**Caracterização molecular do HTLV-1: impacto no desenvolvimento de doenças associadas**

**Introdução**: O HTLV-1 é um patógeno distribuído globalmente, afetando cerca de 5 a 10 milhões de pessoas. Estima-se que no Brasil existem 800 mil pessoas vivendo com o vírus, o que o caracteriza como um dos países com maior número absoluto de casos. No Brasil, as regiões Norte e Nordeste são as mais afetadas pelo vírus. A infecção pelo HTLV-1 pode levar ao desenvolvimento de doenças como ATLL, HAM/TSP e IDH. O genoma do HTLV-1 possui genes estruturais e uma região responsável pela codificação de proteínas acessórias e regulatórias, importantes para o estabelecimento, disseminação e ação do vírus. O sequenciamento pode elucidar fatores desse vírus que influenciam a propagação e o desenvolvimento de manifestações clínicas associadas ao HTLV-1. **Objetivo:** O objetivo principal deste trabalho é avaliar o papel das mutações no genoma do HTLV-1 e sua possível associação com o desenvolvimento das doenças associadas ao vírus. **Material e Métodos:** Estudo de corte transversal, com amostras de indivíduos infectados pelo HTLV-1, que foram sequenciadas para identificação de mutações nas regiões LTR e *hbz* do genoma do vírus. Além disso, duas revisões sistemáticas foram realizadas seguindo as instruções do PRISMA para aprofundamento e revisão das áreas de estudo. **Resultados:** Foram observadas duas mutações, V15M e R119Q, na região de *hbz*, sendo a V15M observada exclusivamente em pacientes HAM/TSP, enquanto R119Q aparenta ser um fator de proteção para o desenvolvimento de doença. Em LTR foram encontradas mutações que levaram ao surgimento ou deleção de sítios de ligação de fatores de transcrição. Além disso, foram encontrados pela revisão sistemática quatro trabalhos que relacionaram mutações em *env* e ORF-1 com o desenvolvimento de alguma forma clínica. Em relação ao estudo de sequenciamento, a metodologia Sanger é mais utilizada para o estudo do HTLV-1. **Conclusões**: Foi possível observar a presença de mutações relacionadas ao desenvolvimento de HAM/TSP, como C39Y, P45L, S69G, P86S e R88K (ORF-I), S72G e N93D (gp46) e V15M e R119Q (HBZ). Também foram observadas alterações nos sítios de ligação de fatores de transcrição, que podem estar relacionadas a um desequilíbrio de produção e expressão de genes virais. Além disso, foi possível observar que apesar dos avanços nas tecnologias de sequenciamento, os estudos de HTLV-1 ainda trabalham com a primeira geração de sequenciamento, o Sanger.

**Palavras-chave:** HTLV-1; sequenciamento; mutações, caracterização molecular

**Molecular Characterization of HTLV-1: impact on the development of associated diseases**

***Background:*** *HTLV-1 is a globally distributed pathogen that affects approximately 5 to 10 million people. It is estimated that Brazil has about 800,000 people living with the virus, which characterizes it as one of the countries with the highest absolute number of cases. In Brazil, the North and Northeast regions are the most affected by the virus. HTLV-1 infection can lead to the development of diseases such as ATLL, HAM/TSP and IDH. The HTLV-1 genome has structural genes and a region responsible for encoding accessory and regulatory proteins, important for the establishment, dissemination, and action of the virus. The sequencing of this genome can elucidate factors of this virus that influence the spread and development of clinical manifestations associated with HTLV-1.* ***Objective:*** *The main objective of this work is to evaluate the role of mutations in the HTLV-1 genome and its possible association with the development of HTLV-1-associated diseases.* ***Methods:*** *Cross-sectional study, with samples from individuals infected with HTLV-1, which were sequenced to identify mutations in the LTR and hbz regions of the virus genome. In addition, two systematic reviews were performed following the PRISMA instructions for deepening and reviewing the study areas.* ***Results:*** *Two mutations, V15M and R119Q, were observed in the hbz region, with V15M observed exclusively in HAM/TSP patients, while R119Q appears to be a protective factor for the development of the disease. In LTR, mutations were found that led to the appearance or deletion of transcription factor binding sites. In addition, one of the systematic reviews found four studies that associated mutations in env and ORF-1 with the development of some clinical form. Regarding the sequencing study, the Sanger methodology is more used for the study of HTLV-1.* ***Conclusions:*** *It was possible to observe the presence of mutations related to the development of HAM/TSP, such as C39Y, P45L, S69G, P86S and R88K (ORF-I), S72G and N93D (gp46) and V15M and R119Q (HBZ). Alterations in transcription factor binding sites were also observed, which may be related to an imbalance in the production and expression of viral genes. In addition, it was possible to observe that despite advances in sequencing technologies, HTLV-1 studies still work with the first generation of sequencing, Sanger.*

