In Vivo HIV-1 Hypermutation and Viral Loads Among Antiretroviral-Naive Brazilian Patients

Mariana Leão de Lima-Stein,¹ Wagner Tadeu Alkmim,¹ Maria Clara de Souza Bizinoto,¹
Luis Fernandez Lopez,² Marcelo Nascimento Burattini,² Juliana Terzi Maricato,¹ Leila Giron,¹
Maria Cecília Araripe Sucupira,¹ Ricardo Sobhie Diaz,¹ and Luiz Mario Janini¹

Abstract

Hypermutation alludes to an excessive number of specific guanine-to-adenine (G → A) substitutions in proviral DNA and this phenomenon is attributed to the catalytic activity of cellular APOBECs. Population studies relating hypermutation and the progression of infection by human immunodeficiency virus type 1 (HIV-1) have been performed to elucidate the effect of hypermutation on the natural course of HIV-1 infection. However, the many different approaches employed to assess hypermutation in nucleotide sequences render the comparison of results difficult. This study selected 157 treatment-naïve patients and sought to correlate the hypermutation level of the proviral sequences in clinical samples with demographic variables, HIV-1 RNA viral load, and the level of CD4⁺ T cells. Nested touchdown polymerase chain reaction (PCR) was performed with specific primers to detect hypermutation in the region of HIV-1 integrase, and the amplified sequences were run in agarose gels with HA-Yellow. The analysis of gel migration patterns using the k-means clustering method was validated by its agreement with the results obtained with the software Hypermut. Hypermutation was found in 31.2% of the investigated samples, and a correlation was observed between higher hypermutation levels and higher viral load levels. These findings suggest a high frequency of hypermutation detection in a Brazilian cohort, which can reflect a particular characteristic of this population, but also can result from the method approach by aiming at hypermutation-sensitive sites. Furthermore, we found that hypermutation events are pervasive during HIV-1 infection as a consequence of high viral replication, reflecting its role during disease progression.

Introduction

RETROVIRUSES CHARACTERISTICALLY tolerate high mutation levels at an estimated rate of 3.4×10^{-5} mutations/site/generation^{1,2}; however, even minimal increases above this tolerated threshold might result in the “error catastrophe” that, in biological terms, means the genetic collapse of these populations.³ Infection by HIV-1 is characterized by high turnover, with viral production and elimination estimated to be up to 10^{10} viruses/day⁴; nevertheless, most such viral particles are thought to be nonviable.⁵ Due to its ability to replicate and suffer mutations excessively and close to the error threshold, HIV-1 currently remains as a treatable but incurable infection.⁶

Following their original description, the apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3 family (APOBEC3s) is composed of the first enzymes described whose catalytic activity interferes with viral replication by

leading HIV-1 viral populations into an error catastrophe.⁷ Hypermutation is detected in HIV-1 as a monotonous accumulation of guanine-to-adenine (G → A) substitutions along the virus genome. In this process, the full extension of the cDNA synthesized by viral reverse transcription undergoes deamination of its cytidine-to-uridine bases under the action of cellular APOBEC3s, and while the complementary strand is being synthesized, G → A mutations are fixed in the pre-integrative and the proviral DNA.⁸ On the other hand, one of the functions of the HIV-1 Vif protein is to antagonize the action of cellular APOBEC3s, thus hindering their restrictive effect.^{9,10} Vif is a 43-kDa viral protein that is essential for productive viral infection and that inhibits the action of cellular APOBEC3s (A3B, A3DE, A3F, and A3G) by acting as an adaptive molecule that induces the proteasome-mediated degradation of APOBEC3s.¹¹

In addition, vif counteracts the action of APOBEC3s by inhibiting the packaging of APOBEC3s into the viral particle

¹Federal University of São Paulo, São Paulo, SP, Brazil.

²University of São Paulo Medical School, São Paulo, SP, Brazil.

or by decreasing the translation of APOBEC3G mRNA.¹² Most of the hypermutated genomes were proven to be defective.^{13,14} Studies established that in *vif*-defective HIV particles, there was a significant restriction of the events of the viral cycle following their entrance and integration into the host genome, in addition to the inability of these strains to cause new infections.¹⁰ Evidence indicates that hypermutated viruses are replication incompetent.¹⁴ One study employing real-time polymerase chain reaction (PCR) quantification found up to a 35-fold decrease in the viral infectiveness and a 5- to 7-fold reduction in viral DNA synthesis when HIV DNA was exposed to APOBEC3G in a culture of 293T lymphocytes¹⁵ and a recent study detected that only a single APOBEC3G molecule is able to produce inactivating levels of HIV hypermutation, as a discrete “all or nothing” phenomenon.¹⁶ Different from our findings, other studies take hypermutation as an event that generates sublethal frequencies of viral diversification, including antiretroviral resistance mutations.^{17–19}

Also reported was a correlation between the presence of hypermutated sequences and lower viral load levels,²⁰ or with higher counts of CD4⁺ T lymphocytes.²¹ In addition, an APOBEC3G polymorphism prevalent among African Americans, namely H186R, was associated with a decline in CD4⁺ T lymphocytes.²² It has also been demonstrated that high viral load set points correlate with lower levels of APOBEC3G mRNA expression,²³ and that there is an inverse correlation between HIV-1 plasma viral load and hypermutation or APOBEC3G mRNA levels in long-term non-progressors, but not in the noncontroller group.²⁴ Piantadosi *et al.* (2009),²⁵ however, did not find any correlations between the presence of hypermutation in the viral sequences and the HIV-1 viral load or the decline in CD4⁺ T lymphocytes. Assessments of the levels of mRNA expression have yielded contradictory results. Cho *et al.* (2006)²⁶ found lower levels of APOBEC3G and 3F mRNA in individuals infected with HIV-1 than in noninfected individuals. In turn, the data published by Reddy *et al.* (2010)²⁷ on a population of African women at high risk for HIV subtype C infection suggested that the primary infection by HIV-1 deregulates the expression of APOBEC3G mRNA.

Because the hypermutation of proviral sequences is the most efficient manner of detecting the action of APOBECs, this study sought to assess the frequency of HIV hypermutated sequences in proviral DNA of treatment-naïve patients in a Brazilian cohort vis-à-vis levels of viral replication as inferred by HIV viral loads. This study applied a cost reduced, previously published, and fully validated gel-based assay to detect hypermutation in HIV sequences.¹⁴

In the present study, primers were targeted to short viral genome sequences in regions that are highly conserved among the different HIV-1 subtypes²⁸ and that correspond to hypermutation hot spots. Hypermutation hot spots were addressed based on previous publications from our group in which sliding windows scanned HIV-1 genomes searching for hypermutation-related mutations, its levels, and its contexts. Furthermore, our previously published analysis indicated that the HIV-1 integrase gene is frequently a hot spot for hypermutation.^{29,30} Clones from samples with previously known levels of hypermutation were sequenced and sequences were analyzed by the Hypermut Program (Los Alamos) to access the intensity of hypermutation. The intensity of hypermutation according to Hypermut was corre-

lated with gel migration patterns applying the gel-based methodology described in Janini *et al.*¹⁴ In this system, hypermutated sequences are delayed during electrophoresis runs in agarose gels containing HA-Yellow.³¹

The validation of HA-Yellow migration patterns and the Hypermut Program output by a statistical clusterization method allowed us to characterize hypermutation based on migration patterns visualized on HA-Yellow gels. By applying this methodology to treatment-naïve patients, hypermutated sequences were recovered in 31.2% of their clinical samples. Individuals with hypermutated HIV-1 sequences exhibited greater viral load levels and a tendency toward lower CD4⁺ T lymphocyte counts, thus suggesting that the amount of hypermutated sequences generated in the viral population appears to depend upon the number of replication events that have occurred in this population.

Materials and Methods

Patient selection

Peripheral blood mononuclear cell (PBMC) samples from 157 antiretroviral treatment-naïve individuals with chronic HIV infection, with ages above 18 years old and treated at the HIV outpatient clinic of the Federal University of São Paulo (UNIFESP), were analyzed. Data were collected on viral load [nucleic acid sequence-based amplification (NASBA) method, bio Mérieux, Marcy-l'Etoile, France], CD4⁺ lymphocyte count (fluorescence-activated cell sorting—FACS-Calibur 4-color flow cytometer, BD Biosciences, San Jose, CA), and demographic variables.

