American Medical Journal 1 (2): 140-147, 2010
ISSN 1949-0070

Frequency of P/S(XX)P Duplication and FRFE, Absence of LYP in P6*Gag* of Indian Human Immunodeficiency Virus-1 Subtype C Isolates

Krutika Wadekar, Sudhanshu Pandey, Preeti Jain, Vikas Chandra Roy,
Aanchal Asthana and Ramesh Paranjape
Department of Immunology, National AIDS Research Institute, Plot No.73,
'G' Block, MIDC, Bhosari, Pune 411 026 (MH), India

Abstract: Problem statement: The presence of a p6 domain at the C-terminus of the Gag polyprotein is a characteristic feature of HIV-1 and other primate lentiviruses. The p6Gag protein is considered as a major phosphoprotein in mature HIV-1 virions and is involved in the viral assembly and budding process. Sequence variation in the p6*gag* region in HIV-1 has been associated with changes in viral replication capacity and antiretroviral drug susceptibility. **Approach:** We examined sequence variation in the HIV-1 p6*gag* region using 38 isolates or infected Peripheral Blood Mononuclear Cells of HIV-1 subtype C and 20 additional Indian subtype C *gag* sequences from the Los Alamos database. Different patterns of insertions and deletions were observed in different motifs present in the p6 region. PTAP duplication was found in three strains. **Results:** We found PAP/TAP/LEPTAP/LPTVPTAP/ PAVPAAP like insertions at PTAP duplication site. KQE motif (84%) was observed as one of the most conserved motif in p6Gag while LYP motif was absolutely absent in Indian clade C sequences. Insertion of T/H/TT /PYRE/PYKE/EPKDRE was found instead of LYP motif. **Conclusion:** Further studies are needed to determine whether p6*gag* polymorphisms found in different motifs influence viral replication capacity, antiretroviral drug susceptibility, or other phenotypic properties of these strains.

Key words: p6*Gag*, HIV-1, LYP motif, KQE motif, PTAP duplication

INTRODUCTION

The HIV-1 *gag* gene encodes a polyprotein precursor, which is cleaved by HIV-1 protease to produce the Matrix (MA), Capsid (CA) and Nucleocapsid (NC) proteins, as well as p1, p2 and p6. While MA, CA and NC are common to all retroviruses, the presence of a p6 domain at the C-terminus of the *gag* polyprotein is a characteristic feature of HIV-1 (Gottlinger *et al*., 1991). The p6Gag protein is considered as a major phosphoprotein in mature HIV-1 virions and is involved in the viral assembly and budding process (Gottlinger *et al*., 1991; Muller *et al*., 2002). Amino acid polymorphisms and mutations in p6*gag* may influence viral replication capacity and drug susceptibility (Simon *et al*., 2003; Peters *et al*., 2001; Gatanaga *et al*., 2002).

Among the Gag domains of different subtypes of HIV-1, the p6 domain is by far the most variable, both in length and in sequence (Gottlinger *et al*., 1991). A P-X-X-P amino acid sequence motif (often PTAP or PSAP) near the N terminus of p6 region is conserved in most HIV-1 strains and appears to be important for the release of viral particles from infected cells (Huang *et al*., 1995; Marlowe *et al*., 2004). The 15FRFG and (LXX)4 motifs of p6 are essential for the incorporation of the regulatory viral protein Vpr into assembling HIV-1 virions (Kondo and Gottlinger, 1996; Lu *et al*., 1995; Zhu *et al*., 2004). Interactions between the p6Gag and several cellular proteins, Tsg101 and AIP1 are mediated by PT/SAP and LYP, LRSL motifs in p6Gag respectively which are mainly required for virus budding (Bates *et al*., 2004). The presence of treatment associated protease Cutting Site (CS) mutations in drug-naïve patients might lower the genetic barrier of first-line therapies with protease inhibitors. 449H/F, 451T, 452S, 453A of Gag are considered to be therapy associated mutation sites (Verheyen *et al*., 2009).

In one study it was found that p6 region of *gag* is involved in HIV-1 pathogenicity as small deletions in HIV-1 p6 are associated with non-progressive infection in humans (Alexander *et al*., 2000). In addition, two fairly conserved motifs KELY and KQE are also present in p6Gag.We report here

Corresponding Author: Ramesh Paranjape, Department of Immunology, National AIDS Research Institute, Plot No.73, 'G' Block, MIDC, Bhosari, Pune 411 026 (MH), India Tel: +91 20 2712 1343 Fax: +91 20 2712 1071

identification and characterization of polymorphisms, mutations and insertions in p6Gag in HIV-1 subtype C.

