

# Phylogenetic Profile of HIV-1 Subtype B, CRF01\_AE and the Recently Emerging CRF51\_01B among Men Who Have Sex with Men (MSM) in Singapore

Kim Tien Ng<sup>1</sup>, Kah Ying Ng<sup>2</sup>, Wei Xin Khong<sup>2</sup>, Kuan Kiat Chew<sup>2</sup>, Palvinder Kaur Singh<sup>2</sup>, Joe Kwan Yap<sup>2</sup>, Mei Ting Tan<sup>2</sup>, Yee Sin Leo<sup>2</sup>, Oliver Laeyendecker<sup>3,4</sup>, Thomas C. Quinn<sup>3,4</sup>, Adeeba Kamarulzaman<sup>1</sup>, Kok Keng Tee<sup>1\*</sup>, Oon Tek Ng<sup>2</sup>

**1** Centre of Excellence for Research in AIDS (CERiA), Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, **2** Institute of Infectious Disease and Epidemiology, Communicable Disease Centre, Tan Tock Seng Hospital, Singapore, **3** Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland United States of America, **4** Laboratory of Immunoregulation, Division of Intramural Research, National Institute of Allergy and Infectious Disease, National Institutes of Health, Baltimore, Maryland United States of America

## Abstract

HIV-1 subtype B and CRF01\_AE are the predominant infecting subtypes among men who have sex with men (MSM) in Singapore. The genetic history, population dynamics and pattern of transmission networks of these genotypes remain largely unknown. We delineated the phylogenetic profiles of HIV-1 subtype B, CRF01\_AE and the recently characterized CRF51\_01B strains circulating among the MSM population in Singapore. A total of 105 (49.5%) newly-diagnosed treatment-naïve MSM were recruited between February 2008 and August 2009. Phylogenetic reconstructions of the protease gene (HXB2: 2239 – 2629), gp120 (HXB2: 6942 – 7577) and gp41 (HXB2: 7803 – 8276) of the *env* gene uncovered five monophyletic transmission networks (two each within subtype B and CRF01\_AE and one within CRF51\_01B lineages) of different sizes (involving 3 – 23 MSM subjects, supported by posterior probability measure of 1.0). Bayesian coalescent analysis estimated that the emergence and dissemination of multiple sub-epidemic networks occurred between 1995 and 2005, driven largely by subtype B and later followed by CRF01\_AE. Exponential increase in effective population size for both subtype B and CRF01\_AE occurred between 2002 to 2007 and 2005 to 2007, respectively. Genealogical estimates suggested that the novel CRF51\_01B lineages were probably generated through series of recombination events involving CRF01\_AE and multiple subtype B ancestors. Our study provides the first insight on the phylogenetic profiles of HIV-1 subtype B, CRF01\_AE and CRF51\_01B viral strains circulating among MSM in Singapore.

**Citation:** Ng KT, Ng KY, Khong WX, Chew KK, Singh PK, et al. (2013) Phylogenetic Profile of HIV-1 Subtype B, CRF01\_AE and the Recently Emerging CRF51\_01B among Men Who Have Sex with Men (MSM) in Singapore. PLoS ONE 8(12): e80884. doi:10.1371/journal.pone.0080884

**Editor:** Marco Salemi, University of Florida, United States of America

**Received:** September 2, 2013; **Accepted:** October 17, 2013; **Published:** December 2, 2013

**Copyright:** © 2013 Ng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Singapore National Medical Research Training Fellowship Grant (EDG10nov070 to OTN) and the Malaysia Ministry of Higher Education High Impact Research (HIR) grants (H-500001-00-A000012-000001 to KKT and E-000001-20001 to AK). Additional support was provided by the Division of Intramural Research, NIAID, NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Co-author Adeeba Kamarulzaman is a PLOS ONE Editorial Board member. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: k2tee@um.edu.my

## Introduction

Human immunodeficiency virus type 1 (HIV-1) is known for its extreme genetic diversity, owing to its high mutation rates, high viral turnovers and the persistent nature of its infection [1]. In addition to the 9 major subtypes (A - D, F - H, J and K) of the HIV-1 group M, frequent intersubtype recombinations have further diversified the virus into at least 61 known circulating recombinant forms (CRFs) and many unique recombinant forms (URF) [2]. While HIV-1 prevalence among the general population in low- and middle-income countries has exhibited a decreasing trend in recent years, men who have sex with men (MSM) continue to be disproportionately affected by HIV-1 infection, with approximately 19 times odds of being infected with HIV-1 [3]. Recent surveillance reports have shown significant increase in the prevalence of HIV-1 infections among MSM in several Asian countries [4-6].

