

RESEARCH ARTICLE

# Genetic Characterization of a Novel HIV-1 Circulating Recombinant Form (CRF74\_01B) Identified among Intravenous Drug Users in Malaysia: Recombination History and Phylogenetic Linkage with Previously Defined Recombinant Lineages

Hui Ting Cheong<sup>1</sup>, Wei Zhen Chow<sup>1</sup>, Yutaka Takebe<sup>1,2,3</sup>, Jack Bee Chook<sup>1</sup>, Kok Gan Chan<sup>4</sup>, Haider Abdulrazzaq Abed Al-Darraj<sup>1</sup>, Clayton Koh<sup>1</sup>, Adeeba Kamarulzaman<sup>1</sup>, Kok Keng Tee<sup>1\*</sup>

**1** Centre of Excellence for Research in AIDS (CERiA), Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, **2** AIDS Research Center, National Institute of Infectious Diseases, Toyama, Shinjuku-ku, Tokyo, Japan, **3** School of Medicine, Yokohama City University, Yokohama, Kanagawa, Japan, **4** Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

\* [k2tee@um.edu.my](mailto:k2tee@um.edu.my)



**OPEN ACCESS**

**Citation:** Cheong HT, Chow WZ, Takebe Y, Chook JB, Chan KG, Al-Darraj HAA, et al. (2015) Genetic Characterization of a Novel HIV-1 Circulating Recombinant Form (CRF74\_01B) Identified among Intravenous Drug Users in Malaysia: Recombination History and Phylogenetic Linkage with Previously Defined Recombinant Lineages. *PLoS ONE* 10(7): e0133883. doi:10.1371/journal.pone.0133883

**Editor:** Dimitrios Paraskevis, University of Athens, Medical School, GREECE

**Received:** May 11, 2015

**Accepted:** July 3, 2015

**Published:** July 21, 2015

**Copyright:** © 2015 Cheong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All nucleotide sequences are available from the GenBank database (accession numbers KR019770–KR019775).

**Funding:** The work was supported by grants from the Ministry of Education, Malaysia: High Impact Research H-500001-00-A000012-000001 to KKT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

In many parts of Southeast Asia, the HIV-1 epidemic has been driven by the sharing of needles and equipment among intravenous drug users (IDUs). Over the last few decades, many studies have proven time and again that the diversity of HIV-1 epidemics can often be linked to the route of infection transmission. That said, the diversity and complexity of HIV-1 molecular epidemics in the region have been increasing at an alarming rate, due in part to the high tendency of the viral RNA to recombine. This scenario was exemplified by the discovery of numerous circulating recombinant forms (CRFs), especially in Thailand and Malaysia. In this study, we characterized a novel CRF designated CRF74\_01B, which was identified in six epidemiologically unlinked IDUs in Kuala Lumpur, Malaysia. The near-full length genomes were composed of CRF01\_AE and subtype B', with eight breakpoints dispersed in the *gag-pol* and *nef* regions. Remarkably, this CRF shared four and two recombination hotspots with the previously described CRF33\_01B and the less prevalent CRF53\_01B, respectively. Genealogy-based Bayesian phylogenetic analysis of CRF74\_01B genomic regions showed that it is closely related to both CRF33\_01B and CRF53\_01B. This observation suggests that CRF74\_01B was probably a direct descendent from specific lineages of CRF33\_01B, CRF53\_01B and subtype B' that could have emerged in the mid-1990s. Additionally, it illustrated the active recombination processes between prevalent HIV-1 subtypes and recombinants in Malaysia. In summary, we report a novel HIV-1 genotype designated CRF74\_01B among IDUs in Kuala Lumpur, Malaysia. The characterization of the novel CRF74\_01B is of considerable significance towards the

**Competing Interests:** The authors have declared that no competing interests exist.

understanding of the genetic diversity and population dynamics of HIV-1 circulating in the region.

## Introduction

Group M, the pandemic branch of HIV-1 can be classified into different subtypes named A-D, F-H, J and K. Recombinant lineages generated through recombination between subtypes isolated from at least three epidemiologically unlinked individuals are known as circulating recombinant forms (CRFs), else these recombinants would be known as unique recombinant forms (URFs) [1].

As of April 2015, at least 72 CRFs have been identified globally (<http://www.hiv.lanl.gov/>). Although HIV-1 subtypes A, B and C remained the most prevalent genotypes worldwide, CRFs can be attributed to at least 16% of HIV-1 infections globally and are anticipated to increase in proportion [2]. Similar patterns can be observed in Asia, where the complexity and diversity of HIV-1 increases in parallel to the cumulated number of HIV-1 infections, though the rates and extent varies [3, 4].