MATERIAL AND METHODS

Study subjects and virus isolates: As a part of an ongoing prospective study of full length sequencing of HIV-1, virus isolates from different regions of India and from patients in different stages of the disease, 38 isolates or infected Peripheral Blood Mononuclear Cells (PBMC) of HIV-1 infected persons from the period of 1999-2009 were procured from repository at National AIDS Research Institute. The study was approved by the Ethics Committee of the Institute and the identifiers were removed from the samples prior to sequencing.

Virus cultivation: Fresh PBMCs were isolated from blood obtained from healthy HIV uninfected persons by Ficoll-Hypaque (Sigma Aldrich) density gradient centrifugation and stimulated with Phytohemagglutinin (PHA-P) (5 µg mL^{-1}), in RPMI 1640 containing 10% fetal bovine serum at 37°C in 5% CO_2 for 48-72 h. Activated cells were co-cultivated with an equal number of PBMCs from HIV-1 positive individuals in RPMI 1640 containing 10% fetal calf serum and interleukin-2 (20 U mL^{-1}). The culture supernatants were tested for p24 antigen on 7th, 14th, 21st Post Infection Day (PID) by Enzyme-Linked Immunosorbent Assay (ELISA). Culture supernatants tested positive for p24 antigen, were stored at - 80°C for further use and PBMCS were taken for extraction of proviral DNA using QIAamp DNA blood mini kit (Qiagen, Hilden, Germany).

PCR amplification and cloning: Amplification of virtually full-length HIV-1 genomes was done using LTR specific primers; MSF12-5´GTAAAACGACGGCCAG3´ and MSR5-5´CAGGAAACAGCTATGAC 3´ with expand long template PCR system (Roche applied science). The amplified PCR products were cloned by using TA cloning method into vector pCR2.1 (Invitrogen, San Diego, CA) followed by transformation into *Escherichia coli* (TOP10F´) according to the manufacturer's instructions. Confirmation of clones was carried out by colony PCR using same primers. Colonies that were positive were selected and cultured using LB broth. Plasmid was extracted using Qiagen Miniprep kit (Qiagen, Hilden, Germany).

Sequencing: HIV-1 genome sequencing was carried out with an ABI 3730XL DNA Sequencer (Applied Biosystems Inc., Foster City, California, USA) by BigDye terminator method. Both DNA strands for clones were sequenced using primer walking method. The full length contig were generated and full length sequences were edited using SeqScape v 2.5 (Applied Biosystems) using HXB2 sequence as reference (Accession no. K03455).