***Key words:*** *HTLV-1; sequencing; mutation, molecular characterization*

**ARTIGO 1 -** “***Molecular characterization of HTLV-1 genomic region hbz from patients with different clinical conditions***”. **(doi: 10.1002/jmv.27005)**

**Tabela

Descrição gerada automaticamenteTexto

Descrição gerada automaticamente**

FIGURE 1 HBZ amino acid sequence, mutations, and protein domains. (A) HBZ amino acid sequences. Reference HBZ sequence (ATK1—J02029) on top and mutated HBZ sequence below. The amino acid mutations are highlighted in bold. (B) Location of the amino acid mutations (black line) in regard to the HBZ molecular domain structure. The striped region corresponds to the activation domain, the white corresponds to the central domain, where the basic region 1 is contained (dark gray) and the light gray is the bzip domain.

TABLE 1 Variant frequency of hbz mutations and comparison of frequencies between AC and HAM/TSP, ATLL, and IDH

Note: In bold are the values with statistical support

Abbreviations: AA, amino acid; AC, asymptomatic; ATLL, adult T‐cell leukemia/lymphoma; HAM/TSP: HTLV‐1‐associated myelopathy/tropical spastic paraparesis; IDH, infective dermatitis associated with HTLV‐1.

aBased on ATK1 (J02029.1).

bStatistical difference (p‐value < 0.05) between all groups calculated by χ2 test and statistical correction by Bonferroni method.

cStatistic difference (p‐value < 0.05) between AC and disease (HAM/TSP and ATLL) calculated by Fisher's exact test and statistical correction by Bonferroni method.

dCould not calculate; n: corresponds to the total number of analyzed sequences in each study group. The relative frequency of each mutation represents the proportion of mutated sequences in a total number of sequences in the nucleotide position.

