Variability in HIV-1 partial genomic sequences in Costa Rican patients: analysis with different bioinformatics tools

Lizeth Taylor-Castillo,[1] *María Paz León-Bratti,*[2] *Antonio Solano-Chinchilla,*[3] *Gisela Herrera-Martínez,*[3] *Ricardo Boza-Cordero,*[4] *Bernal León,*[5] *Ronald B. Luftig,*[5] *and Kirsten Visoná*[5]

Suggested citation Taylor-Castillo L, León-Bratti MP, Solano-Chinchilla A, Herrera-Martínez G, Boza-Cordero R, León B, et al. Variability in HIV-1 partial genomic sequences in Costa Rican patients: analysis with different bioinformatics tools. Rev Panam Salud Publica. 2010;27(1):23–31.

ABSTRACT

Objective. *To estimate subtype and genomic variability in the HIV* pol *gene of Costa Rican patients by using different bioinformatics tools and to use this information to establish new policies to better manage these patients.*

Methods. *A total of 113* pol *sequences available from Costa Rican patients under highly active antiretroviral therapy were analyzed by using the Genotyping, REGA, Stanford, and MEGA programs. The* pol *sequences came from 77 virologic failures (VF) and 36 basal samples (BS). Of the 77 VF, 22 also were sequenced in the* env *region.*

Results. *No major differences were found among the variables studied. However, there was a tendency for more variability in VF patients with a high baseline viral load. In the* pol *gene, 75%–83% of BS and 66%–75% of VF samples were pure B subtype by Genotyping and REGA, respectively. The other samples presented variations related mainly to circulating recombinant form CRF12 by genotyping or to CRF17 or –29 by phylogenetic analysis or a new possible BD recombinant with all programs. In the Stanford program, all variable samples showed a subtype B with high polymorphism. The variability in the* env *sequences was lower than that in the* pol *region.*

Conclusion. *The B subtype is predominant in Costa Rican HIV-positive patients. There is high variability within sequences with potential recombination between B and F or D subtypes. The BD recombinant has not been previously reported. This high variability is likely the result of possible recombinant events, nonadherence to antiretroviral therapy, sexual intercourse without protection, and many sexual partners. Similar studies should be done in other countries in the Region, in particular in those places with extensive immigration, in order to decrease the possibility of virus variability as well as the cost of antiretroviral therapy.*

Key words HIV-1; sequence analysis; recombination, genetic; genetic variation; software; Costa Rica.

[1] Centro de Investigación en Enfermedades Tropicales, Microbiology Faculty, University of Costa Rica, San Pedro Montes de Oca, Costa Rica. Send correspondence and reprint requests to: Lizeth Taylor-Castillo, Microbiology Faculty, Virology Section, University of Costa Rica, Ciudad Universitaria Rodrigo Facio, San Pedro Montes de Oca, Costa Rica; e-mail: MAYRA.TAYLOR@ucr.co.cr.

[2] México Hospital, Social Security, Medicine Faculty University of Costa Rica, San José, Costa Rica.

[3] Calderon Guardia Hospital Social Security, San José, Costa Rica.

[4] San Juan de Dios Hospital, Social Security, Medicine Faculty University of Costa Rica, San José, Costa Rica.

[5] Louisiana State University, International Center for Medical Research and Training, San José, Costa Rica.

In Latin America, there were 100 000 new HIV infections and 58 000 AIDS deaths during 2007. The number of people living with HIV is around 1.6 million, and more than 460 000 need antiretroviral therapy (ART) (1). Of them, more than 15 000 (15 000–50 000) live in Central

America, where approximately 14 000 patients were receiving ART in 2007 (2, 3). Each country has its own criterion to enroll patients for ART as well as different procedures for follow-up. The detection of resistant mutants was initiated in Costa Rica during the year 2000; other countries in the region are in the process of implementation (1).

Different guidelines and studies have established that determining the level of CD4 T cells and HIV viral load are the ideal criteria to enroll a patient in ART (4, 5). However, in several developing countries both methods are not available because of the high cost; therefore, only CD4 count is used as the criterion for treatment initiation. In many patients, the basal viral load could be very high, which decreases the possibility of obtaining undetectable levels of virus after ART initiation, with a concomitant greater risk of viral relapse. This situation could also lead to increases in viral diversity because of enhanced replication as well as the possibility of recombination and the emergence of resistant mutants (6, 7). In contrast, in developed countries most patients start treatment according to ideal criteria, which permit control of viral diversity and of the spread of HIV.

Genetic diversity and subtype determination are best done by sequencing the full-length genome. However, this procedure is very costly and time-consuming (8, 9). Partial genomic sequencing with commercial kits allows us to determine resistant mutants of HIV (10–12) and provides an opportunity to analyze sequenced materials in order to study HIV variability and the possibility of detecting recombinants in different geographic areas using several software packages. Sequence analysis of *env* and *pol* together is a frequent choice for determining viral heterogeneity (13–16).

HIV genetic variability leads to quasispecies, which are formed as a result of the highly error-prone reverse transcriptase (RT) enzyme and the ability for RNA/DNA to recombine inside infected cells during retrotranscription and replication (6–8). During genome replication, crossover recombination may occur ranging from three to nine times for each replication cycle, leading to intrasubtype recombinants (8, 9, 17, 18).

A second type of recombination, known as intersubtype recombination, originates through the existence of different subtypes in cocirculation. There are at least 9 subtypes for HIV-1 and 43 circulating recombinant forms (CRFs) reported throughout the world as well as a large variety of equally important unique full-length recombinant forms that have been described. This diversity in the HIV-1 population allows variants to escape from host immunity and also has an impact on ART (6, 7, 9, 17–21).