DNA extraction and PCR amplification

The total DNA of the PBMC samples was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The HIV-1 integrase gene was used as the target of nested touchdown PCR. The external primers were designed containing degenerated and/or hypermutated sites in the locations where hypermutation was expected to increase the sensitivity of hypermutation detection. As shown in Table 1, the degenerated primers contained G-to-A substitutions in certain sites of the component GpA and GpG contexts, whereas the hypermutated primers contained G-to-A substitutions in all of the GpA and GpG sites. The first-round reaction included 28 mM Tris-HCl (pH 8.3), 70 mM KCl, 4.0 mM MgCl₂, 0.4 mM dNTPs, and 0.6 pM of a 1:1 mixture of each external (degenerated and hypermutated) primer and 0.025U AmpliTaq. The first cycle of the first round was composed of an initial DNA denaturation at 94°C for 7 min, annealing at 49°C for 45 s, and extension at 72°C for 55 s. The touchdown approach was used with decreasing annealing temperatures. Starting at 49°C in the first cycle, the annealing temperature was decreased by 0.5°C at each successive cycle until it reached 45°C, at which point it was kept constant until completing the 35 reaction cycles.

The resulting product corresponded to a 782-base pair fragment of the integrase gene. The second round employed 5.0 µl of the PCR product amplified in the first round in a reaction including 28 mM Tris-HCl (pH 8.3), 70 mM KCl, 3.5 mM MgCl₂, 0.4 mM dNTPs, and 0.6 pM of each internal primer. The primers for the second round were designed to

TABLE 1. HIV-1 INTEGRASE PRIMER SEQUENCES, THEIR HXB-2 POSITIONS, AND MELTING TEMPERATURES

Primers	(Sequence 5'-3')	HXB-2 location	Melting temperature (°C)
HIV integrase—first round			
Outer degenerated sense	GCTTTGCARRATTCRRRATTAG	4,002–4,023	47.6–56.2
Outer degenerated antisense	GTCAGYYTGATYTYTTAYYTGTTC	4,716–4,740	48.3–58.7
HIV integrase—first round			
External hypermutated sense	GCTTTGCAAAATTCAAAATTAG	4,002–4,023	47.6
External hypermutated antisense	GTCAGTTTGATTTTTTATTTGTTC	4,716–4,740	48.3
HIV integrase—second round			
Internal sense	AGTGAATCAGAGTTAGTCAATCAAATAAT	4,091–4,114	53.1
Internal antisense	CCCTACAACCCACAAAGTCAA	4,653–4,673	55.0

avoid dinucleotide contexts as hypermutation targets. The second round reaction was composed of an initial denaturation at 94°C for 7 min and 35 subsequent cycles with constant annealing at 53°C for 45 s and extension at 72°C for 50 s, resulting in a 583-base pair final product.

HA-Yellow gel

The amplified 583-base pair products were subjected to electrophoresis in 1% agarose gels in 0.5×TBE (Tris-Borate-EDTA, Invitrogen, Carlsbad, CA) and stained with 10,000× *GelRed* intercalating agent in water (Biotium, Inc., Qiagen, Germany) for band visualization. To detect hypermutation, HA-Yellow was added to the agarose gels at 7 µl/ml of gel. HA-Yellow dye (Resolve-It Kit, Vector Laboratories, Burlington, CA) contains bisbenzimidazole (Hoechst 33258 or 2-[4-hydroxyphenyl]-5-[4-methyl-1-piperazinyl]-2,5-bis-1H-benzimidazole) covalently bound to polyethyleneglycol.³¹ Bisbenzimidazole has an affinity for the genomic nitrogenous bases adenine (A) and thymine (T) and intercalates into these regions. Due to the resulting high molecular weight of the HA-Yellow composite, where bisbenzimidazole is PEGylated, DNA sequences with a higher genomic content of adenine are retarded in the electrophoretic run proportionally to their adenine content. The genome's A+T content serves as an indicator of hypermutation.¹⁴ Therefore, the migration of hypermutated sequences is slower than that of nonhypermutated sequences. This method thus allows for the identification of hypermutation.

Cloning and sequencing

Amplified fragments from two patients (35789 and B190D) who exhibited multiple bands in gels with HA-Yellow were subjected to bacterial cloning using the TOPO TA Cloning kit (Invitrogen Corporation, Carlsbad, CA). The 538-base pair integrase gene fragment was amplified from the clones, and the product was subjected to a further electrophoretic run in a gel with HA-Yellow. The fragments exhibiting different migration profiles were sequenced. The sequences were edited using Sequencer 4.5 software (Genes Code Corporation, Ann Arbor, MI) and aligned using Clustal X software.³² Twenty clones were generated from sample 35789, and the migration position of each clone was recorded. All of these clones were sequenced and subjected to analysis with Hypermut software.³³

Hypermut is an approach useful to characterize the sequences based in their A+T content and hypermutation target sites. The software was used to generate a panel ex-

hibiting the correlation between the migration patterns in the gel with HA-Yellow and hypermutation.³³ The clones were considered to be hypermutated when the *p*-value estimated by the software was lower than 0.05 compared with clone 35789_27. Clone 35789_27 was used in Hypermut as the nonhypermutated internal reference sequence because it exhibited the greatest relative migration in the gel with HA-Yellow and the lowest A+T content among the clones in sample 35789. The use of a sample internal reference in Hypermut, clones in this case, allowed us to define more precisely which G→A mutations should be considered to result from hypermutation. B190D clones were also obtained, sequenced, and subjected to analysis by Hypermut. Three clones were selected from the B190D sample, subjected to electrophoretic migration within each set of investigated clinical samples, and used as references for hypermutation and nonhypermutated sequences: a hypermutated control (B190D_Hypermutated), an intermediate hypermutated control (B190D_Intermediate), and a nonhypermutated control (B190D_not_Hypermutated).

Hypermutation analysis

Hypermutation of clone samples was assessed by Hypermut software³³ (available at www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html). This software estimates a value for each performed test corresponding to the result of a Fisher's exact test that compares a control sample (which must contain the lowest possible index of G-to-A substitutions) to one or more samples. This software algorithm was used as the gold standard. Hypermut normalizes the number of substitutions occurring within the hypermutation context (G-to-A substitutions) relative to other non-specific substitutions and the number of nucleotides in the sequence.

K-means analysis method

The k-means method³⁴ allowed the Hypermut classification to be correlated with the data collected from the migration of the amplified DNA in the presence of HA-Yellow. First, the images were photodocumented using the GelDoc-It TS Imaging Systems Biolmaging system (UVP, Cambridge, CA). Whenever the clinical samples exhibited more than one viral subpopulation with distinct hypermutation levels, that is, with more than one displacement pattern, each distinguishable band was considered separately. The software *ImageJ*³⁵ (available at www.rsweb.nih.gov/ij/) was used to

measure the distances, in pixels, among the highest light-intensity peaks on each sample. Given that each light-intensity peak corresponds to the highest DNA concentration of the amplified product, the distance between it and the point of sample application measures its displacement in the electrophoresis gel. Hence, the hypermutation analysis was turned into a mathematical problem of clustering sets of Euclidean distances. Furthermore, each gel had its internal controls, because electrophoresis conditions may be influenced by temperature, agarose percentage, and other factors.

There are a number of different clustering methods, each one working better with a different set of data. Our problem is new and we have no previous information on the statistical behavior of our variables, so we opted for the method with the minimum set of assumptions. The *k-means* method³⁴ fits very well, with only two acceptable assumptions: independence of observations and normal distribution of the residuals and errors. The idea behind the method is very simple: for each gel, suppose that its n biological sample displacements are partitioned in k clusters $\{S_1, S_2, \dots, S_k\} (k \leq n)$. Then, among all possible sets of k clusters, look for the one that minimizes the within-cluster sum of squares.

$$WCSS = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

where μ_i is the mean of points in S_i .

