



## Research paper

## HTLV-1aA introduction into Brazil and its association with the trans-Atlantic slave trade



Adjile Edjide Roukiyath Amoussa<sup>a,1</sup>, Eduan Wilkinson<sup>b,1</sup>, Marta Giovanetti<sup>a,c</sup>,  
Filipe Ferreira de Almeida Rego<sup>a,d</sup>, Thessika Hialla A Araujo<sup>a,e</sup>, Marilda de Souza Gonçalves<sup>a</sup>,  
Tulio de Oliveira<sup>b</sup>, Luiz Carlos Junior Alcantara<sup>a,\*</sup>

<sup>a</sup> Laboratory of Hematology, Genetics and Computational Biology/Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil

<sup>b</sup> Africa Centre for Health and Population Studies, University of Kwazulu-Natal, Durban, South Africa

<sup>c</sup> University of Rome "Tor Vergata", Italy

<sup>d</sup> Catholic University of Salvador, Salvador, Bahia, Brazil

<sup>e</sup> Bahia School of Medicine and Public Health/Bahia Foundation for Development of Science, Salvador, Bahia, Brazil

## ARTICLE INFO

## Article history:

Received 9 September 2016

Received in revised form 21 November 2016

Accepted 6 December 2016

Available online 8 December 2016

## Keywords:

HTLV

Brazil

Southern Africa

Trans-Atlantic slave trade

## ABSTRACT

**Introduction:** Human T-lymphotropic virus (HTLV) is an endemic virus in some parts of the world, with Africa being home to most of the viral genetic diversity. In Brazil, HTLV-1 is endemic amongst Japanese and African immigrant populations. Multiple introductions of the virus in Brazil from other epidemic foci were hypothesized. The long terminal repeat (LTR) region of HTLV-1 was used to infer the origin of the virus in Brazil, using phylogenetic analysis.

**Methods:** LTR sequences were obtained from the HTLV-1 database (<http://htlv1db.bahia.fiocruz.br>). Sequences were aligned and maximum-likelihood and Bayesian tree topologies were inferred. Brazilian specific clusters were identified and molecular-clock and coalescent models were used to estimate each cluster's time to the most recent common ancestor (tMRCA).

**Results:** Three Brazilian clusters were identified with a posterior probability ranged from 0.61 to 0.99. Molecular clock analysis of these three clusters dated back their respective tMRCAs between the year 1499 and the year 1668. Additional analysis also identified a close association between Brazilian sequences and new sequences from South Africa.

**Conclusion:** Our results support the hypothesis of a multiple introductions of HTLV-1 into Brazil, with the majority of introductions occurring in the post-Colombian period. Our results further suggest that HTLV-1 introduction into Brazil was facilitated by the trans-Atlantic slave trade from endemic areas of Africa. The close association between southern African and Brazilian sequences also suggested that greater numbers of the southern African Bantu population might also have been part of the slave trade than previously thought.

© 2016 Published by Elsevier B.V.

## 1. Introduction

Human T-lymphotropic virus type 1 (HTLV-1) was the first human pathogenic retrovirus to be discovered (Gallo, 2011). HTLV-1 is associated with many severe diseases, such as Adult T-cell leukemia/lymphoma (ATL) (Poiesz et al., 1980; Yoshida et al., 1982) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) (Gessain et al., 1985). However, the majority of infected people remain asymptomatic (Bandeira et al., 2015). It is estimated that at least between 5 and 10 million people are infected with the virus, with specific

areas of endemicity (Gessain and Cassar, 2012). The endemic areas for this virus are sub-Saharan Africa, Central and South America, the Caribbean, Japan, Melanesia and the Middle East (Proietti et al., 2005; Gonçalves et al., 2010). Japan and the African continent are the most important areas where HTLV-1 is endemic, followed by South America (Gessain and Cassar, 2012).

Phylogenetic analysis of the LTR region of HTLV-1, the most diverse region of the viral genome, classifies the virus into seven major genetic subtypes (a–g). These are: a, or the cosmopolitan subtype (Miura et al., 1994); b, or the Central African subtype (Vandamme et al., 1994); c, or the Melanesian subtype (Gessain et al., 1991); d, or the Cameroonian subtype (Chen et al., 1995; Mahieux et al., 1997); e, or the Democratic Republic of Congo subtype; f, or the Gabonese subtype (Salemi et al., 1998); and g, also from Cameroon (Wolfe et al., 2005). The cosmopolitan subtype is divided into five subgroups: A, or transcontinental, B or

\* Corresponding author at: Fundação Oswaldo Cruz, Salvador, Bahia, Brazil.

E-mail addresses: [amoussaroukiyath@yahoo.fr](mailto:amoussaroukiyath@yahoo.fr) (A.E.R. Amoussa),

[ewilkinson83@gmail.com](mailto:ewilkinson83@gmail.com) (E. Wilkinson), [lalcan@bahia.fiocruz.br](mailto:lalcan@bahia.fiocruz.br) (L.C.J. Alcantara).

<sup>1</sup> Same contribution.

Japanese, C or West African, D or North African (Miura et al., 1994) and E or Black Peruvian (Van Dooren et al., 1998). However, the transcontinental subgroup of the cosmopolitan subtype (HTLV-1aA) is spread worldwide (Gessain and Cassar, 2012), mainly in Latin America and southern Africa (Alcantara et al., 2006). The HTLV-1aC genotype is described in West Africa while recent reports of HTLV-1aD have been documented in West African countries like Guinea-Bissau, Gambia, Ghana, Ivory Coast and Senegal (Zehender et al., 2008; van Tienen et al., 2012; Desrames et al., 2014). Furthermore, a recent report described viral recombination between HTLV-1 strains currently present in North Africa and West Africa (Desrames et al., 2014).

Brazil is the country in South America that received the bulk of immigrants hailing from different parts of the world. An admixed population comprising people of African, European, Asian and native America backgrounds, characterizes the population of Brazil today. HTLV-1 in Brazil is concentrated amongst individuals who descend from immigrant populations from areas where the virus is endemic, such as Africa and Japan (Paiva and Casseb, 2015). The city of Salvador is estimated to be home to the largest number of HTLV-1 infections in Brazil and it shares many socio-demographic characteristics of some African countries (Alcantara et al., 2003; Mota et al., 2007; Rego et al., 2008). A large proportion of the city's population are the descendants of African immigrants. It is estimated that Brazil received roughly four million slaves during the trans-Atlantic slave trade (Alcantara et al., 2003). It is believed that the bulk of slaves were sourced from Portuguese ports in western Africa (e.g. present day Angola and Guinea Bissau). However, some recent studies reported a close association between HTLV-1 isolates from Brazil and the southern African region (Alcantara et al., 2006; Mota et al., 2007). This close relationship suggests that HTLV-1 was also brought to Brazil via the slave trade from southern African ports (e.g. Mozambique and Madagascar). On a human genetic level, it has also been demonstrated that 26.5% of the descendants of African slaves living in Salvador, Bahia, carry the BEN/CAR (Benin/Central African Republic)  $\beta$ -globin haplotype. This haplotype of the  $\beta$ -globin gene circulates primarily amongst Bantu speakers from the southern African region (Alcantara et al., 2006; Rego et al., 2008).

