

The Indonesian Variants of CRF33_01B: Near-Full Length Sequence Analysis

Ivo N. SahBandar,^{1,2} Kiyomi Takahashi,¹ Kazushi Motomura,³ Zubairi Djoerban,² Iman Firmansyah,⁴
Katsuhiko Kitamura,⁵ Hironori Sato,³ Herdiman T. Pohan,² Shigehiro Sato¹

Abstract

Cocirculation of subtype B and CRF01_AE in Southeast Asia has led to the establishment of new recombinant forms. In our previous study, we found five samples suspected of being recombinants between subtype B and CRF01_AE, and here, we analyzed near full-length sequences of two samples and compared them to known CRFs_01B, subtype B, and CRF01_AE. Five overlapped segments were amplified with nested PCR from PBMC DNA, sequenced, and analyzed for genome mosaicism. The two Indonesian samples, 07IDJKT189 and 07IDJKT194, showed genome-mosaic patterns similar to CRF33_01B references from Malaysia, with one short segment in the 3' end of the p31 integrase-coding region, which was rather more similar to subtype B than CRF01_AE, consisting of unclassified sequences. These results suggest gene-specific continuous diversification and spread of the CRF33_01B genomes in Southeast Asia.

AS A RETROVIRUS, HIV-1 is prone to natural error of viral reverse transcriptase. In combination with a high rate of virion production in infected individuals and genome recombination during genome synthesis, HIV-1 exhibits an extensive genetic diversity in nature.^{1,2} Cocirculation of two or more different subtypes and/or circulating recombinant forms (CRFs) in HIV-1 high-risk populations can lead to HIV-1 intersubtype recombination.³⁻⁵ To date, 9 different subtypes and 43 CRFs have been reported.⁶ In addition to the reported CRFs, numerous unique recombinant forms (URFs) were also identified worldwide, including Southeast Asia,^{5,7,8} where cocirculation of subtype B and CRF01_AE as the main circulating subtypes resulted in the generation of CRFs, such as CRF15_01B,⁹ CRF33_01B,¹⁰ and CRF34_01B.¹¹ These CRFs originated and mostly circulated in their countries of origin.⁹⁻¹¹

Our previous study showed that CRF01_AE was the main circulating subtype in Indonesia, followed by small number of subtype B. Furthermore, we also found five samples suspected of being CRF_01B URFs.¹² In this current study we analyzed near full-length sequences of two samples showing recombination patterns similar to CRF33_01B previously reported in Malaysia. These samples were obtained during a cross-sectional study of molecular epidemiology in 2007, from the Sulianti Saroso Infectious Diseases Hospital in Northern

Jakarta, Indonesia. 07IDJKT189 is a 25-year-old male with numerous risk factors; the subject was an injecting drug user (IDU) who also had a history of sexual contact with commercial sex workers whose HIV status was unknown. The patient presented with opportunistic infections such as lung tuberculosis, oral candidiasis, and cytomegalovirus (CMV) retinitis, which led to the diagnosis of HIV. The CD4 count was 17 cells/ μ l at admission. 07IDJKT194 is a 21-year-old female with heterosexual risk factors. She was screened for HIV infection due to her spouse's HIV status. Her CD4 count was 360 cells/ μ l at admission. Both subjects were antiretroviral (ARV) naive, without any known relationship to each other (Table 1).

Blood samples were collected during enrollment; peripheral blood mononuclear cells (PBMCs) were separated, and DNA was extracted as described previously.¹² First round amplifications of the 9.0-kbp genome cDNAs were performed by long range polymerase chain reaction (PCR) amplification using the Takara ExTaq Hot Start Version (Takara Bio Inc, Otsu, Japan) with the following primer sets: 172A (sense, 5'-ATCTCTAGCAGTGGCGCCCGAACAG-3') and 9KU5B (antisense, 5'-GGTCTGAGGGATCTCTAGTTACCAG-3'), followed by nested PCR of five overlapping genome segments from 1 kbp to 2.3 kbp using the primers listed in Table 2. The nested PCR

¹Department of Microbiology, School of Medicine, Iwate Medical University, Morioka, Japan.

²Department of Internal Medicine, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia.

