Infection, Genetics and Evolution 12 (2012) 1458–1465

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Evidence of multiple introductions of HIV-1 subtype C in Angola

Joana Morais Afonso [a,b], Mariza G. Morgado [a], Gonzalo Bello [a,*]

[a] Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, Brazil
[b] Faculdade de Medicina, Universidade Agostinho Neto, Luanda, Angola

ARTICLE INFO

Article history:
Received 16 March 2012
Received in revised form 10 May 2012
Accepted 12 May 2012
Available online 24 May 2012

Keywords:
HIV-1
Subtype C
Angola
Phylogeography

ABSTRACT

HIV-1 subtype C is the most prevalent group M clade in southern Africa and some eastern African countries. Subtype C is also the most frequent subtype in Angola (southwestern Africa), with an estimated prevalence of 10–20%. In order to better understand the origin of the HIV-1 subtype C strains circulating in Angola, 31 subtype C *pol* sequences of Angolan origin were compared with 1950 subtype C *pol* sequences sampled in other African countries. Phylogenetic analyses reveal that the Angolan subtype C sequences were distributed in 16 different lineages that were widely dispersed among other African strains. Ten subtype C Angolan lineages were composed by only one sequence, while the remaining six clades contain between two and seven sequences. Bayesian phylogeographic analysis indicates that most Angolan clades probably originated in different southern African countries with the exception of one lineage that most likely originated in Burundi. Evolutionary analysis suggests that those Angolan subtype C clades composed by $\geqslant 2$ sequences were introduced into the country between the late 1970s and the mid 2000s. The median estimated time frame for the origin of those Angolan lineages coincides with periods of positive migration influx in Angola that were preceded by phases of negative migratory outflow. These results demonstrate that the Angolan subtype C epidemic resulted from multiple introductions of subtype C viruses mainly imported from southern African countries over the last 30 years, some of which have been locally disseminated establishing several autochthonous transmission networks. This study also suggests that population mobility between Angola and southern African countries during civil war (1974–2002) may have played a key role in the emergence of the Angolan subtype C epidemic.

1. Introduction

Analysis of the global distribution of human immunodeficiency virus type 1 (HIV-1) group M subtypes and recombinants worldwide reveals that subtype C is the most prevalent variant, accounting for nearly half (48%) of all global infections (Hemelaar et al., 2011). Subtype C is mainly found in southern Africa, east Africa and India, with smaller numbers in central Africa and Brazil (Hemelaar et al., 2011).

Angola is a southwestern African country bordered by the Democratic Republic of Congo (DRC) and the Republic of Congo on the north, Zambia on the east and Namibia on the south. Angola displays a particularly complex HIV-1 molecular epidemiologic scenario characterized by the circulation of most group M subtypes, a large collection of inter-subtype recombinants, and other unclassified group M strains (Abecasis et al., 2005; Bartolo et al., 2005, 2009; Castelbranco et al., 2010). The extraordinary degree of viral genetic diversity observed in Angola probably reflects its close geographical proximity with the DRC that has been identified as the epicenter of the HIV-1 group M clade (Vidal et al., 2000).

Subtypes C and F1 are the most prevalent HIV-1 clades in Angola, accounting for nearly 30–40% of HIV-infections (Abecasis et al., 2005; Castelbranco et al., 2010). Previous studies revealed that the subtype F1 epidemic in Angola resulted from the successful dissemination of a single viral strain probably originated in the DRC at around the early 1980s (Bello et al., 2012; Guimaraes et al., 2009; Mehta et al., 2011). Whether the high prevalence of subtype C in Angola also resulted from the efficient expansion of a local virus or from the intense viral influx from neighboring countries is unclear. In this study, we use a combination of phylogeographic and evolutionary analyses to trace the origin and time-scale of the HIV-1 subtype C epidemic in Angola.

2. Material and methods

2.1. Angolan HIV-1 subtype C sequences

A total of 31 HIV-1 subtype C *pol* (PR/RT) sequences obtained from drug-naive Angolan patients were used in this study. Sixteen sequences sampled in 2001 were retrieved from the Los Alamos

* Corresponding author. Lab. de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz – FIOCRUZ, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil. Tel.: +55 21 3865 8154; fax: +55 21 3865 8173.
E-mail address: gbello@ioc.fiocruz.br (G. Bello).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.05.005

HIV sequence database (http://www.hiv.lanl.gov) and described in a previous work (Bartolo et al., 2009). The remaining 15 sequences were collected from drug-naive patients attending the São Lucas Medical Center (CSSL) in Kifangondo city (Bengo province) between 2008 and 2010, with the approval of the local Ethical Committee. A fragment of ~1100 bp from the *pol* (*PR/RT*) gene region (nucleotides 2253–3272 relative to HXB2) was amplified and sequences as described elsewhere (Bello et al., 2008).

2.2. HIV-1 subtype C reference sequences

Subtype C Angolan sequences were combined with subtype C sequences from southern, east, and central African countries available at the Los Alamos HIV database that matched the selected genomic region. In order to minimize the effect of drug-induced convergent evolution on the phylogenetic analyses, only sequences retrieved from antiretroviral therapy naïve individuals were selected. Only one sequence per individual was included. Sequences retrieved from Los Alamos HIV database were submitted to the REGA HIV subtyping tool v.2 (de Oliveira et al., 2005) and all sequences with evidence of erroneous subtype assignment or evidence of inter-subtype recombination were excluded. Sequences containing frame-shift mutations or deletions and those sequences from countries poorly represented ($n < 10$ sequences) were also removed.

