Bioinformatics Computational & Systems Biology

http://elynsgroup.com

Short Communication Journal of Bioinformatics, Computational and Systems Biology Open Access

A Fully Annotated Genome Sequence of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1)

Fernanda K. Barreto[1], Thessika HA. Araújo[1,2], Filipe F. Almeida Rego[3], and Luiz CJ. Alcantara

[1]Oswaldo Cruz Foundation, Institute Gonçalo Moniz , Salvador, Bahia, Brazil

[2]Bahian School of Medicine and Public Health, Salvador, Bahia, Brazil

[3]Catholic University of Salvador, Salvador, Bahia, Brazil

Received Date: May 09, 2017, **Accepted Date:** June 16, 2017, **Published Date:** June 23, 2017.

***Corresponding author:** Luiz Carlos Junior Alcantara, Oswaldo Cruz Foundation, Institute Gonçalo Moniz, Salvador, Bahia, Brazil, E-mail: lalcan@bahia.fiocruz.br .*

Abstract

In the next generation of genome sequencing, sequence annotation plays an important role with respect to genome evaluation. The aim of annotation is to identify key features in the genome, such as genes and their products. Although annotation tools are available and some sequence features have been published, annotation information for many complete and partial genomes of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) remains unavailable from GenBank. Sequence analysis is critical to the understanding of the pathogenesis of HTLV-1, and a well-annotated reference sequence is an essential component in this analysis. More accurate and complete information about the HTLV-1 genome can assist the scientific community in investigations on possible therapeutic and prophylactic vaccines, as well as aid studies on the pathogenesis of HTLV-1-associated diseases. Here we describe for the first time the complete nucleotide position annotation of the frequently used HTLV-1 reference sequence, ATK1 (accession number: J02029.1).

Keywords: HTLV-1; ATK1; Complete Genome; Annotation

Introduction

Human T-cell lymphotropic virus type 1 (HTLV-1) is present throughout the world and it is estimated that 5–10 million individuals are infected [1]. This retrovirus has been mainly linked to Adult T Cell Leukemia/Lymphoma (ATLL), Tropical Spastic Paraparesis/HTLV-associated myelopathy (HAM/TSP) and infective dermatitis [2–4].

One of the challenges faced by researchers in the development of an HTLV-1 vaccine is to determine why some individuals develop pathological processes, while others remain asymptomatic. Genomic studies have indicated that HTLV-1 mutations may be associated with infection outcome, yet the GenBank database contains relatively few complete genomes available [5–8]. In addition, the most used HTLV-1 genome (ATK1) is incomplete with regard to the start and end nucleotide position of each gene. ATK1 was the first human retrovirus genome described and to date has not been fully annotated [9]. Here, we performed the complete nucleotide position annotation of the full ATK1 genome available at GenBank. We hope that this information will support future HTLV-1 research efforts by the scientific community.

Materials and Methods

To perform the complete nucleotide position annotation of the most used HTLV-1 genome (ATK1), this sequence was downloaded from GenBank (accession number: J02029.1) and all available features were recorded. Next, we identified in GenBank other complete and partial HTLV-1 sequences with some nucleotide position information, through the "HTLV-1 complete sequence; HTLV-1 and LTR; HTLV-1 and HBZ; HTLV-1 and p12; HTLV-1 and p30" keywords.

After downloading these sequences, Clustal X 2.0 software was used to align all sequences, including ATK1 [10,11]. The alignment was manually edited and the correct nucleotide positions of the HTLV-1 genes in the complete and partial sequences was analyze in relation to ATK1 sequence. The nucleotide position annotation of ATK1 was performed using Geneious R6 software [12]. Finally, Universal Protein Resource (UniProt, www.uniprot.org) was used to confirm coding region annotations through the alignment of HTLV-1 protein sequences available in the UniProt and the ATK1 sequence translated based on our annotations [13]. Figure 1 explains the workflow of ATK1 nucleotide annotation.

Results and Discussion

Complete nucleotide annotations provide the scientific community with the necessary data to better interpret biological processes. In the case of HTLV-1, this information is particularly important since the literature is controversial with respect to the nucleotide position of each of the protein products, especially those produced by pX.

