

Description of the new HIV-1 intersubtype B/C circulating recombinant form (CRF146_BC) detected in Brazil

Rodrigo Cunha Oliveira¹, Darren Martin², Juliana Sacramento Mota de Souza¹, Luiz Carlos Júnior Alcântara³, Monick Lindenmeyer Guimarães⁴, Carlos Brites⁵, Joana Paixão Monteiro-Cunha^{1/+}

¹Universidade Federal da Bahia, Departamento de Bioquímica e Biofísica, Núcleo de Bioinformática, Salvador, BA, Brasil

²University of Cape Town, Institute of Infectious Disease and Molecular Medicine, Department of Integrative Biomedical Sciences, Computational Biology Group, Cape Town, South Africa

³Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de Flavivírus, Rio de Janeiro, RJ, Brasil

⁴Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de AIDS e Imunologia Molecular, Rio de Janeiro, RJ, Brasil

⁵Universidade Federal da Bahia, Complexo Hospitalar Prof Edgard Santos, Laboratório de Pesquisa em Infectologia, Salvador, BA, Brasil

BACKGROUND The human immunodeficiency virus 1 (HIV-1) infections in Brazil are predominantly caused by two subtypes, B and C.

OBJECTIVES Here we present the characterisation of a novel HIV-1 recombinant form, indicating a new Brazilian CRF_BC, named CRF146_BC.

METHODS RDP, JphMM and Simplot recombination tools were used to evaluate the mosaic pattern.

FINDINGS In this work, we identified three HIV-1 nucleotide sequences previously classified as unique recombinant forms (URFs), plus one new partial genome sharing the same BC recombination pattern. The mosaic genome is almost entirely represented by the subtype C sequence, with a small subtype B recombination region in the pol gene, at the Integrase level. The phylogenetic analyses strongly indicate a common origin between the strains, which were isolated in Rio Grande do Sul, Rio de Janeiro and Bahia states.

MAIN CONCLUSIONS Thus, the new HIV-1 CRF146_BC is circulating in three different Brazilian regions: South, Southeast and Northeast.

Key words: CRF146_BC - HIV-1 - recombination - Brazil

The human immunodeficiency virus (HIV) discovered in the early 1980s is classified into two types: HIV-1 and HIV-2.^(1,2) HIV-1 is highly diversified, being represented by four groups (M, N, O and P). The M group is further classified into 10 subtypes (A, B, C, D, F, G, H, J, K and L), sub-subtypes (A1-A8 and F1, F2) and circulating or unique recombinant forms (URF).^(2,3) *In vivo* co-infection or super-infection with different strains can result in the emergence of hybrid viruses.^(2,4,5) These viruses are generated through the genic recombination mechanism, which occurs mainly during retrotranscription, from pure or recombinant parental strains, resulting in recombinant forms.⁽⁵⁾ When recombinant viruses sharing an identical mosaic structure are isolated from three or more epidemiologically unrelated individuals, they are designated circulating recombinant forms (CRF).⁽²⁾ To define a new CRF, three near-full length genomes (NFLG) or two NFLGs in conjunction with a partial sequence of a third strain are sufficient. Recombinant forms that do not meet these criteria are designated as URF.^(2,5)

The subtype C accounts for nearly one-half of HIV-1 infections worldwide (46.6%), whereas the B subtype is globally distributed.⁽⁴⁾ Recombinant strains are widely spread and estimated to be responsible for nearly 23% of infections worldwide.⁽⁴⁾ Numerous URFs and 157 CRFs have been described over the years (<https://www.hiv.lanl.gov/components/sequence/HIV/crfdb/crfs.comp>). Among these, 14 are CRF_BC, 11 of which were identified in China,⁽⁶⁾ but some were also identified in Brazil and Italy.^(7,8) The first (CRF07_BC) was described in 2000⁽⁵⁾ and the last (CRF118_BC) in 2021.⁽⁹⁾

In Brazil, the subtype B is the prevalent HIV-1 genotype, followed by the C and F1 subtypes, together with distinct recombinant forms (BF1 and BC mainly).⁽¹⁰⁾ In this country, B and C subtypes cause 81% of the HIV-1 infections and a BC CRF (CRF31_BC), in addition to several BC URF have been described.^(7,11) Here we present the characterisation of a novel HIV-1 recombinant form, indicating a new Brazilian CRF_BC, named CRF146_BC.

