The Journal of Microbiology (2012) Vol. 50, No. 5, pp. 869–873
DOI 10.1007/s12275-012-2136-z

NOTE

Envelope Diversity, Characteristics of V3 Region and Predicted Co-Receptor Usage of Human Immunodeficiency Viruses Infecting North Indians

Raiees Andrabi[1], Rajesh Kumar[1], Manju Bala[2], Ambili Nair[1], Prakash SS[1], Vandana Kushwaha[1], and Kalpana Luthra[1]*

[1]Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India
[2]Regional STD Teaching Training & Research Centre, Safdarjung Hospital, New Delhi, India

(Received March 12, 2012 / Accepted June 12, 2012)

Subtypes of human immunodeficiency virus type 1 circulating in 21 north Indian patients were characterized based on the partial sequence of the gp120 envelope protein. A majority of viruses (85.7%, 18/21) were subtype C, while 14.3% (3/21) were subtype A. Sequence analysis revealed that the V3 region was highly conserved compared with V4 and V5. The predicted use of co-receptors indicated exclusive usage of R5, except for two subtype A viruses (AIIMS279 and AIIMS281). Our results demonstrate conservation within the V3 loop of subtype C viruses, and suggest the emergence of non-clade C viruses in the north Indian population.

Keywords: HIV-1, subtypes, envelope glycoprotein, variable regions, co-receptor usage

Viral diversity has emerged as a major challenge in the development of a vaccine against human immunodeficiency virus type 1 (HIV-1) (Taylor *et al.*, 2008). The first description of HIV-1 in India was from Tamil Nadu state in 1986 (John *et al.*, 1987). The virus has rapidly spread throughout the country, with subtype C the most predominant, and a fraction of other subtypes circulating in the population (Lakhashe *et al.*, 2008). Subtype C is the major cause of the current global HIV-1 pandemic, accounting for more than 50% of worldwide infections (Hemelaar *et al.*, 2006), with most infections occurring in Africa, then followed by India. Even between these two geographic regions, the envelope sequences of the subtype C viruses are distinctly different (Shankarappa *et al.*, 2001). The binding of HIV-1 gp120 to the CD4 receptor, and subsequent binding to the co-receptor (CXCR4 or CCR5) on host cells, facilitates viral entry. The crucial involvement of the gp120 envelope protein for viral entry makes it a potential target for HIV-1 vaccines (Pantophlet and Burton, 2006). The HIV-1 gp120 consists of five constant (C1–C5) and five variable (V1–V5) regions (Modrow *et al.*, 1987). The third variable region (V3), usually a 35-mer, cysteine-bound, variable loop of the viral envelope gp120, has been shown to be a critical determinant of co-receptor selectivity. Based on which co-receptor is used, HIV-1 can be classified as X4-tropic (using CXCR4) or R5-tropic (using CCR5) (Berger *et al.*, 1999). Phenotypic and genotypic analysis has shown that the viruses from India are R5- and X4-tropic, with mostly a non-syncytium inducing (NSI) phenotype (Cecilia *et al.*, 2000; Shankarappa *et al.*, 2001). R5-tropic viruses are predominant during the early stages of infection, and persist throughout subtype C infections. Approximately half of HIV-1 subtype B viruses switch from R5- to X4-tropism as disease progresses (Richman and Bozzette, 1994). We previously showed that the V3 region in HIV-1 subtype C-infected individuals was highly antigenic compared with the membrane proximal external region (MPER). Anti-V3 plasma antibodies were dependent on virus persistence (Andrabi *et al.*, 2011). We undertook this work to determine the subtypes present, and to assess HIV-1 envelope diversity in infected individuals residing in the northern states of India.

