

First Report of Human Immunodeficiency Virus Type 1 Circulating Recombinant Form 02_AG Recombinant Strain of African Origin from North India

^{1,2}Ujjwal Neogi, ¹Vikas Sood, ¹Arpita Choudhury, ³Shukla Das, ³V.G. Ramachandran,
⁴Ajay Wanchu and ¹Akhil C. Banerjea

¹Division of Virology, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi-110067, India

²HIVIND Unit, St. John's National Academy of Health Sciences, Bangalore-560034, India

³Department of Microbiology, University College of Medical Sciences,
Guru Teg Bahadur Hospital, Delhi-110095, India

⁴Department of Internal Medicine,

Post Graduate Institute of Medical Education and Research, Chandigarh-160012, India

Abstract: Problem statement: Human Immunodeficiency Virus Type-1 (HIV-1) displays extensive genetic diversity globally which poses a serious challenge in making efficacious vaccine and developing effective therapeutic interventions. HIV-1 epidemic in India is largely driven by genetic subtype C but lately a number of reports suggest the presence of circulating recombinant forms and unique recombinant forms from different states of India. **Approach:** We analyzed 13 envelope sequences spanning the region from V1-V5 from three regions of North India and subjected to extensive genetic analysis. **Results:** We identified equal numbers (n = 6) of subtype C and CRF02_AG recombinant strain and a single subtype B-specific envelope. The extensive phylogenetic analysis of our CRF02_AG strains indicate that they are closely related to Central and Central-West African strains. The analysis of intra-subtype nucleotide distance revealed that the subtype C sequences are significantly more divergent and diverse than CRF02_AG strains (p<0.004). **Conclusion:** Appearance of CRF02_AG strain of African origin among commercial sex workers from Delhi and Punjab region of India is reported for the first time which needs to be monitored for its pathogenesis and its ability to recombine with the circulating strains.

Key words: Viral genotyping, molecular epidemiology, HIV-1 genes, HIV-1, CRFs in India

INTRODUCTION

First incident of CRF02_AG strain was reported from Gabon, a South Western country of Africa in the year 1988 (Delaporte *et al.*, 1996). Subsequently it was reported from Djibouti, Nigeria, Mali, Niger and other neighboring countries (Carr *et al.*, 1998; McCutchan *et al.*, 1999; Cornelissen *et al.*, 2000). It was first reported in Europe from France (Vachot *et al.*, 2004) in the year 1992, followed by Netherlands, Sweden, Spain (mainly from the immigrants from west and central west Africa) (Holguin *et al.*, 2000). Among the Asian countries the first report was from Republic of Korea (Kim *et al.*, 1999). The first case of HIV/AIDS in India appeared in the year 1985. The progress of HIV epidemic has relatively been slow initially. Over the years the number of HIV infected persons have increased and reported from almost all the states of India and is a major public health problem

now. The estimated number of HIV-1 infected individuals was 2.5 million in the year 2008 which placed India in the third position behind South Africa and Nigeria (UNAIDS, 2008). Global HIV-1 spread coupled with extraordinary ability of the virus to mutate poses a great challenge to the clinicians and researchers in terms of devising successful therapy and in the development of effective vaccine. In India, the predominant HIV-1 genetic subtype is C followed by Subtypes B, A and E along with several Circulating Recombinant Forms (CRFs) and Unique Recombinant Forms (URFs) co-circulating in different parts of India (Neogi *et al.*, 2009). India is a large country and it is quite possible that isolated pockets of HIV-1 pandemic may develop independently. The geographic location of North India is important for shaping the HIV-1 epidemic because of the presence of other genetic subtypes in countries like China, Pakistan and Afghanistan. The inhabitants of these regions are

Corresponding Author: Ujjwal Neogi, HIVIND Unit, St. John's National Academy of Health Sciences, Sarjapur Road, Bangalore-560034, India Tel: +91-80-22065059

also frequent traveler to the European countries (Neogi *et al.*, 2008).

We have analyzed partial envelope sequences from North India from four different but adjacent states (Delhi, Punjab, Haryana and Himachal Pradesh, Delhi being the capital of India). We found existence of subtypes B, C and CRF02_AG strains in this region. In this study we analyzed in detail the ancestral origin of the CRF02_AG strains from North India using phylogenetic approach and predict the possible route of spread of these strains from Africa to Europe and Asia.