SUBJECTS AND METHODS

Sample 223 presented here (accession number OR260538) was collected in 2019 from an HIV-1 infected patient followed at Professor Edgard Santos University Hospital in Salvador (Bahia), a capital city in the Northeast region of Brazil. Both Hospital and the Federal University of Bahia Ethics Committee approved the study (Ethics Committee Number: 2.831.993). The

doi: 10.1590/0074-02760230214

Financial support: RCO and JSM are PhDs scholarship (FAPESB); MLG is recipient of fellowship from CNPq (number 07854/2021-3).

+ Corresponding author: joana.monteiro@ufba.br

ORCID: <https://orcid.org/0000-0001-9074-457X>

Received 20 November 2023

Accepted 23 July 2024



participant supplied written informed consent for specimen collection and subsequent analyses. The nested-polymerase chain reaction (PCR) and DNA sequencing protocol were applied as previously described,⁽¹²⁾ using the specific HIV-1 primers, covering four overlapping regions. However, only genomic fragments corresponding to *integrase/vif* and *nef* genes (positions 4172 to 5198 and 8694 to 9512 relative to HXB2 reference genome, respectively) were successfully obtained, while other genomic regions did not amplify. RDP,⁽¹³⁾ JphMM⁽¹⁴⁾ and Simplot⁽¹⁵⁾ recombination tools were used to evaluate the mosaic pattern. The Simplot analysis were conducted based in the follow parameters: 350 nucleotides sliding window and steps of 30 bases, in accordance with Passaes et al. and Pessôa et al.^(16,17) A breakpoint between C and B subtypes was found in the *integrase/vif* region (positions 4837+77 - 5071+21 HXB2 coordinates), while the *nef* gene region reveals a pure C fragment.

A search for complete genome sequences of BC recombinants has been performed in the Los Alamos National Laboratory (LANL) database (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>), using the following criteria: complete genome; subtypes B + C; include recombinants. Then, a manual search for sequences identified as pure BC with more than 7000 nt in length was conducted using the BioEdit program V.7.2.5.⁽¹⁸⁾ As a result, 225 NFLG BC sequences were obtained. In order to investigate and compare the recombination patterns, sequences were analysed using the aforementioned recombination tools. Out of the 225 BC recombinants found in the database, 16 NFLG sequences were isolated in Brazil. Of these, three samples, previously described as URF_BC, were identified with similar breakpoints in the *integrase/vif* genes and classified as pure subtype C in the *nef* gene: GQ365652 isolated in Rio Grande do Sul, KT427778 and KT427790, both isolated in Rio de Janeiro. In order to investigate the phylogenetic relationships, these four samples (OR260538, GQ365652, KT427778 and KT427790) were aligned with

134 recombinant sequences representing 14 CRF_BC, plus 42 pure subtype reference sequences collected from LANL database. The alignment was performed using MAFFT online program⁽¹⁹⁾ (<http://mafft.cbrc.jp>) under the command: `mafft --thread 8 --threadtb 5 --threadit 0 --reorder -- auto input > output` and manually edited using BioEdit software.⁽¹⁸⁾ The maximum likelihood (ML) trees were reconstructed using IQ-TREE 1.6.8 webserver⁽²⁰⁾ (Command: `path_to_iqtree -s aln.fasta -m TEST -bb 1000 -alrt 1000`) and RDP5⁽¹³⁾. The reliability of each cluster was evaluated by analysing 1,000 bootstrap (BS) replicates and with the approximate likelihood ratio test (aLRT) based on the Shimodaira-Hasegawa-like procedure.⁽²¹⁾ The ML trees were visualised with the FigTree program version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). In order to confirm the recombination, the alignment was fragmented at the breakpoints and the segments (subtype C concatenated fragment - *integrase/vif* and *nef* - and subtype B *integrase* fragment) were submitted to ML analysis as described above.

