AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 32, Number 3, 2016
© Mary Ann Liebert, Inc.
DOI: 10.1089/aid.2015.0273

Deep Sequencing Analysis of Human T Cell Lymphotropic Virus Type 1 Long Terminal Repeat 5′ Region from Patients with Tropical Spastic Paraparesis/Human T Cell Lymphotropic Virus Type 1-Associated Myelopathy and Asymptomatic Carriers

Filipe Ferreira de Almeida Rego,[1–3] Tulio de Oliveira,[2] Marta Giovanetti,[1]
Bernardo Galvão-Castro,[4] Marilda de Souza Gonçalves,[1] and Luiz Carlos Alcantara[1]

Abstract

The aim of this study was to analyze patients by deep sequencing the human T cell lymphotropic virus type 1 (HTLV-1) long terminal repeat (LTR) region in order to determine if minor and/or major mutations in this promoter region might be associated with tropical spastic paraparesis (TSP)/human T cell lymphotropic virus type 1-associated myelopathy (HAM) outcome or proviral load or HTLV-1 expression. This study is a cross-sectional analyze of 29 HTLV-1-infected patients with TSP/HAM or asymptomatic carriers. Proviral DNA from those subjects was submitted to a nested PCR for the HTLV-1 LTR5′ region. The HTLV-1 LTR5′ purified products were submitted to deep sequencing using the Ion Torrent sequencing technology (Life Technologies, Carlsbad, CA). We found that samples with low proviral load showed more detected minor mutations than the samples with high proviral load. Mutations in 136 positions were found over the 520-bp analyzed fragment of HTLV-1 LTR5′ with at least 1% frequency. Eleven mutations were present in the previously determined major transcription factor binding sites (TFBS) and in more than one patient, indicating that there might be a differential HTLV-1 expression comparing individuals or in comparing different cells from the same individual. Three mutations were statistically significant using the Fisher nonparametric test between the groups but were not present in previously determined TFBS (G126C/T, G306C, and C479T). Those mutations that were not present in previously determined TFBS were statistically significant in this study and were most frequent in patients with low proviral load or in asymptomatic carriers. Although those mutations were not present in previously determined TFBS, one of those mutations (G306C/A) was present in an Sp-1 binding site determined by *in silico* analysis, and its presence abrogated the site for Sp-1 binding and created a new possible ATF binding site.

Human T cell lymphotropic virus type 1 (HTLV-1) is known to be the etiological agent of tropical spastic paraparesis/HTLV-1-associated myelopathy (TSP/HAM),[1,2] adult T cell leukemia/lymphoma (ATL),[3] and other inflammatory diseases such as HTLV-1-associated uveitis (HAU)[4] and infective dermatitis associated with HTLV-1 (IDH).[5]

The development of TSP/HAM in HTLV-1-infected individuals remains unknown, as few infected individuals develop this disease; it is probably related to environmental, host, and viral factors, such as proviral load, the methylation profile of proviral DNA, the integration site of HTLV-1 into the host genome, and viral mutations.[6–8] Recent studies have shown through Sanger sequencing and real-time PCR that mutations in the LTR region may increase HTLV-1 proviral load[9,10] and could consequently affect TSP/HAM outcome.

The HTLV-1 genome encodes the *gag*, *pol*, *env*, and *pX* genes, regulated by the LTR flanking region. The flanking regions are divided into three regions: U3, R, and U5, which

[1]Laboratory of Hematology, Genetic and Computational Biology, Gonçalo Moniz Research Center, Oswaldo Cruz Foundation (LHGB/CPqGM/FIOCRUZ), Salvador, Bahia, Brazil.
[2]Africa Centre for Health and Population Study, University of KwaZulu-Natal, KwaZulu-Natal, South Africa.
[3]Universidade Católica do Salvador, Salvador, Bahia, Brazil.
[4]HTLV Center/Bahia School of Medicine and Public Health, Salvador, Bahia, Brazil.

contain signals to HTLV-1 expression control, such as the TxRE (Tax-responsive elements) region in U3 LTR5′,[11] responsible for proviral genome transcription through the interaction with transcription factors such as CREB/ATF, the TATA box motif, and the poly(A) signal.

