Rare HIV-1 Subtype J Genomes and a New H/U/CRF02_AG Recombinant Genome Suggests an Ancient Origin of HIV-1 in Angola

Inês Bártolo,[1,*] Rita Calado,[1,*] Pedro Borrego,[1] Thomas Leitner,[2] and Nuno Taveira[1,3]

Abstract

Angola has an extremely diverse HIV-1 epidemic fueled in part by the frequent interchange of people with the Democratic Republic of Congo (DRC) and Republic of Congo (RC). Characterization of HIV-1 strains circulating in Angola should help to better understand the origin of HIV-1 subtypes and recombinant forms and their transmission dynamics. In this study we characterize the first near full-length HIV-1 genomic sequences from HIV-1 infected individuals from Angola. Samples were obtained in 1993 from three HIV-1 infected patients living in Cabinda, Angola. Near full-length genomic sequences were obtained from virus isolates. Maximum likelihood phylogenetic tree inference and analyses of potential recombination patterns were performed to evaluate the sequence classifications and origins. Phylogenetic and recombination analyses revealed that one virus was a pure subtype J, another mostly subtype J with a small uncertain region, and the final virus was classified as a H/U/CRF02_AG recombinant. Consistent with their epidemiological data, the subtype J sequences were more closely related to each other than to other J sequences previously published. Based on the *env* gene, taxa from Angola occur throughout the global subtype J phylogeny. HIV-1 subtypes J and H are present in Angola at low levels since at least 1993. Low transmission efficiency and/or high recombination potential may explain their limited epidemic success in Angola and worldwide. The high diversity of rare subtypes in Angola suggests that Angola was part of the early establishment of the HIV-1 pandemic.

ACCORDING TO THE World Health Organization (WHO) there are currently about 35 million people living with HIV/AIDS worldwide and it is estimated that ∼39 million people have died of AIDS since 1981.[1] HIV can be divided into two types: HIV-1, which is responsible for the worldwide AIDS pandemic, and HIV-2, which is less prevalent and virulent.[2] Based on extreme genetic diversity, HIV-1 variants can be classified into four major groups: M (main), O (outlier), N (non-M, non-O), and more recently P.[3] The M group, which is responsible for the pandemic, includes nine different subtypes (A–D, F–H, J, and K), more than 70 circulating recombinant forms (CRFs), and many unique recombinant forms (URFs).[4] Groups O, N, and P have infected a relatively small number of patients worldwide.[5]

In sub-Saharan Africa almost all subtypes, CRFs, and URFs circulate; the highest genetic diversity has been observed in West Central Africa where the HIV-1 pandemic is believed to have originated.[5–7] On a global perspective, the most prevalent HIV-1 subtypes are C (50%), A (12%), B (11%), followed by CRF02_AG (8%), G (5%), CRF01_AE (5%), and D (2%). The remaining subtypes and recombinant strains represent less than 1% of HIV-1 infections.[5,6,8] Currently there are only three full-length genomes of subtype J and four of subtype H available in the Los Alamos HIV sequence database.[4] Full-length subtype J genomes are from Sweden and Cameroon (GenBank accession numbers AF082394, AF082395, GU237072) whereas subtype H genomes are from Belgium, United Kingdom, and Central African Republic (AF190127, AF190128, FJ711703, AF005496). In addition, subtype J appears as fragments in several CRFs and shorter sequences.[4]

Angola is a South-western African country surrounded by Namibia, Zambia, Democratic Republic of Congo (DRC), and Republic of Congo (RC). According to the UNAIDS report, HIV/AIDS prevalence among adults in Angola was 2.4% in 2014.[9] Despite this low prevalence, all subtypes except B and many CRFs and URFs have been detected in Angola, in particular in the Provinces of Luanda and Cabinda where most studies have been done. This high diversity is a direct

[1]Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Lisboa, Portugal.
[2]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.
[3]Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Instituto Superior de Ciências da Saúde Egas Moniz, Caparica, Portugal.
*These two authors have contributed equally to the work.

consequence of the long-standing presence of HIV-1 in the country and the frequent interchange of people with the DRC and RC.[10–14] In this study we describe the first near full-length HIV-1 genomic sequences from HIV-1 infected individuals from Cabinda, a province of Angola, which is an exclave surrounded by DRC in the south and east, and by the RC in the north.