HIV-1 seropositive patients (n=21) were recruited for this study from the Regional STD Teaching Training & Research Centre of Safdarjung Hospital (New Delhi, India). The study was approved by the institutional ethics committee, and written informed consent was obtained from all participants. Of the 21 patients, 5 were females and 16 were males, with ages ranging from 25–52 years. Fourteen patients were drug naïve and seven were undergoing antiretroviral therapy for a period of up to two years. The viral load in naïve patients (n=10) ranged from 1660–2,180,000 RNA copies/ml of plasma (median=43700). For treated patients (n=5), the viral load was low (128–703,000 RNA copies/ml of plasma, median=248). The viral load could not be estimated in the remaining six patients. The median CD4 count was 190 (range, 14–442) and 197 (104–405) cells/mm^3, while the mean total immunoglobulin G (IgG) levels were 12.21 and 12.27 mg/ml for the naïve and treated patients, respectively.

Genomic DNA was isolated from whole blood using a QIAamp Blood Mini kit (QIAGEN, Germany). The viral envelope from part of the C2 through a part of the C5 re-

*For correspondence. E-mail: kalpanaluthra@gmail.com; Tel.: +91-11-26546431

gion of gp120 was amplified using nested PCR. All PCRs were performed with high fidelity Taq DNA polymerase (New England Biolabs) using primers *Env*-1 gp120 (5′-TCA GCA CAG TAC AAT GTA CAC ATG GAA T-3′ and *Env*-2 gp120 (5′-GTG CTT CCT GCT CCC AAG AAC CCA-3′), yielding an 861 bp amplicon. A 675 bp fragment (C2–C5) was amplified in the second round PCR using primers *Env*-3 gp120 (5′-TGT TAA ATG GCA GTC TAG CAG AA-3′) and *Env*-4 gp120 (5′-TTA TAT AAT TCA CTT CTC CAA TT-3′). The first round reverse transcription and amplification were performed using a Superscript One-Step RT-PCR® kit for long templates (Invitrogen, USA). The second round of the nested PCR was performed using platinum PCR Supermix High Fidelity® 120 (Invitrogen). The PCR amplicons were purified using a gel purification kit (QIAGEN) and cloned into a TA cloning vector (Real Biotech Corp.). Three randomly selected clones were sequenced (Macrogen, Korea) from each sample and one representative clone was selected for analysis. The sequence data obtained was submitted to GenBank with the following accession numbers assigned: JF300177; JF320891–JF320892 JF320894–JF320899; JF320903; JF320907–JF320908; JF320910; JF320913; JF320915; JF320920; and JN019801–JN019805.

Changes in HIV-1 diversity are posing a major challenge in the development of an effective vaccine. Viral heterogeneity is ever-increasing with the emergence of more complex viral mosaics or circulating recombinant forms (CRFs) arising through complex biological processes. In this study, we performed a detailed genetic analysis of part of the envelope sequence (C2–C5) of gp120 from 21 HIV-1-infected individuals living in north India. Subtypes were determined using HIV-1 genotyping tools [REGA version 2.0, Recombinant Identification Program 3.0 (RIP) and NCBI Genotyping tool]. Subtyping revealed that 85.7% (18/21) of the viruses detected in the patients were subtype C, with subtype A present in the remaining 14.3% (3/21) of subjects (Table 1 and Fig. 2). For phylogenetic analysis, viral sequences were aligned with reference sequences retrieved from the Los Alamos HIV database (http://hiv.lanl.gov) using CLUSTAL X. The phylogenetic tree was constructed with Mega (version 4) using the neighbor-joining method. Bootstrap values in the phylogenetic tree revealed that 18/21 viral sequences clustered with a subtype C Indian isolate, while three subtype A viruses aligned with subtypes A1 and A2 (Fig. 2).

We observed a range (11–17) of potential N-linked glycosylation sites (PNGS) within the C2–C5 region we examined. We further assessed the sequence and length corresponding to the V3, V4, and V5 regions and found that V3 was highly conserved as compared with V4 and V5 (Fig. 1). Sequence analysis of the donor viruses showed high sequence variability in the C3 region of gp120, especially in the 18 amino acid α2-helix residues (Fig. 1) consistent with previous findings in subtype C viruses (Gaschen *et al*., 2002).