Brazil also received large numbers of Japanese migrants during the first half of the 20th century (Alcantara et al., 2003). In Brazil, HTLV-1 was first described in the Japanese immigrant population from Okinawa, Southern Japan, residing in the city of Campo Grande, state of Mato Grosso do Sul. Recently, a prevalence of 6.8% of infection was found amongst adult Japanese immigrants and their descendants living in Campo-Grande (Mato Grosso do Sul). Phylogenetic analysis performed in this population showed that the cosmopolitan subtype of Transcontinental and Japanese subgroup are the most circulated amongst this population (Bandeira et al., 2015).

Several studies have investigated the origin of HTLV-1 in Latin America, particularly in Brazil (Alcantara et al., 2003; Magalhães et al., 2008; Aleluia et al., 2015). The majority of these studies support the hypothesis of multiple introductions of HTLV-1aA into Brazil during the post-Columbian period (> 1492) (Van Dooren et al., 1998; Alcantara et al., 2003; Rego et al., 2008). On the basis of this consideration, the aim of the present study was to estimate the date of the introduction of HTLV-1 into Brazil (i.e. time to the most recent common ancestor) and investigate the relationship amongst African and Brazilian isolates in order to better understand the ethnic origin of people of African descent residing within Brazil today.

## 2. Methods

### 2.1. Dataset composition

A total of 390 HTLV-1 LTR sequences were retrieved from the HTLV database (<http://htlv1db.bahia.fiocruz.br>), including 120 sequences from Brazil (Araujo et al., 2012). A breakdown of the 390 sequences, their Genbank accession numbers, country of origin and sampling

dates are provided in Supplementary materials. Each sequence was annotated as follows: (i) two-letter country code corresponding to the International Organization for Standardization (ISO) codes, (ii) their unique Genbank accession numbers, (iii) the subtype and (iv) their isolation date (if available). Sequences without a date of sampling were assigned an arbitrary isolation date (e.g. 2007 or 2015). We feel this to be acceptable because of the low mutation rate of HTLV-1 and the fact that the majority of sequences were derived from contemporary sampling (~1990 to the present). Genbank accession numbers of the LTR sequences obtained from the HTLV database can be found in Supplementary materials.

### 2.2. Preliminary phylogenetic analysis of the LTR dataset

All the 390-LTR sequences were aligned against one another in ClustalW (Thompson et al., 1997) using a quick tree method to speed up the alignment process. The alignment was manually edited using the Geneious v 8.0.5 software package (Kearse et al., 2012). Prior to any phylogenetic inference, we performed recombination analysis in order to rule out viral recombination. To this end, we aligned the 390 LTR sequences, along with 41 LTR sequences recently described by Desrames et al., 2014 as a positive control, as described previously (Desrames et al., 2014). Desrames and colleagues described the presence of viral recombination amongst the 41 LTR sequences which they collected from 19 different African countries. GenBank accession numbers of the 41 LTR sequences recently described by Desrames et al., 2014 are presented in Supplementary materials.

The alignment was imported into Recombination Detection Programme (RDP) v 4.0 (Martin et al., 2015) and a detailed scan for viral recombination was performed. Once recombination had been ruled out, we inferred a maximum likelihood (ML)-tree topology in RAXML v 8.0.0 (Stamatakis, 2014). The general time reversible (GTR) model of nucleotide substitution (Lanave et al., 1984; Rodríguez et al., 1990) and an estimated gamma shape parameter (Anderson and Ray, 1975) was used to infer the tree in RAXML. The multiple or rapid bootstrap resampling method ( $n = 1,000$ ) was implemented on a 12-core MacPro to infer support for splits. The ML-tree topology was visualized in FigTree v 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>) and rooted on the Subgroup C ancestral branch. The ML-tree topology was further analysed in Path-O-gen to evaluate the molecular clocklikeness of the dataset. To this end, all non-transcontinental sequences were removed from the tree topology with the ape package in R prior to analysis in Path-O-gen. This ensured that reliable clocklike parameters for the HTLV-1aA group were obtained.

Due to poor bootstrap support values in the ML-tree topology, a Bayesian tree topology was inferred for the entire LTR dataset in MrBayes v 3.2.5 (Huelsenbeck and Ronquist, 2001) using the GTR model of nucleotide substitution (Lanave et al., 1984; Rodríguez et al., 1990). To speed up the analysis, we inferred Bayesian tree topologies using the multi-threaded version of MrBayes with four independent chains spread over four computational nodes. Posterior trees and parameters were sampled in a Metropolis Hastings Markov Chain Monte Carlo (MCMC) chain with a total length of 100 million steps and a sample frequency of every 10,000 steps. Following 100 million steps in the analysis, the MCMC was evaluated based on the average standard deviation of splits (<0.01) and the posterior parameters, which were evaluated in the Tracer v 1.6 software application (<http://beast.bio.ed.ac.uk/Tracer>). The resulting Bayesian tree topology was visualized in FigTree v 1.4.2 and manually edited for better visual interpretation.

### 2.3. Phylodynamic inference of the HTLV epidemic in Brazil

Three Brazilian clusters were identified in the Bayesian tree topology. The date of origin of each clusters was estimated using the BEAST v 1.8.2 software package (Drummond and Rambaut, 2007). Molecular clock analysis was implemented under the Tamura-Nei model of

nucleotide substitution, an estimated gamma shape parameter (TN93 + G) and a strict molecular clock assumption. Since HTLV-1 is an endemic virus with no signs of spreading, we chose to infer epidemic histories under a constant population size tree prior (Drummond and Rambaut, 2007) and a fixed mutation rate of  $2.0 \times 10^{-5}$  substitutions/site/year. MCMC analysis for each cluster were run for 1 billion steps, while posterior parameters and trees were sampled every 100,000 step in the chain. MCMC convergence and overall performance (based on large Effective Sample Sizes, ESS > 200) were assessed in Tracer v 1.6. Following the discarding of the first 10% of samples from the posterior distribution, the estimated root height for each cluster was recorded along with the 95% highest posterior density (HPD).

Our justification for using a mutation rate of  $2.0 \times 10^{-5}$  substitutions/site/year was based on the estimated mutation rate that was calculated by Alcantara and colleagues (Alcantara et al., 2006). This rate was estimated from a vertical transmission pair between a mother and an infant containing a single point mutation between the two. When Alcantara and colleagues added an additional 16-transmission pairs, the estimated mutation rate was approximately ten times slower ( $4.49 \times 10^{-5}$  substitutions/site/year, 95% confidence interval  $1.34 \times 10^{-5}$ – $1.08 \times 10^{-6}$ ). However, out of the combined 17 transmission chains, only three point mutations were informative. Our rate was within the lower 95% confidence interval of the 17 pair estimated rate. Furthermore, our molecular clock analysis of our ML-tree topology in Path-O-Gen, suggested a mutation rate of  $3.67 \times 10^{-5}$  substitutions/site/year, based on the slope of the root to tip regression.

