

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

E box motifs as mediators of proviral latency of human retroviruses

Retrovirology 2009, **6**:81 doi:10.1186/1742-4690-6-81

Jean-Michel Terme (jeanmichel.term@ibmb.csic.es)
Sebastien Calvignac (sebastien.calvignac@univ-lyon1.fr)
Madeleine Duc Dodon (madeleine.duc.dodon@ens-lyon.fr)
Louis Gazzolo (louis.gazzolo@ens-lyon.fr)
Albert Jordan (ajvm@bc.crg.es)

ISSN 1742-4690

Article type Commentary

Submission date 29 April 2009

Acceptance date 16 September 2009

Publication date 16 September 2009

Article URL <http://www.retrovirology.com/content/6/1/81>

This peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Retrovirology* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Retrovirology* or any BioMed Central journal, go to

<http://www.retrovirology.com/info/instructions/>

For information about other BioMed Central publications go to

<http://www.biomedcentral.com/>

E box motifs as mediators of proviral latency of human retroviruses

Jean-Michel Terme[£], Sébastien Calvignac[§], Madeleine Duc Dodon^{#*}, Louis Gazzolo[#] and Albert Jordan^{£,*}

[£] Institut de Biologia Molecular Barcelona IBMB-CSIC, Parc Científic de Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain

[§] Université de Lyon, UMR CNRS 5023, Laboratoire d'Ecologie des Hydrosystèmes Fluviaux, Campus de la Doua, 43, Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

[#] INSERM U758 Virologie Humaine, IFR 128 Biosciences Lyon-Gerland, Ecole Normale Supérieure de Lyon, 4, Allée d'Italie, 69364 Lyon Cedex 07, France.

JMT and SC contributed equally to this work.

* Corresponding author

Emails:

JMT : jeanmichel.terme@ibmb.csic.es

AJ : ajvbmc@ibmb.csic.es

SC : sebastien.calvignac@univ-lyon1.fr

MDD : madeleine.duc.dodon@ens-lyon.fr

LG : louis.gazzolo@ens-lyon.fr

Abstract

The palindromic sequence motifs (CANNTG) known as E boxes are considered as binding sites for the basic helix-loop-helix (bHLH) class of DNA-binding proteins. Their presence has been reported in the long terminal repeats (LTR) of the HIV-1 and HTLV-1 proviruses. Their close proximity with the TATA region of both LTRs indicates that the bHLH proteins may act as important regulators of the function of proviral transcription. Indeed, observations on HIV-1 and recent results on HTLV-1 underline that these E boxes may be critically involved in the regulation of the proviral transcription of these human retroviruses. Indeed, of the two E boxes flanking the TATA sequences of the HIV-1 provirus, the 3' E box has been implicated in the transcriptional inhibition of viral gene expression. Such a role might also be played by the unique 5' E box present in the HTLV-1 LTR. In both cases, the expression of tissue-specific bHLH proteins, like TAL1 might counteract the inhibitory effect exerted by E box proteins, thereby increasing proviral transcription. Finally, a phylogenetic study encompassing several subtypes of these two human retroviruses underlines that these E box motifs have recently appeared in the proviral LTRs and may be considered as potential mediators in the establishment of proviral latency.

Introduction

The two prototypic human pathogenic retroviruses HIV-1 (Human Immunodeficiency Virus, type 1) and HTLV-1 (Human T cell Leukaemia Virus, type 1) infect their hosts on a long-term basis that relies on their abilities to infect latently specific cellular subsets, such as memory CD4⁺ T cells. These viruses are thus able to efficiently escape the immune responses as well as the effects of anti-retroviral drugs that are included in highly efficient therapeutic protocols. Consequently, understanding the mechanisms that promote the establishment of latency is critical to the design of future therapies. A promising avenue of investigation is the search for host factors that

would decrease proviral transcription. Hence, by considering data on HIV-1 and recent results on HTLV-1, we discuss lines of evidence showing that bHLH proteins may critically intervene in the proviral transcription of these two human retroviruses. Such an intervention is made possible through the presence of E-boxes in the long terminal repeats (LTRs) of the respective provirus.

