AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 30, Number 6, 2014

DOI: 10.1089/aid.2013.0275

Phylogenetic Characterization of Six Full-Length HIV-1 Subtype C Molecular Clones from Three Patients: Identification of Rare Subtype C Strains Containing Two NF-κB Motifs in the Long Terminal Repeat

Luke Elizabeth Hanna,[1,*,†] Ujjwal Neogi,[2,*] Udaykumar Ranga,[3] Soumya Swaminathan,[4] and Vinayaka R. Prasad[1]

Abstract

Molecular surveillance is the backbone of HIV-1 vaccinology. Full-length HIV-1 sequences are useful tools that can provide a better understanding of the epidemiology in a given region. A limited number of full-length HIV-1 sequences are available from India, where >95% of the HIV infections are due to HIV-1 subtype C (HIV-1C), which is distinct from the prototype African HIV-1C. In this study, we sequenced six full-length clones isolated from three patients. Extensive phylogenetic analyses of the full-length viral sequences using bioinformatic tools identified a separate cluster of Indian strains, thus confirming the distinct phylogenetic identity of the Indian HIV-1C. Notably, the long terminal repeat (LTR) of two of the six molecular clones contained only two NF-κB binding sites. The sequences also displayed features characteristic of HIV-1C including a Tat dicysteine motif, a shortened Rev open reading frame, and a predicted CCR5 coreceptor tropism for gp120 of three of the proviral sequences.

Human immunodeficiency virus type 1 subtype C (HIV-1C) accounts for nearly 50% of global HIV infections and is the predominant HIV subtype in South Africa and India.[1] We and others reported that the Indian HIV-1C (referred to as HIV-$1C_{IN}$) epidemic originated in the 1970s from a single or a few genetically related strains from Africa.[2–4] Following its introduction into India, HIV-1C evolved independently and is responsible for the current HIV-1C epidemic in India, which is distinct from that in Africa.[2,4–6] HIV-$1C_{IN}$ displays several interesting genotypic and phenotypic properties. For example, the Tat protein of HIV-$1C_{IN}$ contains a C31S polymorphism in the otherwise highly conserved C30C31 motif.[7] Studies in a severe combined immune deficiency mouse HIV encephalitis (SCID-HIVE) model have shown a reduced neuropathogenic potential of HIV-$1C_{IN}$ isolates.[8] HIV-1C isolates circulating in Southern African countries contain higher frequencies of variants with an intact Tat-C30C31 motif correlating with the higher prevalence of HAD and increased neuropathogenesis.

Conversely, the HIV-$1C_{IN}$ isolates uniformly contain a C31S polymorphism and display lower neuropathogenic potential.[5,8] Furthermore, subtype C appears to demonstrate extensive variations in the viral enhancer unlike other subtypes. Furthermore, the viral enhancer of subtype C is characterized by the presence of three NF-κB binding sites unlike other subtypes, which mostly contain only two such elements. Recent evidence shows the rapid emergence of several enhancer variant viral strains of HIV-1 in India.[9] The enhancer variant viral strains contain an additional NF-κB site (four NF-κB sites) and/or an RBEIII motif, thus demonstrating a higher magnitude of genetic variation not seen in other viral subtypes.

Given the complexity of the HIV-1 epidemic driven by intrasubtype and intersubtype recombinants, accurate genetic identification of HIV is important for population-based molecular epidemiology studies. The availability of full-length molecular clones and sequences is critical to address the evolving subtype C HIV epidemic. Currently, only 21 full-length HIV-$1C_{IN}$ sequences are available from 15 individuals and most of the samples were collected in the 1990s (between 1993 and 1999). Given the rapidly evolving HIV-1 subtype C epidemic in India,[10] a periodic phylogenetic

[1]Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York.
[2]Hematology Research Unit, Department of Molecular Medicine, St. John's Research Institute, Bangalore, India.
[3]Molecular Biology Unit, Jawaharalal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India.
[4]National Institute for Research in Tuberculosis, Chennai, India.
*These authors contributed equally to this work.
[†]Current affiliation: National Institute for Research in Tuberculosis, Chennai, India.

analysis of full-length genome sequences is important to understand viral evolution.