#### 2.4. Phylodynamic inference of the HTLV infection in South Africa

In order to analyse the phylogenetic relationship between Brazilian and South African sequences, another dataset was constructed by adding seven recently genotyped HTLV-1 isolates from South Africa. These sequences were derived from HTLV-1/HIV-1 co-infected

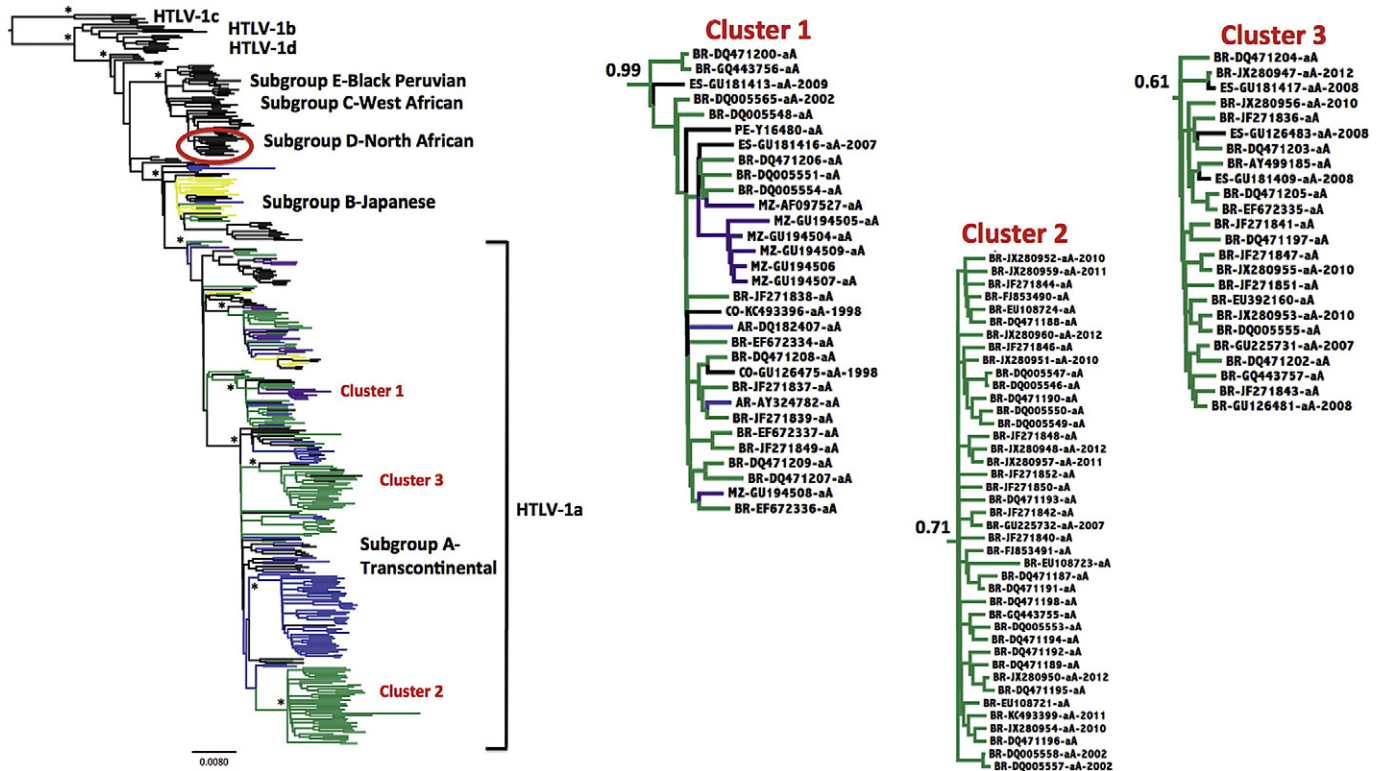
individuals in a rural area of northern KwaZulu-Natal, South Africa, close to the Mozambican border. Rego and colleagues previously described the genotyping of these seven isolates (Rego et al., 2013). The 390 LTR HTLV-1 sequences were aligned against the seven South African isolates in ClustalW and manually edited, as previously described. Due to the shorter length of the seven South African sequences, the alignment was reduced to 468 base pairs (bp). A Bayesian tree topology was inferred from this new alignment in MrBayes under the same models and parameters as previously described. The resulting tree topology was visually inspected to determine the association between Brazilian and South African isolates and to verify the clustering of Brazilian clades. Molecular clock and coalescent analysis of these Brazilian clusters were performed as previously described to validate our previous estimates.

### 3. Results

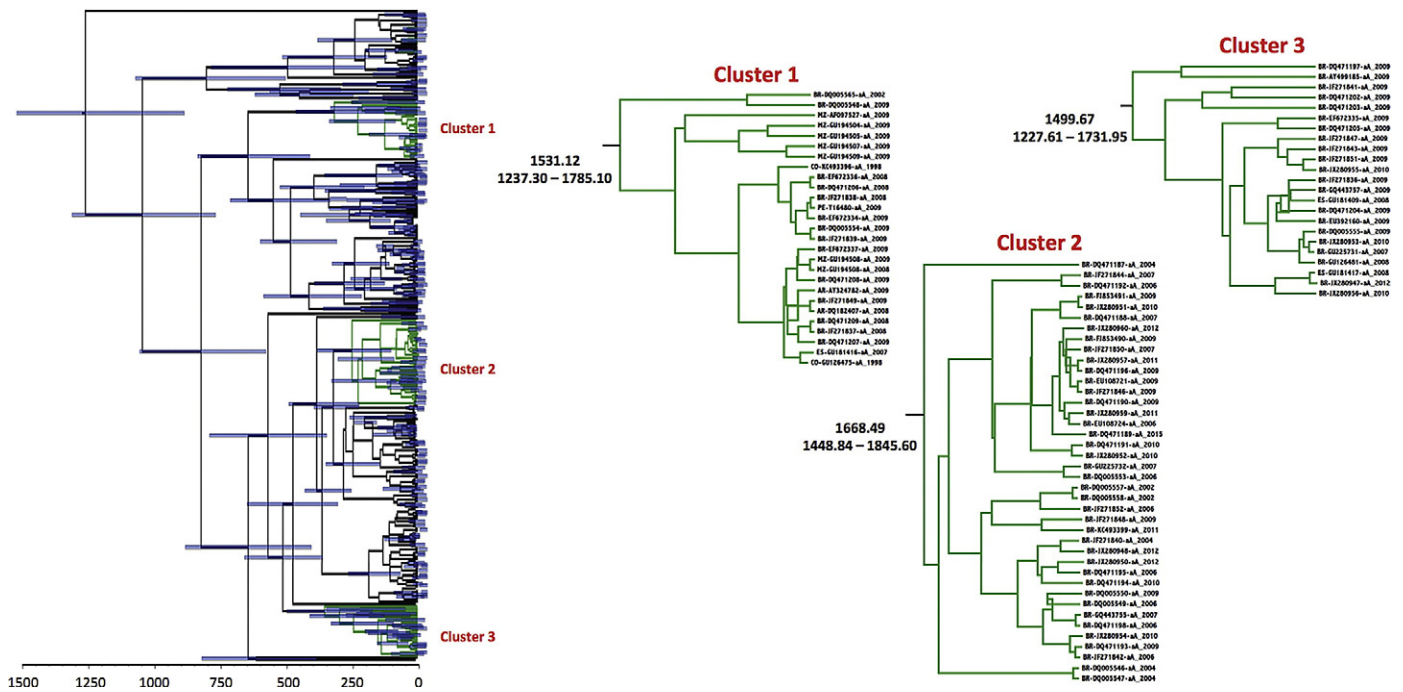
#### 3.1. Recombination and phylogenetic inference