In Latin America, HIV-1 subtype B is predominant, but CRFs such as CRF12 recombinant BF have been described in Brazil, Argentina, Cuba, and Uruguay (8, 13, 16, 19, 22, 23). In Costa Rica and other Central American countries, there are no reports describing the distribution of HIV-1 subtypes or recombinant forms. In a study with patients receiving treatment at Costa Rica's Social Security System (Caja Costarricense de Seguro Social, CCSS) where more than 90% of Costa Rica's HIV patients needing treatment receive it, it was determined that virologic failure (VF) in ART was due to HIV-resistant mutants, and the appearance of these mutants was related to lack of adherence to ART or high viral load at the initiation of treatment (12).

We took advantage of the sequences available in Costa Rica to determine HIV-1 variability in the country and to analyze the material with existing software. These results have provided useful information for better management of patients and improved intervention policies.

MATERIALS AND METHODS

Study population

A total of 113 samples from patients infected with HIV were collected between 1997 and 2004 at the CCSS in Costa Rica; 77 samples were from patients receiving highly active ART (HAART) with VF and 36 basal samples (BS) were from patients before starting treatment. VF was defined as an increase in viral load from < 50 copies/milliliter (mL) to $\geq 1 \times 10^3$ copies/mL in two consecutive samples in patients receiving HAART. A blood sample was collected from each patient (VF and BS) and the plasma was referred to the International Center of Medical Research and Training and kept frozen below –40 °C until analyzed. All 113 samples (VF and BS) were sequenced with a TruGene HIV-1 genotyping kit (Visible Genetics, Toronto, Ontario, Canada) to detect resistant mutants of HIV RT and protease (PR) (12). Additionally, 22 samples from the VF patients were sequenced in the *env* region V3 loop.

A group of 68 HIV-infected individuals with HAART and without VF were used as controls; they were matched in age and period of treatment.

Data collection

Each patient signed a consent form and the following information was obtained from the patient's clinical record: sex, age, place of residence, risk behavior (men who have sex with men, heterosexual, intravenous drug users), occupation (professional, technical businessman, manual worker, housewife), infection acquired in Costa Rica (yes or no), number of sexual partners, sexual intercourse with or without a condom, baseline viral load, actual viral load in VF patients, adherence (good or bad; good adherence was defined as more than 90% compliance with the prescribed regimen), time under treatment, patient progression (good, VF, change of treatment), and number of previous treatment changes.

RNA extraction, RT–polymerase chain reaction (PCR), and sequencing

RNA was extracted from 1 mL of previously centrifuged (23 500 × *g* for 1 h at 8 °C) plasma samples, using a silica gel system QIAamp viral RNA minikit (Qiagen Inc., Valencia, California, and Hamburg, Germany). RT-PCR of the *pol* gene of HIV was performed and then sequenced for both the RT and PR regions of interest with the TruGene HIV-1 genotyping kit (Visible Genetics, Atlanta, Georgia, United States of America) according to the manufacturer's instructions. Sequence analysis was performed with the Gene Open System and the VG Gene Librarian HIV Module, version 3.1.6 (Visible Genetics, Atlanta, Georgia, United States of America).

For the *env* region, a 325-nucleotide (nt) region spanning the V3 loop was amplified and sequenced. The primers were derived from HIV-1 strain HXB2 (outer primer bases 6953 to 6971, 5′-acagtacaatgtacacatgg-3′, and bases 7749 to 7715, 5′-ccactcttctctttgccttggtggg-3′; inner primer bases 7006 to 7025, 5′-/Cy5/aaatggcagtctagcagaag-3′, and bases 7336 to 7317 5′/Cy5.5/aatttctgggtcccctcctg-3′). RT-PCR of the *env* gene was performed under the

following conditions: 40 microliters (µL) of master mix with 0.1 micromolar (µM) each primer, 200 units of SuperScript II RT (Gibco, Carlsbad, California, United States of America), 1× Superscript II buffer, 2.5 units of Taq polymerase (Promega, California, USA), 1× Taq buffer (Promega, San Luis Obispo, California, United States of America), 2.5 millimolar (mM) $MgCl_2$, 0.2 mM each dNTP, and 10 µL of DNA template. The mixture was run in a cycler (2400 Perkin Elmer, Waltham, Massachusetts, United States of America) as follows: 95 °C for 5 min, 42 °C for 55 min, 70 °C for 10 min, 95 °C for 5 min; 45 cycles at 95 °C for 50 s, 55 °C for 50 s, and 72 °C for 90 s; and one extension of 7 min at 72 °C. After that, sequencing was performed according to the Cy5/Cy 5.5 sequencing kit protocol (Visible Genetics, Bayer, Frankfort, Germany).

HIV-1 sequence analysis

All sequences obtained from the RT-PR and *env* genes were then used to determine the subtype, using database information from the following websites: National Center for Biotechnology Information (NCBI) Genotyping tool (15, 24), REGA HIV subtyping tool (25), and Stanford HIV-1 drug-resistance database (26). RT-PR sequences were analyzed using the RIP HIV sequence database tool (27) to confirm the presence of possible recombinants.

Sequence alignments were done with the BioEdit program (28) and phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4 (29) and the PhyML HIV sequence database tool (30).

Statistical analysis

Analysis of different variables was carried out using SPSS software. The chi-square test with $\alpha < 0.05$ was defined as significant. A logistic regression analysis was used to identify which variables were associated with VF and sequence variability.

RESULTS

The patients in this study were predominantly males (> 80%): 60.0% in the BS group and 74.1% in the VF group; 61.0% of the patients in the control group were men who had sex with men. In all groups, more than 80% of the participants were less than 50 years old, most lived in San José—the capital of Costa Rica—and half of the population were manual workers (housewives included); fewer than 23% were professionals. Good adherence was determined in only 13.6% of the VF group versus 98% in the control group ($P < 0.001$). In 53 patients, 23 (43.4%) have been in treatment more than 2 years, 9 of them for more than 5 years (Table 1). Of the 45 patients for whom information was available, 10 (22.2%) were infected outside of Costa Rica, half of them showed variability in their sequences without significant association, 23 had between two and five sexual partners, and 22 had more than five sexual partners. Only 14 of 33 patients (42.4%) used condoms in sexual intercourse regularly.