Blood samples from HIV-1 infected individuals were collected in 1993 from Hospital Distrital de Cabinda, Cabinda, Angola. Samples were collected anonymously with oral consent. The study was approved by the ethics committees of the participating institutions. Virus isolates were obtained using the co-cultivation method as described previously.[15] All three isolates used the CCR5 co-receptor.[16] Viral genomic RNA was extracted from cell culture supernatant and reverse transcriptase–polymerase chain reaction (RT-PCR) was performed using Titan One Tube RT-PCR System (Roche Diagnostic Systems). All amplifications were performed using the Expand Long Template PCR system (Roche Diagnostic Systems) according to the manufacturer's instructions. New primers were designed to amplify the full-length genomes (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/aid). Amplified DNA fragments were purified using JETQUICK Gel Extraction Spin Kit (GenoMed) and sequenced. The nucleotide sequence data were deposited in GenBank with accession numbers KU310618-20. Genomic sequences were aligned with a set of reference sequences representative of all HIV-1 group M subtypes obtained from the Los Alamos HIV Sequence Database[17] using Clustal ×2.1 (www.clustal.org/clustal2/) and the alignment was manually edited with GeneDoc (http://iubio.bio.indiana.edu/soft/molbio/ibmpc/genedoc-readme.html). Maximum likelihood (ML) trees were inferred with program PhyML using the Seaview software (http://pbil.univ-lyon1.fr/software/seaview.html) using the best-fit substitution model identified by Modeltest v3.7 using the Akaike information criterion. To find the ML tree an iterative heuristic method combining nearest neighbor interchange and subtree pruning and regrafting tree rearrangement methods was used. The reliability of the obtained tree was estimated with the approximate likelihood-ratio test. Potential recombination patterns of our new near full-length sequences were analyzed by bootscanning using SimPlot with a sliding window of 500 bp advanced in 100 bp increments (http://sray.med.som.jhmi.edu/SCRoftware/simplot/). For each window 100 bootstrap replicates were generated. Potential recombination breakpoints between subtypes were considered when the percentage of permuted trees for a given subtype was above 70%. We also used RIP, jpHMM, and the branching index (BI) to confirm the bootscanning results.[18–21]

An overall genome-wide tree analysis of 93AOHDC250 and 93AOHDC253 showed that they clustered with subtype J reference sequences and were more closely related to each other than to the other J isolates, which is consistent with their epidemiological data (Fig. 1). Note that this tree is not a true
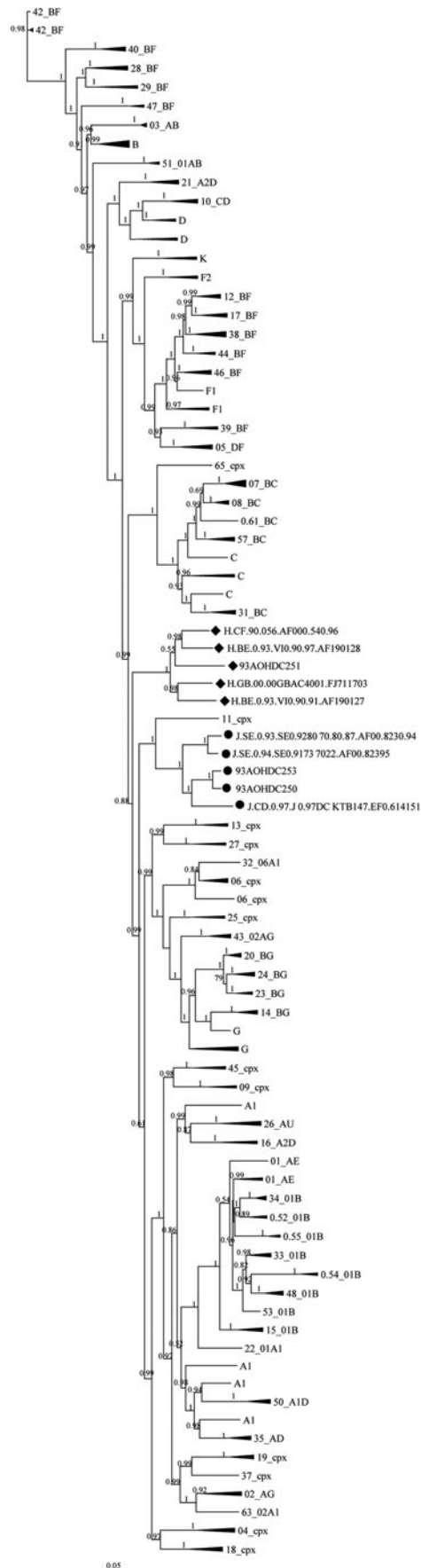