Our analysis in RDP4 could not identify any signs of viral recombination amongst the 390 LTR sequences that were retrieved from the HTLV-1 sequence database. Phylogenetic inference in a maximum-likelihood framework produced a phylogeny consistent with clustering observed in previous studies (Alcantara et al., 2003). However, the branch supports for splits in this tree topology were largely unresponsive (bootstrap < 70), with the exception being branches for major subtypes (i.e. cosmopolitan subtype HTLV-1a). The overall low bootstrap support in this dataset can be attributed to the slow mutation rate of the HTLV-1 virus, which has resulted in a very homogenous viral population within major subtypes.

The removal of non-transcontinental sequences from this tree topology with the ape package and the subsequent analysis in Path-O-gen revealed that the dataset carried enough clocklike signal. The slope of the



**Fig. 1.** Tree topology of 390 LTR sequences of HTLV-1 isolates. The sequence alignment covered 552-bp and the tree topology was inferred under a GTR substitution model with gamma site heterogeneity in MrBayes. Support values for branches with a posterior support > 0.6 are marked with an asterisk. All Brazilian sequences are marked green, Japanese sequences yellow, Mozambican purple and Argentinian in blue. The genetic distance is shown at the bottom.



**Fig. 2.** Time scaled tree topology of the 290 sequences in the transcontinental clade of HTLV-1. The tree was inferred through sampling in BEAST under a Tamura-Nei model of nucleotide substitution. A strict molecular clock was used along all branches with a fix mutation rate of  $2.0 \times 10^{-5}$  substitutions/site/years and a constant population size tree prior. On the right hand side are the three LTR clusters with their estimated time of origin. On the left hand side is a big transcontinental Bayesian time resolved phylogenetic tree with 111 LTR Brazilian sequences.

regression analysis was  $3.67 \times 10^{-5}$  substitutions/site/year, which broadly correlates with the known mutation rate for the LTR region of HTLV-1. The x-intercept for the regression analysis was 1188.268, while the R squared ( $R^2$ ) for the Path-O-gen analysis was only  $3.9 \times 10^{-3}$ . The low  $R^2$  for this dataset can be attributed to the slow mutation rate of HTLV-1, in comparison with other viruses (e.g. HIV).

The Bayesian phylogenetic tree of the 390 LTR sequences was consistent with the clustering pattern observed in the ML-tree topology. For the epidemic reconstruction, three Brazilian clusters were identified (Cluster 1, 2 and 3). All the Brazilian isolates clustered within the main transcontinental subgroup A of HTLV-1 (Fig. 1), together with sequences from Mozambique and other South American sequences (e.g. Argentina, Colombia, or Peru). Only five of the Brazilian isolates clustered with Japanese sequences (Fig. 1). Cluster 1 (posterior probability = 0.99) included 28 sequences; 15 from Brazil, seven from Mozambique, two from Colombia, two from Argentina, one from Peru and one from Spain. Cluster 2 (posterior probability = 0.71) included only Brazilian sequences ( $n = 41$ ), while Cluster 3 (posterior probability = 0.61) included 24 sequences, 21 from Brazil, and three from Spain. Descriptions of accession numbers contained in each cluster are detailed in Supplementary materials.

### 3.2. Phylodynamic reconstruction of the Brazilian epidemic

Fig. 2 shows a time resolved phylogenetic tree, which was constructed for the entire transcontinental clade (HTLV-1a) in the BEAST package. The clustering of the three Brazilian clades in this time resolved phylogenetic tree was consistent with the clustering in the Bayesian tree topology that was constructed in MrBayes (Fig. 1). The estimated tMRC for each of the Brazilian clusters as well as for the entire transcontinental group are summarized in Table 1. The estimated date of origin for Cluster 1 was around the year 1531.12 CE (95% HPD: 1237.30–1785.10) and includes sequences from Brazil, Argentine and Mozambique. The estimated tMRC for Cluster 2, which contained only Brazilian isolates, was around the year 1668.49 CE (95% HPD: 1448.84–1845.60) Finally, the origin of Cluster 3 dated back to the year 1499.67 CE (95% HPD: 1227.61–1731.95).