We aligned the V3 sequences of viruses in patients with their corresponding consensus sequences (consensus-A and consensus-C V3) using Seqpublish. It was observed that V3 was conserved (35 amino acids; Table 1) except in two samples (AIIMS212 and AIIMS279), which had a one amino

Table 1. Comparison of V3 sequences and co-receptor usage in patients studied

AIIMS ID	Subtype	V3 sequence	Amino acid at 25/11 position of V3	Net charge on V3	V3 crown motif	Co-receptor used	Accession number
Con_A		CTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHC					
AIIMS235	A	---------R--H---------SNI-T------Y-	SI	4	GPGQ	CCR5	JF320896
AIIMS279	A	----------GVH----RV--*-NN------K-Y-	GN	6	GPGR	CXCR4	JF320908
AIIMS281	A	-------IIRRRIHIGPGRAFY-SNTR------Y-	RN	7	GPGR	CXCR4	JF320910
Con_C		CTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHC					
AIIMS287	C	----------	SD	4	GPGQ	CCR5	JF320913
AIIMS212	C	--------------------F--*----------	SD	4	GPGQ	CCR5	JF300177
AIIMS265	C	-I--S------------------------------	SD	4	GPGQ	CCR5	JN019803
AIIMS275	C	------------------------E---N------	SE	5	GPGQ	CCR5	JF320907
AIIMS278	C	----------------------R-E----------	SE	5	GPGQ	CCR5	JN019805
AIIMS245	C	-------------------A-----E----------	SE	4	GPGQ	CCR5	JN019801
AIIMS239	C	-I----------------------E----------	SE	4	GPGQ	CCR5	JF320897
AIIMS234	C	-I----------------------A--------Y-	SA	4	GPGQ	CCR5	JF320895
AIIMS224	C	---------R--------------A----------	SA	5	GPGQ	CCR5	JF320920
AIIMS264	C	----G-------------------E------K-Y-	SE	4	GPGQ	CCR5	JN019802
AIIMS261	C	----------------------------P--E---	SD	4	GPGQ	CCR5	JF320903
AIIMS269	C	-A--G------V--------------T--------	SD	4	GPGQ	CCR5	JN019804
AIIMS290	C	-I--S----------------------T------Y-	SD	3	GPGQ	CCR5	JF320915
AIIMS213	C	-V--G--------------------VT-N------	SD	5	GPGQ	CCR5	JF320891
AIIMS240	C	------------GM---R----------N--E---	SD	4	GPGR	CCR5	JF320898
AIIMS215	C	-----------V----------------------Y-	SD	3	GPGQ	CCR5	JF320892
AIIMS241	C	-----------T----------M------------	SD	4	GPGQ	CCR5	JF320899
AIIMS226	C	-I--------------------R------------	SD	5	GPGQ	CCR5	JF320894

The V3 sequences of envelopes from patients were aligned with their corresponding consensus sequences using Seqpublish (http://hiv.lanl.gov).
Based on the V3 sequences, the co-receptor in use [CCR5 (R5), CXCR4 (X4)] was predicted using Web PSSM. This program also showed the net charge on V3, and amino acids present at the co-receptor determining positions within V3 (11 and 25). All subtype C viruses were R5-tropic. AIIMS279 and AIIMS281 were predicted to be X4-tropic.
*indicates a deletion mutation in the V3 sequence

acid deletion (Table 1). Among the 18 clade C viruses analyzed, the GPGQ motif was conserved throughout, which is characteristic of subtype C viruses (Korber *et al.*, 1994). In AIIMS240, the glutamine residue (Q) was substituted for an arginine (R) at the apex of the crown (Table 1). Of the three clade A viruses, two (AIIMS279 and AIIMS281) contained a GPGR instead of a GPGQ motif (found in consensus-A V3) at the tip of the crown. Additionally, AIIMS281