³Center for Genomic Pathogen, National Institute of Infectious Diseases, Tokyo, Japan.

⁴Sulianti Saroso Infectious Disease Hospital, Jakarta, Indonesia.

⁵Yokohama City University School of Medicine, Yokohama, Japan.

TABLE 1. SUBJECTS' CLINICAL CHARACTERISTICS

Characteristic	07IDJKT189	07IDJKT194
Gender	Male	Female
Age (years)	25	21
Risk factor	IDU and heterosex ^a	Heterosex
Initiation of risky behavior	2004	2006
Ethnicity	Javanese	Batavian
Education	High School	High School
Occupation	Employee	Housewife
Marital status	Not married	Married
Sex partner HIV status	Unknown	Positive
Treatment status	Naive	Naive
CD4 count (cells/ μ l)	17	360
Disease stage	AIDS	HIV
Opportunistic infection	Present ^b	Absent
Hepatitis B infection	Positive ^c	Unknown
Hepatitis C infection	Negative	Unknown

^aIDU, injecting drug use; heterosex, heterosexual transmission.

^bOpportunistic infections: lung tuberculosis, oropharyngeal candidiasis, and cytomegalovirus retinitis.

^cBased on positive HBsAg result.

products were purified and directly sequenced as reported previously,¹² using primers designed to sequence overlapping regions. Five-nucleotide sequence segments were assembled using Sequencer v.4.8 to form a near full-length genome sequence, and aligned with reported genome sequences of various HIV-1 major group M in the world by MEGA v.4¹³ for phylogenetic tree construction and recombination analysis. SimPlot v.3.5¹⁴ and RDP v.3 software packages were used to analyze possible recombination events.

A representative neighbor-joining tree of the near full-length genome sequences showed that both 07IDJKT189 and 07IDJKT194 were clustered closely to each other, and formed a branch node with other CRF33_01B Malaysian sequences with high bootstrap values (Fig. 1). This branch joined the CRF01_AE, CRF34_01B, and CRF15_01B branches and formed a major branch node (Fig. 1). These results raised the

possibility that 07IDJKT189 and 07IDJKT194 had mosaic genomes similar to the Malaysian CRF33_01B genomes.¹⁰ To assess this possibility, we performed bootscanning plot analysis as described previously.¹⁵ The analysis confirmed that 07IDJKT189 and 07IDJKT194 indeed were mosaics of CRF01_AE and subtype B in Southeast Asia (subtype B') and had a mosaic pattern similar to the CRF33_01B genomes (Fig. 2A, bootscanning plots and phylogenetic trees): the four putative recombination breakpoints located around the 3' end of the *gag* gene or the 5' end of the *pol* gene constantly resulted in the highest statistical significance with various recombination detection tools in the RDP software package ($p < 0.01$).

Notably, the short genome segments corresponding to the border of the integrase and *Vif* coding regions were phylogenetically similar to subtype B and highly diverse among the 07IDJKT189, 07IDJKT194 (subsegment Va in Fig. 2B), and CRF33_01B genomes.¹⁶ We suspected that there was a possibility of recombination events in this region. However, we could not define appreciable recombination breakpoints and parental sequences for recombination by using available recombination detection tools in the RDP v.3 software package. This might be because the mosaic segment is too short to accumulate sufficient informative sites for assessing the statistical significance of the recombination events. Therefore it remains unclear at present how and why this region tends to be more diversified than the other regions. Interestingly, bootscanning analysis of this region in Malaysian CRF33_01B sequences obtained from <http://hiv-web.lanl.gov/> also shows similar results (data not shown), suggesting this region might be the new breakpoint site for recombination to occur.