Here, we describe the isolation of six full-length molecular clones of HIV-1 subtype C from three different subjects from Tamilnadu, one of the states of India with the highest HIV prevalence rates. Furthermore, the sequences of these full-length molecular clones were subjected to extensive molecular and phylogenetic analyses. Importantly, we show that the long terminal repeat (LTR) of two of the six molecular clones contains only two NF-κB binding sites, a feature that is uncommon for subtype C that is rarely reported.[11]

The study was approved by the institutional ethical review boards of the National Institute for Research in Tuberculosis (NIRT), Chennai, India, and the Albert Einstein College of Medicine, Bronx, NY. Written informed consents were obtained from HIV-1-infected individuals and healthy donors prior to blood collection.

Single peripheral blood samples were collected between 2004 and 2005 from three subjects attending the clinic at the National Institute for Research in Tuberculosis (NIRT), Chennai, India. Subject characteristics are presented in Table 1. Among the three subjects, two were therapy naive (NIRT379 and NIRT333) while patient NIRT723 stopped therapy during the time of blood collection. Peripheral blood mononuclear cells (PBMCs) were isolated from blood using the density-gradient centrifugation technique, and the PBMC were cocultured with CD8 T cell-depleted and PHA-stimulated PBMCs obtained from HIV-negative healthy donors. Genomic DNA was extracted from cultured cells using the GenElute blood genomic DNA kit (Sigma-Aldrich, USA). The full-length proviral DNA was amplified as two fragments, a 3.0-kb fragment extending from the 5' LTR through the reverse transcriptase (RT) region and a 6.5-kb fragment spanning the rest of the provirus until the end of the 3'-LTR by nested PCR, as described previously.[12]

The DNA fragments were digested with *Not*-I and *Mlu*-I and *Mlu*-I and *Sbf*-I, respectively, and cloned into the commercial pNEB193 vector (New England Biolabs, USA) using *Not*-I and *Sbf*-I. Full-length clones were identified by restriction digestion and subjected to full-length HIV genome sequencing by Sanger sequencing using 40 individual sequencing primers placed approximately 500 nucleotides apart. The sequence contigs were assembled using Lasergene software (DNASTAR, Inc., Madison, WI) to generate the complete nucleotide sequence of individual open reading frames. From these contigs, the full-length sequences were generated manually. The sequences were submitted to GenBank under the accession numbers KF766537–KF766542.

The six full-length sequences were subjected to extensive phylogenetic analysis. To this end, a total of 68 full-length reference HIV-1 subtype C sequences were downloaded from the Los Alamos database from different continents representing 16 different countries (Africa: Brazil, Botswana, South Africa, Zambia, Djibouti, Ethiopia, Malawi, Tanzania, Kenya; North America: United States; Latin America: Brazil, Uruguay, and Argentina; and Southeast Asia: India, China, and Myanmar). Phylogenetic and cluster analyses were performed in Molecular Evolutionary Genetics Analysis software version 5.2.2 (MEGA 5.2.2) with general time reversible with inverse gamma distribution (GTR + G + I), which has been predicted as the best-fit model.[13] The individual protein sequences were aligned using Clustal W2.[14] Genotypic coreceptor tropism was analyzed using Geno2Pheno$_{[co\text{-}receptor]}$ software with a 10% false-positive rate, which has been shown to be highly sensitive and specific in determining X4 tropism in subtype C (unpublished data). Similarity analysis was performed using SimPlot version 3.5.1.[15] Drug resistance mutations were identified using the Stanford University HIV Drug Resistance Database (HIVdB program, http://hivdb.stanford.edu) accessed in September 2013.

The HIV-1 subtyping analysis confirmed the subtype C nature of all the six molecular clones. The phylogenetic analysis, which is based on the full-length viral sequences, identified a separate cluster consisting of only the Indian (HIV-1C$_{IN}$) strains, thus confirming a distinct phylogenetic identity of the Indian HIV-1 subtype C (Fig. 1).

