CHARACTERIZATION OF THE HIV-1 SUBTYPES PRESENT IN CÓRDOBA (ARGENTINA) FROM NEARLY THE BEGINNING OF THE INFECTION, USING THE POL AND ENV GENES

CARACTERIZACIÓN DE LOS SUBTIPOS DE VIH-1 PRESENTES EN CÓRDOBA (ARGENTINA) DESDE CASI EL INICIO DE LA INFECCIÓN, UTILIZANDO LOS GENES POL Y ENV.

Raúl E. González-Ittig[1,2], María A. Lucca[2], Luciana Caeiro[2], Nidia M. Modesti[2]

Abstract:

Objective: In the present study we characterize, using a retrospective approach, the HIV-1 infection in Córdoba, Argentina, by determining the viral subtypes circulating in the period 1986-2001. Methods: The *pol and env genes* were analyzed using two different approaches: 1) comparing the sequences from Córdoba with those of pure subtypes and circulating recombinant forms of the HIV-1 M group; 2) subdividing the sequences into pure-subtype data sets (avoiding the effects of recombination) to determine if the subtypes from Córdoba have a single or multiple geographic origins. Results: The first approach revealed that eighteen strains were subtype B, seven B/F1, one C and one F1. Only two of the B/F1 corresponded to the CRF12_B/F1, previously described for Argentina; the remaining B/F1 sequences could be considered different forms of URF_B/F1. The Phylogenetic results of the second approach suggest that in Córdoba subtype B would have multiple geographic origins, while subtypes F1 and C would have been introduced directly from Brazil or indirectly from Buenos Aires. Conclusions: The great subtype variability indicates that the HIV-1 infection in Córdoba was complex since its very beginning. Our results also reveal the close contact of people of this city, through tourism and commercial activities, with those from Buenos Aires and Brazil.

Keywords: HIV-1 infection; Córdoba, Argentina; HIV-1 subtypes; phylogenetic analyses; inter subtype recombinant strains

Resumen:

Objetivo: En el presente estudio, utilizando un enfoque retrospectivo, caracterizamos la infección de VIH-1 en Córdoba (Argentina) a través de la determinación de los subtipos virales circulantes en el período 1986-2001. Métodos: Se analizaron los genes pol y env utilizando dos enfoques diferentes: 1) comparando las secuencias de Córdoba con aquellas de subtipos puros y formas recombinantes circulantes del grupo M del VIH-1; 2) subdividiendo las secuencias en matrices de subtipos puros (evitando los efectos de la recombinación) a fin de determinar si los subtipos de Córdoba tienen uno o varios orígenes geográficos. Resultados: El primer enfoque reveló que dieciocho cepas correspondían al subtipo B, siete a B/F1, una a C y una a F1. Solo dos de los B/F1 correspondían a la forma CRF12_B/F1, previamente descripta para Argentina; las secuencias B/F1 restantes serían diferentes formas de URF_B/F1. Los análisis filogenéticos del segundo enfoque sugieren que el subtipo B de Córdoba tendría múltiples orígenes geográficos, mientras que los subtipos F1 y C habrían sido introducidos directamente desde Brasil o indirectamente desde Buenos Aires. Conclusiones: La gran variabilidad de subtipos indica que la infección de VIH-1 en Córdoba fue compleja desde sus inicios. Nuestros resultados también revelan el estrecho contacto entre la gente de esta ciudad, a través del turismo y actividades comerciales, con las de Buenos Aires y Brasil.

Palabras clave: Infección de VIH-1; Córdoba, Argentina; subtipos de VIH-1; análisis filogenéticos; cepas recombinantes inter-subtipo

1. Instituto de Diversidad y Ecología Animal (CONICET-UNC) and Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba. Argentina. E-mails: raulgonzalezittig@yahoo.com.ar and raulgonzalezittig@unc.edu.ar
2. CEPROCOR – Ministry of Science and Technology of the Córdoba Province Government, Argentina.

Introduction

The HIV/AIDS epidemic is characterized by a high diversity of circulating strains around the world, which is a consequence of both high mutation and viral recombination rates[1]. The main concerns regarding increases in HIV diversity include the reduced efficiency of diagnostic and monitoring assays, the unknown impact on transmissibility, and/or pathogenicity and vaccine development, thus reinforcing the necessity for subtype surveillance studies. The molecular epidemiology of HIV-1 has become increasingly important as viral subtypes are becoming more dispersed worldwide. There are four major HIV-1 phylogenetic groups: M, O, N and P[2]. Contemporary human migration has determined the global dynamics of HIV-1 causing that not all viruses are geographically widespread; the M group is recognized as the most extensively distributed.

Evolutionary analyses of viral sequences have identified eleven circulating genetic subtypes within the M group: A1, A2, B, C, D, F1, F2, G, H, J and K. Besides, there has been reported more than sixty circulating recombinant forms (CRFs) and several unique recombinant forms (URFs), generated by recombination events between two or more subtypes[2]. Besides, the worldwide distribution of HIV-1 subtypes and inter-subtype recombinants is not homogeneous. In North and Central America and the Caribbean, HIV-1 subtype B predominates. In South America, northern and pacific coast countries show a similar epidemic to that of North America, but in Brazil and Argentina subtypes B, F1, C and several recombinant forms are also represented[3].

In Argentina, the first AIDS case was identified in 1982 and since then the disease has been expanding. In the 1990s, the *env* gene (V3 and C2-V3) was mostly used to characterize the circulating HIV-1 subtypes in Buenos Aires and Rosario cities; these studies detected the presence of subtypes B and F, with subtype B as the most abundant[4,5,6]. The full-length sequencing allowed the characterization of the recombinant CRF12_B/F1 together with other diverse URF_B/F1[7,8], showing that almost all F1 strains, according to the *env* gene, were in fact recombinant forms with breaking points with subtype B in the *pol* gene. Other studies using only strains from Buenos Aires detected recombinants URF_B/F1, CRF16_A2D, B/D, B/C, B/K, B/A, CRF06_cpx, and a triple recombinant B/C/F[9,10,11,12,13].

The epidemic in Buenos Aires has been exhaustively studied, but little is known about the HIV-1 epidemic in the rest of Argentina. The city of Córdoba is the capital of the province with the same name and is nearly 700 km (435 miles) away from Buenos Aires. The first HIV-1 infected person was identified in 1985, and the first case of AIDS was diagnosed in 1986. During the period 1985-1999 the number of infected people increased steadily, but in the last years it has stabilized in around 710 cases per year. Regarding the HIV-1 subtypes circulating in Córdoba, most of the knowledge of the subtypes circulating in the population corresponds to studies in recent years. Pando et al.[14] in a study on the genetic variability of HIV-1 in sex workers from several cities of Argentina included two women from Córdoba: one presented subtype B/F1 and the other presented a dual infection B and B/F1. More recently, Pando et al.[15] subtyped seventeen strains from Córdoba; of these 5 were B, 9 B/F1 and 3 C (Dr. María A. Pando, personal communication). Considering the subtypes of these relatively "modern" sequences, a question arises about how was the epidemiology of HIV in Córdoba in its beginnings. Therefore, in the present study we characterize the subtypes of HIV-1 seropositive patients from Córdoba diagnosed between 1986 and 2001, using sequences of the *env* and *pol* genes. By using phylogenetic analyses, we contribute to the knowledge about whether the different subtypes have one or multiple geographic origins.