In 2002, Kalish et al. [7] observed a pattern of HIV-1 subtype segregation by the main transmission risk groups in Singapore, namely MSM and heterosexual, that were commonly infected with HIV-1 subtype B and CRF01\_AE, respectively. A more recent study further confirmed that HIV-1 subtype B and CRF01\_AE continue to be the predominant genotypes circulating in Singapore, with a higher CRF01\_AE prevalence among MSM [8]. In addition, Ng et al. [9] reported the emergence of a novel HIV-1 CRF51\_01B lineage among MSM in Singapore which is genetically distinct from other previously established subtypes or recombinants. Despite the significant clinical and epidemiological impact attributed to HIV-1 subtype B, CRF01\_AE and CRF51\_01B in Singapore [10], the genetic history, population dynamics and pattern of transmission networks of these genotypes remains largely unexplored. Through better understanding of the origin and evolutionary behaviours of these genotypes and sexual network analysis, effective preventive measures can be customized

to prevent HIV-1 transmission among MSM [11,12]. In the present study, based on the HIV-1 protease and *env* (gp120 and gp41) genes, we attempt to delineate the genetic history and the phylodynamic profiles of HIV-1 subtype B and CRF01\_AE, as well as the little known CRF51\_01B lineages isolated from the MSM population in Singapore, by using a suite of phylogenetic tools that involve maximum likelihood and Bayesian coalescence approaches.

## Materials and Methods

### Ethics Statement

The study was approved by the National Healthcare Group Ethics Committee. Written informed consent was obtained from all subjects prior to sample collection.

### Study Subjects and Specimens

A total of 212 treatment-naïve HIV-1 positive subjects were interviewed and recruited from the Singapore Communicable Disease Centre (CDC) outpatient clinic between February 2008 and August 2009. Demographic information, namely age, gender and ethnicity, as well as transmission risk factors were obtained, followed by collection of plasma specimens that were serologically tested and confirmed to be HIV-1 positive.

### Phylodynamic Inference and Divergence Times of Transmission Clusters

As previously described by Ng et al [8], HIV-1 RNA was extracted, reverse-transcribed, followed by nested PCR amplification and population sequencing of the protease gene (HXB2: 2239 – 2629, 390 bp), gp120 (HXB2: 6942 – 7577, 636 bp) and gp41 (HXB2: 7803 – 8276, 474 bp) of the *env* gene. HIV-1 genetic subtype and potential transmission networks in this dataset were first deduced from neighbour-joining tree of each genomic region [13] using MEGA version 5.05 based on Kimura-2 parameter model [14]. The reliability of each transmission networks were further evaluated by a more robust maximum likelihood phylogenies, heuristically inferred using subtree pruning and regrafting, as well as nearest neighbour interchange algorithms implemented in PAUP version 4.0 [15]. The statistical significance of the branching orders was evaluated by bootstrap analysis of 1000 replicates. The most appropriate nucleotide substitution model was determined using FindModel, a web implementation of Modeltest available at the HIV database (<http://www.lanl.gov.com>). In addition, Bayesian maximum clade credibility (MCC) trees were constructed using BEAST 1.7 [16] to determine the posterior probability values for each network in subtype B, CRF01\_AE and CRF51\_01B. As previously described elsewhere, transmission network is defined as a network consisting of at least 3 isolates from different individuals [17] of the same geographical (i.e. country) origin [18], and a phylogenetic clade supported by bootstrap value of more than 90% and Bayesian posterior probability value of 1.0 at the tree node [17,19,20].

The time of most recent common ancestor (tMRCA) of the respective transmission networks observed in each genomic region were estimated by the Bayesian coalescent-based relaxed molecular clock model, conducted in BEAST 1.7.4 [16]. The uncorrelated lognormal model [21] nested in general time-reversible (GTR) nucleotide substitution model [22] with a proportion of invariant sites and four rate categories of gamma-distribution model of among site rate heterogeneity [23] was employed to estimate HIV-1 phylogenies, nucleotide substitution rates and to date the tMRCAs of each transmission networks. Different parametric demographic models namely constant population size,