To clarify whether the mosaicism and the diversity in the Indonesian CRF33_01B genomes exist at the clonal genome level, we cloned both the *gag-pol* and *pol-vif* regions from blood samples. Corresponding segments were amplified with segment 1 and 2-3 primer sets (Table 2) and cloned into plasmids using the TOPO XL PCR Cloning Kit (Invitrogen, Carlsbad, CA). The patterns of the bootscanning plots of the 2.3 kb containing *Gag* and part of the *Pol* coding region and the 1.9 kb containing the integrase and *Vif* coding region were almost identical between the sequences of the uncloned and

TABLE 2. PRIMERS USED FOR NESTED PCR AMPLIFICATION OF FIVE OVERLAPPED SEGMENTS

Segment	Primer	Gene	Location ^a	Sequence	
1	F: 174A	<i>gag-pol</i>	683 \Rightarrow 707	CTCTCGACGCAGGACTCGGCTTGCT	
	R: RT2-28		3012 \Leftarrow 3031	TGGAATATTGCTGGTGATCC	
2	F: RT3-27	<i>pol-int</i>	2798 \Rightarrow 2818	AACTCAAGACTTCTGGGAAGT	
	R: Pol-3'		4899 \Leftarrow 4919	GCTGTCCCTGTAATAAACCCG	
3	F: <i>pol/tat-5'</i>	<i>int-env</i>	4808 \Rightarrow 4829	GTACAGTGCAGGGGAAAGAATA	
	R: <i>env SB-3'</i>		6990 \Leftarrow 7010	ATTTAACAGCAGTTGAGTTGA	
4	F: 106A	<i>env</i>	6867 \Rightarrow 6889	ATACATTATTGTRCTCCRGCTGG	
	R: <i>env (Y,Y)</i> ^b		8288 \Leftarrow 8308	AAACCTAYYAAGCCTCCTACT	
	SK69 ^c		7911 \Leftarrow 7937	CCAGACTGTGAGTTGCAACAG	
5	F: <i>env/nef (R,Y)</i> ^b	<i>env-3'L</i>	8221 \Rightarrow 8243	GGRCAAGTYTGTGGAATTGGTTT	
	SK68 ^c		TR	7796 \Rightarrow 7815	AGCAGCAGGAAGCACTATGG
	R: MSR5		9606 \Leftarrow 9627	CAAGGCAAGCTTTATTGAGGCT	
2-3	F: <i>pol/seq2</i>	<i>pol-vif</i>	3876 \Rightarrow 3901	GATGGGGCAGCTARYAGGGARACTAA	
	R: <i>vif365B</i>		5782 \Leftarrow 5805	GGGTGTTCGACATAGCAGAATAGGC	

^aLocation is in reference to HIV HXB2 (K03455).

^bPrimer used for 07IDJKT189.

^cPrimer used for 07IDJKT194.

