Global HIV-1 molecular epidemiology with special reference to genetic analysis of HIV-1 subtypes circulating in North India: Functional and pathogenic implications of genetic variation


Laboratory of Virology II, National Institute of Immunology, New Delhi 110 067, India
Guru Nanak Dev University, Amritsar 143 005, India
University College of Medical Sciences & GTB Hospital, Delhi 110 097, India
Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh 160 012, India

HIV-1 displays extensive genetic diversity globally which poses challenge in designing a suitable antigen/immunogen to provoke desired protective immune response in host. HIV-1 mediated pathogenesis is complex and involves host genes, virus genes and other factors. A number of genetic subtypes have been identified based on sequence variations, largely in envelope region. Different genetic subtypes display variation in amino acid sequences with increasing incidence of subtype B, C, D and mosaic recombinants in India. They can potentially alter the functions of several proteins like Rev, Tat, Vpr, Vif etc and thereby, influence HIV-1 mediated pathogenesis. Recent study has shown that LTR promoter region exhibits novel mosaic structures with segments from B/C Myanmar and India. This indicates rapid evolving nature of HIV-1 and causing epidemics due to existence of multiple subtypes in Indian region. These multiple subtypes show significant differences in various functions (gene activation, cell cycle arrest, RNA binding activities) compared to prototype subtype B genes. These differences may help in better understanding of unique features of HIV-1 epidemic in India.

Keywords: HIV/AIDS epidemic, HIV-1 genes, Molecular epidemiology

Prevalence of HIV-1 subtype in India

In 2006, UNAIDS/WHO has reported that 39.5 million people are infected in the world, but using improved methods of estimation it is reported to be 33.2 million in 2007. Although Sub-Saharan Africa contributed to this decrease to an extent, the major reason for this decline was the reduction in reported cases of HIV from 4-5 million in 2006 to 2.5 million in year 2007 in India using more sophisticated models (especially improved surveillance system) of estimation. It must be emphasized that these reduced number do not change the severity of public health problem in India. It is estimated that national adult HIV-1 prevalence in India is in the range of ~ 0.36%. According to the national population based survey (NFHS-3, 2007), the prevalence rates vary among different states of India ranging from 0.07% in Uttarpradesh to ~ 1% in Andhrapradesh to 1.13% in Manipur. The reasons for relatively high prevalence of HIV-1 in Southern states compared to Northern states in India are not clear. It must be pointed out that HIV-1 epidemic in India is largely limited to high risk population (men having sex with men, commercial sex worker, injecting drug users etc.). This is in contrast to sub-Saharan Africa where the epidemic has affected the general population. It is estimated that approximately 6000 people get infected with HIV-1 every day and the same number of people die, mostly in poor countries, due to not having access to anti-retroviral drugs. Three distinct groups (M- major; N – non-M, non-O or new; and O – outlier) of HIV-1 have been identified. Of these three groups, it is the M group that is driving the global epidemic. The M subgroup comprises several subtypes (A-D, F-H, J, K) besides unique circulating forms. Global distribution of genetic subtypes is also interesting in the sense that UK and US show predominance of genetic subtype B, subtype C is dominant in Asia (including India) and some parts of Africa. Subtypes A, B and F are further divided into A1, A2, B’, F1 and F2. It is important to mention that within subtype about 15% in Gag and approximately 30% in Envelope sequence may vary. Intersubtype changes may vary from 30 - 40% in the same two genes. If
more than two subtypes are involved, cpx terminology is used that denotes complex structure. Full-length genome sequence information is available on limited samples from India, but it may not be unreasonable to conclude that pure genetic subtypes in M-group (subtype C) are going to be rare event due to extensive ongoing recombination (intra-and inter subtype). In some instances, for example, in China and South East Asia, subtype B’ is replaced by subtype B of American lineage, then subtype C entered in 1990 and presently, two unique recombinant strains, CRF07_BC and CRF08_BC, are mainly responsible for sustaining the current epidemic. It has been hypothesized that virus subtype with enhanced replication fitness ultimately gains superiority over others. Therefore, in principle, it is possible in India or any other region of the world, the prevalent subtype may be replaced with other subtypes or novel recombinants. Therefore, genetic analysis of circulating HIV-1 strains should be an ongoing essential exercise for any country as a national public health policy. It is also likely that independent “foci” of HIV epidemic may be developing in more than one region of India (or any other country that sharing its borders with many nations). Epidemic in Manipur, India is influenced by subtypes prevalent in injecting drug users in that region (junction of China, Mynmar and India). Same may be applicable to other regions, like Goa (located South of Mumbai, India) where the beaches are inhabited by many foreigners using illicit drugs.