The aim of this study was to analyze patients by deep sequencing the HTLV-1 LTR region in order to determine whether minor and/or major mutations in this promoter region might be associated with TSP/HAM outcome, proviral load, or HTLV-1 expression.

To accomplish this a cross-sectional analysis of 29 HTLV-1-infected patients with TSP/HAM or asymptomatic carriers from the HTLV Center, Salvador, Brazil was carried out. Basic demographic and clinical information (age, sex, TSP/HAM status, and proviral load) was obtained from the patient's clinical record. The patients were then classified according to their proviral load (high or low) and TSP/HAM status (present or absent). The Research Ethics Committee of the Oswaldo Cruz Foundation approved the project (Protocol Number 368/2011).

Patient nuclear DNA was extracted from peripheral blood mononuclear cell (PBMC) samples using a QIAGEN QIAamp DNA Blood kit following the manufacturer's protocol. DNA samples were submitted to nested PCR using the HTLV-1 long terminal repeat (LTR) 5′ region primers as described previously.[12] The PCR products were purified using the PureLink PCR Purification kit. HTLV-1 sequencing was performed using the Taq FS Dye Terminator Cycle Sequencing kit (Applied Biosystems) on an automated 3130xl Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA) using the identical nested PCR inner primers in order to analyze the major strains and create the consensus sequence for further deep sequencing analysis.

The HTLV-1 LTR5′ purified products were submitted to deep sequencing using Ion Torrent sequencing technology (Life Technologies, Carlsbad, CA). At least 500 sequences with a minimum of 100 base pairs (bp) were produced for each specimen, which allowed characterization of the viral population at the 1% level. For Ion Torrent analysis the purified PCR products were submitted to enzymatic fragmentation following the manufacturer's conditions for the 100-ng-long amplicon input (Ion Xpress Plus Fragment Library Kit Part No. 4471269, Life Technologies, Carlsbad, CA), with 15 min of incubation time in a heat block followed by Agencourt AMPure XP purification (Beckman Coulter, Brea, CA). The ligation of the Ion Xpress P1 and barcode adapters and nick repair was performed (Ion Xpress Barcode Adapters 1–16 Catalog No. 4471250 and Ion Xpress Barcode Adapters 17–32 Catalog No. 4474009) followed by the second Agencourt purification. The fragmented DNA was then size selected (E-Gel SizeSelect 2% Agarose, Life Technologies, Carlsbad, CA) for the 200-bp sequencing run. The samples were amplified and purified for the third time using Agencourt. The samples were then quantified using the Qubit dsDNA HS Assay kit (Invitrogen).

The amplicons were pooled and then submitted to the One Touch system for clonal amplification through emulsion PCR (Ion OneTouch 200 System Template Kit Catalog No. 4471263, Life Technologies, Carlsbad, CA) followed by the enrichment step. Finally, the pooled enriched amplicons bound to the Ion Sphere Particles were submitted to sequencing for 200-bp reads (Ion PGM 200 Sequencing Kit Catalog No. 4474004, Life Technologies, Carlsbad, CA) in a 314 chip (Ion 314 Chip 8-pack Catalog No. 4462923, Life Technologies, Carlsbad, CA).

The generated sequence data were trimmed and then assembled with the consensus reference sequence using the highest sensitivity plus fine tuning followed by manual alignment of the mapped reads and single nucleotide polymorphisms (SNP) using the Geneious R6 Software application.[13] The SNP were manually filtered and statistically compared between groups using STATA.