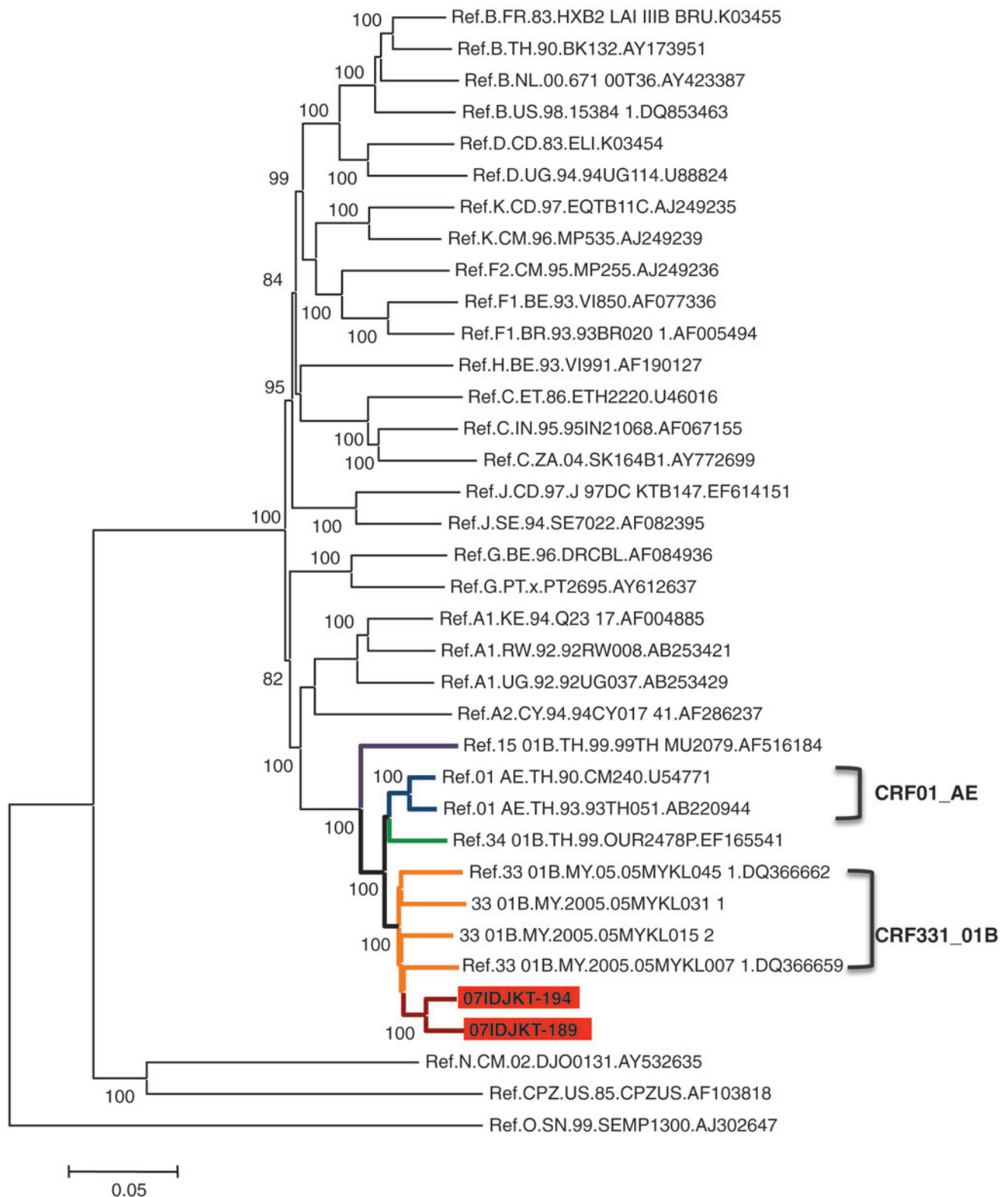


FIG. 1. Phylogenetic classification of near full-length genome sequences of 07IDJKT189 and 07IDJKT194. The tree was constructed with the neighbor-joining method using MEGA v.4. Reference sequences of HIV-1 major group M were obtained from the Los Alamos National Laboratory HIV Database (<http://hiv-web.lanl.gov/>). Two more sequences of CRF33_01B from Malaysia (accession numbers DQ366660 and DQ366661) were included in this analysis. The reliability of interior branches in the phylogenetic tree was assessed by the bootstrap method with 500 resamplings.