Discussion

The consensus hexanucleotide sequence known as the E-box motif (5'-CANNTG-3') represents the core DNA sequence capable of binding the basic helix-loop-helix (bHLH) class of proteins [1, 2]. These proteins contain a basic DNA-binding region juxtaposed to the HLH domain that functions in protein dimerization. bHLH proteins include ubiquitous (class I, e.g. E47) and tissue-specific (class II, e.g. TAL1) transcription factors and play a prominent role in regulatory networks that control a diversity of processes from cell proliferation to cell differentiation. They can form homodimers or heterodimers with class II bHLH proteins and can act as transcriptional activators or repressors through the recruitment of distinct co-activator or co-repressor complexes, respectively [2, 3]. Accordingly, the presence of E boxes in the LTRs of HIV-1 and HTLV-1 underlines the possible intervention of bHLH proteins in regulating proviral transcription.

E box motifs in the LTR of human retroviruses.

Indeed, four E box motifs have been described in the LTR of HIV-1; two are located 11 base pairs (bp) upstream and 6 bp downstream of the TATA sequence [4, 5]. These two palindromic sequence motifs (CAGATG and CAGCTG) have been referred to as the 5' E box and the 3' E box, respectively [4]. More recently, the presence of a unique E box sequence (CATATG) in the LTR of HTLV-1, located 28 bp upstream the TATA box, has been reported [6]. The close proximity of these E boxes to the TATA region appears to be specific to the LTRs of HIV-1 and HTLV-1. In the 5' LTR

of bovine leukemia virus, (an HTLV-1 related retrovirus), three E box motifs overlapping the cyclic AMP responsive elements (CREs) have been shown to be involved in transcriptional repression of BLV basal gene expression [7].

Disruptive mutagenesis experiments have underlined the functional importance of the TATA sequences and the flanking E boxes of the HIV-1 LTR, and more particularly of the 3' E box in regulating basal and Tat-induced gene expression [4]. Thus, it has been observed that one natural clone of HIV-1 carrying two mutated E boxes exhibited a high LTR basal activity in U937 cells [8]. More interestingly, experiments have been performed to characterize the bHLH proteins that bind to these E boxes in order to determine their role in the regulation of proviral transcription. Gel retardation analysis demonstrated that the specific binding of E box proteins (as either E47 homodimers or HEB homodimers or HEB/E47 heterodimers), and AP-4, a bHLH-ZIP protein, to the 3' E box of HIV-1 LTR [4, 5]. Recently, the binding of AP-4 (and possibly other bHLH proteins) to the 3' E box was found to exclude the binding of TATA-binding proteins (TBP) to the TATA box and to inhibit the LTR-mediated transcription of the HIV-1 provirus *in vitro* [4, 9]. Consequently, E boxes in the LTR may account for a modulation of viral replication, and even for the establishment and maintenance of latency in HIV-1 infected cells. The presence of an E box motif in the LTR of HTLV-1 and the observation that E47 was repressing both basal and Tax-induced LTR activity suggest that this bHLH protein indeed favors HTLV-1 proviral latency, possibly by interfering with the binding of TBP to the proximal TATA element [10]. The functional significance of the HTLV-1 putative E box has not been assessed yet. However, the above observations suggest that these E boxes are able to mediate the effect of bHLH proteins on the LTR activity of human retroviruses. Of interest, the overlap of E box elements and CREs in the BLV LTR has been proposed as a strategy to allow better silencing of viral transcription. In this model, suppression of viral gene expression has been shown to contribute not only to the impairment of immune

surveillance, but also to the onset and progression of lymphoid tumours in BLV-infected sheep [11].

