

# Identification of a Novel HIV Type 1 Circulating Recombinant Form (CRF52\_01B) in Southeast Asia

Yongjian Liu,<sup>1</sup> Lin Li,<sup>1</sup> Zuoyi Bao,<sup>1</sup> Hanping Li,<sup>1</sup> Daomin Zhuang,<sup>1</sup> Siyang Liu,<sup>1</sup> Xiaolin Wang,<sup>1</sup> Tianyi Li,<sup>1</sup> Lei Jia,<sup>1</sup> Shaomin Yang,<sup>2</sup> and Jingyun Li<sup>1</sup>

## Abstract

Thirty HIV-1 URF\_01AE/ B' complete or nearly full-length genome sequences sampled within Southeast Asia were obtained from the Los Alamos HIV Sequence Database. Phylogenetic and recombinant analyses revealed that three sequences indeed displayed the identical recombinant structure. Of note, the three subjects, harboring novel CRF01\_AE/B recombinants, did not have apparent epidemiological linkage. They fulfilled the criteria for the designation of a new circulating recombinant form (CRF) and constituted the 52nd CRF identified in the worldwide HIV-1 pandemic. In this chimera, two short subtype B segments were inserted into a backbone of CRF\_01AE. The breakpoints corresponded to HXB2 nucleotide positions 2930, 3251, 8521, and 9004 approximately. This CRF is the first one identified by neatening and analyzing the sequences already presented in the Los Alamos HIV Sequence Database. This indicates that we should pay attention not only to explicit subtype sequences but also to those classified as a unique recombinant form (URF) so far.

**M**UTATION, COUPLED WITH RECOMBINATION, has resulted in the extensive genetic diversity of HIV-1. The high genetic diversity of HIV-1 allows its classification into groups, subtypes, subsubtypes, and circulating recombinant forms (CRFs).<sup>1</sup> Group M HIV-1 dominates the AIDS pandemic with at least nine subtypes (A to D, F to H, J, and K) and multiple intersubtype recombinants. Currently, a total of 51 CRFs have been represented in the Los Alamos HIV database web site (<http://hiv-web.lanl.gov/CRFs/CRFs.html>). Meanwhile multiple unique recombinant forms (URFs) have been reported in many regions around the world.

Southeast Asia was thought to be an epicenter of the HIV-1 epidemic in Asia. The AIDS epidemic in this region was initiated by two different subtype strains in two separated risk groups, with subtype B among injecting drug users (IDUs) and CRF01\_AE among those heterosexually exposed, throughout the early phase.<sup>2,3</sup> CRF01\_AE can clearly cross over from the heterosexual epidemic to IDUs and by 1995 it also became the predominant subtype in IDUs.<sup>4</sup> The co-circulation of the two different strains in this region led to the generation of more complex recombinant strains that emerged in almost every country in Southeast Asia.<sup>5-10</sup> Furthermore, some of the recombinant strains have spread widely in populations and have become CRFs. Up to now,