exponential and logistic growth model as well as a nonparametric Bayesian skyline plot (BSP) was tested, with the best fits coalescent model selected by means of a Bayes factor (BF), determined by Tracer version 1.5 (<http://tree.bio.ed.ac.uk/software/tracer>), using marginal likelihood [24]. Bayes factors can be interpreted as follows:  $2 \ln BF < 2$  no evidence;  $2 - 6$  weak evidence;  $6 - 10$  strong evidence;  $> 10$  very strong evidence [25]. The Markov chain Monte Carlo (MCMC) analysis was computed for 50 million states sampled every 10,000 states and output was assessed for convergence by means of effective sampling size (ESS) after a 10% burn-in using Tracer. To minimize the effects of standard errors, only traces with ESS of at least 200 were accepted. Prior to the inference of tMRCA for each transmission network [26], tMRCA for subtype B' of Thai origin (isolates 93CNRL42, 96M145, 96TH\_NP1538, 96M081, 99TH\_C1416, 99MMmSTD101, 00TH\_C3198, 01CNHN24, 02CNHNsc11, 02CNHNsmx2 and 02CNHNsq4), which were thought to be the descendent of the ancestral subtype B lineages that emerged around 1980 to 1991 [27,28], were co-estimated and checked for aberration with the current tMRCA estimations.

Given the fact that each HIV-1 sequence in this time-scaled phylogeny represents a different subject, the branches linking two sequences through their most recent common ancestor (MRCA) represents one transmission event. Consequently, the distance between the MRCA and the previous node deduced the upper bound of the year between transmission events. To obtain a more representative picture on the magnitude of transmission in this cohort, the transmission events (in years) of these networks were estimated.

### Sequences

Partial sequences of the HIV-1 protease gene, gp120 and gp41 of *env* gene analyzed in this study have been deposited in GenBank under the accession numbers listed in **Table S1**.

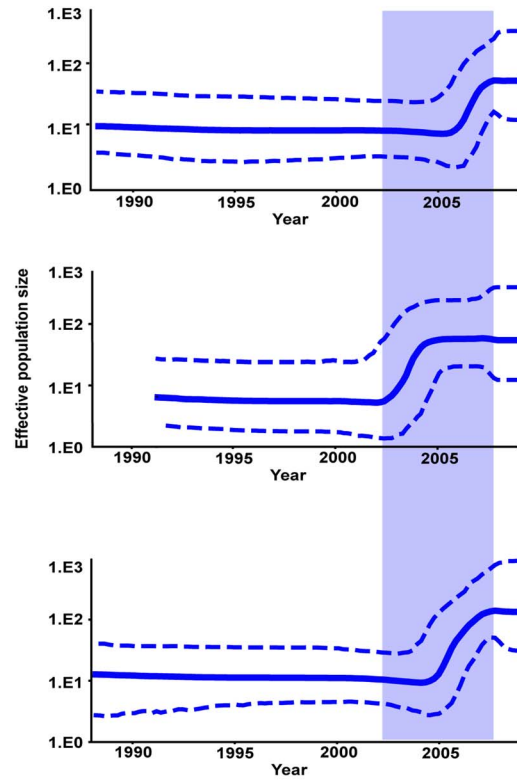
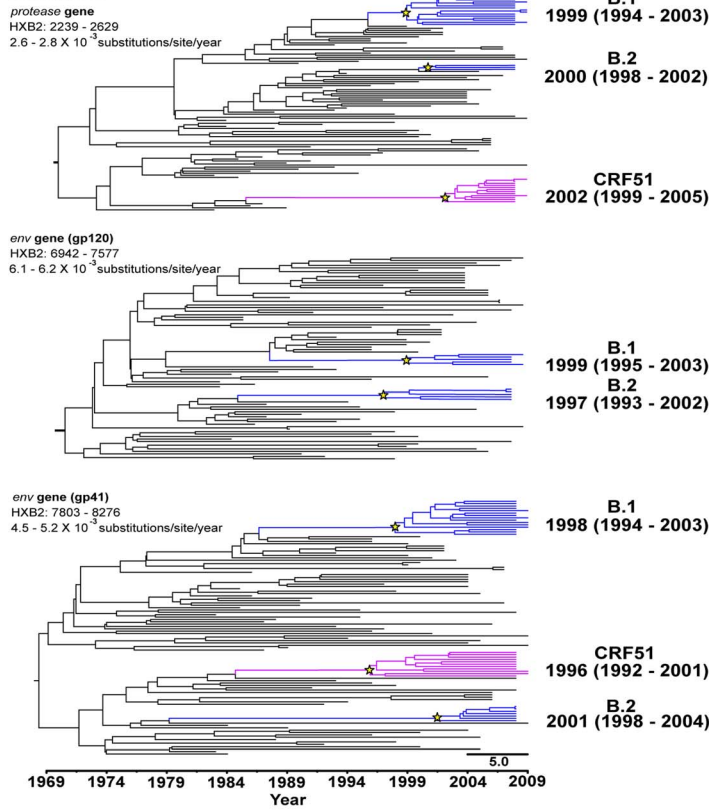
## Results

