Multiple NF-κB sites in HIV-1 subtype C LTR confer superior magnitude of transcription and thereby the enhanced viral predominance

Mahesh Bachu†, Swarupa Yalla†, Mangaiarkarasi Asokan†, Anjali Verma†, Ujjwal Neogi‡, Shilpee Sharma†, Rajesh V Murali†, Anil Babu Mukthey†, Raghavendra Bhatt‡, Snehajyothi Chatterjee‡, Roshan Elsa Rajan†, Narayana Cheedarlal†, Venkat S Yadavalli†, Anita Mahadevan§, Susarla K Shankar§, Nirmala Rajagopalan§, Anita Shet†, Shanmugam Saravanan%, Pachamuthu Balakrishnan%, Suniti Solomon%, Madhu Vajpayee†, Kadappa Shivappa Satish*, Tapas K. Kundu†, Kuan-Teh Jeang†, and Udaykumar Ranga†

†HIV/AIDS Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India, ‡St. John’s National Academy of Health Sciences, Bengaluru, India, §Freedom Foundation, Bengaluru, India. †Transcription and Disease Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India, ‡Molecular Virology Section, Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, §The Department of Neuropathology, National Institute of Mental Health and Neurological Sciences, Bengaluru, India, %YRG Centre for AIDS Research and Education (YRG CARE), Voluntary Health Services Hospital Campus, Taramani, Chennai-600 113, India, *HIV and Immunology Division, Department of Microbiology, All India Institute of Medical Sciences (AIIMS), Ansari Nagar, New Delhi 110029, India, %Seva Free Clinic and Chest and Maternity Centre, Bengaluru, India

Running title: Role of NF-κB in HIV subtype C evolution

*To whom correspondence should be addressed: Udaykumar Ranga, Professor, HIV/AIDS Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur (PO), Bangalore 560 064, INDIA, Ph. +91-80-2208 2830 (Office),Fax: +91-80-2208 2766, Email: udaykumar at jncasr.ac.in

Keywords: HIV-1, Subtype C, NF-κB, Replication competence, Viral evolution, LTR

Background: Viral evolution of HIV-1 is dynamic and moving towards higher order of replicative fitness.

Results: HIV-1 subtype C acquires an extra (4th) NF-κB site to achieve a higher degree of transcription and in turn enhances its replicative fitness and preponderance.

Conclusion: Subtype C with an extra NF-κB site adopts a novel strategy of strengthening its promoter to gain fitness.

Significance: Learning how the new strains could impact viral prevalence and pathogenesis and disease management strategies is critical.
SUMMARY

We demonstrate that at least three different promoter variant strains of HIV-1 subtype C have been gradually expanding and replacing the standard subtype C viruses in India, and possibly in South Africa and other global regions, over the past decade. The new viral strains contain an additional NF-κB, NF-κB-like or RBEIII site in the viral promoter. While the acquisition of an additional RBEIII site is a property shared by all the HIV-1 subtypes, acquiring an additional NF-κB site remains an exclusive property of subtype C. The acquired κB-site is genetically distinct, binds the p50-p65 heterodimer and strengthens the viral promoter at the levels of transcription initiation and elongation. The 4-κB viruses dominate the 3-κB ‘isogenic’ viral strains in pairwise competition assays in T-cell lines, primary cells, and the EcoHIV- mouse model. The dominance of the 4-κB viral strains is also evident in the natural context when the subjects are coinfected with κB-variant viral strains. The mean plasma viral loads, but not CD4 counts, are significantly different in 4-κB infection suggesting that these newly emerging strains are probably more infectious. It is possible that higher plasma viral loads underlie selective transmission of the 4-κB viral strains. Several publications previously reported duplication or deletion of diverse transcription factor binding sites (TFBS) in the viral promoter. Unlike previous reports, our study provides experimental evidence that the new viral strains gained a potential selective advantage as a consequence of the acquired TFBS, and importantly that these strains have been expanding at the population level.

Human Immunodeficiency Virus Type 1 (HIV-1) strains are classified into four groups (M, N, O and P) of which group M, responsible for much of the global pandemic, is further classified into 9 genetic subtypes (A, B, C, D, F, G, H, J and K). In addition, at least 55 circulating recombinant forms (CRF) and a large number of unique recombinant forms (URF) have been documented (1). Of the various genetic subtypes of HIV-1 that are unevenly distributed globally, subtype C and its recombinant forms are responsible for more than half of all global HIV infections (2).

Although subtype C shares the time and space of its origin with all other HIV-1 genetic subtypes from the African continent (1), the dissemination pattern of this subtype seems to demonstrate several interesting properties. First, subtype C appears to establish monophyletic epidemics in countries of its origin in Africa and in countries such as India where its introduction precedes that of other subtypes. Second, subtype C appears to rapidly replace or dominate ‘founder’ viral strains where its introduction follows that of other subtypes as is manifested in its competition against several other subtypes in The Democratic Republic of Congo, Tanzania and South Africa; against subtype B in the South American continent especially in southern Brazil; and against subtypes B and CRF01_AE in China. Lastly, since the time of its origin, subtype C has expanded at a faster rate than any other viral subtype. For instance, in southern Brazil, the incidence of subtype C increased from 3% in 1990s (3) to 30% in 2002 (4) and eventually to 45-48% in recent years (5). Furthermore, since the time of their introduction in India, the subtype C strains of India appear to remain genetically stable without undergoing a major genetic recombination with other subtypes, unlike that seen in China or southern Brazil, even though several CRFs have been reported from India (6).

The regulatory elements of viruses could play an important role in conferring differences in replication fitness as previously observed in the HIV-1 subtypes. The long-terminal repeat (LTR) sequences of the viral subtypes are highly diverse, differing up to 20-25% between subtypes (7). A comparison of subtype C LTR (C-LTR) with that of others identified several distinct differences in the composition of the TFBS including NF-κB (8,9), NF-AT, USF (10), and other regulatory elements such as the TATA box, and the TAR region (11-13). Of these variations in the LTR, subtype-specific patterns within the enhancer element, exclusively consisting of the NF-κB motifs, are important given the profound impact NF-κB has on gene expression regulation from the
viral promoter. The enhancer in most of the viral subtypes, including the prototype subtype B virus, consists of two identical and canonical κB motifs with the exception of subtypes A/E and C. While in subtype A/E, the upstream NF-κB site is replaced with a GABP-binding motif (14), the genetic variation within the subtype C enhancer is more complex. While a large proportion of C-LTRs contains three NF-κB sites (8,15,16), a small minority of these LTRs contain a fourth motif that is either a canonical κB site or κB-like site (8,17-20). Additionally, C-LTRs with only two κB sites have also been reported (16,21). C-LTR containing three NF-κB sites can demonstrate stronger transactivation activity compared to the LTRs of other subtypes containing two NF-κB sites (10,12). Thus, unlike other HIV-1 subtypes, the C-LTR has a significant variation of κB-site number and sequence changes within these sites.

In subtype B, immediately upstream of the viral enhancer element, an insertion of unique sequences of 15 to 34 bp length was reported in approximately 38% of isolates in 1990s (22), or in at least 14.2% of the viral isolates in our present analysis; this finding adds an additional level of complexity to the genetic diversity of this important regulatory region. These insertions commonly known as the most frequent naturally occurring length polymorphism (MFNLP), predominantly generate an RBEIII motif, a binding site for the RBF2 transcription factor (23). It was proposed that MFNLP is a compensatory mechanism to ensure the presence of at least one functional RBEIII site in the LTR (24). Although a small number of such sequences have been deposited into the databases, MFNLP insertions have not been reported previously for the subtype C LTR.

The divergence of HIV-1 into several genetic subtypes, the uneven distribution of these subtypes across the globe, and the evident domination of certain subtypes - especially of subtype C - over others raise two important questions. First, which molecular properties of the viral subtypes may underlie the wide range of biological differences observed? Second, are the HIV-1 subtypes, likely to undergo additional evolutionary modifications in the future?

In a previous screening, from a total of 608 primary clinical viral isolates collected at several clinics in southern India in 2000-2003, 34 viral strains containing sequence insertions in the LTR were identified (25). A subsequent sequence analysis of the viral enhancer in 25 of the 34 viruses demonstrated that the sequence insertions in C-LTR acquired at least two different types of TFBS- predominantly NF-κB (12 of 25) and in a minority of the strains the RBEIII site (6 of 25) (26). Additionally, a few viral strains (4 of 25) contained sequence insertions that broadly resembled a canonical NF-κB motif hence referred to here as the κB-like sites. Two of the 25 sequences contained the insertion of both of the RBEIII and the κB-like motifs. A single viral strain contained a dual insertion of a canonical NF-κB site and an RBEIII-like site. Our previous analysis thus identified five different types of sequence insertions in the subtype C viral promoter. It is, however, not known whether any of the promoter variant viral strains is endowed with altered infectivity and/or pathogenic properties.

Here, we demonstrate that HIV-1 subtype C appears to be undergoing further modifications in the viral promoter by the acquisition of additional TFBS in the enhancer region - possibly gaining higher levels of replication fitness. Importantly, no associated variations have been found in any of the viral gene products consisting of all of the structural, regulatory, and accessory proteins. At least three different viral promoter variant strains of subtype C have emerged in the early 1980s in India and South Africa independently (data not shown) and possibly at other geographical locations, and appear to have been replacing the standard subtype C strains at substantial pace. Amongst these, the viral strains containing four κB-binding sites in the enhancer demonstrate the fastest expansion in India. The emergence of subtype C viral variants containing four κB-sites is not merely of academic interest but may have implications for viral fitness, pathogenesis, and
overall viral evolution considering the magnitude of viral recombination characteristics of HIV-1.

**EXPERIMENTAL PROCEDURES**

**Ethics statement** - Ethical clearance for the study was obtained from the institutional human ethics committees at the participating institutions including Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), St. John’s Hospital, Freedom Foundation, Y. R. Gaitonde Centre for AIDS Research and Education (YRG CARE), and All India Institute of Medical Sciences (AIIMS). The blood samples were collected after obtaining a written informed consent form each participant. The BALB/c mice used in EcoHIV-infection experiments were handled in an animal facility (Registration No. 201/CPCSEA) in strict accordance with rules and guidelines defined by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), The Government of India and approved by the Institutional Animal Ethics Committee of JNCASR.

**Clinical samples** – The study participants consisted of only adult subjects, over 18 years of age, representing both the genders (83/214 females, 39%), believed to have acquired the infection primarily through heterosexual transmission and a large majority were drug-naïve (166/214, 78%) (Supplemental Table 1). A single vial of 6-8 ml peripheral blood was collected after obtaining informed consent from each of the participants. The samples from St. John’s Hospital (n=46) and Freedom Foundation (n=30) both based in Bengaluru, YRG CARE, Chennai (n=55) and AIIMS, New Delhi (n=30) were collected during 2010-2011. For the southern Indian cohort, a total of 607 samples were collected during 2000-2003 from multiple urban centers (25). Samples for the JNCASR cohort comprising of 57 volunteers were collected between 2006 and 2007 (27).

**Amplification, cloning and bioinformatics analysis of the LTR sequences** - The amplification and cloning of the LTR sequences from the patient isolates were performed as described previously by us (26). Briefly, full-length or U3 region of the viral LTR were amplified from the proviral DNA or plasma viral RNA, respectively. The full length LTR sequences amplified from the proviral DNA were directionally cloned between the restriction sites MluI and EcoRI into the reporter vector pcDNA3.1 (+) Luciferase-IRES-EGFP thus substituting the CMV promoter with the amplified LTR. The sequence information for each of the viral promoters was obtained in both the directions using multiple plasmid clones for each of the clinical samples. Primers N1007 (5’-TTAAGCTACAAGGCAAGGC-3’) and N1009 (5’-GTTGTTTCGTGGCGTTGG-3’) were used for sequence determination of the LTR insert. Multiple recombinant clones were sequenced for each subject. Alternatively, the viral RNA was extracted from the frozen plasma samples using a commercial kit (Cat.No: 740956, Nuclisense Viral RNA isolation kit, Machery-Nagel, USA). First-strand cDNA was synthesized using random hexamers and Superscript II Reverse Transcriptase (Cat.No: 18064-022, Invitrogen BioServices, Bengaluru, India). A 400 bp fragment in the U3 region of the LTR was amplified using the primers N1031 (5’–GCTTCTTTTTAAGAAAAAGGGGGACTGGA-3’) and N1024 (5’-TGTACTGGGTTCATGCTAGGTAAGGA-3’). The PCR products were purified and sequenced directly using the primer N698. DNA sequencing was performed on the ABI 377 automated sequencer using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit. Every individual LTR sequence was subjected to the BLAST analysis against the global and laboratory sequence database to confirm authenticity. The sequences were deposited to GenBank sequence database and the information related to the patient details and GenBank accession numbers is provided in Supplemental Table 1. Of a total of 3,054 full-length LTR sequences belonging to all HIV-1 genetic subtypes either downloaded from the databases or generated through the present work, 479 sequences contained insertions up stream of the viral enhancer. Of the total 1,113 sequences in subtype C, 213 sequences contained such insertions. The
multiple sequence alignment was performed using ClustalW in the BioEdit software. The transcription factor binding sites in the LTR sequences were identified using the TF Search program (www.rwcp.or.jp).