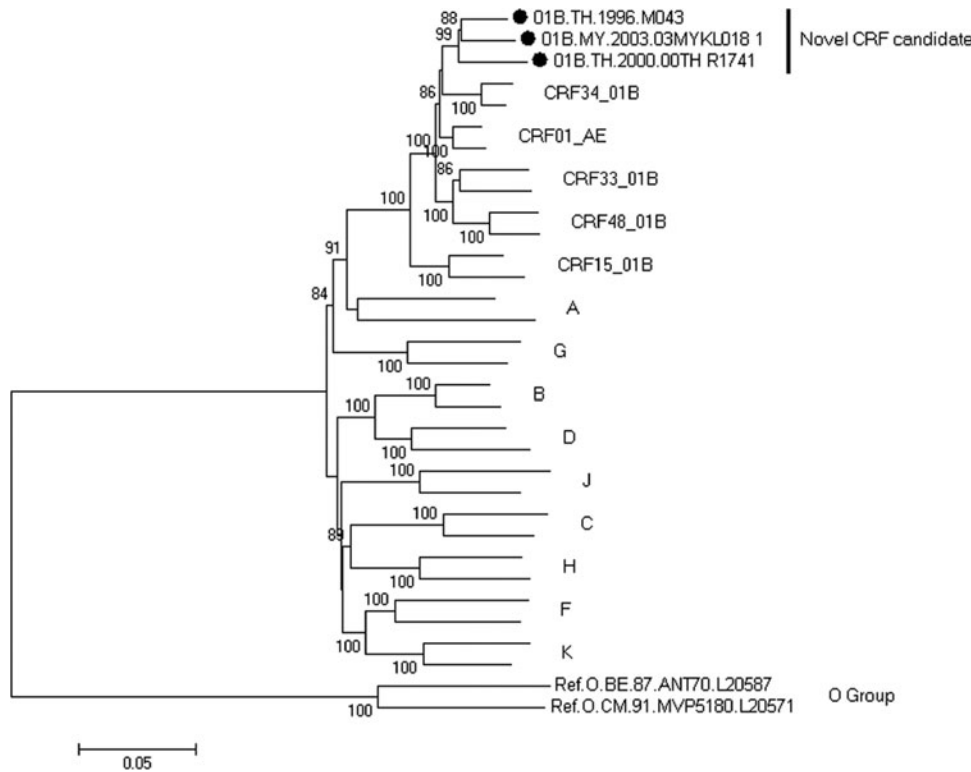
five CRFs, originated from CRF01\_AE and subtype B' lineages, have been reported in Southeast Asia: CRF15\_01B<sup>11</sup> and CRF34\_01B<sup>12</sup> from Thailand, CRF33\_01B<sup>13</sup> and CRF48\_01B<sup>14</sup> from Malaysia, and CRF51\_01B<sup>15</sup> from Singapore. The HIV-1 epidemic in this region has now become increasingly heterogeneous.

The extensive use of complete and nearly full-length genome sequencing of HIV-1 strains has provided a powerful and accurate approach to the molecular epidemiology of regional epidemics. Consequently, the number of complete or nearly full-length HIV-1 genome sequences was growing larger and larger. The majority of these can be classified as subtypes and CRFs definitely, while the minority remains as URFs. To define a CRF, at least three epidemiologically unlinked HIV-1 sequences with identical mosaic structures should be characterized, at least two of them in the nearly full-length genome (8 kb)<sup>1</sup>. For the present study, we identified a novel CRF\_AE/B' candidate, which emerged in Southeast Asia, using sequences obtained from the Los Alamos HIV Sequence Database ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)).

Thirty HIV-1 URF\_01AE/ B' complete or nearly full-length genome sequences sampled within Southeast Asia were obtained from the Los Alamos HIV Sequence Database. These sequences were generated by sequencing from plasma or

<sup>1</sup>Department of AIDS Research, State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China.

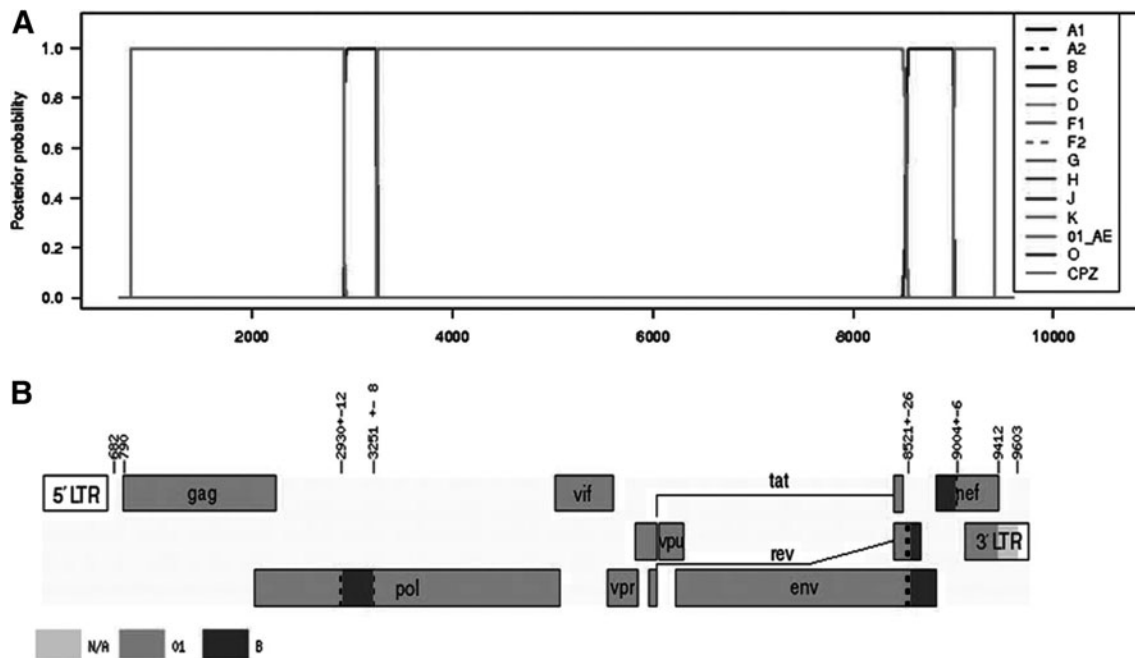
<sup>2</sup>Yunnan Provincial Hospital Infectious Disease, AIDS Care Center (YNACC), Kunming, Yunnan, China.



