Supplementary data

Methods

Construction of a full-length CRF07_BC molecular clone

A full-length CRF07_BC molecular clone was constructed with the strategy delineated in Supplementary Figure 1A. Briefly, a full-length HIV-1 genome was constructed by ligation of two PCR products, representing the 5'-LTR and gag-pol-env-nef-U3R fragments, respectively. They are SpeI-LTR-1 (5'-ACT AGT TGG AAG GGC TAA TTC ACT CC-3') and MSF12-RC (5'-CTG TTC GGG CGC CAC TGC TAG AGA TTT-3') for the 5'-LTR, and MSF12 (5'-AAA TCT CTA GCA GTG GCG CCC GAA CAG-3') and ofm19-NotI (5'-ATA AGA ATG CGG CCG CGG CAA GCT TTA TTG AGG CTT A-3') for the gag-pol-env-nef-U3R fragment. The pCRF07_BC+p6 plasmid was derived from pCRF07_BC with an addition of 7 amino acids in p6 ($_{30}$PIDKELY$_{36}$) by site-directed mutagenesis.

Plasmids

The HXB2RU3 and CI plasmids were derived from Dr. Tun-Hou Lee's laboratory (Department of Immunology and Infectious Diseases, Harvard School of Public Health) as described [1]. NL4.3 and P125 plasmids were kindly

provided by Dr. Li-Ming Huang (Department of Pediatrics, National Taiwan University Hospital). The pRK5-Tat B and pRK5-Tat CRF07_BC plasmids were derived from pRK5 expressing the HXB2RU3-derived or CRF07_BC-derived full-length Tat, with their C-termini tagged with Flag [2]. The pGL2–B LTR and pGL2–CRF07_BC LTR plasmids, containing a luciferase gene driven by HIV-1 HXB2RU3 or CRF07_BC LTR, were used as the Tat-mediated transactivation reporter system. The pGL2–CRF07_BC LTRm plasmid, whose third NF-kBiii binding site has been disrupted by deletion of nt 340-343(GGGA), was generated by site-directed mutagenesis.

Results

Decreased Tat-mediated transactivation of long terminal repeat (LTR)

The activation of long terminal repeat (LTR)-directed gene expression by HIV-1-encoded RNA-binding protein Tat is critical during HIV-1 replication [3]. The function of full-length CRF07_BC LTR, which is a B'/C recombinant with 5' end (nt 1-472) from subtype C and 3' end (nt 473-633) from subtype B (Supplementary Figure 1C), was first characterized in this study. Like subtype C LTR, CRF07_BC LTR has one additional potential NF-κBiii binding site. Therefore, the ability of CRF07_BC Tat to transactivate LTR and the role of the

additional NF-κB binding site in Tat-mediated transactivation was determined. As shown in Supplementary Figure 3, CRF07_BC Tat has a significantly lower transactivation activity on either authentic LTR or subtype B LTR compared to subtype B Tat. The decreased transactivation activity of CRF07_BC Tat was not due to reduced expression of Tat proteins since comparable Tat protein expression was detected. Site-directed mutagenesis was further conducted to mutate the potential third NF-κB site on CRF07_BC LTR, and the mutated LTR, CRF07_BC LTRm, exhibited similar activity as the CRF07_BC LTR, implicating the third NF-κB site might not increase the promoter activity of CRF07_BC LTR.