Between February 2008 and August 2009, 105 treatment-naïve HIV-1 positive MSM were recruited. In the present study, 89 (37 subtype B, 40 CRF01\_AE and 12 CRF51\_01B) interpretable protease sequences, 54 (16 subtype B, 29 CRF01\_AE and 9 CRF51\_01B) gp120 and 96 (43 subtype B, 42 CRF01\_AE and 11 CRF51\_01B) gp41 of *env* gene sequences were used for analysis. In addition to unique transmission lineages that do not cluster (data not shown), substantial clustering patterns (protease region: 64.0%; 57/89, gp120 region: 70.4%; 38/54 and gp41 region: 62.5%; 60/96) were observed.

Based on the significant statistical supports generated at the internal tree nodes of maximum likelihood and Bayesian's maximum clade credibility (MCC) tree analysis for the three genetic regions, phylogenetic reconstruction uncovered four monophyletic transmission networks, two each within subtype B lineages, denoted B.1, with 5 – 17 MSM subjects and B.2, with 3 – 7 MSM subjects (**Fig. 1a**) and CRF01\_AE lineage, denoted AE.1, with 16 – 23 MSM subjects and AE.2, with 3 – 4 MSM subjects (**Fig. 1b**). In addition, one transmission network was observed within CRF51\_01B lineages, denoted CRF51, with 9 – 12 MSM subjects (**Fig. 1a and 1b**). CRF51\_01B is a recently described recombinant, involving subtype B in protease and gp41, and CRF01\_AE in gp120 of *env* gene [8,9]. Chi-square analysis indicated that such clustering pattern have no significant association with ethnicity ( $P = .69$ ), age ( $P = .86$ ), viral load ( $P = .097$ ) and CD4 count ( $P = 0.81$ )

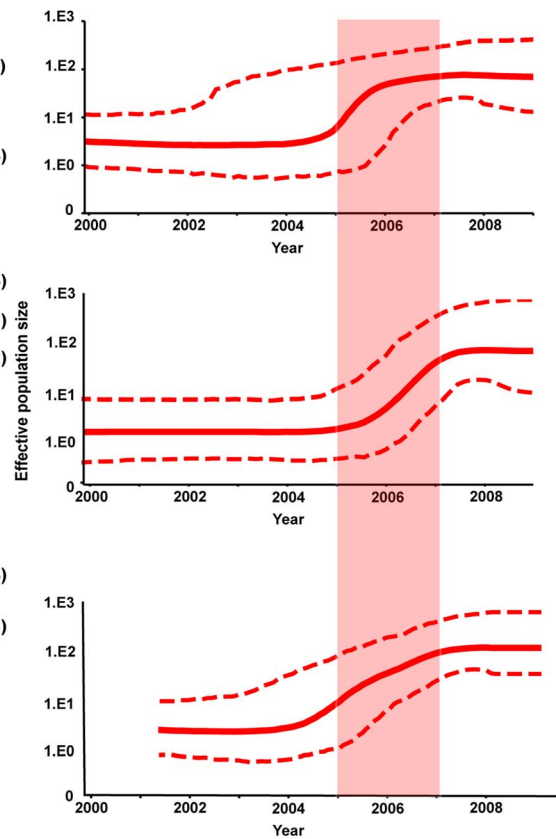
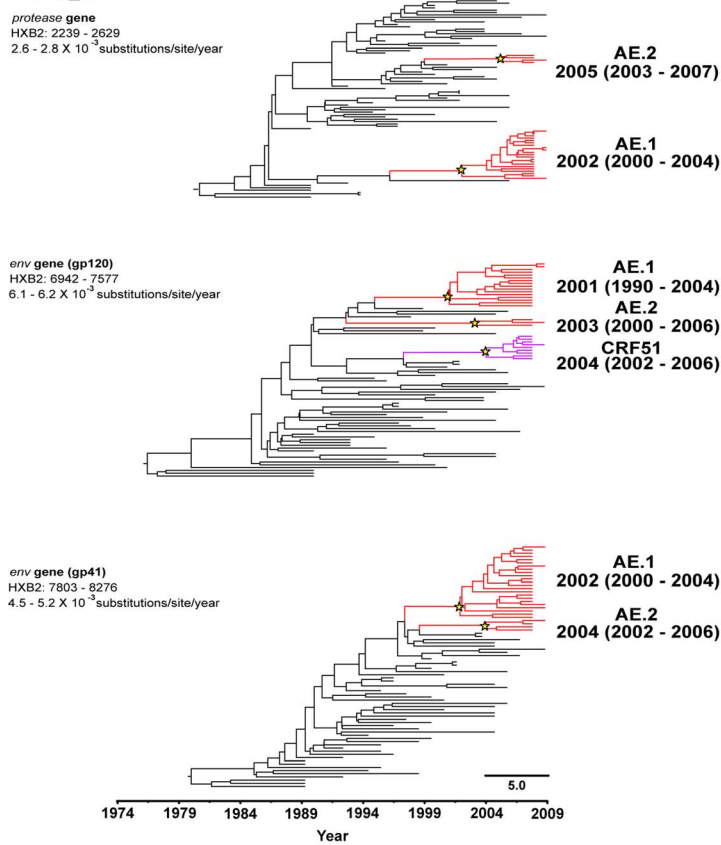
1a