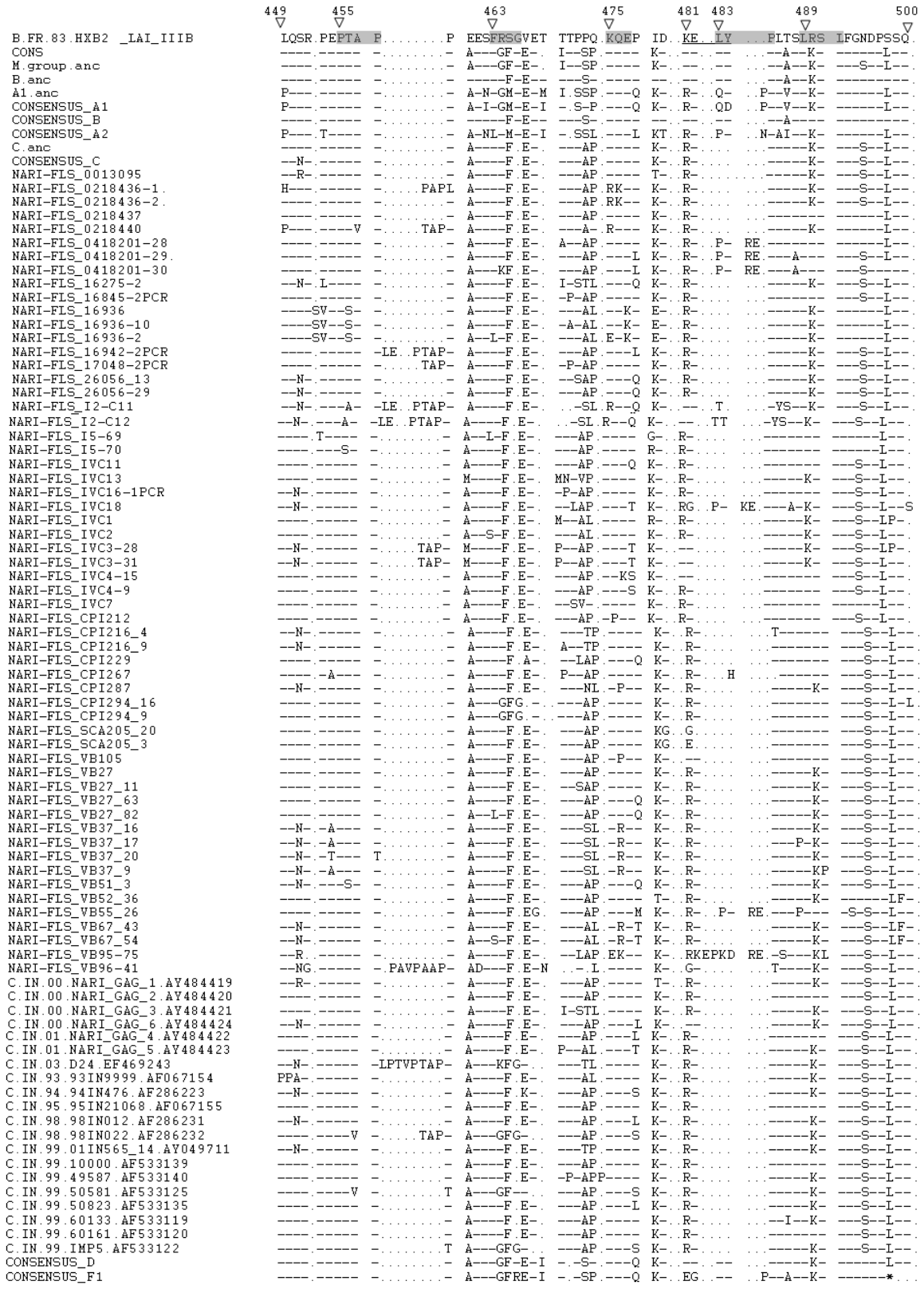
Analysis: When multiple clones were obtained from a sample, apart from representative, those clones showing dissimilarity in sequences were included for analysis. The pairwise and multiple alignments were done using Clustal W. Sequence alignment was done with the representative subtype sequences and previously reported HIV-1 *gag* sequences (Los Alamos National Laboratory, 2008) from Indian subtype C sequences. Phylogenetic analysis was done using neighbor joining method, maximum likelihood model and kimura-2-parameter. Finally, p6*gag* region was located and analyzed from the 38 full length sequences and 20 additional Indian subtype C *gag* sequences from the Los Alamos database of HIV-1 subtype C (Los Alamos National Laboratory, 2008).

RESULTS

Alignments of Gag amino acid sequences revealed extensive length variation with complex patterns of insertions and deletions at different positions throughout the p6 region (Fig. 1). The HXB2 p6 region is 52 amino acids in length while in Indian strains it was found to be 48-56 amino acids in length.

L449H, considered to be a therapy associated cutting site mutation, was found in one of the Indian sequence. L449P is a common feature of subtype A but it was found only in 3 subtype C strains. S451N was observed as most common polymorphism, found in seventeen (28.8%) Indian sequences. Also S451A (2 clones out of four) and S451R was found in three (5.2%) strains. There was deletion at position R452 in one strain and R452G substitution was found in one strain.

While PTAP motif was found to be highly conserved (90%), in some strains PSAP was found. Amino acid position 458 of *gag* is the most frequent site for insertion. Insertion of PAP/TAP/LEPTAP/LPTVPTAP/PAVPAAP was found in 1 (1.6%), 4 (6.8%), 2 (3.4%), 1 (1.6%), 1 (1.6%) cases respectively (Fig. 2). Duplication, partial duplication and variable amino acid sequences were also observed at 458th position of *gag*. The size and sequence of the amino acid insertions varied among the Indian HIV-1 strains. Among all sequences, PTAP duplication was found in only three strains. The alternative motif, PSAP was present in five sequences (8.4%). Out of 5 patients, PSAP motif was found in one clone each from the two samples, but not found in other clones.


