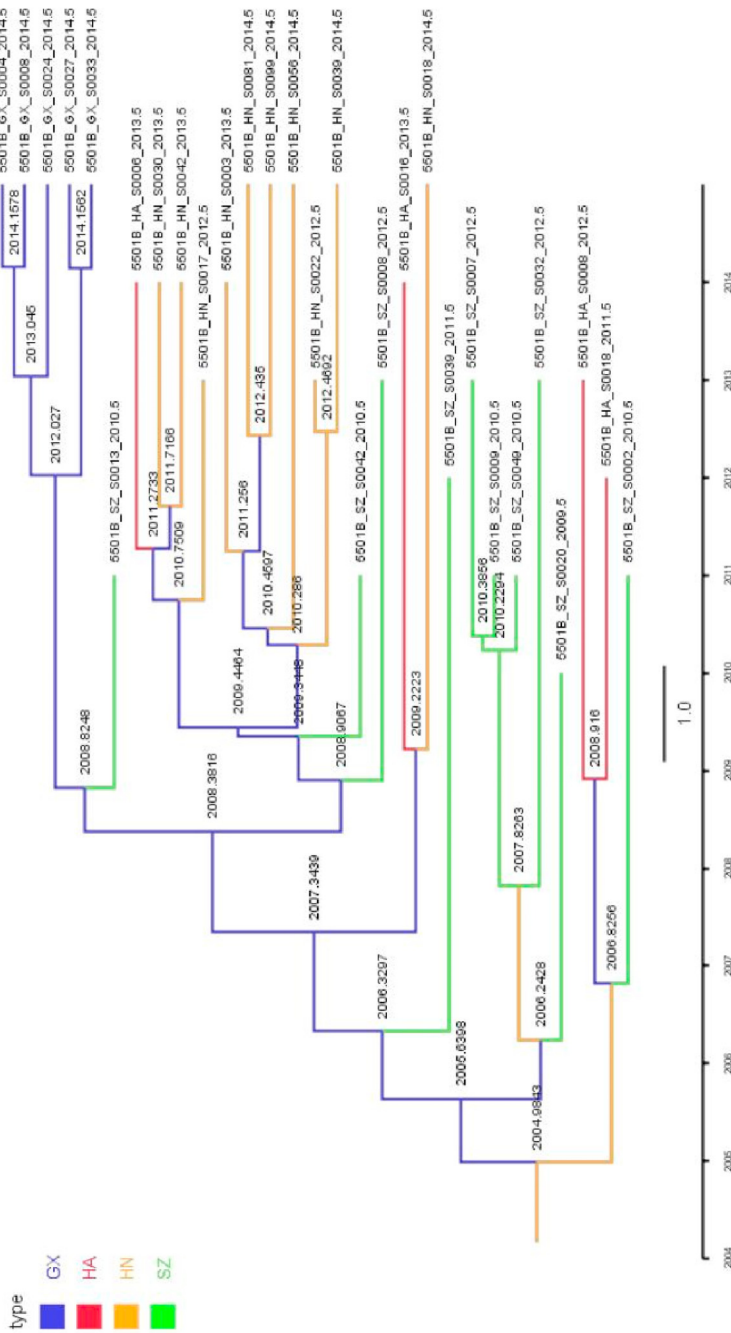Fig. 2. The MCC tree of CRF07_BC with time scale.

Fig. 3. The MCC tree of CRF55_01B with time scale.

We also studied CRF55_01B's migration characteristics among Guangxi province, Hunan province, Henan province and Shenzhen city through BDMM model since CRF55_01B is a rising star in recent years (Fig. 3). Here the node is represented by the ancestral time traced. It can be seen that the transmission of CRF55_01B in the MCC tree started from Henan MSM population, and gradually spread to other three provinces in 2005. It was introduced into Guangxi province in the next half of 2005 along the migration of ancestors, and continued to be introduced into other provinces from Guangxi province in the following three years. The Guangxi strain of this subtype was introduced into Shenzhen in early 2006, and gradually stabilized in Shenzhen MSM population after 2007. It was introduced into Henan in 2010 and soon became stable in this area. The Hunan subtype strain was introduced in at different times, also from Guangxi province. Around 2012, CRF55_01B formed a stable transmission in Guangxi Province. It can be inferred that the Guangxi MSM population acted as a bridge among the four areas in terms of CRF55_01B's transmission. Moreover, we found that the median migration rates from Henan province to Hunan province (about 0.3291) and to Shenzhen city (about 0.3345) are higher than that to Guangxi province (about 0.2917). While the median migration rates from Guangxi province to other three areas are all higher, at around 0.35.

4. Discussion

HIV gene sequence contains all the information of virus heredity and variation, and even hides the transmission dynamics of infectious diseases. The effective use of genetic data can facilitate research on the spread of epidemics, but only few study that used bioinformatic techniques to track changes of HIV phylogenetic dynamics among Chinese MSM (Li et al., 2015).