RESULTS

Comparing sample OR260538 BC recombinant fragment with the other three NFLG sequences from Brazil [Fig. 1, Supplementary data (Fig. 1)], we identified similar breakpoints in: GQ365652 (4834+75 - 5085+21), KT427778 (4831+80 - 5080+19) and KT427790 (4851+75 - 5085+21). However, sample GQ365652 presented two other small subtype B fragments in the reverse transcriptase region (2986-3089 and 4172-4315). The detection of these subtype B insertions could only be detected by informative site analysis, since bootscanning analysis of these samples, only showed a discrete decrease in the similarity with subtype C (data not shown).⁽¹⁶⁾

Next, these BC recombinant viruses were submitted to phylogenetic analyses with 134 HIV-1 nucleotide sequences of other BC CRFs and 42 pure subtypes, retrieved from LANL database.⁽²⁰⁾ ML trees (Fig. 2) were reconstructed based on the fragment size of sample

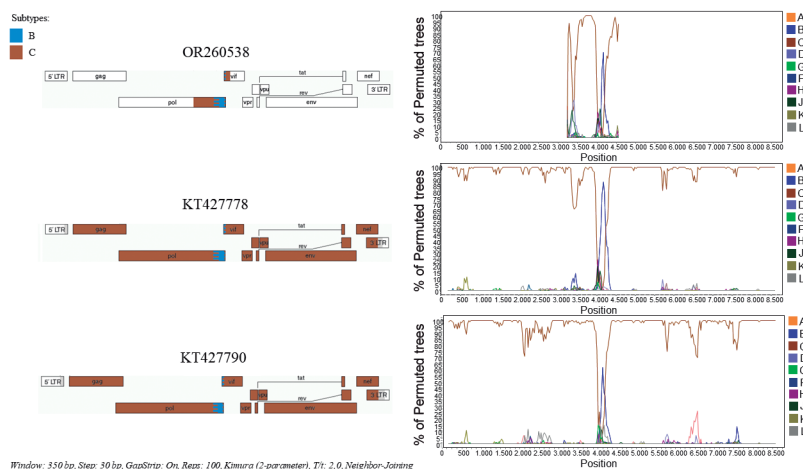


Fig. 1: recombination pattern of human immunodeficiency virus 1 (HIV-1) CRF146_BC strains. CRF146_BC breakpoints were detected using the jumping profile hidden Markov model (JpHMM) and simplot tools. Simplot analysis was performed using a sliding window of 350 nucleotides and steps of 30 bases

OR260538 (Fig. 2A) and NFLG sequences (Fig. 2B). Moreover, ML analyses of the individual fragments (subtype C concatenated fragment - *integrase/vif* and *nef* - and subtype B *integrase* fragment) using the appropriate evolutionary model confirmed the Bootscanning recombination results [Supplementary data (Fig. 2)]. Sequences OR260538, GQ365652, KT427778 and KT427790 grouped together with aLRT (approximated likelihood ratio test) > 90 and BS values > 94%. The ML trees generated in RDP5 software, showed similar topologies, reinforcing the close genetic relationship between these viruses and their branching apart of HIV-1 pure subtypes and other BC CRFs [Supplementary data

(Fig. 3)]. Bootscanning analysis (Fig. 3) showed that BC recombinant sequences inside this monophyletic group were more similar to each other (OR260538, KT427778, KT427790 and GQ365652) than to reference sequences of pure HIV-1 subtypes. In addition, two other trees were generated using a panel of pure subtype B and C sequences from Brazil (Fig. 4). The four BC recombinant strains also grouped together as monophyletic clusters in both ML trees: inside the larger subtype C clade in the *integrase/vif* concatenated tree (Fig. 4A) and inside the subtype B clade in the smaller *integrase* fragment (Fig. 4B). These observations reinforce the straight phylogenetic relationships among these viral isolates.