Collectively, these observations support that E box motifs in the LTRs of HIV-1 and HTLV-1 might represent important mediators of proviral transcription, by allowing the binding of bHLH proteins that might interfere with the transcriptional complex recruited at the TATA element. Consequently, some of these motifs would contribute to post-integration latency by turning off proviral expression. Furthermore, epigenetic mechanisms could also be implicated in the long-term suppression of viral expression, as E47 is known to bring chromatin-remodeling complex to specific promoters and therefore could induce epigenetic changes in proviral genome [12]. In return, that inhibition could be relieved by the action of tissue-specific class II bHLH proteins, such as TAL1 (T-cell Acute Leukemia 1), providing a way to exit proviral latency (Figure 1). Indeed, the binding of E47/HEB heterodimers to the HIV-1 3' E box is abrogated by the over-expression of a tissue-specific class II bHLH factor, TAL1, a functional inhibitor of E proteins [13]. Similarly, the over-expression of TAL1 is also able to counteract the E47 protein-mediated inhibition of the HTLV-1 LTR [10]. The regulation of proviral transcription during the early steps of T lymphocyte maturation might, therefore, be an important event contributing to the pathogenesis of this retroviral infection [14, 15]. As an alternative to this scenario, one can imagine that the tissue-specificity of class II bHLH proteins may be linked to the differential ability of retroviruses to establish post-integration latency. For example, HIV-1 is known to establish latency in mature, but not in immature thymocytes [16]. Such a difference may be linked to the expression of TAL1 that is restricted to immature thymocytes. Clearly, much more experimental work is needed to assess the functional significance of these E boxes; however, the observations that have been devoted to delineate their intervention in the control of proviral latency are suggestive of their critical contributions.

Evolutionary significance of E box motifs in the LTR of human retroviruses.

As outlined above, the functionality of the E box elements needs further experimental evidence. Nevertheless, one can already address their evolutionary significance through the analysis of their distribution and conservation in LTRs of retroviruses that infect individuals, populations and species. Indeed such an analysis might provide valuable information on their involvement as mediators of proviral latency and on the evolutionary significance of the mechanisms highlighted above. At the scale of one to a few individuals, nearly all HIV-1 quasispecies exhibit intact E boxes (85 to 100% of intact E boxes) [8, 17, 18]. At a larger phylogenetic scale, E box mutants are strictly restricted to HIV-1 subgroups E, F and G (most virological studies have been performed with subtype B isolates) [19]. Interestingly, isolates of subtypes E and G while accounting for less than 9% of all HIV-1 infections were found to display a two- to three-fold higher LTR basal activity than those of subtype B, consistent with a role of the intact E box in decreasing viral expression [19, 20]. Thus, the presence of intact E boxes in the HIV-1 LTR might account for a mechanism that regulates HIV-1 gene expression, as well as proviral latency.

Considering that HTLV-1 replication occurs mainly by clonal expansion of infected cells (rather than by reverse transcription), this retrovirus is expected to lack intra-individual sequence variations. However, variations have been observed at the population level. On the basis of phylogenetic analyses, 6 groups of HTLV-1 (HTLV-1a to f) have been identified [21]. As the early diverged HTLV-1c sequences lack the putative E box [6], the ancestral state of HTLV-1 should be characterized by the absence of a TATA-neighbouring putative E box thus arguing for an active proviral transcription and reverse transcription. Of note, this putative E box (CANNTG) can only be found in the LTR sequences of HTLV-1a and HTLV-1e (as deduced from the analysis of alignments [6]). HTLV-1a, often referred to as the Cosmopolitan clade, represents by far the most widely distributed strain in HTLV-1 infected persons. It is striking to observe that an overwhelming majority (97%) of HTLV-1a isolates harbour this putative E box (as deduced from the analysis of alignments [22]), suggesting a

selection for this motif during evolution (Figure 2). Such an observation might suggest that the presence of an E box might have contributed to the spread of the retrovirus, possibly by favouring the transcriptional repression of viral genes and thus facilitating virus- escape from the host immune system. Indeed, such a strategy might be critical to proviral maintenance via clonal expansion which necessarily precedes leukemia development.