It is worth noting that the *k-means* method, widely used in signal processing, presents two strong drawbacks: first, there is no prescription on how to choose the number k of clusters; second, there is no efficient algorithm to calculate the minimum of the sum of squares (it is an NP-hard computational problem). Nevertheless, those drawbacks do not apply in the present analysis. First, we look for exactly two clusters (hypermutated or nonhypermutated) and we do know one reference member in each cluster. These are the hypermutated and normal controls, respectively. Second, as our set of points is small, we can solve the problem by hand, using spreadsheet software. We started the clustering analysis ordering the elements by displacement size, attributing the two largest ones to the hypermutation cluster and the others to the normal cluster. In the next steps, we moved the next larger displacement value from the normal to the hypermutation cluster and continued to do so until we had two elements in the normal cluster. Then we test for the statistical significance of the minimum within-cluster sum of squares. The *k-means* clustering method is a type of ANOVA, with the different groups characterized first, and then each element is analyzed to determine the group to which it belongs. To classify the samples as hypermutated, the maximum likelihood method was adopted to avoid false-positive results. There are a number of statistical software packages that implement *k-means* calculation. In particular, the R statistical package (www.r-project.org) has a *k-means* library that implements a variety of algorithms suitable to solve additional related problems.

Quantification of proviral loads and number of cells in the analyzed clinical samples

HIV proviral loads were measured by qPCR using the HIV-1 integrase gene as the target and the human CCR5 gene as the normalizer. We extrapolated copy numbers from

standard curves generated with plasmids harboring the CCR5 and the HIV integrase gene sequences, as described by Sharkey *et al.* (2000) and Komninakis *et al.* (2012).^{36,37} Briefly, the reaction mixture consisted of 20 μ l of Gene expression master mix (Life Technologies, Carlsbad, CA, and USA Applied Biosystems, Coutabouef, France) 2 $\times$, 0.75 μ M of each oligonucleotide, 0.3 μ M of each probe, 5 μ l of sample, and DNase RNase-free water q.s.p. 20 μ l. Sequences of primers and probes and cycling conditions are available at the above mentioned citations.

Detection of hypermutation in clinical samples

After the samples were classified as hypermutated or not by electrophoretic runs in gels with HA-Yellow and validated by the *k-means* method, the mean, median, and quartiles of the viral load indexes were calculated in the hypermutated and nonhypermutated sample groups. These indexes were compared using Sigma Stat 4 software with a significance level of $p < 0.05$. Characteristics such as sex and CD4⁺ T cell count were also assessed.

Results

Analysis of hypermutation using a gel-based approach with HA-Yellow dye has already been fully validated before with protease sequences. In this context, protease clone sequences with an increasing number of G $\rightarrow$ A substitutions were subjected to electrophoretic runs in agarose gels with HA-Yellow dye in order to determine HA-Yellow's sensitivity and specificity.¹⁴ By the application of this methodology, two patients exhibiting fragments with different hypermutation levels were selected. The amplified sequences from patients 35789 and B190D were subjected to bacterial cloning and then sequenced. A total of 20 clones from patient 35789 and three clones from patient B190D (B190D_Hypermutated, B190D_Intermediate, and B190D_not_Hypermutated) were sequenced. Each clone was assessed in an agarose gel with HA-Yellow (Fig. 1) to characterize its hypermutation level.

This study considered the classification of hypermutation supplied by Hypermut as the standard to define hypermutation. Samples exhibiting a $p < 0.05$ in the above mentioned test were considered to be hypermutated (Table 2). The

FIG. 1. The assessment of migration profiles in an agarose gel with HA-Yellow of the clones generated from patient 35789 and the hypermutation controls from patient B190D.

TABLE 2. SEQUENCE CHARACTERISTICS DESCRIPTION OF 35789 AND B190D CLONES ACCORDING TO HYPERMUT PARAMETERS

Sequence	Fisher exact p-value	#G→A	Dinucleotide context			
			GG	GA	GC	GT
35789_01	0.959167	5	1	1	1	2
35789_02	0.860379	7	2	1	2	1
35789_04	5.12E-14	60	21	35	1	2
35789_05	0.71822	7	3	3	0	1
35789_06	0.591371	7	3	2	0	1
35789_08	0.262761	6	2	2	1	1
35789_09	0.703421	7	2	2	1	2
35789_10	2.05E-06	34	21	11	0	1
35789_13	0.202824	18	9	6	0	2
35789_15	0.71822	7	3	3	0	1
35789_17	5.12E-14	60	21	35	1	2
35789_21	2.57E-06	26	18	5	0	2
35789_25	0.635934	4	1	2	0	1
35789_26	3.82E-05	28	15	10	1	1
35789_28	1.07E-14	61	22	35	1	2
35789_30	0.591371	6	2	2	1	1
35789_31	0.707485	2	1	1	0	0
35789_33	4.04E-15	60	22	35	1	1
35789_35	1.69E-06	26	18	5	0	2
B190D_NOT_HYPERMUTATED	0.542087	9	2	3	1	2
B190D_INTERMEDIATE	0.800222	9	2	2	2	2
B190D_HYPERMUTATED	1.47E-17	58	21	30	1	6

Hypermut results were correlated with the results of the migration of the amplified fragments in agarose gels with HA-Yellow through a k-means clustering analysis. This correlation made it possible to determine which patterns of migration in the gel might be associated with hypermutated samples. Table 2 shows a panel with the values obtained from the Hypermut analysis of the clones from patients 35789 and B190D. Clone 35789_27 of sample 35789 was used as an intrinsic nonhypermutated control in the Hypermut analysis because it exhibited the longest displacement among all of the clones in the gel with HA-Yellow and the lowest A+T content. The number of G→A substitutions is described in the column #G→A. *p*-values lower than 0.05 indicate that the sequences were considered to be hypermutated compared with the control 35789_27. These values are shown in the Fisher exact *p*-value column. The right-most columns, titled "Dinucleotide Context," describe the number of G→A substitutions occurring in every possible dinucleotide (GpG, GpA, GpC, and GpT). Mutations typical of hypermutation and random mutations were observed throughout the sequences.

Analyzing the images of the gels with HA-Yellow using ImageJ software made it possible to measure the position of each sequence in the gel from the origin of the electrophoretic run to the point with the highest luminosity. The point with the highest luminosity corresponds to the region with the highest DNA concentration in the sample and represents the final position of the sequence in the gel for analysis.

Figure 2 indicates how the position of each sample was identified in the gel. Twelve clinical samples and the previously sequenced patient B190D controls containing known hypermutation levels are represented on the left side of Fig. 2 after an electrophoresis run on a gel containing HA-Yellow.

The samples that displaced a shorter distance from the point of origin (samples 03, 07, 09, 11, and 12 and the hypermutated control) exhibit the highest hypermutation levels. The graphs on the right side show the migration position (abscissa) and intensity (ordinate) of the samples. The depressions in the panel correspond to the regions with maximal luminosity. In the graphs, a greater shift of the depressions to the left means a shorter migration of the samples in the gel and, thus, a higher A+T content and hypermutation level.

To validate the k-means technique associated with the HA-Yellow gels, the hypermutated and nonhypermutated controls of sample B190D were used. Thus, it was possible to classify each clone of patient 35789 as either corresponding to the migration pattern of the hypermutated or the nonhypermutated sequences (see Fig. 5). For all of the clones assessed, there was a correlation between Hypermut analysis and the k-means method except for clone 35789_35, which was rated as hypermutated in Hypermut ($p = 2.40E-03$) and borderline by the k-means method (see Fig. 5). However, we successfully validated the correlation between both methods by linear regression, where $R^2 = 0.98$ (data not shown).

After a satisfactory correlation was established between the data from the electrophoretic migration in the gels with HA-Yellow and the k-means clustering and the results of the same 23 cloned sequences in Hypermut, we processed the clinical samples and correlated the results with the associated data.