### 3.3. Phylodynamic relationship between southern African and Brazilian isolates

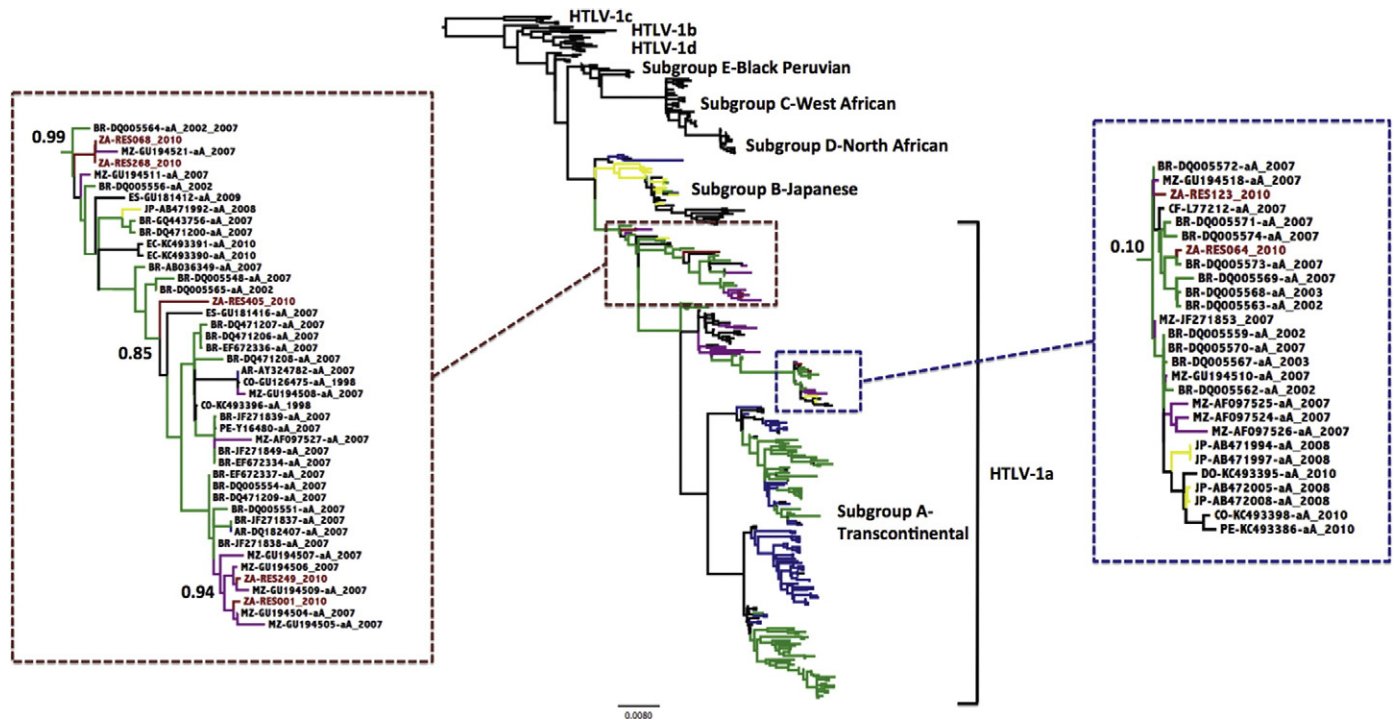
The phylogenetic analysis of the dataset, enriched by the seven sequences from South Africa (Genbank accession numbers: KF042345–KF042351), identified these seven isolates as belonging to the cosmopolitan subtype transcontinental subgroup of HTLV-1a. The Bayesian tree topology analysis showed that the seven South African sequences grouped into two different monophyletic clusters in the cosmopolitan subtype transcontinental subgroup cluster. Five sequences grouped with Mozambican and South American sequences (PP = 0.99) and the remaining two sequences with South American and Mozambican sequences (PP = 0.10). Within these two clades, a close clustering between the seven southern African sequences and sequences from Mozambique and Brazil was further identified (Fig. 3). The phylogenetic analysis showed that the clade which contained seven sequences from South Africa was formed with Brazilian sequences isolated in Salvador, Bahia/Molecular clock analysis of these two clusters produced estimated dates of origin that were slightly more recent but still within the 95% HPD confidence intervals of our previous analysis. The posterior support for the splits in this tree was dramatically lower than the previous tree

**Table 1**

Estimated dates of origin of each of the identified clusters.

	Models parameters	tMRC	95% HPD interval	ESS
Cluster 1	strict.const.fix.01	1531.12	1237.30–1785.10	9001
	strict.const.fix.02	1529.61	1236.28–1784.28	9001
Cluster 2	strict.const.fix.01	1668.49	1448.84–1845.60	8635.44
	strict.const.fix.02	1668.27	1456.32–1842.79	8300.93
Cluster 3	strict.const.fix.01	1499.67	1227.61–1731.95	9001
	strict.const.fix.02	1498.10	1232.03–1727.51	9001
Transcontinental	strict.const.fix.01	728.05	306.61–1086.73	9001
	strict.const.fix.02	732.82	323.97–1090.51	9001

tMRC – time to the most recent common ancestor; HPD – highest posterior density; ESS – Effective Sample Size; fix – fixed mutation rate; strict – strict molecular clock assumption; const – constant population size tree prior.



**Fig. 3.** Tree topology of 397 LTR sequences of HTLV-1 isolates, including the seven samples from South Africa. The sequence alignment covered 468-bp and the tree topology was inferred under a GTR substitution model with gamma site heterogeneity in MrBayes. Support values for branches with a posterior support >0.6 are marked with an asterisk. All Brazilian sequences are marked green, Japanese sequences yellow, Mozambican purple and Argentinian in blue. The genetic distance is shown at the bottom.

presented in Fig. 1, due to the shorter length of the alignment (468 bp vs. 552 bp).

#### 4. Discussion

The phylogenetic inference of the LTR sequences retrieved from the database (<http://htlv1db.bahia.fiocruz.br>) was performed to investigate the time of introduction of the transcontinental subtype of HTLV-1 in Brazil using Bayesian inference. Our results support the hypothesis of an African origin of the HTLV-1 epidemic in Brazil, in particular in Salvador, Bahia, which was closely linked to the slave trade era between the 16th to 20th centuries. The dispersal of Brazilian sequences throughout the transcontinental clade, and the existence of large Brazilian clusters in the Bayesian tree topology (Fig. 1), confirm the hypothesis of multiple introductions of HTLV-1aA into Brazil. The estimated tMRCAs of all of the clusters we identified was around the year 1500 CE, which suggests that the bulk of introductions of HTLV-1aA occurred during the post-Columbian period (Table 1).

Considering that the Brazilian population is composed of at least of three ancestral population groups (European, African and Amerindian), and that the HTLV-1aA subgroup is found in many parts of the world, we suggest that the multiple introduction of HTLV into Brazil is the result of different sources of immigrant populations, primarily from Africa. Approximately 80% of the population of Salvador have ethnic African ancestry. This is reflected in the language, religion and customs of individuals from Salvador (Alcantara et al., 2003). Moreover, a close relationship between Brazilian and southern African isolates was previously described (Alcantara et al., 2006; Mota et al., 2007). Our analysis was also able to show one cluster (PP > 0.99) in which Brazilian and Mozambique sequences appear closely related. This relationship seems to confirm the hypothesis that southern Africans were brought to Brazil as part of the slave trade during the 16th and 19th centuries (Alcantara et al., 2003; Rego et al., 2008).

According to historical reports, HTLV-1 emerged in Central Africa and then spread throughout the world (Van Dooren et al., 2001).

Therefore, Africa is host to most of the genetic variability of HTLV-1 (Van Dooren et al., 2001). Some studies have already confirmed the presence of HTLV-1aA in Maputo (Mozambique), with a prevalence of 4.5%, amongst HIV co-infected patients (Bhatt et al., 2009; Vicente et al., 2011). The genetic relationship between South African and Mozambican HTLV-1aA strains suggests a common origin of HTLV-1 in these two countries. Bantu populations in South Africa are ethnically related to populations in current day southern Mozambique, where the port of Maputo is located (Vicente et al., 2011). The characterization of HTLV-1aA as a prevalent subgroup deriving from Mozambique, suggests that perhaps the transcontinental subgroup could also have been introduced to Latin America independently, mainly from Mozambique to Brazil. This is supported by the Portuguese colonial history and the use of Portuguese as an official language in Mozambique.

