Molecular Characterization of Unique Intersubtype HIV Type 1 A1/C Recombinant Strain Circulating in Pune, India

Sudhanshu Pandey,¹ Srikanth Tripathy,² and Ramesh Paranjape³

Abstract

An increasing number of circulating recombinant forms (CRFs) and unique recombinant forms (URFs) all over the world has necessitated being vigilant about new recombinants. Since the first report of a recombinant virus with an A1/C mosaic in 1998 more and more B/C and A/C recombinant viruses are being reported from India. Here we report the identification and characterization of a unique HIV-1 A1/C recombinant circulating in Western India. Analysis of the full-length genome using RIP, SimPlot, and jpHMM@Gobics has confirmed its mosaic structure with insertion of subtype A1 in the backbone of subtype C at three positions: *gag-pol* (1973 ± 15 – 2617 ± 47), *pol-vif* (4879 ± 37 – 5582 ± 32), and *gp41* (8437 ± 106 – 8811 ± 8); however, RIP and SimPlot showed one more small insertion in *integrase* (4343–4519). Phylogenetic analysis confirmed that the recombinant virus has an insertion of clade A1 in the backbone of subtype C, which has come from Indian subtype C.

Introduction

MORE THAN 95% OF THE HIV strains circulating in India belong to subtype C, although subtypes A and B have been reported as early as 1996¹ and 1994,² respectively. Although reports of HIV-1 A and B infections appeared only sporadically thereafter, the mosaic A/C recombinant virus was reported as early as 1998. A recent analysis of the global prevalence of recombinant viruses has revealed that between the years 2004 and 2007 the combined proportion of all circulating recombinant forms (CRFs) increased by 4.5% globally, which corresponds to a 50% increase in the number of CRF infections in comparison to the years 2000–2003. In contrast, the proportion of infections caused by unique recombinant forms (URFs) diminished by 3.1% resulting in a 39% decrease in the burden of URFs. Together, all recombinants (CRFs and URFs) increased by 17%, resulting in a 1.4% increase in the proportion of recombinant infections for a total of 20.5%.³ Full-length genome sequencing of HIV-1, coupled with new techniques for the identification and mapping of recombinant genomes, has resulted in greater information about prevalent strains, their evolution, and their global spread. India does not have a surveillance mechanism in place to monitor the clade distribution in the country.

A study of the intravenous drug user (IDU) cohort from Northeastern India showed the effectiveness of the Multiregion Hybridization Assay (MHA) for genotype-based sur-

veillance, and along with full genome sequencing the assay can serve as an efficient tool for the characterization of a recombination pattern among the newly emerging HIV-1 recombinants.⁴ Most of the other molecular characterization studies in India are based on sequencing of a part of the genome. Very few reports of a full genome sequence have emanated from India.

Six A1C recombinant viruses have been reported from India, out of which only two are characterized based on a full-length genome sequence^{5,6} and four based on partial genome data from Calcutta,⁷ Bengaluru,⁸ and Pune.⁹ B/C recombinant viruses were reported from Manipur by Lakhashe *et al.*^{10,11} based on full-length genome sequencing and by Bhanja *et al.*¹² based on partial gene sequencing. B/C recombinant viruses were also reported from West Bangal,¹³ Karnataka,¹⁴ and Maharashtra based on partial genome sequence data. CRF02_AG has also been reported from Delhi.¹⁵

Materials and Methods

As a part of the project “full-length genome analysis of HIV-1 isolates from India” a blood sample was collected from a 27-year-old male seropositive individual (NARI Patient Id: 09-387) having a CD4 count of $220/\text{mm}^3$ from Pune, India in July 2009. Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation using Ficoll-Hypaque (1.077 g/ml , H8889, Sigma, St. Louis, MO). Isolated

¹Division of Immunology, National AIDS Research Institute (ICMR), Pune, India.

²Central JALMA Institute for Leprosy and other Mycobacterial Diseases, Agra, India.