Each set of clinical samples was PCR amplified and subjected to electrophoretic runs together with the three control clones from sample B190D (Fig. 3). The electrophoretic runs in agarose gels with HA-Yellow samples were then classified as hypermutated or nonhypermutated by the k-means method. Clinical samples were not subjected to sequencing or

FIG. 2. The measurement of the position of each sequence in the gel. Equal reference regions represented by the *boxes* were defined for each sequence using ImageJ. A graph was generated for each sample migration distance from the origin in the abscissa and the intensity of the light signal in the ordinate.

Hypermur software because we were able to validate our study based on HA-Yellow gels and the k-means clusterization method. As shown in Fig. 3, this method identified hypermutation in samples 27, 28, 29, 31, and 33. Figure 3 also shows the different hypermutation levels in samples 27, 31, and 34. The samples rated as borderline in the k-means classification were classified as nonhypermutated to avoid false-positive results since Hypermur was considered the gold standard.

Hypermutation was detected in 49/157 (~31.2%) of the investigated samples, from which 29/157 (18.5%) exhibited more than one viral population with different hypermutation levels. Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com/aid) presents the

demographic and clinical data of the patients and Supplementary Figs S1, S2, and S3 assess the HA-Yellow-based approach for all clinical samples. As mentioned above, samples classified as “borderline” by the k-means method were considered nonhypermutated.

To investigate whether detection of hypermutation could be associated with proviral load values and number of cells present in each clinical sample, we quantified proviral loads by qPCR from the studied samples. We addressed the HIV-1 integrase gene as target and the human CCR5 gene as normalizer. We extrapolated copy numbers from standard curves generated with plasmids harboring the CCR5 and the HIV-1 integrase gene sequences. We were able to analyze 118 (75%) of the total number of samples, with 58 classified as hypermutated and 60 without hypermutation according to HA-Yellow gel/k-means analysis. Mean values for samples with and without hypermutation detection were 303.66 (minimum value = 74.3 – maximum value = 533) $\times 10^6$ cells and 295.38 (minimum value = 117.3 – maximum value = 473.4) $\times 10^6$ cells, respectively. We did not find differences with statistical significance when proviral load values from samples with hypermutation and from samples without hypermutation were compared. *t*-test and ANOVA generated values of $p = 0.8404$ and $p = 0.955$, respectively.

The full case series was composed of 70 females (44.6%) and 87 males (55.4%); hypermutation was detected in 18/87 (20.7%) of the males and in 31/70 (44.3%) of the females, and it was more frequent in the females (Chi-squared test, $p = 0.003$). The mean viral load of the patients was 214,279 (standard deviation 378,583) copies/ml, and the median was 41,400 copies/ml. The detection of hypermutation was higher in patients with higher viral load levels (Mann–Whitney test, $p = 0.035$) (Fig. 4 and Supplementary Table S1).

The overall case series mean of the $CD4^+$ T-lymphocyte count was 366 (standard deviation 279) cells/mm³. In patients

FIG. 3. An agarose gel with HA-Yellow illustrating the results of clinical samples 26 to 36. The *arrow* indicates a sample with viral subpopulations exhibiting different hypermutation levels.

FIG. 4. Statistical analysis comparing the viral load in patients with or without hypermutation. It was performed with the Mann-Whitney test, $p=0.035$.

with samples exhibiting hypermutation, the mean of the T cell count was 308 (standard deviation 256) cells/mm³, and the median was 208 cells/mm³; in samples in which hypermutation was not detected, the mean number of CD4⁺ T lymphocytes was 393 (standard deviation 285) cells/mm³, and the median was 347 cells/mm³. The patients with higher hypermutation levels also exhibited lower mean and median CD4⁺ T-lymphocyte counts; however, this difference was not statistically significant (Fisher's test, $p=0.068$).

Quantitative analysis

The mean viral load was higher in the patients with hypermutation (Mann-Whitney, $p=0.023$). An HIV-1 viral load above 100,000 RNA copies/ml relates to a faster progression to AIDS.³⁸ In our case series, patients were divided into two groups for analysis, based on viral load values higher or lower than 100,000 RNA copies/ml. A total of 100 patients exhibited serum viral load values lower than 100,000 RNA copies/ml, 23% of which exhibited virus with hypermutation; 53 patients exhibited a viral load higher than 100,000 RNA copies/ml, 46.3% of which exhibited virus with hypermutation. The frequency of hypermutation detection was higher in the patients with higher viral load levels (Mann-Whitney test, $p=0.007$). When we compare patients with or without hypermutation detection and their respective viral loads, a correlation was found between higher viral loads and hypermutation detection (Mann-Whitney test, $p=0.035$) (Fig. 4).

In summary, out of a total of 157 investigated patients, hypermutation was detected in 49 (~31%). Females exhibited a higher frequency of hypermutation compared with males (Chi-square, $p=0.003$). A quantitative analysis revealed that the patients with samples in which hypermutation was detected exhibited higher viral load levels (Mann-Whitney test, $p=0.035$) and a nonsignificant tendency toward lower CD4⁺ T-lymphocyte counts.

Discussion

This study was the first to assess hypermutation frequency in a Brazilian naive cohort. A method for hypermutation study was formulated to detect G-to-A substitutions in target sites and allowed k-means clustering analysis of electro-

phoretic runs in HA-Yellow gels. The high sensitivity and specificity of hypermutation detection in the sequences were due to selection of a hypermutation hot spot (viral integrase gene),³⁹ the touchdown approach, and the high magnesium concentration (4.0 mM) in the first round of nested PCR. Experimental standardization showed that higher magnesium chloride concentrations increase the detection of hypermutated sequences.

To validate the method, two clinical samples, B190D and 35789, were selected to be cloned and sequenced because they exhibited viral subpopulations with different hypermutation levels (Fig. 1). The clone sequences were subjected to Hypermut and tested with agarose gels with HA-Yellow dye. The results thus obtained with HA-Yellow gel were correlated with the results of the classification supplied by the Hypermut software (Fig. 5). Through the use of the k-means method, the displacement of the PCR-amplified sequences in the gel with HA-Yellow was perfectly correlated with the classification supplied by the Hypermut software, thus validating systematically the application of this approach to the detection of hypermutation in clinical samples using gels with HA-Yellow. The k-means classification system supplied a robust, reproducible, and rapid method of detecting hypermutation, which is less expensive than sequencing, and thus allowed for the assessment of a larger number of samples.

The results were correlated with the data of the corresponding population (Tables 3 and 4 and Supplementary Table S1). By applying this method to 157 investigated patients, HIV-1 hypermutation was detected in 49 individuals (~31%). Hypermutation detection was higher in females than in males (Chi-Square, $p=0.003$). The patients with samples in which hypermutation of viral sequences was detected exhibited higher viral load levels (Mann-Whitney test, $p=0.007$).

Hypermut detects A → G substitutions within the GpG and GpA contexts of a subgenomic sequence when compared with a control sequence. The proper manner of obtaining classifications in this software is by setting the least hypermutated sample as the control sample, which, as a rule, is the one with the lowest A + T content in the investigated set. When this rule is not observed, errors in classification occur.

FIG. 5. Comparison between the results generated by the Hypermut software and k-means method of classification for clones 35789 and BD190D.

The software itself recommends carefully choosing the reference sequence. With regard to the process of the natural diversification of HIV-1, G→A transitions are the most commonly found,⁴⁰ and although they might occur within the GpG and GpA hypermutation contexts, it is not possible to rule out that these transitions might occur independently of this phenomenon. Thus, it is not impossible for Hypermut, when assessing unrelated viral sequences, to count G→A transitions in HIV-1 natural diversification as typical hypermutation substitutions. Therefore, one approach to minimizing this error is to sequence clones of the same patient that exhibit different patterns of migration in a gel, that is, different hypermutation levels revealed by a gel with HA-Yellow.

The two samples in this study, namely 35789 and B190D, exhibited multiple bands in gels with HA-Yellow, indicating different hypermutation levels in the viral populations, and thus they were cloned (Fig. 1), sequenced, and analyzed using Hypermut software (Table 2). The generation of clones allowed for the identification of a less hypermutated sequence or a sequence without hypermutation, which was used as the sample internal reference. The use of sequences from a sample primarily distinguished for exhibiting different hypermutation levels allows for the minimization of the effect that other substitutions might have on the calculation of hypermutation-specific mutations performed by Hypermut software. Hypermutation is not an erratic process, and it is characterized by a specific accumulation of G-to-A substitutions in a definite context and should not be confounded with the mutational background of a virus such as HIV-1,

which replicates under the influence of a highly diversifying force.