**Cell Culture** - The Jurkat, CEM-CCR5 and EL4 cells were cultured in RPMI 1640 medium (Cat. No: R0883, Sigma Aldrich, St. Louis, Missouri, USA) supplemented with 10% fetal bovine serum (Cat. No: 04-222-1, Biological Industries, Kibbutz Beit Haemek, Israel), 2 mM glutamine, 100 units/ml penicillin G, and 100 g/ml streptomycin. The peripheral blood mononuclear cells (PBMC) were purified from 10 ml of fresh blood from healthy donors by density-gradient centrifugation using Ficoll-Hypaque (Cat. No 10771, Sigma-Aldrich, St. Louis, USA). The CD8 cells were depleted from the PBMC using the StemSep Human CD8\(^+\) depletion kit (Cat. No. 14662, Stem Cell Technologies, Vancouver, Canada). PBMCs were cultured in complete RPMI medium supplemented with 10 U/ml of Interleukin-2 (IL-2, Cat. No: 136, NIH AIDS Reagent Program) and 5 µg/ml of Phytosohemagglutinin P (PHA-P, Cat. No: L9132, Sigma-Aldrich, St. Louis, USA) for three days. PBMC were subsequently used for viral infection in a medium supplemented with only IL-2. The human embryonic kidney HEK293, 293T and TZM-bl cells were grown in Dulbecco’s modified Eagle’s medium (Cat. No: D5546, Sigma-Aldrich, St. Louis, USA) supplemented with 10% FBS.

**Preparation of the nuclear extract** - Preparation of the nuclear extract: Jurkat cells (50 x 10^6) were suspended in 1 ml of ice cold 1 X PBS (Phosphate buffered saline) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.47 mM KH₂PO₄, pH 7.4) in a 1.5 ml plastic vial and centrifuged at 500 g for 5 min at 4°C. PBS was aspirated and the cells were resuspended in 500 µl of the sucrose buffer (0.32 M sucrose, 10 mM Tris-Cl, pH 8.0, 3 mM CaCl₂, 2 mM MgOAc, 0.1 mM EDTA, 0.5% Nonidet P-40 (NP-40), 1 mM Dithiothreitol (DTT) and 0.5 mM phenylmethanesulfonyl fluoride (PMSF)). Using a 1 ml wide-bore plastic tip, cells were mixed gently by pipetting the suspension up and down several times. The cell suspension was centrifuged at 500 g for 5 min at 4°C and the supernatant was aspirated. The nuclear pellet was washed once with 1 ml cold sucrose buffer devoid of NP-40. Using a 1 ml wide-bore plastic tip, the nuclear pellet was gently pipetted up and down to disrupt it. The nuclei were resuspended in 150 µl of the low salt buffer (20 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 20 mM KCl, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT and 0.5 mM PMSF) and subjected to gentle vortexing. Subsequently, an equal volume of the high salt buffer (20 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 800 mM KCl, 0.2 mM EDTA, 25% glycerol, 1% NP-40, 0.5 mM DTT, 0.5 mM PMSF, 4.0 µg/ml each of leupeptin, aprotinin and pepstatin) was added slowly drop by drop while mixing the contents with a plastic tip. The samples were then incubated for 30-45 min at 4°C with gentle mixing on a rotator. The samples were centrifuged at 14,000 g for 15 min at 4°C. The protein content of the extracts was determined using a commercial BCA Protein Assay Kit (Cat. No. 23227, Pierce biotech, Rockford, USA). The nuclear extract was snap frozen in liquid nitrogen and stored in a deep-freezer until use.

**Preparation of the radiolabeled probe** - One of the two oligonucleotides of a double-stranded DNA probe was radiolabeled using 50 ng of DNA and T4 polynucleotide kinase (New England Biolabs, Massachusetts, USA) in a solution containing 30 µCi of [γ^32P] ATP. The mixture was incubated for 30 min at 37°C and the enzyme was heat-inactivated at 65°C for 20 min. The oligonucleotide was passed through a pre-packed Sephadex G-50 column. The oligonucleotide was then hybridized to 200 ng of the complementary strand in the annealing buffer (100 mM NaCl, 5 mM Tris pH 7.5, 10 mM MgCl₂, 20 µM EDTA, and 1 mM DTT). The annealed double-stranded probe was gel-purified using a 6% polyacrylamide gel prior to use in the EMSA. An autoradiogram was developed by exposing the gel to an X-ray film (GBX-2, Kodak, Colorado, USA) for 1 h. The developed autoradiogram was superimposed on the gel and the corresponding gel portion was excised out using a scalpel. The probe was extracted from the excised gel using the crush and
soak method (28). The following probe sequences were used in the mobility shift experiments. F-κB (5’-cagaaGGGACTTTCTgtg-3’ and 5’-cagcAGAAAGTCCCctctg-3’), H-κB (5’-ccgctGGGACTTTCCagga-3’ and 5’-tcctGGAAAGTCCCagcgg-3’), F-κB* (5’-ccgctGGGACTTTCTagga-3’ and 5’-tcctAGAAAGTCCCagcgg-3’), H-κB* (5’-cagaaGGGACTTTCCgctg-3’ and 5’-cagcGGAAAGTCCCttctg-3’), Mutant of H-κB (5’-ccgctGCCACTTGGCagga-3’ and 5’-tcctGCCAAGTGGCagcgg-3’), IκB (5’-gtagGGGACTTTCCgagctcgagatcctatg-3’ and 5’-cataggatctcgagctcGGAAAGTCCCctac-3’ and 5’-TTCTAATGATTGCATTCGAC-3’).

Gel-shift assays - For the electrophoretic mobility shift assays (EMSA), 10 µg of nuclear extracts were mixed with 30,000 CPM of the labeled probe in the binding buffer (10 mM HEPES pH 7.9, 4% glycerol, 25 mM KCl, 1 mM DTT, 0.5 mM EDTA, and 25 mM NaCl) and in the presence of poly dIPdC (1 µg/ml) (Cat. No: P4929P5UN, SigmaAldrich, St. Louis, Missouri, USA), BSA (1 µg/ml). The sample was incubated for 20 min at room temperature and resolved on a 6% polyacrylamide gel at 150 V in a cold room. Competition experiments were performed with 100Pfold excess of unlabeled oligonucleotides. For the supershift assay, nuclear extracts were pre-incubated with 1 µg of affinity-purified rabbit polyclonal antibodies raised against p50, p65, c-Rel, p52 or RelB proteins. After 30 min of incubation, radiolabeled probes were added and the electrophoresis was performed as described above. Gels were dried and exposed to a Kodak Biomax film at -85 °C. Anti-rel antibodies were raised against each of the five rel family members using antigenic peptides conjugated to the Keyhole Limpet Hemocyanin (KLH) carrier in New Zealand White rabbits. Antigen-specific antibodies were purified on immune-affinity columns using peptide affinity chromatography. The peptide sequences used for raising the antibodies are as follows - p50 (NH2-YNPGLDGIIEYDDFKLNSSIC-COOH), p65 (NH2-QASALAPAPPQVLQPAPAC-COOH), c-Rel (NH2-FQVLPDEHGNLTTALPPVVC-COOH), p52 (NH2-AEDDPYLGRPEQMFHLDPSLC-COOH) and RelB (NH2-NHSGPFLPPSALLPDPFSSGTVCSC-COOH). The specificity of the affinity-purified antibodies was confirmed using Western blot and EMSA analyses.

Protein expression and purification - The prokaryotic expression plasmid pGEX CD-p50 was a generous gift from Dr. Neil D Perkins (University of Dundee, UK). The expression and purification of full length p50-GST fusion protein was accomplished as described (29). Briefly, p50-GST protein was expressed in Escherichia coli BL21 (DE3) cells. The cells were grown in the LB medium supplemented with 100 µg/ml ampicillin for 3–5 h at room temperature until the cells reached an optical density of 0.5 at 600 nm and then induced for 1 h with 1 mM of isopropyl-1-thio-β-D-galactopyranoside. The induced cells were pelleted, washed and resuspended in GST-binding buffer (20 mM HEPES, pH 7.5, 10 mM MgCl2, 20% glycerol, 0.1% NP-40, 1 mM DTT, Protease inhibitor cocktail (Cat. No: P 8465, Sigma-Aldrich, St. Louis, Missouri, USA) and 0.5 mM PMSF). The bacterial cell suspension was sonicated in a Sonics Vibra Cell sonicator at amplitude of 40% with a 10-second pulse alternated with 10 seconds of resting on ice for 5 cycles. The lysate was centrifuged at 45,000 × g for 30 min to remove the cell debris. The supernatant from the lysate was mixed with 500 µL of Glutathione Sepharose 4B beads (Cat. No: 17-0756-01, GE Healthcare Bio-Sciences Corp, NJ, USA) pre-incubated with GST-binding buffer (A 50:50 slurry of glutathione-Sepharose 4B beads in GST-binding buffer). The lysate along with the beads was allowed to tumble on a rotator for 2 h at 4 °C. The suspension was centrifuged at 100 x g and the supernatant was carefully aspirated without disturbing the beads. The beads were washed three times with 20 ml of GST-binding buffer each time and the slurry was transferred to a fresh 1.5 ml plastic vial. The elution was performed using 500 µL of GST-elution buffer (10 mM reduced glutathione, 50 mM Tris-HCl, pH 8.0 and 5% glycerol). The beads were incubated with...
the elution buffer for 15 min on ice, the vial was centrifuged at 100 x g and the supernatant was transferred to a fresh vial. The supernatant was subjected to dialysis in ITC buffer (10 mM sodium phosphate, pH 7.0, 150 mM NaCl, 1 mM EDTA, and 5 mM DTT), overnight at 4°C with three changes of the ITC buffer. Following dialysis, the protein concentration of the sample was determined using a commercial kit (Cat No: 23225, BCA Protein Assay Kit, Pierce) and the sample was stored in aliquots at -20°C until required.

Isothermal titration calorimetry (ITC) – The ITC measurements were recorded using the ITC200 instrument (MicroCal, Inc. California, USA). Titrations were performed at 25°C in ITC buffer (10 mM sodium phosphate, 150 mM NaCl, 1 mM Na₂EDTA, and 5 mM DTT at pH 7.0). A micro cuvette was loaded with 240 µL of 10 µM of the recombinant p50 protein GST-tagged and 2 µl aliquots of the H-κB, F-κB or mutant double-stranded oligonucleotides, at a concentration of 100 µM, were injected into the cell using a syringe. The protein solutions were dialyzed against the ITC buffer prior to the assay to minimize the contribution of the dilution to the binding heat. Injections were made at an interval of 120 sec, and the duration of each injection was 0.4 sec. A stirring speed of 1,000 rpm was maintained during the assay to ensure proper mixing after each injection. The heat change versus the molar ratio of the titrated products was plotted and analyzed using the manufacturer’s software. Origin software within the ITC data analysis package was used for the data analysis.

The chromatin immunoprecipitation assay (ChIP) - Jurkat cells (5 x 10⁷) were infected with 5,000 infectious units of VSV-G pseudotyped pCLIGIT viruses and harvested at 72 h. Prior to harvesting, cells were treated in the absence or presence of Tumor necrosis factor -α (TNF-α, 20 ng/ml) for 1 h. The cells were resuspended in 10 ml of 1 X PBS and a formaldehyde stock solution of 37% was added to the suspension drop wise and with continuous mixing to a final concentration of 1%. After 10 min of incubation at room temperature, the cross-linking reaction was stopped by adding a glycine stock solution of 2.5 M to a final concentration of 0.125 M. The cells were centrifuged at 2,000 rpm for 5 min at 4°C, supernatant was aspirated and the cells were washed once using cold PBS. The cells placed on ice were resuspended in 0.5 ml the lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, 0.5% NP-40). Protease Inhibitor Cocktail (Cat. No. 11836170001, Roche Applied Science, Indianapolis, USA) was added to the buffer just prior to use. The tube was centrifuged at 2,000 rpm at 4°C for 5 min and the crude nuclear fraction was resuspended in 500 µl of the RIPA buffer supplemented with the protease inhibitors. The nuclear suspension, in 100 µl fractions, was subjected to sonication using the Sonics Vibra cell sonicator for 10 cycles, at 10% amplitude, using 15-second-on, 15-second-off pulses for the duration of 5 min with 1 min intervals. DNA was sheared to achieve an average size of 200 - 700 bp, as confirmed in electrophoresis. The sonicated mixture was centrifuged at 14,000 rpm for 15 min at 4°C and the supernatant was collected. Immunoprecipitation was performed using 5 µg each of anti-p50 (sc-8414 X, Santacruz Biotechnology, California, USA), anti-p65 (sc-8008 X) and IgG (as negative control, generated in-house) polyclonal antibodies. A 240 bp region flanking the κB and Sp1 elements in the LTR was amplified from the immunoprecipitated DNA using the primer pair N1054 (5'-GAAATTTAAAGTGGAAGTTGGACATTGAATTC-3') and N1056 (5'-AGAGACCCAGTACAGGCGAAAAGC-3'). A 300 bp PCR fragment targeting the cellular IκB-α promoter (N1057-5'GACGACCCCAATTCAAATCGTTC-3' and N1058 5'-TCAGGCTCGGGAATTTCC-3') was used as an internal control for NF-κB recruitment in the assay. Chromatin isolated from Jurkat cells in the absence of the retroviral infection served as a negative control (Data not shown).