Material and methods

Study population

Blood samples were taken from twenty seven unrelated HIV-1 infected individuals who attended CEPROCOR (Centro de Excelencia en Productos y Procesos, Córdoba, Argentina) for HIV-1 viral load analysis (Table 1). Samples were taken between 1996 and 2001, but the known diagnosis date varied between 1986 and 2001. This study was conducted in compliance with all federal regulations of Argentina governing the protection of human subjects.

The ages ranged between 23 and 57 years, except for three newborn children. The transmission route was sexual (including MSMW: men who have sex with men and women, MSM: men who have sex with men, HET: heterosexual): 13 individuals (48.1%); IDU (injecting drug user): 2 (7.4%); PAR (parenteral or accidental infection

by a puncture object): 1 (3.7%); VER (vertical transmission): 3 (11.1 %), not known: 8 (29.7%) (Table 1).

Sample processing and amplification of env and pol genes

Peripheral blood mononuclear cells (PBMCs) were collected from EDTA whole blood and purified by Ficoll gradient centrifugation. After two washes in phosphate-buffered saline (PBS), cells were pelleted and stored at – 20º C. DNA was extracted with a standard phenol-chloroform procedure. The *env* and *pol* regions were amplified by nested polymerase chain reaction (PCR). For the *env* gene, the first PCR round was run using primers ED3 and ED14. The second round primers were ED5 and ED12 and amplified a 1.2 kb fragment encompassing the V1-V5 region of the *env* gene (positions 6556-7792 of the HXB2 genome, GenBank accession number K03455). For the outer primers the reaction started with denaturation at 94 ºC for 4 min, followed by 35 cycles of denaturation at 94ºC for 1 min, annealing at 55ºC for 1 min and extension at 72ºC for 2 min. Finally, there was a hold period of 5 min at 72ºC. For the second-round PCR, 5 μL of the first-round PCR mixture was used, with the inner primers ED5 and ED12 under identical reaction conditions.

A portion of the *pol* gene was amplified in a first PCR round using primers Pro3F and RT3474R. The inner primers, Pro3F and ProRT, were used to amplify a ± 1.1 kb fragment (positions 2205-3263 of the HXB2 genome, positions 2205-3263 of the HXB2 genome. For the outer primers the reaction started with an initial denaturation for 10 min at 94ºC, 45 cycles of 30 sec at 94ºC, 30 sec at 55ºC, and 1.5 min at 72ºC, with a final extension of 7 min at 72ºC. For the second-round PCR, 5 μL of the first-round PCR mixture was used, with the inner primers Pro3F and ProRT, under identical reaction conditions. Amplifications were performed in a GeneAmp PCR System 9600 (Perkin Elmer Corporation, Norwalk, CT). The PCR products were detected on agarose gel electrophoresis and visualized by ethidium bromide staining.

Sequences of the *env* and *pol* genes were obtained directly from the PCR products. Nucleotide sequencing was performed with the fluorescent terminator technology (Big Dye Terminator Cycle Sequencing with AmpliTaq DNA polymerase FS, Perkin Elmer) according to the manufacturer instructions. Sequencing was performed on an Applied Biosystems 310 automatic DNA sequencer (PE Applied Biosystems, Foster City, California, USA).

Sequence alignment and Recombination analyses

Each sequence was screened using the BLAST program of the Los Alamos HIV database (http://www.hiv.lanl.gov) to search for similarities with previously reported HIV-1 reference strains; subtype reference sequences were obtained from that database. Nucleotide alignments among the reference sequences and those of the present study were produced using MAFFT v7[16]. This program implements progressive alignments with extra steps to improve the alignment guide tree. We used the strategies L-INS-i and G-INS-I, with gap opening penalty: 1.53, gap extension penalty: 0.1.

Given the high complexity involved in HIV-1 sequences not all recombinants are detected by all methods and different methods can infer different breaking points; therefore, for both the *env* and *pol* genes evidences of recombination were investigated using several approaches: 1) The Recombinant Identification Program (RIP) v3.0[17](http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) with a minimum confidence threshold of 0.95 with a window size of 200 nt was used. 2) The REGA HIV-1[18] subtyping tool (http://www.bioafrica.net/rega-genotype/html/subtypinghiv.html), using a window size of 400 and step size of 20. 3) The Simplot v3.5.1[19]; a window size of 300bp and a step size of 50 were used. Sequences were gap-stripped and genetic distances were calculated using Kimura 2-p parameters. Simplot analyses were performed using the pure subtype sequences; bootstrap supporting values were based on 100 resamplings. 4) With the program jpHMM[20] (http://jphmm.gobics.de) recombination breaking points were set using the highest statistically significant X_2 value around the 50% crossover point between subtypes. The statistical significance of the identified breaking points was assessed using Fisher's exact test. 5) The Subtype Classification Using Evolutionary Algorithms (SCUEAL)[21] program (http://www.datamonkey.org/dataupload_scueal.php) was used only for the *pol* gene.

Phylogenetic analyses

We used two different approaches for the analyses:

1) comparing the complete *pol* and *env* sequences from Córdoba with those of pure and circulating recombinant forms of Los Alamos database. Both of the *env* and *pol* sequence data sets were analyzed using maximum parsimony (MP), Bayesian inference analyses (BI) and Neighbor Joining (NJ). MP was performed with PAUP 4.0b10[22] using a heuristic search of 1000 replicates of random taxon addition and the TBR (tree bisection-reconnection) branch swapping algorithm. For BI analyses, the best-fitting model of sequence evolution was selected using jModeltest 2[23]: for the *"env"* data set, the GTR+G model was selected using the Bayesian information criterion (BIC), and the following starting parameters were employed for subsequent Bayesian inferences: a base frequency of A = 0.4041, C = 0.1839, G = 0.1990, T = 0.2131; a rate matrix of [A-C]=1.3974, [A-G]= 3.293, [A-T]=0.6203, [C-G]=0.9025, [C-T]=3.3938, [G-T]=1.000; a gamma distribution with alpha = 0.547. For the *"pol"* data set, the GTR+I+G model was selected using the BIC criterion, and the following starting parameters were used: a base frequency of A = 0.4075, C = 0.1580, G = 0.2110, T = 0.2235; a ratio matrix of [A-C]=2.3574, [A-G]= 7.8380, [A-T]=0.8228, [C-G]=1.4359, [C-T]=10.7780, [G-T]=1.000; a proportion of invariable sites of 0.2960, and a gamma distribution with alpha = 0.8270. BI analyses were performed using MrBayes 3.2[24] with two independent Markov chain Monte Carlo (MCMC) runs, with one cold and three heated chains each. Runs were performed for three million generations for the *pol* data set and four million generations for the *env* data set; trees were sampled every 1000 generations. We discarded the first 25% of the samples as "burn in" and the two runs converged on very similar posterior estimates with an average standard deviation of split frequencies of 0.006 for *pol* and 0.004 for *env*. NJ phylogenetic trees were constructed using the Kimura 2-parameter (K2p) distance with PAUP 4.0b10. Bootstrap analysis was performed in order to assess the support of the nodes using 1000 replicates.