As mentioned above, a direct comparison of the results reported by different studies is often not possible, due to the use of different criteria to analyze sequences and discriminate hypermutation. Thus, this difference in criteria influences the results and conclusions of each study. Hypermut is used as the gold standard for hypermutation classification in a large portion of studies; however, the results generated by this algorithm depend essentially on the reference sequence against which all of the other sequences are compared. The study by Piantadosi *et al.* (2009),²⁵ which found the highest hypermutation frequency reported in the literature up to the present time, used a software-produced sequence known as “most recent common ancestor” (MRCA) as its reference sequence. Although obtained from actual sequences, MCRA is a virtual sequence. Furthermore, most hypermutated genomes are not viable and cannot replicate yielding no descendants.

The use of MRCAs to gauge hypermutation does not seem to be the best approach. The study by Wei *et al.* (2004)⁴¹ used HXB-2 as a reference sequence. Despite being widely known and used, in our opinion, this sequence also fails to provide the most appropriate standard for the investigation of hypermutation in Hypermut since the use of nonrelated reference sequences that are not essentially related to the investigated sequences might underestimate or overestimate the occurrence of hypermutation. Other studies used algorithms other than Hypermut to classify the samples. The studies by Pace *et al.* (2006) and Ulenga *et al.* (2008)^{20,23} estimated hypermutation based on a statistical analysis that,

TABLE 3. POPULATIONAL DATA AND PERCENTAGE OF HYPERMUTATION IN THE STUDIED POPULATION

Sex	Female: 70 (44.6%) Male: 87 (55.4%)		
Viral load	Mean: 214,270 ± 378,583 Median: 41,400	Maximum: 1,980,000 Minimum: 80	Normal distribution $p=0.02$ Kolmogorov–Smirnov
CD4	Mean: 366 ± 279 Median: 328	Maximum: 1,437 Minimum: 2	Normal distribution $p=0.054$ Kolmogorov–Smirnov
Hypermutated sample	Yes: 49 (31.2%) No: 108 (68.7%)		

TABLE 4. COMPARATIVE DATA OF THE STUDIED POPULATION BASED ON DETECTION OF HYPERMUTATION

	Sex	Viral load	CD4
Patients with hypermutation	Males	Mean	Mean
	18 (36.74%)	285,782 ± 437,662	308 ± 256
	Females	Median	Median
	31 (63.26%)	110,000	208
Patients without hypermutation	Males	Mean	Mean
	56 (51.9%)	181,825 ± 343,618	393 ± 285
	Females	Median	Median
	52 (48.1%)	30,550	347

similarly to Hypermut, took into account the number of nucleotides in the sequence, the number of G→A mutations within the hypermutation context (GpG and GpA), the number of occurrences of the second most prevalent mutation in the HIV-1 genome (A→G), and the total number of random mutations related to the reference sequence. Our study selected the sequence with the lowest A+T content as the hypermutation reference and classified samples only after validating the correlation between the migration distance of a PCR product in a gel with HA-Yellow and the Hypermut gold standard algorithm.

The detection method used to locate hypermutated sequences might or might not favor the detection of this type of sequence. The method used in at least two studies mentioned above [Pace *et al.* (2006). and Ulena *et al.* (2008)^{20,23}] was sequencing; however, the detection of minority populations requires a specific approach, such as cloning. In this sense, the study by Kieffer *et al.* (2005)⁴² sought to assess not only the majority but also the minority sequences by sequencing the clones obtained by serial dilution and reported that hypermutation was detected in approximately 9–10% of the total number of proviral particles present in a patient and that this same phenomenon was not detected at the level of viral RNA.

Finally, the threshold defining whether a sequence is hypermutated or not also influences the counting of hypermutated sequences. The study by Land *et al.* (2008)²¹ considered the presence of stop codons in proviral sequences as a marker of hypermutation and did not take into account that stop codons could have been caused by random mutations and that sometimes hypermutation might not result in stop codons but restrict viral replication by other means.

For the present study, we chose to validate the presence of hypermutation based on real sequences obtained from clones of a single patient. Thus, we could subtract the substitutions present in the mutational background of HIV-1 more efficiently and consider only G→A substitutions within the hypermutation context in the assessment of samples with a higher degree of accuracy. Thus, the total correlation between the displacement of a sample in an agarose gel with HA-Yellow and the results obtained with the software Hypermut according to the maximum likelihood method in k-means clustering allowed for the validation of a displacement pattern in an agarose gel with HA-Yellow, distinguishing between hypermutated and nonhypermutated samples (Fig. 1). The samples classified as “borderline” by the k-means method were counted as nonhypermutated to avoid false-positive results. If, instead of the maximum likelihood method, classic considerations were used in the k-means

variance analysis ($p < 0.01$), the number of samples classified as hypermutated would be even higher (52.2%).

The ability of HA-Yellow to detect hypermutation is due to its ability to retard sequences proportionally to their A+T content. Clinical samples of patients who were treated or nontreated, progressing or nonprogressing, and receiving or not receiving antiretroviral therapy in which hypermutation was detected exhibited substitutions in the specific GpG or GpA contexts described^{17,43–47} and, consequently, an increased A+T content in their sequences. In contrast to Hypermut, which selectively discriminates GpG and GpA contexts that are hypermutation targets, HA-Yellow is not selective for hypermutation and might also bind to nonspecific A+T sites. However, no other biological process for increasing the viral A+T content is known other than hypermutation, and for this reason, HA-Yellow can be used to detect hypermutation. In this sense, an important goal of this study was to develop a method that, once validated, no longer requires the sequencing of samples but uses the displacement of sequences in gels to detect hypermutation. Thus, we found a perfectly applicable correlation between two different methods, namely, sequencing and gels with HA-Yellow followed by k-means clusterization.

The k-means method defined a threshold that allowed samples to be classified as hypermutated or nonhypermutated. In contrast, no threshold in the number of G→A substitutions was ever defined to establish the minimum number of substitutions required to classify a sample as hypermutated. Different regions of the HIV-1 genome are known to exhibit a differential susceptibility to hypermutation. Polypurine tracts, for instance, do not suffer hypermutation, whereas viral negative factor (nef) and integrase genes are hypermutation hot spots.^{29,30,48,49} Hypermut itself takes into account characteristics of sequences, such as size, G→A substitutions within the hypermutation context, and the number of guanine bases in the sequence, among others, to infer a p -value based on Fisher's exact test that estimates the probability that a sequence is hypermutated. For instance, if the polypurine tract of a hypermutated sequence was to be subjected to Hypermut analysis because the hypermutation levels tend to be zero in this genomic region, hypermutation would not be detected. Therefore, it is critically important to establish a threshold, and the threshold should be linked to the particular genomic region under investigation.

Concerning the biological and experimental significance of samples that cannot be easily classified as either hypermutated or nonhypermutated, in the present study, these samples corresponded to the clones and samples classified as “borderline” by the k-means method (Fig. 5). As mentioned above, we considered the clones and samples classified as

“borderline” by the k-means method as nonhypermutated. Thus, upon observing the “borderline” sequences, we may not rule out the hypothesis that these samples represent viral populations that suffered partial hypermutation but still successfully integrated into the host genome.

As shown in Fig. 5, with the sole exception of clone 35789_35, all of the clones considered being hypermutated in the Hypermut analysis ($p < 0.05$) received the same classification by the k-means method. This graph further shows that clone 35789_35 exhibits a vertical drag in electrophoresis, and thus we may not rule out that the drag might have placed the highest luminosity point in the region corresponding to the nonhypermutated samples. Therefore, we may not dismiss the hypothesis that analysis by the k-means method classified hypermutated clinical samples as “borderline” and, consequently, gave false-negative results. Considering the frequency of differences between the classifications supplied by both methods as 1/20 (5%), if this difference were extended to the clinical samples, we would have detected hypermutation in up to 37% of the investigated population. However, the fact that the k-means method did not detect hypermutation in one sample does not diminish the reliability of the analysis, whose error is approximately 5%; consequently, this method of classification is valid.