A detailed molecular analysis of the six full-length viral sequences identified several interesting and unusual features in the viral LTR and protein coding sequences including unique genetic polymorphisms, deletions, and insertions. The LTR in four of the six molecular clones contained three NF-κB (3-κB) sites, typical of subtype C. In the other two clones (C.IN.05.NIRT333.1 and C.IN.05.NIRT723.1), interestingly, the LTR contained only two NF-κB (2-κB) sites (Fig. 2A), a molecular feature uncommon for subtype C with a small number of such variants reported previously.[11] Importantly, as was observed previously, the C-κB motif (5′ GGGGCG TTCC 3′), which is unique for subtype C, was retained in

Table 1. Demographic Profile of Study Subjects

Patient ID	*Age (years)/sex*	*Time since diagnosis (years)*[a]	*Year of sampling*	*History of prior ART*	*CD4 count (cells/μl of blood)*[b]	*Viral load (copies/ml of plasma)*[b]	*Clone (accession no.)*	*Coreceptor usage*
NIRT379	22/M	2.5	2004	No	486	296,000	C.IN.04.NIRT379.1(KF766537)	CCR5
							C.IN.04.NIRT379.2(KF766538)	CCR5
							C.IN.04.NIRT379.3(KF766539)	CCR5
NIRT333	27/F	4.0	2005	No	343	121,000	C.IN.05.NIRT333.1(KF766540)	CXCR4
NIRT723	32/F	4.5	2005	Yes	392	502.000	C.IN.05.NIRT723.1(KF766541)	CCR5
							C.IN.05.NIRT723.2(KF766542)	CD[c]

[a]Time since diagnosis of HIV infection at the time of sampling.
[b]CD4 count and viral load measurement were performed on the same sample that was used for cloning.
[c]CD, cannot be determined.
ART, antiretroviral therapy.