Reasons for selective expansion and spread of certain subtypes are not clear, but can be multifactorial including global trade and travel. Earlier, unique circulating forms, like CRF01_AE was predominant in South East Asia among commercial sex workers, but now spreading faster in all risk groups. The possible entry point of various genetic subtypes in India has been shown in Fig. 1.

---

**Fig. 1** — Plausible route of HIV-1 entry to India. The arrows indicate the probable route of spread of different HIV-1 subtypes. It is noteworthy that subtype C is the predominant subtype in India and occasionally recombinants (1 to 10%) have been observed. [External boundaries of the countries in the map as depicted are neither correct nor authentic]
Genetic characterization of HIV-1 circulating strains in India

HIV-1 epidemic in India is largely driven by genetic subtype C but lately other genetic subtypes including Circulating Recombinant forms (CRFs) and Unique Recombinant Forms (URFs) have been reported. It has now spread to almost all the states of India. Although the HIV-1 epidemic came later in Asia than Africa, it is now recognized as the second largest epidemic that accounts for more than 20% infections in the world. Genetic studies carried out earlier in India on full-length HIV-1 genome have suggested that recombinants possess one or more gene sequences of HIV-1 derived from different genetic subtypes. The genetic characterization of HIV-1 subtypes has been carried out mostly from Maharashtra (Mumbai and Pune), West Bengal and Karnataka (Bangalore or Manipur), India which clearly suggests that subtype C is most predominant subtype along with some recombinants. Some isolates possess mosaic structure with genetic components from subtypes A and C. Unique B/C recombinants have been reported from Southern India and West Bengal and also from Manipur. It is, thus, clear that existence of multiple subtypes or recombinants (Fig. 2) create a favorable condition for generation of intra-subtype (within the same subtype) or inter-subtype (within different subtypes) recombinants. This is possible when an individual is co-circulating two or multiple subtypes in large number due to extensive replication and nature of replication of genome of HIV-1 RNA with error-prone reverse transcriptase. Since HIV-1 recombination occurs in a random manner, it is likely that cross over break points can occur outside or within the open reading frames (ORFs) of HIV-1 genes (data unpublished). It is interesting to note that extensive intra-subtype recombination has been reported recently in South

![Image of India with HIV-1 subtype distribution](image-url)

Fig. 2 — Sequence (partial or full-length) based characterization of HIV-1 subtypes and Circulating Recombinant Forms (CRFs) in different states of India. [External boundaries of India as depicted are neither correct nor authentic]
African HIV-1 subtype C genome. It is also noteworthy that B/C recombinants have been detected in several states of India (Fig. 2).