Subtype B



1b

CRF01\_AE



**Figure 1. Phylodynamic profile of HIV-1 subtype B, CRF01\_AE and CRF51\_01B among men who have sex with men (MSM) in Singapore.** For clarity, only monophyletic transmission networks are shown. Bayesian's MCC trees of protease gene (HXB2: 2239 – 2629, 390bp), gp120 (HXB2: 6942 – 7577, 636bp) and gp41 (HXB2: 7803 – 8276, 474bp) of *env* gene from Singapore MSM sequence data collected between February 2008 and August 2009 are shown, with Singapore HIV-1 subtype B branches in blue (1a), CRF01\_AE in red (1b) and CRF51\_01B in purple and reference sequences in black branches. Of note, CRF51\_01B is a recombinant involving subtype B in protease and gp41 of *env* gene, and CRF01\_AE in gp120. The Bayesian coalescent-based relaxed molecular clock model was performed in BEAST 1.7, with uncorrelated lognormal model coupled with general time-reversible (GTR) nucleotide substitution model and a proportion of invariant sites. The Markov chain Monte Carlo (MCMC) analysis was computed for 50 million states sampled every 10,000 states and output was assessed for convergence by means of effective sampling size (ESS) after a 10% burn-in. Transmission networks are defined based on strong statistical inferences generated at the internal nodes of the maximum likelihood and MCC tree reconstructions (bootstrap values of more than 90% and posterior probability of 1, respectively). The mean tMRCAs and 95% highest posterior distribution (HPD) for each network are indicated in parentheses. The scale bar indicates the time in year. Bayesian skyline plots (BSPs) generated based on the time-stamped protease gene (HXB2: 2239 – 2629, 390bp), gp120 of *env* gene (HXB2: 6942 – 7577, 636bp) and gp41 of *env* gene (HXB2: 7803 – 8276, 474bp) from the MSM population in Singapore are shown. The origin and changes in effective population size through time for subtype B and CRF01\_AE in the country were estimated. The 95% HPD of the effective population size is indicated in dashed lines. doi:10.1371/journal.pone.0080884.g001

Analyses of these three genomic regions under a relaxed clock (uncorrelated lognormal) and constant coalescent model revealed a mean evolutionary rate of 2.6 – 2.8 (95% credible region: 2.1 – 3.5)  $\times 10^{-3}$  substitutions/site/year in protease region, 6.1 – 6.2 (5.0 – 7.6)  $\times 10^{-3}$  substitutions/site/year in gp120 region and 4.5 – 5.2 (3.8 – 6.4)  $\times 10^{-3}$  substitutions/site/year in gp41 region, overlaps with estimates reported elsewhere [29]. Based on these evolutionary parameters and estimates, the most recent common ancestor (tMRCAs) of each transmission network was estimated. The transmission networks of subtype B (B.1 and B.2) mainly surged around late 1990s to early 2000s with the oldest network most likely emerging as early as 1993 (Fig. 1a), while CRF01\_AE transmission networks (AE.1 and AE.2) diverged in early 2000s (Fig. 1b). Divergence times estimated for each subtype B and CRF01\_AE networks were concordant between genetic regions. In contrast, the tMRCAs for the recently isolated CRF51\_01B were estimated around 2002 (1999 – 2005), 2004 (2002 – 2006), and 1996 (1992 – 2001) in the protease gene, gp120 and gp41 of *env* gene, respectively. In addition, the tMRCAs for reference subtype B' isolates was estimated at around the late 1980s in all three genetic regions and in congruence with the findings from other studies [27,28,30].