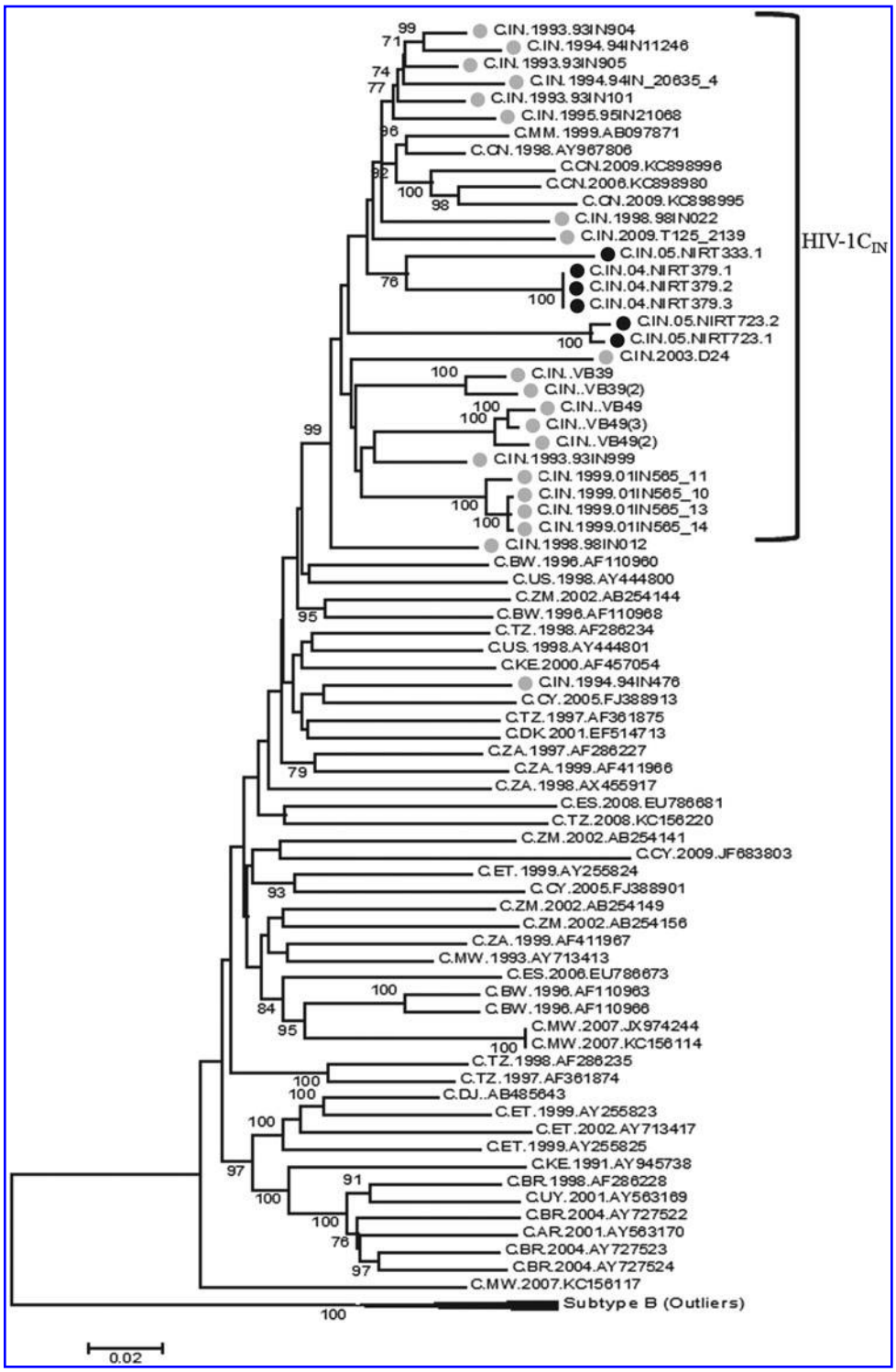

FIG. 1. Phylogenetic analysis of full-length HIV-1 sequences. Phylogenetic analysis was performed using full-length sequences of six clinical isolates and 68 sequences downloaded from the Los Alamos database. The database-derived sequences represent HIV isolates from 16 different countries in Africa, North America, Latin America, and Southeast Asia. All the available 21 full-length sequences from India were also included. The phylogenetic tree was constructed with general time reversible with inverse gamma distribution (GTR + G + I) in MEGA 5.2.2 software. All positions containing gaps and missing data were eliminated. There were a total of 7,472 positions in the final dataset. Sequences of clones that are the subject of this study are indicated with black circles and other Indian strains are indicated with gray circles. The cluster of Indian HIV-1C isolates is indicated with a square bracket.

these two clones suggesting its critical function in subtype C. In all subtype C strains, including the two we isolated that contained only two NF-κB sites, it is invariably one of the two H-κB sites (5′ GGGACTTTCC 3′), which is deleted but never the C-κB site.

Furthermore, of the two clones isolated from the subject NIRT723, one contained two NF-κB sites (C.IN.05.-NIRT723.1) and the second clone contained three sites (C.IN.05.NIRT723.2) suggesting a mixed infection in this subject. However, in the subject NIRT_333, we were unable to sequence a second clone to determine if it again represented a mixed population. The SimPlot identified >90% similarity with the consensus subtype C sequence for the LTR of all the six clones, thus precluding the possibility of a short-fragment recombination event (Fig. 2B). A similar result was obtained when the full-length viral genome was used for