2) To determine a single or multiple geographic origins of the subtypes present in Córdoba, we divided the sequences of the *pol* and *env* genes into "pure" subtype data sets in order to avoid confounding effects produced by recombination. For these last analyses we used only pure subtype sequences of Los Alamos database with different geographic origins. Both the *env* and *pol* sequences were divided into three data sets; each one was analyzed using only BI. The best-fitting model of sequence evolution was selected using jModeltest 2: for the three *pol* data sets, the TPM3uf+I+G model was selected using the BIC criterion and for the three *env* data sets, the GTR+G model was selected using the BIC criterion. To ensure convergence, all the analyses were performed as described above, but using 10 million generations.

Results

Identification of recombinant sequences

All the DNA samples were maintained in freezers at -20º C for 15 to 20 years, therefore some technical problems arose in PCR amplifications: we could recover 27 sequences for the *env* gene, but only 16 for the *pol* gene. We attempted to amplify this last gene using different conditions and annealing temperatures at least three times before discarding the sample. The poor amplification success in the *pol* gene could be attributed to several causes:

a) DNA degradation due to the passage of time; b) the 5' position of the amplified fragment is in the *gag/pol* junction, which is known that present indels; c) in 8 of the 11 negative *pol* samples plasma viral load was known and 5/8 had plasma viral load below 300 copies/ml, but this is less probable because the *env* gene of the same samples was amplified (Table 1).

With the positive HIV-1 sequences from Córdoba, different methods (RIP, REGA, Simplot, jpHMM, SCUEAL) were used to detect recombination breaking points. The mosaic structures of the *env* and *pol* genes are shown in Fig. 1: ten strains corresponded to subtype B in both genes. None of the strains was F1 for both genes. We detected several recombinants B/F1: one was B for *pol* and F1 for *env* (Argco 033), one presented a recombination points in *pol* and was B in *env* (Argco 010), and four presented recombination points in *pol* and were F1 in *env* (Argco 018, Argco 025, Argco 035 and Argco 034). All the programs mentioned above detected that the strains Argco 018, Argco 025 and Argco 035 presented the same mosaic structure of the CRF12_B/F1. In the strain Argco034 all the programs detected the existence of two breaking points in the *pol* gene but at different positions,

indicating that there was an uncertain breaking point region; in Fig. 1 we show the positions detected by jpHMM and RIP. The strains of Argco 034 and Argco 010 would correspond to cases of URF_B/F1.

For those strains where only the *env* gene was available, we detected the following mosaic structure: eight were B, one F1 and one C. The strain Argco030 was the only B/F1, which showed a short portion of subtype F1 flanked by segments of subtype B (the two breaking points were detected with RIP, REGA, Simplot and jpHMM).

Considering the results obtained with the two genes, eighteen (66.7%) patients presented subtype B, seven (25.9%) presented different B/F1 recombination patterns, one C (3.7%) and one F1 (3.7%; in the *env* gene).

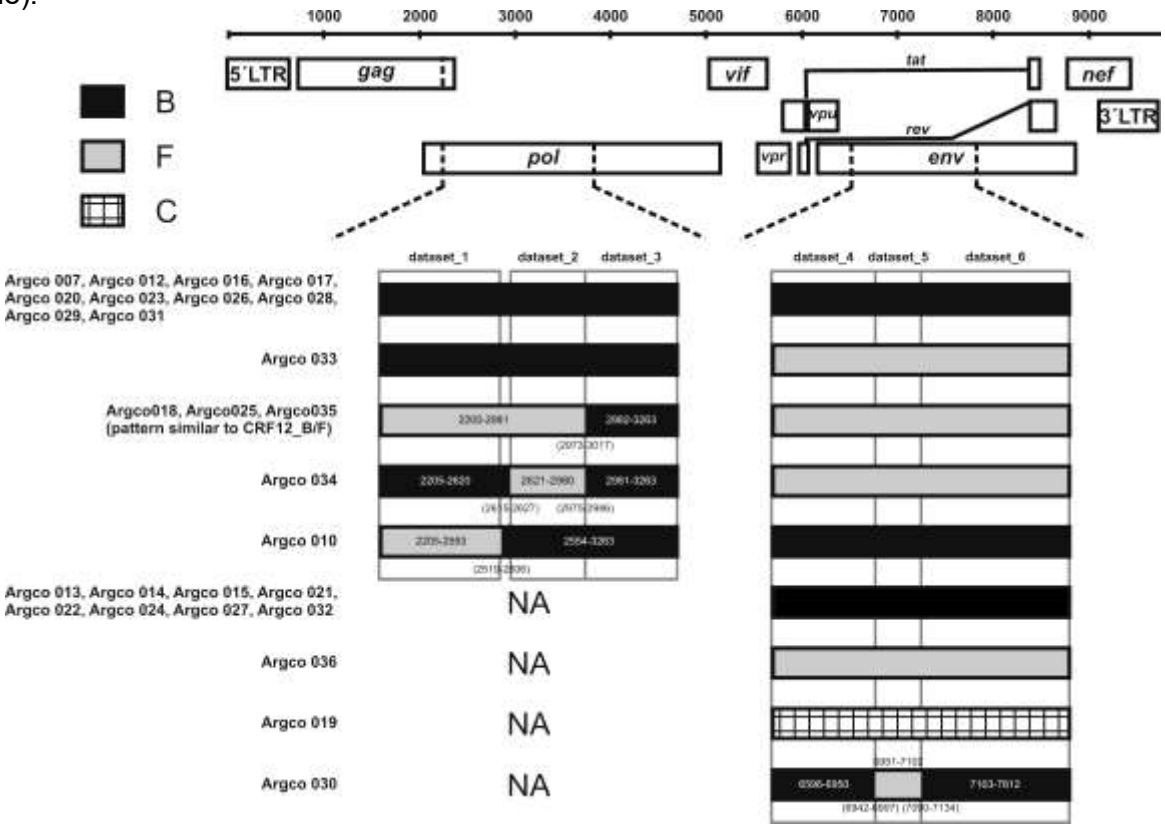

Fig. 1. Representation of the mosaic structure identified in the *pol* and *env* genes of the 27 HIV-1 sequences from Córdoba, Argentina. Breaking point locations and 95% confidence intervals (in parenthesis) are numbered according to the HXB2 sequence. The three datasets of the *pol* and in the *env* genes used in phylogenetic analyses are highlighted.

Phylogenetic analyses of complete sequences

For the *pol* matrix we performed preliminary phylogenetic analyses including representative sequences of the recombinants forms CRF17_B/F1, CRF40_B/F1, CRF42_B/F1, CRF44_B/F1, CRF46_B/F1 and CRF47_B/F1 obtained from Los Alamos HIV database. None of our HIV-1 samples grouped with these CRF_B/F1 (results not shown), so these sequences were discarded from the comparisons listed below. Of the sixteen strains from Córdoba, eleven where classified as subtype B, with high statistical support (Fig. 2), confirming the results showed in Fig. 1. The sequence of Argco 034, which presented a unique mosaic structure, did not group with any other sequences. Two sequences (Argco 018 and 025) formed a moderately supported clade with representatives of the recombinant subtype CRF12_B/F1 (Fig. 2) supporting the mosaic structure showed in Fig. 1. Interestingly, the sequence of Argco 035, which also showed the mosaic structure of CRF12_B/F1, clustered with high support with the CRF38_B/F1.FJ213781 (Fig. 2); however, all the recombinant analyses (RIP, REGA, Simplot, jpHMM, SCUEAL) performed on the CRF38_B/F1 revealed different breaking points from those of Argco 035 (not shown) and those of CRF12_B/F1. Something analogous occurred with the sequence of Argco 010 that grouped with the Brazilian CRF28_B/F1 and CRF29_B/F1. The statistical support of the clade was low with MP, and moderate with NJ and BI (Fig. 2). Although the overall mosaic structure is similar, the recombination breaking points of Argco 010 are not exactly at the same position of those detected in the CRF28_B/F1 and CRF29_B/F1 sequences.