Conclusion

Even if HIV-1 and HTLV-1 have distinct characteristics and strategies to survive in the infected T cells, the capacities of these two human retroviruses to establish a latent infection is considered to be crucial for their pathogenesis. HIV-1 latency has been described to occur at both pre-integration and post-integration levels in infected CD4+ T lymphocytes [23-25]. More specifically, the presence of E box motifs in the HIV-1 LTR has been reported several years ago and raised their implication in the transcriptional regulation of the provirus. The E boxes in the LTR of HIV-1 might represent the binding sites of positively- and negatively-acting bHLH factors that arbitrate between episodes of active viral transcription and silenced gene expression [4]. Under these conditions, it is expected that bursts of viral replication will expose infected cells to the host's immune response, and lead to the gradual depletion of the CD4+ T-cell compartment, favoring immunodeficiency. Concerning HTLV-1, the primary infection is considered as a two-step process involving a transient step of reverse transcription followed by the Tax-induced polyclonal proliferation of infected cells, during genetic instability occurs [26]. Consequently, it is possible that E box proteins (and particularly E47) could contribute to decreased and silenced proviral transcription, that may favor the selection of a restricted number of latently-infected clones that escape the immune response and survive the genomic insults. Thus, proviral silencing might significantly contribute to the initiation of the leukemogenic process, during which the emergence of epigenetic events, such as LTR methylation

[27, 28], would in turn favor the long-term suppression of LTR activity. Finally and more importantly, evolutionary considerations also point out the potential importance of these E box motifs in the LTRs of HIV-1 and HTLV-1. Clearly, they plead for further functional investigation of the mechanisms involved in the recruitment of bHLH proteins to the E boxes and of the importance of these motifs as mediators of proviral latency.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors read and approved the final manuscript.

Acknowledgements

We are grateful to P. Jalinot and A. Calvignac-Spencer for helpful discussions. This work was supported by "Association pour la Recherche sur le Cancer, by the "Comité du Rhône de la Ligue Nationale contre le Cancer", by the Fondation pour la Recherche Médicale and by Fondation de France (Comité Leucémie).

References

1. Massari ME, Murre C: **Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms.** *Mol Cell Biol* 2000, **20**:429-440.
2. Slattery C, Ryan MP, McMorrow T: **E2A proteins: regulators of cell phenotype in normal physiology and disease.** *Int J Biochem Cell Biol* 2008, **40**:1431-1436.
3. Kee BL: **E and ID proteins branch out.** *Nat Rev Immunol* 2009, **9**:175-184.

4. Ou SH, Garcia-Martinez LF, Paulssen EJ, Gaynor RB: **Role of flanking E box motifs in human immunodeficiency virus type 1 TATA element function.** *J Virol* 1994, **68**:7188-7199.
5. Zhang Y, Doyle K, Bina M: **Interactions of HTF4 with E-box motifs in the long terminal repeat of human immunodeficiency virus type 1.** *J Virol* 1992, **66**:5631-5634.
6. Calvignac S, Terme JM, Hensley SM, Jalinot P, Greenwood AD, Hanni C: **Ancient DNA identification of early 20th century simian T-cell leukemia virus type 1.** *Mol Biol Evol* 2008, **25**:1093-1098.
7. Calomme C, Dekoninck A, Nizet S, Adam E, Nguyen TL, Van Den Broeke A, Willems L, Kettmann R, Burny A, Van Lint C: **Overlapping CRE and E box motifs in the enhancer sequences of the bovine leukemia virus 5' long terminal repeat are critical for basal and acetylation-dependent transcriptional activity of the viral promoter: implications for viral latency.** *J Virol* 2004, **78**:13848-13864.
8. Estable MC, Bell B, Merzouki A, Montaner JS, O'Shaughnessy MV, Sadowski IJ: **Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients representing all stages of infection display a wide range of sequence polymorphism and transcription activity.** *J Virol* 1996, **70**:4053-4062.
9. Imai K, Okamoto T: **Transcriptional repression of human immunodeficiency virus type 1 by AP-4.** *J Biol Chem* 2006, **281**:12495-12505.
10. Terme JM, Wencker M, Favre-Bonvin A, Bex F, Gazzolo L, Duc Dodon M, Jalinot P: **Crosstalk between expression of the HTLV-1 Tax transactivator and the oncogenic bHLH transcription factor TAL1.** *J Virol* 2008, **82**:7913-22.
11. Merimi M, Klener P, Szydal M, Cleuter Y, Bagnis C, Kerkhofs P, Burny A, Martiat P, Van den Broeke A: **Complete suppression of viral gene**