We performed the ChIP analyses for RNA polymerase II and its phosphorylated form to confirm the recruitment of the transcription-competent complexes to the viral promoters. Immunoprecipitation was performed using 5 µg each of the antibodies specific to the
unphosphorylated RNA polymerase II C-terminal domain (CTD) repeat (YSPTSPS, ab5408, Abcam, Cambridge, USA), and S2-phosphorylated RNA polymerase II CTD repeat (phospho S2, ab5095). An IgG1 antibody (PP64B, Upstate, CA, USA) was used as an isotype negative control. The presence of the unphosphorylated form of RNA polymerase II on the viral promoter was confirmed using the primer pair N1054 and N1056. The presence of the S2-phosphorylated form of RNA polymerase II was confirmed approximately 3000 bp from the transcription start site (TSS) by amplifying a 300 bp PCR fragment in the TatP coding sequence using primers N1140 (5'-TCCAGTCCACAACCATGGATGGAGCCAGT AGATCCTAAC-3') and N1141 (5'-GGGCCCTCTGAGCTAAGTCGAAGGGGTCT GTCTC-3'). A 173 bp fragment of the cellular GAPDH promoter was amplified using primers N2014 (5'-TACTAGCGGTTTTACGGGCG-3') and N2015 (5'-TCGAACAGGGAGGAGCGAGCGA-3') as an internal control for S2-phosphorylation of RNA polymerase II. The recruitment of the NFAT transcription factors to the viral promoters was confirmed using the primer pair N1054 and N1056 and 5 µg each of anti-NFAT1 (ab2722), anti-NFAT2 (ab2796) or an IgG isotype control antibodies (PP64B, upstate, CA, USA) for immunoprecipitation. As a positive control for NFAT recruitment, a 147 bp fragment of the cellular TNF-α promoter containing an NFAT binding site was amplified using primers N2016 (5'-AGGATGGGGAGTGTGAGGGG-3') and N2017 (5'-CCTTGGTGGAACCCATGAGCTCATC-3') (30).

Construction of the reporter vectors - A dual-reporter vector pLTR-sLuc-IRES-EGFP that simultaneously expresses two different reporter genes - secreted Gaussia Luciferase (sLuc) and enhanced green fluorescent protein (EGFP) - was constructed as follows. Gaussia Luciferase was amplified with primers N712 (5’-CCAGGCGGAATTCACCATGGAGTCAAAGTTCTG-3’) and N713 (5’-GGCCGCGGATCTTATGCACCACCGCCGCCCTT-3’) from pCMV-Gluc (Nanolights, NanoLight Technologies, Panama, USA) and cloned directionally into pIRES-EGFP (BD Biosciences Clontec, NJ, USA) between the EcoRI and BamHI sites. The luciferase-IRES-EGFP cassette was moved to the pcDNA 3.1(+) vector directionally using the NotI and EcoRI sites. The CMV promoter was later replaced by the full-length HIV-1 subtype C LTR sequence, consisting of 656 bp, amplified from a patient BL42 (Gen Bank Acc No: HQ202921) using primers N698 and N854, between the restriction sites MluI and EcoRI.

One group of the promoter variant viral strains of subtype C of India contains four functional NF-κB binding sites in the viral enhancer referred to here as the κB viral strains as opposed to the standard subtype C viral strains that contain only 3 such binding sites designated here as the 3κB strains. Furthermore, the four NF-κB binding sites in the subtype C viral promoter fall into three genetically distinct types and for the sake of clarity we label them accordingly. The two genetically identical canonical κB-sites (5’-GGGACTTTCC-3’) found in all the HIV-1 strains are designated here as ‘H-κB’ sites. The Sp1-proximal variant κB-site (5’-GGGGCGTTC-3’) unique for subtype C alone, and not found in any other HIV or SIV, is designated here as ‘C-κB’ site. The fourth variant κB-site (5’-GGGACTTTCT-3’) found further upstream and inserted through the sequence duplication and the focus of the present work is designated here as the ‘F-κB’ (F for fourth) site. The LTR from the viral isolate BL42 was selected here as it represented all the subtype-specific molecular properties including the presence of the F-κB site. A panel of isogenic variant LTR reporter vectors was generated from this parental vector using the overlap PCR strategy. The 21 bp tandem duplicated sequence, consisting of the F-κB site, was deleted from the native viral promoter (FHHC-LTR, see the main text) to generate the HHC-LTR that represented the standard configuration of the subtype C LTR. Additionally, two different viral LTRs that contained H- or F-κB sites at all the four locations (HHHH and FFFF) were also constructed to examine the transactivation properties from the viral enhancers homogenous for the κB sites. In all the above 4-κB
LTRs, the sequence context flanking the individual NF-κB binding sites is identical to that of the original viral promoter. A null vector devoid of all the NF-κB binding sites was also constructed. Furthermore, for comparison, two additional reporter vectors were made containing viral promoters derived from NL4-3 (subtype B, accession number: M19921) or Indie C1 (subtype C, accession number: AB023804) molecular clones.

The analysis of the reporter gene expression - Jurkat cells were transiently transfected with the Lipofectamine 2000 transfection reagent (Cat. No. 11668-019, Invitrogen BioServices, Bengaluru, India). The cells were seeded into 48-well clusters at a density of 5 x 10⁶ cells/well in 400 µl of RPMI-1640 medium supplemented with 10% fetal calf serum. A plasmid DNA pool of 500 ng, containing 300 ng of one of the reporter plasmids, 100 ng of the pGL3 plasmid expressing Firefly luciferase (Cat. No. E1741, Promega Corporation, Wisconsin, USA) and 100 ng of the subtype C Tat-expression vector was prepared in 50 µl of serum-free RPMI medium. For the Tat-minus transfection, 100 ng of pcDNA3.1 (+) plasmid was used as carrier DNA. One µl of Lipofectamine was mixed with 49 µl of serum-free RPMI medium to prepare the lipid transfection reagent which in turn was mixed with 50 µl of the plasmid pool. The plasmid-lipid mix was incubated for 20 min at room temperature and then added to appropriate wells. Twelve hours following the transfection, the cells were washed to remove the lipid complexes and resuspended in 500 µl of the complete RPMI medium. For cell activation, transiently transfected Jurkat cells in a volume of 500 µl were incubated in the absence or presence of one or a combination of the activators: TNF-α (20 ng/ml; Cat. No: 210-TA-010, R&D Systems, Minneapolis, MN), Phorbol 12-myristate 13-acetate (PMA, 20 ng/ml, Cat. No: 8139, Sigma-Aldrich, St. Louis, Missouri, USA), PHA (5 µg/ml, Sigma-Aldrich, Cat. No: L9132, St. Louis, Missouri, USA) or anti-CD3 plus anti-CD28 antibodies (at 1:1 bead to cell ratio, Cat. No. 111.31D, Invitrogen BioServices, Bengaluru, India). BioLux Gaussia Luciferase Assay Kit (Cat. No. E3300L, New England Biolabs, Massachusetts, USA) was used to monitor the levels of the Gaussia Luciferase secreted into the culture supernatant at 24 h. The luciferase assay was performed using a SpectraMax L Luminescence 96-well Microplate Reader (MDS, Inc Model No: s/n Lu 03094, Sunnyvale, CA, USA) by mixing 25 µl of the culture supernatant and an equal volume of the 1 X BioLux GLuc substrate reagent. The experiments were performed in triplicate wells and every experiment was repeated at least two times. Transfection efficiency was monitored by measuring the expression of the Firefly luciferase in the cell extracts using a commercial kit (Bright-Glo Luciferase Assay kit, Cat. No. E2620, Promega Corporation, Wisconsin, USA) as per the manufacturer’s instructions. The primary data were normalized for the tranfection efficiency.

Construction of the viral molecular clones ‘isogenic’ for the F-κB site in the LTR - The viral molecular clone Indie-C1 was used for the construction of the ‘isogenic’ HHC and FHHC paired viruses. Except for the 22 bp difference in the viral promoter, the paired viruses contain identical genetic context. Such viral pairs are referred to as ‘isogenic’ viral pair for the sake of convenience. Throughout the manuscript we use the expression ‘isogenic’ to represent viral promoters or infectious viral clones that are genetically identical with the difference of containing or not containing the 22 residues constituting the F-κB motif in the viral enhancer. Using overlap PCR, the original LTR at the 3' end of Indie was replaced with the LTR derived from the primary viral isolate BL42 in three successive steps. This procedure also introduced an MluI site immediately upstream of the 3'-LTR, to permit subsequent LTR exchanges. First, the BlpI fragment of 4.1 kb, spanning the base pairs 4709 and 8840, was deleted from Indie and the backbone was self-ligated to make the BlpI site unique. Second, two different PCRs were performed, one to amplify the 230 bp fragment between the unique BlpI site and up to the beginning of the 3' LTR and the other to amplify the BL42 LTR. The first PCR was performed using primers N1075 (5'-GAGAGAATGAGACGAGCTGAGCCAG-3')
and N1077 (5'-ACGCGTCCCTCTTTTCTTTTAAAAAGAGCT-3') and the second PCR using primers N1076 (5'-GAAAAGGGGGAGCCTGGAAGGTTAA TTTACTC-3') and N1078 (5'-AGCTCCACCGCGGCATCTCP-3'). The internal primers N1076 and N1077 introduced the MluI site. The two PCR products contained a 17 bp overlap between them in addition to introducing the MluI site upstream of the 3' LTR. A third-round overlap PCR was performed using the first- and second-round PCR products for template and the primer pair N1075 and N1077. The product of the third-round PCR was cloned directionally between the BlpI and SacII sites of the Indie intermediate vector from step 1 above. Lastly, the original 4.1 kb BlpI fragment from Indie was restored into the BlpI site and the clones containing the insert in the correct orientation were identified using restriction analysis. As a consequence of the MluI site introduction, the recombinant clones contain an additional amino acid alanine (GCG) at position 98 in Nef. The recombinant viral clones were confirmed for p24 production in HEK293T cells and for infectivity in CEL5 cells.

Using a similar strategy, we engineered the 3' LTR of the EcoHIV molecular clone to generate viruses isogenic for the T-κB site. The EcoHIV molecular clone is basically a subtype B virus containing an LTR derived from NL4-3 at the 3' end. We substituted the B-LTR at the 3' end with that of BL42, a subtype C viral strain, as well as engineered an MluI site immediately upstream of the 3'PLTR. A 1,558 bp fragment spanning the XhoI and NaeI sites was assembled in three steps using an overlap PCR. A 188 bp fragment spanning the nef-LTR junction was amplified with primers N1333 (5'-GAGCAGTATCTAGACATGGAG-3') and N1335 (5'-TTAGCCCTTCCAGCGTCCTCCTTTTTT TAAAAGT-3') to introduce the MluI site. The full-length LTR of BL42 was amplified with primers N1334 (5'- AAAGGGGGGACCGCTGGAAGGCTAATTC ACTCCCA-3') and N1337 (5'-CCTCTAGCTAGCCCCGCGGTGCTAGAGAT TTTCCCACTAG-3'). A 736 bp fragment up stream of the 3'-LTR on the vector backbone was amplified with primers N1336 (5'-AAATCTTAGACCCGCCGCTAGCTAGG AGGTAGAGTTGCAG-3') and N1338 (5'-CCGACACAGGCGCTGATGACTTACT-3'). Successive overlap PCRs performed using the amplified products generated a 1,538 bp fragment that was cloned directionally between the XhoI and Nael sites. An additional amino acid alanine was introduced into Nef at the position 98 due to the engineering of the MluI site. The recombinant viral clones were confirmed for p24 production in HEK293T cells and for infectivity in CEL5 cells.

Preparation of the viral stocks - HEK293T cells were transiently transfected with different viral molecular clones using the standard calcium phosphate protocol (31). Cells were seeded in a 90 mm dish at low confluency and transfected with 10 µg of the viral plasmid DNA along with 0.2 µg of the CMV-EGFP expression vector - the latter as an internal control for the transfection efficiency. Culture supernatants were harvested at 72 h, passed through a 0.22 µ filter and stored in a deep freezer in multiple 1 ml aliquots. The p24 levels of the viral stocks were evaluated using a commercial ELISA kit (Cat. No: NEK050B, HIV-1 p24 ELISA kit, PerkinElmer Life Sciences, Massachusetts, USA). The infectious titre of the viral stocks was determined using TZM-bl cells. Briefly, 10^4 TZM-bl cells were seeded in 100 µl of DMEM in a flat-bottom 96-well culture plate. After 24 h, 100 µl of serially diluted viral stocks (a serial 4-fold dilution) were added to appropriate wells in complete DMEM medium supplemented with 10 µg/ml of polynucleotide. Following 2 h of incubation, the medium in the wells was replaced with 200 µl of complete DMEM medium, and the plates were incubated for 2 days at 37°C in the presence of 5% CO₂. To examine β-galactosidase expression, on day 3, the culture medium in each well was replaced with 100 µl of PBS and the plates were incubated for 5 min at room temperature. The cells were washed two times with PBS, 100 µl of freshly prepared β-
galactosidase staining solution (4 mM potassium ferrocyanide, 4 mM M potassium ferricyanide, 2 mM MgCl$_2$, and 1 mM X-gal in PBS) was added to each well and the plates were incubated for 2-3 h at 37°C. Plates were washed two times in PBS and the blue-stained cells were counted manually under a low-resolution microscope. The infectious units of each viral stock were determined by multiplying the cell-count with the dilution factor. Pseudotyped lentiviral vectors used in the ChIP analysis were packaged in HEK293T cells. Cells seeded in 90 mm culture dishes were transfected using the Ca$^{2+}$-phosphate transfection protocol with a total of 20 µg plasmid DNA pool consisting of 10 µg of pcLGIT reporter virus containing one of the variant LTRs at the 3' end (HHC, FHHC, FFFF or HHHH), 5 µg pMDLg/pRRE, 3.5 µg pVSV-G, and 1.5 µg pRSV-Rev. The infectious units of the pseudotyped viruses were determined as mentioned above.