Table 1. Age, Proviral Load Count, and Sequencing Data of Analyzed Groups

	Asymptomatic		*TSP/HAM*	
Characteristic	*Low proviral load (n = 9)*	*High proviral load (n = 8)*	*Low proviral load (n = 7)*	*High proviral load (n = 6)*
Age, years				
Median (IQR)	33.1 (23.5–45.6)	55.3 (39.3–68.6)	53.1 (27.8–55.8)	45.8 (40.9–64.4)
Proviral load, copies per 10^6 cells				
Median (IQR)	964 (386–1,591)	248,792 (196,473–507,858)	36,776 (18,929–44,396)	188,060 (164,213–232,227)
Proviral load, $\log_{10}$				
Median (IQR)	2.98 (2.85–3.71)	5.39 (5.29–5.69)	4.57 (4.28–4.65)	5.27 (5.22–5.37)
Sanger sequencing mutations				
Median	3	2.5	3	2
Ion Torrent depth coverage				
Median (IQR)	445 (340–759)	389 (244–1,139)	551 (196–886)	352 (165–1,203)
Ion Torrent major mutations				
Median	5	2.5	3	2.5
Ion Torrent minor mutations				
Median	7	3.5	7	6

IQR, interquartile range; n, number of patients; TSP/HAM, tropical spastic paraparesis/human T cell lymphotropic virus type 1-associated myelopathy.

Table 2. Mutations Frequency by Groups and Transcription Factor Binding Sites Change According to Mutation

Mutation position	*Change*	*Previously described TFBS*	*TFBS by TFSCAN*	*TFBS change*		*Asymptomatic* (n=*17*)	*TSP/HAM* (n=*13*)	*Low pvl* (n=*16*)	*High pvl* (n=*14*)	*Asy and low pvl* (n=*9*)	*Asy and high pvl* (n=*8*)	*TSP/HAM and low pvl* (n=*7*)	*TSP/HAM and high pvl* (n=*6*)
				Creation	*Abrogation*								
126	G → C/T	None	None	—	—	0.35	—	0.25	0.14	0.44	0.25	—	—
151Delta11	Delta11	TxRE-2	Sp-1, CREB, ATF	—	Sp-1, CREB, ATF	0.06	—	—	0.07	—	0.13	—	—
153	G → A	TxRE-2	Sp-1	—	Sp-1	0.35	0.46	0.44	0.36	0.33	0.38	0.57	0.33
165	T → C	TxRE-2	None	c-ets-2, CACCC-binding factor	—	0.12	—	0.13	—	0.22	—	—	—
167	C → T	TxRE-2	c-ets-2	—	c-ets-2	0.12	—	0.06	0.07	0.11	0.13	—	—
199	G → A	c-ets-2	c-ets-2	—	c-ets-2	—	0.15	—	0.14	—	—	—	0.33
207	A → G	Sp1	TGT-3	—	TGT-3	0.12	—	0.13	—	0.22	—	—	—
215	C → T	Sp1	GammaCAC2, CACCC-binding factor	—	GammaCAC2, CACCC-binding factor	0.06	0.15	0.19	—	0.11	—	0.29	—
219	C → G/T	c-ets-2	GammaCAC2, CACCC-binding factor	H4TF-1	GammaCAC2, CACCC-binding factor	0.29	0.38	0.31	0.36	0.22	0.38	0.43	0.33
268	C → T	TxRE-3	c-ets-2	—	c-ets-2	0.12	0.08	0.13	0.07	0.22	—	—	0.17
298	A → G	E-Box	None	—	—	—	0.15	0.13	—	—	—	0.29	—
301	T → A/C	E-Box	None	Sp-1	—	—	0.15	0.06	0.07	—	—	0.14	0.17
306	G → C/A	None	Sp1	ATF	Sp-1	0.24	0.08	0.31	—	0.44	—	0.14	—
345	T → C	Inr	None	—	—	0.12	—	0.13	—	0.22	—	—	—
479	C → T	None	None	—	—	0.35	0.08	0.31	0.14	0.56	0.13	0.00	0.17

Position related to J02029. Pvl, proviral load; Asy, asymptomatic; TFBS, transcription factor binding site; TSP/HAM, tropical spastic paraparesis/human T cell lymphotropic virus type 1-associated myelopathy.

The HTLV-1 sequences were submitted to the LASP HTLV-1 Automated Subtyping Tool Version 1.0[14] for genotyping in order to exclude subtype-specific mutations. To analyze the transcription factor binding site (TFBS) modifications related to specific mutations we used the TFScan plugin of the Geneious software.