FIG. 1. Overall tree classification of HIV-1 near full-length genomes of isolates 93AOHDC250, 93AOHDC251, and 93AOHDC253. ML trees were constructed with our sequences (highlighted by *dots*) and 136 reference sequences representative of all HIV-1 group M subtypes. Cladistics support is indicated by aLRT values. The scale is in units of substitutions per site. aLRT, approximate likelihood-ratio test; ML, maximum likelihood.
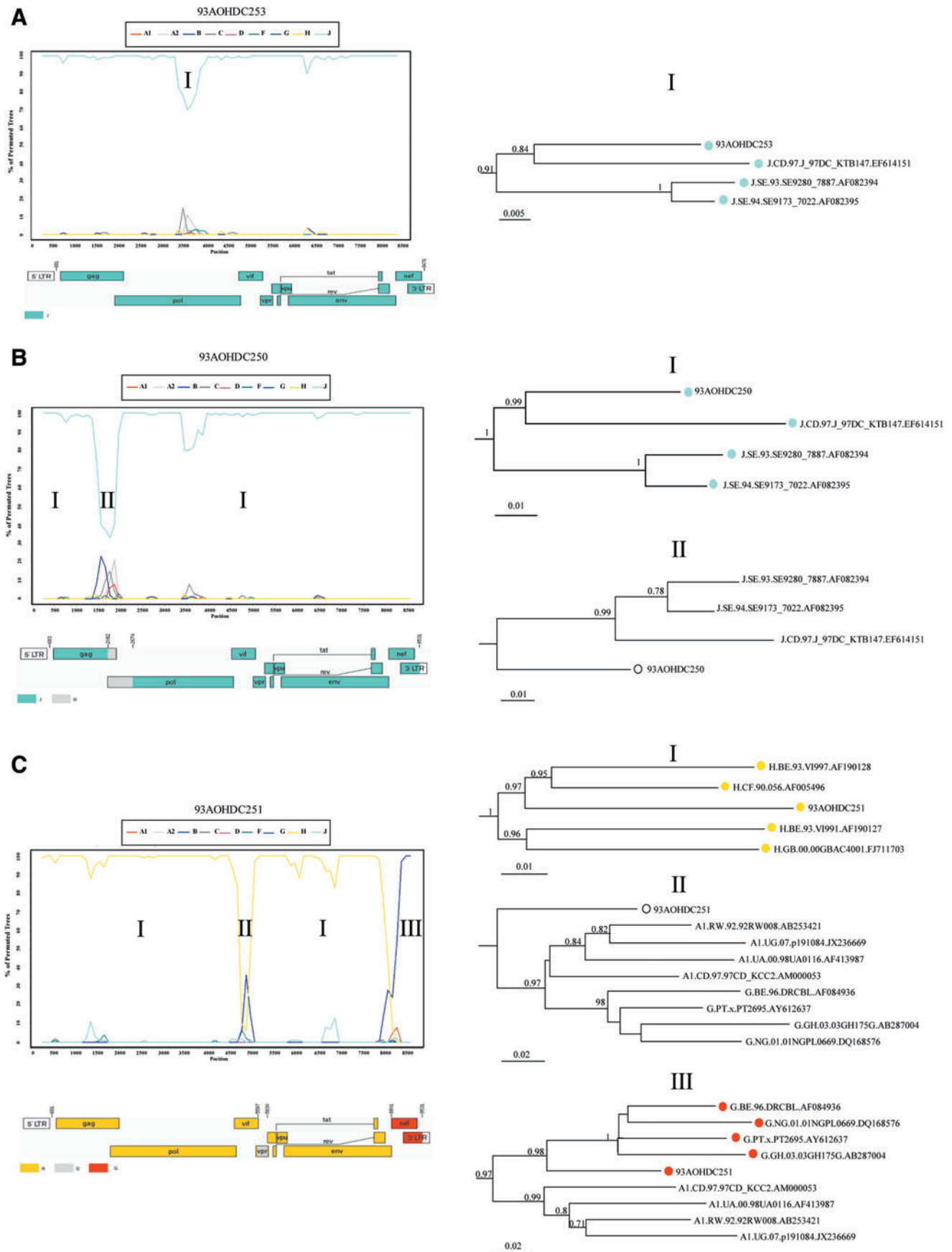