MATERIALS AND METHODS

Patient population and genetic characterizations:

We were able to amplify 13 Envelope sequences (V1-V5 region) from 45 enrolled patients from Delhi, Haryana, Punjab and Himachal Pradesh following all the requisite ethical clearances. The genomic DNA (pro-virus) was obtained from whole blood genomic DNA using QIAamp Blood DNA mini kit (Qiagen, Germany) and was used to amplify the envelope gene as described by us previously using nested PCR technology (Neogi *et al.*, 2008; Bano *et al.*, 2009). All the Polymerase Chain Reactions (PCRs) were performed with high fidelity Taq DNA polymerase (Ex-Taq, Takara, Japan) using the specific primers. Approximately 1.25 Kb DNA fragment corresponding to V1-V5 region was amplified initially using the following primers (FPEnv-01 and RPEnv-02). Thereafter, 700 bp fragment (V3-V5) was amplified using two internal sets of primers (FPEnv-03 and RPEnv-04). The two pairs of primer sequences were as follows:

- FPEnv-01: 5'-ATGGGATCAAAGCCTAAAGCCATGTG
- RPEnv-02: 5'-AGTGCTTCCTGCTGCTCCCAAGAACCCAAG
- FPEnv-03: 5'-CTGTAAATGGCAGTCTAGC
- RPEnv-04: 5'-CACTTCTCCAATTGTCCCTCA

The cycling conditions for amplifying both the fragments were: 35 cycles at 98°C for 15 sec, 55°C for 30 sec and 72°C for 1 min with a final extension at 72°C for 10 min. PCR-amplified DNA was cloned into pGem-T (Ez) vector (Promega Biotech. WI, USA) and sequenced in both directions using T7 and SP6-specific primers. At least 4 independent clones were analyzed for each sample and only one representative clone from each sample was subjected to extensive genotypic

analysis. Mother and child samples were processed separately to avoid cross contamination. It is important to mention that we were unable to amplify envelope sequences from several samples which may be due to extreme genetic variability, poor yields of genomic DNA, or due to low proviral DNA content in the cells of the infected individuals as most of them were under first line of antiretroviral treatment. The characteristics and some of the clinical parameters of 13 patients are given in Table 1.

Sequences were compared with reference strains downloaded from Los Alamos database (<http://www.hiv.lanl.gov>). Multiple sequence analysis was performed in ClustalW 1.8.3 located in DNA data Bank of Japan (DDBJ) website (<http://clustalw.ddbj.nig.ac.jp/top-e.html>). The phylogenetic analysis was carried out using MEGA 4 software (Tamura *et al.*, 2007). Genotyping was carried out using viral genotyping tools located at NCBI (<http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>), REGA subtyping tool ver 2.0 (<http://www.bioafrica.net/subtypetool/html>) and RIP 3.0 (<http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>). *In silico* co-receptor usage predictions were performed using Geno2pheno (<http://coreceptor.bioinf.mpi-inf.mpg.de/>) and WebPSSM (<http://indra.mullins.microbiol.washington.edu/webpssm/>) programs.

Table 1: Characteristics of HIV-1 infected individuals (n = 13) from North India

Characteristic	Value
Age (mean ± SD) (years)	28.6±11.8
Gender	
Male	5.0
Female	8.0
Residence	
Punjab	5.0
Haryana	2.0
Himachal Pradesh	1.0
Delhi	5.0
Route of transmission	
Heterosexual	10.0
Vertical	2.0
Unknown	1.0
CD4 count*	
Median (range)	480.0 (96-834)
ART status	
Naive	4.0
Experienced	9.0
Year of sampling	
2007	2.0
2008	10.0
2009	1.0

*: Median CD4 count is calculated from 10 patients only

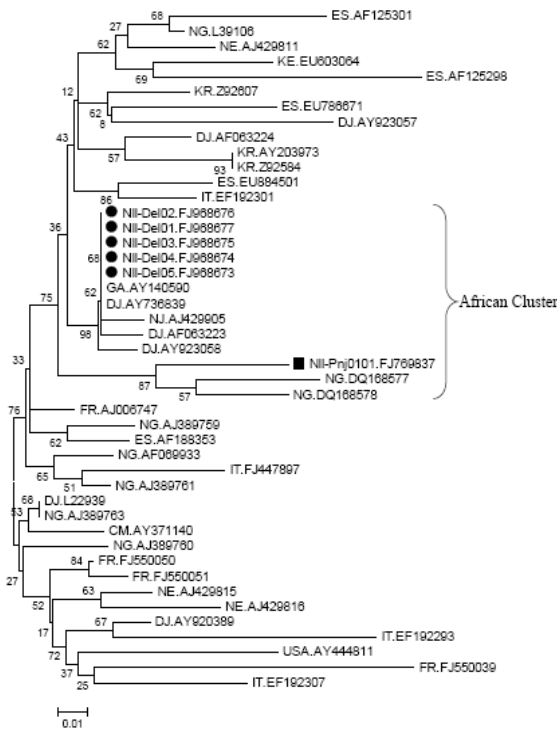


Fig. 2: Evolutionary relationships of CRF02_AG strains from different continent

Note: The evolutionary history was inferred using the Neighbor-Joining method. Confidence probability was multiplied by 100 so that the interior branch length is greater than 0, as estimated using the bootstrap test (1000 replicates is shown next to the branches). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The accession numbers are given followed by country initial (DJ: Djibouti, NG: Nigeria, NJ: Niger, GA: Ghana, CM: Cameroon, KR: Korea, IT: Italy, NE: Netherlands, FR: France, ES: Spain). Indian strains were indicated by ■ (Punjab) and ● (New Delhi)