The selected patients were analyzed as to age and proviral load. The median age of asymptomatic carriers with a low proviral load was lower than the other groups but had no statistical significance (Table 1). The log proviral load was similar among the groups characterized as high proviral load (5.39 Asymptomatic X 5.27 TSP/HAM, $p=0.1551$) but slightly different in the low proviral load group (2.99 Asymptomatic X 4.57 TSP/HAM, $p=0.0172$) possibly due to the age difference.

The Ion Torrent sequencing showed a variable coverage depth between the samples ranging from 108.6x to 3740.5X (Table 1). To avoid detecting subtype-specific mutations we submitted the sequences to the LASP HTLV-1 Automated Subtyping Tool Version 1.0 and all the sequences were classified as subtype a subgroup A. We described the mutations found on Ion Torrent as minor (if the mutation prevalence was lower than 50%) and major (if the mutation prevalence was higher or equal to 50%). We found that samples with a low proviral load showed more minor mutations than the samples with high proviral load.

We found mutations in 136 positions over the 520-bp analyzed fragment of HTLV-1 LTR5′ with at least 1% frequency. Out of 136 mutations 52 were present in more than one patient, with 11 out of 52 present in the previously determined major TFBS (Table 2). The other three mutations were statistically significant using the Fisher nonparametric test between the groups but were not present in previously determined TFBS (G126C/T, G306C, and C479T).

We performed an *in silico* analysis using TFScan of 11 mutations in the previously described TFBD plus those mutations with a statistical difference between groups and one deletion of 11 base pairs to determine possible modifications in the TFBS. We found that A153G, C167T, G199A, A207G, C215T, C219G/T, C268T, and G306C abrogated possible TFBS while T165C, C219G/T, T301A/C, and G306C/A created new possible TFBS (Table 2).

The Ion Torrent DNA sequencing was able to provide more information on HTLV-1 quasispecies than the Sanger sequencing, which lacks the sensitivity to show mutations at lower levels (20%). For HTLV-1 studies it is important to analyze quasispecies instead of viral populations because the virus is highly conserved with low evolutionary rates.[15–17] In addition, only around 1% of the infected cell will produce HTLV-1 proteins in a day,[18] which makes it important to analyze the mutations at 1% prevalence or even lower in studies, whereas the HTLV-1 mutations are being analyzed as disease outcome. The detection of mutations in approximately one in four analyzed sites by Ion Torrent sequencing with at least 1% frequency supports the use of this technique to search for minor mutations.

It was found that samples with low proviral load showed more detected minor variants than samples with high proviral load, which could be partly explained by the highest rate of clonal expansion of HTLV-1-infected cells.[7]

Eleven mutations were present in the previously determined major TFBS and in more than one patient, indicating that there might be a differential HTLV-1 expression between individuals or in different cells from the same individual. Studies indicated that two mutations might be a marker for an increase in viral load[9,10] and those mutations were usually found in the TFBS. We have searched one of those mutations to confirm the data previously described but we did not find any associations with proviral load; unfortunately all studies lacked sample size.

Three mutations that were not present in previously determined TFBS were statistically significant in this study and were most frequent in patients with low proviral load or in asymptomatic carriers. Despite the fact that those mutations were not present in previously determined TFBS, one of those mutations (G306C/A) was present in an Sp-1 binding site determined by *in silico* analysis, and its presence abrogated the site for Sp-1 binding and created a new possible ATF binding site.

The HTLV-1 promoter binding sites are not completely understood mainly because of the absence of chromatin immune precipitation analysis in the LTR promoter of HTLV-1-infecting T CD4^{+} cells, and those mutations could decrease or increase the viral expression abrogating or creating TFBS. However, a further controlled study must be conducted in order to determine if there are transcription factors bound to those positions as well as the 11 positions in which mutations were found.

Sequence Data

The GenBank accession numbers of the new HTLV-1 LTR fragments included in the study are SRR2182833 to SRR2182854.

Acknowledgments

The authors are grateful to Ms. Prevashine Padayache for technical assistance and Justen Manasa M.Sc. for DNA Sanger sequencing support. This project was supported by Brazil Ciencia Sem Fronteiras project from CNPQ. F.F.A.R. was supported by CAPES.