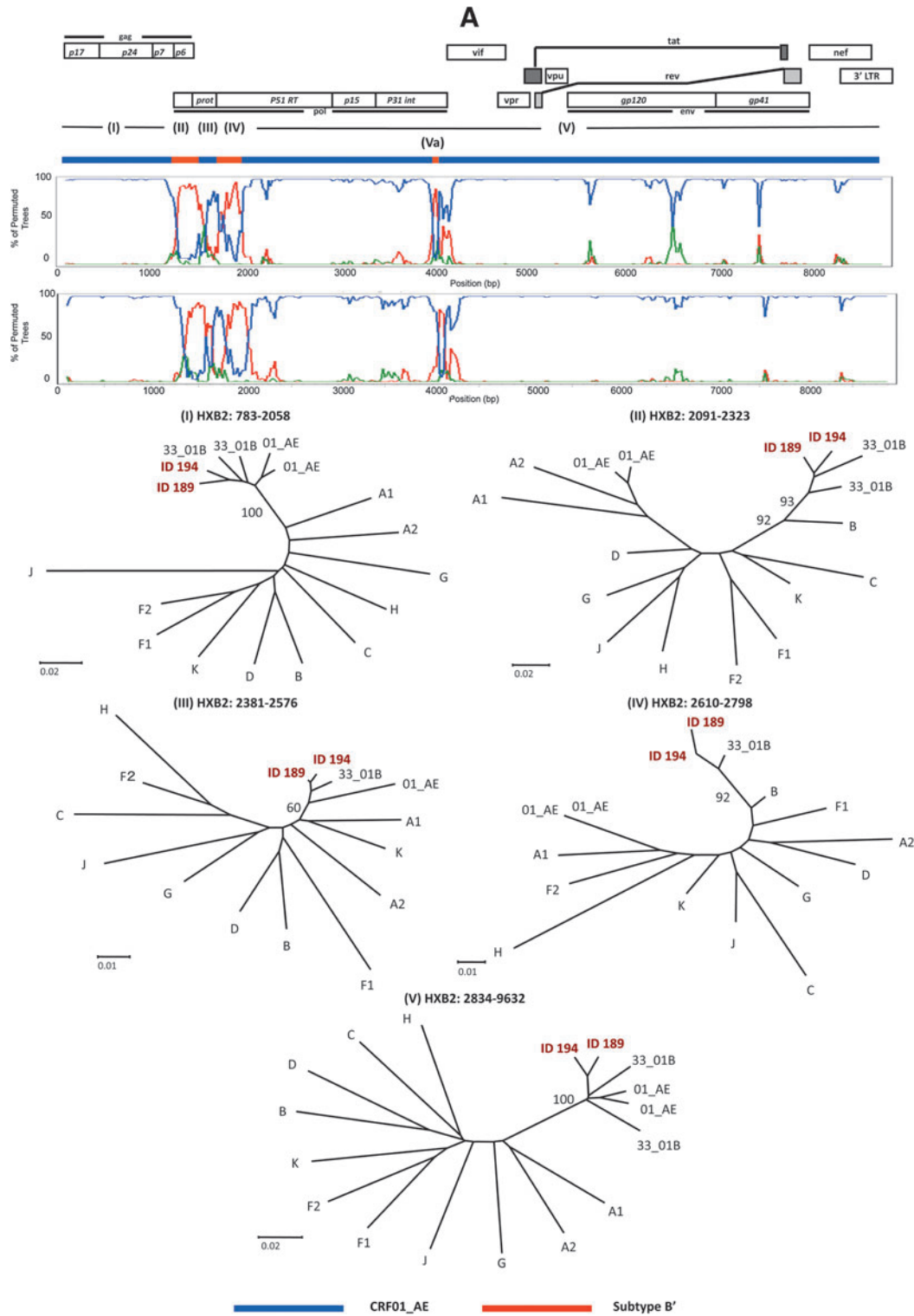
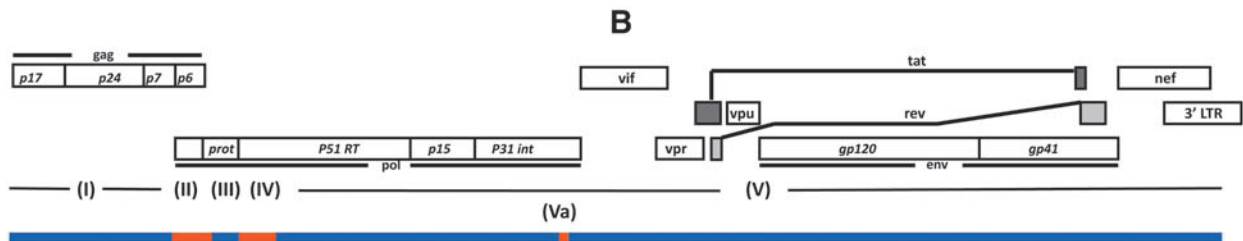
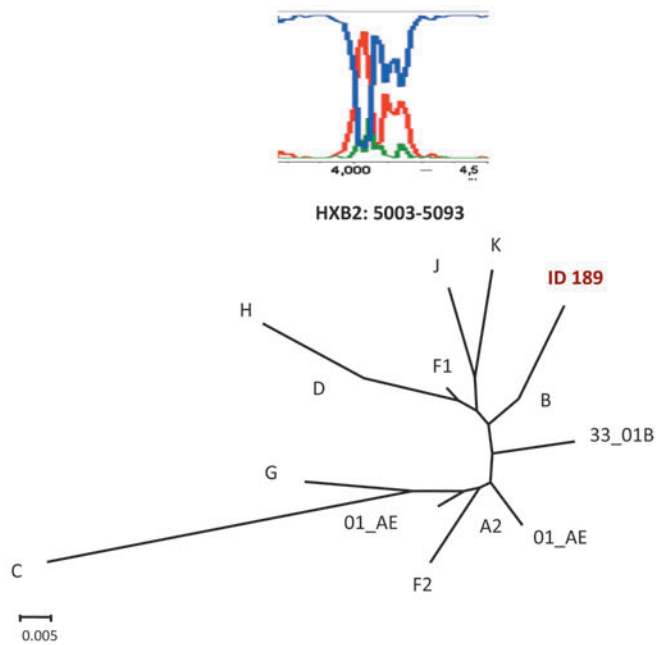