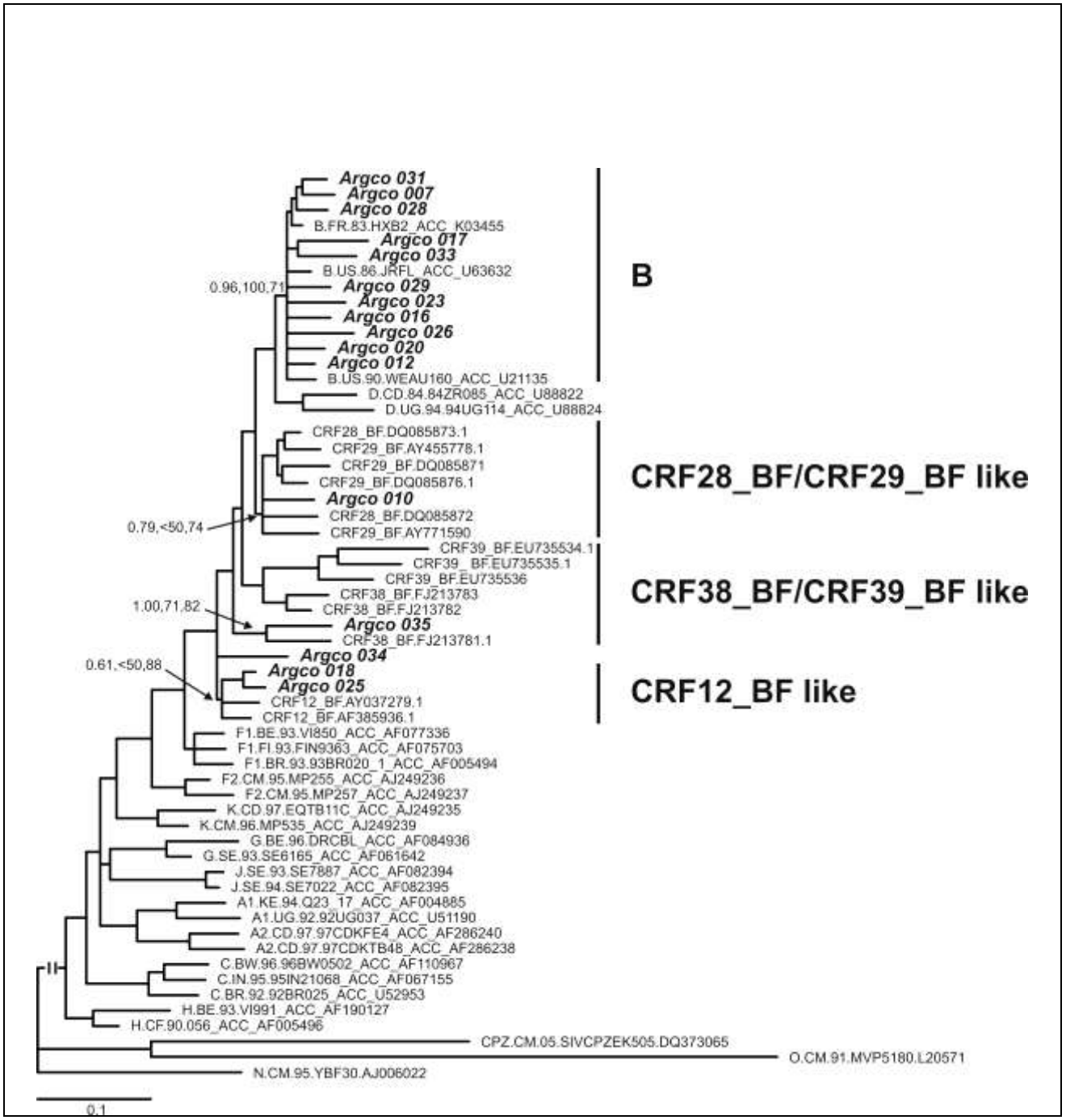

Fig. 2. Phylogram of the Bayesian consensus tree obtained from the *pol* gene data set after three million generations. The node supporting values are indicated only for the clades that include the HIV-1 strains characterized in this study (Argco). The order of the values is: Bayesian posterior probabilities/Maximum Parsimony/Neighbor joining. Reference sequences were obtained from Los Alamos HIV database; HIV-1 sequences of strains of CPZ, group N and group O were used as outgroups.

For the *env* gene data set, the different phylogenetic analyses (BI, PM and NJ) produced similar results (Fig. 3): twenty sequences belonging to patients from Córdoba were classified as subtype B with high statistical support. The HIV strain of patient Argco 010 was an atypical subtype B, which presented a very long branch length; all the programs used for detecting recombination breaking points showed that this sequence corresponds to a pure subtype B and according to the *pol* gene it would correspond to a URF_B/F1. On the other hand, six sequences were classified as subtype F1 (Fig. 3); these sequences where more related to those from Brazil and Argentina, than to those from Finland included in the analyses. The sequence of Argco 019 was classified as subtype C (Fig. 3); this subtype presented two subclades, one that included the sequence from Córdoba and other sequences from Argentina and Brazil, whereas the other subclade grouped sequences from Zambia, Botswana, Tanzania, South Africa and India.

Considering the diagnosis years, in the period 1986-1994, the 6 cases were subtype B (5 MSMW and one unknown); in the period 1995-2001, ten were B (53%; 1 MSM, 1 PAR, 5 HET and 3 unknown), seven were B/F1 (37%; 2 IDU, 1 HET, 2 VER and 2 unknown), one C (5%; 1 VER) and one F1 (5%; at least for the *env* gene; unknown mode of infection). With unknown year of diagnosis, two isolates were B.