In Malaysia, the early phase of the HIV-1 epidemic was driven mostly by intravenous drug users (IDUs) that were associated with subtype B' of Thai origin and CRF01\_AE. Since then, numerous HIV-1 CRFs such as CRF33\_01B, CRF48\_01B, CRF52\_01B, CRF53\_01B, CRF54\_01B, and CRF58\_01B have been reported in Malaysia in the past decade [5–10], all of which have arisen from the predominant subtype B' and CRF01\_AE co-circulating in the country. These new CRFs could in fact further recombine with other subtypes or CRFs, generating second-generation recombinant descendants. One example is the CRF48\_01B discovered in 2007, a CRF descended from the CRF33\_01B lineage [6]. More recently, a new CRF candidate closely related to CRF33\_01B was also detected through molecular screening [11], indicating an increasing trend of second-generation CRFs emerging in the region. Therefore, one could anticipate that HIV-1 infections caused by CRFs would increase not only in proportion, but also in complexity over time [2]. In this study, we described a novel CRF candidate discovered among IDUs in Malaysia.

## Materials and Methods

### Ethics Statement

The study was approved by the University Malaya Medical Centre (UMMC) Medical Ethics Committee. Standard, multilingual consent forms allowed by the Medical Ethics Committee were used. Written consent was obtained from all study participants. Being an especially vulnerable population, all interviews and data collected from study participants were kept confidential. Potential participants who declined to take part in the study were not in any way disadvantaged from receiving medical attention, treatment and care.

### Amplification and genetic characterization of a novel second-generation CRF

Patients with history of unsafe injecting drug practices were recruited in Kuala Lumpur, Malaysia between 2010 and 2011. HIV-1 RNA was extracted from plasma samples through magnetic silica-based method implemented in the automated NucliSENS easyMAG platform (BioMérieux, France). Reverse transcription was carried out with random hexamers using SuperScript

III RNase H Reverse Transcriptase (Invitrogen, California, USA) according to manufacturer's instruction. Through amplification of the protease-reverse transcriptase (PRRT) genes and neighbour-joining analysis described in another study conducted earlier [11], four of 128 (3.1%) PRRT sequences from epidemiologically-unlinked study subjects forming a distinct cluster were identified. In addition, we identified two other isolates from our routine genotyping, which shares a clade with the four sequences identified previously. In order to study the detailed mosaic structures for this monophyletic cluster, near-full length genome amplification were carried out through nested PCR strategies, using sets of primers described in [S1 Table](#). PCR products were purified and sequenced through the ABI PRISM 3730XL DNA Analyzer (Applied Biosystems, California, USA). Nucleotide sequences were assembled to produce near-full length genomes of approximately 9 kb in size.

### Phylogenetic and recombination analysis

In order to investigate the phylogenetic relationship of the sequences, reference sequences relevant to HIV-1 epidemics in Southeast Asia, such as CRF01\_AE, subtype B of Western origin, subtype B' of Thai origin, CRF07\_BC, CRF08\_BC, CRF15\_01B, CRF33\_01B, CRF34\_01B, CRF48\_01B, CRF51\_01B, CRF53\_01B, CRF54\_01B and CRF58\_01B were downloaded from the Los Alamos National Laboratory HIV sequence database (<http://www.hiv.lanl.gov>) for phylogenetic analysis. References and query sequences (including near-full length and partial genomes) were codon-aligned in accordance to HIV Sequence Compendium 2013 (<http://www.hiv.lanl.gov>). Subsequently, bootscanning and informative sites estimation [12] using SimPlot version 3.5.1 [13] were carried out. HIV-1 CRF01\_AE (90THCM240) and subtype B' (93CNRL42) were used as putative parental genotypes, as well as subtype C (95IN21068) as an outlier group, to determine the recombination structures as well as the location of breakpoints in the near-full length and partial sequences. Genomic sequences with similar mosaic structures and recombination breakpoints were categorised as novel CRFs. On top of that, breakpoints for the previously reported CRF33\_01B and CRF53\_01B were also determined concurrently to ascertain the accuracy of the estimated breakpoints for the novel CRF. Next, sub-genomic region neighbor-joining trees were constructed using MEGA version 6, on Kimura two-parameter model with a transition-to-transversion ratio of 2.0 [14] to reaffirm the parental lineages of each sub-genomic segment. The reliability of the branching orders were analysed by bootstrap replicates of 1000. Maximum likelihood tree analyses were performed using PAUP version 4.0 [15].