**Genetic characterization of HIV-1 LTR genes in North India**

We have recently reported a novel recombinant HIV-1 LTR sequence from North India whose 5’-half consisted of intra-subtype recombinant between Mynamar BC and Indian BC and 3’-half of this long terminal repeat (LTR) has shown relatedness with consensus B subtype. These HIV-1 infected individuals have been reported from Chandigarh, India (Northern region) by A. Wanchu, Immunodeficiency Clinic, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh, India after obtaining all requisite clearances. Mode of transmission has been reported through heterosexual for all of them. This is a relatively affluent region with most families having relatives settled in USA, UK or Canada. Appropriate primers (subtype B or C-specific - obtained on request) were designed to amplify full-length HIV-1 long terminal repeat (LTR) promoter region from the genomic DNAs isolated from PBMCs as described by us earlier and cloned into a T-tailed vector (pTarget-TM – Promega) under T7 promoter for sequencing purposes using T7 promoter-specific primer. The plasmid was subjected to sequencing and detail genetic analysis was carried out using modern genetic tools as described by us recently using bootstrap analysis, simplot and viral genotyping tools (NCBI) etc. (details given in the legend to Fig. 3). Several interesting and novel features were observed with these recombinants. LTR region was approximately 650 bases long present on either end of 9kb long genomic RNA. This region typifies a cellular pol II promoter which was rich in several transcription factor binding sites (TFBs). It was noteworthy that genetic subtype-specific changes in this region were observed. Usually there are two Nf-kb sites in prototype LTR subtype B (the most studied because it is prevalent in US and UK), but subtype C LTRs increasingly possess 3 Nf-kb TFBs. The other TFBs are 3 Sp1, NFAT, TATA, TAR etc. We analyzed HIV-1 LTR region from 6 HIV-1 infected individual and observed region specific changes in several of TFBs, especially the NF-AT-II and III sites were highly polymorphic. Most notably, we observed two unique and mosaic B/C recombinants. It turned out be a combination of intra- and inter-subtype B/C recombinants derived from Indian B/C and Mynmar B/C. A boot scan analysis with informative site has been shown in Figure 3. Single nucleotide changes

![Fig. 3— Novel mosaic recombinant LTR- B/C NII-PGI-IND-S3 was analyzed using viral genotyping tools located in NCBI. The bar above the graph represents intra- and inter-subtype recombination with multiple cross-over points (nucleotide positions – 267, and 336) involving B/C-Myanmar, B/C-India and subtype B after performing bootscan analysis. The accession number of this unique B/C recombinant is EU574917. The analysis was performed in Simplot ver 3.5.1 using the kimura-2 parameter with a window size of 100 bases and a step size of 20bp.](image-url)
were observed in the TAR region also. Figure 4 depicts a plausible mechanism by which a recombinant B/C LTR can be generated. Usually, lysine t-RNA primes the short –ve sense DNA at the end of 5’LTR and then it jumps to 3’-LTR to form a double-stranded DNA (the provirus; step 1). But in a cell that is co-infected with subtype C, -ve sense short DNA can jump and hybridize to LTR C (step 2) giving rise to LTR/B/C recombinant.

**HIV-1 Tat B and C functional analysis**

There are two exons for *Tat* gene and usually the first exon is sufficient for its remarkable activity to transactivate HIV-1 LTR promoter. There are significant changes in the sequence of amino acids between the prototype B Tat (derived from pNL4-3) and 93IN905 (derived from an Indian isolate in NARI, Pune). Most remarkably, the RGD domain present in Tat B is replaced by QGD in subtype C. We made several mutant constructs by swapping the first and second exons between subtypes B and C. We also made a construct of subtype B that possessed the RGD motif. These changes profoundly affected the ability of the new construct to activate HIV-1 LTR promoter. RGD seemed to promote apoptosis in subtype B, but the same could not be concluded for the QGD motif present in subtype C. Earlier Campbell *et al.* have shown that HIV-1 Tat C protein derived from the same Indian isolate (93IN905) not only failed to induce intracellular calcium flux, but induced reduced tumor necrosis factor production (TNF) and Tat protein of subtype C is a defective cytokine and it may have neurological implications.

**HIV-1 Vpr B and C functional analysis**

Vpr is an important accessory protein of HIV-1. It is important for replication of virus in certain cell types especially the macrophages. It is 96 aa long and is known to possess three alpha helix structure. It is known to influence several other functions, notably in causing cell cycle arrest, promote HIV-1 LTR activation and causes apoptosis. Our comparative study using prototype subtype B Vpr and Vpr C (derived from an Indian isolate- 93IN905) showed significant variations. Remarkably, Vpr B dose-dependent HIV-1 LTR activation was observed in a transient cellular assay which was not the case with Vpr C. On the contrary, Vpr C always caused about 10 -20% more apoptosis. This differential impact on the two vital functions (transactivation and apoptosis) will obviously have varying effects on gene expression and pathogenesis. When both the proteins were tested for their ability to cause cell cycle arrest, interesting differences were observed between them. We have designed an efficient 10-23 catalytic motif containing DNA-enzymes that can cleave the full-length Vpr B and C RNA in a sequence-specific manner and used them to reverse the cell cycle arrest mediated by them.