Figure S1A

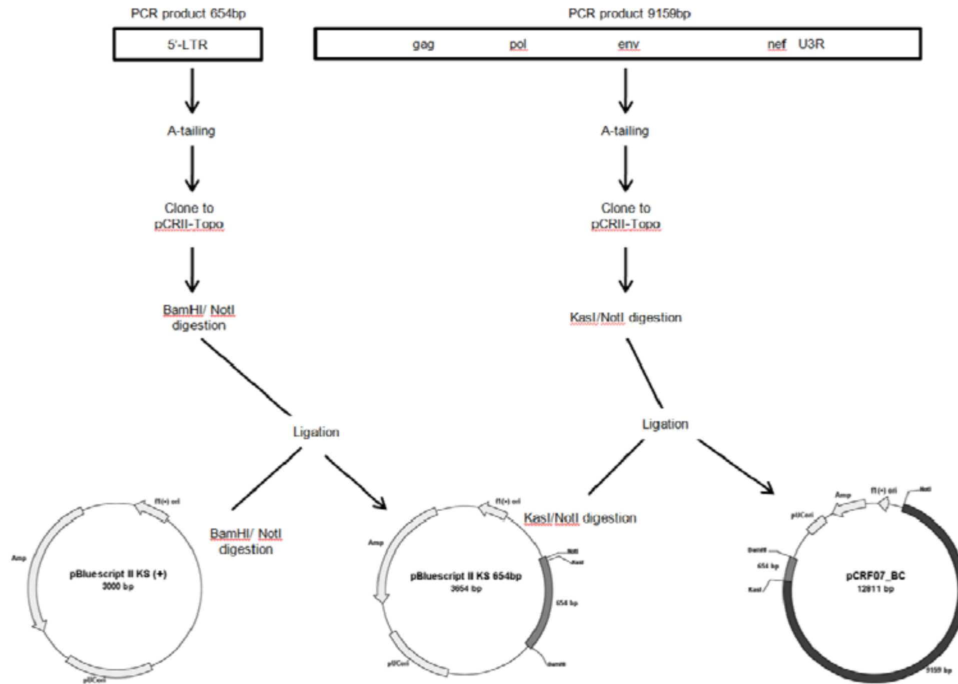


Figure S1B

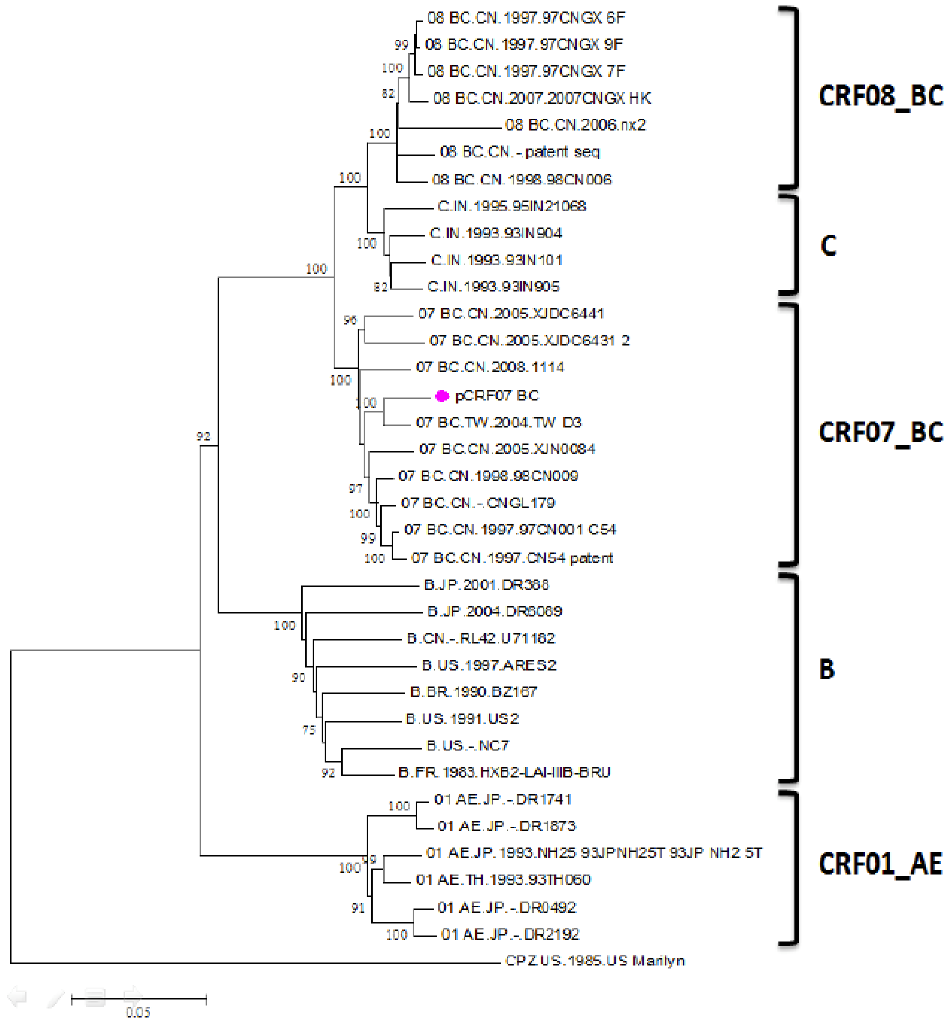
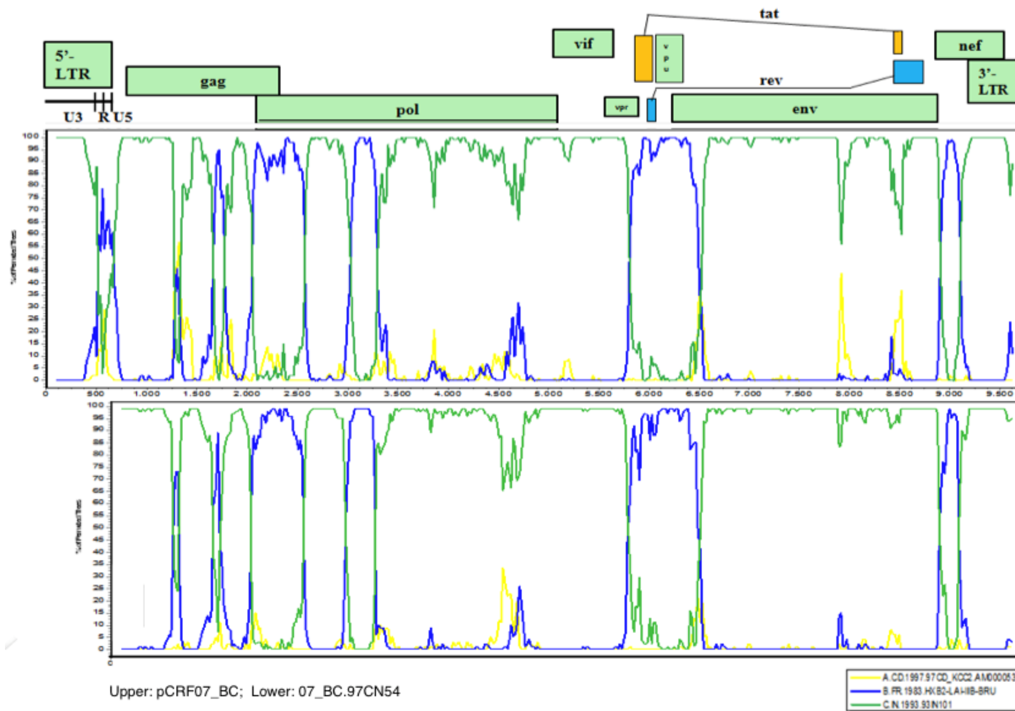