The heteroduplex tracking assay (HTA) - The LTR-HTA consisted of a nested-PCR that amplified in the second-round of amplification the fragments of 340 or 362 bp spanning the U3PR region, from 3- and 4-kB viruses, respectively. The primer pairs N558 (5' TGGAAGGGTTAATTACTCTAAGGAAAAGGAAAGAGATCCT 3') and N424 (5' GACACCAARGAAGCYTTAGAYAARATAGA 3') were used in the first-round and N419 (5' GTATGTTGCTTCAAGCTAGTRCCAGTTGAA-3') and N1024 (5' TGTACTGGGTCTCTCTAGGTAGA-3') in the second-round. A homologous fragment of 336 bp amplified from NL4-3 (subtype B) LTR, using the primer pair N419 and N1024, was used as a probe to form differential heteroduplex complexes between the FHHC and HHC PCR fragments. When the heteroduplexes are resolved in a polyacrylamide gel, the identity of the two competing viral strains could be unequivocally distinguished (Supplemental Fig. 4) and quantitated using a phosphorimager. The reverse primer N1024 was end-labeled using 30 µCi of $\gamma$-32P dATP (Cat. No. LCP-101, Board of Radiation and Isotope Technology, India) and in combination with the unlabeled N419 forward primer was used for the amplification of the B-LTR fragment. The amplified PCR fragment was gel purified and used as probe in the HTA. For the HTA, the PCR products of the clinical samples were column purified (Cat. No: 28104, QIAquick PCR purification kit, Qiagen India, New Delhi, India) and the DNA concentration of each sample was determined using UV spectrophotometry as well as confirmed in agarose gel electrophoresis. A typical HTA reaction of 25 µl volume consisted of 100-125 ng of the amplified DNA hybridized to approximately 1,000 CPM of the B-LTR probe in the annealing buffer (100 mM Tris-HCl, pH 7.8, 100 mM NaCl, and 2 mM EDTA). The samples were incubated at 95°C for 3 min and snap-chilled by placing the reaction vials on wet ice. Three µL of HTA loading dye (50% Glycerol, 0.02 M Tris-Cl, 0.5 M DTT, 0.25% Bromophenol blue, 0.25% Xylene Cyanol) were added to each tube, the tubes were vortexed and centrifuged at 12,000 rpm for 1 min. The entire HTA reaction mix was applied to a Whatman chromatography paper, wrapped in plastic film, dried and scanned using a phosphor imager (FLA-5000, Fujifilm, Japan). The band intensities were quantified using the ImageJ software. The HTA analysis was slightly modified when this technique was applied to the clinical samples. Since the inclusion of mono-infection controls was not a possible option in the natural infection, unlike in the experimental models, we directly compared the band intensities to measure viral domination.

Viral proliferation and the pair-wise viral competition assay - CEM-CCR5 cells or PBMC (5 x 10^6) were infected with viral mixes representing different ratios of the 3- or 4-kB isogenic strains at an approximate MOI of 0.0001 (see schematic diagram, Figure 6B). The infectious titres of the viral stocks were determined as described above and the competing viruses were used at equal multiplicity of infection (MOI) (1: 1, 500 infectious units each) or one of the viruses at an MOI 10 times higher than the other (50 units of one virus and 450 units of the other). In addition, mono-infections were also included in the assays.
for comparison (500 infectious units/assay). The PBMC were CD8 cell-depleted and activated with PHA for 72 h prior to viral infection. The cells were incubated with the viruses in complete RPMI medium supplemented with 10 µg/ml of polybrene for 2 h at 37°C. Subsequently, the cells were washed three times in PBS and incubated in RPMI-1640 complete medium. The cells were monitored, the medium was replenished twice a week and fresh cells - activated PBMC from the same donor - were added to the cultures as required. The secretion of p24 into the culture medium was monitored periodically using a commercial kit (Perkin Elmer Life Sciences, Boston, MA, USA) and viral growth curve for each monoinfection was constructed. Two weeks following the infection, the genomic DNA was extracted from the cells using a commercial kit (QIAamp Blood Mini Kit, Cat. No. 69504, Qiagen India Pvt. Ltd., New Delhi, India) and 250 ng of the DNA was used in the LTR-HTA. For the EcoHIV mouse infection model, each animal, 4 female BALB/c mice per group, received intraperitoneally 5 µg of p24 equivalent of the virus generated in HEK293 cells. Mono- and dual-infections were established essentially as described above for the T-cell infection. Genomic DNA was extracted from the splenocytes two weeks following the viral infection. The relative replicative fitness of the competing viral strains was evaluated essentially as described previously (32) and as depicted (Supplemental Fig. 4A). Briefly, the replicative fitness of a viral strain was calculated as the ratio of the band intensities in a coinfection to that in the monoinfection which in turn was compared to the summation of the two ratios of the two competing viruses. For instance, production of the 3-κB virus was calculated to be the ratio of the band intensities of the coinfection ‘c’ to that of the monoinfection ‘a’. Relative fitness of the 3-κB viral strain ‘W₂’ was then expressed as the ratio of 3-κB band comparison (c/a) to the summation of 3- and 4-κB band ratios (c/a + b/d). A similar strategy was used to determine the relative fitness of the 4-κB viral strain ‘W₄’ which was expressed as the ratio of 4-κB band comparison (b/d) to the summation of 3- and 4-κB band ratios (c/a + b/d).

The analysis of the post-entry events of the viral infection - CEM-CCR5 (5 x 10⁶) cells were infected independently with 5,000 infectious units (MOI = 0.001) of 3- or 4-κB isogenic viruses pretreated for 1 h with 10 U/mL of DNase I (Cat. No: M0303S, New England Biolabs, MA, USA) and harvested at different time points for the evaluation. Cellular DNA was extracted using a commercial kit (QIAamp Blood Mini Kit, Qiagen India Pvt. Ltd., New Delhi, India). The reverse transcription products were detected 12 h following the viral infection using a primer pair N1734 (5'-TGTGTGCCGCTCTGTTGTG-3') and N1735 (5'-GAGTCCCTCGTCTAGGATC-3') that amplifies a 142 bp fragment from the U5-ω region. A total of 250 ng of the genomic DNA was used for the real-time PCR analysis using a commercial kit (SensiFAST SYBR mastermix kit, Bioline, London, UK). A standard curve for the PCR products was prepared using a 10-fold serial dilution of the plmdie-C1 plasmid ranging from 10⁸ to 1 copy, diluted using a salmon-sperm DNA solution (50 ng/µl). A real-time PCR for the detection of the 2-LTR-circles was performed 24 h following the viral infection using a primer pair N1736 (5'-TGGGTAGACCAGATCGAGCCT-3') and N1737 (5'-AGGGTTGACACTCCCAGTCCC-3') that amplified a 223 bp fragment spanning the U5-U3 regions in the LTR. A 1,268 bp fragment representing the 2-LTR circle target was generated using PCR from the Indie LTR. A serial 10-fold dilution of this fragment ranging from 1-10⁹ copies was used to construct the standard curve and the copy number of the 2-LTR circles was determined using the regression analysis. The analysis for the number of viral integration events was performed at 48 h following the viral infection using the nested Alu-PCR strategy. The forward primer N1739 (5'-AGCTAGGGAACCACGGTCTAAGC-3') used in the first-round of the amplification is located in the R region of the LTR and the reverse primer N1740 (5'-TGCTGGATTACAGGCGTCG-3') in the Alu repeat elements. The first-round of the PCR amplified a mixed population of PCR products depending on the distance between the proviruses that may have integrated randomly into...
various genomic locations and the Alu sequences dispersed throughout the genome. The inner primer set amplified a target sequence of 200 residues within the RPU5 region of the LTR using the primer pair N1720 (5'-GTTAGACCAGATCTGAGCCTP3') and N1741 (5'-GGTAGTGTGGAAAATCTCTAGCAGP3'). To generate a standard curve for the Alu-PCR, we infected HEK 293 cells with vesicular stomatitis virus G protein (VSV-G) pseudotyped LGIT virus containing the Indie-C1 LTR. The cells were infected at a high MOI and selected over a period of one month to ensure representation of diverse integration events. At the end of the selection period, the infected cell pool contained an integration incidence of 1.4 ± 0.6 copies per cell. The genomic DNA extracted from these cells was used to generate an integration DNA standard curve ranging from 1-10^3 copies/cell. The number of the viral integration events in the samples was determined by regression analysis using the standard curve. The magnitude of the viral transcription from the 3- or 4-kb viral promoters was evaluated at 48 h following the viral infection. Cells were activated with 100 ng/ml of TNF-α 1 h prior to harvesting or left without activation. Total cellular RNA was extracted from 1×10^6 cells using the RNeasy Plus kit (Cat. No. 74124, Qiagen India, New Delhi, India). The first-strand cDNA was synthesized using the SuperScript II Reverse Transcriptase (Cat. No. 18064-014, Invitrogen BioServices, Bengaluru, India). Following this, two different PCRs were performed one for the proximal viral transcripts within the TAR element and the other for distal transcripts in Tat located approximately 5.4 kb downstream of the transcription start site (see schematic diagram, Fig. 7A). Proximal viral transcripts were detected with the primer pair N1720 (5'-GTTAGACCAGATCTGAGCCTP3') and N1721 (5'-GGTAGTGTGGAAAATCTCTAGCAGP3') that amplified a 89 bp fragment in the TAR element. The primer pair N1728 (5'-TTGCGACAGAGAAGAGCAAGP3') and N1729 (5'-GATACTTACTGCTTGTAGATP3') amplified a distal viral transcript of 226 bp in Tat. Transcription from the cellular gene β-actin was used for the gene expression normalization of the viral promoters using a primer pair N1730 (5'-GTCGACACGGCTCCGGC-3') and N1731 (5'-GGTGTTGGTGCCAGATTCTC-3') that amplified a 239 bp fragment. Quantitative real-time PCR for each primer pair was performed using the Comparative C_t (ΔΔC_t) strategy and using the Corbett Rotor-Gene 2000 Real-Time PCR machine (Corbett Lifescience, Hilden, Germany).

**Determination of the CD4 T-cell count and plasma viral load** - The CD4 cell count was determined in fresh blood samples within 4 h from the time of sample collection. The counts were determined using MultiTest antibodies, counting beads (340499 and 349480) and the Lyse-No-Wash protocol as recommended by the manufacturer (BD Biosciences, CA, USA). The samples were acquired on a BD FACSCalibur flow cytometer. The plasma was separated from the whole blood within 6 h of collection and stored in a deep freezer as 0.5 ml aliquots. Plasma viral load (PVL) was determined using a commercial kit (NucliSens EasyQ v1.1.1.1, bioMerieux, France) and PVL below the detection limit of <25 IU/ml, was assigned a value of 25 IU/ml. A plasma sample of known viral load and a sample negative for the virus were included in each run and the assay was validated only if the internal controls gave consistent results. Samples were thawed only once and used immediately for viral load estimation.

**RESULTS**

**Rapid expansion of variant viral strains in India and worldwide** – Previous analysis form our laboratory identified five different types of sequence insertion in the subtype C viral promoter between the transcription factor binding sites RBEIII and NF-κB (26). The sequence insertion generated an additional binding site for the transcription factors NF-κB or RBEIIIPlike motifs). Samples (n = 607) for the previous analysis were collected nearly a decade ago between 2000 and 2003 from all the four southern states of India (33). The
prevalence of 3 of the 5 promoter variant viral strains containing the insertions of NF-κB, κB-like and RBEIII sites was 2, 1 and 1%, respectively, in 2000-2003 (Fig. 1A). To identify the contemporary prevalence of the subtype C LTR variant viral strains in a cross-sectional analysis, fresh blood samples were collected from four different clinics (three in southern and one in northern India) in 2010-2011 (Table S1). Sequences of the U3 region or the full-length LTR were determined from plasma viral RNA and proviral genomic DNA, respectively. The results indicated a rapid expansion of all three LTR variant viral strains at all the four clinics (Fig. 1A). Of the variant strains, the 4-κB viruses expanded at a faster rate with their prevalence increasing dramatically from 2% in 2000-2003 to as high as 20-30% in 2010-2011. Using a different and well-defined clinical cohort (27), under which several clinical samples were collected from a different clinic (Seva Free Clinic, Bengaluru) in 2005, we found that the prevalence of each of the 3 variants was intermediate, with a level of 5, 2 and 4% prevalence for κB-, κB-like and RBEIII insertions, respectively (Fig. 1A).