TABLE 2 Physicochemical alterations according to HBZ amino acid change

Texto

Descrição gerada automaticamente

**Genotipagem e identificação de mutações em LTR de amostras do Centro de Atendimento ao Indivíduo com HTLV-1 (CHTLV)**

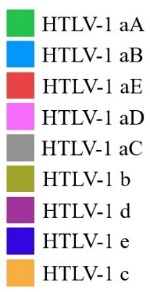
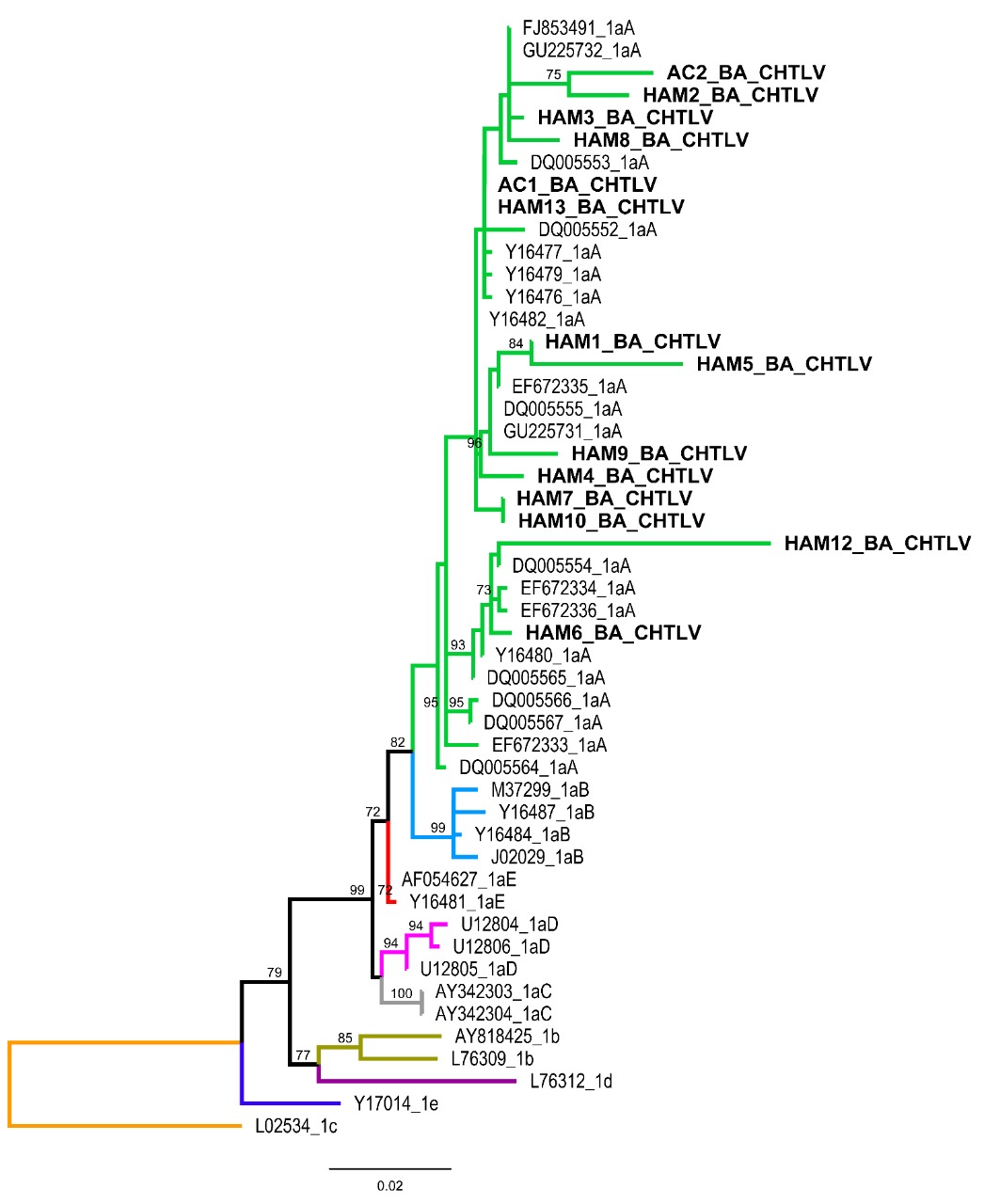


Figura 1. Árvore de Máxima Verossimilhança enraizada do HTLV-1. Árvore filogenética de Máxima Verossimilhança do HTLV-1 baseada na região LTR com 760pb contendo 14 novas sequências (nomes em negrito) e 36 sequências referências. Sequência L02534 (Isolado Mel 5) foi usada como grupo externo para enraizar a árvore.

Tabela 2 Frequência das modificações dos sítios de ligação dos fatores de transcrição, causadas pelas mutações em seis sequências de LTR do CHTLV.



**ARTIGO 2 -** “***Mapping variants in HTLV-1 genome to analyze their impacts on the HAM/TSP development: a systematic review***”.

Table 1. Summary of the information collected on the eleven included studies. In bold, the HTLV-1 mutations possibly associated with HAM/TSP development are highlighted.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Diagrama

Descrição gerada automaticamente

Figure 1. Flow diagram of the systematic selection.

Gráfico

Descrição gerada automaticamente com confiança baixa

Figure 2. HTLV-1 non-synonymous mutations possibly associated with HAM/TSP development.