Human genetic testing in Salvador Bahia, has reported a large presence (26.5%) of the BEN/CAR  $\beta$ -globin haplotype amongst individuals of African heritage. This haplotype of the  $\beta$ -globin gene is prevalent amongst Bantu speakers from the southern African region (Alcantara et al., 2006; Rego et al., 2008). Moreover, viral sequences derived from patients that were part of the human ancestry study from Salvador, Bahia (Alcantara et al., 2006; Rego et al., 2008) was present in all three of the clusters. Our results presented here confirm previous findings that suggested a post-Columbian African origin of this HTLV in Salvador, Bahia (Alcantara et al., 2003).

Portugal had three main African colonies from where slaves could be sourced. These are the present day countries of Mozambique, Angola and Guinea Bissau. In Angola and Mozambique the HTLV-1aA variant is prevalent (Van Dooren et al., 2001; Bhatt et al., 2009; Vicente et al., 2011), while in Guinea Bissau, the HTLV-1aD variant is the most prevalent (Fig. 1). HTLV-1aD variants from Guinea Bissau cluster in a monophyletic clade along with other West African strains from Senegal (red circle Fig. 1). It is curious to note that there is no genetic relationship between Brazilian and Guinea Bissau HTLV variants, either in the tree topologies we inferred in the present study or in any previous reports

(van Tienen et al., 2012). This could imply that slaves from Guinea Bissau were exported to other slave colonies in North America and the Caribbean in preference to colonies in South America. However, we can negate this hypothesis as official reports indicate that the majority of Africans brought to Bahia (Brazil) during the slave trade came from West African ports (Curtin, 1969; Verger, 1976). The absence of West African HTLV-1 strains in Salvador, Bahia, and the close genetic relationship between viral isolates from southern Africa and isolates from Salvador, Bahia, suggests that the bulk of HTLV-1aA introductions came from slaves originating from ports in the southern African region (e.g. in Angola, Mozambique and the Indian ocean).

Apart from the association between Brazilian and African sequences, a close relationship between Brazilian and Argentinian isolates was also observed (Fig. 1). Brazil and Argentina are neighbouring countries in Latin America and several studies have already found HTLV-1aA to be endemic in Argentina amongst Amerindians (Berini et al., 2013; Gastaldello et al., 2008; Pineda et al., 2015). These studies suggested that HTLV-1aA in Argentina spread from the Amerindians of South America to other ethnic groups within the country (Gastaldello et al., 2008). Our analysis showed at least one cluster containing Brazilian and Argentinian sequences (Cluster 1), but the estimated date of the cluster correlated with the trans-Atlantic slave trade period, which would suggest a much more recent transmission between the two countries (Fig. 1).

Considering the ancestral human migrations, one hypothesis suggests that the introduction of HTLV-1aA in Latin America would have happened following the migration of the Mongoloid population of Asia to the New World via the Bering Strait at the last glaciation period, approximately 10,000–15,000 years ago (Paiva and Casseb, 2015). The evidence that reinforce this theory is the identification of an HTLV-1 isolate in an Andean mummy dated approximately 1500 years old. This led to the conclusion that the mummy strain was carried into the New World during the ancestral migrations of humans from Asia across the Bering Strait (Li et al., 1999). Our phylogenetic and molecular clock analysis do not supported the idea that HTLV-1aA was introduced to the new world through the land bridge of Beringia, as was suggested by Miura et al. (1994) & Yamashita et al. (1998), but rather through the slave trade that occurred during the colonial era. Therefore, based on these conflicting reports, it is possible that HTLV-1 may have been introduced multiple times over a much wider timescale than previously thought.

Additionally, a close association between three Spanish sequences and a Brazilian cluster was observed (Fig. 1). Historical data reports that Salvador also has a higher Spanish ancestry than others states of Brazil, especially immigrants from the Galicia region of Spain (Bacelar, 1994; Leira, 2002). This hypothesis supports the fact that Spanish immigrants could also have participated in the introduction of the HTLV-1aA into Brazil. It is important to note that HTLV-1aA is the subtype also characterized amongst Spain infected individuals (Vallejo et al., 2003; Treviño et al., 2014). However, over the past 10 years, Spain has received a considerable flow of immigrants from South American countries, who have to the ethnic and cultural heterogeneity of the population. Many of the foreigners come from Brazil, with which Spain shares historic links, and to a lesser extent from Sub-Saharan Africa (Pádua et al., 2011). Likewise, we can be suggested that the immigrants from South America who were infected with HTLV-1aA introduced the virus in Spain. The route of introduction of the virus between Brazil and Spain should be investigated, because of the grouping in the same cluster of Brazilians and Spain sequences observed in our study. In addition, haplotype analysis of the infected individuals from Spain should be carried out to clarify the clustering of the Brazilian and Spanish sequences.

Finally, our data also confirms the hypothesis that the migrations of Japanese migrants to Latin America during the first half of the 20th century also brought HTLV-1 to Brazil (Van Dooren et al., 1998). This is evident in the association between Japanese and Brazilian sequences

within the transcontinental clade. This association has already been highlighted by several studies and the HTLV-1 virus was described for the first time in Brazil amongst the Japanese immigrant population (Kitagawa et al., 1986; Alcantara et al., 2003).

In conclusion, in this highly admixed population, our results seem to support the hypothesis of multiple introductions of HTLV-1aA into Brazil mainly via and during the Africans slave trade. Our results support the hypothesis that southern African Bantu populations infected with HTLV-1aA participated in the introduction of this subgroup into Latin America, which was facilitated through the trans-Atlantic slave trade. However, because of the high number of migrations from Africa, the investigation of HTLV-1 isolates and haplotypes from other Bantu-speaking groups (Central and West Africa) should be conducted to increase knowledge about the origin of Brazilian Afro descendants. Likewise, the origin of the HTLV-1aA in Spain should be investigated because of the close relationship observed in our study between Brazilian and Spanish strains. Therefore, it is important to continue to study the migratory routes of infected individuals, and understand how the virus spread around the world not only for the evolutionary history of the virus, but also in order to curb the spread of the virus.

### Acknowledgements

This work was supported by CNPq (200328/2015-8) Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil. E.W. and T.d.O are funded by a Medical Research Council flagship grant of the Republic of South Africa (MRC-RFA-UFSP-01-2013/UKZN HIVEPI).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.meegid.2016.12.005>.