449 455 463 475 481 483 489 500
B.FR.83.HXB2 _LAI_IIIB
CONS
M.group.anc
B.anc
A1.anc
CONSENSUS_A1
CONSENSUS_B
CONSENSUS_A2
C.anc
CONSENSUS_C
NARI-FLS_0013095
NARI-FLS_0218436-1
NARI-FLS_0218436-2
NARI-FLS_0218437
NARI-FLS_0218440
NARI-FLS_0418201-28
NARI-FLS_0418201-29
NARI-FLS_0418201-30
NARI-FLS_16275-2
NARI-FLS_16845-2PCR
NARI-FLS_16936
NARI-FLS_16936-10
NARI-FLS_16936-2
NARI-FLS_16942-2PCR
NARI-FLS_17048-2PCR
NARI-FLS_26056_13
NARI-FLS_26056-29
NARI-FLS_I2-C11
NARI-FLS_I2-C12
NARI-FLS_I5-69
NARI-FLS_I5-70
NARI-FLS_IVC11
NARI-FLS_IVC13
NARI-FLS_IVC16-1PCR
NARI-FLS_IVC18
NARI-FLS_IVC1
NARI-FLS_IVC2
NARI-FLS_IVC3-28
NARI-FLS_IVC3-31
NARI-FLS_IVC4-15
NARI-FLS_IVC4-9
NARI-FLS_IVC7
NARI-FLS_CPI212
NARI-FLS_CPI216_4
NARI-FLS_CPI216_9
NARI-FLS_CPI229
NARI-FLS_CPI267
NARI-FLS_CPI287
NARI-FLS_CPI294_16
NARI-FLS_CPI294_9
NARI-FLS_SCA205_20
NARI-FLS_SCA205_3
NARI-FLS_VB105
NARI-FLS_VB27
NARI-FLS_VB27_11
NARI-FLS_VB27_63
NARI-FLS_VB27_82
NARI-FLS_VB37_16
NARI-FLS_VB37_17
NARI-FLS_VB37_20
NARI-FLS_VB37_9
NARI-FLS_VB51_3
NARI-FLS_VB52_36
NARI-FLS_VB55_26
NARI-FLS_VB67_43
NARI-FLS_VB67_54
NARI-FLS_VB95-75
NARI-FLS_VB96-41
C.IN.00.NARI_GAG_1.AY484419
C.IN.00.NARI_GAG_2.AY484420
C.IN.00.NARI_GAG_3.AY484421
C.IN.00.NARI_GAG_6.AY484424
C.IN.01.NARI_GAG_4.AY484422
C.IN.01.NARI_GAG_5.AY484423
C.IN.03.D24.EF469243
C.IN.93.93IN9999.AF067154
C.IN.94.94IN476.AF286223
C.IN.95.95IN21068.AF067155
C.IN.98.98IN012.AF286231
C.IN.98.98IN022.AF286232
C.IN.99.01IN565_14.AY049711
C.IN.99.10000.AF533139
C.IN.99.49587.AF533140
C.IN.99.50581.AF533125
C.IN.99.50823.AF533135
C.IN.99.60133.AF533119
C.IN.99.60161.AF533120
C.IN.99.IMP5.AF533122
CONSENSUS_D
CONSENSUS_F1

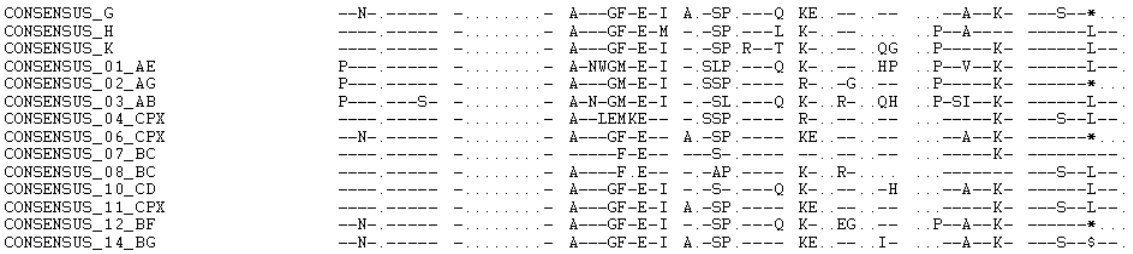


Fig. 1: Alignment of p6Gag region amino acid sequences of HIV-1 subtype C from India. Alignment of 38 HIV-1 subtype C sequences and 20 available *gag* subtype C sequences from Los Alamos database (Los Alamos National Laboratory, 2008) with concensus and ancestral sequences of different subtypes and CRFs from Los Alamos database (Los Alamos National Laboratory, 2008) in comparison to the HXB2 sequence were done using the Clustal W algorithm. Positions of different motifs in p6*gag* region are indicated in boxes that are shaded (grey) in HXB2 at the top for comparison. KELY motif is underlined