**FIG. 1.** Identification of a novel circulating recombinant form (CRF) candidate. The neighbor-joining phylogenetic tree is constructed based on the nucleotide sequence of nearly full-length genomes using MEGA 5.04. The tree is rooted with group O. All the reference strains were retrieved from the Los Alamos HIV Sequence Database. Values on the branches represent the percentage of 1000 bootstrap replicates and bootstrap values over 70% are shown in the tree. The novel CRF candidate is marked with a solid circle. The scale bar indicates 5% nucleotide sequence divergence.

peripheral blood mononuclear cell (PBMCs) samples collected from Thailand ( $n=19$ ), Malaysia ( $n=8$ ), Myanmar ( $n=2$ ), and China ( $n=1$ ). In addition to these 30 sequences, we also retrieved subtype reference sequences from the same database. All the sequences were incorporated into one se-

quence dataset using SynchAlign, software [http://www.hiv.lanl.gov/content/sequence/SYNCH\\_ALIGN/SynchAligns.html](http://www.hiv.lanl.gov/content/sequence/SYNCH_ALIGN/SynchAligns.html), and edited using BioEdit software. A phylogenetic tree was constructed by the neighbor-joining method based on Kimura's two-parameter distance matrix with 1000



**FIG. 2.** Recombinant analysis of the novel identified CRF52\_01B. (A) Posterior probabilities of the subtypes at each sequence position (original sequence positions) calculated by jpHMM. Inset: The different symbols represent different reference strains. (B) Genome map (based on HXB2 numbering). Numbers in the figure above denote intervals for recombination breakpoints based on HXB2 numbering. The white regions denote missing information due to input fragment sequences. Subtype N/A denotes missing information due to uninformative subtype models (subtype: N/A).