In 71 of the total patients analyzed, where the therapy scheme could be evaluated, 87.4% had two or more treatment changes, and only 22.5% had good adherence according to the physicians' evaluation. No significant difference was found between the number of changes of treatment and increases in sequence variability.

Analysis of *pol* gene sequences was performed using three different programs: Genotyping (24), REGA HIV subtyping tool (25), and the Stanford HIV-1 drug-resistance database (26) (Table 2). As there were some differences with the analysis using Genotyping programs, and those using pure references or pure plus CRF references (Figure 1), it was decided to perform the analysis using only the pure sequences. There are different proportions of B subtypes in BS and VF samples with both programs. Variation was detected in the sequences as follows: 27 (75.0%) and 30 (83.3%) BS and 51 (66.2%) and 58 (75.3%) VF samples were classified as B pure subtype by Genotyping and REGA, respectively. In BS and VF samples, the most frequent variable noted is an association with CRF12 (BF) in 13.8% and 11.7% of the sequences, respectively, using the Genotyping program. Figure 1 shows the variability found at the 3′ and 5′ ends of a sample with this program, using the CRF references; however, only the 5′ variability was confirmed using the pure reference sequences. In the REGA analysis, this sample showed a 5′ end matching with F1, a low bootstrap between 600 and 800 nt, and a best matching with subtype B at the 3′ end.

The REGA program showed possible recombination with the D subtype as the most frequent in all the analyzed samples (18.6%) (13.8% in BS and 20.7% in VF) (Table 2). This variation was also detected by the Genotyping program in 10% of VF samples, but none of these values showed any significant difference between the BS and VF groups. One of the VF samples was classified as a possible CRF12 or BF recombinant by both methods (Figure 1) and it showed 90%–93% association with a B subtype by using the Stanford program. Most of the samples classified as B pure by Genotyping showed a bootscan ≥ 95% in the Stanford program, with the RT region showing a bigger variation than the PR region (Table 2). Analysis of these sequences with the RIP program did not

TABLE 1. Characteristics of study populations, Costa Rica, 1997–2004

Variable	BS[a]		VF[b] samples	
	No.	%	No.	%
Male	30	83.3	67	87.0
Female	6	16.6	10	12.9
18 to 50 years old	25/31	80.6	57/66	86.4
Living in San Jose	15/20	75.0	29/48	60.4
Men who have sex with men	18/30	60.0	43/58	74.1
Professional	3/20	15.0	11/48	22.9
Manual worker	10/20	50.0	23/48	47.9
Low adherence	None		51/59	86.4
Basal viral load ≤ 10^4	7/33	21.2	15/66	22.7
Basal viral Load > 10^4	26/33	78.8	51/66	77.3
≤ 2 years under treatment	NA[c]		30/53	56.6
> 2 years under treatment	NA[c]		23/53	43.4

[a] Basal samples.
[b] Virologic failure.
[c] Not applicable.

TABLE 2. HIV-1 subtypes of 36 basal samples (BS) and 77 virologic failure samples (VF) in *pol* gene using three different informatics tools, Costa Rica, 1997–2004

	Genotyping						Stanford[b]			
	BS		VF		REGA[a]		BS > 95% bootscan		VF > 95% bootscan	
Subtype	No.	%	No.	%	BS	VF	PR[c]	RT[d]	PR[c]	RT[d]
B[e]	27	75.0	51	66.2	22 B	44 B, 1BG	20/22	11/22	31/45	25/45
					5 BD	6 BD	4/5	4/5	3/6	4/6
BF RT[d]	5	13.8	9	11.7	1 BF, 4 B	1 BF, 8 B	5/5	0/5	4/9	2/9
BF PR[c] RT			1	1.3		1 BD			0/1	0/1
BD PR-RT	2	5.5	8	10.4	2 B	8 BD	1/2	0/2	3/8	1/8
BC RT	1	2.8	2	2.6	1 B	1BD, 1B	1/1	0/1	0/2	0/2
BC PR-RT	1	2.8	1	1.3	1 B	1 B	1/1	0/1	0/1	0/1
19 cpx[f] RT			2	2.6		2 B			1/2	0/2
19 cpx[f] PR-RT			1	1.3		1 B			0/1	0/1
BAF[g]			1	1.3		1 BG			0/1	0/1
BFC[h]			1	1.3		1 B			0/1	0/1
Total BD	2	5.5	8	10.4	5 (13.8%)	16 (20.7%)				
	27	75.0	51	66.2	30	58	32	15	42	32
Total B					(83.3%)	(75.3%)	(88.8%)	(41.7%)	(54.5%)	(41.5%)

[a] Mixes of subtypes are shown with a bootscan between 20% and 70%.
[b] Numbers to the right of the / in the Stanford program correspond to total samples in each row according to the Genotyping and REGA programs.
[c] Protease region.
[d] Reverse transcriptase region.
[e] B subtype only by genotyping.
[f] CRF19cpx (A1, D, G).
[g] Possible B, A, and F mix.
[h] Possible B, F, and C mix.