³Division of Immunology, National AIDS Research Institute (ICMR), Pune, India.

PBMCs were cocultured for 21 days with an equal number of mitogen-activated ($5 \mu\text{g/ml}$ of PHA-P, Sigma-Aldrich) PBMCs from a healthy HIV infection free individual. During the process both types of PBMCs were suspended in RPMI 1640 (Hi-Media) supplemented with 10% fetal bovine serum

(FBS) (Sigma-Aldrich), 20 U/ml human interleukin-2 (Roche), 100 U/ml penicillin, 100 $\mu\text{g/ml}$ streptomycin (Hi-Media), and 2 mM L-glutamine (Sigma-Aldrich) and were incubated in 5% CO_2 . Then 3×10^6 fresh mitogen-activated PBMCs from the same healthy donor were added after every 7 days.

FIG. 1. Phylogenetic analysis of newly derived full genome sequences of recombinant isolates from India. The sequence of NARI-FLS_09-387 was aligned with Indian clade C (accession numbers AY049708, EF694036, EF469243, AF067154, AF286232, AB023804, AY713414, AF067157, AF067159) and reference sequences of all pure subtypes (A-K) available at the Los Alamos HIV database. The scale bar represents 2% genetic distance.

FIG. 2. Identification of recombinant isolate 09-387. Similarity plots were drawn using RIP 3.0 software with window size: 300 nucleotides; confidence threshold: 90%; gap handling: strip all gaps and plot all window values; and multistate character: Yes. Similarity with subtypes C and A1 shows insertion of subtype A1 at four positions (denoted as 1, 2, 3 and 4) in the backbone of subtype C. Color images available online at www.liebertpub.com/aid

FIG. 3. Subtype identities of isolate 09-387. The similarity plot was created using SimPlot 3.5.1 software configured with window size: 300 nucleotides; step: 30; GapStrip: On, F84 (maximum likelihood). Breakpoints are shown at the bottom of the plot, which indicates insertion at four places (nts 1168–1720, 3415–3590, 3920–4614, and 7379–7619). Reference sequences of the subsubtypes A1, A2, B, C, D, F1, F2, G, H, J, and K, as well as of subtype A1 and C from India, are listed on the right side. Color images available online at www.liebertpub.com/aid

FIG. 4. Breakpoints in HXB2 coordinates. **(a)** Breakpoints based on jpHMM@gobics and **(b)** breakpoints based on SimPlot [map drawn with the help of HIV-1 drawing tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html)] show a similar pattern of recombination. The genomic structure is divided into seven segments including four subtype C (I, III, V, and VII) and three subtype A1 (II, IV, and VI). However, SimPlot shows one more small insertion in segment III and further divided it into three blocks (A, B, and C). Color images available online at www.liebertpub.com/aid

The culture supernatant was tested for p24 antigen on days 7, 14, and 21 postinfection (PID) by HIV-1 antigen micro ELISA (Biomérieux). Proviral DNA was extracted from PBMCs from p24-positive culture using the QIAamp DNA blood mini kit (Qiagen, Hilden, Germany). Isolate NARI-FLS_09-384 was amplified with two fragment strategy with an overlap of 0.88 kb nucleotides. The 5' terminal 2.88-kb fragment of the 5' LTR through the RT region (0001–2878 in the HXB2 coordinate) was amplified by nested polymerase chain reaction (PCR) using outer primer pairs N558F and N453R in the first round and inner primer pairs N533F and N502R¹⁶ in the second round, and a 7.5-kb fragment p7Gag through the 3' LTR (1996–9632 in the HXB2 coordinate) was amplified using primer pairs Uninef7' and pro5F.¹⁷