FIG. 2. Genomic segment analysis of near full-length genomes. Panels show bootscanning and tree analyses of isolates 93AOHDC253 (**A**), 93AOHDC250 (**B**) and 93AOHDC251 (**C**). Tree analysis was done for regions suggested by bootscanning to belong to different subtypes. A schematic genome map of the subtype composition of the isolates was produced based on the results obtained by bootscanning. Color images available online at www.liebertpub.com/aid
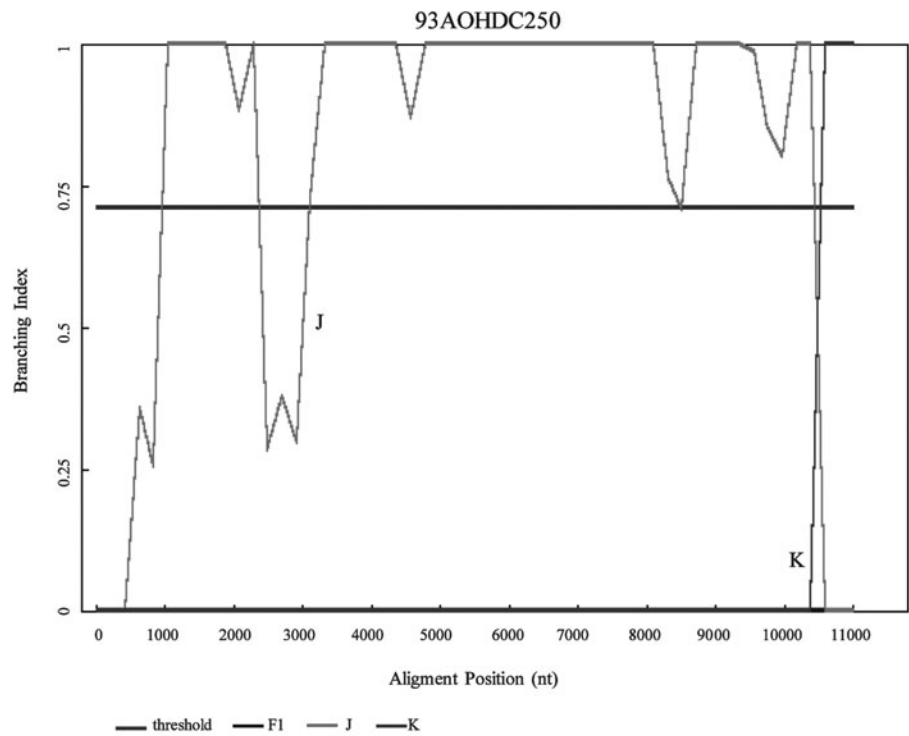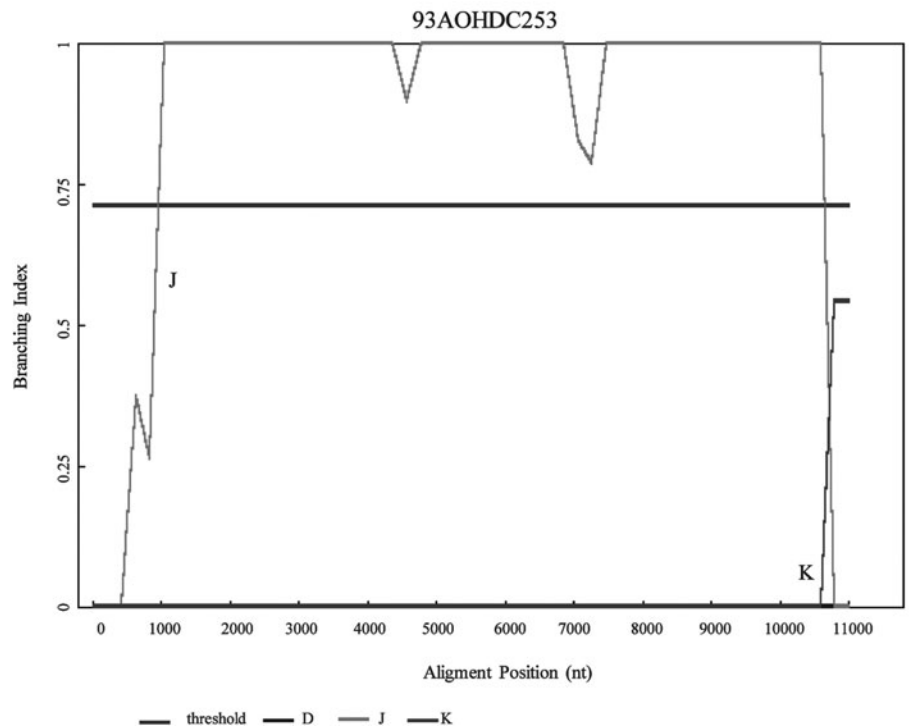
FIG. 3. Branching Index (BI) analysis of 93AOHDC250 and 93AOHDC253.