2.3. Sequence alignments and phylogenetic analysis

Sequence alignments were created using the Clustal X program (Thompson et al., 1997) and are available from the authors upon request. Phylogenetic trees were inferred by the maximum likelihood (ML) method under the GTR+I+Γ_4 nucleotide substitution model, selected using the jModeltest program (Posada, 2008). ML trees were reconstructed with program PhyML (Guindon and Gascuel, 2003) using an online web server (Guindon et al., 2005). Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (*aLRT*) (Anisimova and Gascuel, 2006) based on the Shimodaira-Hasegawa-like procedure. ML trees were visualized using the FigTree v1.3.1 program (Rambaut, 2009).

2.4. Phylogeographic and evolutionary analysis

The evolutionary rate (μ, units are nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{mrca}, years), and the probable geographic origin of Angolan subtype C lineages were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in the BEAST software package v1.6.2 (Drummond et al., 2002; Drummond and Rambaut, 2007). The temporal structure of the selected dataset was not sufficient to reliably estimate the evolutionary rate employing the dates of the sequences. Therefore, the interval of mean substitution rates at *pol* gene previously estimated for HIV-1 subtype C ($1.5–2.5 \times 10^{-3}$ substitutions/site/year) (Bello et al., 2008; Dalai et al., 2009) was incorporated as a prior probability distribution in our analyses. Analyses were performed using the GTR+I+Γ nucleotide substitution model, a relaxed (uncorrelated Lognormal) molecular clock model (Drummond et al., 2006), a Bayesian skyline coalescent tree prior (Drummond et al., 2005), and a discrete phylogeographic model in which all possible reversible exchange rates between locations were equally likely (Lemey et al., 2009). Two separate MCMC chains were run for 10×10^7 generations with a burn-in of 1×10^7. Adequate chain mixing was checked by calculating the Effective Sample Size (ESS) using TRACER v1.4 (Rambaut and Drummond, 2007), after excluding an initial 10%. All parameter estimates for each run showed ESS values >100. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator after removing an initial 10%, and visualized with FigTree v1.3.1.

2.5. GenBank accession numbers

The new Angolan HIV-1 subtype C *pol* sequences used in this study have been deposited in GenBank under Accession Nos. JN937017, JN937018, JN937022, JN937024, JN937032, JN937040, JN937055, JN937067, JN937071, JN937076, JN937091, JN937094, JN937095, JN937105 and JN937110.

3. Results

3.1. Multiple introductions of HIV-1 subtype C in Angola

To trace the geographical origin of the HIV-1 subtype C circulating in Angola, 31 subtype C *pol* sequences sampled in Angola between 2001 and 2010 were analyzed in the present study. HIV samples collected in 2001 ($n = 16$) were mainly obtained from patients living in the Angolan provinces of Luanda and Cabinda (Bartolo et al., 2009); while new subtype C sequences collected from 2008 to 2010 ($n = 15$) were retrieved from patients living in different Angolan provinces from the southern (Benguela = 4, Huíla = 1 and Namíbe = 5), central (Luanda = 1, Bengo = 1 and Cuanza Norte = 1), and northern (Cabinda = 1 and Zaire = 1) country regions (Table 1). Most new Angolan subtype C sequences were retrieved from females (73%) and heterosexual (93%) individuals (Table 1). Subtype C Angolan sequences were analyzed in comparison to 1950 subtype C *pol* references sequences retrieved from Los Alamos HIV database that were collected at 12 different African countries with an estimated prevalence of this variant higher than 5% (Hemelaar et al., 2011) (Table 2).

The Angolan subtype C strains were initially compared with those sequences from South Africa that represent the majority (52%) of subtype C sequences in our dataset. The close relatives South African sequences (descendant of nodes that are first to fifth level ancestor of at least one sequence from Angola) were selected up to a maximum of 100 (Supplementary Fig. S1) and combined with subtype C sequences from the other African countries. The final ML phylogenetic tree shows that the 31 Angolan strains were distributed in 16 different lineages (I–XVI) that were interspersed with African sequences from other countries (Fig. 1), thus demonstrating that subtype C was introduced into Angola on multiple occasions. Nine subtype C Angolan lineages were composed by only one sequence, while the remaining seven Angolan clades contain between two and seven sequences. Most Angolan lineages

Table 1
Epidemiological characteristics of HIV-1 subtype C infected Angolan patients enrolled from 2008 to 2010.

Sample ID	Sex	Age	Risk group	Sampling year	Province
JN937017	Female	21	Heterosexual	2008	Namibe
JN937018	Female	27	Heterosexual	2008	Namibe
JN937022	Female	40	Heterosexual	2008	Benguela
JN937022	Female	27	Heterosexual	2008	Namibe
JN937032	Female	19	Heterosexual	2008	Cuanza Norte
JN937040	Male	41	MSM	2008	Luanda
JN937055	Male	30	Heterosexual	2009	Huíla
JN937067	Female	35	Heterosexual	2010	Namibe
JN937071	Female	28	Heterosexual	2010	Bengo
JN937076	Female	30	Heterosexual	2010	Zaire
JN937091	Male	40	Heterosexual	2010	Namibe
JN937094	Female	31	Heterosexual	2010	Benguela
JN937095	Male	45	Heterosexual	2010	Benguela
JN937105	Female	36	Heterosexual	2010	Cabinda
JN937110	Female	43	Heterosexual	2010	Benguela