### Bayesian coalescent analysis

In order to study the phylogenetic relationship between the novel CRF as well as the approximate time of emergence of the novel CRF and other published CRFs in Malaysia, Bayesian coalescent analysis using the Markov chain Monte Carlo (MCMC) sampling method were performed through BEAST version 1.7.5 [16] under the uncorrelated log-normal relaxed clock model with the general time-reversible (GTR) nucleotide substitution model with constant and exponential coalescence model. Each MCMC analysis was executed for 50 million states, sampled at every 50,000 states. The model with effective sampling size (ESS) >200 was selected. Posterior probabilities densities were determined in Tracer v1.5 (available in <http://tree.bio.ed.ac.uk/software/tracer/>) and 10% of each chain will be discarded as burn-in. All sequences studied have been deposited in GenBank, as listed in [Table 1](#).

**Table 1. Demographic characteristics of study subjects infected by HIV-1 CRF74\_01B.**

Sample ID	Gender	Risk group	Collection year	HXB2 location (nt)	Accession number
10MYKJ052	Male	IDU*	2010	643–9606	KR019770
10MYPR268	Male	IDU	2010	695–9606	KR019771
11MYPR416	Male	IDU	2011	682–9533	KR019772
10MYKJ016	Male	IDU	2010	733–4898 <sup>#</sup>	KR019773
11MYJJ741	Male	IDU	2011	708–4898 <sup>#</sup>	KR019774
11MYYC672	Male	IDU	2011	708–4898 <sup>#</sup>	KR019775