Figure S1C



Supplementary Figure S1 Construction of pCRF07_BC molecular clone and sequence analysis of the full-length CRF07_BC sequence. (A) Cloning strategy of pCR07_BC. Two PCR amplicons representing the 5'LTR and the gag-pol-env-nef-U3R fragments were cloned respectively to the pCRII-Topo vectors. Each of the plasmids were digested with the indicated restriction enzymes, and the inserts were subsequently ligated to the pBluscript II KS (+) vector to obtain the full-length pCRF07_BC clone (accession number: KF234628). **(B) Phylogenetic tree analysis of full-length CRF07_BC sequences.** Phylogenetic tree analysis was

conducted to analyze the the full-length CRF07_BC sequences available from the NCBI database (http://www.ncbi.nlm.nih.gov/nuccore/). The phylogenetic tree was constructed by the neighbor-joining method based on the Kimura 2-parameter distance matrix listed in the MEGA software. Reference sequences of various subtypes and recombinants were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/nuccore/). **(C) Bootscanning analysis of full-length CRF07_BC sequences.** In the bootscanning analysis, the full-length CRF07_BC sequences from pCRF07_BC and 07_BC.97CN54 (AX149647) were first aligned and gap stripped. The recombination breakpoints of full-length CRF07_BC sequences from pCRF07_BC and 07_BC.97CN54 were assessed by the SimPlot software (version 3.5.1), using subtype B strain B.FR.1983.HXB2 (K03455) and subtype C strain C.IN.1993.93IN101 (AB023804), as references. Analysis was performed using the maximum parsimony and a sliding window of 200 nucleotides (nt) advanced in 20-nt increments. Nucleotide positions were indicated in relation to HXB2 using the HIV/SIV Sequence Locator Tool.