Importantly, near full-length sequences of 256 viral strains from South Africa, collected between 2000 and 2005, have been deposited in the databases (34), thus providing an opportunity to examine the molecular nature of the insertions in a different clinical cohort dominated by subtype C. We found that the South African samples, as those of India, contained all three types of insertions in the viral promoter (Fig. 1B). Among the insertions, κB, κB-like and RBEIII site insertions constituted of 4.3, 5.1, and 8.2% of the total number of samples, respectively. Furthermore, an analysis of the global sequences of HIV-1 available in the databases identified the κB-site insertions from other countries including China (n=2/81), South Africa (n=16/315), Tanzania (n=11/58), Botswana (n=3/72), Zambia (n=2/52), and Spain (n=2/117) (Fig. 2A). These data collectively suggested that the emergence of 4-κB viruses is a universal phenomenon applicable to subtype C regardless of the geographical location. Since the 4-κB strains of subtype C demonstrated the fastest expansion rate in India, we next characterized the molecular and biological properties of this HIV-1 variant.

The new NF-κB site characterized by a unique genetic variation is exclusive for subtype C - Multiple sequence alignments of a subset of the representative 4-κB containing C-LTR sequences of India and several similar sequences downloaded from the databases revealed several molecular properties unique to subtype C (Fig. 2A). Several of the Indian LTR sequences used for this analysis have been previously reported (26). A 4-κB subtype C promoter contains three types of genetically distinct κB sites, and for clarity they are labelled accordingly. The two genetically identical canonical κB-sites (5'-GGGACTTTCC-3') found in all the HIV-1 strains are designated here as ‘H-κB’ sites. The Sp1-proximal variant κB-site (5'-GGGGCGTCC-3') unique for subtype C alone, and not found in any other HIV or SIV, is designated here as ‘C-κB’ site. The fourth variant κB-site (5'-GGGACTTTCT-3') inserted through MFNLP is designated here as the ‘F-κB’ (for fourth) site. A careful examination of the sequence alignment revealed a tandem duplication of 22 residues, consisting of 9 residues of the H-κB II site and 12 upstream residues, inserted immediately upstream of the viral enhancer (Fig. 2A). The original and the duplicated sequences were separated by a single ‘T’ residue that constituted the 10th position of the newly created F-κB site.

The insertion of specific sequences in the viral promoter has been studied extensively in the context of the subtype B LTR (22), but not in any other viral subtype including subtype C. Importantly, sequence duplication of the NF-κB binding site has not been reported previously in any HIV subtype including subtypes B and C although a few LTR sequences have been found in the extant databases. To examine if an association existed between the nature of the viral subtype and the duplication of any of the three transcription factor binding sites (NF-κB, NF-κB like or RBEIII), we performed an extensive search and downloaded several thousands of the LTR sequences belonging to the major genetic subtypes of HIV-1 available in the databases. Of the total
3,054 full-length LTR sequences belonging to all the HIV-1 genetic subtypes from the databases or derived through the present work, 479 (15.7%) contained sequence insertions up stream of the viral enhancer. We determined the percent prevalence of the different MFNLP in each of the viral subtypes and found that the acquisition of the F-κB site is an exclusive property of subtype C (Fig. 2B). Of the total 1,113 sequences in subtype C, 213 sequences (19.1%) contained the insertions of which 81, 56, and 76 sequences were characterized as containing the F-κB (7.3%), κB-like (5.0%) and RBEIII (6.8%) sites, respectively. In contrast, viral subtypes A, B, D, F, and group O predominantly contained the RBEIII site insertions. Subtype G, the recombinant CRF AG_02, and group N contained nearly equivalent numbers of RBEIII and κB-like or HP-κB site insertions although the sample size available for the analysis was small. The insertion of the 22 residue sequence and the creation of the C-to-T variation in the F-κB site could be a consequence of the propensity of subtype C reverse transcriptase for sequence duplication and the non-template mediated base addition at the growing end of the viral DNA in a RNA:DNA hybrid molecule.

The F-κB site is biologically functional and recruits p50:p65 heterodimer - To understand if the genetically variant F-κB site is biologically functional in the context of the viral promoter and recruits NF-κB, we used EMSA. Using radiolabeled double-stranded DNA probes representing the canonical H-κB site or the variant F-κB site, we asked if any cellular factors from Jurkat cell extracts were bound to the DNA probes. Cell extracts were prepared from Jurkat T-cells with or without TNF-α activation for 60 min. Two distinct complexes were found under control conditions, and these complexes were enhanced several fold following TNF-α activation (Fig. 3A, lanes 2 and 3). These specific complexes were out competed with 25-fold molar excess of cold probes representing H-, F-κB or a κB-site derived from the IκB cellular promoter (lanes 4-6), but not with a mutant F-κB probe (lanes 7). The nature of the complexes was comparable between H- and F-κB probes suggesting similar function regardless of the variation. Furthermore, supershift analysis using affinity-purified rabbit antibodies identified the presence of p50-p65 heterodimers in the complexes but not other members of the rel family (Fig. 3B). Probes containing swapped flanking sequences (H* and F*) produced comparable results thus confirming that the flanking sequences did not influence NF-κB binding to the probes (Supplemental Fig. 1A). Importantly, densitometric analysis of the supershifted p50 and p65 band intensities revealed comparable intensities between H- and F-κB probes suggesting similar binding affinities (Supplemental Fig. 1B).

To compare the binding profiles of NF-κB to the viral enhancer in the chromatin context of the integrated provirus, we performed a ChIP assay. Given the close proximity of the genetically distinct and multiple κB-sites in the viral enhancer and the repetitive nature of the sequences flanking the F-κB site, it was practically not possible to position primers for the ChIP analysis. To circumvent this problem, we generated two different HIV-1 reporter viruses, based on a previously reported pcLGIT vector (35). In these reporter viruses, the viral enhancer in the 3' LTR was engineered to contain four tandem repeats of the H- (HHHH) or F-κB site (FFFF) with the spacer sequences and other molecular features maintained and comparable to the wild type 4-κB viruses (Fig. 3D, top panel). Daughter viruses produced from these vectors would copy the 3’ U3 into the 5’ viral promoter thus placing the expression of the reporter gene EGFP and that of the viral transactivator protein Tat under the control of the engineered LTRs. The nature of the NF-κB binding to the viral enhancers could therefore be compared in a homogenous κB-site context following viral integration into the chromatin. Importantly, the HHHH- and FFFF-
LTRs are functional and drive the expression of a reporter gene with efficiencies comparable to that of the wild type FHHC-LTR (see below). VSV-G env pseudotyped viruses prepared in HEK293 cells were used for the infection of Jurkat T-cells at an MOI of 1 and 72 h later, the cells were stimulated with TNF-α for 60 min or left without activation. Fragmented chromatin complexes were precipitated using antibodies specific to p50, p65, or unphosphorylated C-terminal repeat domain (CTD, Y1S2P3T4S5P6S7) of RNA polymerase II (36). Isotype-matched control antibody was used as a negative control. An amplification was performed to detect and quantitate the presence of the transcription-competent complexes along with the NF-kB on the viral (HHHH or FFFF-LTR) or a control cellular IkB-α promoter (Figure 3D). In the absence of TNF-α activation, only p50, but not p65, was found recruited to the IkB-α, H-kB, and F-kB enhancers (lanes 2 and 3, Fig. 3D). Following cell activation, however, recruitment of both p50 and p65 was evident at all the promoters, at levels comparable between H- and F-kB enhancers (lanes 7 and 8). Likewise, the unphosphorylated form of RNA polymerase II complex was detected following TNF-α activation on both the viral as well as the cellular promoters (compare lanes 4 and 9). The PCR amplifications above targeted a region proximal to the TSS on both the viral and cellular promoters. The Ser 2 phosphorylated form of RNA polymerase II complex was detected on the viral promoters following TNF-α activation by amplifying a region approximately 3000 bp downstream of the TSS (lanes 15). The complex was also detected on the GAPDH promoter, a constitutive cellular promoter (36) regardless of the cellular activation status (lanes 12 and 15). Furthermore, end-labeled H- or F-kB DNA probes recruited NF-kB from Jurkat cell nuclear extracts at comparable affinities in southwestern blot analysis (data not shown). These results collectively suggested that the F-kB site regardless of the variation at position 10 behaves functionally in a manner identical to the canonical H-kB site with respect to NF-kB binding.

In the absence of Tat, all the viral promoters were responsive to extracellular stimulation with the response to PMA being the strongest. FHHC-LTR consistently expressed significantly higher levels of the reporter gene under all the conditions of cell activation as compared to HHC-LTR (Fig. 4B, p<0.001). Importantly, HHHH-LTR, but not FFFF-LTR, performed at efficiencies comparable to that of FHHC viral promoter. The under-performance of the FFFF-LTR, however, was the consequence of the loss of the overlapping NFAT site due to the C-to-T variation (Supplemental Fig. 2). In the presence of Tat, the profile of reporter gene expression essentially remained the same except that the magnitude increased by 5- to 10-fold (Fig. 4C). With the exception of the FFFF-LTR, the gene expression pattern of the other viral
promoters was quite comparable when the cells were exposed to combinations of two different activators (Fig. 4D). Nevertheless, when compared to the null vector, the FFFF-LTR demonstrated significant magnitude of reporter gene expression ($p<0.001$). Collectively, these data proved quantitative gain-of-function by the C-LTR due to the acquisition of the F-κB site. The gain although moderate was statistically significant. It also appears that the natural context of κB-site genetic variation and the order in which these sites are arranged have biological significance.

To determine whether the genetic difference between H- and F-κB sites could lead to qualitative gain-of-function, a few signalling pathways of the T-cell activation were examined using small molecule inhibitors. To this end, three different small-molecule inhibitors Rotterlin, Stauorospine, and G06976 were used to intercept T-cell signalling mediated through different protein kinases (38). Jurkat cells were transfected with the LTR reporter vectors, activated with PMA in the absence or presence of one of the inhibitors, and the expression of luciferase was monitored (Supplemental Fig. 3, left panel). All the viral promoters demonstrated enhanced transactivation following activation, and significant reduction in the presence of the inhibitor molecules. The pattern of gene modulation was comparable between FHHC and HHC-LTRs except for the quantitative differences. Likewise, TNF-α induced reporter gene expression was efficiently blocked by leptomycin in a manner identical to all the variant viral promoters including FHHC and HHC or HHHH (Supplemental Fig. 3, right panel). Collectively, the inhibition of cell signalling in Jurkat cells or HEK293 cells (data not shown) failed to identify notable differences between H- and F-κB sites. However, given that HIV-1 is capable of infecting a wide range of target cells in the natural context, a more thorough analysis would be required to evaluate the biological significance of the C-to-T variation in the F-κB site.

The subtype C 4-κB viral strains out-compete the 3-κB ‘isogenic’ counterparts in experimental systems and in the natural infection - To examine replication fitness of the 4-κB viral strains, using a standard subtype C molecular clone Indie-C1, paired infectious viruses, with or without the F-κB site in the enhancer but otherwise having an isogenic background (hereafter referred to as ‘isogenic’), were generated (Fig. 5A, top panel). Replication kinetics of the two viruses in CEM-CCR5 T-cells (Fig. 5B) or in PBMCs (Fig. 5C) from several healthy donors demonstrated the superiority of the 4-κB viral strain at every time point, and this difference remained significant throughout the study period ($p<0.001$). Furthermore, the replicative fitness of the two viral strains was compared in pair-wise competition assays at different ratios with respect to the MOI as depicted schematically (Supplemental Fig. 4A). The HTA for the LTR was used to monitor the competing viral strains essentially as described previously for HIV-1 env (32). The LTR-HTA consisted of a nested-PCR that amplified a 340 or 362 bp fragments in the second-round of the PCR comprising of the viral enhancer, from the 3- and 4-κB viral strains, respectively (Supplemental Fig. 4B). The PCR fragments were hybridized to an end-labeled 330 bp DNA probe amplified from an identical location from subtype B LTR of the NL4-3 virus. When the hetero-duplexes are resolved in a polyacrylamide gel, the identity of the two competing viral strains can be unequivocally distinguished (Supplemental Fig. 4C), and quantified using phosphorimaging; the relative proliferation of each virus was then calculated as depicted (see mathematical formula, Supplemental Fig. 4D). Of note, the detection strategy of HTA employed here eliminates the possible experimental artifacts at two different levels. First, the primer pairs bind identical sites on both of the competing viruses thus eliminating amplification differences. Second, the same probe binds both of the amplified PCR products under identical experimental conditions in the same vial thus eliminating hybridization differences. Thus, the HTA detection format offers a greater level of confidence to the data in not discriminating between the competing viral strains in addition to being a quantitative assay (Supplemental Fig. 6). In pair-wise competition assays, at an equivalent MOI, the 4-κB virus out-competed the 3-κB counterpart at the second week in CEM-CCR5.
cells (Fig. 5D) or PBMCs (Fig. 5E), and the differences were statistically significant (p<0.05 by two-tailed student t test). When PBMC were infected at an equal MOI (1:1), and examined using HTA 48 h after the infection, the magnitude of infection by 3- and 4-κB viral strains was comparable suggesting that the 4-κB virus did not dominate its counterpart at the early step of viral entry but only at the later stages (Supplemental Fig.5).