Amplified products were cloned into a pCR2.1 vector and eight clones of 2.88 kb (0001–2878 in the HXB2 coordinate)

and one clone of 7 kb (1996–9632 in the HXB2 coordinate) were generated. DNA sequencing of amplified products of both the fragments from proviral DNA and all the clones was done using an ABI PRISM Big Dye terminator v3.1 cycle sequencing kit (Applied Biosystems) with the help of an automated genetic analyzer ABI 3730XL DNA Sequencer (Applied Biosystems Inc., Foster City, CA). A total of 45 primers was used to cover the entire length of the viral genome and a contiguous sequence was made using the primer walking method with the help of Seqscape v2.6 (Applied Biosystems). A full-length genome sequence (NARI-FLS_09-387) of a 9615 nucleotide-long position (1–9603 relative to the HXB2 genome) and spanning 5' LTR to 3' LTR was obtained by sequencing the amplified product directly.

The full genome sequence NARI-FLS_09-387 obtained by sequencing the PCR product was aligned with HIV-1 strains

TABLE 1. INTERSUBTYPE RECOMBINATION BREAKPOINTS IN FULL GENOME DNA SEQUENCES OBTAINED FROM SEQUENCING POLYMERASE CHAIN REACTION PRODUCTS AND CLONAL SEQUENCES BASED ON jpHMM@GOBOICS

Clone ID	Sequence length (HXB2 coordinate)	NA	Clade C	A1	C	A1	C	A1	C	NA
NARI-FLS_09-387	0001–9603	1–789	790–1976	1977–2656	2657–4879	4880–5585	5586–8332	8333–8802	8803–9407	9408–9603
NARI-FLS_09-387-25	1996–9603			1996–2656	2657–4879	4480–5569	5570–8518	8519–8802	8803–9407	9408–9603
NARI-FLS_09-387-1	0001–2900	1–789	790–1976	1977–2656	2657–2900					
NARI-FLS_09-387-4	0001–2900	1–789	790–1976	1977–2631	2632–2900					
NARI-FLS_09-387-13	0001–2900	1–789	790–1976	1977–2656	2657–2900					
NARI-FLS_09-387-18	0001–2900	1–789	790–1976	1977–2656	2657–2900					
NARI-FLS_09-387-21	0001–2900	1–789	790–1976	1977–2656	2657–2900					
NARI-FLS_09-387-22	0001–2900	1–789	790–1976	1977–2631	2632–2900					
NARI-FLS_09-387-23	0001–2900	1–789	790–1976	1977–2656	2657–2900					
NARI-FLS_09-387-24	0001–2900	1–789	790–1976	1977–2656	2657–2900					

of reference subtypes and subtype C from India obtained from the Los Alamos HIV Database (<http://hiv-web.lanl.gov/>) using Clustal W software. A phylogenetic tree was constructed by MEGA 5.05 to confirm the ancestral subtype using the maximum likelihood method under the General Time Reversible Model (GTR model) and Gamma distributed with Invariant sites (G+I) with the number of discrete gamma category 5.

Results

The intracluster average genetic distance of Indian clade C was 6.34%, the average genetic distance of NARI-FLS_09-387 with the Indian clade C group was 9.42%, while the sub-subtype distances between B-D and F1-F2 were 10.7% and 11.2%, respectively. The sequence NARI-FLS_09-387 clustered with the clade C cluster but away from Indian subtype C

and the distance from Indian clade C suggesting the possibility of a recombinant genome (Fig. 1).

The full genome sequence was analyzed with RIP (<http://hiv.lanl.gov/content/sequence/RIP/RIP.html>) to determine the subtype and it was found that this strain is recombinant of subtypes C and A1. A plot of distance (similarity) using window size 300 and confidence threshold 90% shows insertion of subtype A1 in the backbone of clade C at four positions for NARI-FLS_09-387 (Fig. 2A). The recombination event was further confirmed along with breakpoints with the help of SimPlot (version 3.5.1; <http://sray.med.som.jhmi.edu/RaySoft/SimPlot/>) and jpHMM@Gobics.¹⁸ Discrepancies were observed in results obtained by bootscan analysis using SimPlot and the jumping profile hidden Markov model for NARI-FLS_09-387. SimPlot analysis using distance model F84 (maximum likelihood), window size 300 bp, step 30 bp, showed three major insertions at *gag-pol* (2039–2635),