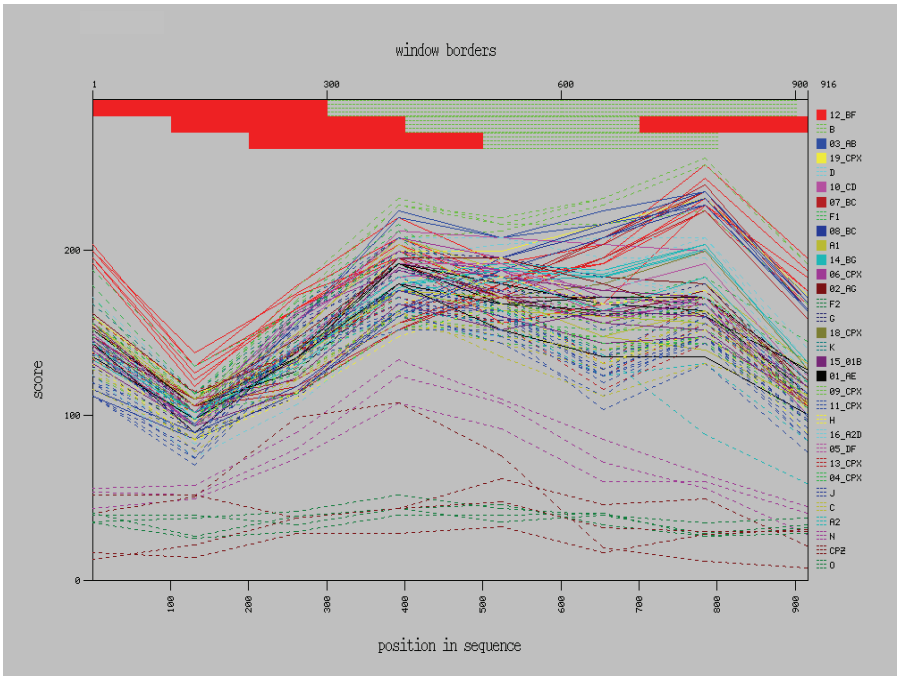
confirm the presence of any recombinants using the pure subtype sequences as a reference.

Alignment with the BioEdit program showed a sequence identity matrix with values between 90% and 99% for these samples and the subtype B references used (B_K03455, B_AY037268, B_M17449), the range being 93%–99% and 90%–97% in BS and VF samples, respectively, with the VF samples showing a greater genetic distance. Identity of the B subtype with the D subtypes selected (D_K03454, D_M27323) was 92%–94%, and it was 89%–90% with the F1 subtypes used (F1_AF005494, F1_AF077336).

Phylogenetic tree analysis of samples with a possible correlation to subtype F using the MEGA (Figure 2) and PhyML (data not shown) programs demonstrated one sample associated with CRF17 BF and four associated with CRF29 BF, while in the phylogenetic tree samples with a possible correlation to subtype D, there are two samples near CRF05 DF and a clade of nine samples between subtype B and CRF05 DF (Figure 3).

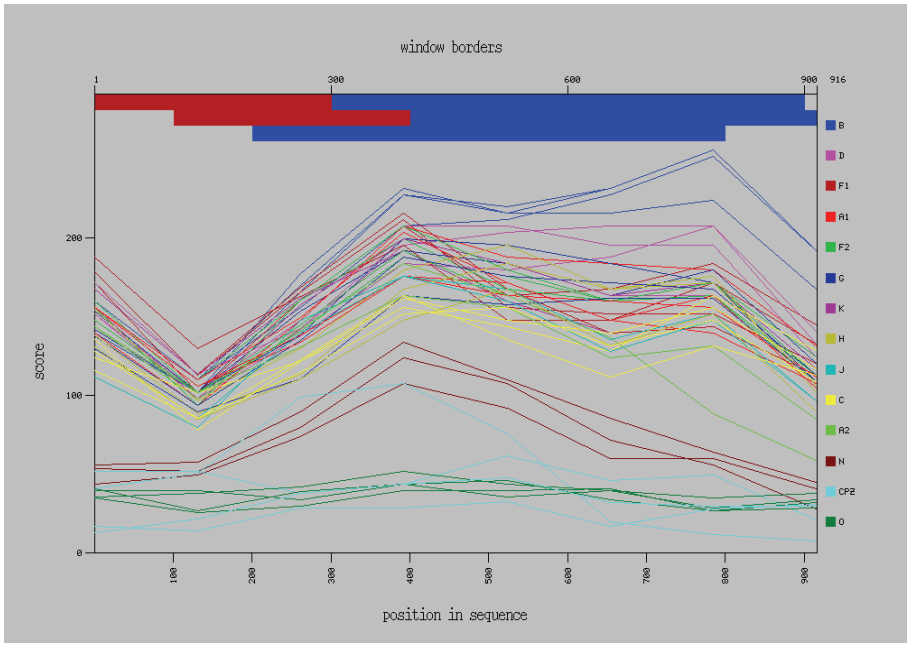
The baseline viral load levels in both groups for this study as well as the control group were very similar (> 75% with > 10^4 copies/mL). The relationship between baseline viral load of the patients in BS and VF groups and variability of their sequences were analyzed and no significant difference was found ($P = 0.5$ and 0.3, respectively), but there was a tendency in the VF group to increased genetic variability in patients with a basal viral load of > 10^4 copies/mL (Table 3). Twelve sequences of these BS correspond to a viral load of > 10^4

FIGURE 1. Graphic representation of a possible CRF12 sample analyzed in subtype Genotyping and REGA programs

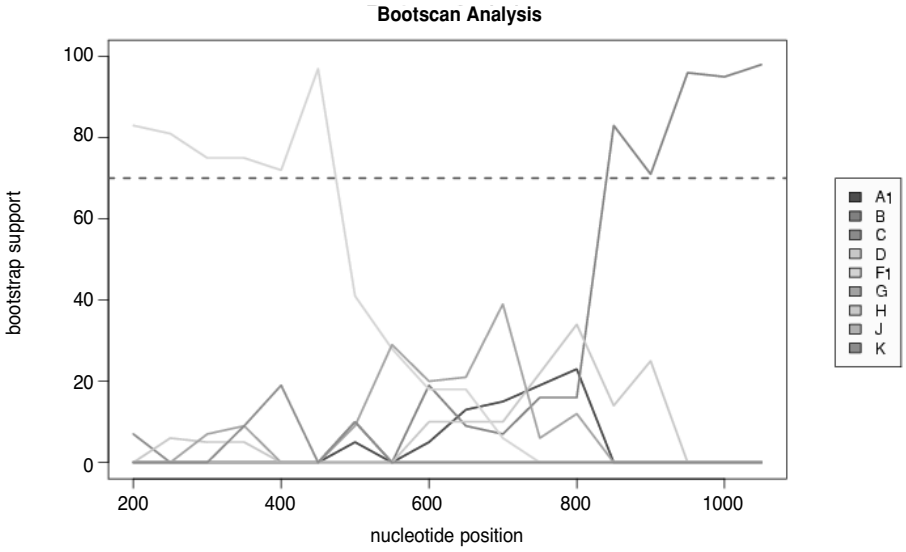


(A) Using 2005 pure and CRF reference sequences; the sample showed CRF12 regions at both ends (red regions) with B in the middle. *(continued)*

FIGURE 1. ***Continued***



(B) Using 2005 pure reference sequences, the sample showed only a 3' region like CRF12, red region.