The pair-wise competition experiments were also performed in an experimental animal model, the EcoHIV-mouse system (39). In EcoHIV which predominantly represents the prototype subtype B NL4-3 molecular clone, HIV-1 gp120 was replaced by gp80 of the ecotropic murine leukemia virus (MLV), a retrovirus that infects only rodents. The chimera virus EcoHIV can productively infect immune-competent mice and establish spreading viral infection that could be detected and quantified using a real-time PCR. EcoHIV challenge system offers a low-cost experimental animal model for evaluating anti-retroviral drugs and examining vaccine efficacy (40,41). The EcoHIV clone was further engineered to replace the original subtype B LTR at the 3’ end with subtype C LTRs ‘isogenic’ for the F-κB site (Fig. 6A). Each animal received intra-peritoneally 5 µg of p24 equivalent of the virus generated in HEK293 cells. Mono- and dual-infections were established essentially as described above for the T-cell infection. Two weeks following the viral infection, genomic DNA was extracted from the splenocytes and the HTA detection strategy was used to determine fitness differences between the viral strains in the in vivo conditions. Viral infection was typically peaked around week 2 hence we examined viral fitness at this time point. As in the in vitro system, the 4-κB virus out-competed the 3-κB counterpart when both viruses were inoculated into the mice at equivalent p24 concentrations (Fig. 6B, p<0.001). The data from both in vitro and in vivo models collectively confirmed that the 4-κB virus exhibits superior replication fitness due to the acquisition of the F-κB motif. Importantly, given the ‘isogenic’ nature of the competing viruses, the replication advantage could be entirely ascribed to the presence of the F-κB-site in the viral enhancer.

The HTA detection strategy can efficiently differentiate between the 3- and 4-κB viral strains in a mixed infection. Taking advantage of the HTA assay, several clinical samples that contained mixed infections with both 3- and 4-κB viruses were examined to determine which of the two viruses would be the dominant strain in natural infections. The HTA analysis was performed as described above, however, with one difference. Since mono-infection controls were not a possible option in the natural infection, unlike in the experimental models, the mathematical formula could not be applied. We, therefore, directly compared the band intensities of the competing viral variant strains to measure viral domination. The direct comparison of the band intensities must faithfully represent the natural distribution of the competing viral strains in the clinical sample given the uniform hybridization of the probe to either of the viral LTRs in the HTA (supplemental Fig. 5 and 6). In a subset of 6 samples randomly selected from the 2000-2003 cohort, in assays for proviral DNA, the band intensities representing the 4-κB viral strains were significantly higher than those representing the 3-κB viruses (Fig. 6C, p<0.001). Importantly, in each of the clinical samples, the 4-κB viral variant was the dominating partner. From a subset of three clinical samples, we cloned the 3- and 4-κB LTRs into plasmid vectors, determined the sequences and examined the phylogenetic relationship between the variant viral pairs. The 3- and 4-κB viral sequences clustered together based on the identity of the clinical sample confirming distinct identity (Supplemental Fig. 7). Additionally, this analysis also ruled out the possibility of the 3-κB amplification being a laboratory generated contamination. In the case of a contamination, all the 3-κB sequences are expected to cluster together in the phylogenetic analysis. Viral amplification from the genomic DNA represents the viral species archived in the proviral compartment but not the active virus present in the plasma. We therefore extended the HTA analysis to plasma viral RNA extracted from three clinical samples available from the 2005 JNCASR cohort. The 4-κB virus was found to be
the dominant virus in the plasma viral RNA in all three clinical samples (Fig. 6D, left panel, p<0.001). Additionally, from one of the subjects S189, where two plasma samples were available one year apart, the 4-κB virus remained dominant over this period suggesting that the replicative difference established between the two viruses is stable over extended periods (Fig. 6D, right panel).

The 4-κB LTR manifests superior magnitude of transcription initiation and elongation - NF-κB plays a critical role in regulating basal level transactivation from the viral promoter in the absence of Tat and overall gene expression in its presence. Acquisition of an additional and functional κB-site therefore, must confer a significant replicative advantage on the 4-κB viral strains of subtype C. To understand at what level of the viral life cycle, the 4-κB viral strains might achieve replication advantage, CEM-CCR5 T-cells were infected independently with HHC or FHHC ‘isogenic’ viral strains (Fig. 7A) and viral proliferation was compared at four different stages using real-time PCR. Equivalent TCID₅₀ units of the viruses were used for cell infection. We compared the generation of the reverse transcription products in the cell extract, the two-LTR circle formation in the nucleus, the extent of viral integration using Alu-PCR, and the formation of proximal versus distal viral transcripts, as depicted schematically (Fig. 7A). No significant differences were found between the two viruses at the level of the reverse transcription (Fig. 7B), nuclear translocation (Fig. 7C), or proviral integration (Fig. 7D). A significant difference, however, was evident at the level of transcription, at both transcription initiation and elongation (Fig. 7E). Two different PCRs were performed; one for the proximal viral transcripts within TAR and the other for distal transcripts in Tat located approximately 5.4 kb downstream of the transcription start site (see the schematic diagram, Fig. 7A). Significantly higher levels of viral transcripts, from both proximal and distal transcripts, were generated from the 4-κB LTR from cells in the absence of activation suggesting stronger basal level promoter activity. TNF-α activation of the cells induced 1 or 2 orders of magnitude higher gene expression from both the viral promoters; however, the 4-κB LTR manifested significantly higher transcription as compared to the 3-κB LTR. Overall, the data suggested that the acquisition of an additional κB-binding site by the HIV-1 LTR significantly enhanced promoter strength. In summary, transcription enhancement from the viral promoter containing four κB-binding sites was improved at two different levels, transcription initiation and elongation, but not at other phases of the viral life cycle.

Higher plasma viral load and comparable CD4 cell count are associated with the 4-κB viral infection - If the 4-κB LTR can generate more viral transcripts, then, HIV-1 infections with the 4-κB strains may lead to higher plasma viral loads as compared to the 3-κB viral infections. To test this hypothesis, plasma viral loads and CD4 cell count were measured at a single time point in 60 and 20 subjects of 3- and 4-κB viral infections, respectively. Natural infection with the 4-κB strains contained evidently higher mean plasma viral loads as compared to those with the 3-κB viruses (Fig. 8A), and the difference between the means was found to be statistically significant (p<0.01). Of note, there was considerable overlap between the two groups, and data scatter was indeed large. Importantly, a similar comparison of the CD4 cell count failed to detect a statistically significant difference between the two groups (Fig. 8B). A significantly higher magnitude of plasma viral loads in the 4-κB infections is suggestive of greater probability of viral transmission since viral transmission is directly correlated to plasma viral load (42,43).

DISCUSSION

Positive selection of variant viral strains captured at the population level - Current results suggest the emergence of at least three different and novel viral strains of subtype C in India, South Africa and probably other geographical locations where this viral subtype is dominant (Fig. 1). The common theme that underlies the emergence of the variant viral strains is the acquisition of an additional TFBS, primarily an NF-κB or an RBEIII motif, through sequence duplication. The
acquisition of the RBEIII site has been studied extensively in the context of subtype B (22). Furthermore, our analysis shows that RBEIII duplication is universal among all the HIV-1 genetic subtypes including subtype C (Fig. 2B). In contrast, the acquisition of the NF-κB (F-κB) site appears to be a property exclusive to the subtype C virus and has not been reported previously. Importantly, duplication or deletion of diverse TFBS in the viral promoter, including that of NF-κB and Sp1, by itself is not a novel theme. A large number of publications have previously reported such LTR modifications in HIV-1 (10,12,44-52), HIV-2 (53) and SIV (54,55). However, unlike previous reports, our study provides experimental evidence that the new viral strains gained a potential selective advantage as a consequence of the acquired TFBS, and importantly that these strains have been expanding at the population level. To the best of our knowledge, this is the first report demonstrating a variant viral strain (4-κB or RBEIII duplication variant) progressively spreading at the population level and replacing the canonical parental strains (3-κB subtype C strains).

The data presented here are limited by a few technical issues primarily the quality and sample size of the clinical samples and sequences. Most of the analyses performed here were based on the sequences primarily generated through this work and sequences available from extant databases. Given this limitation, our results could be considered only as inferential evidence suggesting a positive evolutionary selection of the variant viral strains. However, our work embodies the data collected over a period of a decade. Furthermore, a larger number of viral promoter sequences belonging to the variant viral strains were generated through the present work than available in the extant databases. Additionally, several observations support the progressive expansion of the variant viral strains. First, in the city of Bengaluru, a gradual increase in the prevalence of the variant viral strains was observed spanning across a decade or longer using samples collected at three different time periods 2000-2003, 2005 and 2010-2011. Second, the prevalence of the variant viral strains reported from South Africa during 2000-2005 was comparable to that of India during the same approximate period. Third, the prevalence of the variant viral strains during 2010 and 2011 was comparable at four different clinics in three different cities of India suggesting a uniform distribution of the variant viral strains across the country.

The F-κB site is biologically functional in the viral context - We confirmed the biological function of the F-κB site using a range of experimental formats. The presence of the F-κB site conferred a quantitative gain-of-function on the viral promoters in the reporter gene expression analyses (Fig. 4). Additionally, binding of p50-p65 heterodimer to the F-κB site was unequivocally demonstrated in the supershift (Fig. 3B), ChIP (Fig. 3D) and South-western blot assays (data not shown). However, in the context of the artificial viral promoter FFFF-LTR, the F-κB site demonstrated significantly low magnitude reporter gene expression as compared to the H-κB site in the HHHH-LTR (Fig. 4), suggesting that the F-κB site cannot be a functional substitute for the H-κB site in the natural viral promoter. Using ChIP analysis we further demonstrated that the diminished strength of the F-κB site in the FFFF-LTR is the consequence of the absence of NFAT recruitment to this element due to the ‘C to T’ substitution at position 10 (Supplemental Fig. 2). It therefore appears that the F-κB site confers a quantitative-gain-of-function on the transcriptional strength of the viral promoter by the selective recruitment of NF-κB, but not NFAT. The data presented here prove that the F-κB site is fully functional in the context of the viral promoter and confers a gain-of-function advantage on the variant viral strains.

A little controversy surrounds the functionality of a different NF-κB site in the subtype C promoter which is located proximal to the Sp1 sites and referred to here as the C-κB site (5'-GGGGCGTTCC-3'). Using radiolabeled DNA probes and gel-shift analyses, a couple of publications previously claimed that the NF-κB site in subtype C LTR may not be functional. Unfortunately, neither of the publications used
authentic sequences with the natural flanking sequences. For instance, Lemieux et al used a double-stranded synthetic probe that differed from the authentic sequence at two different sites (5'-GGGCGGTTCT-3', differences underlined) (56). Likewise, Roof et al used a probe containing two tandem sequences of the NF-κB site that are separated by a single ‘T’ residue (57). Importantly, the probe containing the C-κB site sequences was not used directly for the probe-binding but only in the cold competition of the other probes. Given the technical limitations, it may not be appropriate to conclude that the subtype C specific NF-κB sequence is not functional especially in the natural context. In contrast to the above reports, two other publications demonstrated biological functioning of the C-κB site. While Naghavi et al demonstrated strong binding of the cellular factors from HeLa nuclear extracts to the DNA probes (10), Montano et al showed an efficient competition between the labeled canonical NF-κB probe and the cold C-κB sequence in the gel shift assay (12). Furthermore, in an extensive analysis using protein-binding microarrays and diverse NF-κB binding sequences, Siggers et al demonstrated that a sequence nearly comparable to the HIV-1 subtype C unique C-κB probe (5'-GGGCGGTTCC-3') is endowed with a potential to bind NF-κB (58). Lastly, in our hands, the C-κB site efficiently binds NF-κB from the T-cell nuclear extracts (data not shown).