**HIV-1 Vif B and C differ in their ability to degrade APOBEC3G**

HIV-1 Vif protein has recently been recognized to selectively and efficiently degrade APOBEC3 family of proteins which are known to deaminate HIV-1 genome, leading to disruption of open reading frames (ORFs). Vif variants have earlier been reported to possess varying activities towards the degradation of APOBECs often contributing to viral diversity. When one compares the amino acid sequences of prototype Vif B with that of Vif C, significant changes are observed throughout the entire sequence. We have reasoned that these changes may impact upon its functions, particularly its ability to interact with proteins (Elongins, Cull 5 etc) involved in ubiquitin-mediated pathways responsible for degradation. To broadly define the Vif domains responsible for degradation of APOBECs, we have constructed HIV-1 Vpr B/C chimeric constructs such that the N-terminal half consists of either B or C Vpr.

---

**Fig. 4— Proposed mechanism for formation of LTR B/C recombinant.** Lysine t-RNA is used for priming the synthesis of short –ve SS DNA. This short –ve SS DNA can jump and hybridize to either homologous (step 1) LTR B or to heterologous, in this case LTR C (step 2), due to similarity in the sequences of the two subtypes.
gene and the same is true for the C-terminal half. Presently, Vif regions or amino acids involved with direct interaction with APOBEC protein are not well defined but believed to involve several regions of the polypeptide chain. We have observed that C-terminal half from subtype C possessed the major determinant for APOBEC degradation\(^{18}\).

**More efficient interaction of Rev B protein and RRE B RNA than Rev C protein and RRE C RNA**

We expressed Rev B and C protein in *E. coli* as GST-fusion proteins and purified them using standard procedures. We amplified 375 bases long RRE region using subtypes B and C specific primers. Labeled RRE RNA was incubated with increasing amounts of protein and binding was monitored by gel shift analysis. We observed that Rev B – RRE B RNA interaction was more efficient than Rev C and RRE C RNA. Remarkably, Rev B protein showed good binding with RRE C, but under similar conditions, Rev C protein failed to show any binding with RRE B RNA (Banerjea, unpublished data). This has enormous physiological relevance. In situations where dual infections (mainly subtype B and C) are common, subtype B specific gene expression would be more. This obviously has implications for subtype specific pathogenesis.

**Functional implications of genetic variations in HIV-1 subtype C envelope**

Sundaravaradan et al.\(^{19}\), have carried out functional analysis of V3 to V5 region from subtype-C by placing them in the background of the most studied NL4-3 virus (subtype B)\(^{19}\). Their study has suggested that these regions are responsible for increased viral replication in peripheral blood mononuclear cells and monocyte derived macrophages. It is noteworthy that V3 region is the major determinant for conferring the T-lymphocyte-tropic (X4 tropic) and macrophage-tropic (R5 tropic) phenotypes besides harboring epitopes for neutralization. Earlier studies with Indian isolates have suggested that almost all isolates use CCR5 chemokine receptor (R5 tropic) and show no evidence for co-receptor switch from R5 to X4 with progression of the disease\(^{36}\). Lately, some dual tropic, X4R5 or X4, isolates have also been observed (J Bhattacharya, NARI, Pune, and Banerjea unpublished data). It is widely known that R5 tropic virus initiates the infection in mucosal tissues and at later stages, X4 tropic viruses appear that leads to development of HIV/AIDS rapidly. Presently, we do not understand the determinants for this important observation, but it may seem reasonable to speculate that changes in envelope region may have a major impact. Individuals not expressing a functional CCR5 (due to a remarkable 32 base pair deletion mutation in its ORF) are largely protective against HIV-1 infection\(^{10}\). Thus, CCR5 chemokine receptor is an attractive target for designing novel antiviral approaches (structure based analogs, siRNAs, ribozymes etc) against HIV-1.