(C) Analysis by REGA program showed the sample like F1 in the 3' region (yellow line) and B in the 5' region (blue line) (possible BF recombinant).

copies/mL and only two showed some variability at the beginning.

Of the 22 VF samples sequenced in the *env* region V3 loop, subtype B was found in 21 samples (95%) and 1 (5%) was classified as BG by Genotyping. Fifteen samples were assigned to subtype B with bootstrap support ≥ 95%, and the other samples used showed a bootstrap support < 91% by the REGA program (Table 4).

DISCUSSION

Costa Rica has been able to provide better support for HIV patients undergoing treatment with a follow-up that showed a level of only 10% VF in the overall population treated. In an earlier study, we showed that all BS from patients who responded well to antiviral treatment had a viral load of < 10^4 copies/mL (12). In spite of that, we found a high viral variability in the VF patients, with low treatment adherence (Table 2). This variable, together with resistant mutants, was associated with VF in this early study (12). Variables such as sex, age, place of residence, risk group, occupation, and whether the persons were infected outside of Costa Rica were not found to correlate with HIV-1 intra- or intersubtype.

It might be expected that recombinant diversity would be greater in samples from the VF group versus the BS group as the results showed, because the number of drug-resistant strains increases proportionally to residual virus replication, which often reaches high levels after VF and correspondingly exhibits high levels of recombinants. In an earlier study, Taylor-Castillo et al. (12) analyzed mutant genotypes resistant to antiviral treatment using a TruGene System (15) (which identifies 99% of non-subtype B as well as the B subtype) (10, 11), and they reported 78% resistance to nucleoside RT inhibitors, and 50% resistance to both nonnucleoside RT inhibitors and PR inhibitors. In this study, we used the same sequences (32 new) and the reported mutations were checked with the Stanford program (26). These analyses confirmed the previous report of mutation (data not shown) and also demonstrated a higher sequence polymorphism for VF samples; however, it was concluded that the variability found was not related to antiviral resistant mutants because the percentage of possible recombinants in BS, where there are fewer resistant mutants, and VF groups did not show significant differences in spite of the higher percentage of pure subtype B in BS.

A tendency toward increased sequence variability and high viral loads at the time of treatment initiation correlated well with evolutionary viral variation; however, these data are not statistically significant. This trend and a major percentage of variability in VF samples may be associated with other factors such as time of infection before initiation of treatment or time needed to reduce the viral load to undetectable levels or to adherence itself. This observation reinforces the need to implement good prevention policies that should include the availability of screening tests for early detection as well as tests for viral load and CD4 quantification. These tests can

FIGURE 2. Phylogenetic neighbor-joining circular tree of the *pol* region generated using the MEGA program with evolutionary distances determined by the maximum composite likelihood method (5 000 bootstraps) for samples with variability associated with genotypes B and F. Scale represents numbers of substitutions per site. Arrows show association of sample 4724 CV with CRF17 BF and some samples with CRF29 BF

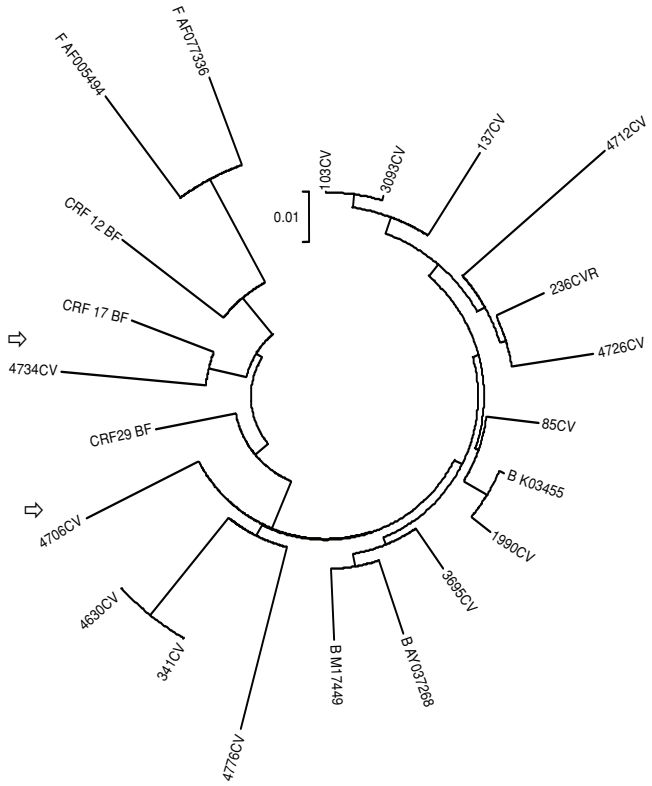
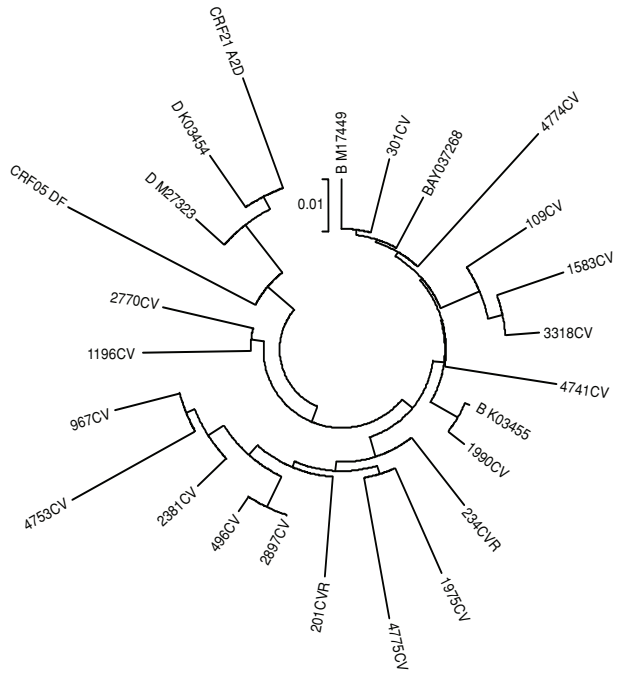