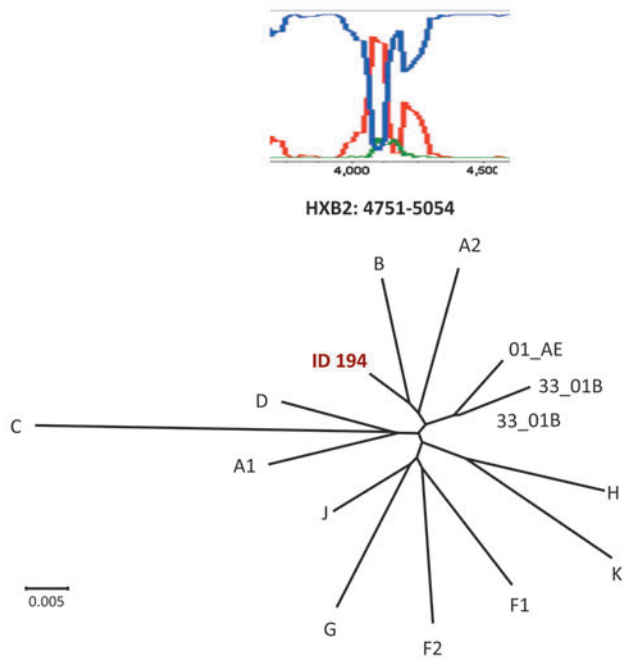
FIG. 2. Phylogenetic evidence for genome mosaicism of Indonesian HIV-1 strains. **(A)** Bootscanning plots of nucleotide sequences of near full-length genomes of 07IDJKT189 and 07IDJKT194 from Indonesian blood samples. Each query genome sequence was aligned with three reference sequences with CLUSTAL W software, using subtype B Thai variant (subtype B'; AY173951) and CRF01_AE (U54771) as putative parental strains, with subtype C (AF067155) as the outgroup. Bootstrap values were plotted for a window with 200 bp moving in increments of 20 bp. Neighbor-joining trees were constructed with segment (I), (II), (III), (IV), and (V) sequences and are shown under the bootscanning plots. **(B)** Phylogenetic relationship of the subgenome segment. The 1.9-kb segment containing portions of the integrase and *Vif* coding regions of 07IDJKT189 and 07IDJKT194 was amplified, cloned into plasmid, and subjected to bootscanning plot and phylogenetic tree analyses. Trees were constructed with Va region sequences (91 and 304 nt for 07IDJKT189 and 07IDJKT194, respectively). Reference sequences used for phylogenetic tree construction include HIV-1 A1 (AF004885), A2 (AF286238), B (AY173951), C (AF067155), F1 (AF005494), F2 (AY371158), G (U88826), H (AF005496), J (EF614151), K (AJ249239), CRF01_AE (U54771) and (AB220944), and CRF33_01B (DQ366662) and (DQ366659).



(1) Sub-segment Va of 07IDJKT189



(2) Sub-segment Va of 07IDJKT194



CRF01_AE Subtype B'

FIG. 2. (Continued)

cloned genome segment (Fig. 2, bootscanning plots). Similarly, results of phylogenetic tree analysis of both regions were consistent with those obtained with uncloned sequences (Fig. 2, phylogenetic trees). Taken together, these results suggest the presence of the mosaicism and diversity at these regions in the single genome segment.

Both Indonesian subjects came from different risk factor groups, and had no connection with each other with no history of traveling to Malaysia or contact with Malaysian people. This suggests that CRF33_01B might have been circulating in Indonesian long before both patients initiated their risky behaviors and have been crossing between risk factor groups. Moreover, the end-of-integrase variations of subtype B' in Indonesian CRF33_01B sequences suggest an evolutionary process that might have started in Malaysia and further developed in and spread to other countries, including Indonesia. So far several new CRFs in Southeast Asia have been reported, and they tended to circulate locally in the beginning of their endemic period. After a certain period of time, however, they started to spread to other countries.⁶ The Indonesian CRF33_01B variants not only support this phenomenon, but also suggest that these new CRFs are still evolving in each country and might contribute to the generation of other new CRFs by recombining with local strains of HIV-1.