phylogeny as it cannot depict the evolutionary history of all recombinant sequences; we used it here merely to investigate the overall sequence-based similarity for our classification purpose.

Bootscanning analysis showed no evidence of recombination in isolate 93AOHDC253 (Fig. 2A). Isolate 93AOHDC250 displayed an untypable region between positions 1450 and 2050 corresponding to the end of *gag* gene, the protease (PR) region, and the first 192 nucleotides of RT, which was further

confirmed by phylogenetic analysis (Fig. 2B). BI results confirmed that 93AOHDC253 was a pure subtype J and that 93AOHDC250 had a genomic region with no known subtype in this *gag/pol* region (Fig. 3).

Comparing our new J sequences to existing J sequences in the Los Alamos HIV database revealed that subtype J is quite diverse (Fig. 4). In the region where most J sequence fragments exist (HXB2 positions 7041–7358), 93AOHDC250 and 93AOHDC253 cluster together with a previous sequence from
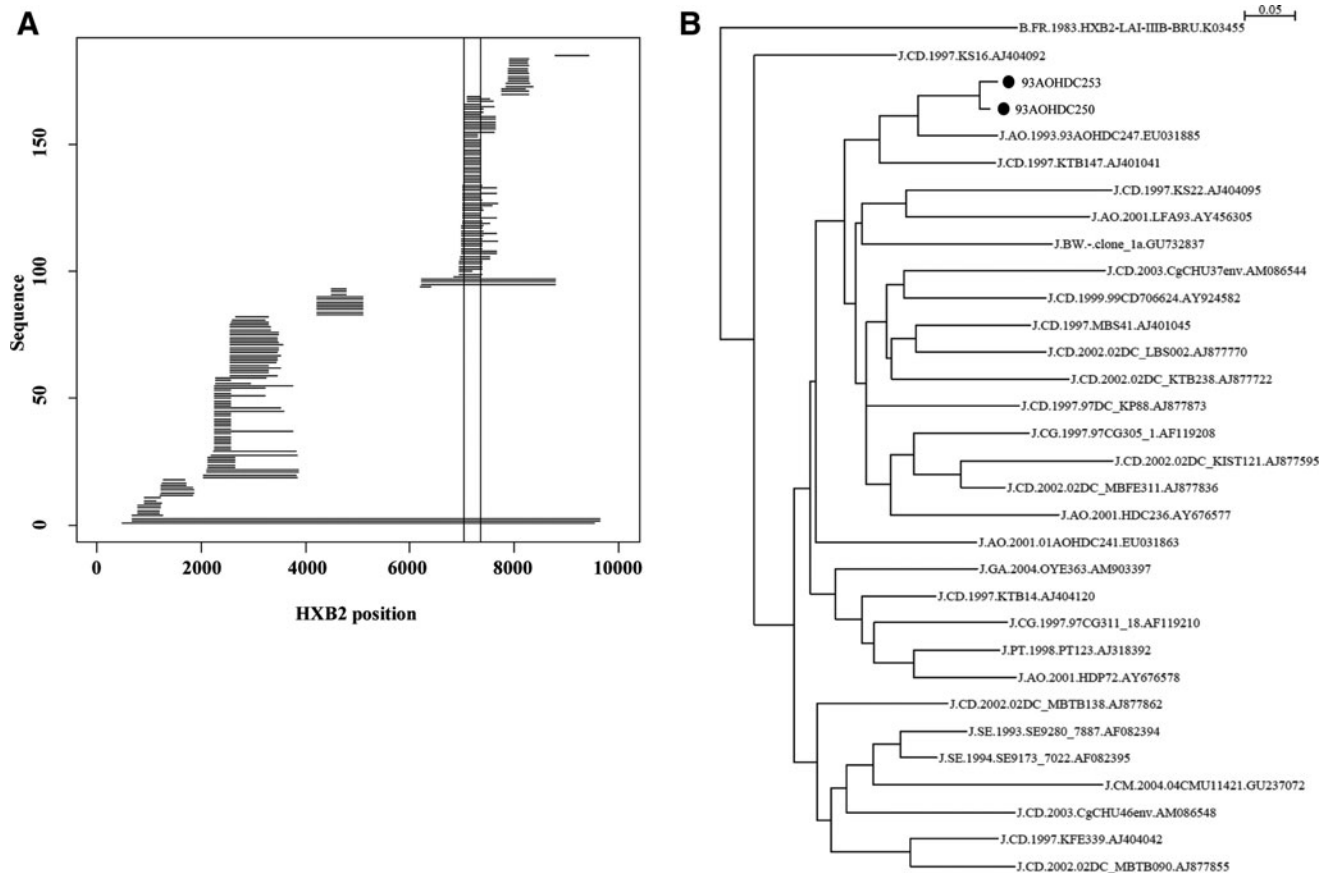