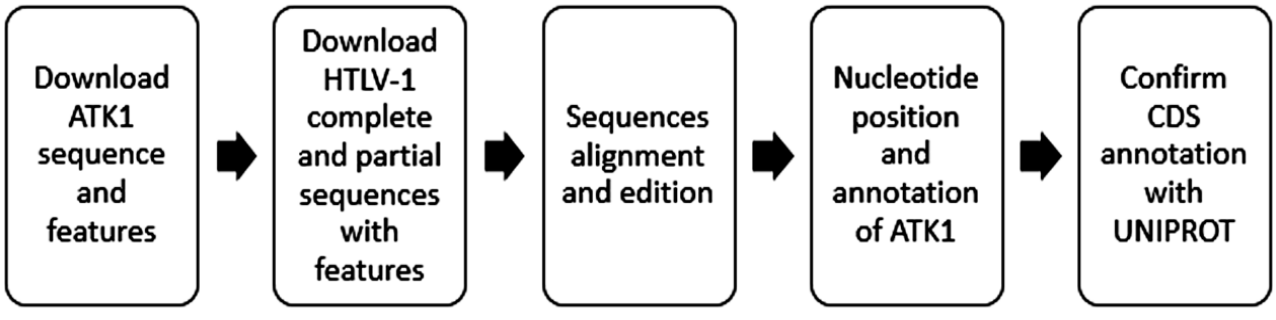

Figure 1: The workflow of ATK1 nucleotide position annotation. The ATK1 accession numbers is J02029.1. The accession numbers of the HTLV-1 complete and partial sequences downloaded are NC_001436, Y16487.2, U19949, JX184913.1 and KM436104.1.

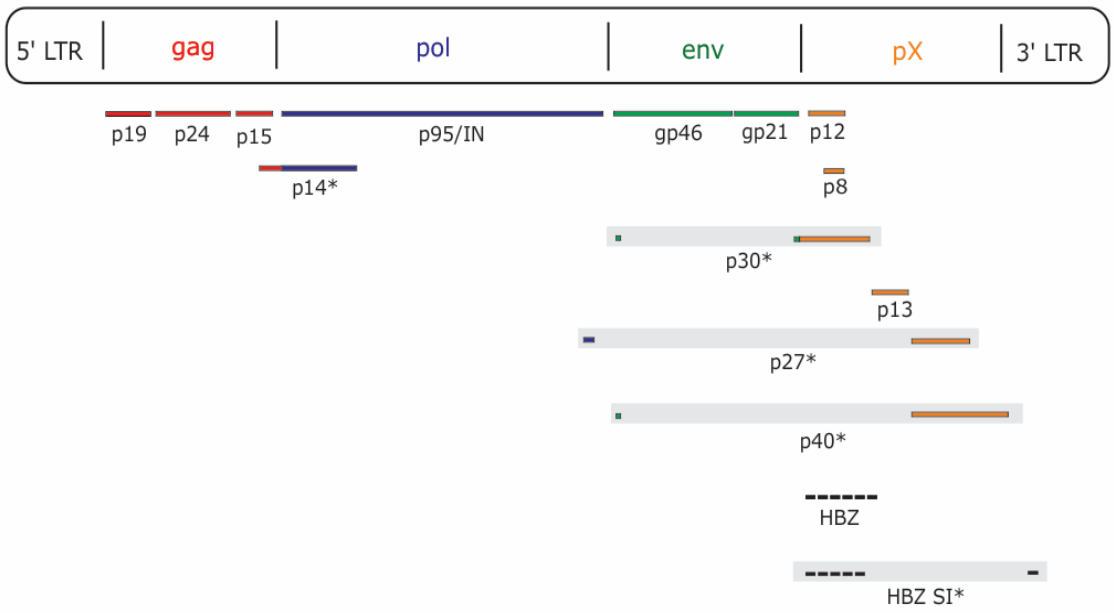

Figure 2: The fully annotated genome of human T-cell lymphotropic virus type 1 (HTLV-1). (The complete HTLV-1 genome is represented as a box. The LTR promoter regions and the genes encoded by sense mRNA are shown in different colors. Each protein is represented by a line with a color specific to its corresponding gene. HBZ and HBZ-SI are encoded by an antisense mRNA, represented by a dotted line. Proteins labeled with asterisks (*) are encoded by more than one gene).

Region	Repeat Features		Location
5' LTR	U3		23...374
	R		375...603
	U5		604...777
3' LTR	U3		8301...8652
	R		8653...8881
	U5		8882...9055
Gene	**Mature Peptides**	**CDS**	**Location**
gag	gag-pro-pol	p19	824...1213
		p24	1214...1855
		p15	1856...2113
pro	gag-pro-pol	p14	1960...2778
pol	gag-pro-pol	p95/IN	-...5210
env	gp63	gp46	5203...6138
		gp21	6139...6669
pX	p12	p12	6857...7156
	p8	p8	6944...7156
	p30	p30	5203-5206...6853-7574
	p13	p13	7311...7574
	Rex	p27	5147-5206...7325-7834
	Tax	p40	5203-5206...7325-8382
	HBZ	HBZ	7312...6686
	HBZ SI	HBZ SI	7290-6686...8702-8690

Table 1: Nucleotide positions of HTLV-1 genes in the ATK1 sequence (accession number: J02029.1). (The sites of reverse transcriptase (p95) and integrase translation initiation have not been determined; LTR = long terminal repeat).