Although the EcoHIV mouse model is a highly artificial system because multiple differences exist between the human and mouse hosts, viral proliferation could be reproducibly demonstrated in this setting. The EcoHIV virus used in the competition assays is essentially an HIV molecular clone NL4-3 the native envelope of which was substituted with that of a mouse virus MLV. In other words, the chimera virus EcoHIV is like the SHIV endowed with its own merits. One of the advantages of EcoHIV is that an immunocompetent mouse strain (eg. regular BALB/c or C57BL) could be used for the study. This animal model therefore offers an inexpensive, simple and powerful in vivo system for certain applications especially in resource-constrained countries where access to humanized mouse or primate models is a practical problem. In addition to the two experimental models used here, importantly, the dominance of the 4-κB variant viruses was consistent in the clinical samples of co-infection that appears to remain stable over extended periods (Fig. 6C and D). The 4-κB viruses clearly dominated the 3-κB counterparts in all the clinical samples tested, in both the proviral compartment and in the plasma virus. The data from the natural infection and the experimental models performed in the NF-κB ‘isogenic’ context collectively present a strong case that the fitness advantage of the 4-κB viral strains could be attributed to the additional κB site insertion in the viral enhancer. Furthermore, extension of these observations to the canonical subtype C viral strains implicates a critical role the additional NF-κB site (the C-κB site) may have played in the past moderate, but not profound, differences in the replication fitness drive viral evolution - We measured biological differences between the κB variants using a panel of reporter vectors or paired viruses that were constructed in an ‘isogenic’ background. In all these experiments, reporter gene expression analyses and pair-wise viral competition assays, we found modest, often 1.5 to 2 fold differences, between the 4- and 3-κB viruses with the former typically dominating the latter. Several research groups previously reported such modest differences in reporter gene expression when comparing viral promoters of different subtypes (56). Importantly, the positive selection of the variant forms of the pathogenic organisms is typically driven by small, not profound, differences (59). As a function of time, the minor differences per replicative cycle of the competing viral strains will result in the measurable replicative advantage for one of the viral strains. The data presented in our work are statistically significant and offer a convincing explanation for the positive selection of the 4-κB viral strains in the natural infection (Fig.6C and D). Importantly, even in a highly artificial in vivo model of EcoHIV, the domination of the 4-κB viral strains is sustained (Fig. 6B) essentially corroborating the in vitro data obtained using the T-cell lines and PBMC (Fig. 5) which are the acceptable standards in the field currently.
for the successful expansion of the 3-κB containing subtype C strains in comparison to the 2-κB containing strains, although currently the latter viruses were not found.

*The replicative fitness of the 4-κB viral strains probably confers a higher transmission rate* - The analysis of the correlation between the 4-κB strains with viral load and CD4+ T-cell count in patient samples should be interpreted with a word of caution as the patient samples are derived from a single time point only due to sampling limitations. Despite the sampling limitations, the 4-κB viral strains appear to have gradually increased in prominence over a decade or longer in India (Fig. 1). Successful expansion of the 4-κB strains at the population level could be the outcome of the relatively higher plasma viral load of these variant viruses in the body fluids (Fig. 8A); and higher plasma viral loads are directly correlated with enhanced transmission efficiency (42,43). Recent studies demonstrated that HIV-1 transmission to a new host involves the transmission of a single or as few as 2-5 viral strains in the majority of cases, even though the donor may contain a large number of genetically distinct viral strains (60,61). The κB variant strains sharing the same env (data not shown) are expected to maintain identical biological properties including cell tropism, preferred route of transmission and target cell populations. Given the higher plasma viral load of the 4-κB viral strains in a mixed infection, the probability of transmission of the 4-κB strain to a new host is likely to be superior to that of the 3-κB strain. Nevertheless, given the sampling limitations, we have been cautious not draw unwarranted inferences from the data.

*Is a stronger viral promoter essential for successful viral expansion?* - Increased gene expression from the viral promoter is expected to cause at least two different problems for viral fitness potentially offsetting the advantages gained by higher transmission rate. First, a higher magnitude of gene expression and enhanced viral load is expected to cause enhanced immune activation, although only an indirect association between these factors has been shown (62). Second, an increased number of the NF-κB sites in the viral promoter should be disfavoured for the establishment and maintenance of viral latency given the profound influence this transcription factor has on viral gene transcription (63,64). It therefore appears paradoxical that subtype C virus opted for a perplexing strategy of strengthening its promoter to increase replicative competence. This paradox perhaps could be resolved if subtype C virus indeed represents an advanced form of viral attenuation (65,66). Whether or not subtype C represents an evolved form of viral attenuation is highly controversial, and substantial evidence is lacking to support such hypothesis. Likewise, it is also not known whether a higher magnitude of viral attenuation is essential for positive evolutionary selection of the variant viral strains. In spite of these limitations, experimental proof examining the molecular basis underlying phenotypic differences between viral genetic subtypes has been gradually emerging in recent years (67-69). Subtype-dependent differences in cytokine induction profile have been demonstrated (70-72). We would like to propose that the relatively less virulent nature of subtype C could offer this subtype a window of opportunity to accommodate relatively higher levels of gene expression, as has been demonstrated here by increased plasma viral load (PVL), without significantly increasing its virulence - as suggested by the absence of a difference in the CD4 cell count (Fig. 8C). We caution that the data presented here are from a cross-section analysis which needs confirmation with prospective studies; nevertheless, the findings are consistent with the 4-κB viral strains gaining replication advantage at the population level as a consequence of their higher infectivity. At present, it is not known whether the 4-κB viruses are more pathogenic than the standard subtype C strains.

Our data also raise several important questions. First, are the variant viral strains of subtype C likely to alter the landscape of the HIV demographics in India in the coming years? In the recent past, the rate of viral expansion has slowed or even declined in several global regions including India (UNAIDS Global Report: 2010). How are the rates of viral prevalence going to be affected as a consequence of the emerging viral
strains in India and elsewhere? Second, are the HIV-1 subtypes, especially subtype C given its supposedly high levels of replication competence, likely to undergo additional evolutionary modifications in the coming years? Considering the high impact the enhanced gene expression could have on immune activation and disease progression (73), it is rather unlikely that 4-kB subtype C would repeat the same evolutionary strategy to acquire additional NF-κB or other positive regulatory sites in future to enhance the promoter strength further. Third, given that the variant strains have begun to emerge as mono-infections in India (Supplemental Fig. 7A), is an evolutionary divergence between the circulating and emerging viral strains a possibility in the coming years? Almost all the viral infections of the 4-kB viral strains during the period of 2000-2003 in India were found as coinfections with the 3-kB viral strains. In 2010-2011, however, a large number of the 4-kB viral infections were found as monoinfections suggesting a progressive epidemiological segregation of the two viral variant strains (Supplemental Fig. 8).

Furthermore, an extensive sequence comparison of 200 and 10 full-length viral sequences of 3- and 4-kB viral strains, respectively, of the South African origin (Rousseau et al., 2006) using the Viral Epidemiology Signature Pattern Analysis tool (VESPA) software available at the HIV Sequence Database (www.hiv.lanl.gov) failed to identify differential association of any amino acid or nucleotide residue with either of the κB promoters (3- versus 4-kB) in any of the viral proteins, including Gag, Pol, Env, Tat, Rev, Vif, Vpr, Vpu and Nef (data not shown). A similar analysis of 10 each envelope sequences of 3- and 4-kB viral strains from our own clinical cohort too failed to identify any preferential association of the residues with the κB number. A large number of the 4-kB viral infections in the early years were found invariably as a coinfection with the 3-kB viral strains (Supplemental Fig. 7). The enormous magnitude of the viral recombination in vivo may have obliterated any possible variations between the two variant viral strains. Now that the 4-kB viral strains are found as a monoinfection in a significantly large number of contemporaneous samples, the manifestations of the divergent viral evolution may be identified in the coming years. Prospective observational studies will be required to find answers for many of these questions.

In conclusion, we demonstrated here that at least three different promoter variant strains of HIV-1 subtype C have been gradually expanding and replacing the standard subtype C viruses in India, and possibly in South Africa and other global regions, over the past decade. The new viral strains contain an additional NF-κB, NF-κB-like or RBEIII site or a combination of the last two sites. While the acquisition of an additional RBEIII site is a property shared by all the HIV-1 subtypes, duplication of an additional NF-κB site remains an exclusive property of subtype C virus. The acquired κB-site is genetically distinct; it binds the p50-p65 heterodimer and strengthens the viral promoter at the levels of transcription initiation and elongation. The 4-kB viruses dominate the 3-kB ‘isogenic’ viral strains in pairwise competition assays in T-cell lines, primary cells, and the EcoHIV- mouse model. The dominance of the 4-kB viral strains is also evident in the natural context when the subjects are coinfectcd with the κB-variant viral strains. The mean plasma viral loads, but not CD4 counts, are significantly different in the 4-kB infection suggesting that these newly emerging strains are probably more infectious. It is possible that higher plasma viral loads underlie selective transmission of the 4-kB viral strains. Our current results propose that subtype C virus exploits a small window of opportunity to make higher viral load without probably eliciting a higher magnitude of immune activation. Future studies remain needed to further validate this idea.
Acknowledgements - MB, MA and AV are recipients of the Council of Science and Industrial Research fellowship from the Government of India. Several reagents were obtained through the AIDS Research and Reference Reagent Program. We thank Drs. Vinayaka Prasad and Brian Wigdahl for critical reading of the manuscript. The clinical samples provide by the Freedom Foundation were generated in the Foundation’s Center for HIV/AIDS Medicine and Research at Bengaluru funded through a unique partnership with the Institute for Molecular Medicine and Infectious Disease at the Drexel University College of Medicine in Philadelphia, PA, USA under the direction of Dr. Brian Wigdahl.

FOOTNOTES

1This work was supported by grants to TKK and UR from The Department of BioTechnology, Government of India (Chromatin and Disease: Program Support, Grant No. BT/01CEIB/10/III/01) and intramural funds from JNCASR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

2To whom correspondence may be addressed: HIV-AIDS Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur (PO), Bangalore 560 064, INDIA, Ph. +91-80-2208 2830 (Office), Fax: +91-80-2208 2766, Email: udaykumar@jncasr.ac.in

3This article contains supplemental Figs. S1-S8 and a supplemental Table. 1.

4The abbreviations used are: ATP, Adenosine 5'-triphosphate; BCA, Bicinchoninic acid; ChIP, Chromatin immunoprecipitation assay; CTD, C-terminal doamin; DTT, Dithiothreitol; DMEM, Dulbecco's modified Eagle's medium; ELISA, Enzyme-linked immunosorbent assay; EMSA, Electrophoretic mobility shift assay; EcoHIV, Ecotropic Human Immunodeficiency Virus; EGFP, Enhanced Green fluorescent protein; GST, Glutathione S-transferase; GABP, GA-binding protein; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HIV, Human Immunodeficiency Virus; HTA, Heteroduplex tracking assay; IL-2, Interleukin-2; ITC, Isothermal titration calorimetry; KLH, Keyhole limpet hemocyanin; LTR, Long terminal repeat; MLV, Murine leukemia virus; MFLN, Most frequent naturally occurring length polymorphism; NFAT, Nuclear factor of activated T-cells; NP-40, Nonyl phenoxypolyethoxylethanol; PBS, Phosphate buffered saline; PBMC, Peripheral blood mononuclear cells; PMSF, Phenylmethylsulfonyl fluoride; PMA, Phorbol 12-myristate 13-acetate; PHA-P, Phytohemagglutinin from Phaseolus vulgaris; RIPA, Radioimmunoprecipitation Assay; RPMI, Roswell Park Memorial Institute medium; RBEIII, Ras-responsive binding element III; RBF2, Ras-responsive binding factor 2; SHIV, Simian-human immunodeficiency virus; TFBS, Transcription factor binding site; TSS, Transcription start site; TNF-α, Tumor necrosis factor-alpha; USF, Upstream stimulatory factor; VSV-G, Vesicular Stomatitis Virus-G; VESPA, Viral Epidemiology Signature Pattern Analysis.
REFERENCES


immunodeficiency virus type 1: distinct patterns of viral growth are determined by T-cell types. *J. Virol.* 65, 4350-4358


**FIGURE LEGENDS**

**Figure 1.** Transcription factor binding site polymorphism in the viral promoter of HIV-1 subtype C. (A) A progressive increase in the prevalence of the variant viral strains across India during the periods 2000-2003, 2005 and 2010-2011. The time of sample collection, geographic location, sample number and the nature of the insertion have been illustrated. All the clinical samples below the dotted horizontal line were collected during 2010-2011 from four different clinics in India. The 2000-2003 (southern India) and 2005 (JNCASR) clinical cohorts have been reported previously. The nature of the sequence insertion in the viral promoter of each viral isolate was determined by sequencing the PCR product directly or after cloning into a plasmid vector. The pie charts represent the percentage prevalence of the variant viral strains in a color-coded fashion, wild type 3-kB (cream), 4-kB (red), kB-like (pink), RBE-III (green) and kB-like plus RBE-III (yellow). (B) The molecular nature of the sequence insertion in a South African clinical cohort as reported in Rousseau et al (34).

**Figure 2.** The molecular nature of the NF-kB site acquisition and distribution among HIV-1 genetic subtypes. (A) Multiple sequence alignment of the viral enhancer and upstream sequences of the representative 4-kB viral isolates. The sequences presented here are representative of 159 sequences determined through the present work and 54 sequences downloaded from the databases. The kB motifs have been highlighted in red shading, and the RBE-III site in green. The original (blue arrow) and the duplicated (red arrow) 21-residue sequences are shown. The ‘C’ and ‘T’ variations of the respective kB sites are highlighted in black shade and yellow text color. The cohort identity, sample collection time, geographic location and the total number of the sequences available are shown. The asterisk represents the viral sequences derived through the present work. (B) The global distribution of the MFNLP insertions. A large number of the LTR sequences belonging to the various genetic subtypes of HIV-1 and containing any one of the three (NF-kB, NF-kB like or RBEIII) sequence insertions were downloaded from the extant databases. The color code of the sequence insertions is consistent with that of the Fig.1. The subtype identity, total number of sequences used in the present analysis and the number of viral
isolates containing additional NF-κB or RBE-III sites are shown. The percent prevalence of the variant viral isolates under each genetic subtype is illustrated.