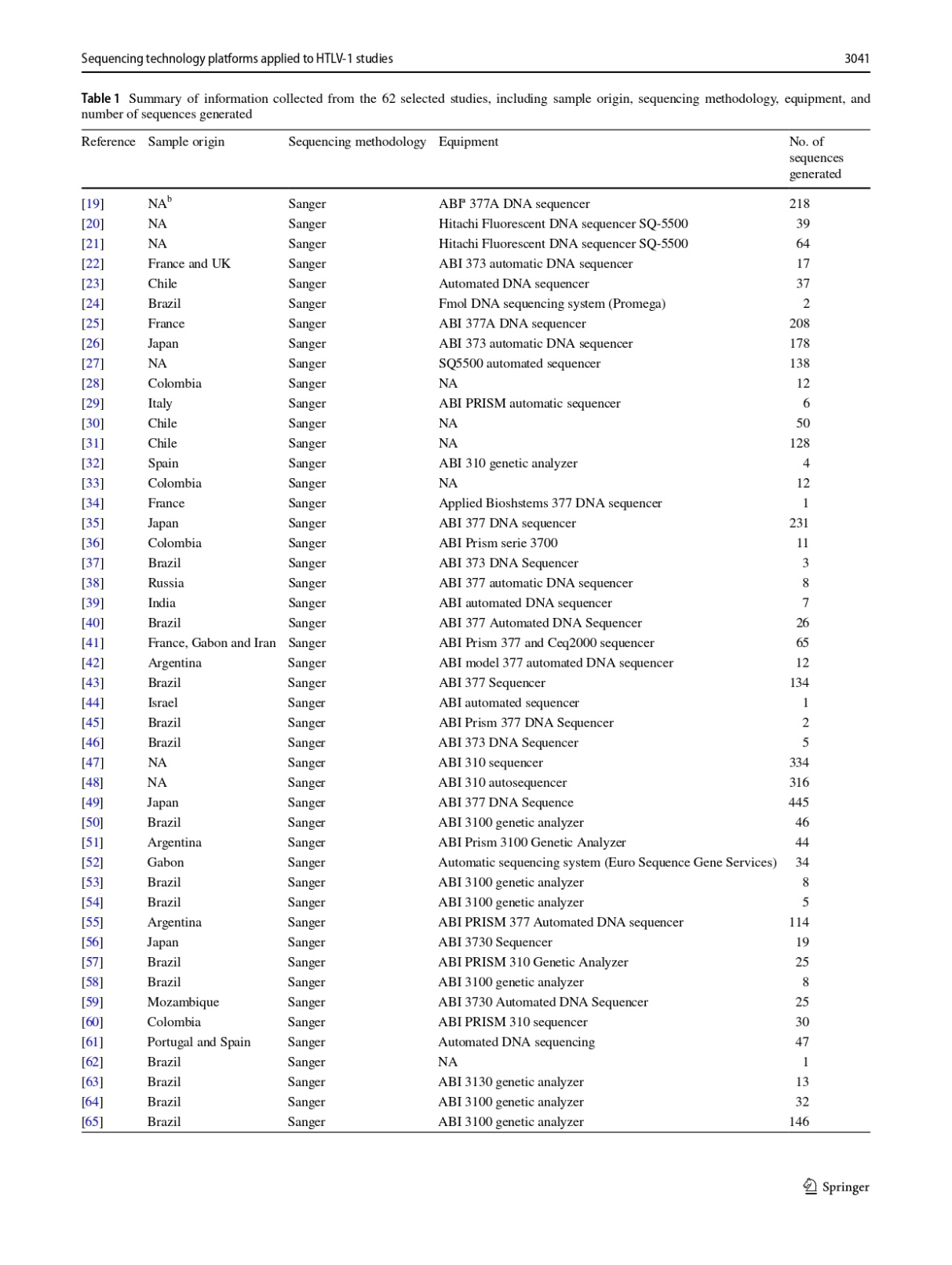
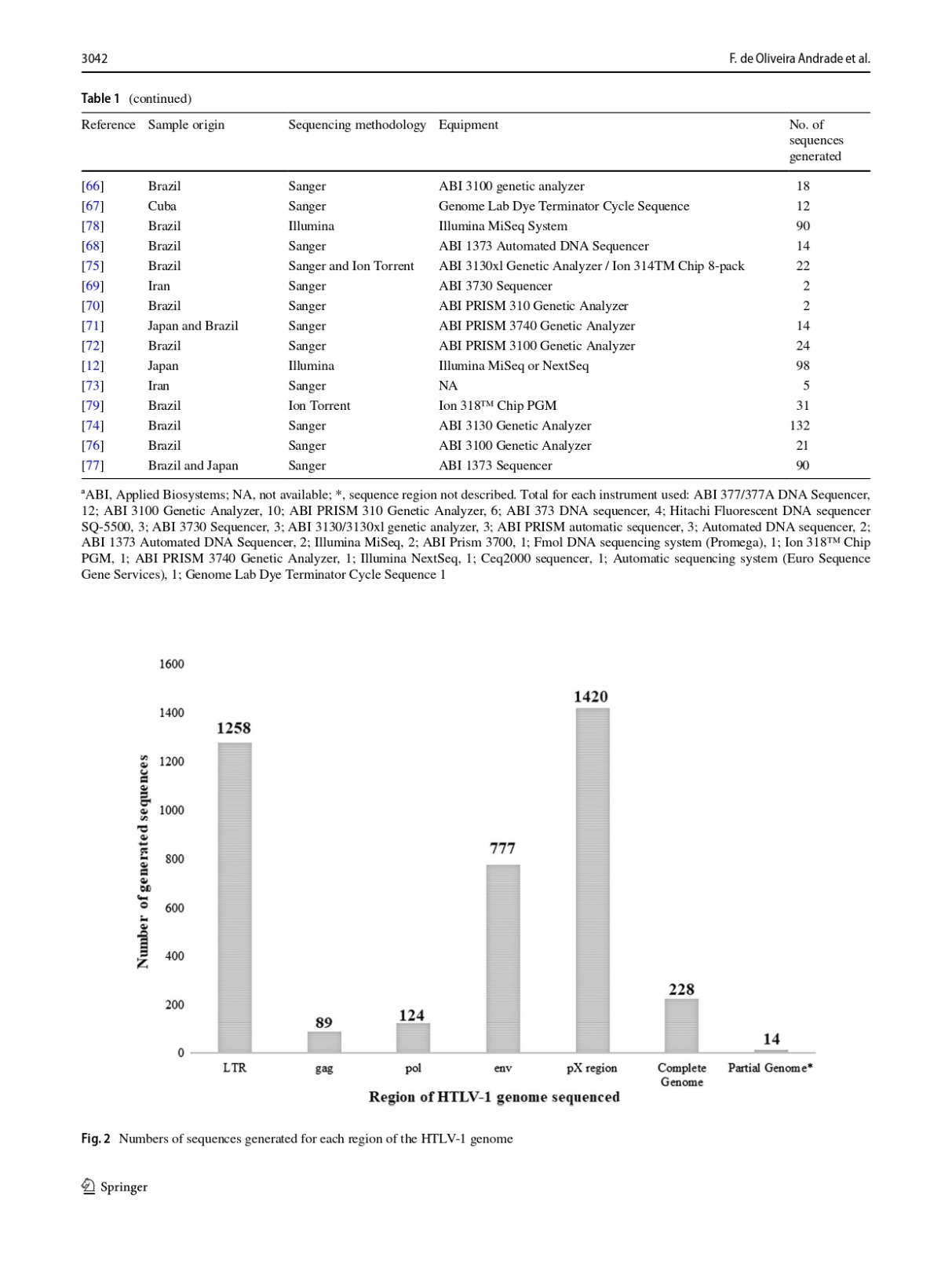
**ARTIGO 3 -** “***An overview of sequencing technology platforms applied to HTLV‑1 studies: a systematic review***”. **(doi: 10.1007/s00705-021-05204-w)**