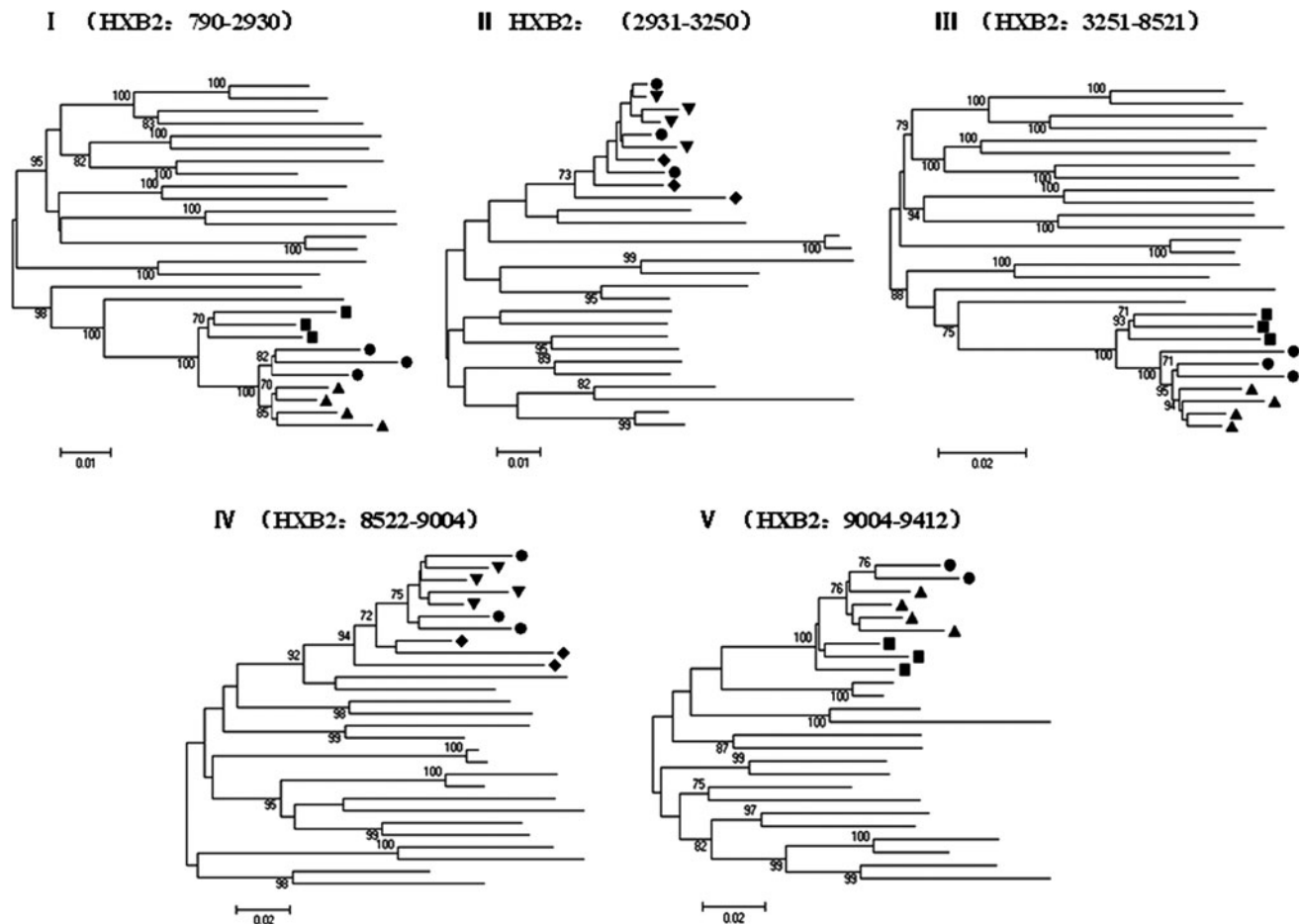
bootstrap replicates using Mega 5.04. With respect to recombination analysis, RIP, <http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>, software was applied at first; jpHMM software (<http://jphmm.gobics.de/>) was used to further scrutinize and define the recombination breakpoints subsequently. The origin of each segment was analyzed by subregion neighbor-joining tree analysis. Detailed information about the three subjects harboring CRF52\_01B recombinants was collected from the studies<sup>13,16,17</sup> in which the three subjects were reported. Furthermore, we retrieved the records for the three subjects from the Los Alamos HIV Sequence Database in order to supplement elements such as patient name and risk factor.

Using phylogenetic analysis, we found that three nearly full-length genome sequences from three patients formed a monophyletic branch supported by a high bootstrap value of 99%, located outside any HIV-1 subtypes and known CRFs in Southeast Asia (Fig. 1). Both RIP and jpHMM revealed that the three URFs nearly full-length genome sequences indeed displayed the identical recombinant structure composed of CRF\_01AE and subtype B. In this chimera, two short subtype B segments were inserted into a backbone of CRF\_01AE. The

breakpoints corresponded to HXB2 nucleotide positions 2930, 3251, 8521, and 9004 approximately (Fig. 2).

To confirm the subtype structure of CRF52\_01B and to estimate likely parental lineages, we divided the HIV-1 genomes into five segments according to the breakpoints. Subregion phylogenetic analysis confirmed the four breakpoints identified using jpHMM. The genomic segments I, III, and V clustered with reference subtype CRF01\_AE while the other two segments II and IV branched with reference subtype B (Fig. 3). As shown in Fig. 3, the subregion tree analysis demonstrated that CRF01\_AE segments (I, III, and V) belonged to the cluster of Thailand CRF01\_AE other than African CRF01\_AE, showing that the three CRF01\_AE segments indeed originated from CRF01\_AE of Thailand origin. Similarly, subtype B segment IV analysis revealed that it originated from Thailand B rather than Western B. However, we could not determine the origin of the remaining subtype B segment II because the length was too short to obtain high bootstrap values.