F.F.A.R., T.O., and M.G. wrote the paper and did the analysis. F.F.A.R. did the experiments. T.O., M.G., B.G., and L.C.A. helped with the discussion. F.F.A.R., B.G.C., and L.C.A. designed the study. B.G. provided the samples.

Author Disclosure Statement

The authors certify that they have no affiliations with or involvement in any organization with any financial or nonfinancial interest in the subject matter.

References

1. Gessain A, Barin F, Vernant JC, *et al.:* Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet 1985;2:407–410.
2. Osame M, Usuku K, Izumo S, *et al.:* HTLV-I associated myelopathy: A new clinical entity. Lancet 1986;1: 1031–1032.
3. Yoshida M, Miyoshi I, and Hinuma Y: Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc Natl Acad Sci USA 1982;79(6):2031–2035.
4. Mochizuki M, Ono A, Ikeda E, *et al.:* HTLV-I uveitis. J Acquir Immune Defic Syndr Hum Retrovirol 1996;13(Suppl 1):S50–56.

5. La Grenade L: HTLV-I-associated infective dermatitis: Past present and future. J Acquir Immune Defic Syndr Hum Retrovirol 1996;13(Suppl 1):S46–49.
6. Taniguchi Y, Nosaka K, Yasunaga J, *et al.:* Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. Retrovirology 2005;2:64.
7. Gillet NA, Malani N, Melamed A, *et al.:* The host genomic environment of the provirus determines the abundance of HTLV-1-infected T cell clones. Blood 2011;117(11): 3113–3122.
8. Olavarria VN, Gomes AN, Kruschewsky RA, *et al.:* Evolution of HTLV-1 proviral load in patients from Salvador, Brazil. Braz J Infect Dis 2013;16(4):357–360.
9. Neto WK, Da-Costa AC, De Oliveira AC, *et al.:* Correlation between LTR point mutations and proviral load levels among human T cell lymphotropic virus type 1 (HTLV-1) asymptomatic carriers. Virol J 2011;8:535.
10. Magri MC, Costa EA, and Caterino-De-Araujo A: LTR point mutations in the Tax-responsive elements of HTLV-1 isolates from HIV/HTLV-1-coinfected patients. Virol J 2012;9:184.
11. Grant C, Barmak K, Alefantis T, Yao J, *et al.:* Human T-cell leukemia virus type I and neurologic disease: Events in bone marrow, peripheral blood, and central nervous system during normal immune surveillance and neuroinflammation. J Cell Physiol 2002;190:133–159.
12. Mboudjeka I, Zekeng L, Yamashita M, *et al.:* Prevalence and phylogenetic analysis of HTLV-I isolates in Cameroon, including those of the Baka Pygmy. Jpn J Cancer Res 1997;88(7):619–624.
13. Kearse M, Moir R, Wilson A, *et al.:* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012;28(12):1647–1649.
14. Alcantara LC, Cassol S, Libin P, *et al.:* A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res 2009;37:634–642.
15. Gessain A and Gout O: Chronic myelopathy associated with human T-lymphotropic virus type I (HTLV-I). Ann Intern Med 1992;117(11):933–946.
16. Sherman MP, Saksena NK, Dube DK, *et al.:* Evolutionary insights on the origin of human T-cell lymphoma/leukemia virus type I (HTLV-I) derived from sequence analysis of a new HTLV-I variant from Papua New Guinea. J Virol 1993;66(4):2556–2563.
17. Switzer WM, Salemi M, Qari SH, *et al.:* Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. Retrovirology 2009;6:9.
18. Asquith B, Zhang Y, Mosley AJ, *et al.:* In vivo T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection. Proc Natl Acad Sci USA 2007;104(19):8035–8040.

Address correspondence to:
Luiz Carlos Alcantara
LHGB/CPqGM/FIOCRUZ
Rua Waldemar Falcão 121
Candeal, Salvador, Bahia, 40295-001
Brazil

E-mail: lalcan@bahia.fiocruz.br

This article has been cited by:

1. V. Sauvage, M. Eloit. 2016. Viral metagenomics and blood safety. *Transfusion Clinique et Biologique* **23**, 28-38. [CrossRef]