The present study sought to assess the impact that hypermutation might have on the replicative ability of HIV *in vivo* as inferred by the patients' viral loads. For this reason, 157 samples were selected from an antiretroviral treatment-naïve population. Because the HIV-1 integrase region is extremely conserved among the several viral subtypes and it is a hypermutation hot spot, it was selected to sort the hypermutation levels of samples.^{39,50} We chose a hypermutation hotspot as reporter in order to increase our power of detection, allowing us to identify hypermutation even in less represented sequences in viral populations. Because each sample was treated equally and hypermutation controls were present in each electrophoresis run our method could really differentiate samples according to their hypermutation status.

The presence of HIV-1 hypermutated sequences was detected in 31.2% of the analyzed samples. This value is the second highest frequency reported in the literature, following only that of Piantadosi *et al.* (2009),²⁵ who used MRCA as a reference sequence. The use of MRCA further assumes an ancestry relationship between it and the assessed sample sequence. This notion assumes the existence of generations between MRCA and the hypermutated sequence. However, hypermutation is not a process of gradual and generational accumulation of G → A substitutions but an abrupt process resulting in nonviable viral particles that have lost their replicative ability.

If, instead of using the maximum likelihood method, we had used $p < 0.01$ in this study, which is statistically acceptable, we would have encountered a higher frequency of this phenomenon (82/157, corresponding to 52.2%). To avoid false-positive results, we chose a more stringent classification according to Hypermut; however, due to the likelihood method, we observed once our sample 35789_35 was considered hypermutated based on k-means displacement method and not hypermutated according to Hypermut. So, we cannot rule out the possibility that we have underestimated the true number of hypermutated samples.

Other studies suggested that not all of the hypermutated variants are able to integrate into the host genome. Cell enzymes known as DNA glycosylases (e.g., UNG and SMUGs)⁴⁸ are associated with the degradation of uracylated viral cDNA, although this mechanism has not yet been fully elucidated. If, indeed, the effect of these cell DNA repair enzymes is effective, our findings represent a further indication of the fact that the frequency of hypermutation traceable by detection methods based on viral DNA might be lower than the frequency occurring in the intracellular environment.

Because different values of proviral loads could have an impact on the detection of hypermutated sequences, we ran qPCR analysis to measure proviral loads in our clinical samples. We could not find statistical significant differences between proviral loads in hypermutated and nonhypermutated samples, meaning that identification of hypermutation was not biased in our study on the size of the proviral compartment. Due to the difficulty in establishing straight correlations between viral and proviral loads as a result of the differential kinetics of each compartment, we could not speculate about the comparison between these values in our clinical samples.

The method described in this study allowed for the detection of a high number of samples exhibiting hypermutation of the viral sequences (31.2% of our target population). To date, there are no other systematic studies on the Brazilian population surveying the prevalence of hypermutation in this country. We suspect that although the Brazilian target population is genetically quite diversified and might exhibit unique characteristics compared with other and more homogeneous populations with different ethnoracial compositions, the sensitivity of the detection method was possibly critical for the high prevalence of hypermutation found, which, in fact, was higher than that found by most studies.^{14,20,22} Nevertheless, we may not rule out that the frequency of hypermutation in the Brazilian population might be higher than in other populations that have already been investigated.

A study by An *et al.* (2004)²² showed that the H186R polymorphism in African Americans is related to faster progression to AIDS. Based on that finding, it is possible to suggest that classification by skin color should represent a factor in the assessment of hypermutation effects among the different Brazilian racial types, especially when the significance of miscegenation in this country is taken into account. Nevertheless, the factor “skin color” is not believed to be the most accurate attribute to characterize the ethnic origin of the patients in this study because, particularly in Brazil, skin color is not a good predictor of genomic ancestry.⁵¹ Although we did not find an apparent correlation between any ethnic group and the frequency of hypermutation, more specific studies are needed to establish the prevalence of this polymorphism in Brazil and its influence on this country's population regarding progression to AIDS.

The only study performed in Brazil searched for A3G polymorphisms among the Brazilian population among antiretroviral-naïve HIV-1-infected patients. That study detected one polymorphism, namely rs5757463, which in G/G homozygosis or C/G heterozygosis is associated with lower CD4 levels compared with allele C/C carriers.⁵² Another important factor is the method chosen to select the hypermutated

sequences. In studies using sequencing as the method of selection, because it is not the best approach to detect minority populations,⁵³ the sequences these studies considered to be hypermutated were possibly also the most represented sequences in the sample. These studies actually assessed the viral load levels of patients in whom the majority of sequences exhibited hypermutation. Regarding the difference between the present study and other studies, the method we employed assessed short sequences using primers targeted to hypermutated sites to detect hypermutation even when present in few sequences of the viral population. This difference is one of the reasons that we found a high number of hypermutated samples. A further issue concerns the HIV-1 genomic region subjected to analysis. We chose integrase because it is extremely conserved among the different viral strains and is a hypermutation hot spot. Regarding the possible occurrence of hypermutation in both majority and minority viral populations, our study design allowed for the tracing of short sequences of the viral genome known to be hypermutation targets,³⁹ and there was no need to assess the full genome because the amplification of integrase sufficed to trace hypermutation. Therefore, we employed a PCR specifically modified to ensure high sensitivity and specificity for hypermutated sequences, and thus it was able to detect hypermutation at all possible levels whenever it was present (Table 2).

Finally, based on the validation of the analysis of the migration of amplified sequences in agarose gels with HA-Yellow by the k-means method, the sequencing of the amplified material to detect hypermutation was not needed because detection was possible based on the migration distance of the samples from the point of application.

The analysis of the migration of the amplified fragments of the clinical samples in the gels with HA-Yellow revealed the presence of more than one viral subpopulation in the sample. Figure 3 shows that samples 27, 31, and 34 contain viral populations exhibiting different hypermutation levels in a single patient. HIV-1 is known to suffer the influence of host factors^{54,55} and it is likely that hypermutation is also a viral event. Consequently, these findings imply that one single clinical sample might exhibit viral subpopulations that are more or less susceptible to hypermutation. Several host factors are known to influence the progression of the disease, such as the APOBEC3G H186R polymorphism, which in the African American population, is associated with rapid progression to AIDS; haplotypes htSNP 6 and ctSNP 5G present in CUL5 in this same population, which are associated with a higher CD4 depletion rate in HIV-1-infected individuals⁵⁶; and the APOBEC3G rs5757463 polymorphism, which in the Brazilian population is associated with lower CD4 levels.⁵² Although innate immunity is regulated by the host, we observed that hypermutation might affect viral subpopulations in a differentiated manner. It is estimated that 7 ± 4 APOBEC3G molecules are incorporated into each HIV-1 viral particle.⁴⁹

As far as viral diversity, the Brazilian HIV-1 epidemic is known to be genetically diverse. Several HIV-1 subtypes and recombinant strains circulate in Brazil with subtype B the most prevalent. Most of the HIV-1 subtyping done in Brazil has addressed the *pol* region of the viral genome. Because of this we sought to study the genetic diversity of *vif* in a set of 214 samples containing the same 157 samples used in the study presented here. The HIV-1 virion infectivity factor

(Vif) is required during viral replication to inactivate the host cell antiviral factor, APOBEC. Phylogenetic analysis characterized 165 (77.1%) samples as subtype B, 27 (12.6%) as subtype F, 14 (6.5%) as BF recombinants, and 8 (3.8%) as subtype C. The prevalence of *vif* subtypes was similar to what has been observed for *pol* in the Brazilian epidemic.⁵⁷

With regard to the demographic data (Table 3 and Supplementary Table S1) and gender, in our case series, the male:female ratio was 1:0.8. The mean total viral load was 214,270 viral RNA copies/ml (median 41,400 copies/ml), and the mean CD4⁺ count was 366 cells/mm³ (median 328 cells/mm³). Upon including the data on hypermutation, the prevalence of hypermutation was higher in females because, from the total number of samples in which hypermutation was detected, 63.26% corresponded to females (Chi-square, $p=0.003$) (Table 4). This study is the first to detect a higher prevalence of hypermutation in females, and further studies are needed to better elucidate this finding.