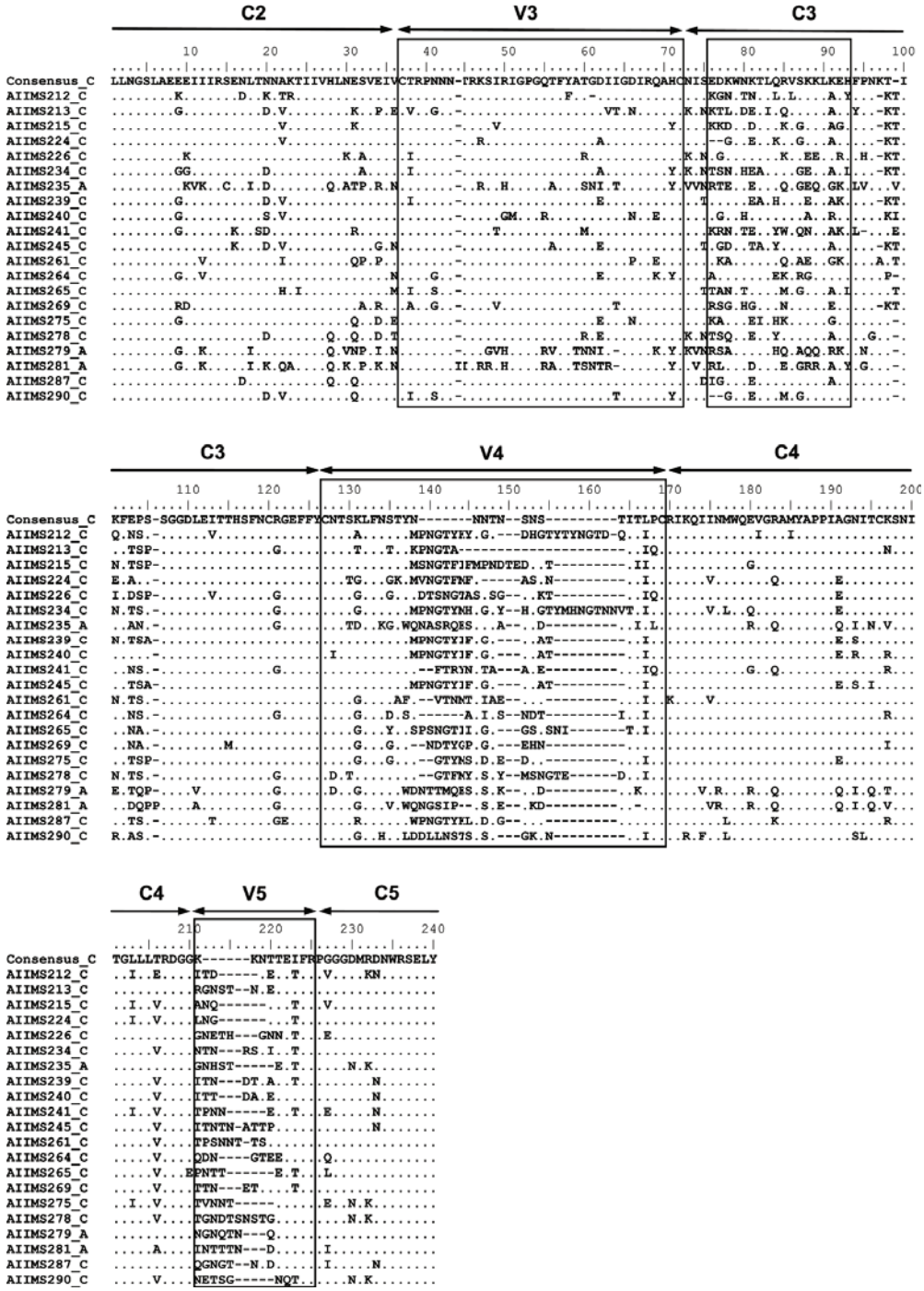

Fig. 1. Comparison of variable regions (V3, V4, and V5) from the viruses of patients studied. Deduced envelope sequences of viruses (partial C2–C5 of gp120) were aligned with corresponding consensus subtype C sequences using Clustal X program. The regions corresponding to third (V3), fourth (V4) and fifth (V5) variable segments of envelope gp120 protein are shown in boxes. A high conservation in V3 compared to V4 and V5 was observed. The median number of amino acids for V3, V4 and V5 was 35 (range, 34–35), 31 (22–41) and 12 (9–15), respectively, compared with 35, 25 and 9 in the corresponding consensus-C envelope sequence. Also the 18 amino acid residues forming the α2-helix in third constant (C3) region of gp120 is shown, which displays high variability particularly in subtype C viruses. Within the alignment, dots indicate identity with the consensus sequence and dashes indicate deletions.