FIGURE 3. Phylogenetic neighbor-joining circular tree of the *pol* region generated using the MEGA program with evolutionary distances determined by the maximum composite likelihood method (5 000 bootstraps) for samples with variability associated with genotypes B and D. Scale represents numbers of substitutions per site. Arrow shows possible association of some samples with CRF05 DF



serve as important markers for initiation of antiviral treatment in order to provide maximal benefit for patients by using ART to lower VF and avoid overall viral variability (5). According to information from the Ministry of Health and other professionals in Central America (2, 31, 32), the availability of screening tests is variable, but in general it is low and the markers used as criteria to initiate treatment are not optimal—that is, lower levels of CD4 (≤ 250 cells/mL) and higher levels of viral load (≥ 100 000 copies/mL). This is despite the demonstrated importance of starting the treatment when CD4 is between 350 and 200 cells/mL in correlation with low viral loads (4, 5).

Although different studies have shown that patients with a sustained viral load of < 50 copies/mL will not develop VF due to resistant mutants, they have not taken into account the effect produced by drug toxicity or other treatment complications or the effect of transient viremia or viral "blips" because of activation of the immune system by infectious agents in these patients (33).

Most of the samples in this study were classified as pure subtype B (Tables 2–4), similar to the situation in Honduras, which has reported 99% of subtype B (34). However, the high variability shown in Costa Rican samples (90%–99% of identity with genotype B reference strains) probably corresponds to the recombination capacity of HIV due to the high percentage of sexual intercourse without protection (58%) or having more than five sexual partners (49%), higher than has been reported overall in Latin America (25%–35%) (1).

A region matching CRF12 (BF) in the *pol* gene (Table 2) was found to be the most frequent possible recombinant by Genotyping, while with REGA the most frequent was a possible recombination between B and D subtypes (Table 2). Both programs detected a sample that was identified as a possible CRF12 or BF recombinant in the *pol* region (Figure 1), and by phylogenetic analysis it was better correlated with CRF17 BF (Figure 2). Recombination between B and F subtypes had been described earlier in South America (Brazil and Argentina) (22, 23), while recombination between B and D is not part of the 43 CRFs described until now and the D subtype has been reported only in 0.4% and 0.6% of samples in North and South America, re-

TABLE 3. Relationship between viral load and *pol* gene subtype in 66 virologic failure samples by program, Costa Rica, 1997–2004

	Virologic failure samples							
	Genotyping				REGA			
	Subtype B		B with variability		Subtype B		B with variability	
Basal viral load	No.	%	No.	%	No.	%	No.	%
≤ 10^4 copies/mL	11	73.3	4	26.7	17	80.9	4	19.1
> 10^4 copies/mL	30	58.8	21	41.2	28	62.2	17	37.8

TABLE 4. HIV-1 subtypes of 22 patients using *env* and *pol* in the Genotyping and REGA programs, Costa Rica, 1997–2004

Genotyping			REGA		
Subtype	No.	%	Subtype bootstraps	No.	%
B	21	95.4	B > 95	16	72.7
			B 90–91	2	13.6
			B low	3	9.1
BG	1	4.6	B low	1	4.6
Total	22			22	

spectively (21). Because of variation observed in both programs, an individual analysis was done with the RIP program using the HIV database (27) and no real recombination was found—only 95%–97% of similarity with the B subtype and 92%–95% with the D subtype. In other words, by using the RIP program important variations are noted in the *pol* sequences, but they do not correspond to true recombination. It is notable that analysis using this program involves only pure subtype sequences.

In contrast, the phylogenetic analysis showed a good correlation between one sample with CRF17 and four with CRF29 (Figure 2) as well as a possible correlation with some samples and subtypes B and D. This analysis, however, was not well defined because this type of recombination has not yet been reported and the comparison was done using DF and AD recombinants. Subtypes B and D are the most closely related and this fact could introduce noise into the comparison (Figure 3).

Less variability was detected in the *env* region (Table 4), although more conventional *env* sequenced-based subtyping analysis of HIV-1 is a good indicator for differentiating subtypes and CRFs. Sequence heterogeneity in this region makes it valuable for phylogenetic differentiation, but it is a relatively poor selection criterion for viral recombination and chimeric virus detection (34). In that study, the true value of such an analysis is not clear because only a small *env* fragment was analyzed.

In spite of limitations, such as an inability to obtain some patients' clinical records and not having the possibility of making complete sequences, our study supports the hypothesis that there is a high level of variability and possibly multiple recombinatorial events occurring in Costa Rican patients. It will be important to determine the origin of these variations and whether they are related to polymorphisms or whether they could be novel sequences that provide a selective advantage to the virus.

It is important to consider that while HIV-drug-resistant mutants have been detected in Costa Rica since 2000, other regional countries are still in the initiation process. In Nicaragua (< 0.5% general prevalence (31)) and Honduras (1.5% general prevalence (32)), ART was started only in 2007 because of the high cost of these drugs. Also, in Nicaragua, CD4 level is the only criterion used to enroll patients, while in El Salvador, the number of patients under treatment is low despite the high number of HIV-positive persons (2). In Central America, there are at least 20 000 persons under treatment. The high-level availability of screening tests in Costa Rica occurs because it is part of the service provided by the Social Security System with more than 95% of the population covered, and screening is also made available for students at the largest state university (Universidad de Costa Rica). The lack of detection and treatment protocols noted above for regional countries could increase viral diversity throughout Central America because it is a small geographic area with active migration among countries (1), and it is further complicated by intermingling of people from Nicaragua and other countries who come to Costa Rica to obtain better medical support.