Interestingly, patients with samples in which hypermutation was detected exhibited lower numbers of CD4⁺ cells (308 ± 256) compared to patients with samples in which hypermutation was not detected (393 ± 285). Although this difference was not statistically significant (Fisher's exact test, $p=0.068$), it represented a trend and may suggest that hypermutation is a cumulative process becoming more evident as the disease progresses.

The basal viral load is a biological marker of HIV replication in which levels correlate to the pace of disease progression, and levels above 100,000 RNA copies/ml are associated with rapid progression to AIDS.³⁸ Surprisingly, the patients in this study with samples in which hypermutation was detected exhibited higher mean ($285,782 \pm 437,662$) and median viral load levels compared to patients with samples in which hypermutation was not detected ($181,825 \pm 343,618$) (Table 4). Because our results differed from others published in the literature, we divided our case series into patients with viral loads above and below 100,000 RNA copies/ml. Nevertheless, the detection of hypermutation was higher in the patients with higher viral load levels (Mann-Whitney test, $p=0.007$). Due to the high mean standard deviation of viral load we performed another statistical test. We divided the patients in two groups according to the presence or absence of hypermutation detection. It was found that patients with hypermutation presented higher viral loads (Mann-Whitney test, $p=0.035$). Therefore, our results strongly suggest that hypermutation is linked to persistently high levels of viral replication. Hypermutated proviruses from elite suppressors were found at a frequency that was not significantly different from that observed in patients on highly active antiretroviral therapy with suppressed virus replication⁴⁷ and RNA hypermutated virus sequences are not found.^{15,17} However, hypermutated *vif* could be easily detected in long-term nonprogressors possibly because of the lack of superimposing replication-competent virus sequences.⁵⁸

Thus, it is conceivable that hypermutation exhibits a higher frequency when viral replication is also higher. Our method is highly sensitive; thus, we cannot rule out the possibility that the detection of hypermutation traced a minority portion of viral particles that do not contribute to total viral replication. However, we cannot rule out the possibility of a correlation among higher viral load values, higher replication, and higher frequencies of hypermutation.

Our strategy did not require sequencing and was carefully based on the analysis of the migration profiles of amplified samples in gels with HA-Yellow and the k-means classification method. This data analysis allowed us to identify a higher frequency of hypermutation in females ($p=0.003$) than in males. More studies are needed to determine if hypermutation is gender related. Furthermore, we were able to establish a direct correlation between higher viral replication levels and the presence of hypermutation ($p=0.035$), as well as a tendency toward decreased CD4⁺ T cell levels ($p=0.068$) and hypermutation. We have yet to consider that we studied only one viral load measurement of a naive cohort that probably presents heterogeneous AIDS progression, and other studies, such as those by Kourteva *et al.* (2012) and Ulena *et al.* (2008),^{23,24} considered two or more groups based on their different rates of AIDS progression.

Although our results differ from other published results, based on these findings, we can suppose that the accumulation and production of hypermutation occur progressively in the course of HIV-1 infection and increase when HIV replication is more intense. Based on the fact that viral load is a viral replication marker and CD4⁺ T cell levels predict clinical outcome but also work as a time marker of infection, it is expected that viruses under replication processes or longer duration infections present a higher accumulation of sequences with hypermutation.^{59,60} Based on these findings, we cannot rule out the possibility that hypermutation might eventually act as a factor that correlates with the generation of viral genetic diversity because viruses under the influence of immunological pressure or drugs escape in an attempt to produce resistant varieties and achieve immunological escape.⁶¹

Our interpretation of the set of findings in this study indicates that hypermutation is a pervasive process that can be associated with high levels of viral replication, and we cannot rule out that the viral load values detected in an individual represent the number of viral particles that escaped from the hypermutation process. In the same vein, we could suppose that for hypermutation to occur, viral replication is a requirement. Thus, the occurrence of hypermutation depends upon viral replication and might also result from viral replication. In this sense, it is conceivable that the mediated innate immunity of APOBECs might not be a determinant against HIV, and hypermutation might not be a relevant marker of restriction to disease progression.

Acknowledgments

This study was supported by the FAPESP research grant (2007/50650-0) and by the Infectious Diseases Department of the Federal University of São Paulo (UNIFESP).

Author Disclosure Statement

No competing financial interests exist.

References

- Rathore A, Chatterjee A, Yamamoto N, and Dhole TN: Absence of H186R polymorphism in exon 4 of the APOBEC3G gene among North Indian individuals. *Genet Test* 2008;12:453–456.
- Mansky LM and Temin HM: Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 1995;69:5087–5094.
- Eigen M: Error catastrophe and antiviral strategy. *Proc Natl Acad Sci USA* 2002;99:13374–13376.
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, and Ho DD: HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996;271:1582–1586.
- Li Y, Hui H, Burgess CJ, *et al.*: Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: Evidence for limited defectiveness and complementation. *J Virol* 1992;66:6587–6600.
- Johnston R: HIV cure: Controversy, consensus, and a consortium. *AIDS Res Hum Retroviruses* 2010;26:943–946.
- Sheehy AM, Gaddis NC, Choi JD, and Malim MH: Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002;418:646–650.
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, and Trono D: Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 2003;424:99–103.
- Gabuzda DH, Lawrence K, Langhoff E, *et al.*: Role of vif in replication of human immunodeficiency virus type 1 in CD4⁺ T lymphocytes. *J Virol* 1992;66:6489–6495.
- Gaddis NC, Chertova E, Sheehy AM, Henderson LE, and Malim MH: Comprehensive investigation of the molecular defect in vif-deficient human immunodeficiency virus type 1 virions. *J Virol* 2003;77:5810–5820.
- Ribeiro AC, Maia e Silva A, Santa-Marta M, *et al.*: Functional analysis of Vif protein shows less restriction of human immunodeficiency virus type 2 by APOBEC3G. *J Virol* 2005;79:823–833.
- Opi S, Kao S, Goila-Gaur R, *et al.*: Human immunodeficiency virus type 1 Vif inhibits packaging and antiviral activity of a degradation-resistant APOBEC3G variant. *J Virol* 2007;81:8236–8246.
- Julias JG and Pathak VK: Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J Virol* 1998;72:7941–7949.
- Janini M, Rogers M, Birx DR, and McCutchan FE: Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J Virol* 2001;75:7973–7986.
- Mbisa JL, Barr R, Thomas JA, *et al.*: Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. *J Virol* 2007;81:7099–7110.
- Armitage AE, Deforche K, Chang CH, *et al.*: APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete “all or nothing” phenomenon. *PLoS Genet* 2012;8:e1002550.
- Russell RA, Moore MD, Hu WS, and Pathak VK: APOBEC3G induces a hypermutation gradient: Purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA. *Retrovirology* 2009;6:16.