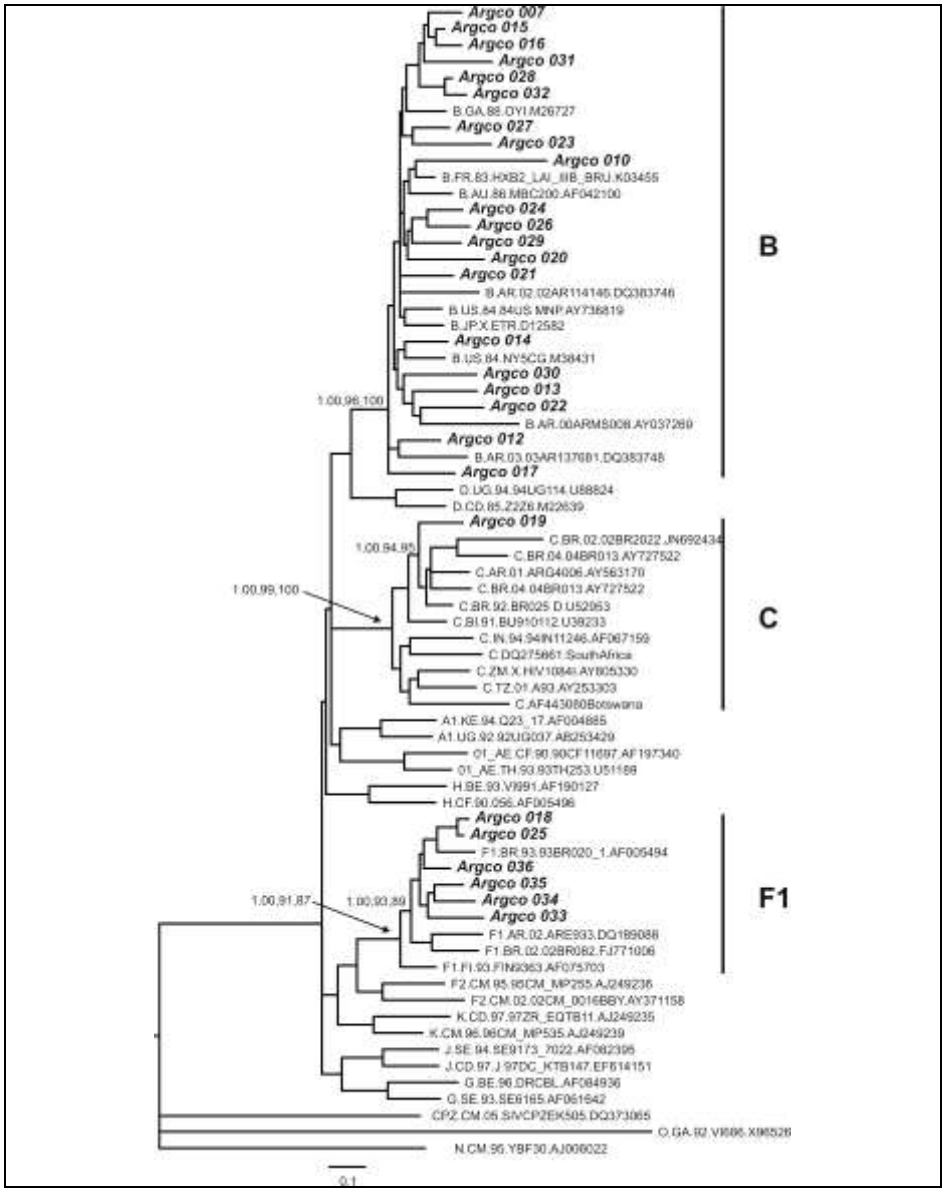

Fig. 3. Phylogram of the Bayesian consensus tree obtained from the *env* gene data set after four million generations. The node supporting values are indicated only for the clades that include the HIV-1 strains characterized in this study (Argco). The order of the values is: Bayesian posterior probabilities/Maximum Parsimony/Neighbor joining. Reference sequences were obtained from Los Alamos HIV database; HIV-1 sequences of strains of CPZ, group N and group O were used as outgroups.

Phylogenetic analyses of "pure" data sets

To determine a single or multiple geographic origins for the subtypes from Córdoba, we divided the *pol* and *env* sequences into different "pure" data sets to avoid the puzzling effects produced by recombination. The BI analyses of the data sets were congruent with the classification obtained in Fig. 1 for the different segments of the mosaic structure. The *pol* data sets 1, 2, and 3 presented 367, 283 and 219 nucleotides, respectively. In these data sets, sequences classified as subtype B presented a posterior probability of 0.95, 0.63 and 0.68, respectively (Fig. 4a, b, c). Strains corresponding to subtype F1 presented a posterior probability 0.78 and 0.99 for data sets 1 and 2; in data set 3, none of the sequences from Córdoba corresponded to subtype F1. The analyses revealed different geographic origins for the subtypes from Córdoba; sequences corresponding to subtype B from Córdoba clustered with other sequences from diverse geographic origins in the three data sets. Sequences F1 from Córdoba were related to other from Brazil and Argentina and not to sequences from Europe (Belgium, Spain, Romania, etc.) or Africa (Angola, Democratic Republic of Congo) in dataset 1 (Fig. 4a). However, this pattern was not recovered in dataset 2, probably because this segment presented very low phylogenetic signal (Fig 4b).