The profiles of the subtype structure of the three sequences were distinct from the other five CRFs originated from CRF01\_AE and subtype B' reported in Southeast Asia.



**FIG. 3.** Phylogenetic trees for genome segments of the novel identified CRF52\_01B. The neighbor-joining tree is constructed using MEGA 5.04. Values at the nodes indicate the percent bootstraps in which the cluster to the right was supported. Bootstraps of 70% and higher only are shown. The scale bars are shown at the bottom. ● = CRF52\_01B; ▲ = Thailand CRF01\_AE; ■ = African CRF01\_AE; ▼ = Thailand B; □ = Western B.

TABLE 1. INFORMATION ON SUBJECTS HARBORING NOVEL CRF01\_AE/B RECOMBINANTS IN SOUTHEAST ASIA

Patient	Age	Gender	Country	Year of first positive HIV-1 test	Year of collection	Risk factor	Sequence name	Accession number
M043	n/a	Female	Thailand	n/a	1996	Heterosexual	01B.TH.1996.M043	DQ354113
MYKL018	32	Male	Malaysia	1997	2003	Bisexual	01B.MY.2003.03MYKL018_1	DQ366664
R1741	27	Female	Thailand	1999	2000	Heterosexual	01B.TH.2000.00TH_R1741	AY945734

n/a, not available.

Of note, the three nearly full-length genome sequences (01B.MY.2003.03MYKL018\_1, 01B.TH.2000.00TH\_R1741, and 01B.TH.1996.M043) were presented in three different studies<sup>13,16,17</sup> from three different patients (Table 1) with no apparent epidemiological linkage. These data established that the three sequences show the same recombinant structure over the whole genome, fulfilling the criteria for designation of a new CRF. They constitute the 52nd CRF identified in the worldwide HIV-1 pandemic, herein called CRF52\_01B.

Detailed information about the three subjects harboring CRF52\_01B recombinants is shown in Table 1. Unlike other CRFs identified worldwide, the most interesting element is that the three subjects belonged to two different countries, with two in Thailand and one in Malaysia, within Southeast Asia. However, the subjects were reported by year of first positive HIV-1 test, although that of subject M043 was not available. From Table 1, it is apparently that subject M043 was infected in or before 1996. The obviously wide geographic separation and early infection date of the three subjects suggested that CRF52\_01B arose at least 15 years ago. To obtain accurate estimates regarding the time of origin and transmission direction of the novel CRF in Southeast Asia, more sequences with different locations and sampling times are needed in Bayesian coalescent analysis using BEAST software.<sup>18,19</sup>

All of the three subjects reported sexual exposure, with two being heterosexual and one bisexual. It has been reported that most women infected in Southeast Asia have been the monogamous wives or regular partners of higher risk men.<sup>20</sup> Of interest, subject R1741 was one of the participants recruited from public family planning clinics in Thailand. Her only reported risk factor for HIV-1 infection was heterosexual exposure.<sup>16</sup> Taking subject R1741 into account, she should have acquired the HIV-1 infection from her male sexual partner. It was suggested that the novel CRFs have spread into low-risk people early.

In this study, we identified a novel CRF (CRF52\_01B) composed of CRF01\_AE and subtype B in Southeast Asia by analyzing the 30 complete or nearly full-length genome sequences obtained from the Los Alamos HIV Sequence Database. CRF52\_01B is the sixth identified CRF that has originated from CRF01\_AE and subtype B lineages within Southeast Asia in the past several years. What is most interesting is that this CRF is the first one identified by neatening and analyzing the sequences already presented in the Los Alamos HIV Sequence Database. This emphasizes that we should pay attention not only to explicit subtype sequences but also to those classified as URFs so far. The prevalence and public significance of the novel CRF remain to be elucidated in the future.

## Acknowledgments

This work was supported by the National Key S&T Special Projects on Major Infectious Diseases (Grants 2008ZX10001-004 and 2008ZX10001-012).

## Author Disclosure Statement

No competing financial interests exist.