18. Fourati S, Lambert-Niclot S, Soulie C, *et al.*: HIV-1 genome is often defective in PBMCs and rectal tissues after long-term HAART as a result of APOBEC3 editing and correlates with the size of reservoirs. *J Antimicrob Chemother* 2012;67:2323–2326.
19. Fourati S, Malet I, Lambert S, *et al.*: E138K and M184I mutations in HIV-1 reverse transcriptase co-emerge as a result of APOBEC3 editing in the absence of drug exposure. *AIDS* 2012;26:1619–1624.
20. Pace C, Keller J, Nolan D, *et al.*: Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J Virol* 2006;80:9259–9269.
21. Land AM, Ball TB, Luo M, *et al.*: Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. *J Virol* 2008;82:8172–8182.
22. An P, Bleiber G, Duggal P, *et al.*: APOBEC3G genetic variants and their influence on the progression to AIDS. *J Virol* 2004;78:11070–11076.
23. Ulena NK, Sarr AD, Hamel D, Sankale JL, Mboup S, and Kanki PJ: The level of APOBEC3G (hA3G)-related G-to-A mutations does not correlate with viral load in HIV type 1-infected individuals. *AIDS Res Hum Retroviruses* 2008;24:1285–1290.
24. Kourteva Y, De Pasquale M, Allos T, McMunn C, and D'Aquila RT: APOBEC3G expression and hypermutation are inversely associated with human immunodeficiency virus type 1 (HIV-1) burden in vivo. *Virology* 2012;430:1–9.
25. Piantadosi A, Humes D, Chohan B, McClelland RS, and Overbaugh J: Analysis of the percentage of human immunodeficiency virus type 1 sequences that are hypermutated and markers of disease progression in a longitudinal cohort, including one individual with a partially defective Vif. *J Virol* 2009;83:7805–7814.
26. Cho SJ, Drechsler H, Burke RC, Arens MQ, Powderly W, and Davidson NO: APOBEC3F and APOBEC3G mRNA levels do not correlate with human immunodeficiency virus type 1 plasma viremia or CD4+ T-cell count. *J Virol* 2006;80:2069–2072.
27. Reddy K, Winkler CA, Werner L, *et al.*: APOBEC3G expression is dysregulated in primary HIV-1 infection and polymorphic variants influence CD4+ T-cell counts and plasma viral load. *AIDS* 2010;24:195–204.
28. Sanabani S, Kleine Neto W, Kalmar EM, Diaz RS, Janini LM, and Sabino EC: Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. *Infect Genet Evol* 2006;6:368–377.
29. Kijak GH, Janini LM, Tovanabutra S, *et al.*: Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. *Virology* 2008;376:101–111.
30. Kijak GH, Janini LM, Tovanabutra S, *et al.*: HyperPack: A software package for the study of levels, contexts, and patterns of APOBEC-mediated hypermutation in HIV. *AIDS Res Hum Retroviruses* 2007;23:554–557.
31. Muller W, Hatteshohl I, Schuetz HJ, and Meyer G: Polyethylene glycol derivatives of base and sequence specific DNA ligands: DNA interaction and application for base specific separation of DNA fragments by gel electrophoresis. *Nucleic Acids Res* 1981;9:95–119.
32. Thompson JD, Gibson TJ, and Higgins DG: Multiple sequence alignment using ClustalW and ClustalX. *Curr Protocall Bioinform* 2002;Chapter 2:Unit 2.3.
33. Rose PP and Korber BT: Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. *Bioinformatics* 2000;16:400–401.
34. MacQueen J: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967;1:281–297.
35. Schneider CA, Rasband WS, and Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671–675.
36. Sharkey ME, Teo I, Greenough T, *et al.*: Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. *Nat Med* 2000;6:76–81.
37. Komninakis SV, Santos DE, Santos C, Oliveros MP, Sanabani S, and Diaz RS: HIV-1 proviral DNA loads (as determined by quantitative PCR) in patients subjected to structured treatment interruption after antiretroviral therapy failure. *J Clin Microbiol* 2012;50:2132–2133.
38. Egger S, Petoumenos K, Kamarulzaman A, *et al.*: Long-term patterns in CD4 response are determined by an interaction between baseline CD4 cell count, viral load, and time: The Asia Pacific HIV Observational Database (APHOD). *J Acquir Immune Defic Syndr* 2009;50:513–520.
39. Suspene R, Rusniok C, Vartanian JP, and Wain-Hobson S: Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucl Acids Res* 2006;34:4677–4684.
40. Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, and Wain-Hobson S: HIV-1 isolates are rapidly evolving quasispecies: Evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* 1989;2:344–352.
41. Wei M, Xing H, Hong K, *et al.*: Biased G-to-A hypermutation in HIV-1 proviral DNA from a long-term non-progressor. *AIDS* 2004;18:1863–1865.
42. Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, and Siliciano RF: G → A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J Virol* 2005;79:1975–1980.
43. Vartanian JP, Meyerhans A, Sala M, and Wain-Hobson S: G → A hypermutation of the human immunodeficiency virus type 1 genome: Evidence for dCTP pool imbalance during reverse transcription. *Proc Natl Acad Sci USA* 1994;91:3092–3096.
44. Martinez MA, Vartanian JP, and Wain-Hobson S: Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc Natl Acad Sci USA* 1994;91:11787–11791.
45. Wain-Hobson S, Sonigo P, Guyader M, Gazit A, and Henry M: Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* 1995;209:297–303.
46. Dinoso JB, Kim SY, Siliciano RF, and Blankson JN: A comparison of viral loads between HIV-1-infected elite suppressors and individuals who receive suppressive highly active antiretroviral therapy. *Clin Infect Dis* 2008;47:102–104.
47. Gandhi SK, Siliciano JD, Bailey JR, Siliciano RF, and Blankson JN: Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. *J Virol* 2008;82:3125–3130.
48. Sousa MM, Krokan HE, and Slupphaug G: DNA-uracil and human pathology. *Mol Aspects Med* 2007;28:276–306.

49. Xu H, Chertova E, Chen J, *et al.*: Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions. *Virology* 2007;360:247–256.
50. Skinner LM, Lamers SL, Sanders JC, Eyster ME, Goodenow MM, and Katzman M: Analysis of a large collection of natural HIV-1 integrase sequences, including those from long-term nonprogressors. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998;19:99–110.
51. Pimenta JR, Zuccherato LW, Debes AA, *et al.*: Color and genomic ancestry in Brazilians: A study with forensic microsatellites. *Hum Hered* 2006;62:190–195.
52. Bizinoto MC, Leal E, Diaz RS, and Janini LM: Loci polymorphisms of the APOBEC3G gene in HIV type 1-infected Brazilians. *AIDS Res Hum Retroviruses* 2011;27:137–141.
53. Schuurman R, Demeter L, Reichelderfer P, Tijnagel J, de Groot T, and Boucher C: Worldwide evaluation of DNA sequencing approaches for identification of drug resistance mutations in the human immunodeficiency virus type 1 reverse transcriptase. *J Clin Microbiol* 1999;37:2291–2296.
54. Chang J, Naif HM, Li S, Sullivan JS, Randle CM, and Cunningham AL: Twin studies demonstrate a host cell genetic effect on productive human immunodeficiency virus infection of human monocytes and macrophages in vitro. *J Virol* 1996;70:7792–7803.
55. Naif HM, Li S, Alali M, *et al.*: Definition of the stage of host cell genetic restriction of replication of human immunodeficiency virus type 1 in monocytes and monocyte-derived macrophages by using twins. *J Virol* 1999;73:4866–4881.
56. An P, Duggal P, Wang LH, *et al.*: Polymorphisms of CUL5 are associated with CD4+ T cell loss in HIV-1 infected individuals. *PLoS Genet* 2007;3:e19.
57. Simão LNAW, Zukurov JPL, Bueno J, Bizinoto MCS, Nunes ERM, Rodrigues F, Soane MM, Lima ML, Diaz RS, and Janini LM: APOBEC3G-related motifs and subtype analysis of human immunodeficiency virus type 1 Vif sequences in Sao Paulo, Brazil. *Brasília: Virus Rev Res* 2009;14.
58. Simon V, Zennou V, Murray D, Huang Y, Ho DD, and Bieniasz PD: Natural variation in Vif: Differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* 2005;1:e6.
59. Chaudhary M, Kashyap B, Gautam H, Saini S, and Bhalla P: Use of surrogate markers to predict the HIV disease stage and time to initiate antiretroviral therapy in developing countries. *J Int Assoc Physicians AIDS Care (Chic)* 2008;7:259–264.
60. Mellors JW, Rinaldo CR Jr, Gupta P, White RM, Todd JA, and Kingsley LA: Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996;272:1167–1170.
61. Smith RJ, Okano JT, Kahn JS, Bodin EN, and Blower S: Evolutionary dynamics of complex networks of HIV drug-resistant strains: The case of San Francisco. *Science* 2010;327:697–701.

Address correspondence to:

Luiz Mario Ramos Janini
Universidade Federal de São Paulo
Rua Pedro de Toledo, 781, 16º andar
Vila Clementino
São Paulo, SP 04039032
Brazil

E-mail: janini@unifesp.br