FIG. 4. Subtype J *env* fragment comparison. At the time this report was written 185 subtype J sequence fragments were available in the LANL HIV database. Their HXB2 coordinates covered various parts of the HIV-1 genome **(A)**. Previous to our new sequences, only three full subtype J genomes have been described (long *horizontal lines* at *bottom* of graph). The most sequenced region was a part of *env* (HXB2 coordinates 7041–7358), indicated by the *vertical lines* in the graph. An ML phylogeny of this region revealed that subtype J is very diverse **(B)**. Our new sequences are highlighted by *black dots*.
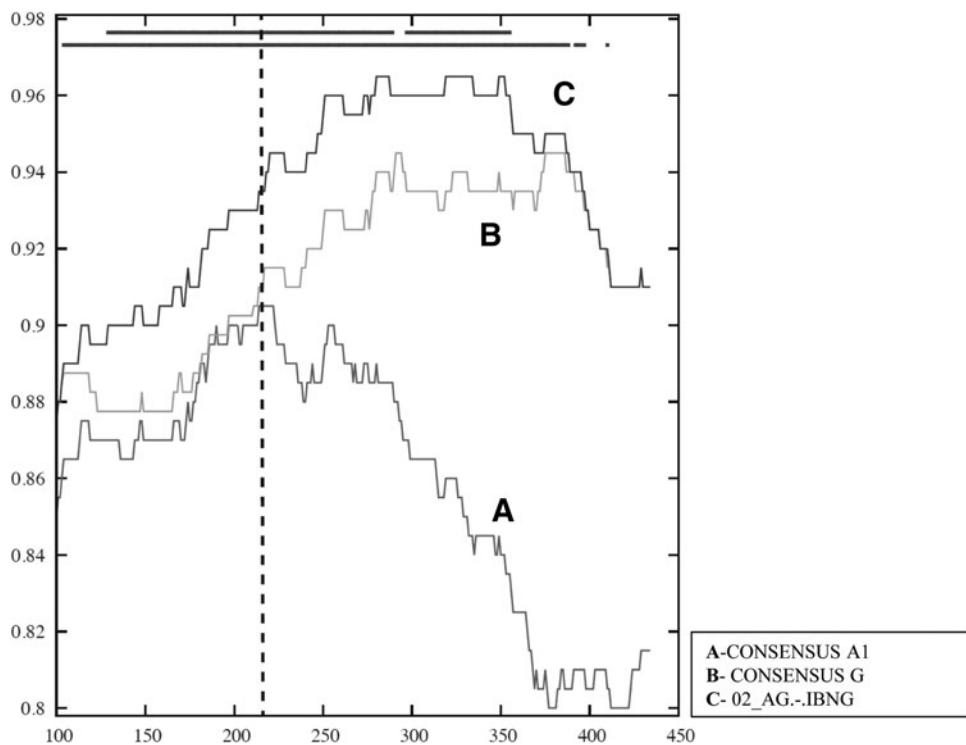

FIG. 5. RIP analysis of segment III of 93AOHDC251. A sliding window analysis (window size 200 nt, step 1 nt; significance threshold 0.90) comparing 93AOHDC251 to consensus subtype sequences A1 and G, and CRF02 prototype sequence IBNG showed that this region was closer to CRF02 in both the A1 and G parts. The CRF02 A1/G breakpoint is indicated by a *bashed vertical line*. The *top bar* indicates statistical significance, and the *lower bar* best match.