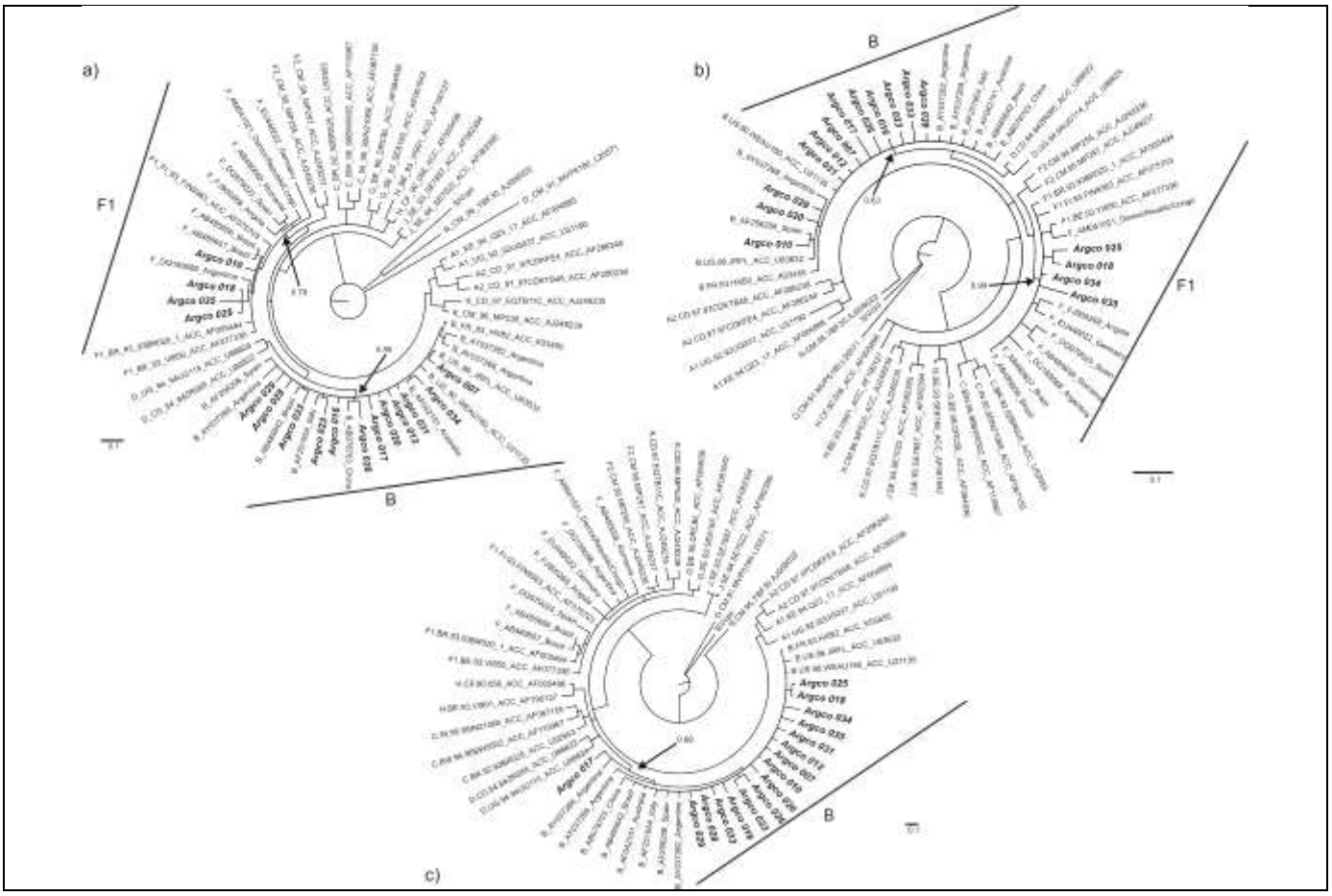

Fig. 4. Phylograms of the Bayesian consensus trees obtained from the three *pol* gene data sets after ten million generations; the location of each data set is shown in Fig. 1: a) result of data set 1; b) result of data set 2; c) result of data set 3. The node supporting values are indicated only for the clades that include the HIV-1 strains characterized in this study (Argco).

The *env* data sets 4, 5, and 6 presented 498, 272 and 800 nucleotides, respectively. In data sets 4 and 6, subtype B was recovered as monophyletic with a posterior probability of 0.99 and 0.73, respectively (Fig. 5a and c). In data set 5, the consensus tree showed a lack of resolution for subtype B, and most sequences were recovered in a basal polytomy (Fig. 5b). For subtype F1, the posterior probability in data sets 4, 5 and 6 was 1, 0.76 and 0.65; six sequences from Córdoba grouped within F1 for data sets 4 and 6, and seven for data set 5 (the small fragment of F1 detected in Argco 030 was clearly part of this subtype). The sequence of Argco 019 was recovered as part of subtype C with a posterior probability of 0.99, 0.76 and 0.61 in the three data sets. In general, the phylogenetic signal of the *env* gene was

stronger than that of the *pol* gene and showed that sequences from Córdoba corresponding to subtype C and F1 are strongly related with other sequences from Argentina and Brazil. Subtype B presents very high levels of variation and within the subtype, sequences from Córdoba were related to other from France, Argentina, Brazil, Australia, Japan and USA (Fig. 5 a, b and c).

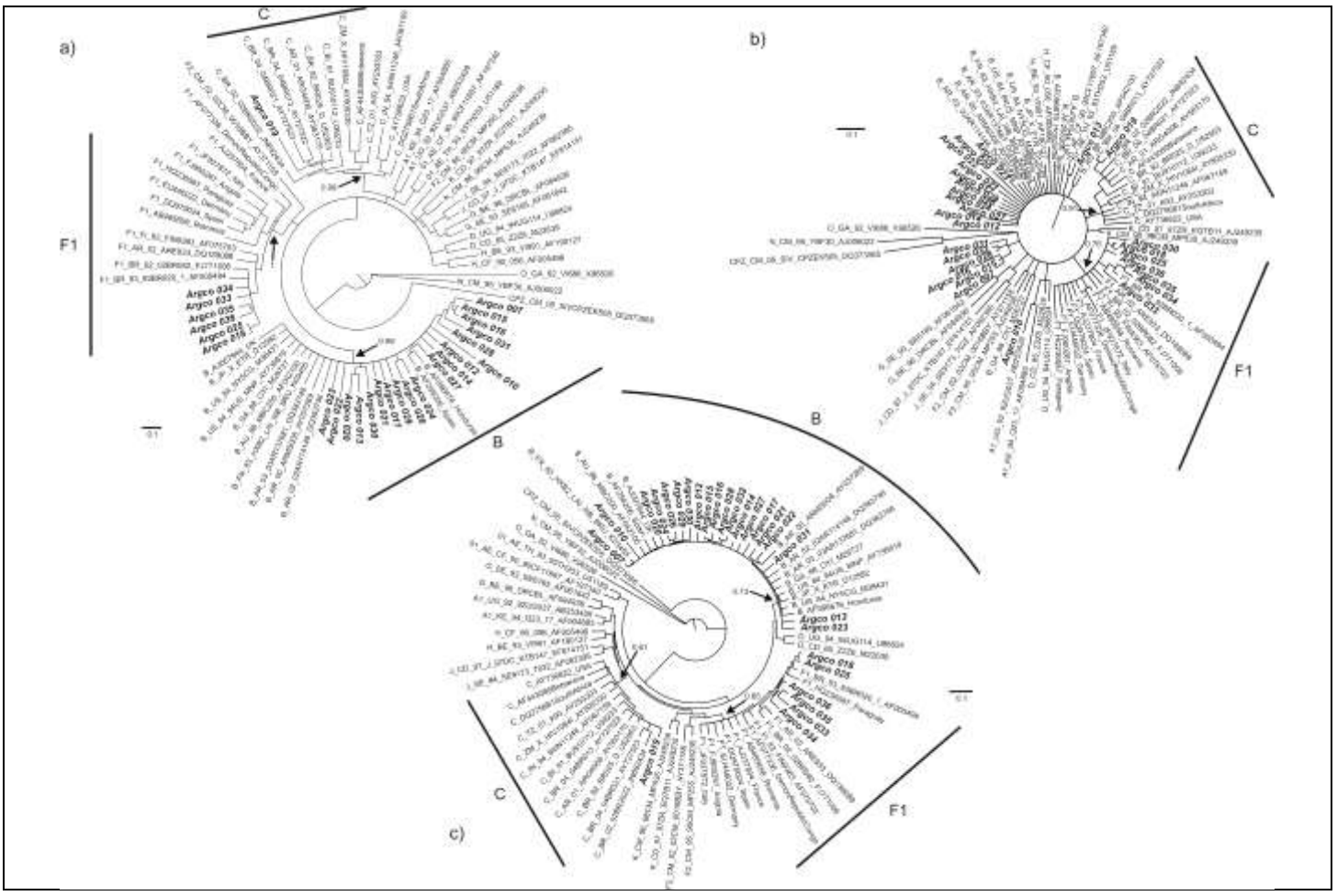

Fig. 5. Phylograms of the Bayesian consensus trees obtained from the three *env* gene data sets after ten million generations; the location of each data set is shown in Fig. 1: a) result of data set 4; b) result of data set 5; c) result of data set 6. The node supporting values are indicated only for the clades that include the HIV-1 strains characterized in this study (Argco).