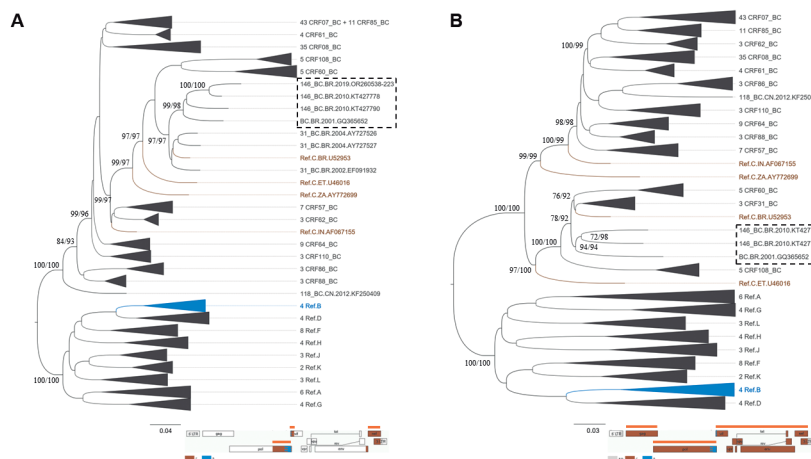


Fig. 2: maximum likelihood (ML) analyses showing the phylogenetic relationships of the Brazilian CRF146_BC recombinant viruses with other human immunodeficiency virus 1 (HIV-1) BC circulating recombination forms (CRFs) (n = 134) and pure subtype reference (n = 42) sequences. Sequences sharing CRF146_BC recombination pattern formed a monophyletic group shown within the box. Trees were rooted by the midpoint. The subtype/CRF classification and the number of sequences within each collapsed cluster are indicated. The statistical support is indicated only at key nodes as approximated likelihood ratio test (aLRT) and bootstrap (BS) values. Trees were built under the GTR+I+G evolutionary model and visualised in the Figtree v1.4.4 software. Horizontal branch lengths were drawn to scale with the bar at the bottom indicating the nucleotide substitutions per site. Trees were built based on (A) BC *integrase/vif* and C *nef* concatenated genomic fragments as in sample OR260538 (4172 - 5198 and 8694 - 9412 relative to HIV-1 reference strain (HXB2) and (B) near-full length genomes (NFLG) sequences without OR260538 (nucleotides 805 - 8912 relative to HXB2 reference strain). The genomic region used to build each tree is shown in the HIV-1 genomic map below the topology (orange line).

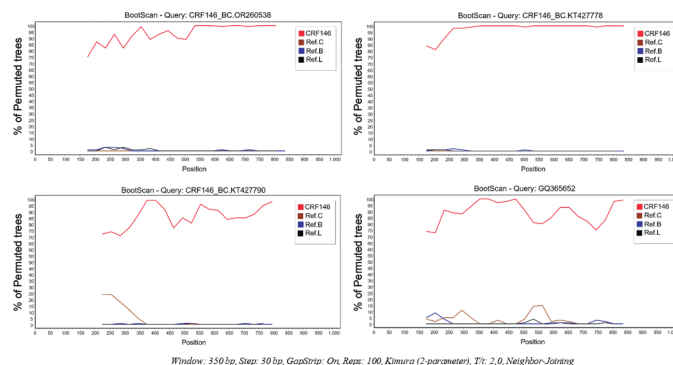


Fig. 3: Bootscan analysis (SimPlot) showing the relationship between Brazilian CRF146 recombinant sequences. In this analysis, each CRF146 query (OR260538, KT427778, KT427790 and GQ365652) is more similar to the CRF146 group (containing the other three recombinant isolates) than to pure subtypes B and C reference sequences. Analyses were performed using the integrase genomic fragment, based on sample OR260538 sequence length.