In conclusion, this study showed a predominance of subtype B in Costa Rican HIV-positive patients and a high variability within HIV sequences with possible recombination between B and F or B and D subtypes; recombination between B and D has not been described (21). Also, at least one sample was identified as CRF12 or related to CRF17 and four were identified with CRF29 based on sequence analysis of the *pol* region. These genetic variations are likely caused by a combination of several factors, such as ART nonadherence, a high percentage of sexual intercourse without protection (58%), and having more than five sexual partners. These factors have been described in other studies as well as reports of a lowered adaptive immune response (6, 34). Intra- and interclade variation within *pol* sequences is particularly relevant, because this region encodes RT and PR proteins, the target of most HIV antiviral drugs. Variations in these regions therefore may affect drug susceptibility and development of drug resistance (22).

Similar studies could be done in other countries, where partial sequences are being generated, and then used to obtain more information about HIV variability in this region. It is especially relevant for tracking the role of cross-country immigration in creating new HIV recombinants. The availability of screening tests, an emphasis on good adherence in taking antiviral drugs, and individual country surveillance analysis and safe sex practices need to be promoted throughout Central America in order to decrease increased virus variability and the cost of ART (35, 36). If the HIV natural history of each country is well documented, a better preventive policy could also be undertaken as part of HIV global spread control.

Acknowledgments. The authors thank Oscar Solano, Laura Chavarría, Johny Obando, and Yadira Martinez from the Costa Rica Social Security System; Anabelle Ferrera from Honduras; Marla Ramos from Panamá; Juan José Amador from the Ministry of Heath in Nicaragua; Carlos Mejía Villatoro from Guatemala; and Gilberto Ascencio from El Salvador for their collaboration to obtain the information for Central America. Also, the authors thank Carlos Vargas from the Virology Section, Microbiology Faculty, University of Costa Rica, for technical assistance and the editorial services provided by Sustainable Sciences Institute (SSI), in particular the revisions and comments from Sondra Schlesinger, board member of SSI and professor emeritus, Department of Molecular Microbiology, Washington University School of Medicine. This work received grant support from Louisiana State University–International Center for Medical Research and Training and financial support from the Florida Ice and Farm from Costa Rica ("Aportes" prize). This study was performed as a partial requirement for the Ph.D. degree of Lizeth Taylor-Castillo at the University of Costa Rica.

REFERENCES

1. United Nations Programme on HIV/AIDS; World Health Organization. 2007 AIDS epidemic update [Internet]. Geneva: UNAIDS; 2007. Available from: http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/default.asp. Accessed 19 April 2008.
2. El Salvador, Ministerio de Salud Pública y Asistencia Social, Programa Nacional de ITS/VIH-sida. Situación del VIH en El Salvador, 1984–2008 [Internet]. El Salvador: Ministerio de Salud Pública y Asistencia Social. Available from: http://www.mspas.gob.sv/pdf/SIDA_DOC/Situacion_VIH_SIDA_El_Salvador_Abril2009.pdf. Accessed 19 April 2009.
3. Nicaragua, Ministerio de Salud, Biblioteca Nacional de Salud. Documentos a textos completos [HIV/AIDS Full Text Documents] [Internet]. Available from: http://www.minsa.gob.ni/bns/sida/doc.html. Accessed 18 September 2008.
4. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. December 1, 2009; 1-161 [Internet]. Available from: http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL.pdf. Accessed April 2008.
5. Wood E, Montaner JSG. When to initiate HIV antiretroviral therapy: do benefits other than survival deserve greater attention? J Acquir Immune Defic Syndr. 2007;45:131–2.
6. Iglesias-Sánchez MJ, López-Galíndez C. Each genomic RNA in HIV-1 heterozygous virus generates new virions. Virology. 2005;333:316–23.
7. Holguín A, Lospitao E, López M, de Arellano ER, Pena MJ, del Romero J, et al. Genetic characterization of complex inter-recombinant HIV-1 strains circulating in Spain and reliability of distinct rapid subtyping tools. J Med Virol. 2008;80:383–91.
8. Schroeder TL, Burger H, Weiser B, Bengualid V, Kimani J, Anzala AO, et al. Characterization of intersubtype recombinant HIV type 1 genomes using a nonradioactive heteroduplex tracking assay. AIDS Res Hum Retroviruses. 2005;21:314–8.
9. Casado G, Thomson MM, Sierra M, Nájera R. Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. J Acquir Immune Defic Syndr. 2005;40:532–7.
10. Grant RM, Kuritzkes DR, Johnson VA, Mellors JW, Sullivan JL, Swanstrom R, et al. Accuracy of TRUGENE HIV-1 genotyping kit. J Clin Microbiol. 2003;41:1586–93.
11. Hirigoyen DL, Cartwrigth CP. Use of sequence data generated in the Bayer TruGene genotyping assay to recognize and characterize non-subtype-B human immunodeficiency virus type1 strains. J Clin Microbiol. 2005;43:5263–71.
12. Taylor-Castillo L, Herrera-Martínez G, León-Bratti MP, Boza R, León-Rodríguez B, Luftig RB, et al. Study of antiretroviral mutants in HIV patients with treatment failures and the effect of risk factors in the virological failures. Rev Inst Med Trop Sao Paulo. 2005;47:327–31.
13. Cuevas MT, Ruibal I, Villahermosa MT, Díaz H, Delgado T, Vázquez de Parga E, et al. High HIV-1 genetic diversity in Cuba. AIDS. 2002;16:1643–53.
14. Posada D. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol. 2002;19:708–17.
15. Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusosa T. A web-based genotyping resource for viral sequences. Nucleic Acids Res. 2004;32:654–9.
16. Eyer-Silva W, Morgado MG. Molecular epidemiology of HIV-1 infection in small Brazilian county usefulness of envelope and polymerase sequences to epidemiologic studies. J Acquir Immune Defic Syndr. 2006;41:664–70.
17. Levy DN, Adrovandi GM, Kutsh O, Shaw GM. Dynamics of HIV-1 recombination in its natural target cells. Proc Natl Acad Sci. 2004;101:4204–9.
18. Bacharov G, Ford NJ, Edwards J, Breining T, Wain-Jobson S, Meyerhans A. A genetic-algorithm approach to simulating human immunodeficiency virus evolution reveals the strong impact of multiply infected cells and recombination. J Gen Virol. 2005;86:3109–18.
19. Aulicino PC, Kopra J, Rocco C, Mangano A, Sen L. Sequence analysis of a South American HIV type 1 BC recombinant. AIDS Res Hum Retroviruses. 2005;21:894–6.
20. Chin MPS, Rhodes TD, Chen J, Fu W, Hu W-Sh. Identification of a major restriction in HIV-1 intersubtype recombination. Proc Natl Acad Sci. 2005;102:9002–7.
21. HIV sequence database. The circulating recombinant form (CRFs) [Internet]. Available from: http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html. Accessed 15 July 2009.
22. Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, Rachid M, Gracie RSG, Chequer-Fernandez SL, et al. Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: assessing subtype and drug resistance associated mutations in HIV-1 infected individuals mailing highly active antiretroviral therapy. Mem Inst Oswaldo Cruz. 2005;100: 73–8.
23. Thomson MM, Delgado E, Herrero I, Villahermosa RL, Vazquez de Parga E, Cuevas MT, et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1BF intersubtypes recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. J Gen Virol. 2002;83:107–19.
24. NCBI genotyping tool [Internet]. Available from: http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi. Accessed 18 September 2008.
25. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Sebregts C, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics. 2005;21:3797–800.
26. Stanford University HIV drug resistance database [Internet]. Available from: http://hivdb.stanford.edu/. Accessed 19 April 2008.
27. HIV sequence databases. Recombinant Identification Program (RIP 3.0) [Internet]. Available from: http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html. Accessed 17 July 2009.
28. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41:95–8.
29. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24:1596–9. Available from: http://www.kumarlab.net/publications.
30. HIV sequence database. PhyML interface [Internet]. Available from: http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html. Accessed 17 July 2009.