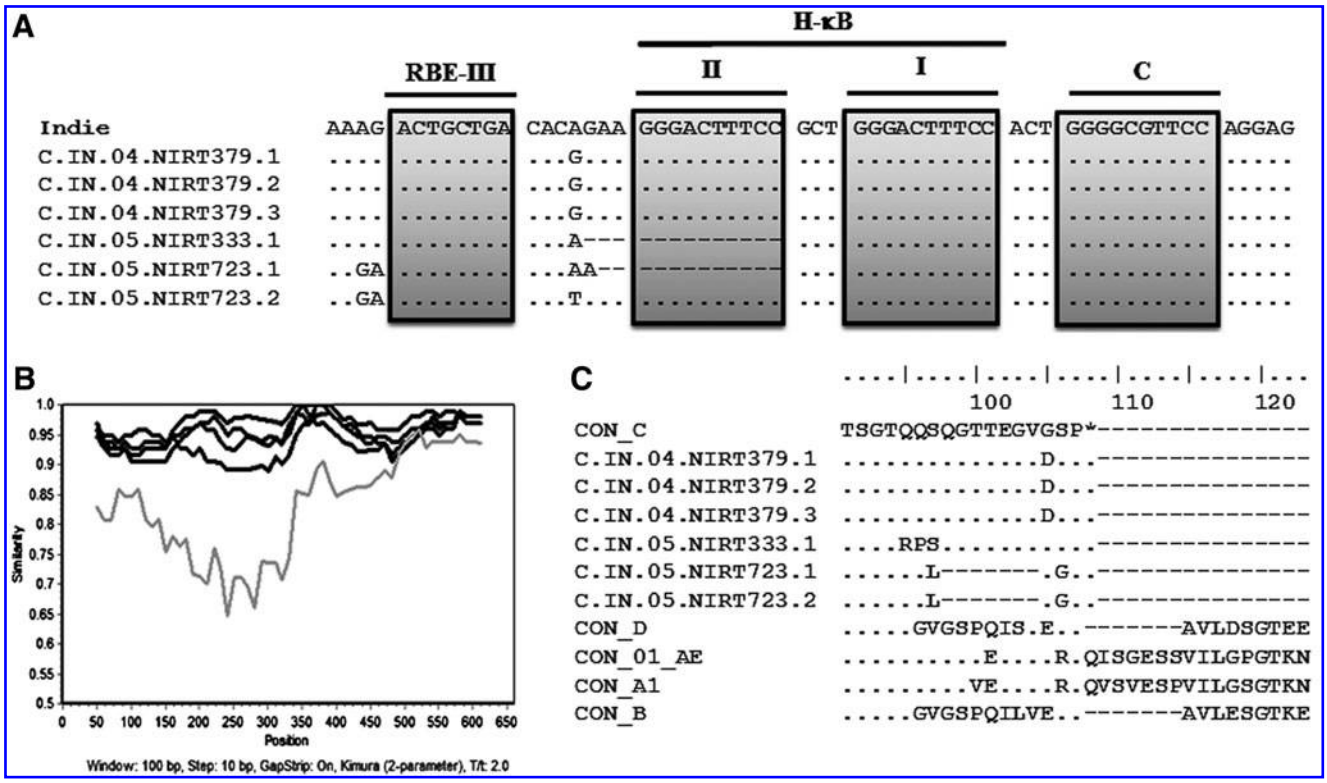

FIG. 2. Genetic characterization of HIV-1C long terminal repeat (LTR) and Rev. **(A)** Multiple sequence alignment of the NF-κB region of 5′-LTR. **(B)** SimPlot analysis of the full-length 5′-LTR of the clinical isolates was performed to detect short fragment recombination events. The distance from consensus C is plotted; the light gray line indicates the HXB2 (K03455) subtype B sequence used as a control. Each of the black lines indicates genetic similarity of one of our new Indian sequences to the subtype C consensus. **(C)** Multiple sequence alignment of C-terminal Rev Exon-II identified a premature stop codon and a large seven amino acid deletion in the Indian strains. Consensus sequences (2004) were downloaded from the Los Alamos database.

SimPlot analysis with a 500 nt window size and 20 nt step size (data not shown).

In addition to cis-acting sequences, we also examined predicted amino acid sequences of key HIV polyproteins, Gag, Gag-Pol, and Env, which revealed another important signature—the absence of L35Y36 residues in the ALIX binding site of the HIV-1C Gag p6 late domain. It is well known that subtype B HIV-1 Gag p6 protein encodes two late domain motifs that assist in the recruitment of host proteins that facilitate HIV budding. The PTAP motif recruits TSG101, while the LYPXnL motif recruits ALIX protein for budding. HIV-1C, however, lacks the L35Y36 residue in the LYPXnL motif and thus appears to depend solely on PTAP.[16] All of the clones studied here showed the LY deletion in the Gag p6 late domain, a signature of HIV-1 subtype C.

We next examined the Pol region of the Gag-Pol polyprotein sequence. One of the study subjects (NIRT723) was previously treated with reverse transcriptase inhibitors (RTIs). Therefore, we examined the RT region for the presence of RTI-resistance mutations. RTI-resistance mutations were observed only in NIRT723 but not the other two subjects (NIRT379 and NIRT333). Both the clones derived from NIRT723 showed the presence of the nucleoside analog reverse transcriptase inhibitor (NRTI) resistance mutations M184V, T215F, and K219E, and the non-nucleoside reverse transcriptase inhibitor (NNRTI) resistance mutations Y181C, G190A, and N348I. These six mutations together confer resistance to all the available NRTIs and NNRTIs. No mutations conferring resistance to protease inhibitors or integrase inhibitors (in the Pol region) were observed, as expected.