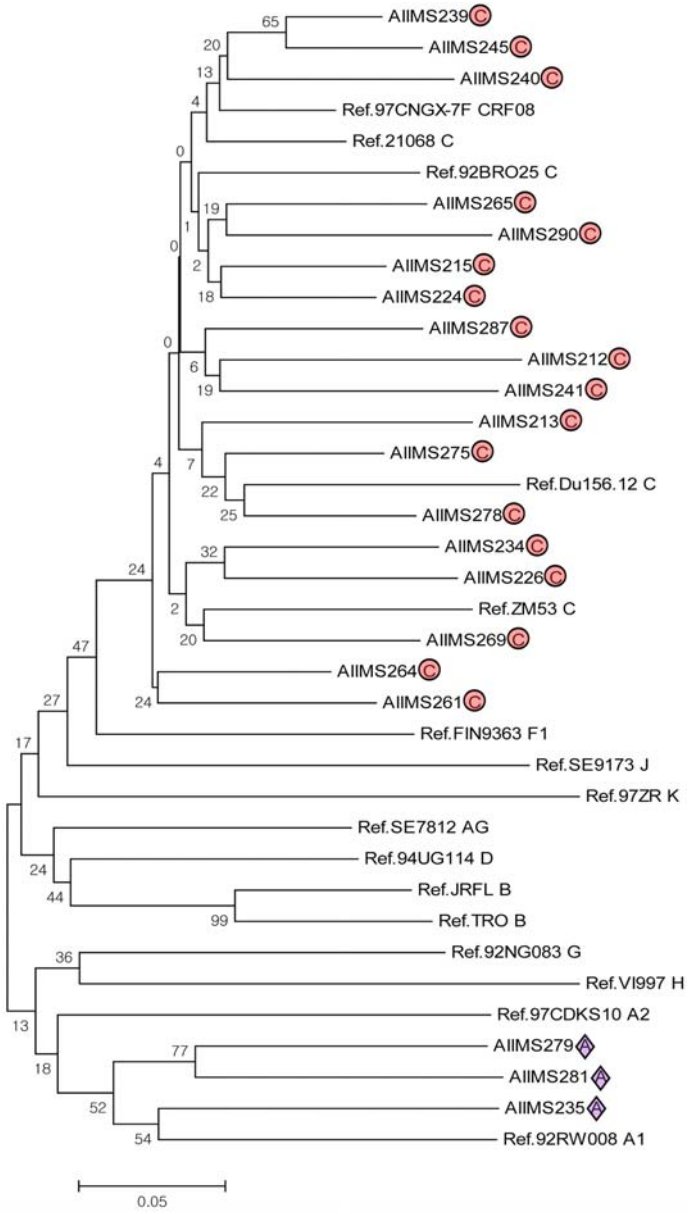

Fig. 2. Neighbor-joining phylogenetic tree constructed from HIV-1 gp120 envelope protein sequences. The Indian viral envelope sequences were aligned with HIV-1 reference sequences from major subtypes (A–K) and two circulating recombinant forms (CRF02_AG and CRF08_BC) using CLUSTAL X. The phylogenetic tree was constructed using Mega (version 4) employing the neighbor-joining method. Red circles indicate subtype C, and the purple diamonds indicate subtype A.

also contained an amino acid insertion and deletion at positions 8 and 27, respectively (Fig. 1).