Among the HIV-1 infected individuals there were 2 mother-to-child transmission cases. In both the cases the mother and child were infected with subtype C viruses and successful transmission (judged by high bootstrap value among individual mother-child pairs) was detected (data not shown). 5 CRF02_AG and 4 subtype C showed CCR5 usage. One each subtype C (Acc No. FJ769835) and subtype B (Acc. No. FJ769836) showed CXCR4 co-receptor usage (data not shown, manuscript under preparation).



Fig. 3: Plausible route of spread of CRF02_AG strain in Africa, Europe and Asia

Note: The route of spread of 02_AG strain from West and Central West Africa to Europe (Netherlands, France, Spain, Sweden) and Asia (Punjab and Delhi of India) is shown by arrows. (External boundaries of the countries shown in the map as depicted are neither correct nor authentic)

The most remarkable finding was the predominance of CRF02_AG strain among the unrelated Commercial Sex Workers (CSWs) from Delhi (Capitol of India) region. Most of the isolates from Punjab, Himachal Pradesh and Haryana region showed relatedness with consensus C. We also found one CRF02_AG (FJ769837) from Punjab (the rest from CSWs) and one subtype B (FJ769836) envelope-specific sequence from Himachal Pradesh.

The neighbor joining tree conducted in MEGA4 of 02_AG strains showed clustering of Indian CRF02_AG isolates with African CRF_02 clades (Fig. 2). Isolates from Punjab region formed a direct branching with two Nigerian isolates (accn. no. DQ168577 and DQ168578) with high bootstrap value (87, >70 considered significant) whereas isolates from Delhi formed a monophyletic cluster with the reference strains reported from Djibouti, Niger and Ghana. Therefore, it is tempting to speculate that Indian 02_AG strains may have spread from Africa. The UPGMA and ME methods of analysis also gave similar results (data not shown). Blast search also showed greater relatedness of Indian strains with African strains (~90-93%) compared to the European strains.

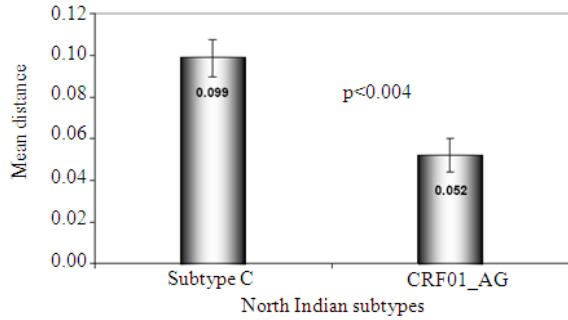


Fig. 4: Nucleotide distance (Kimura two-parameter model) comparisons within V3-V5 sequences of subtype C and Indian CRF02_AG strains

Note: Numbers within columns are the mean distances and error bars are standard errors of the mean. p-values refer to difference between group's means (unpaired t-test with Welch's correction). Mean and pairwise distances were calculated using the Poisson correction method in MEGA4

Taking into consideration the origin and movement of CRF_02AG in Europe and Asia, we predict the most likely route of introduction of this strain in India in Fig. 3. Since we found equal number of subtype C and CRF02_AG (6 each) sequences, we analyzed the intra-subtype differences between these strains. The analysis of intra-subtype nucleotide distances revealed that the subtype C population is significantly more divergent and diverse than CRF02_AG population ($p < 0.004$; Fig. 4).

DISCUSSION

Although subtype C and other recombinant forms of different HIV-1 strains have been reported from different regions from India, this is the first report of CRF02_AG strain from India. This probably indicates a recent introduction of CRF02_AG strain in India. Co-circulations of variety of genetic subtypes were reported from different states of India from the very beginning of HIV-1 pandemic. We recently reported the presence of B/C recombinant LTR (Neogi *et al.*, 2008) and B/C/D recombinant Vpr from north India (Bano *et al.*, 2009). The geographical location of North India plays an important role in influencing the magnitude and nature of HIV-1 epidemic of India. This region is surrounded by China, Pakistan and Nepal where the HIV-1 epidemic is driven by different genetic subtypes (Pakistan: A and China: 07_BC and 08_BC). Therefore, due to extensive traveling, tourism and trade, by HIV-1

infected individuals of the neighboring countries, ideal situation may be created for generating novel and complex recombinant strains. The epicenter of global spread of 02_AG strain is West and central west of Africa. The Indian 02_AG strain has very likely spread from this region of Africa. This strain has potential to recombine with the most dominant subtype of India, the subtype C, along with the various recombinants that have been reported to date from India (Neogi *et al.*, 2009). Our data also underscores the importance of rapidly evolving scenario of HIV-1 epidemic in this region which will have enormous impact on the design of vaccine antigens.