To establish a localization standard for the HTLV-1 genes and proteins, we analyzed the ATK1 sequence and performed the complete nucleotide annotation of this sequence. As shown in figure 2, the HTLV-1 genome is composed of genes gag, pol, env and the pX region, flanked by two Long Terminal Repeat (LTR) regions at both 5' and 3' ends. The gag precursor protein is cleaved into products. The pol gene encodes polymerase p95 (reverse transcriptase) and an integrase, although the sites of translation initiation have yet to be determined. A frameshift occurs at the 3' gag termination and the beginning of the pol gene, which encodes p14 (protease). The env precursor protein is also cleaved to generate two products: gp46 and gp21. The pX region contains four overlapping open reading frames (ORF) that encode regulatory and accessory proteins and an antisense mRNA that generates the basic leucine zipper (HBZ) protein and the isoform of HBZ (HBZ-SI). ORF-I produces the p12 protein, which can be further cleaved into p8 protein, while ORF-II produces two proteins: p13 and p30, with part of the p30 protein being coded by env. ORF-III and ORF-IV produce proteins p27 (Rex) and p40 (Tax), respectively, both also partially coded by env. The size of this genome is approximately 9 kilobase (kb) and the start and end nucleotide positions of each gene are described below in Table 1.

The RefSeq database of GenBank suggests another sequence as reference (accession number: NC_001436). However, most of HTLV-1 papers used the ATK1 as reference sequence in their analysis [14–16]. Nevertheless, both of these sequences do not have information about all the HTLV-1 products, as p14, p12, p8, p30, p13 and HBZ. Therefore, our complete results can be used as a reference for the alignment and annotation of other HTLV-1 genomes.

Conclusion

The present study attempted to perform a complete nucleotide annotation of the most used HTLV-1 complete genome, ATK1. There are many questions that remain to be answered in the field of HTLV-1 research, and we hope that these data will assist other investigations carried out by the scientific community.

Availability of Data and Material

All sequences are available in the GenBank database (accession numbers: J02029.1; NC_001436; Y16487.2; U19949; JX184913.1; KM436104.1).

Authors' Contribution

All authors wrote, read and approved the final manuscript.

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. Front Microbiol. 2012;3:388. doi: 10.3389/fmicb.2012.00388.
2. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet. 1985;2(8452):407-10.
3. Yoshida M, Miyoshi I, 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 U S A. 1982;79(6):2031-5.
4. Goncalves DU, Guedes ACM, Proietti AB de FC, Martins ML, Proietti FA, Lambertucci JR. Dermatologic lesions in asymptomatic blood donors seropositive for human T cell lymphotropic virus type-1. Am. J. Trop. Med. Hyg. 2003;68:562–565.
5. Ilinskaya A, Heidecker G, Derse D. Opposing effects of a tyrosine-based sorting motif and a PDZ-binding motif regulate human T-lymphotropic virus type 1 envelope trafficking. J Virol. 2010;84(14):6995-7004. doi: 10.1128/JVI.01853-09.
6. Barreto FK, Khouri R, Rego FF, Santos LA, Castro-Amarante MF, Bialuk I, et al. Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome. Infect Genet Evol. 2016;45:420-425. doi: 10.1016/j.meegid.2016.08.020.
7. Neto WK, Da-Costa AC, de Oliveira AS, Martinez VP, Nukui Y, Sabino EC, 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. doi: 10.1186/1743-422X-8-535.
8. Magri MC, Costa EAS, 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. doi: 10.1186/1743-422X-9-184.
9. Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA. 1983;80:3618–3622.
10. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.
11. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41:D36–42. doi: 10.1093/nar/gks1195.
12. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9. doi: 10.1093/bioinformatics/bts199.
13. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge A, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. In: Edwards D, editor. Plant Bioinforma. [Internet]. Springer New York; 2016 [cited 2017 Jan 9]. p. 23–54. Available from: https://link.springer.com/protocol/10.1007%2F978-1-4939-3167-5_2.
14. Pessôa R, Watanabe JT, Nukui Y, Pereira J, Casseb J, de Oliveira AC, et al. Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by Illumina massively parallel sequencing technology. PLoS One. 2014;9(3):e93374. doi: 10.1371/journal.pone.0093374.
15. Rosadas C, Vicente ACP, Zanella L, Cabral-Castro MJ, Peralta JM, Puccioni-Sohler M. Evidence of Sexual Transmission of Human T Cell Lymphotropic Virus Type 1 with ORF-I G29S Mutation Among Humans. AIDS Res Hum Retroviruses. 2017;33(4):328-329. doi: 10.1089/AID.2016.0238.
16. Mota-Miranda AC, de-Oliveira T, Moreau DR, Bomfim C, Galvão-Castro B, Alcantara LC. Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design. Mem. Inst. Oswaldo Cruz. 2007;102:741–9.

Corresponding author: Luiz Carlos Junior Alcantara, Oswaldo Cruz Foundation, Institute Gonçalo Moniz, Salvador, Bahia, Brazil, E-mail: lalcan@bahia.fiocruz.br .

Received Date: May 09, 2017, **Accepted Date:** June 16, 2017, **Published Date:** June 23, 2017.

Citation: Barreto FK, Araújo THA, Rego FFA, Alcantara LC (2017) A Fully Annotated Genome Sequence of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1). J Bioinf Com Sys Bio 1(1): 105.