The choice of co-receptor selectivity (CXCR4 or CCR5) is primarily determined by overall charge of the V3 region, but certain amino acid positions in V3 are particularly crucial. Based on the V3 sequence of viruses, we predicted the co-receptor in use through a web-based program (PSSM). All subtype C viruses were R5-tropic, while two of three subtype A viruses (AIIMS279 and AIIMS281) were predicted to be using CXCR4 (X4) as their co-receptor (Table 1). For confirmation of the co-receptor prediction, we compared a few patient virus V3 sequences with the V3 sequences of viruses tested previously for *in vitro* co-receptor usage and observed consistent results (Ndung'u *et al.*, 2006).

Phylogenetic analysis of the subtypes of donor viruses in this study showed that 85.7% of patients were infected with subtype C, consistent with previous reports from India (Hemelaar *et al.*, 2006; Lakhashe *et al.*, 2008). The remaining subjects were infected with subtype A viruses. The occurrence of a fraction of viruses not contained within clade C for these patients was important in the context of overall HIV-1 diversity across India where greater than 95% of infections are caused by subtype C viruses. Interestingly, subtype C viruses in this study showed a high degree of sequence similarity with CRF08_BC compared with subtype C. This is to be expected, as both of these subtypes have similar envelope sequences, but it may also be cautionary with respect to the possible emergence of more complex viruses in India. These findings could also be attributed to the small sample size, a limitation of the study, with assessment required in a greater number of patients from different regions across India. A few recent studies have shown the existence of non-

clade C viruses, including CRFs, from India, but again a small sample size and region-specific sample collection cannot consolidate these facts, and more data must be generated (Lakhashe *et al*., 2008; Neogi *et al*., 2009).

The V3 region of the HIV-1 envelope gp120 protein is important for co-receptor binding. This region is exposed as a consequence of the conformational changes induced by interaction of the CD4 binding site on gp120, and the CD4 receptor on host cells. Based on the V3 sequence of viruses, the co-receptor in use (X4 or R5) was predicted by web PSSM, which has a high degree of reproducibility. All viruses were predicted to be R5-tropic except for AIIMS279 and AIIMS281, which were predicted to be X-4 tropic. Coincidently, both these viruses had a GPGR at the V3 crown instead of GPGQ, which is commonly found in subtype A viruses. Further, V3 sequence analysis revealed that AIIMS240, a clade C virus, had an unusual Q to R transition at the crown apex (observed in less than 1% of subtype C viruses) (Patel *et al*., 2008) and a substitution from "RI to GM" just before the GPGR motif. A similar kind of a mutation has been found to be associated with dual co-receptor tropism (viruses using both CXCR4 and CCR5 as co-receptor) (Zhang *et al*., 2010). These observations correspond with those described in some recent studies, which showed that HIV-1 exploited several co-receptors in addition to CCR5 as a consequence of unique amino acid substitutions in the V3 crown region (Gharu *et al*., 2009; Zhang *et al*., 2010).

Overall, we observed high conservation in the V3 region, especially in subtype C viruses (Table 1). A reason for this could be that it is primarily involved in interactions with the co-receptor required for viral entry into host cells. Another plausible reason for the limited variation or conserved nature of V3 could be a low immune selection pressure within the V3 loop of the native virus. The latter hypothesis is further supported by the presence of a highly conserved hydrophobic core in the V3 crown (amino acid I14 in V3; Table 1), which favors immune evasion by epitope masking in subtype C viruses (Lynch *et al*., 2010). Our study is also suggestive of the emergence of non-subtype C virus infections in this population.

We thank all study participants. We acknowledge Prof. Miroslaw K. Gorny and Prof. Susan Zolla Pazner for their constant technical advice and support. We thank the Department of Biotechnology (BT/PR 10511/MED/29/66/2008) and Indian Council for Medical Research (61/7/2008-BMS) for funding this work. The AITRP (USA) and JRF fellowships provided by ICMR are gratefully acknowledged.

References

Andrabi, R., Choudhary, A.K., Bala, M., Kalra, R., Prakash, S.S., andey, R.M., and Luthra, K. 2011. Relative reactivity of HIV-1 polyclonal plasma antibodies directed to V3 and MPER regions suggests immunodominance of V3 over MPER and dependence of high anti-V3 antibody titers on virus persistence. *Arch. Virol.* **156**, 1787–1794.