A genotypic analysis to predict coreceptor tropism of the six molecular clones, based on the V3 region of envelope, revealed that the three clones from subject NIRT_379 and one of the two clones (C.IN.05.NIRT723.2) obtained from NIRT_723 harbored the CCR5-tropic envelope. In contrast, the single clone (C.IN.05.NIRT333.1) isolated from subject NIRT333 encoded the CXCR4-tropic envelope. This is due to a mutation in position 13 (H13G) and 25 (E25D). However, the signature GPGQ motif in HIV-1 subtype C is conserved in this sequence. We could not, however, determine the tropism of the second clone (C.IN.05.NIRT723.2) of NIRT_723 due to an inability to align V3 region sequences due to multiple mismatches after position 17 in the V3 region just downstream of the GPGQ motif.

We also examined Tat and Rev, the two key regulatory proteins required for viral replication and pathogenesis. In the predicted Tat protein sequence, among the six subtype C specific signature residues (31S, 35L, 39Q, 57S, 60P, and 63E), with the exception of 60P, the other five residues were conserved.[7] The 60P is replaced with 60Q, which is the signature residue for the Indian Tat subtype C sequences.[17] All six HIV clones contained a premature termination at position

108 (Rev Exon-II), while the consensus clade B Rev sequence is 116 amino acids long (Fig. 2C). Further analysis of a large number of subtype C ($n=1329$), B ($n=3502$), A1/A2 ($n=320$), D ($n=164$), and 01_AE ($n=572$) sequences downloaded from the Los Alamos database revealed that this stop codon was unique to subtype C and that 94.3% (1254/1329) of global HIV-1C strains displayed this stop codon, and this was statistically significant ($p<0.001$) when compared to non-C strains [B and D (0.6%), A1/A2 (5.9%), and 01_AE (14.3%)]. This feature was previously observed in the South African subtype C strains as well.[18] An additional seven amino acid deletion was observed in Rev exon II in both the clones isolated from subject NIRT723, which is also seen in the infectious subtype C molecular clone pIndie-C1. However, the functional consequences of this deletion are not known (Fig. 2C). The absence of the C-terminal tail (residues 103–123) in the Rev protein from a majority of subtype C isolates suggests that this sequence is dispensable for subtype C Rev function.

Of note, our attempts to culture viruses derived from the molecular clones in target cells were not successful, suggesting that the clones contain debilitating mutations. A large number of previous reports found only a small minority of the viral clones to be replication competent. For instance, only one out of eight[12] molecular clones was found to be infectious in a previous attempt. It is therefore not surprising that none of the six clones was replication competent in our hands and the clones may need a corrective manipulation to make them replication competent.

In summary, in the present study, we identified two different viral strains from two different individuals that are characterized by the presence of only two NF-κB binding motifs in the viral promoter, which is quite uncommon for subtype C. Phylogenetic characterization of the Indian molecular clones using full-length viral sequencing clustered these viral strains with the Chinese HIV-1C strains that are distinct from that of the southern and eastern African subtype C strains. This result is consistent with our previous hypothesis, based on individual HIV-1 *tat*, *gag*, *pol*, and *env* genes, that the southeast Asian HIV-1C isolates are genetically distinct from those circulating in Africa.[2,5] A detailed molecular analysis of viral genes and their predicted amino acid sequences confirmed several molecular signature motifs of subtype C such as the truncated Rev, signature amino acid residues of Tat, and the preferred coreceptor tropism of env. As molecular surveillance is the backbone for HIV-1 vaccinology, characterization of more full-length sequences of HIV-1 viruses is likely to contribute toward a better understanding of the epidemiology in this region and will guide efficient disease intervention strategies.

Acknowledgments

The work was supported by NIH Grants R01 MH083579 and R37 AI030861 (to V.R.P.). L.E.H. and U.N. wish to acknowledge training support from D43 TW-001403.

Author Disclosure Statement

No competing financial interests exist.