Through BDSKY model, the origin of each of the five subtypes can be found in Fig. 1, which is quite close to the related literature (Li et al., 2015) by Bayesian skyline plot demographic model, except that tMRCA of subtype 01_AE_4 is about 4 years latter (1998) than that in literature (Li et al., 2015) (1994). In addition, through MCMC sampling, we obtained important transmission parameter for the epidemic: the effective reproductive number R_e varying over time, and also two important subtypes' temporal and spatial migration characteristics. The obtained results showed that the spread of HIV among young MSM population was still severe. It further indicates that the free treatment policy and prevention strategies in China did not effectively inhibit the spread of HIV virus.

Our research traced transmission history of HIV in young MSM population in China. But it is quite difficult to give some very clear information about how to control the virus spread just from the gene data. If we connect the results from gene data with some macroscopic (patch or not) propagation models, maybe our results will show power effect. In the future, we will do some research in this direction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Akaike, H. (1974). *A new look at the statistical model identification*. IEEE Trans-actions on Automatic Control.

Alizon, S., Lion, S., Murall, C. L., et al. (2014). Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. *Virulence, 5*(8), 825−827.

Beyrer, C., et al. (2012). Global epidemiology of HIV infection in men who have sex with men. *Lancet, 380*, 367−377. https://doi.org/10.1016/S0140-6736(12)60821-6

Centers for Disease, C. (1981). Pneumocystis pneumoniaCLos Angeles. *Morbidity and Mortality Weekly Report, 30*, 250−252.

Denise, K., Tanja, S., Vaughan, T. G., et al. (2016). Phylodynamics with migration: A computational framework to quantify population structure from genomic data. *Molecular Biology and Evolution, 33*(8), 2102−2116.

Grenfell, B. T., Pybus, O. G., Gog, J. R., et al. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 327−332.

Hastings, W. K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*.

Li, Z., Liao, L., Feng, Y., et al. (2015). Trends of HIV subtypes and phylogenetic dynamics among young men who have sex with men in China, 2009-2014. *Scientific Reports, 5*, 16708.

Li, H. M., Peng, R. R., Li, J., Yin, Y. P., Wang, B., Cohen, M. S., et al. (2011). HIV incidence among men who have sex with men in China: A meta-analysis of published studies. *PloS One*, Article e23431.

Lu, L., Jia, M., Ma, Y., Yang, L., Chen, Z., Ho, D. D., et al. (2008). The changing face of HIV in China. *Nature, 455*, 609−611.

Metropolis, N., et al. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*(6), 1087−1092.

Pomeroy, L., Bjornstad, O., & Holmes, E. (2008). The evolutionary and epidemiological dynamics of the paramyxoviridae. *Journal of Molecular Evolution, 66*, 98−106.

Pybus, O., Charleston, M., Gupta, S., Rambaut, A., Holmes, E., & Harvey, P. (2001). The epidemic behavior of thehepatitis C virus. *Science, 292*, 2323−2325.

Qian, H. Z., Vermund, S. H., Kaslow, R. A., Coffey, C. S., Chamot, E., Yang, Z., et al. (2006). Coinfection with HIV and hepatitis C virus in former plasma/blood donors: Challenge for patient care in rural China. *AIDS, 20*, 1429−1435.

Ragonnet-Cronin, M., Jackson, C., Bradley-Stewart, A., et al. (2018). Recent and rapid transmission of HIV among people who inject drugs in Scotland revealed through phylogenetic analysis. *The Journal of Infectious Diseases, 217*(12), 1875−1882.

Shang, H., et al. (2012). HIV prevention: Bring safe sex to China. *Nature, 485*, 576−577. https://doi.org/10.1038/485576a

Stadler, T., et al. (2013). BirthCdeath skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences, 110*(1), 228−233.

Villabona-aRENAS, C. J., DE Oliveira, J. L., DE Sousa-Capra, C., & etal. (2016). Epidemiological dynamics of an urban dengue 4 outbreak in Sao Paulo, Brazil. *PeerJ, 4*, Article e1892.

Wu, Z., Sullivan, S. G., Wang, Y., Rotheram-Borus, M. J., & Detels, R. (2007). Evolution of China's response to HIV/AIDS. *Lancet, 369*, 679—690.

Ye, S., Xiao, Y., Jin, C., Cassell, H., Blevins, M., Sun, J., et al. (2012). Effectiveness of integrated HIV prevention interventions among Chinese men who have sex with men: Evaluation of a 16-city public health program. *PloS One*, Article e50873.

Zhang, F., Dou, Z., Ma, Y., Zhang, Y., Zhao, Y., Zhao, D., et al. (2011). Effect of earlier initiation of antiretroviral treatment and increased treatment coverage on HIV-related mortality in China: A national observational cohort study. *The Lancet Infectious Diseases*, 516—524.