Discussion

In the present study we determined the HIV-1 subtypes circulating in Córdoba since the earliest years of the epidemic by sequencing both the *pol* and *env* genes of strains belonging to patients diagnosed with the infection between 1986 and 2001. The numerous methods for detecting recombination breaking points and phylogenetic analyses revealed the presence of eighteen pure subtype B, seven B/F1, one C and one F1 (at least in the *env* gene) (Figs. 1, 2 and 3). These results are congruent with those of Pando et al.[15] that studied HIV-1 strains collected between 2006 and 2008 from volunteers (female sex workers and trans sex workers) from Córdoba city. The authors identified 5 strains as subtype B (29%), 9 as B/F1 (53%) and 3 as C (18%) (Dr. Maria Pando, personal communication). In Buenos Aires, until the beginning of the 1990's, subtype B was predominant, but a rapid phase of exponential growth allowed the increment in frequency of B/F1 recombinants[25]. Samples collected in that city from 1995 to 1998, showed 60% of subtype B, 40% of recombinant B/F1 viruses, and the pure F1 subtype was not detected[26]. Later studies detected that recombinant B/F1 subtypes increased to 76%-80%[7,27]. These data from Buenos Aires are consistent with those presented in this study from Córdoba, where a change in the epidemic can also be

observed: between 1986 and 1994 the circulating subtype was a 100% B (mainly in MSM or MSMW); whereas between 1995-2001 we found 56% B, 39% B/F1, 5% C and 5% F1, involving a more varied transmission route (Table 1). Although the sample size is limited, it is noticeable that the occurrence of B/F1 recombinants increased over the years replacing subtype B as the most frequent. This observation is in agreement with the fact that currently heterosexual infections (which are infected predominantly with B/F1 strains) are more common than homosexual ones (which are infected predominantly with B strains)[14].

Regarding the epidemic of subtypes C, B, F1 and recombinants B/F1 in South America[28,29,30,31], most studies stated the need of analyzing sequences of "pure" subtypes to determine if introductions in different countries occurred once or multiple times from different geographic origins. Following these recommendations, we divided the sequences of the *pol* and *env* genes into "pure" data sets. Two of the *pol* data sets were small sized, with insufficient phylogenetic signal to determine the geographic origin of the sequences (Fig 4b and c). However, data set 1 shows that F1 sequences from Córdoba are related to others from Brazil and Argentina, while subtype B sequences are clustered with others from multiple geographic origins (Fig. 4a). A similar result was obtained with the *env* gene where the three data sets show that sequences of subtype C and F1 from Córdoba are strongly related with other sequences from Argentina and Brazil, whereas sequences of subtype B are related to strains with multiple geographic origins and do not form a monophyletic cluster. According to Junqueira et al.[31] and Gilbert et al.[32] subtype B was introduced in the Caribbean directly from Africa around 1964-1966, and then spread to USA with an important expansion of the epidemic. Junqueira et al.[31] recovered three different clades of subtype B and hypothesized that dissemination of this subtype into South America occurred from both USA and the Caribbean countries. In more recent times, the introduction from Europe cannot be ignored. Therefore in South America there are several independent clades, compatible with different introduction events at different time points. The results obtained in the present study are in agreement with those observations, supporting the idea that sequences of subtype B from Córdoba would have different geographic origins.

Regarding to subtype F1, it is believed that Brazil was the center of the epidemic in South America. There are different estimations of the date of the most recent common ancestor (MRCA) for this subtype in the region: Aulicino et al.[33] suggested that it could be traced back until 1976 and Dilernia et al.[30] until 1969. In both studies a monophyletic origin for subtype F1 was proposed and the geographic origin would be Central Africa; this subtype probably colonized Argentina from Brazil. In the present study, all the data sets of the *env* gene and data set 1 of the *pol* gene showed that F1 fragments clustered with sequences from Brazil and Argentina, supporting previous hypotheses (Figs. 4a and 5). The introduction of subtype F1 could have occurred directly from Brazil or indirectly from Brazil to Buenos Aires and from Buenos Aires to Córdoba. Apparently, soon after its arrival to South America, the pure subtype F1 probably recombined with pure strains of subtype B and by the mid-1980s the new B/F1s arrived to Buenos Aires[34]. Dilernia et al.[30] estimated that the MRCA of Argentinean subtype B can be traced to 1972 (10 years before the first AIDS case was diagnosed in this country) and probably this subtype recombined with F1 in the late 70's. The recombinant CRF12_B/F1 would have a monophyletic origin which occurred in 1982. In the present study, we detected two strains that correspond to this recombinant form (Argco 018 and Argco 025) (Fig. 1). In the six data sets of the pure *pol* and *env* genes (Figs. 4 and 5), these two sequences clustered together in the trees, reinforcing the monophyletic origin proposed by Dilernia et al.[30]. However, more than one recombination event between subtypes B and F1 must have occurred, since several strains could be considered as URF_B/F1. For example, in the present study the sequence of Argco 035, a child infected in 2001 by vertical transmission, presented a mosaic structure identical to CRF12_B/F1 (Fig. 1), but the complete *pol* sequence was phylogenetically related to the recombinant form CRF38_B/F1 (described relatively recently in Uruguay[35]) and not to the sequences of CRF12_B/F1 (Fig. 2). Although, the sequences of pure *pol* and *env* genes of Argco 018, Argco 025 and

Argco 035 clustered with those of Argentina and Brazil, Argco 035 did not form a monophyletic group with the other two strains. Argco 035 shows a CRF12_B/F1 mosaic structure but the phylogenetic position suggest that another recombination event could have occurred (it must be considered that the complete genome of this strain was not obtained and it could have differences with CRF12_B/F1). It is also interesting to note that the strain of Argco 010 clustered (with moderate node support and different breaking points) with sequences of both CRF28_B/F1 and CRF29_B/F1 (Fig. 2) described in Brazil[36], providing more evidences that recombination between subtypes B and F1 occurred more than once.

In addition to the B and B/F1 strains, it is important to consider that the HIV-1 epidemic in Argentina is also composed by an appreciable proportion of subtype C. In the present study we found that the *env* sequence of Argco 019 clustered within subtype C. This HIV-1 strain was collected in 1998 and corresponded to a newborn child infected from vertical transmission; this collection date is one year earlier than the previously reported subtype C for Argentina[37,38]. Unfortunately, we could not sequence the *pol* gene of this strain to analyze if it corresponds to a pure subtype C or to the recombinants B/C and B/C/F, described in other studies from Argentina and Brazil[12,28,39]. According to the geographic origin and phylogenetic relationship, Jones et al.[28], using the *pol* gene, found that subtype C in South America could be divided in three clades, suggesting that at least three independent introductions occurred: a) most of the subtype C strains grouped in a major clade with a Brazilian origin (the clade named SAM), b) the clade that included one sequence from Venezuela with those from Zambia, Burundi and Bostwana, and c) the clade that clustered one strain from Argentina with sequences from Zimbabwe, Botswana, Tanzania, Zambia, Malawi and Senegal. In the present study, the *env* sequence of Argco 019 clustered with other sequences from Brazil and Argentina and would correspond to the SAM clade originated in Brazil.

In summary, the great variability in the subtype composition indicates that the HIV-1 epidemic in Córdoba was complex since its very beginnings. The historical approach used in the present analysis allowed us to identify the same relative progression pattern of HIV-1 infection detected in Buenos Aires, probably produced by the tight connection of people of the two most important cities of Argentina. We also detected a strong relationship between HIV-1 strains from Córdoba (Argentina) with those from Brazil, which would also reflect the close contact of these two countries, through tourism and commercial activities.

Acknowledgments

We thank Dr. N. Luna (Hospìtal Rawson) for collaborating with samples from her patients. We thank Dr. María Pando for providing unpublished results of her ongoing research. We thank Juan Pablo Carreras for his help in literature search. R.E. González Ittig is a Career Investigators of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

Bibliografía

1. Nájera R, Delgado E, Pérez-Alvarez L, Thomson MM. Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 2002;16:S3-16.

2. HIV Sequence Compendium; Los Alamos National Laboratory. En: http://hiv-web.lanl.gov; consultado el 17/12/2014.

3. Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol 2007;81:10209-19.