- expression is associated with the onset and progression of lymphoid malignancy: observations in Bovine Leukemia Virus-infected sheep. *Retrovirology* 2007, **4**:51.
12. Xu Z, Meng X, Cai Y, Koury MJ, Brandt SJ: **Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors.** *Biochem J* 2006, **399**:297-304.
 13. Doyle K, Zhang Y, Baer R, Bina M: **Distinguishable patterns of protein-DNA interactions involving complexes of basic helix-loop-helix proteins.** *J Biol Chem* 1994, **269**:12099-12105.
 14. Murre C: **Helix-loop-helix proteins and lymphocyte development.** *Nat Immunol* 2005, **6**:1079-1086.
 15. O'Neil J, Shank J, Cusson N, Murre C, Kelliher M: **TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB.** *Cancer Cell* 2004, **5**:587-596.
 16. Brooks DG, Kitchen SG, Kitchen CM, Scripture-Adams DD, Zack JA: **Generation of HIV latency during thymopoiesis.** *Nat Med* 2001, **7**:459-464.
 17. Delassus S, Cheynier R, Wain-Hobson S: **Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro.** *J Virol* 1991, **65**:225-231.
 18. Delassus S, Meyerhans A, Cheynier R, Wain-Hobson S: **Absence of selection of HIV-1 variants in vivo based on transcription/transactivation during progression to AIDS.** *Virology* 1992, **188**:811-818.
 19. Jeeninga RE, Hoogenkamp M, Armand-Ugon M, de Baar M, Verhoef K, Berkhout B: **Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G.** *J Virol* 2000, **74**:3740-3751.
 20. Arien KK, Vanham G, Arts EJ: **Is HIV-1 evolving to a less virulent form in humans?** *Nat Rev Microbiol* 2007, **5**:141-151.

21. Van Dooren S, Verschoor EJ, Fagrouch Z, Vandamme AM: **Phylogeny of primate T lymphotropic virus type 1 (PTLV-1) including various new Asian and African non-human primate strains.** *Infect Genet Evol* 2007, **7**:374-381.
22. Coulthart MB, Posada D, Crandall KA, Dekaban GA: **On the phylogenetic placement of human T cell leukemia virus type 1 sequences associated with an Andean mummy.** *Infect Genet Evol* 2006, **6**:91-96.
23. Zamborlini A, Lehmann-Che J, Clave E, Giron ML, Tobaly-Tapiero J, Roingard P, Emiliani S, Toubert A, de The H, Saib A: **Centrosomal pre-integration latency of HIV-1 in quiescent cells.** *Retrovirology* 2007, **4**:63.
24. Han Y, Wind-Rotolo M, Yang HC, Siliciano JD, Siliciano RF: **Experimental approaches to the study of HIV-1 latency.** *Nat Rev Microbiol* 2007, **5**:95-106.
25. Marcello A: **Latency: the hidden HIV-1 challenge.** *Retrovirology* 2006, **3**:7.
26. Matsuoka M, Jeang KT: **Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation.** *Nat Rev Cancer* 2007, **7**:270-280.
27. Ishida T, Hamano A, Koiwa T, Watanabe T: **5' long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals.** *Retrovirology* 2006, **3**:69.
28. Taniguchi Y, Nosaka K, Yasunaga J, Maeda M, Mueller N, Okayama A, Matsuoka M: **Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms.** *Retrovirology* 2005, **2**:64.

Figure legends

Figure 1. Schematic representation of the proposed regulation of the HTLV LTR by E box proteins. Binding of E box proteins to the adjacent E box disrupts the interaction of TBP with the TATA element and may inhibit basal expression thus favoring proviral latency. In turn, the expression of a bHLH class II protein (like TAL1) should interfere with the binding of bHLH factors to the E box, thus restoring the basal retroviral expression and turning off latency.