**Diagrama

Descrição gerada automaticamente**

Fig. 1 Flow diagram for the systematic selection of studies to review.

Table 1 Summary of information collected from the 62 selected studies, including sample origin, sequencing methodology, equipment, and number of sequences generated.



aABI, Applied Biosystems; NA, not available; \*, sequence region not described. Total for each instrument used: ABI 377/377A DNA Sequencer,12; ABI 3100 Genetic Analyzer, 10; ABI PRISM 310 Genetic Analyzer, 6; ABI 373 DNA sequencer, 4; Hitachi Fluorescent DNA sequencer SQ-5500, 3; ABI 3730 Sequencer, 3; ABI 3130/3130xl genetic analyzer, 3; ABI PRISM automatic sequencer, 3; Automated DNA sequencer, 2; ABI 1373 Automated DNA Sequencer, 2; Illumina MiSeq, 2; ABI Prism 3700, 1; Fmol DNA sequencing system (Promega), 1; Ion 318™ Chip PGM, 1; ABI PRISM 3740 Genetic Analyzer, 1; Illumina NextSeq, 1; Ceq2000 sequencer, 1; Automatic sequencing system (Euro Sequence Gene Services), 1; Genome Lab Dye Terminator Cycle Sequence 1.

**Gráfico

Descrição gerada automaticamente**

Fig. 2 Numbers of sequences generated for each region of the HTLV-1 genome.

**Gráfico

Descrição gerada automaticamente**

Fig. 3 Geographic origin of HTLV-1 sequences and methodology used for sequence generation.

**Gráfico

Descrição gerada automaticamente**

Fig. 4 Number of HTLV-1 sequences generated between 2000 and 2020.