FIG. 5. Phylogenetic analysis based on breakpoints of segments showing subtype C. Each segment was analyzed separately by the phylogenetic maximum likelihood method under the General Time Reversible Model (GTR model) and the discrete Gamma was distributed with Invariant sites (G+I) with the number of discrete gamma category 5. The genetic distance corresponding to the lengths of the branches is shown by the bottom line. Segments I, V, and VII clustered within clade C radiation but segment III clustered away from subtype C radiation.

pol-vif (4849–5549), and *gp41* (8481–8739) of subtype A1 and a small insertion in the *pol* region (4343–4519) (Fig. 3). A genome map was drawn with the recombinant breakpoint data using the recombinant HIV-1 drawing tool (http://hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html) (Fig. 4b). However, the jumping profile hidden Markov model (jpHMM@Gobics) has revealed insertion of subtype A1 only at three positions: *gag-pol* (1973 ± 15 – 2617 ± 47), *pol-vif* (4879 ± 37 – 5582 ± 32), and *gp41* (8437 ± 106 – 8811 ± 8) (Fig. 4a). All the clones also showed a similar pattern of recombination, which suggests that infection was caused by a single isolate (Table 1).

Phylogenetic analysis was performed next using a full-length genome sequence obtained from PCR products based on the breakpoints received from jpHMM@Gobics and SimPlot to confirm potential ancestor subtypes from which the recombinant strain was derived. Segment II spanning p7⁸RT (1980–2600), IV spanning the end of IN-Vif (4850–5549), and VI spanning *gp41* (8400–8800) genes clustered

within clade A1 radiation with A1 isolates sequenced in the study with significant bootstrap value (≥ 90) (Fig. 6). The segment I Gag (791–1970), V Vif-Gp41 (5560–8430), and VII Gp41-Nef (8481–9400) clustered within clade C radiation along with Indian isolates but segment III Pol (2605–4845) clustered away from subtype C (Fig. 5) radiation suggesting the possibility of insertion of clade A1 and corroborating the insertion of clade A1 in the *integrase* gene (4343–4519 HXB2 coordinates) as revealed in SimPlot analysis.

Phylogenetic analysis of segment III was further done with the breakpoints based on SimPlot results. Based on results obtained by SimPlot analysis segment III is composed of three blocks (A, B, and C) showing a small insertion of A1 (block B 4343–4519) within the clade C backbone. Block A (2604–4342) and block C (4520–4848) clustered within subtype C radiation with significant bootstrap values of 100 and 83, respectively (Fig. 7). Block B showed A1 in bootscan analysis, with a fall in A1 radiation with a bootstrap value of 63 with the isolate from Kenya but with a nonsignificant bootstrap value of 38 at the

FIG. 6. Phylogenetic analysis of segments (II, IV, and VI) showing subtype A1. Isolate NARI-FLS_09-387 clustered within subtype A1 radiation with Indian clade A1 sequences.

node of clade A1. BLAST search results indicated that block B had the highest similarity (97%) with subtype A1 sequences from Kenya.

Similarity with Kenyan isolates may be the result of the availability of very few HIV-1 subtype A sequences from India. Sliding window size has an important role in recombination identification. Zhang *et al.*¹⁹ suggested that a window size of 300 was best suited for the identification of HIV-1 recombination. Discrepancies in the results by jpHMM@GOBICS and SimPlot may be due to the use of different window sizes. All the clones and DNA sequences clustered very closely (data not shown) and exhibited similar breakpoints as well as infection established from a single recombinant isolate.

The mosaic isolate NARI-FLS_09-387 shows recombination at four positions whereas Lole *et al.*⁵ reported the A1/C mosaic virus with insertion of the clade A1 sequence only in the *env-nef* region (6320–9100 in the HXB2 coordinate) with the backbone of clade C. The other mosaic virus reported by Rodriguez *et al.*⁶ showed recombination of C in the backbone of A1 at one position each in the *gag* (1134–1814) and *pol* (2623–4394) genes. Comparison with other mosaic viruses

suggests *gag* (1800–2000 position) and *pol* (2600–2650) are the positions that favor the possibility of recombination.