To illustrate the past population dynamics of HIV-1 circulating among the Singapore MSM, temporal changes in effective population size over the last two decades for subtype B and CRF01\_AE were estimated. Bayesian skyline plot indicated the initial growth phase for subtype B lineages started in the late 1980s, followed by an exponential growth started between 2002 and 2007, and reached its peak from 2007 onwards (Fig. 1a). For CRF01\_AE lineages, the initial growth phase radiated in the early 2000s, and later undergone regular increase around 2005 to 2007, and reached sustained growth thereafter (Fig. 1b).

To estimate the time between transmission events within networks, each HIV-1 sequences in the time-scaled phylogeny represents a different patient, and distance between two adjacent tMRCAs denotes the upper boundary of the time between 2 transmission events. Estimation of the divergence times for each node revealed that the median interval between transmission events within these networks were short: 0.9 years (0.1 – 4.7) in subtype B, 0.75 years (0.1 – 1.7) in CRF01\_AE and 0.65 years (0.1 – 2.2) in CRF51\_01B.

## Discussion

Although HIV-1 transmission is increasing among the MSM population in Singapore and other countries in the region, a comprehensive analysis on the viral diversity, growth and temporal dynamics in this high-risk group remains scarce particularly in Southeast Asia. To our knowledge, this is the first study in Singapore investigating the transmission behaviour and evolutionary history of

the co-dominating HIV-1 subtype B, CRF01\_AE and newly emerged CRF51\_01B among the MSM population based on the time-stamped sequence data collected between 2008 and 2009. Here, we highlight the possible key factors that might have association with the pattern of HIV-1 infections among MSM in Singapore, namely the establishment of transmission networks, active spread of HIV-1 through MSM network and lastly, the emergence of CRF51\_01B.

The HIV-1 *pol* gene has proven informative and important in evaluating and defining transmission networks within a population of interest [31–33]. In the present study, gp120 and gp41 of *env* gene were co-analyzed. By definition, transmission networks are phylogenetic networks consisting of 3 or more nucleotide sequences from different individuals that are supported by strong statistical evidence, namely bootstrap value of more than 90% and Bayesian posterior probability of 1.0. Based on these inclusion criteria, significant clustering pattern was observed. Bayesian coalescent analysis estimated that the tMRCAs for all transmission networks originated between 1995 and 2005, with subtype B radiated around late 1990s, followed closely by CRF01\_AE that emerged in early 2000s. The short median interval (less than 1 year) between transmission events within these transmission networks indicate an ongoing active HIV-1 transmission and suggest possible epidemic expansion due to high-risk sexual behaviour among the MSM population [5,34,35]. The identification of viral transmission clusters coupled with the estimated tMRCAs suggest that HIV-1 subtype B and CRF01\_AE lineages were introduced into the MSM population in Singapore through at least 5 known sub-epidemics, which occurred between 1995 and 2005. Of note, similar findings in the same time frame were recently reported in neighbouring country, Malaysia [36] and in Western Europe [20,34,35], indicating that the emergence of transmission networks was likely caused by the increasing exposure to high-risk behaviours among MSM. Taken together, these findings suggest that most of HIV-1 infections among MSM in Singapore were likely to be linked to local networks. In addition, Brenner et al [37] had illustrated the direct correlation between networks size, transmission events and expansion rate, revealing the importance of understanding these transmission networks during assessment of ongoing epidemic or expanding outbreak. In the present study, such correlation was not observed, possibly due to the sample size of the study, sampled between February 2008 and August 2009 from one clinical centre. In order to have a clearer representation, an in-depth sampling approach spanning a longer recruitment period and involving all clinical centres in Singapore should be employed.

The first case of HIV-1 infection in Singapore was first identified in 1985, with the first case of AIDS reported in September 1986 [38]. Of note, Bayesian skyline plot for subtype B lineage indicated the initial growth phase started in late 1980s,

corroborating with epidemiological data that HIV-1 might have been introduced into Singapore MSM population around the mid 1980s. Genetic data also demonstrated the evolutionary history of HIV-1 subtype B and CRF01\_AE lineages spanning a period of almost two decades. Collectively, the Bayesian skyline plot showed an epidemic expansion from early to late 2000s, with a 10-fold exponential growth in effective population size of subtype B and CRF01\_AE observed between 2002 – 2007 and 2005 – 2007, respectively.