### References

- Alcantara Jr., L.C., Van Dooren, S., Gonçalves, M.S., Kashima, S., Costa, M.C.R., Santos, F.L.N., Bittencourt, A.A.L., Dourado, I., Filho, A.A., Covas, D.T., Vandamme, A.M., Galvão-Castro, B., 2003. Globin haplotypes of human T-cell lymphotropic virus type I-infected individuals in Salvador, Bahia, Brazil, suggest a post-Columbian African origin of this virus. *J. Acquir. Immune Defic. Syndr.* 33, 536–542.
- Alcantara, L.C., de Oliveira, T., Gordon, M., Pybus, O., Mascarenhas, R.E., Seixas, M.O., Gonçalves, M., Hlela, C., Cassol, S., Galvão-Castro, B., 2006. Tracing the origin of Brazilian HTLV-1 as determined by analysis of host and viral genes. *AIDS* 20, 780–782.
- Aleluia, M.M., Mello, M.A.G., Alcântara, L.C.J., Rego, F.F.A., de Souza Santos, L.P., Galvão-Castro, B., de Souza Gonçalves, M., de Oliveira, T., Marin, L.J., Sousa, S.M.B., Gadelha, S.R., 2015. The origin of HTLV-1 in southern Bahia by phylogenetic, mtDNA and  $\beta$ -globin analysis. *Virology* 5, 63–74.
- Anderson, C.W., Ray, W.D., 1975. Improved maximum likelihood estimators for the gamma distribution. *Commun. Stat.* 4, 437–448.
- Araújo, T.H.A., Souza-Brito, L.L., Libin, P., Deforche, K., Edwards, D., de Albuquerque-Junior, A.E., Vandamme, A.-M., Galvão-Castro, B., Alcantara, L.C.J., 2012. A Public HTLV-1 Molecular Epidemiology Database for Sequence Management and Data Mining. *PLoS ONE* 7, e42123.
- Bacelar, J.A., 1994. *Galegos no paraíso racial*. Centro Editorial e Didático.
- Bandeira, L.M., Uehara, S.N.O., Asato, M.A., Agüena, G.S., Maedo, C.M., Benites, N.H., Puga, M.A.M., Rezende, G.R., Finotti, C.M., Cesar, G.A., Tanaka, T.S.O., Castro, V.O.L., Otsuki, K., Vicente, A.C.P., Fernandes, C.E., Motta-Castro, A.R.C., 2015. High prevalence of HTLV-1 infection among Japanese immigrants in non-endemic area of Brazil. *PLoS Negl. Trop. Dis.* 9, e0003691.
- Berini, C.A., Delfino, C., Torres, O., García, G., Espejo, R., Pianciola, L., Juárez, M., Arribere, G., Nadal, M., Eirin, M.E., Biglione, M.M., 2013. HTLV-1 cosmopolitan and HTLV-2 subtype b among pregnant women of non-endemic areas of Argentina. *Sex. Transm. Infect.* 89, 333–335.
- Bhatt, N.B., Gudo, E.S., Sema, C., Bila, D., Di Mattei, P., Augusto, O., Garsia, R., Jani, I.V., 2009. Loss of correlation between HIV viral load and CD4+ T-cell counts in HIV/HTLV-1 co-infection in treatment naive Mozambican patients. *Int. J. STD AIDS* 20, 863–868.
- Chen, J., Zekeng, L., Yamashita, M., Takehisa, J., Miura, T., Ido, E., Mboudjeka, I., Tsague, J.M., Hayami, M., Kaptue, L., 1995. HTLV type I isolated from a pygmy in Cameroon is related to but distinct from the known central African type. *AIDS Res. Hum. Retrovir.* 11, 1529–1531.
- Curtin, P.D., 1969. *The Slave Atlantic Trade: A Census*. The University of Wisconsin Press, Milwaukee.
- Desrames, A., Cassar, O., Gout, O., Hermine, O., Taylor, G.P., Afonso, P.V., Gessain, A., 2014. Northern African strains of human T-lymphotropic virus type 1 arose from a recombination event. *J. Virol.* 88, 9782–9788.

- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Gallo, R.C., 2011. Research and discovery of the first human cancer virus, HTLV-1. *Best Pract. Res. Clin. Haematol.* 24, 559–565.
- Gastaldello, R., Iñiguez, A.M., Otsuki, K., Lamas, G., Balangero, M., Barbas, M.G., Mangano, A., Sen, L., Maturano, E., Remondegui, C., Vicente, A.C.P., Gallego, S., 2008. HTLV type 1 genetic types among native descendants in Argentina. *AIDS Res. Hum. Retrovir.* 24, 1139–1146.
- Gessain, A., Barin, F., Vernant, J.C., Gout, O., Maurs, L., Calender, A., de Thé, G., 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* 2, 407–410 (Lond. Engl.).
- Gessain, A., Cassar, O., 2012. Epidemiological Aspects and world distribution of HTLV-1 infection. *Front. Microbiol.* 3, 388.
- Gessian, A., Yanagihara, R., Franchini, G., Garruto, R.M., Jenkins, C.L., Ajdukiewicz, A.B., Gallo, R.C., Gajdusek, D.C., 1991. Highly divergent molecular variants of human T-lymphotropic virus type I from isolated populations in Papua New Guinea and the Solomon Islands. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7694–7698.
- Gonçalves, D.U., Proietti, F.A., Ribas, J.G.R., Araujo, M.G., Pinheiro, S.R., Guedes, A.C., Carneiro-Proietti, A.B.F., 2010. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. *Clin. Microbiol. Rev.* 23, 577–589.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755 Oxf. Engl.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649 Oxf. Engl.
- Kitagawa, T., Fujishita, M., Taguchi, H., Miyoshi, I., Tadokoro, H., 1986. Antibodies to HTLV-1 in Japanese immigrants in Brazil. *JAMA* 256, 2342.
- Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20, 86–93.
- Leira, L.P., 2002. Galegos na Bahia de todos os Santos. *Galicia en El Mundo Edicións. Editora Graciela Alba burgos, Spain*, pp. 5–12.
- Li, H.C., Fujiyoshi, T., Lou, H., Yashiki, S., Sonoda, S., Cartier, L., Nunez, L., Munoz, I., Horai, S., Tajima, K., 1999. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. *Nat. Med.* 5, 1428–1432.
- Mahieux, R., Ibrahim, F., Mauclore, P., Herve, V., Michel, P., Tekai, F., Chappey, C., Garin, B., Van Der Ryst, E., Guillemin, B., Ledru, E., Delaporte, E., de The, G., Gessain, A., 1997. Molecular epidemiology of 58 new African human T-cell leukemia virus type 1 (HTLV-1) strains: identification of a new and distinct HTLV-1 molecular subtype in Central Africa and in Pygmies. *J. Virol.* 71, 1317–1333.
- Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1, vev003.
- Magalhães, T., Mota-Miranda, A.C., Alcantara, L.C.J., Olavarria, V., Galvão-Castro, B., Rios-Grassi, M.F., 2008. Phylogenetic and molecular analysis of HTLV-1 isolates from a medium sized town in northern of Brazil: tracing a common origin of the virus from the most endemic city in the country. *J. Med. Virol.* 80, 2040–2045.
- Miura, T., Fukunaga, T., Igarashi, T., Yamashita, M., Ido, E., Funahashi, S., Ishida, T., Washio, K., Ueda, S., Hashimoto, K., 1994. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc. Natl. Acad. Sci. U. S. A.* 91, 1124–1127.
- Mota, A.C.A., Van Dooren, S., Fernandes, F.M., Pereira, S.A., Queiroz, A.T.L., Gallazzi, V.O., Vandamme, A.-M., Galvão-Castro, B., Alcantara, L.C.J., 2007. The close relationship between South African and Latin American HTLV type 1 strains corroborated in a molecular epidemiological study of the HTLV type 1 isolates from a blood donor cohort. *AIDS Res. Hum. Retrovir.* 23, 503–507.
- Pádua, E., Rodés, B., Pérez-Piñar, T., Silva, A.F., Jiménez, V., Ferreira, F., Toro, C., 2011. Molecular characterization of human T cell leukemia virus type 1 subtypes in a group of infected individuals diagnosed in Portugal and Spain. *AIDS Res. Hum. Retrovir.* 27, 317–322.
- Paiva, A., Casseb, J., 2015. Origine and prevalence of Human Lymphotropic Virus Type 1 (HTLV-1) and type 2 (HTLV-2) among indigenous populations in the Americas. *Rev. Inst. Med. Trop* 57, 1–13.
- Pineda, M., Bouzas, M.B., Golemba, M., Remesar, M., Aulicino, P., Mammana, L., Tadey, L., Sen, L., Mangano, A., 2015. Molecular epidemiology of HTLV-1 in Argentina. *Retrovirology* 12, 85.
- Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., Gallo, R.C., 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* 77, 7415–7419.
- Proietti, F.A., Carneiro-Proietti, A.B.F., Catalan-Proaires, B.C., Murphy, E.L., 2005. Global epidemiology of HTLV-1 infection and associated diseases. *Oncogene* 24, 6058–6068.
- Rego, F.F., Alcantara, L.C.J., Moura Neto, J.P., Miranda, A.C. O. de S., P., M. de S., G., Galvão-Castro, B., 2008. HTLV type 1 molecular study in Brazilian villages with African characteristics giving support to the post-Columbian introduction hypothesis. *AIDS Res. Hum. Retrovir.* 24, 673–677.
- Rego, F.F.A., Galvão-Castro, B., Alcantara, L.C.J., de Oliveira, T., 2013. HIV and HTLV-1 Co-infection Effect on CD4 + Level of Patients Failing Anti-retroviral Therapy in Rural KwaZulu-Natal (Unpublished).
- Rodríguez, F., Oliver, J.L., Marín, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142, 485–501.
- Salemi, M., Van Dooren, S., Audenaert, E., Delaporte, E., Goubau, P., Desmyter, J., Vandamme, A.M., 1998. Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have closest relatives among African STLV-1 strains. *Virology* 246, 277–287.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 Oxf. Engl.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Treviño, A., Alcantara, L.C., Benito, R., Caballero, E., Aguilera, A., Ramos, J.M., de Mendoza, C., Rodríguez, C., García, J., Rodríguez-Iglesias, M., Ortiz de Lejarazu, R., Roc, L., Parra, P., Eiros, J., del Romero, J., Soriano, V., on behalf of the HTLV Spanish Study Group, 2014. Molecular epidemiology and clinical features of human T cell lymphotropic virus type 1 infection in Spain. *AIDS Res. Hum. Retrovir.* 30, 856–862.
- Vallejo, A., Capote, F.J., Guisado, F., Leal, M., Calderón, E., 2003. Cosmopolitan HTLV-1a subtype among Spanish native patients. *AIDS Res. Hum. Retrovir.* 19, 609–611.
- Vandamme, A.M., Liu, H.F., Goubau, P., Desmyter, J., 1994. Primate T-lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-1. *Virology* 202, 212–223.
- Van Dooren, S., Gotuzzo, E., Salemi, M., Watts, D., Audenaert, E., Duwe, S., Ellerbrok, H., Grassmann, R., Hagelberg, E., Desmyter, J., Vandamme, A.M., 1998. Evidence for a post-Columbian introduction of human T-cell lymphotropic virus [type I] [corrected] in Latin America. *J. Gen. Virol.* 79, 2695–2708.
- Van Dooren, S., Salemi, M., Vandamme, A.M., 2001. Dating the origin of the African human T-cell lymphotropic virus type-i (HTLV-1) subtypes. *Mol. Biol. Evol.* 18, 661–671.
- Van Tienen, C., de Silva, T.L., Alcantara, L.C.J., Onyango, C.O., Jarju, S., Gonçalves, N., Vincent, T., Aaby, P., Whittle, H., van der Loeff, M.S., Cotten, M., 2012. Molecular epidemiology of endemic human T-lymphotropic virus type 1 in a rural community in Guinea-Bissau. *PLoS Negl. Trop. Dis.* 6, e1690.
- Verger, P., 1976. Mouvement des navires. P.12. In: Verger, P. (Ed.), *Trade Relations Between the Bight of Benin and Bahia, 17th to 19th century*. Trans. Ibadan University Press, Ibadan, pp. 24–26 Evelyn Crawford.
- Vicente, A.C.P., Gudo, E.S., Iñiguez, A.M., Otsuki, K., Bhatt, N., Abreu, C.M., Vubil, A., Bila, D., Ferreira Jr., O.C., Tanuri, A., Jani, I.V., for the HTLV in Mozambique Study Group, 2011. Genetic characterization of human T-cell lymphotropic virus type 1 in Mozambique: transcontinental lineages drive the HTLV-1 endemic. *PLoS Negl. Trop. Dis.* 5, e1038.
- Wolfe, N.D., Heneine, W., Carr, J.K., Garcia, A.D., Shanmugam, V., Tamoufe, U., Torimiro, J.N., Prosser, A.T., LeBreton, M., Mpoudi-Ngole, E., McCutchan, F.E., Birx, D.L., Folks, T.M., Burke, D.S., Switzer, W.M., 2005. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc. Natl. Acad. Sci.* 102, 7994–7999.
- Yamashita, M., Picchio, G., Veronesi, R., Ohkura, S., Bare, P., Hayami, M., 1998. HTLV-1s in Argentina are phylogenetically similar to those of other South American countries, but different from HTLV-1s in Africa. *J. Med. Virol.* 55, 152–160.
- Yoshida, M., Miyoshi, I., Hinuma, Y., 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2031–2035.
- Zehender, G., Ebranati, E., De Maddalena, C., Gianelli, E., Riva, A., Rusconi, S., Massetto, B., Rankin, F., Acurie, M., Galli, M., 2008. Description of a “trans-Saharan” strain of human T-lymphotropic virus type 1 in West Africa. *J. Acquir. Immune Defic. Syndr.* 47, 269–273.