**Functional studies with HIV-1 Vpu protein C**

Vpu is one of the accessory proteins that may play an important role in modulating the severity of HIV/AIDS disease. Besides playing an important role in proteosomal degradation of CD4 receptor via E3 ubiquitin ligase complex, it is also known to facilitate the release of viruses from the infected cell surface. Hussain et al.\(^{37}\), have carried out studies using Vpu C from an Indian isolate (R5 tropic) with respect to intracellular localization and its ability to form oligomers (monomer to pentamer). Their study has confirmed the various oligomeric states of Vpu and suggested that it is localized in golgi or intracellular vesicles, but not in endoplasmic reticulum.

**Conclusion**

Several important conclusions can be drawn from our studies in this North Indian region. HIV-1 epidemic in this region is rapidly evolving due to extensive recombination among circulating HIV-1 strains, mainly due to subtypes B and C. This is not surprising because extensive intra-subtype recombination has earlier been reported in African Subtype C genome\(^{9}\). This is not unexpected as both intra- and inter-subtype recombinants are generated using similar RT mediated mechanisms responsible for replication of their genomes. Our genetic analysis on HIV-1 LTR regions clearly suggests that genetic elements from Mynmar region contribute to unique mosaic structure. This observation raises interesting questions about the spread of genetic elements from North-Eastern border of India (China- Mynmar) to North India. Remarkably, this recombinant, which is due to intra- and inter-subtype cross-over events, possesses 3 NF-κB sites, which is increasingly found in subtype C. Thus, this unique recombinant may express HIV-1 genes more efficiently than the ones with only 2 copies of NF-κB which is predominantly the case with subtype B LTRs. Recombinants can
further evolve by combining with other subtypes or recombinants prevalent in that region which can potentially result in very complex mosaic structures. Occasionally, new recombinant forms do play a major role in driving the regional epidemic. Such temporal studies are important to monitor the spatio-dynamics of emerging strains. They will obviously have enormous implications in designing region specific T-cell based vaccines. Most studies on HIV-1 proteins have been carried out using a prototype B strain of HIV-1 (NL4-3 or HXB2), but similar information for subtype C is lacking and needs to be studied in more details to understand the molecular basis of pathogenesis where subtype C predominates, especially in India. We have selected 93IN905 Indian strain which is > 95% identical to the predicted consensus subtype C. Our comparative studies with Tat, Vpr Rev and Vif suggest that significant functional differences with respect to apoptosis (with Tat and Vpr), RNA binding (with Rev) or transactivation (Vpr and Tat) exist and this may significantly impact on pathogenesis. HIV-1 envelope changes may influence its binding efficiency to host cells and its ability to elicit neutralizing antibody etc. Most of the accessory proteins of HIV-1 seem to influence several biological reactions. Increase or decrease in the known functions (apoptosis, RNA binding or HIV-1 gene expression) of accessory HIV-1 genes will clearly impact on the progression, pathogenesis and overall replication of the virus. To gain insight into co-receptor switch or the absence of it, additional long-term monitoring of the envelope sequences is warranted.

Acknowledgement

The present work was supported by grants to the author (ACB) from Department of Biotechnology, Government of India, New Delhi, India, and National Institute of Immunology, New Delhi, India. This work was also partially funded by awarding National Bio-science award for career development, and also by DBT-ICMR, and Indo-US RO3-NIH to the authors (ACB and AW). Several HIV-1 related research materials were procured from AIDS Research and Reference Reagent Program, NIH, MD, USA. Dr A. Wanchu, PGI, Chandigarh, India supplied the samples with clinical evaluation and Dr Akhil C Banerjea was Principle Investigator in the project. [This mini review is not meant to be comprehensive and we apologize if important references have not been quoted].

References


Bano A S, Gupta N, Sood V & Banerjea A C, Vpr from HIV-1 subtypes B and C exhibit significant differences in their ability to transactivate LTR-mediated gene expression and also in their ability to promote apoptotic DNA ladder formation, *AIDS*, 21 (2007)1832.