Angola, 93AOHDC247.[10] Interestingly, Angolan J sequences occur throughout the phylogeny of this *env* region, suggesting that the J epidemic in Angola is either the origin of subtype J or that there has been a lot of influx of subtype J from other geographic regions. The great diversity of subtype J has been previously noted in analyses of J-containing CRF11 and CRF13 genomes.[22,23]

The overall tree analysis showed that 93AOHDC251 clustered with subtype H reference sequences (Fig. 1). Bootscanning analysis indicated an unclassifiable (U) region between positions 4936 and 5167 (corresponding to *vpr* gene), and a region between positions 8151 and 8888 (corresponding to the *nef* gene and 3′LTR) appeared to cluster between subtypes A1 and G (Fig. 2C). BI analyses again confirmed these results, and suggested that the *nef*/LTR region was below the subtype-defining threshold for subtype G (Fig. 3). Upon closer inspection using RIP, comparing this region to subtypes A1, G, and CRF02, it became clear that this region was in fact derived from CRF02; 93AOHDC251 was closer to CRF02 than either A1 or G in all parts of this region, which included a CRF02 A/G breakpoint (Fig. 5). Together, these analyses led to a classification of this genome sequence as the first H/U/CRF02_AG recombinant.

In this report we describe the first three HIV-1 genomic sequences from Angola, a country that with the DRC and RC had a crucial role in the early dissemination of the HIV-1 epidemic.[7,10] Sequences were derived from isolates obtained in 1993 from patients living in Cabinda, a province in the North that borders both DRC and RC.

One isolate was a pure subtype J, another a J with an at this point unclear segment, and one was an H-based recombinant. The unclear region in 93AOHDC250 may be a divergent J segment or a segment of an as yet undiscovered subtype. As mentioned above, subtype J has previously been described as very diverse and may hide further sub-subtypes or recombinants of sub-subtype nature.[22]

Despite being in circulation in Cabinda since 1993, as of February 2015 only 27 J and 73 H sequences from Angola had been deposited in the Los Alamos HIV Sequence Database (3.1% and 8.3% of respective subtype sequences from Angola). Globally, pure subtypes J and H are also very rare, which suggests low biological fitness, low transmissibility, high recombination potential, and unsuccessful introductions into high-risk populations. However, the isolates described herein replicate to normal levels in different cell types suggesting that their biological fitness is no different from other subtypes.[16] Interestingly, we have found that our J and H isolates are much more sensitive (about nine times) to the CCR5-antagonist TAK-779 when compared with isolates from other subtypes that are more prevalent in Angola and Portugal.[16] This suggests poor binding to the main HIV co-receptor (CCR5), which may indicate transmissibility problems of these subtypes. On the other hand, J and H subtypes have been found in many recombinants such as CRF04_cpx, CRF06_cpx, CRF11_cpx, CRF13_cpx, CRF18_cpx, CRF27_cpx, and CRF49_cpx.[17] Thus, the rarity of pure subtype J and H, together with the large HIV-1 diversity in Angola and neighboring countries, which suggests they have been around for long times, suggests that these subtypes are less fit to transmit or establish infection by themselves. Similar to how latent virus may survive within a host through recombination with nonlatent immune escaping

plasma virus, subtype J and H virus may do better if they recombine with more fit subtypes.[24]

HIV-1 subtypes J and H are present in Angola at low levels since at least 1993. The high diversity among Angolan subtype J *env* sequences and the fact that the rare subtype H has recombined in Angola together suggest that Angola is either the origin of subtype J or, more complicated, that there has been a lot of influx of subtype J from other geographic regions. Low transmission efficiency and/or high recombination potential may explain their limited epidemic success in Angola and worldwide.

Acknowledgments

Sequence Data

Sequences have been assigned with GenBank accession numbers KU310618, KU310619, and KU310620.

Authors' Contributions

Conceived and designed the experiments: I.B., T.L., and N.T. Performed the experiments: I.B., P.B., T.L., and R.C. Analyzed the data: I.B., T.L., R.C., and N.T. Wrote the article: I.B., R.C., T.L., and N.T. The final text was read and approved for submission by all authors.

Author Disclosure Statement

No competing financial interests exist.