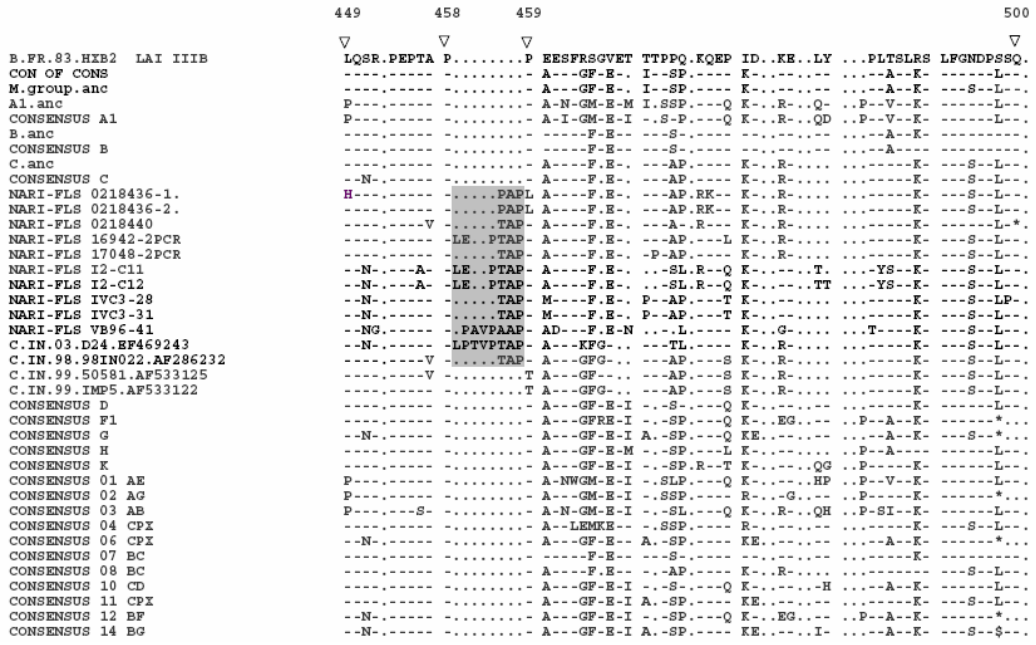


Fig. 2: Amino acid alignments in p6*Gag* Indian HIV-1 subtype C sequences showing duplication or partial duplication (shaded in grey) of PTAP motif (PAP /TAP/LEPTAP/LPTVPTAP/PAVPAAP) with consensus and ancestral sequences of different subtypes and CRFs from Los Almos database (year 2004) in comparison to HXB2 p6 sequence which is shown in the top

LYP motif was not found in any of the Indian sequences. Insertion of T/H/TT /PYRE/PYKE/EPKDRE was found instead of LYP motif (Fig. 3). All the above amino acid insertions were found in only one sample each except PYRE which was found in two samples. LRSL motif was seen in some of the Indian strains (35%) but instead of that, LKSL motif was also observed in more sequences (62%) (Fig. 3). (LXX)4 motif was replaced by (LXX)3 in 95% of the Indian strains. Insertion of FRFE/SRFG/LRFE/FKFE/FGFE was found in place of FRFG, out of which FRFE was the most common and was found in 91% of Indian strains. KELY was seen only in one HIV-1 strain whereas there was an insertion of RE in many samples (79%). It was observed that KQE motif was the highly conserved motif (84%) but variations of the KQE motif were seen in some sequences like RKE/RQE/EQK/KQK/KPE/KRE/EKE in 1 (1.6%), 2 (3.4%), 1 (1.6%), 1 (1.6%), 3 (5%), 2 (3.4%), 1 (1.6%) cases respectively.