Discussion

HIV-1 subtype A was reported in India as early as 1998. There is evidence available that subtype C has a fitness advantage over subtype A1 and it may be a possible cause for the relatively low prevalence of HIV-1 subtype A.²⁰ However, subtype A and C recombinants, described since 1998, have been reported from different parts of the country. The replication fitness and transmission potential of such strains need to be studied.

Since most of HIV-1 subtype A strains reported from India have been typed based on Heteroduplex Mobility Assay (HMA) or sequencing of a short genome fragment it is possible that they may not be pure HIV-1 subtype A but could be recombinants.

Recombination produces mosaic virus from parental strains and thus provides a shortcut to speed up viral evolution. All the recombinations may not necessarily provide a

FIG. 7. Phylogenetic analysis of blocks (A, B, and C) of segment III. Blocks A and C clustered within subtype C radiation with bootstrap values of 100 and 83, respectively. Block B clustered within A1 radiation with a bootstrap value of 38.

fitness benefit. However, rapidly increasing the proportion of CRFs globally indicates that the recombinant viruses may enjoy advantages over the parent strains. The relationship between recombination, mutation, and fitness is highly complex. Syncytium induction and multiple coreceptor usage are strongly associated with a greater decrease in CD4⁺ T cell count over time and a greater risk of disease progression and drug resistance. CTL escape mutants are major challenges to antiretroviral therapeutics and vaccine development. The phenotype of the newly defined CRF14_BG has been described as syncytium inducing, with usage of CXCR4, CCR5, CCR3, and CCR2b coreceptors.²¹

In this study we characterized the full-length genome of the mosaic HIV-1 isolate NARI-FLS_09-387. The recombinant isolate shows insertion of subtype A1 at four positions in the backbone of subtype C. Clustering of the subtype C portion with Indian HIV-1 isolates suggests its Indian origin. There are very few partial and no full-length sequences of subtype A reported from India. This made the ancestral analysis of the subtype A1 portion difficult. This emphasizes the need to generate full-length genome sequence data and elaborate molecular characterization of all clade A1 isolates from India. In conclusion, this isolate is recombinant between subtype A1 and Indian clade C. This novel mosaic structure of the isolate is different from all other reported CRFs and URFs.

Sequence Data

The nucleotide sequences NARI-FLS_09-387 and its nine clones have been submitted to GenBank and the accession numbers are KC911635, KC911636, KC911637, KC911638, KC911639, KC911640, KC911641, KC911642, KC911643, KC911644.

Acknowledgments

We thank NARI clinics for providing us with clinical samples. This work was supported in part by a grant from the Department of Biotechnology (DBT), Government of India (BT/PR7054/Med/14/936/2006) to Ramesh Paranjape. We thank the Council of Scientific and Industrial Research (CSIR), Government of India for providing a senior research fellowship to Sudhanshu Pandey.

Author Disclosure Statement

No competing financial interests exist.

References

1. Tripathy S, Renjifo B, Wang WK, *et al.*: Envelope glycoprotein 120 sequences of primary HIV type 1 isolates from Pune and New Delhi, India. *AIDS Res Hum Retroviruses* 1996;12(12):1199–1202.
2. Baskar PV, Ray SC, Rao R, Quinn TC, Hildreth JE, and Bollinger RC: Presence in India of HIV type 1 similar to North American strains. *AIDS Res Hum Retroviruses* 1994;10(8):1039–1041.
3. Hemelaar J, Gouws E, Ghys PD, *et al.*: Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 2011;25(5):679–689.
4. Sarkar R, Sengupta S, Mullick R, Singh NB, Sarkar K, and Chakrabarti S: Implementation of a multiregion hybridiza-