\* IDU, intravenous drug users

<sup>#</sup> Partial genome

doi:10.1371/journal.pone.0133883.t001

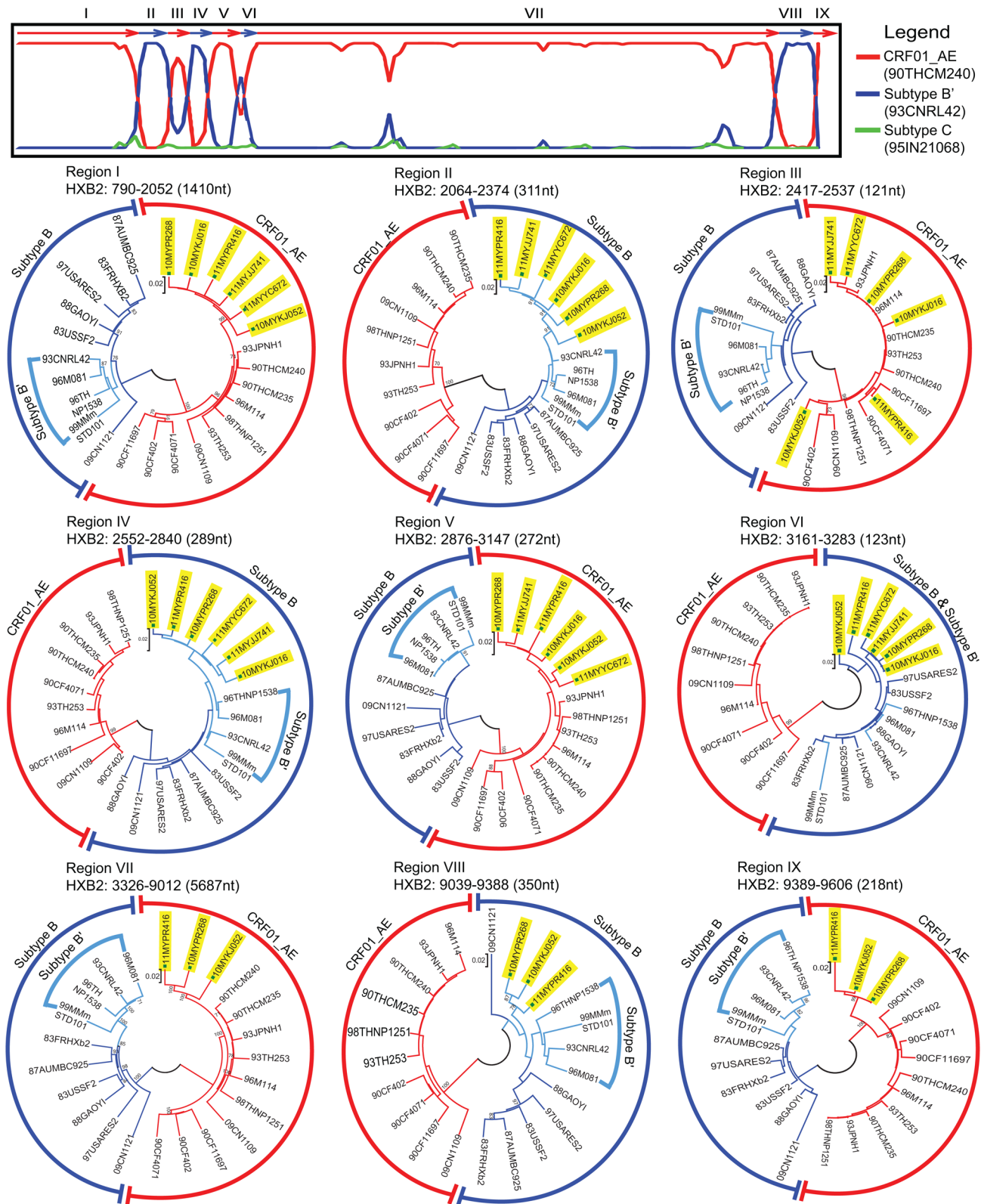
## Results and Discussion

### Characterization of novel HIV-1 CRF74\_01B isolated in Kuala Lumpur, Malaysia

A total of six isolates from epidemiologically-unlinked individuals formed a monophyletic cluster based on phylogenetic analysis of the PRRT sequences reported previously [11]. Among these six isolates, three near-full length and three partial genomes were successfully amplified and sequenced (Table 1). The near-full length genomes of isolates 10MYKJ052, 10MYPR268 and 11MYPR416 spanned the *gag*, *pol*, *env*, *tat*, *rev*, *vif*, *vpr*, *vpu* and *nef* genes. The partial genomes of isolates 10MYKJ016, 11MYYC672, 11MYJJ741 spanned the *gag* and *pol* region (HXB2: 790–4898) (attempts to amplify the near-full length genomes was unsuccessful due to limited specimens quantity and poor specimens quality). The recombination structures were determined through bootscanning and informative sites analysis. All three near-full length genomes have identical mosaic recombination structures, which are composed of nine sub-regions alternating between CRF01\_AE and subtype B'. Sub-region neighbour joining analyses also confirmed the putative parental subtype of each sub-region (Fig 1). A total of eight breakpoints were estimated at HXB2 positions 2053–2063, 2375–2416, 2538–2551, 2841–2875, 3148–3160, 3284–3325, 9013–9038 and 9388–9389. Specifically, five sub-regions of CRF01\_AE were identified by informative site analysis at HXB2 positions 643 to 2052, 2417 to 2537, 2876 to 3147, 3326 to 9012 and 9389 to 9606, while four sub-regions of subtype B' were observed in 2064 to 2374, 2552 to 2840, 3161 to 3283 and 9039 to 9388. However, we note that sub-region VI, which was only 123bp in size (HXB2 position: 3161 to 3283), did not provide sufficient phylogenetic signal for definitive distinction of subtype B (Western origin) and subtype B' (Thai origin). In particular, the first four recombination breakpoints were similar to those in CRF33\_01B, a CRFs previously identified in Malaysia. Similarly, the first two breakpoints were similar to CRF53\_01B [8] (Fig 2). Recombination and sub-region tree analysis of the three partial genomes (isolates 10MYKJ016, 11MYYC672, 11MYJJ741) also showed identical mosaic structures and shared six recombination breakpoints in the *gag-pol* region (HXB2: 790–4898) with the near-full length genomes (10MYKJ052, 10MYPR268 and 11MYPR416) (Fig 1). Recombinant structures of these strains are different from any known CRFs reported to date. Taken together, the novel recombination lineage identified in this report was designated CRF74\_01B by the HIV Los Alamos National Laboratory according to the standardized nomenclature proposal [1].