Interestingly, the genealogical analysis of the recently described CRF51\_01B revealed tMRCA estimates of 2002 (1999 – 2005), 2004 (2002 – 2006), and 1996 (1992 – 2001) in the protease, gp120 and gp41 of *env* gene, respectively. Such tMRCA estimates may suggest the following: a) the significantly older gp41 region indicates that there were at least two recombination events involving subtype B lineages of different evolutionary histories i.e. distinct divergence times, and b) the possible existence of a yet identified intermediate recombinant form (a founding precursor for CRF51\_01B) that was circulating at low level or have since gone extinct.

From the perspective of public health, sexual networks within the MSM population could serve as an important factor in onward HIV-1 transmission, thus it functions as the core entry point for the delivery of intervention strategies. However, the effectiveness and the success rate of any approach largely depends on factors such as ethnicity, race and most importantly, the mindset of general population towards MSM population [39], that may decrease the health-seeking behaviours among MSM [40–42]. In this context, educating both general population and MSM population on disease transmission and antiretroviral-based prevention strategy is imperative to prevent onward HIV-1 transmission.

## References

- Malim MH, Emerman M (2001) HIV-1 sequence variation: drift, shift, and attenuation. *Cell* 104: 469–472.
- Li X, Ning C, He X, Yang Y, Li F, et al. (2013) Genome Sequences of a Novel HIV-1 Circulating Recombinant Form (CRF61\_BC) Identified among Heterosexuals in China. *Genome Announc* 1.
- Baral S, Sifakis F, Cleghorn F, Beyrer C (2007) Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000–2006: a systematic review. *PLoS Med* 4: e339.
- Ananworanich J, Phanuphak N, de Souza M, Paris R, Arroyo M, et al. (2008) Incidence and characterization of acute HIV-1 infection in a high-risk Thai population. *J Acquir Immune Defic Syndr* 49: 151–155.
- Kanter J, Koh C, Razali K, Tai R, Izenberg J, et al. (2011) Risk behaviour and HIV prevalence among men who have sex with men in a multiethnic society: a venue-based study in Kuala Lumpur, Malaysia. *Int J STD AIDS* 22: 30–37.
- Nguyen TA, Nguyen HT, Le GT, Detels R (2008) Prevalence and risk factors associated with HIV infection among men having sex with men in Ho Chi Minh City, Vietnam. *AIDS Behav* 12: 476–482.
- Kalish ML, Korber BT, Pillai S, Robbins KE, Leo YS, et al. (2002) The sequential introduction of HIV-1 subtype B and CRF01AE in Singapore by sexual transmission: accelerated V3 region evolution in a subpopulation of Asian CRF01 viruses. *Virology* 304: 311–329.
- Ng OT, Munshaw S, Lamers SL, Chew KK, Lin L, et al. (2011) Molecular epidemiology of HIV type 1 in Singapore and identification of novel CRF01\_AE/B recombinant forms. *AIDS Res Hum Retroviruses* 27: 1135–1137.
- Ng OT, Eyzaguirre LM, Carr JK, Chew KK, Lin L, et al. (2012) Identification of New CRF51\_01B in Singapore Using Full Genome Analysis of Three HIV Type 1 Isolates. *AIDS Res Hum Retroviruses* 28: 527–530.
- Ng KY, Chew KK, Kaur P, Kwan JY, Khong WX, et al. (2013) High prevalence of CXCR4 usage among treatment-naive CRF01\_AE and CRF51\_01B-infected HIV-1 subjects in Singapore. *BMC Infect Dis* 13: 90.
- Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, et al. (2011) Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 365: 493–505.
- Grulich RM, Gilks CF, Dye C, De Cock KM, Williams BG (2009) Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet* 373: 48–57.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
- Swofford DL (2003) PAUP\*: phylogenetic analysis using parsimony (\*and other methods), version 4 beta. Sinauer Associates, Sunderland, MA.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
- Dennis AM, Hue S, Hurt CB, Napravnik S, Sebastian J, et al. (2012) Phylogenetic insights into regional HIV transmission. *AIDS* 26: 1813–1822.
- Aldous JL, Pond SK, Poon A, Jain S, Qin H, et al. (2012) Characterizing HIV transmission networks across the United States. *Clin Infect Dis* 55: 1135–1143.
- Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, et al. (2010) Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infect Dis* 10: 262.
- Zehender G, Ebranati E, Lai A, Santoro MM, Alteri C, et al. (2010) Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J Acquir Immune Defic Syndr* 55: 156–160.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88.
- Rodríguez F, Oliver JL, Marin A, Medina JR (1990) The general stochastic model of nucleotide substitution. *J Theor Biol* 142: 485–501.
- Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39: 306–314.
- Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18: 1001–1013.
- Kass R, Raftery A (1995) Bayes Factors. *J Am Stat Assoc* 90: 773–795.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307–1320.
- Deng X, Liu H, Shao Y, Rayner S, Yang R (2008) The epidemic origin and molecular properties of B': a founder strain of the HIV-1 transmission in Asia. *AIDS* 22: 1851–1858.
- Li Z, He X, Wang Z, Xing H, Li F, et al. (2012) Tracing the origin and history of HIV-1 subtype B' epidemic by near full-length genome analyses. *AIDS* 26: 877–884.