4. Campodónico M, Janssens W, Heyndrickx L, Fransen K, Leonaers A, et al. HIV type 1 subtypes in Argentina and genetic heterogeneity of the V3 region. AIDS Res Hum Retrov 1996;12:79-81.

5. Marquina S, Leitner T, Rabinovich RD, Benetucci J, Libonatti O, Albert J. Coexistence of subtypes B, F, and an B/F env recombinant of HIV type 1 in Buenos Aires Argentina. AIDS Res Hum Retrov 1996;12:1651-54.

6. Fernández-Medina D, Jansson M, Rabinovich RD, Libonatti O, Wigzell H. Identification of human immunodeficiency virus type 1 subtypes B and F, B/F recombinant and dual infection with these subtypes in Argentina. Scand J Infect Dis 1999;31:235-42.

7. Carr JK, Avila MM, Gómez-Carrillo M, Salomon H, Hierholzer J, et al. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. AIDS 2001;15:F41-47.

8. Thomson MM, Delgado E, Herrero I, Villahermosa ML, Vázquez-de Parga E, et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 B/F

intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. J Gen Virol 2002;83:107-119.

9. Aulicino PC, Kopka J, Rocco C, Mangano AM, Sen L. Sequence analysis of a South American HIV Type 1 BC Recombinant. AIDS Res Hum Retrov 2005;21:894-96.

10. Aulicino PC, Bello G, Guimaraes ML, Ruchansky D, Rocco C, et al. Longitudinal analysis of HIV-1 BF1 recombinant strains in vertically infected children from Argentina reveals a decrease in CRF12_BF pol gene mosaic patterns and high diversity of BF unique recombinant forms. Infect Genet Evol 2011;11:349-57.

11. Aulicino PC, Gómez-Carrillo M, Bello G, Rocco C, Mangano A, et al. Characterization of full-length HIV-1 CRF17_BF genomes and comparison to the prototype CRF12_BF strains. Infect Genet Evol 2012;12:443-47.

12. Pando MA, Eyzaguirre LM, Segura M, Bautista CT, Marone R, et al. First report of an HIV-1 triple recombinant of subtypes B, C and F in Buenos Aires, Argentina. Retrovirology 2006;3:59 doi:10.1186/1742-4690-3-59.

13. Bello G, Aulicino PC, Ruchansky D, Guimarães ML, Lopez-Galindez C, et al. Phylodynamics of HIV-1 Circulating Recombinant Forms 12_BF and 38_BF in Argentina and Uruguay. Retrovirology 2010;7:22 doi: 10.1186/1742-4690-7-22.

14. Pando MA, Eyzaguirre LM, Carrion G, Montano SM, Sanchez JL, et al. High genetic variability of HIV-1 in female sex workers from Argentina. Retrovirology 2007;4: 58 doi:10.1186/1742-4690-4-58

15. Pando MA, Gómez-Carrillo M, Vignoles M, Rubio AE, dos Ramos Farias MS, et al. Incidence of HIV Type 1 infection, antiretroviral drug resistance, and molecular characterization in newly diagnosed individuals in Argentina: a global fund project. AIDS Res Hum Retrov 2011;27:17-23.

16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772-80.

17. Siepel AC, Halpern AL, Macken C, Korber BTM. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. AIDS Res Hum Retrov 1995;11:1413-16.

18. Alcántara LCJ, Cassol S, Libin P, Deforche K, Pybus OG, et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res 2009;37:W634-42.

19. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 1999;73:152-60.

20. Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, et al. jpHMM: Improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res 2009;37:W647-51.

21. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 2006;23:1891-01.

22. Swofford DL. PAUP Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA. 2003.*

23. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 2012;9:772 doi:10.1038/nmeth.2109

24. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 2012;61:539-42.

25. Aulicino PC, Holmes EC, Rocco C, Mangano AM, Sen L. Extremely rapid spread of Human Immunodeficiency Virus Type 1 BF Recombinants in Argentina. J Virol 2007;81:427-29.

26. Thomson MM, Villahermosa ML, Vázquez de Parga E, Cuevas MT, Delgado E, et al. Widespread circulation of a B/F intersubtype recombinant form among HIV-1 infected individuals in Buenos Aires, Argentina. AIDS 2000;14:897-99.

27. Aulicino PC, Kopka J, Mangano AM, Rocco C, Iacono M, et al. Circulation of novel HIV type 1 A, B/C, and F subtypes in Argentina. AIDS Res Hum Retrov 2005;21:158-64.

28. Jones LR, Dilernia DA, Manrique JM, Moretti F, Salomón H, Gómez-Carrillo M. In-depth analysis of the origins of HIV Type 1 subtype C in South America. AIDS Res Hum Retrov 2009;25:951-59.

29. de Oliveira T, Pillay D, Gifford RJ. The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. PLoS One 2010;5:e9311 doi:10.1371/journal.pone.0009311

30. Dilernia DA, Jones LR, Pando MA, Rabinovich RD, Damilano GD, et al. Analysis of HIV Type 1 BF recombinant sequences from South America dates the origin of CRF12_BF to a recombination event in the 1970s. AIDS Res Hum Retrov 2011;27:569-78.

31. Junqueira DM, de Medeiros RM, Matte MCC, Araújo LAL, Chies JAB, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. PLoS ONE 2011;6:e27489 doi:10.1371/journal.pone.0027489.

32. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci USA 2007;104:18566-70.

33. Aulicino PC, Bello G, Rocco C, Romero H, Mangano A, et al. Description of the first full-length HIV-1 subtype F1 strain in Argentina: Implications for the origin and dispersion of this subtype in South-America. AIDS Res Hum Retrov 2007;23:1176-82.

34. Gómez-Carrillo M, Avila M, Hierholzer J, Pando MA, Martinez PL, et al. Mother-to-child HIV type 1 transmission in Argentina: BF recombinants have predominated in infected children since the mid-1980s. AIDS Res Hum Retrov 2002;18:477-83.

35. Ruchansky D, Casado C, Russi JC, Arbiza JR, Lopez-Galindez C. Identification of a new HIV Type 1 Circulating Recombinant Form (CRF38_BF1) in Uruguay. AIDS Res Hum Retrov 2009;25:351-56.

36. De Sa Filho DJ, Sucupira MC, Caseiro MM, Sabino EC, Diaz RS, Janini LM. Identification of two HIV type 1 circulating recombinant forms in Brazil. AIDS Res Hum Retrov 2006;22:1-13.

37. Gómez-Carrillo M, Salomón H, Pando MA, Kijak G, Avila MM. Distribución de subtipos y recombinantes del Virus de la Inmunodeficiencia Humana (HIV): situación en la Argentina. Medicina (Buenos Aires) 2001;61:881-89.

38. Carrión G, Eyzaguirre L, Montano SM, Laguna-Torres V, Serra M, et al. . Documentation of subtype C HIV type 1 strains in Argentina, Paraguay, and Uruguay. AIDS Res Hum Retrov 2004;20:1022-25.

39. Santos AF, Sousa TM, Soares EAJM, Sanabani S, Martinez AMB, et al. . Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. AIDS 2006;20:2011-19.