Berger, E.A., Murphy, P.M., and Farber, J.M. 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **17**, 657–700.

Cecilia, D., Kulkarni, S.S., Tripathy, S.P., Gangakhedkar, R.R., Paranjape, R.S., and Gadkari, D.A. 2000. Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virology* **271**, 253–258.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T., and *et al.* 2002. Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360.

Gharu, L., Ringe, R., Pandey, S., Paranjape, R., and Bhattacharya, J. 2009. HIV-1 clade C env clones obtained from an Indian patient exhibiting expanded coreceptor tropism are presented with naturally occurring unusual amino acid substitutions in V3 loop. *Virus Res.* **144**, 306–314.

Hemelaar, J., Gouws, E., Ghys, P.D., and Osmanov, S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**, W13–23.

John, T.J., Babu, P.G., Jayakumari, H., and Simoes, E.A. 1987. Prevalence of HIV infection in risk groups in Tamilnadu, India. *Lancet* **1**, 160–161.

Korber, B.T., MacInnes, K., Smith, R.F., and Myers, G. 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. *J. Virol.* **68**, 6730–6744.

Lakhashe, S., Thakar, M., Godbole, S., Tripathy, S., and Paranjape, R. 2008. HIV infection in India: epidemiology, molecular epidemiology and pathogenesis. *J. Biosci.* **33**, 515–525.

Lynch, R.M., Rong, R., Li, B., Shen, T., Honnen, W., Mulenga, J., Allen, S., Pinter, A., Gnanakaran, S., and Derdeyn, C.A. 2010. Subtype-specific conservation of isoleucine 309 in the envelope V3 domain is linked to immune evasion in subtype C HIV-1 infection. *Virology* **404**, 59–70.

Modrow, S., Hahn, B.H., Shaw, G.M., Gallo, R.C., Wong-Staal, F., and Wolf, H. 1987. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. *J. Virol.* **61**, 570–578.

Ndung'u, T., Sepako, E., McLane, M.F., Chand, F., Bedi, K., Gaseitsiwe, S., Doualla-Bell, F., Peter, T., Thior, I., Moyo, S.M., and *et al.* 2006. HIV-1 subtype C *in vitro* growth and coreceptor utilization. *Virology* **347**, 247–260.

Neogi, U., Sood, V., Chowdhury, A., Das, S., Ramachandran, V.G., Sreedhar, V.K., Wanchu, A., Ghosh, N., and Banerjea, A.C. 2009. Genetic analysis of HIV-1 circulating recombinant form 02_AG, B and C subtype-specific envelope sequences from Northern India and their predicted co-receptor usage. *AIDS Res. Ther.* **6**, 28.

Pantophlet, R. and Burton, D.R. 2006. GP120: target for neutralizing HIV-1 antibodies. *Annu. Rev. Immunol.* **24**, 739–769.

Patel, M.B., Hoffman, N.G., and Swanstrom, R. 2008. Subtype-pecific conformational differences within the V3 region of subtype B and subtype C human immunodeficiency virus type 1 Env proteins. *J. Virol.* **82**, 903–916.

Richman, D.D. and Bozzette, S.A. 1994. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J. Infect. Dis.* **169**, 968–974.

Shankarappa, R., Chatterjee, R., Learn, G.H., Neogi, D., Ding, M., Roy, P., Ghosh, A., Kingsley, L., Harrison, L., Mullins, J.I., and *et al.* 2001. Human immunodeficiency virus type 1 env sequences from Calcutta in eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. *J. Virol.* **75**, 10479–10487.

Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E., and Hammer, S.M. 2008. The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **358**, 1590–1602.

Zhang, H., Tully, D.C., Zhang, T., Moriyama, H., Thompson, J., and Wood, C. 2010. Molecular determinants of HIV-1 subtype C coreceptor transition from R5 to R5X4. *Virology* **407**, 68–79.