Continuous study is necessary to see how HIV-1 has spread and evolved in Indonesia and in Southeast Asia in general. Recombination may alter the viral characteristics and thus influence the pathogenesis and/or the ability of the virus to survive, including the ability to generate resistance to antiretroviral drugs, which may affect the regional and global HIV-1 epidemic.

Sequence Data

GenBank accession numbers of nucleotide sequences reported in this article are AB547463 for 07IDJKT189 and AB547464 for 07IDJKT194.

Acknowledgments

We thank Dr. Gunawan Martinus, Dr. Dyah Waluyo, and the staff at Cipto Mangunkusumo National Hospital, Kramat 128 Hospital, and Sulianti Saroso Infectious Disease Hospital for their help during specimen collection. We thank Dr. Wataru Habano for technical advice and Ms. Kumi Furusawa for technical assistance. This work was supported by a grant from the Imai Memorial Trust for AIDS Research, a grant-in-aid for the Strategic Medical Science Research Center of Japan Ministry of Education, Culture, Sports, Science and Technology (2009–2013 period), and a fellowship from the Takeda Science Foundation.

Author Disclosure Statement

No competing financial interests exist.

References

1. Malim MH and Emerman M: HIV-1 sequence variation: Drift, shift, and attenuation. *Cell* 2001;104:469–472.
2. Blackard JT, Cohen DE, and Mayer KH: Human immunodeficiency virus super infection and recombination: Current state of knowledge and potential clinical consequences. *Clin Infect Dis* 2002;34:1108–1114.
3. Tovanabutra S, Beyrer C, Sakkhachornphop S, *et al.*: The changing molecular epidemiology of HIV type 1 among Northern Thai drug users, 1999 to 2002. *AIDS Res Hum Retroviruses* 2004;20:465–475.
4. Tee KK, Saw TL, Pon CK, *et al.*: The evolving molecular epidemiology of HIV type 1 among injecting drug users (IDUs) in Malaysia. *AIDS Res Hum Retroviruses* 2005;21:1046–1050.
5. Motomura K, Kusagawa S, Kato K, Nohtomi K, Lwin HH, Tun KM, *et al.*: Emergence of new forms of human immunodeficiency virus type 1 intersubtype recombinants in Central Myanmar. *AIDS Res Human Retroviruses* 2000;16:1831–1843.
6. Los Alamos HIV sequence database. <http://www.hiv.lanl.gov/>.
7. Wang B, Lau KA, Ong LY, *et al.*: Complex pattern 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;367:288–297.
8. Lau KA, Wang B, Kamarulzaman A, Ng KP, and Saksena NK: Near full-length sequence analysis of a unique CRF01_AE/B recombinant from Kuala Lumpur, Malaysia. *AIDS Res Hum Retroviruses* 2007;23:1139–1145.
9. 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;19:561–567.
10. Tee KK, Li XJ, Nohtomi, K, *et al.*: 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;43:523–529.
11. Tovanabutra S, Kijak GH, Betrer 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;23:829–833.
12. SahBandar IN, Takahashi K, Djoerban Z, *et al.*: Current HIV Type 1 molecular epidemiology profile and identification of unique recombinant forms in Jakarta, Indonesia. *AIDS Res Hum Retroviruses* 2009;25:637–646.
13. Tamura K, Dudley J, Nei M, and Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. (Publication PDF at <http://www.kumarlab.net/publications>.)
14. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, and Ray SC: 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.
15. 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.
16. Martin DP, Williamson C, and Posada D: RDP2: Recombination detection and analysis from sequence alignments. *Bioinformatics* 2005;21:260–262.

Address correspondence to:

Kiyomi Takahashi

Department of Microbiology

Iwate Medical University School of Medicine

19-1 Uchimaru

Morioka 020 8505

Japan

E-mail: kiyomit@iwate-med.ac.jp