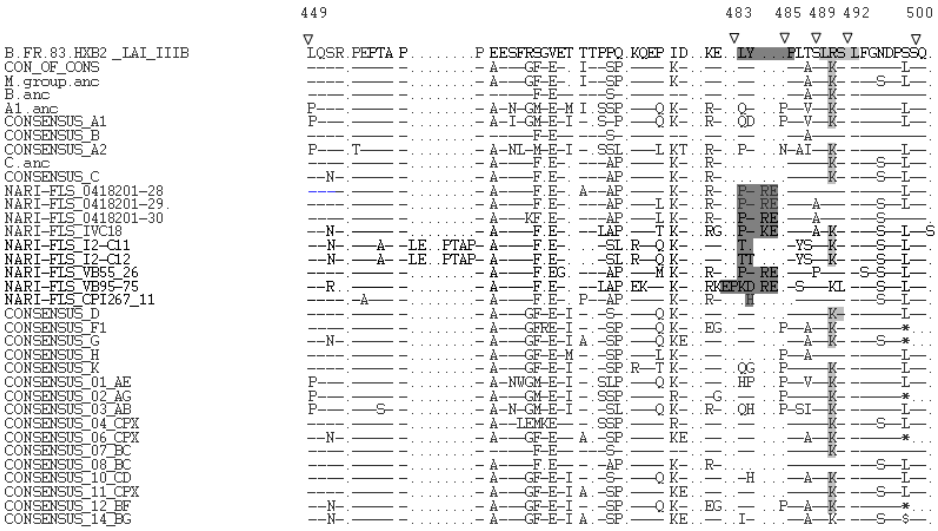


Fig. 3: Amino acid alignments in p6Gag Indian HIV-1 subtype C sequences showing insertions (shaded grey) (T/H/TT/PYRE/PYKE/EPKDRE) in place of LYP motif with concensus and ancestral of different subtypes and CRFs from Los Alamos database (Los Alamos National Laboratory, 2008) in comparison to HXB2 p6 sequence which is shown in the top. LRSL motif is replaced by LKSL in some samples (highlighted in grey)

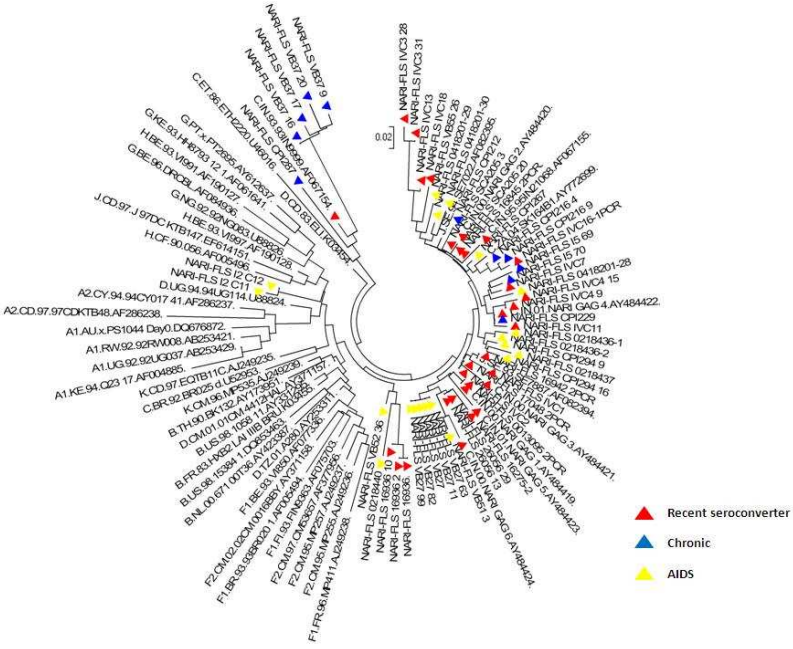


Fig. 4: Phylogenetic tree showing clustering pattern of p6*gag* of HIV-1 from India. Different colors show different clinical stages of the collected sample

The phylogenetic analysis of p6*gag* Indian HIV-1 subtype C sequences along with the previously reported Indian subtype C sequences based on different clinical status (Fig. 4) and year wise collection of samples (Fig. 5) demonstrated that all the sequences were randomly distributed and no clustering was seen associated with year of collection of samples or the stage of infection.