**Figure 3. The F-κB site is an authentic NF-κB motif.** (A) The H- and F-κB probes bind cellular factors in the electrophoretic mobility shift assay. Radiolabeled double-stranded probes containing the H- or F-κB sites were incubated with Jurkat nuclear extracts prepared from cells after 60 min of TNFα activation. Cold competition was performed by pre-incubating the nuclear extracts with 25 molar excess of cold oligonucleotides consisting of H, F, lxB-α or a mutant κB site. A probe for the Oct-1 cellular factor was used as a loading control. Free probe (FP) and non-specific (NS) and specific complexes are indicated. (B) The H- and F-κB probes bind the p50-p65 heterodimer in the supershift assay. Nuclear extracts prepared from control or TNFα-treated Jurkat cells were pre-incubated with affinity-purified rabbit antibodies specific to the rel family members as indicated at the top of the lanes. EMSA was performed as described in panel A. (C) Binding affinity determination of the H- and F-κB probes for the recombinant p50 protein in isothermal calorimetry. Double-stranded H-, F- or a mutant κB-site containing oligonucleotides were used in the assay. The raw ITC traces from a representative titration are presented with respect to the baseline and the heat change versus the molar ratio of the titrated products is plotted. (D) Chromatin immunoprecipitation analysis. A schematic representation of the VSV-G pseudotyped viruses expressing a dual-expression cassette of EGFP and Tat is illustrated in the upper panel. Jurkat cells were infected at an MOI of 1 and 72 h later stimulated with TNF-α for 60 min (left panel) or left without activation (middle panel). Immunoprecipitation of the complexes was performed using 5 µg of antibodies as indicated. A reference lxB-α cellular promoter containing three NF-κB sites was used as a positive control for the ChIP assay. Enrichment of the NF-κB subunits and transcription competent RNA polymerase II on the viral (HHHHH- or FFFF-LTR) or the cellular promoter was evaluated using PCR. One tenth of the input chromatin was uncross-linked and used as an input control. IgG, isotype-matched control antibody. Enrichment of the S2 phosphorylated form of the RNA polymerase II subunit was evaluated by targeting the Tat region located approximately 3000 bp downstream of the TSS of the viral promoters (right panel). The GAPDH cellular promoter was used as a positive control.

**Figure 4. The F-κB site confers quantitative gain-of-function on the C-LTR.** (A) Schematic representation of the dual-expression vectors. The isogenic LTRs originated from a representative subtype C LTR BL42 (FHHC). Reference LTRs from subtype B (NL4.3) and subtype C (Indie-C1) were also included for comparison. (B) Induced reporter gene expression from the viral promoters in the absence of Tat. Jurkat cells were transfected with one of the reporter vectors illustrated in panel A above, 12 h later subjected to diverse activation conditions as indicated and the luciferase secretion at 24 h was evaluated from the medium. Each assay was performed in triplicate wells, and the data are presented as mean relative light units +/- S.D. The data are from one of the three representive experiments. (C) Induced reporter gene expression from the viral promoters in the presence of Tat. Jurkat cells were co-transfected with a plasmid pool containing a Tat-expression vector and one of the reporter vectors. (D) Induced reporter gene expression from the viral promoters under synergistic activation conditions. Jurkat cells were treated as in ‘B’ above and subjected to combinations of two or three different activation conditions as indicated.

**Figure 5. The 4-κB viruses out-compete the 3-κB strains in T-cells** (A) The schematic representation of the paired ‘isogenic’ viral constructs used in the competition assays. The 22 bp sequence consisting of the F-κB site was engineered into the 3’-LTR of the Indie-C1 molecular clone (HHC) to generate the 4-


κB LTR (FHHC). Replication profiles of the viral strains in vitro using (B) CEM-CCR5 T-cells or (C) the CD8-depleted, mitogen-activated PBMC from a representative subject. The cells were infected with 500 infectious units of FHHC or HHC viruses and the secretion of p24 into the medium was monitored for several weeks as indicated. The data are presented as the mean of triplicate wells +/- 1 SD and representative of three independent experiments. Pair-wise competition between the HHC and FHHC isogenic viruses in (D) CEM-CCR5 cells or (E) activated PBMC. The paired viruses were competed against each other at different ratios of MOI as schematically depicted in Supplemental Fig. 4. Genomic DNA was extracted on day 10 following the viral infection and subjected to the HTA analysis. The heteroduplex band intensities are plotted as relative values compared to the mono-infections (see formula in panel D of Supplemental Fig. 4).

Figure 6. The 4-kb viruses out-compete the 3-kb strains in a small animal model and in the natural infection. (A) The schematic representation of the paired ‘isogenic’ viral constructs used in the competition assays. The 22 bp sequence consisting of the F-kb site was engineered into the 3’-LTR of the EcoHIV molecular clone (HHC) to generate the 4-kb LTR (FHHC). (B) Pair-wise competition between the HHC and FHHC ‘isogenic’ viruses in the EcoHIV-mouse model. Mice, 4 animals per group, were infected intra-peritoneally with a total of 5 µg of p24 equivalent of the viruses using the strategy schematically depicted in Supplemental Fig. 4A. Two weeks later, genomic DNA was extracted from splenocytes and subjected to the HTA analysis. The data are representative of three independent experiments. (C) Relative prevalence of the 3- and 4-kb viral variants in the genomic DNA of the clinical samples. Genomic DNA was extracted from PBMC of six different subjects all containing a mixed infection with 3- and 4-kb viral strains. The relative band intensities of the HHC and FHHC viruses were determined in HTA. The mean value of band intensities of the 4-kb strains is statistically significant, p<0.001. (D) Relative prevalence of the 3- and 4-kb viruses in the plasma viral RNA of 3 different subjects. Plasma viral RNA was reverse transcribed, and the proportion of the 3- and 4-kb viruses was determined as above. Longitudinal analysis of the plasma samples separated by a year from a single patient is shown (right panel).

Figure 7. Comparison of the post-entry events between the 3- and 4-kb isogenic viruses. (A) Schematic representation of the early events of HIV replication in CEM-CCR5 T-cells following infection with the HHC or FHHC isogenic viruses. Viruses were generated as in Figure 5A. The filled boxes represent regions targeted for amplification. (B) Quantification of the late reverse transcription products. Cellular DNA was extracted from infected CEM-CCR5 cells 12 h following viral infection, and a 142 bp sequence within the U5-Ψ region was amplified. The data are presented as the mean of three independent reactions +/- S.D. (C) Quantification of the 2-LTR-circles. Cellular DNA was extracted from infected CEM-CCR5 cells 24 h following viral infection, and a 223 bp sequence across the U5-U3 junction was amplified. (D) Quantification of the integrated proviruses using a nested Alu-LTR PCR. Cellular DNA was extracted from infected CEM-CCR5 cells 48 h following viral infection, and a 200 bp sequence within the R-U5 region was amplified. (E) Quantification of the proximal and distal viral transcripts. Forty eight hours following infection of CEM-CCR5 cells with the HHC or FHHC isogenic viruses, total RNA was extracted; first strand cDNA was synthesized, and the quantitative real-time PCR was performed as described. Proximal transcripts were detected with primers amplifying an 89 bp fragment in the TAR region, and the distal transcripts with primers to amplify a 226 bp fragment in Tat.

Figure 8. Comparative analysis of the correlates of disease progression between the 3- and 4-kb viral infections. Subjects categorized by the presence of 3- or 4-kb viral infections have been drawn from all the four Indian clinical cohorts. Each point in the plot represents an individual HIV-1 seropositive subject. (A) The plasma viral load and (B) the CD4 cell count of the subjects are presented in each group. The number of subjects, means and medians of the data, and standard deviation values under
each group are shown. The horizontal lines represent group median values. (C) A schematic model depicting a novel strategy the 4-κB viral strains use to enhance their replication competence by enhancing the infectivity.
FIGURE 1.

A

2000-03
Southern India
N=607

2005
JNCASR cohort
N=57

2010-11

New Delhi

Freedom Foundation
N=30

St. John’s Hospital
N=44

Bengaluru

AIIMS
N=30

Chennai

YRG-CARE
N=55

South African cohort (2005)
N=256

St. John’s Hospital
N=44

Freedom Foundation
N=30

AIIMS
N=30

Chennai

YRG-CARE
N=55

New Delhi

Freedom Foundation
N=30

St. John’s Hospital
N=44

AIIMS
N=30

Chennai

YRG-CARE
N=55

New Delhi

Freedom Foundation
N=30

St. John’s Hospital
N=44

AIIMS
N=30

Chennai

YRG-CARE
N=55

New Delhi

Freedom Foundation
N=30

St. John’s Hospital
N=44

AIIMS
N=30

Chennai

YRG-CARE
N=55

New Delhi

Freedom Foundation
N=30

St. John’s Hospital
N=44

AIIMS
N=30

Chennai

YRG-CARE
N=55

New Delhi
**FIGURE 3.**

A. EMSA analysis

B. Supershift analysis

C. Affinity determination

D. ChIP analysis

**Table:**

<table>
<thead>
<tr>
<th>Affinity Determination</th>
<th>Time (min)</th>
<th>Molar ratio</th>
<th>Molar ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>HxB</td>
<td>0</td>
<td>0.001</td>
<td>0.50</td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>30</td>
<td>1.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>40</td>
<td>2.00</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>50</td>
<td>2.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>60</td>
<td>3.00</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>70</td>
<td>3.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**ChIP analysis**

TFα binding to Oct-1 complex 5'LTR

GFP binding to Oct-1 complex 3'LTR

**Supershift analysis**

NFκB complex

NS

Supershift

FP

H F IκB Mut

Cold competition

Oct-1 complex

**Table:**

<table>
<thead>
<tr>
<th>kbp-oligo</th>
<th>Sequences</th>
<th>No of binding sites</th>
<th>AH</th>
<th>AS</th>
<th>Kd (nM)</th>
<th>Ka (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>ggccgctGGGACTTTCCagga</td>
<td>1</td>
<td>1</td>
<td>1.78x10^6</td>
<td>3.28x10^5</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>gccgctGCCACTTGGCagga</td>
<td>0.79+</td>
<td>0</td>
<td>0.50</td>
<td>1.00</td>
<td>1.2x10^10</td>
</tr>
<tr>
<td>Mutant</td>
<td>ggccgctGCCACTTGGCagga</td>
<td>0.79+</td>
<td>0</td>
<td>0.50</td>
<td>1.00</td>
<td>1.2x10^10</td>
</tr>
</tbody>
</table>

**Notes:**

- H and F represent different binding sites.
- Mutant indicates a mutated oligonucleotide.
- Kd and Ka values are used to determine the affinity of binding.

FIGURE 4.

A

C-LTR  sLuc  IRES  GFP

H  H-κB site  GGGACCTTTCC
C  C-κB site  ---G-G---
F  F-κB site  ---------T
Sp1 site

B

No Tat

p<0.001  ns  p<0.001

p<0.001

synergistic activation

C

p<0.001  p<0.001  ns

With Tat

D

Synergistic activation

p<0.001  ns  p<0.001  p<0.001
FIGURE 5. A Paired viral strains

B CEM-CCR5

C PBMC

D CEM-CCR5

E PBMC

Relative band intensity

Relative band intensity

p<0.001

p<0.05
**FIGURE 6.**

**A**  EcoHIV-3’ LTR engineering

**B**  EcoHIV-Mouse competition

**C**  Proviral DNA

**D**  Plasma viral RNA
FIGURE 7.

A  Post-entry events

B  PCR for late RT products

C  PCR for 2-LTR circles

D  PCR for integrated provirus

E  Viral transcripts

B  Late RT products

C  Two-LTR circles

D  Integrated virus

E  Viral transcripts

<table>
<thead>
<tr>
<th>3-kB</th>
<th>4-kB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

Proviralcopies/10^6 cells

<table>
<thead>
<tr>
<th>3-kB</th>
<th>4-kB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

Fold change

<table>
<thead>
<tr>
<th>3-kB</th>
<th>4-kB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

No activation

<table>
<thead>
<tr>
<th>3-kB</th>
<th>4-kB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

TNF-α activation

<table>
<thead>
<tr>
<th>3-kB</th>
<th>4-kB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>
FIGURE 8.

A

Plasma viral load

RNA copies/ml

$10^8$

$10^6$

$10^4$

$10^2$

3-kB 4-kB

N

Median

Mean

SD

60

60

20

20

1.334,000

1.334,000

1.334,000

1.334,000

60

199.9

199.9

199.9

20

199.2

199.2

199.2

p<0.01

ns

B

CD4 cell count

Cells/mm³

N

Median

Mean

SD

60

20

60

20

199.9

199.9

199.9

199.9

199.2

199.2

199.2

199.2

20

228.02

228.02

228.02

228.02

188.20

188.20

188.20

188.20

? Changes in Virulence

Subtype C

Gain of replication fitness

Fitness difference

3-kB

4-kB

Gain of replication fitness
Multiple NF-κB sites in HIV-1 subtype C LTR confer superior magnitude of transcription and thereby the enhanced viral predominance


*J. Biol. Chem.* published online November 6, 2012

Access the most updated version of this article at doi: 10.1074/jbc.M112.397158

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2012/11/06/M112.397158.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2012/11/06/jbc.M112.397158.full.html#ref-list-1