References

1. Hemelaar J, Gouws E, Ghys PD, Osmanov S, and C W-UNHI: Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 2011;25(5):679–689.
2. Neogi U, Bontell I, Shet A, *et al.:* Molecular epidemiology of HIV-1 subtypes in India: Origin and evolutionary history of the predominant subtype C. PLoS One 2012;7(6): e39819.
3. Tee KK, Pybus OG, Li XJ, *et al.:* Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. J Virol 2008;82(18):9206–9215.
4. Shen C, Craigo J, Ding M, Chen Y, and Gupta P:. Origin and dynamics of HIV-1 subtype C infection in India. PLoS One 2011;6(10):e25956.
5. Rao VR, Neogi U, Talboom JS, *et al.:* Clade C HIV-1 isolates circulating in Southern Africa exhibit a greater frequency of dicysteine motif-containing Tat variants than those in Southeast Asia and cause increased neurovirulence. Retrovirology 2013;10:61.
6. Grez M, Dietrich U, Balfe P, *et al.:* Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. J Virol 1994;68(4):2161–2168.
7. Ranga U, Shankarappa R, Siddappa NB, *et al.:* Tat protein of human immunodeficiency virus type 1 subtype C strains is a defective chemokine. J Virol 2004;78(5):2586–2590.
8. Rao VR, Sas AR, Eugenin EA, *et al.:* HIV-1 clade-specific differences in the induction of neuropathogenesis. J Neurosci 2008;28(40):10010–10016.
9. Bachu M, Yalla S, Asokan M, *et al.:* Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J Biol Chem 2012;287(53):44714–44735.
10. Novitsky V, Wang R, Lagakos S, and Essex M: HIV-1 subtype C phylodynamics in the global epidemic. Viruses 2010;2(1):33–54.
11. Papathanasopoulos MA, Cilliers T, Morris L, *et al.:* Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa. AIDS Res Hum Retroviruses 2002;18(12):879–886.
12. Dash PK, Siddappa NB, Mangaiarkarasi A, *et al.:* Exceptional molecular and coreceptor-requirement properties of molecular clones isolated from a human immunodeficiency virus type-1 subtype C infection. Retrovirology 2008;5:25.
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S: MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28(10): 2731–2739.
14. Larkin MA, Blackshields G, Brown NP, *et al.:* Clustal W and Clustal X version 2.0. Bioinformatics 2007;23(21): 2947–2948.
15. Lole KS, Bollinger RC, Paranjape RS, *et al.:* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 1999;73(1):152–160.

16. Patil A and Bhattacharya J: Natural deletion of L35Y36 in p6 gag eliminate LYPXnL/ALIX auxiliary virus release pathway in HIV-1 subtype C. Virus Res 2012;170(1–2): 154–158.
17. Neogi U, Gupta S, Sahoo PN, *et al.:* Genetic characterization of HIV type 1 Tat exon 1 from a southern Indian clinical cohort: Identification of unique epidemiological signature residues. AIDS Res Hum Retroviruses 2012; 28(9):1152–1156.
18. Scriba TJ, de Villiers T, Treurnicht FK, *et al.:* Characterization of the South African HIV type 1 subtype C complete 5' long terminal repeat, nef, and regulatory genes. AIDS Res Hum Retroviruses 2002;18(2):149–159.

Address correspondence to:
Vinayaka R. Prasad
Department of Microbiology and Immunology
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, New York 10461

E-mail: vinayaka.prasad@einstein.yu.edu

This article has been cited by:

1. Behzad Dehghani, Zahra Hasanshahi, Tayebeh Hashempour, Parvin Afsar Kazerooni. 2021. Subtype Classification by Polymerase and Gag Genes of HIV-1 Iranian Sequences Registered in the NCBI GenBank. *Current Proteomics* **18**:2, 153-161. [Crossref]
2. Shuba Varshini Alampalli, Michael M. Thomson, Raghavan Sampathkumar, Karthi Sivaraman, Anto Jesuraj U. K. J., Chirag Dhar, George D. Souza, Neil Berry, Annapurna Vyakarnam. 2017. Deep sequencing of near full-length HIV-1 genomes from plasma identifies circulating subtype C and infrequent occurrence of AC recombinant form in Southern India. *PLOS ONE* **12**:12, e0188603. [Crossref]