Due to social stigmatization, information on epidemic linkages among the subjects that may assist the transmission network estimation such as numbers of sexual partners, their sexual practices (receptive or insertive anal intercourse), ethnic preference, the use of condoms, location of infection and travel history are not readily revealed by the subjects. However, in this study, we show that the presence of possible predisposing epidemic linkages in a population (particularly those with limited epidemiological data) can be estimated largely based on the genetic information, highlighting the importance (and the ability) of viral genetic information in assessing the epidemic linkages and to study the viral phylodynamic. Since the present study only involved subject recruited between 2008 and 2009, we believe a study spanning a longer period may improve the resolution of HIV-1 genomic diversity and transmission dynamics within the MSM networks in Singapore.

## Supporting Information

**Table S1 Accession numbers and the partial sequences of the HIV-1 protease gene, gp120 and gp41 of *env* gene analyzed in this study.**

(TIF)

## Author Contributions

Conceived and designed the experiments: KKT OTN. Performed the experiments: KTN KKT OTN. Analyzed the data: KTN KKT OTN. Contributed reagents/materials/analysis tools: KTN KYN WXK KKC PS JKY MTT YSL OL TCQ AK KKT OTN. Wrote the paper: KTN KKT.

29. Tee KK, Pybus OG, Li XJ, Han X, Shang H, et al. (2008) Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08\_BC and 07\_BC in Asia. *J Virol* 82: 9206–9215.
30. Li Y, Uenishi R, Hase S, Liao H, Li XJ, et al. (2010) Explosive HIV-1 subtype B epidemics in Asia driven by geographic and risk group founder events. *Virology* 402: 223–227.
31. Brenner BG, Roger M, Moisi DD, Oliveira M, Hardy I, et al. (2008) Transmission networks of drug resistance acquired in primary/early stage HIV infection. *AIDS* 22: 2509–2515.
32. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ (2008) Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 5: e50.
33. Smith DM, May SJ, Tweeken S, Drumright L, Pacold ME, et al. (2009) A public health model for the molecular surveillance of HIV transmission in San Diego, California. *AIDS* 23: 225–232.
34. Giuliani M, Di Carlo A, Palamara G, Dorrucchi M, Latini A, et al. (2005) Increased HIV incidence among men who have sex with men in Rome. *AIDS* 19: 1429–1431.
35. Sullivan PS, Hamouda O, Delpech V, Geduld JE, Prejean J, et al. (2009) Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996–2005. *Ann Epidemiol* 19: 423–431.
36. Ng KT, Ong LY, Lim SH, Takebe Y, Kamarulzaman A, et al. (2013) Evolutionary History of HIV-1 Subtype B and CRF01\_AE Transmission Clusters among Men Who Have Sex with Men (MSM) in Kuala Lumpur, Malaysia. *PLoS One* 8: e67286.
37. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, et al. (2011) Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec. *J Infect Dis* 204: 1115–1119.
38. Chow KY, Ang LW, Verghesse I, Chew SK, Leo YS (2005) Measurable predictive factors for progression to AIDS among HIV-infected patients in Singapore. *Ann Acad Med Singapore* 34: 84–89.
39. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, et al. (2012) Global epidemiology of HIV infection in men who have sex with men. *Lancet* 380: 367–377.
40. Baral S, Burrell E, Scheibe A, Brown B, Beyrer C, et al. (2011) HIV risk and associations of HIV infection among men who have sex with men in peri-urban Cape Town, South Africa. *BMC Public Health* 11: 766.
41. Fay H, Baral SD, Trapence G, Motimedi F, Umar E, et al. (2011) Stigma, health care access, and HIV knowledge among men who have sex with men in Malawi, Namibia, and Botswana. *AIDS Behav* 15: 1088–1097.
42. Poteat T, Diouf D, Drame FM, Ndaw M, Traore C, et al. (2011) HIV risk among MSM in Senegal: a qualitative rapid assessment of the impact of enforcing laws that criminalize same sex practices. *PLoS One* 6: e28760.