- tion assay to characterize HIV-1 strains detected among injecting drug users in Manipur India. *Intervirology* 2009;52(4):175–178.
5. 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.
6. Rodriguez MA, Chen Y, Craigo JK, *et al.*: Corrigendum to "Construction and characterization of an infectious molecular clone of HIV-1 subtype A of Indian origin." *Virology* 2010;397(1):242.
7. Sengupta S, Khetawat D, Jana S, Sarkar K, Bhattacharya SK, and Chakrabarti S: Polymorphism of HIV-1 gag (p17) gene from female sex workers in Calcutta, India. *Arch Virol* 2005;150(10):2117–2124.
8. 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.
9. Lall M, Gupta RM, Sen S, Kapila K, Tripathy SP, and Paranjape RS: Profile of primary resistance in HIV-1-infected treatment-naïve individuals from Western India. *AIDS Res Hum Retroviruses* 2008;24(7):987–990.
10. Lakhashe S, Tripathy S, Paranjape R, and Bhattacharya J: Characterization of B/C recombinants of near full-length HIV type 1 from northeastern India with mosaics identical to ARE195FL but with a different ancestral origin. *AIDS Res Hum Retroviruses* 2008;24(1):92–99.
11. Lakhashe S, Tripathy S, Paranjape R, and Bhattacharya J: Evidence of a novel B/C recombinant exhibiting unique breakpoints of the nearly full-length HIV type 1 genome in northeastern India. *AIDS Res Hum Retroviruses* 2008;24(2):229–234.
12. Bhanja P, Sengupta S, Singh NY, Sarkar K, Bhattacharya SK, and Chakrabarti S: Determination of gag and env subtypes of HIV-1 detected among injecting drug users (IDUs) in Manipur, India: Evidence for intersubtype recombination. *Virus Res* 2005;114(1–2):149–153.
13. Mullick R, Sengupta S, Sarkar K, Saha MK, and Chakrabarti S: Phylogenetic analysis of env, gag, and tat genes of HIV type 1 detected among the injecting drug users in West Bengal, India. *AIDS Res Hum Retroviruses* 2006;22(12):1293–1299.
14. Siddappa NB, Dash PK, Mahadevan A, *et al.*: Identification of unique B/C recombinant strains of HIV-1 in the southern state of Karnataka, India. *AIDS* 2005;19(13):1426–1429.
15. Neogi U, Sood V, Chowdhury A, *et al.*: 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* 2009;6:28–35.
16. Dash PK, Siddappa NB, Mangaiarkarasi A, *et al.*: Exceptional molecular and coreceptor-requirement properties of molecular clones isolated from an human immunodeficiency virus type-1 subtype C infection. *Retrovirology* 2008;5:25–46.
17. Nadai Y, Eyzaguirre LM, Constantine NT, *et al.*: Protocol for nearly full-length sequencing of HIV-1 RNA from plasma. *PLoS One* 2008;3(1):e1420.
18. Zhang M, Schultz AK, Calef C, *et al.*: jpHMM at GOBICS: A web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res* 2006;34:W463–465.
19. Zhang C, Ding N, and Wei J-F: Different sliding window sizes and inappropriate subtype references result in discor-

- dant mosaic maps and breakpoint locations of HIV-1 CRFs. *Infect Genet Evol* 2008;8:693–697.
20. Rodriguez MA, Chen Y, Craig JK, *et al.*: Construction and characterization of an infectious molecular clone of HIV-1 subtype A of Indian origin. *Virology* 2006;345(2):328–336.
 21. Pérez-Alvarez L, Delgado E, Villahermosa ML, *et al.*: Biological characteristics of newly described HIV-1 BG recombinants in Spanish individuals. *AIDS* 2002;16(4):669–672.

Address correspondence to:

Ramesh Paranjape
73 G Block, MIDC
Bhosari, Pune 411026
Maharashtra
India

E-mail: rparanjape@nariindia.org