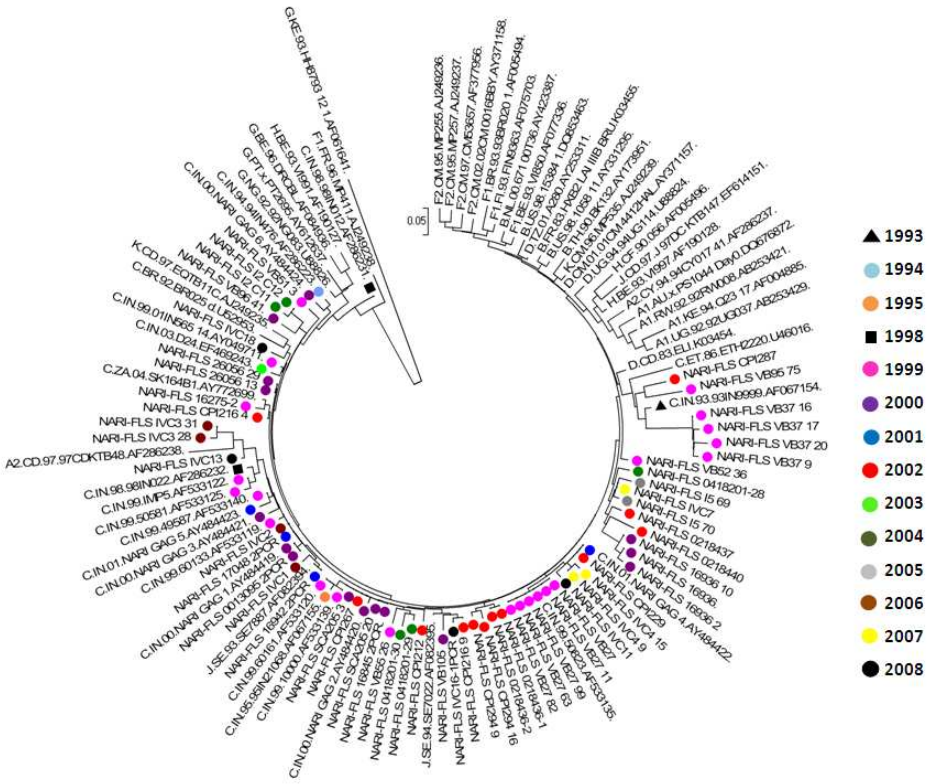


Fig. 5: Phylogenetic tree showing clustering pattern of p6*gag* of HIV-1 from India collected between 1999-2008. Different colors shows different years in which sample was collected

DISCUSSION

Gag is one of the most conserved gene among the HIV-1 virus genes. However, p6 domain of *gag* shows greater variability. Variability in p6*gag* has been shown to influence drug susceptibility and replication capacity of the virus. These observations are based on the studies on non-subtype C HIV-1 viruses. We have sequenced full genome of strains of HIV virus and it has provided opportunity to study the molecular diversity in p6*gag* in Indian HIV-1 subtype C strains. Phylogenetic analysis of p6*gag* did not show any clustering in association with either the clinical stage of the infection or the year of isolation.

The different p6Gag motifs with their role in virus life cycle have been evaluated. Interactions between the p6Gag and cellular protein, Tsg101 is mediated by PT/SAP motif which is required for virus budding (Bates *et al*., 2004). Some findings suggest that PTAP duplications may influence the virological response to antiretroviral therapy (Lastere *et al*., 2003). PTAP duplication was found in three HIV-1 C strains. We have found PAP/TAP/LEPTAP/LPTVPTAP/PAVPAAP like insertions at PTAP duplication site. Duplication or partial duplication of PTAP was seen more frequently in Indian subtype C viruses compared to HIV-1 subtype B reported in literature. Since the duplication is associated with higher replicative capacity, this implies that HIV-1 subtype C may show greater disease progression. However it needs to be substantiated by virus fitness studies in subtype C isolates with these insertions.

FRFG motif is essential for Vpr packaging into the virion and LYP motif is involved in viral budding (Zhu *et al*., 2004; Bates *et al*., 2004). Instead of FRFG in Indian strains FRFE has been found in all the sequences and LYP motif was totally absent in Indian clade C sequences. This suggests that LYP motif may not play any significant role in virus propagation. Subtype specific variation in FRFG motif such as FGMG/FGMR (A1, A2 and AG), WGMG (AE), FGFG (D,G,F1,F2,K,H) FRFG (B) and FRFE (C) has been found. Subtype B with one +ve charge and C with one +ve and one –ve charge at this motif may provide increased Vpr packaging in comparison to other subtypes with no charge.KQE is one of the most conserved motif in p6. In Indian clade C along with

KQE some variations like RKE/RQE/EQK/KQK/KPE/KRE/EKE were also found. Song *et al*. (2007) reported that there was no significant effect on viral replication and fitness even after deletion of KQE motif, near the C-terminus of p6 (Song *et al*., 2007). Substitution of Y36 of KELY in association with a substitution at downstream L41 results in reduced infectivity and failure to incorporate the Env protein into virus particles (Gabriela *et al*., 2004), but Y36/L41 was not present in Indian population except in one sample. Y36 is absent in Indian subtype C however L41 is conserved in Indian subtype C strains. Thus, there is a need to study the functional role of conserved motifs KQE and KELY using new approaches.

CONCLUSION

Thus, our analysis of p6*gag* sequences from Indian strains highlights the distinct patterns of polymorphisms which are not so common in other subtypes. Future studies are needed to test whether *gag* region polymorphisms in different subtypes and CRFs influence phenotypic properties, disease progression, development of antiretroviral drug resistance, or the clinical response to antiretroviral therapy and viral fitness.

ACKNOWLEDGEMENT

The study was supported by Department of Biotechnology (DBT) grant BT/PR7054/Med/14/936/2006. We kindly acknowledge Dr. J.B. Bhattacharya for his help in this study and department of Molecular virology, (NARI) for providing infrastructure. We would also like to thank members of Molecular Virology for their support.