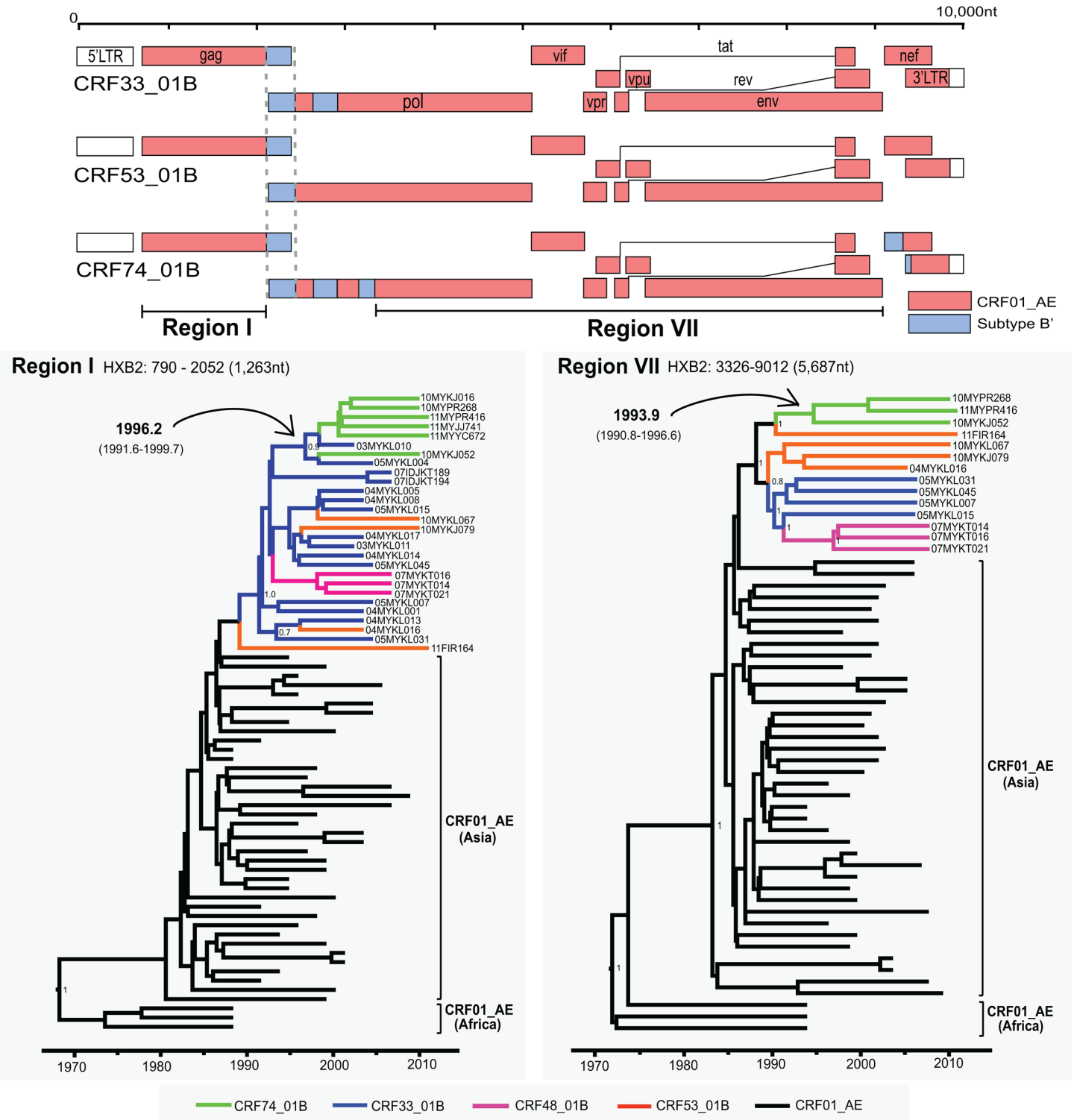
### Phylogenetic analysis of CRF74\_01B

As sharing of breakpoints typically suggest a common ancestry between recombinants [3, 17], it is possible that CRF74\_01B could be a descendent of CRF33\_01B and CRF53\_01B, or vice



**Fig 1. Recombination and neighbour joining analysis of HIV-1 CRF74\_01B.** The mosaic structure of CRF74\_01B was depicted through bootscan using HIV-1 CRF01\_AE (90THCM240) and subtype B' (93CNRL42) as putative parental subtypes, and subtype C (95IN21068) as outlier strain. Subtype assignment for sub-genomic regions of CRF74\_01B (regions I to IX) were determined and depicted using neighbour joining analysis. Red branches indicate CRF01\_AE while blue branches indicate subtype B and B' reference strains. CRF74\_01B isolates were indicated (yellow shades).