CONCLUSION

In conclusion, we show for the first time presence and transmission of HIV-1 02_AG strain in India and presence of subtypes B and C in North India. These observations will impact on the T-cell epitope based vaccine. The existence of multiple HIV-1 genetic subtypes in this region is likely to generate novel and complex recombinants. Since this is the first report of CRF02_AG from India, it will be important to monitor the spread of this strain and its impact on HIV-1 transmission in India.

ACKNOWLEDGEMENT

This study was supported by grants from Department of Biotechnology, Indian Council of Medical Research, Government of India and National Institutes of Health, MD, USA, to ACB, AW and to the National Institute of Immunology, New Delhi. This research was partially supported by the National Bio-Science Award of DBT to ACB. Support to UN from HIVIND Project are highly acknowledged. Authors also like to thank A. Karthika, Bio-statistician, St. John's National Academy of Health Sciences for her help with statistical analysis.

REFERENCES

- Bano, A.S., V. Sood, U. Neogi, N. Goel and V.S. Kuttia *et al.*, 2009. Genetic and functional characterization of HIV-1 VprC variants from North India: Presence of unique recombinants with mosaic genomes from B, C and D subtypes within the ORF of Vpr. *J. Gen. Virol.*, 90: 2768-2776. PMID: 19605589
- Carr, J.K., M.O. Salminen, J. Albert, E. Sanders-Buell and D. Gotte *et al.*, 1998. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology*, 247: 22-31. PMID: 9683568

- Cornelissen, M., R. van Den Burg, F. Zorgdrager and J. Goudsmit, 2000. Spread of distinct human immunodeficiency virus type 1 AG recombinant lineages in Africa. *J. Gen. Virol.*, 81: 515-523. PMID: 10644851
- Delaporte, E., W. Janssens, M. Peeters, A. Buve and G. Dibanga *et al.*, 1996. Epidemiological and molecular characteristics of HIV infection in Gabon, 1986-1994. *AIDS.*, 10: 903-910. PMID: 8828748
- Holguin, A., B. Rodes and V. Soriano, 2000. Recombinant human immunodeficiency viruses type 1 circulating in Spain. *AIDS Res. Hum. Retroviruses*, 16: 505-511. PMID: 10772537
- Kim, Y.B., Y.K. Cho, H.J. Lee, C.K. Kim, Y.K. Kim and J.M. Yang, 1999. Molecular phylogenetic analysis of human immunodeficiency virus type 1 strains obtained from Korean patients: ENV gene sequences. *AIDS Res. Hum. Retroviruses*, 15: 303-307. PMID: 10052762
- McCutchan, F.E., E. Sanders-Buell, J.K. Carr, M. Bajani and T.O. Harry *et al.*, 1999. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology*, 254: 226-234. PMID: 9986789
- Nei, M. and S. Kumar, 2000. Molecular Evolution and Phylogenetics; In: *Phylogenetic Inference: Distance Methods*, Nei, M. and S. Kumar (Eds.). Oxford University Press, New York, ISBN: 0-19-513584-9, pp: 87-111.
- Neogi, U., V. Sood, N. Goel, A. Wanchu and A.C. Banerjee, 2008. Novel HIV-1 Long Terminal Repeat (LTR) sequences of Subtype B and mosaic intersubtype B/C recombinants in North India. *Arch. Virol.*, 153: 1961-1966. PMID: 18818865
- Neogi, U., V. Sood, S. Banerjee, N. Ghosh and S. Verma *et al.*, 2009. 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. *Ind. J. Exp. Biol.*, 47: 424-431. PMID: 19634706
- Rzhetsky, A. and M. Nei, 1992. A simple method for estimating and testing minimum evolution trees. *Mol. Biol. Evol.*, 9: 945-967. PMID: 7804873
- Tamura, K., M. Nei and S. Kumar, 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA.*, 101: 11030-11035. PMID: 15258291
- Tamura, K., J. Dudley, M. Nei and S. Kumar, 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24: 1596-1599. PMID: 17488738
- UNAIDS., 2008. UNAIDS report of the global AIDS epidemic. http://www.unaids.org/en/Knowledgecentre/HIVData/GlobalReport/2008/2008_Global_report.asp
- Vachot, L., Y. Ataman-Onal, C. Terrat, P.Y. Durand and B. Ponceau *et al.*, 2004. Retrospective study to time the introduction of HIV type 1 non-B subtypes in Lyon, France, using ENV genes obtained from primary infection samples. *AIDS Res. Hum. Retroviruses*, 20: 687-691. PMID: 15307910