31. Hoekstra CE, Riedijk M, Matute AJ, Hak E, Delgado E, Alonso RE, et al. Prevalence of HIV and syphilis in pregnant women in León, Nicaragua. Am J Trop Med Hyg. 2006;75:522–5.
32. Lara C, Sällberg M, Johansson B, de Rivera IL, Sönnerborg A. The Honduran human immunodeficiency virus type 1 (HIV-1) epidemic is dominated by HIV-1 subtype B as determined by V3 domain sero- and genotyping. J Clin Microbiol. 1997;35:783–4.
33. Keiffer TL, Finucane MM, Nettles RE, Quinn TC, Bromann KW, Ray SC, et al. Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. J Infect Dis. 2004;189:1452–65.
34. Ying L, Eran E, Ivet B. Sequence analysis. Analysis of correlated mutations in HIV-1 protease using spectral clustering. Bioinformatics. 2008;24:1243–50.
35. Freedberg K, Hirschhorn L, Schackman B, Wolf LL, Martin LA, Weinstein MC, et al. Cost-effectiveness of an intervention to improve adherence to antiretroviral therapy in HIV-infected patients. J Acquir Immune Defic Syndr. 2006;43:S113–8.
36. Gordon ChM. Commentary on meta-analysis of randomized controlled trials for HIV treatment adherence interventions research directions and implications for practice. J Acquir Immune Defic Syndr. 2006;43:S36–40.

Manuscript received on 14 April 2009. Revised version accepted for publication on 2 September 2009.

RESUMEN

Variabilidad de secuencias genómicas parciales del VIH-1 en pacientes costarricenses: análisis con diferentes herramientas bioinformáticas

Objetivos. Determinar el subtipo y la variabilidad genómica del gen *pol* del VIH de pacientes costarricenses mediante diferentes herramientas bioinformáticas y el uso de esta información para establecer nuevas políticas para mejorar el diagnóstico y el tratamiento de estos pacientes.

Métodos. Se analizaron 113 secuencias del gen *pol* de pacientes costarricenses bajo tratamiento antirretrovírico de gran actividad mediante cuatro programas: Genotyping, REGA, Stanford y MEGA. Las secuencias *pol* analizadas provenían de 77 casos considerados fracasos virológicos (FV) y 36 muestras iniciales (MI). También se secuenció la región *env* de 22 de los 77 FV.

Resultados. No se encontraron diferencias importantes entre las variables estudiadas. No obstante, se observó una tendencia a una mayor variabilidad en los pacientes FV que tenían una elevada carga viral inicial. Con respecto al gen *pol*, 77–83% de las MI y 66–75% de las muestras de los FV eran del subtipo B puro según Genotyping y REGA, respectivamente. Las otras muestras presentaron variaciones relacionadas principalmente con la forma recombinante en circulación CRF-12 según Genotyping, con la CRF-17 o la CRF-29 según el análisis filogenético, o una nueva posible forma recombinante BD según todos los programas. Con el programa Stanford, todas las muestras variables reflejaron un subtipo B con elevado polimorfismo. La variabilidad de la secuencia *env* fue menor que la de la región *pol*.

Conclusiones. El subtipo B fue el predominante en los pacientes positivos al VIH en Costa Rica. Existe una alta variabilidad en las secuencias con una posible recombinación entre los subtipos B, y F o D. La forma recombinante BD no se había notificado antes. Esta elevada variabilidad parece ser el resultado de posibles eventos de recombinación, la falta de adhesión al tratamiento antirretrovírico, las relaciones sexuales sin protección y numerosas parejas sexuales. Se deben emprender estudios similares en otros países de la Región, en particular en los lugares con mucha inmigración, para reducir tanto la posibilidad de que el virus varíe como el costo del tratamiento antirretrovírico.

Palabras clave VIH-1; análisis de secuencia; recombinación genética; variación genética; programas informáticos; Costa Rica.