doi:10.1371/journal.pone.0133883.g001

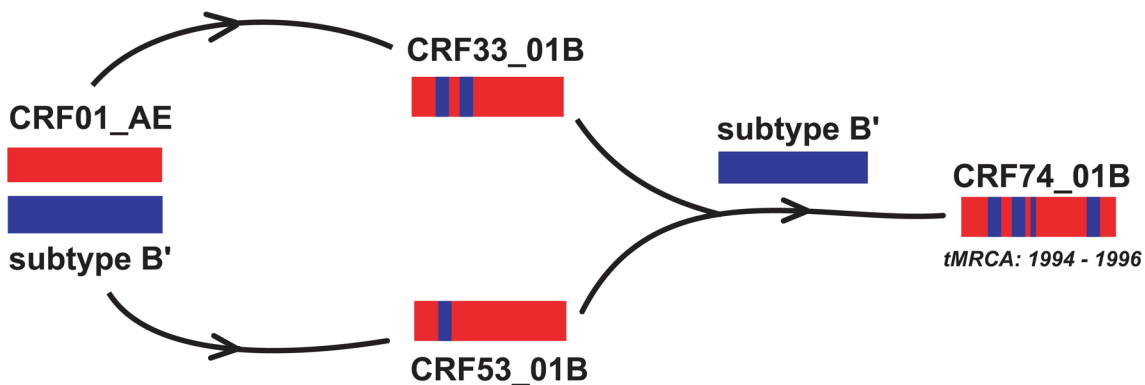


**Fig 2. Recombination structures and sub-genomic maximum clade credibility tree reconstructions.** Recombination structures of HIV-1 CRF33\_01B, CRF53\_01B and CRF74\_01B were shown. Shared recombination breakpoints between these CRFs were indicated by dashed lines. Maximum clade credibility trees for the *gag* region (region I; HXB2: 790–2052) and *pol-env* region (region VII; HXB2: 3326–9012) of CRF74\_01B and other relevant reference sequences were reconstructed. The means of tMRC of CRF74\_01B for region I and region VII were estimate around 1996.2 and 1993.9, respectively, with the 95% credibility intervals stated in parentheses.

doi:10.1371/journal.pone.0133883.g002

versa. The former hypothesis is highly probable, as CRF33\_01B has gradually dispersed and established among the general population since its first report in 2006 [11], therefore having a higher chance to co-circulate with other genotypes and subsequently recombine to generate novel CRF. In order to investigate the phylogenetic relationship of CRF74\_01B with other structurally related recombinant lineages, the longest shared genetic regions (non-recombinant genetic fragment) with sufficient phylogenetic signal were identified among CRF74\_01B, CRF33\_01B and CRF53\_01B. Regions corresponding to HXB2 positions 790–2052 (similar to region I in Fig 1) and 3326–9012 (region VII in Fig 1) were selected for maximum clade credibility (MCC) tree reconstructions in BEAST. MCC tree reconstructions under the constant and exponential demographic models produced similar tree topologies, thus only the constant tree model is shown (Fig 2).

The topologies of the MCC tree showed that CRF74\_01B branched from an internal node along with CRF33\_01B in region I (Fig 2), which could indicate a common ancestry between these two CRFs; though one particular strain of CRF74\_01B, 10MYKL052, is paraphyletic in relation to other CRF74\_01B strains. On the other hand, CRF74\_01B appears to form a monophyletic cluster with CRF53\_01B in the longer sub-genomic region VII (Fig 2). Therefore, it is likely that CRF74\_01B is a second-generation progeny descended from specific lineages of CRF33\_01B and CRF53\_01B. Notably, CRF33\_01B and CRF53\_01B appear to have emerged from a common ancestor, as both group branches from one common internal node (Fig 2). On the other hand, CRF48\_01B that was reported to be descended from CRF33\_01B [6], showed no direct relationship to CRF74\_01B and is clustered independently within CRF33\_01B. Similar observations can be derived from maximum likelihood analyses (S1 Fig). Based on the Bayesian MCMC analysis in this study, the evolutionary rates for region I and region VII were estimated at  $2.3 \times 10^{-3}$  (95% highest posterior density (HPD):  $1.8 \times 10^{-3}$ – $2.8 \times 10^{-3}$ ) and  $2.8 \times 10^{-3}$  (95% HPD:  $2.4 \times 10^{-3}$ – $3.2 \times 10^{-3}$ ), substitutions per site per year, respectively, which are in concordance with substitution rates estimated elsewhere [18, 19]. Following the MCC trees depicted in Fig 2, the estimated time for the most recent common ancestor (tMRCA) for CRF74\_01B placed its emergence at approximately between 1994 and 1996. This outcome supports that notion that CRF74\_01B emerged a few years after the emergence of its putative parental strain CRF33\_01B (Fig 3), which is proposed to have emerged between 1991 and 1993 [18].