References

1. WHO: Fact Sheets HIV/AIDS July 2015. 2015.
2. Santoro MM, Perno CF: HIV-1 genetic variability and clinical implications. ISRN Microbiol 2013;2013:481314.
3. Mourez T, Simon F, Plantier JC: Non-M variants of human immunodeficiency virus type 1. Clin Microbiol Rev 2013; 26:448–461.
4. Foley B, Leitner T, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S, Korber B: *HIV Sequence Compendium 2015*. Los Alamos, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2015.
5. Peeters M, Jung M, Ayouba A: The origin and molecular epidemiology of HIV. Expert Rev Anti Infect Ther 2013; 11:885–896.
6. Buonaguro L, Tornesello ML, Buonaguro FM: Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: Pathogenetic and therapeutic implications. J Virol 2007;81:10209–10219.
7. Faria NR, Rambaut A, Suchard MA, *et al.*: HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science 2014;346:56–61.
8. Abecasis AB, Wensing AM, Paraskevis D, *et al.*: HIV-1 subtype distribution and its demographic determinants in

newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology 2013;10:7.

9. UNAIDS: Epidemiological fact sheets on HIV and AIDS. HIV and AIDS estimates, Angola. Available at www.unaids .org/sites/default/files/epidocuments/AGO.pdf (2013), accessed January 2016.

10. Bartolo I, Rocha C, Bartolomeu J, *et al.*: Highly divergent subtypes and new recombinant forms prevail in the HIV/ AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. Infect Genet Evol 2009;9:672–682.

11. Abecasis A, Paraskevis D, Epalanga M, *et al.*: HIV-1 genetic variants circulation in the North of Angola. Infect Genet Evol 2005;5:231–237.

12. Bartolo I, Epalanga M, Bartolomeu J, *et al.*: High genetic diversity of human immunodeficiency virus type 1 in Angola. AIDS Res Hum Retroviruses 2005;21:306–310.

13. Bartolo I, Rocha C, Bartolomeu J, *et al.*: Antiretroviral drug resistance surveillance among treatment-naive human immunodeficiency virus type 1-infected individuals in Angola: Evidence for low level of transmitted drug resistance. Antimicrob Agents Chemother 2009;53:3156–3158.

14. Bartolo I, Zakovic S, Martin F, *et al.*: HIV-1 diversity, transmission dynamics and primary drug resistance in Angola. PLoS One 2014;9:e113626.

15. Cavaco-Silva P, Taveira NC, Rosado L, *et al.*: Virological and molecular demonstration of human immunodeficiency virus type 2 vertical transmission. J Virol 1998;72:3418–3422.

16. Borrego P, Calado R, Marcelino JM, *et al.*: Baseline susceptibility of primary HIV-2 to entry inhibitors. Antivir Ther 2012;17:565–570.

17. Los Alamos Sequence Database. Available at www.hiv .lanl.gov/ (2016), accessed January 12, 2016.

18. Schultz AK, Zhang M, Bulla I, *et al.*: jpHMM: Improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res 2009;37:W647–W651.

19. 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–W465.

20. Hraber P, Kuiken C, Waugh M, Geer S, Bruno WJ, Leitner T: Classification of hepatitis C virus and human immunodeficiency virus-1 sequences with the branching index. J Gen Virol 2008;89(Pt 9):2098–2107.

21. Wilbe K, Salminen M, Laukkanen T, *et al.*: Characterization of novel recombinant HIV-1 genomes using the branching index. Virology 2003;316:116–125.

22. Zhang M, Wilbe K, Wolfe ND, Gaschen B, Carr JK, Leitner T: HIV type 1 CRF13_cpx revisited: Identification of a new sequence from Cameroon and signal for subsubtype J2. AIDS Res Hum Retroviruses 2005;21:955–960.

23. Wilbe K, Casper C, Albert J, Leitner T: Identification of two CRF11-cpx genomes and two preliminary representatives of a new circulating recombinant form (CRF13-cpx) of HIV type 1 in Cameroon. AIDS Res Hum Retroviruses 2002;18:849–856.

24. Immonen TT, Conway JM, Romero-Severson EO, Perelson AS, Leitner T: Recombination enhances HIV-1 envelope diversity by facilitating the survival of latent genomic fragments in the plasma virus population. PLoS Comput Biol 2015;11:e1004625.

Address correspondence to:
Nuno Taveira
Instituto de Investigação do Medicamento (iMed.ULisboa)
Faculdade de Farmácia
Universidade de Lisboa
Av. Prof. Gama Pinto
Lisboa 1649-003
Portugal

E-mail: ntaveira@ff.ul.pt