## References

- Robertson DL, Anderson JP, Bradac JA, *et al.*: HIV-1 nomenclature proposal. *Science* 2000;288(5463):55–57.
- Ou CY, Weniger BG, Luo CC, *et al.*: Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* 1993;341(8854):1171–1174.
- Kalish ML, Baldwin A, Raktham S, *et al.*: The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: Implications for HIV vaccine trials. *AIDS* 1995;9(8):851–858.
- Subbarao S, Vanichseni S, Hu DJ, *et al.*: Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand. *AIDS Res Hum Retroviruses* 2000;16(8):699–707.
- Tovanabutra S, Polonis V, De Souza M, *et al.*: First CRF01\_AE/B recombinant of HIV-1 is found in Thailand. *AIDS* 2001;15(8):1063–1065.
- Yang R, Xia X, Kusagawa S, Zhang C, Ben K, and Takebe Y: On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China. *AIDS* 2002;16(10):1401–1407.
- Ramos A, Nguyen L, Hu DJ, *et al.*: New HIV type 1 CRF01\_AE/B recombinants displaying unique distribution of breakpoints from incident infections among injecting drug users in Thailand. *AIDS Res Hum Retroviruses* 2003;19(8):667–674.
- Takebe Y, Motomura K, Tatsumi M, Lwin HH, Zaw M, and Kusagawa S: High prevalence of diverse forms of HIV-1 intersubtype recombinants in Central Myanmar: Geographical hot spot of extensive recombination. *AIDS* 2003;17(14):2077–2087.
- Tee KK, Pon CK, Kamarulzaman A, and Ng KP: Emergence of HIV-1 CRF01\_AE/B unique recombinant forms in Kuala Lumpur, Malaysia. *AIDS* 2005;19(2):119–126.
- Wang B, Lau KA, Ong L-Y, *et al.*: Complex patterns of the HIV-1 epidemic in Kuala Lumpur, Malaysia: Evidence for expansion of circulating recombinant form CRF33\_01B and detection of multiple other recombinants. *Virology* 2007;367(2):288–297.
- Tovanabutra S, Watanaveeradej V, Viputtikul K, *et al.*: A new circulating recombinant form, CRF15\_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand. *AIDS Res Hum Retroviruses* 2003;19(7):561–567.

12. Tovanabutra S, Kijak GH, Beyrer C, *et al.*: Identification of CRF34\_01B, a second circulating recombinant form unrelated to and more complex than CRF15\_01B, among injecting drug users in northern Thailand. *AIDS Res Hum Retroviruses* 2007;23(6):829–833.
13. Tee KK, Li X-J, Nohtomi K, Ng KP, Kamarulzaman A, and 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.
14. Li Y, Tee KK, Liao H, *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.
15. Ng OT, Eyzaguirre LM, Carr JK, *et al.*: Identification of new CRF51\_01B in Singapore using full genome analysis of three HIV type 1 isolates. *AIDS Res Hum Retroviruses* 2012 [Epub ahead of print]; DOI: 10/1089/aid.2011.0177.
16. Watanaveeradej V, Benenson MW, Souza MD, *et al.*: Molecular epidemiology of HIV Type 1 in preparation for a phase III prime-boost vaccine trial in Thailand and a new approach to HIV Type 1 genotyping. *AIDS Res Hum Retroviruses* 2006;22(8):801–807.
17. Kijak GH, Tovanabutra S, Sanders-Buell E, *et al.*: Distinguishing molecular forms of HIV-1 in Asia with a high-throughput, fluorescent genotyping assay, MHA bce v.2. *Virology* 2007;358(1):178–191.
18. Drummond A, Rambaut A, Shapiro B, and Pybus O: Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005;22:1185–1192.
19. Drummond A and Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7(1):214.
20. Xu F, Kilmarx PH, Supawitkul S, *et al.*: HIV-1 seroprevalence, risk factors, and preventive behaviors among women in Northern Thailand. *J Acquir Immune Defic Syndr* 2000;25(4):353–359.

Address correspondence to:

Jingyun Li  
Department of HIV/AIDS Research  
State Key Laboratory of Pathogen and Biosecurity  
Beijing Institute of Microbiology and Epidemiology  
20 Dongda Street  
Fengtai District  
Beijing 100071  
China

E-mail: lijy@nic.bmi.ac.cn