**Fig 3. Plausible genetic recombination events leading to the generation of HIV-1 CRF74\_01B.** Using information on the recombination structures and phylogenetic inference of CRF74\_01B and other genetically related CRFs as possible parental lineages, a chronology of recombination events leading to the emergence of CRF74\_01B was proposed. The diagram highlights CRF74\_01B as a second-generation recombinant with an estimated tMRCA between 1994 and 1996.

doi:10.1371/journal.pone.0133883.g003

However, certain limitations that may affect the tMRCA estimation and interpretation in the study should be addressed. For one, the genetic data used for the analyses were limited, as few near-full length genomes of CRF53\_01B and CRF74\_01B are available due to their low prevalence in the country [20]. Even for an established recombinant such as CRF33\_01B, which has an increasingly extensive transmission network in Malaysia [5, 11], only a few near-full length genomes are readily available for comprehensive phylogenetic analyses. Secondly, monophyletic rules [21] suggested that the discrepancies between the two MCC tree topologies could be a result of multiple recombination processes, which could complicate phylogenetic analysis due to limitations in the current theoretical and computational settings [22]. Therefore, until more isolates are detected and sequenced, the intricate relationship between these CRFs must be interpreted with caution, although phylogenetic evidence suggests the ancestral roles of both CRF33\_01B and CRF53\_01B (Figs 2 and 3). In short, CRF74\_01B is probably a second-generation recombinant lineage generated via yet undefined recombination event(s) that involved CRF33\_01B, CRF53\_01B and subtype B'.

The emergence of CRF74\_01B further highlights the continuous inter-genotypic recombination processes between the various genotypes and recombinants that are circulating among the various risk groups in Malaysia. It also prompted the necessity of vigilant surveillance on the genetic diversity of HIV-1 epidemics in Malaysia, since it has increased tremendously with the description of the CRF33\_01B, CRF48\_01B, CRF52\_01B, CRF53\_01B, CRF54\_01B, CRF58\_01B and the latest addition of CRF74\_01B.

## Conclusion

We identified a novel CRF74\_01B among IDUs in Malaysia. Thus far, this is the sixth CRF found in the country. Bayesian MCMC analysis showed that CRF74\_01B could be a second-generation progeny lineage from the recombination between CRF33\_01B, CRF53\_01B and subtype B'. Hence, continuous molecular screening and epidemic surveillance should be implemented in order to comprehend the evolutionary dynamics of the HIV-1 epidemics, which could be crucial in the design of effective vaccines and intervention measures.

## Supporting Information

**S1 Fig. Maximum likelihood analyses for sub-genomic region I and region VII.**  
(PDF)

**S1 Table. List of HIV-1 primers for near-full length genome amplification.**  
(DOCX)

## Acknowledgments

We would like to thank the study subjects for their participation in this study. We would also like to acknowledge Lai Yee Ong, Siti Humaira Razak, Yeat Mei Lee, Azureen Azmel and Kim Tien Ng for assistance and support.

## Author Contributions

Conceived and designed the experiments: HTC KKT. Performed the experiments: HTC WZC KKT. Analyzed the data: HTC WZC YT KKT. Contributed reagents/materials/analysis tools: HTC WZC YT JBC KGC HAAA CK AK KKT. Wrote the paper: HTC KKT.