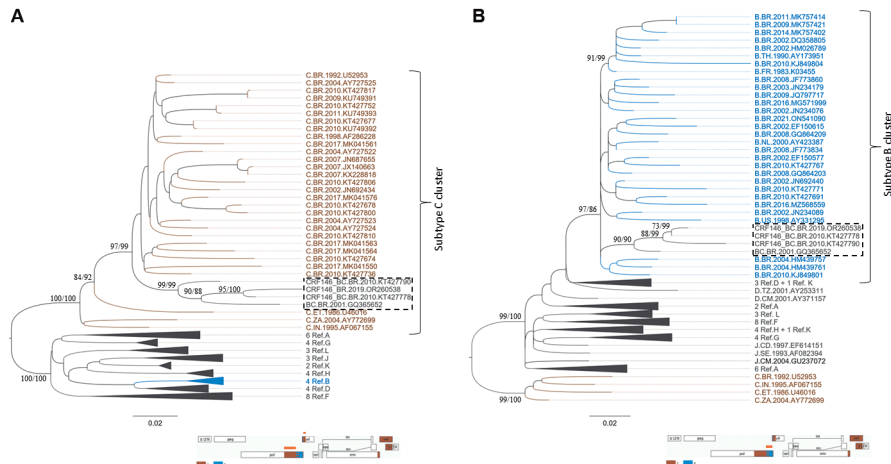


Fig. 4: CRF146_BC recombinants viruses among Brazilian human immunodeficiency virus 1 (HIV-1) subtypes B and C samples. Sequences were fragmented at the breakpoints and the segments were submitted to maximum likelihood (ML) analysis with Los Alamos reference set. (A) *C integrase/vif* concatenated fragment (positions 4172 to 4834 and 5071 to 5190 relative to HIV-1 reference strain (HXB2), and (B) *B integrase* fragment (nucleotides 4833 - 5066 relative to HXB2 reference strain). The statistical support is indicated only at key nodes as approximated likelihood ratio test (aLRT) and bootstrap values. Trees were built under the TIM2+I+G evolutionary model and visualised in the Figtree v1.4.4 software. Horizontal branch lengths were drawn to scale with the bar at the bottom indicating the nucleotide substitutions per site.

DISCUSSION

In Brazil, both subtype C and CRF31_BC were first identified in the Southern region and the subtype C have been spreading northward the country (Southeast, Central West, North and Northeast regions).^(7,22) This region accounts for 53,8% of HIV-1 subtype C infections in Brazil.⁽¹⁰⁾ Out of the four samples representing the new CRF, one (GQ365652) were sampled in Rio Grande do Sul (Southern region), two (KT427778 and KT427790) were from Rio de Janeiro State (Southeastern region) and one (OR260538) from Bahia State (Northeastern region). In addition, Passaes and collaborators,⁽¹⁶⁾ found the same breakpoint (4834+75 - 5085+21) in URF_BC sequences from Rio Grande do Sul State (Southern region) and suggested that second-generation recombination events, involving BC recombinant strains as parental, may be favouring the increase of the new genotypes, once both pure subtypes are prevalent in the country. In this regard, sample GQ365652 presented two additional subtype B fragments (< 200 bp) in *pol* gene,⁽¹⁶⁾ which were not observed in the other two NFLG sequences (KT427778 and KT427790). Since GQ365652 presented very close phylogenetic relationships with all other isolates in CRF146 group, grouping together in all six generated trees, we can speculate that these short (< 150nt) subtype B fragments could have been acquired throughout a second-generation recombination event.

In brief, here we described a novel CRF_BC in Brazil (CRF146_BC), which is circulating in, at least, three out of five geographic regions of the country. The ongoing emergence of new HIV-1 M genotypes emphasises the need to sustain genomic surveillance in Brazil. More specific studies will be necessary to determine whether, relative to its parents this new CRF is either evolutionarily fitter or has altered pathogenicity.

ACKNOWLEDGEMENTS

To Darren Martin, Carlos Brites, Luiz Carlos Júnior Alcântara and Monick Lindermeier Guimarães for advanced support to the research. The authors are grateful for genomic platform DNA sequencing RPT01A (Rede de Plataformas Tecnológicas Fiocruz).

AUTHORS' CONTRIBUTION

RCO performed the conceptualisation, data curation, formal analysis, investigation, methodology, writing - original draft; DM methodology, writing - review & editing; JSMS methodology; LCJA resources; MLG resources, writing - review & editing; CB resources; JPM-C supervision, conceptualisation, visualisation, writing - review & editing. The authors declare there is no conflict of interest among co-authors.