REFERENCES

Alexander, L., E. Weiskopf, T.C. Greenough, N.C. Gaddis and M.R. Auerbach *et al*., 2000. Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. J. Virol., 74: 4361-4376. PMID: 10756051

Bates, M., C. Chappey and N. Parkin, 2004. Mutations in p6gag Associated with Alterations in replication capacity in drug sensitive HIV-1 are implicated in the budding process mediated by TSG101 and AIP1. Proceeding of the 11th Conference on Retroviruses and Opportunistic Infections, Feb. 8-11, NLM Gateway, South San Francisco CA, USA., pp: 1-1. http://gateway.nlm.nih.gov/MeetingAbstracts/ma?f=102270834.html

Gabriela, B., S. Peters, R. Martinez, D. Cmarko, P. Meylan and A. Telenti, 2004. The central region of human immunodeficiency virus type 1 p6 protein (gag residues S14-I31) is dispensable for the virus *in vitro*. J. Gen. Virol., 85: 921-927. PMID: 15039534

Gatanaga, H., Y. Suzuki, H. Tsang, K. Yoshimura and M.F. Kavlick *et al*., 2002. Amino acid substitutions in gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. J. Biol. Chem., 277: 5952-5961. PMID: 11741936

Gottlinger, H.G., T. Dorfman, J.G. Sodroski and W.A. Haseltine, 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc. Natl. Acad. Sci., 88: 3195-3199. PMID: 2014240

Huang, M., J.M. Orenstein, M.A. Martin and E.O. Freed, 1995. p6gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. J. Virol., 69: 6810-6818. PMID: 7474093

Kondo, E. and H.G. Gottlinger, 1996. A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. J .Virol., 70: 159-164. PMID: 8523520

Lastere, S., C. Dalban, G. Collin, D. Descamps and P.M. Girard *et al*., 2003. Impact of amino acid insertions in HIV-1 p6 PTAP region on the virological response to amprenavir in the NARVAL trial. Proceeding of the 10th Conference on Retroviruses and Opportunistic Infections, Feb. 10-14, NLM Gateway, Boston, MA USA., pp: 1-1. http://gateway.nlm.nih.gov/MeetingAbstracts/ma?f=102261223.html

Los Alamos National Laboratory, (LANL), 2008. HIV sequence alignments. http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html

Lu, Y.L., R.P. Bennett, J.W. Wills, R. Gorelick and L. Ratner, 1995. A leucine Triplet repeat sequence (LXX) 4 in p6gag is important for Vpr incorporation into human immunodeficiency virus type particles. J. Virol., 69: 6873-687. PMID: 7474102

Marlowe, N., T. Flys, J. Hackett, M. Schumaker, J.B. Jackson and S.H. Eshleman, 2004. Analysis of insertions and deletions in the gag p6 region of diverse HIV type 1 strains. AIDS Res. Hum. Retroviruses, 20: 1119-1125. PMID: 15585104

Muller, B., T. Patschinsky and H.G. Krausslich, 2002. The Late-Domain Containing Protein p6 is the Predominant Phosphoprotein of human immunodeficiency virus type 1 particles. J. Virol., 76: 1015-1024. PMID: 11773377

Peters, S., M. Munoz, S. Yerly, V. Sanchez-Merino and C. Lopez-Galindez *et al*., 2001. Resistance to nucleoside analog reverse transcriptase inhibitors mediated by human immunodeficiency virus type 1 p6 protein. J. Virol., 75: 9644-9653. PMID: 11559796

Simon, V., N. Padte , D. Murray, J. Vanderhoeven and T. Wrin *et al*., 2003. Infectivity and replication capacity of drug-resistant human immunodeficiency virus type 1 variants isolated during primary infection. J. Virol., 77: 7736-7745. PMID: 12829813

Song, Y.H., Z.F. Meng, H. Xing, Y.H. Ruan and X.P. Li *et al*., 2007. Analysis of HIV-1 CRF07_BC gag p6 sequences indicating novel deletions in the central region of p6. Arch. Virol., 152:1553-1558. PMID: 17458619

Verheyen, J., E. Knops, B. Kupfer, O. Hamouda and S. Somogyi *et al*., 2009. Prevalence of C-terminal gag cleavage site mutations in HIV from therapy-naive patients. J. Inform., 58: 61-67. PMID: 19110315

Zhu, H., H. Jian and L. Zhao, 2004. Identification of the 15FRFG domain in HIV-1 gag p6 essential for Vpr packaging into the virion. Retrovirology, 1: 26-30. PMID: 15363109