## References

1. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science*. 2000; 288(5463): 55–56. PMID: [10766634](#)
2. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS*. 2011; 25(5): 679–689. doi: [10.1097/QAD.0b013e328342ff93](#) PMID: [21297424](#)
3. Kijak GH, Tovanabutra S, Rerks-Ngarm S, Nitayaphan S, Eamsila C, Kunasol P, et al. Molecular evolution of the HIV-1 Thai epidemic between the time of RV144 immunogen selection to the execution of the vaccine efficacy trial. *J Virol*. 2013; 87(13): 7265–7281. doi: [10.1128/JVI.03070-12](#) PMID: [23576510](#)
4. Chen JH, Wong KH, Chen Z, Chan K, Lam HY, To SW, et al. Increased genetic diversity of HIV-1 circulating in Hong Kong. *PLoS One*. 2010; 5(8): e12198. doi: [10.1371/journal.pone.0012198](#) PMID: [20808942](#)
5. Tee KK, Li XJ, Nohtomi K, Ng KP, Kamarulzaman A, Takebe Y. Identification of a novel circulating recombinant form (CRF33\_01B) disseminating widely among various risk populations in Kuala Lumpur, Malaysia. *J Acquir Immune Defic Syndr*. 2006; 43(5): 523–529. PMID: [17031320](#)
6. Li Y, Tee KK, Liao H, Hase S, Uenishi R, Li XJ, et al. Identification of a novel second-generation circulating recombinant form (CRF48\_01B) in Malaysia: a descendant of the previously identified CRF33\_01B. *J Acquir Immune Defic Syndr*. 2010; 54(2): 129–136. doi: [10.1097/QAI.0b013e3181d82ce5](#) PMID: [20386110](#)
7. Liu Y, Li L, Bao Z, Li H, Zhuang D, Liu S, et al. Identification of a novel HIV type 1 circulating recombinant form (CRF52\_01B) in Southeast Asia. *AIDS Res Hum Retroviruses*. 2012; 28(10): 1357–1361. PMID: [22269007](#)
8. Chow WZ, Al-Darraj H, Lee YM, Takebe Y, Kamarulzaman A, Tee KK. Genome sequences of a novel HIV-1 CRF53\_01B identified in Malaysia. *J Virol*. 2012; 86(20): 11398–11399. PMID: [22997419](#)
9. Ng KT, Ong LY, Takebe Y, Kamarulzaman A, Tee KK. Genome sequence of a novel HIV-1 circulating recombinant form 54\_01B from Malaysia. *J Virol*. 2012; 86(20): 11405–11406. PMID: [22997423](#)
10. Chow WZ, Takebe Y, Syafina NE, Prakasa MS, Chan KG, Al-Darraj HA, et al. A newly emerging HIV-1 recombinant lineage (CRF58\_01B) disseminating among people who inject drugs in Malaysia. *PLoS One*. 2014; 9(1): e85250. doi: [10.1371/journal.pone.0085250](#) PMID: [24465513](#)
11. Chow WZ, Ong LY, Razak SH, Lee YM, Ng KT, Yong YK, et al. Molecular diversity of HIV-1 among people who inject drugs in Kuala Lumpur, Malaysia: massive expansion of circulating recombinant form (CRF) 33\_01B and emergence of multiple unique recombinant clusters. *PLoS One*. 2013; 8(5): e62560. doi: [10.1371/journal.pone.0062560](#) PMID: [23667490](#)
12. Salminen MO, Carr JK, Burke DS, McCutchan FE. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses*. 1995; 11(11): 1423–1425. PMID: [8573403](#)
13. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol*. 1999; 73(1): 152–160. PMID: [9847317](#)
14. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30(12): 2725–2729. doi: [10.1093/molbev/mst197](#) PMID: [24132122](#)
15. Swofford DL. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. 2003.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29(8): 1969–1973. doi: [10.1093/molbev/mss075](#) PMID: [22367748](#)
17. McClutchan FE, Carr JK, Murphy D, Piyasirisilp S, Gao F, Hahn B, et al. Precise mapping of recombination breakpoints suggests a common parent of two BC recombinant HIV type 1 strains circulating in China. *AIDS Res Hum Retroviruses*. 2002; 18(15): 1135–1140. PMID: [12402948](#)
18. Tee KK, Pybus OG, Parker J, Ng KP, Kamarulzaman A, Takebe Y. Estimating the date of origin of an HIV-1 circulating recombinant form. *Virology*. 2009; 387(1): 229–234. doi: [10.1016/j.virol.2009.02.020](#) PMID: [19272628](#)
19. Lemey P, Rambaut A, Pybus, OG. HIV evolutionary dynamics within and among hosts. *AIDS Reviews*. 2006; 8(3): 1–16.
20. Cheong HT, Ng KT, Ong LY, Chook JB, Chan KG, Takebe Y, et al. Cross-border sexual transmission of the newly emerging HIV-1 clade CRF51\_01B. *PLoS One*. 2014; 9(10): e111236. doi: [10.1371/journal.pone.0111236](#) PMID: [25340817](#)

21. Abecasis AB, Lemey P, Vidal N, de Oliveira T, Peeters M, Camacho R, et al. Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form. *J Virol*. 2007; 81(16): 8543–8551. PMID: [17553886](#)
22. Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. Eight challenges in phylodynamic inference. *Epidemics*. 2015; 10(2015): 88–92.