REFERENCES

1. Ioachim HL. Immunopathogenesis of human immunodeficiency virus infection. *Cancer Res.* 1990; 50: 5612-7.
2. Robertson DL, Anderson JP, Bradac JA, Karr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science.* 2011; 288(5463): 55-6.
3. Yamaguchi J, Vallari A, McArthur C, Sthresley L, Cloherty GA, Berg MG, et al. Brief report: complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *J Acquir Immune Defic Syndr.* 2020; 83(3): 319-22.
4. Hemelaar J, Loganathan S, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis* 2019; 19(2): 143-55.
5. Santoro MM, Perno CF. HIV-1 genetic variability and clinical implications. *ISRN Microbiol.* 2013; 17: 481314.
6. Su L, Graf M, Zhang Y, von-Briesen H, Kostler J, Melzl H, et al. Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. *J Virol.* 2000; 74(23): 11367-76.

7. Santos AF, Sousa TM, Soares EA, Sanabani S, Martinez AM, Silveira J, et al. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS*. 2006; 20(16): 2011-9.
8. Simonetti FR, Lai A, Monno L, Binda F, Brindicci G, Punzi G, et al. Identification of a new HIV-1 BC circulating recombinant form (CRF60_BC) in Italian young men having sex with men. *Infect Genet Evol*. 2014; 23: 176-181.
9. Gao Q, Feng Y, Gao L, Zhang M, Liu Y, Xiao M, et al. Genetic characteristics of a new HIV-1 subtype B/C intersubtype circulating recombinant form (CRF118_BC) identified in Yunnan, China. *J Infect*. 2021; 83(2): 237-79.
10. Arruda MB, Boullosa LT, Cardoso CC, Costa CM, Alves CRB, Lima STS, et al. Brazilian network for HIV drug resistance surveillance (HIV-BrasNet): a survey of treatment-naïve individuals. *J Int AIDS Soc*. 2018; 21: e25032.
11. Gräf T, Bello G, Andrade P, Arantes I, Pereira JM, da Silva ABP, et al. HIV-1 molecular diversity in Brazil unveiled by 10 years of sampling by the national genotyping network. *Sci Rep*. 2021; 11(1): 15842.
12. Reis MNG, Bello G, Guimarães ML, Stefani MMA. Characterization of HIV-1 CRF90_BF1 and putative novel CRFs_BF1 in Central West, North and Northeast Brazilian regions. *PLoS One*. 2017; 12(6): e0178578.
13. Martin DP, Varsani A, Roumagnac P, Botha G, Suresh M, Schwab KZ, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol*. 2021; 7(1): veaa087.
14. Schultz AK, Zang M, Bulla I, Leitner T, Korber B, Morgenstern B, et al. jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res*. 2009; 37(Web Server issue): W647-51.
15. Salminen MO, Johansson B, Sonnerborg A, Ayeahunie S, Gotte D, Leinikki P, et al. Full-length sequence of an Ethiopian human immunodeficiency virus type-1 (HIV-1) isolate of genetic subtype C. *AIDS Res Hum Retroviruses*. 1996; 12(14): 1329-39.
16. Passaes CPB, Guimarães ML, Bello G, Morgado MG. Near full-length genome characterization of HIV type 1 unique BC recombinant forms from southern Brazil. *AIDS Res Hum Retroviruses*. 2009; 25: 12.
17. Pessôa R, Loureiro P, Lopes ME, Carneiro-Proietti ABF, Sabino EC, Busch MP, et al. Ultra-deep sequencing of HIV-1 near full-length and partial proviral genomes reveals high genetic diversity among Brazilian blood donors. *PLoS One*. 2016; 11(3): e0152499.
18. Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999; 41: 95-8.
19. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005; 33(2): 511-8.
20. Trifinopoulos J, Nguyen LT, Haeseler AV, Minh BQ. W-IQ-TREE: a fast-online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016; 44: W232-5.
21. Anisimova M, Gascuela O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol*. 2006; 55(4): 539-52.
22. Csillag C. HIV-1 subtype C in Brazil. *Lancet*. 1994; 344: 1354.