Figure 2. Mapping putative E-boxes (CANNTG) onto a global HTLV-1 phylogeny.

HTLV-1 strains have been gathered into 6 groups of common ancestry (clades) named a to e, with HTLV-1a being often referred to as the Cosmopolitan clade [21]. This phylogenetic tree depicts the relationships of 127 strains (appearing as terminal, non-interior branches) among which >100 belong to the Cosmopolitan clade (the only pandemic HTLV-1 clade, solid line polygon). HTLV-1a is here largely overrepresented, which fits the reality of the pandemic; the other group distribution areas being much more restricted (e.g. Central Africa). Putative E-boxes have been mapped onto this phylogeny through the appending of a color code (blue=CANNTG consensus sequence; black=other sequences). This makes evident the presence of putative E-boxes in nearly all Cosmopolitan strains, which sharply contrasts with the situation prevailing in the other clades. This tree has been obtained through the analysis of a dataset derived from Coulthart *et al.*[22].

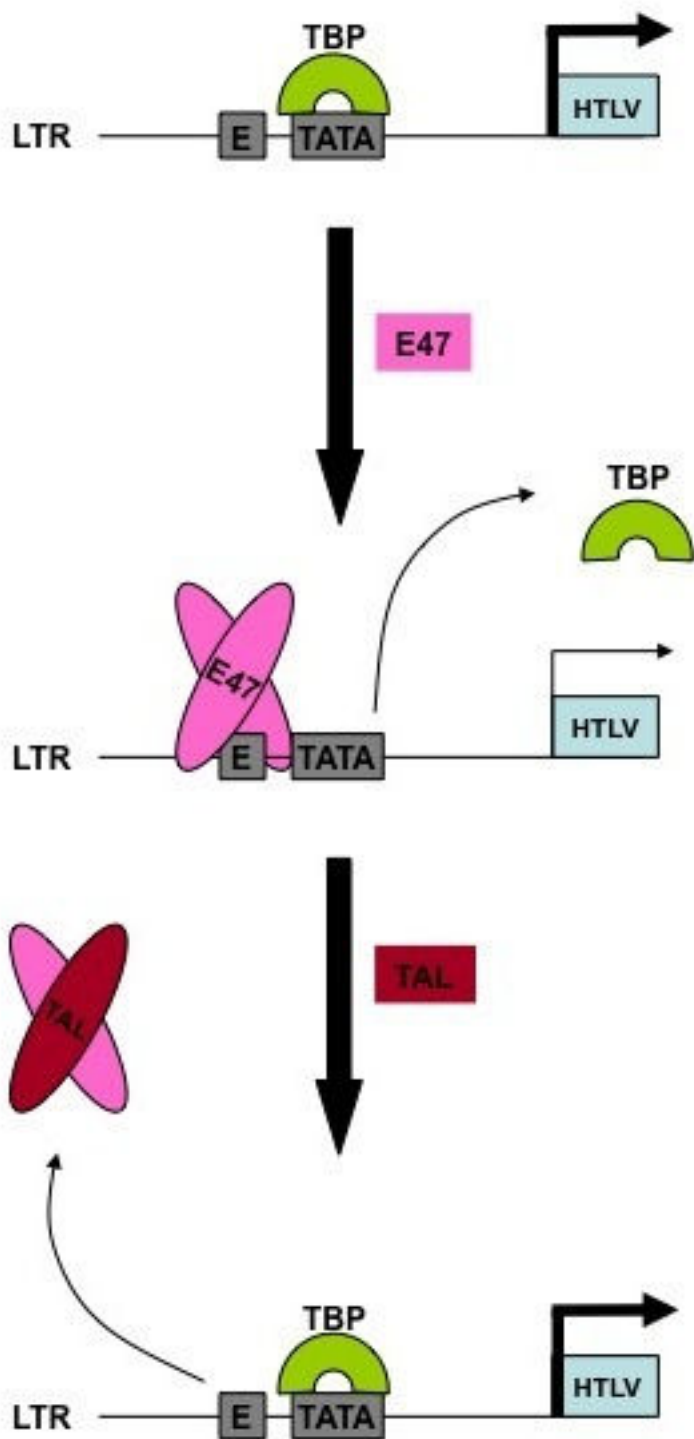


Figure 1

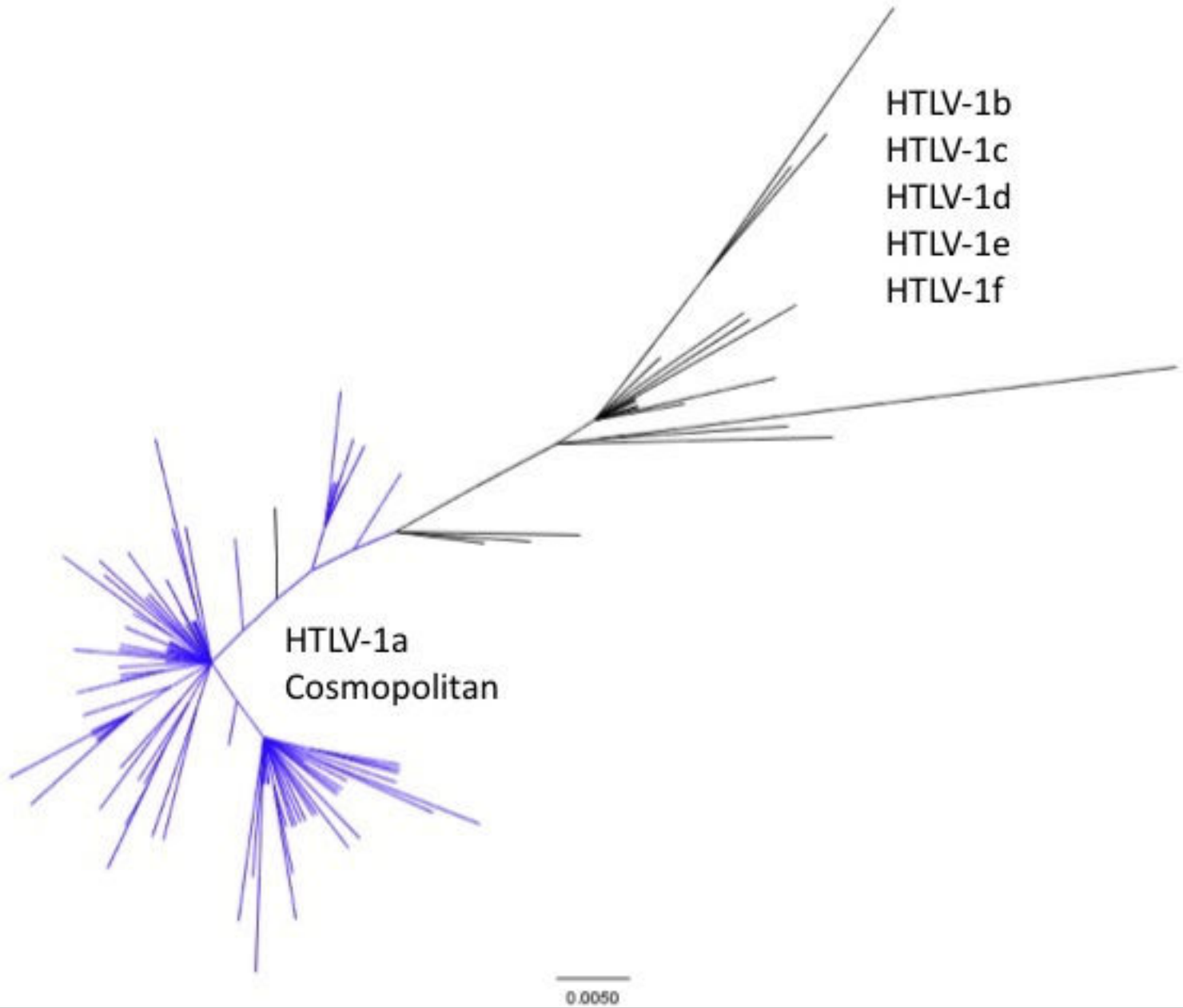


Figure 2