Heliyon 6 (2020) e05679

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Longitudinal characterization of HIV-1 pol-gene in treatment-naïve men-who-have-sex-with-men from acute to chronic infection stages

Cintia Cevallos [a,d], Andrés C.A. Culasso [b,d], Carlos Modenutti [e,f], Ana Gun [g], Omar Sued [g], María M. Avila [a,d], Diego Flichman [a,d], M. Victoria Delpino [c,d], Jorge Quarleri [a,d,*]

[a] Instituto de Investigaciones Biomédicas en Retrovirus y Sida (INBIRS), Facultad de Medicina, Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina
[b] Instituto de Bacteriología y Virología Molecular (IBaViM), Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina
[c] Instituto de Inmunología, Genética y Metabolismo (INIGEM), Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina
[d] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
[e] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina
[f] Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), CONICET, Buenos Aires, Argentina
[g] Fundación Huésped, Pasaje Angel Peluffo 3932, C1202ABB, Ciudad Autónoma de Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Genetics
Microbiology
Virology
Viruses
Viral disease
Viral genetics
HIV-1
Phylogeny
Pol-gene
Next-generation-sequencing
Viral diversity
MSM

ABSTRACT

HIV-1 is characterized by its ability to mutate and recombine even at polymerase (pol) gene. However, pol-gene diversity is limited due to functional constraints. The aim of this study was to characterize longitudinally, by next-generation sequencing (NGS), HIV-1 variants based on pol-gene sequences, at intra- and inter-host level, from acute/early to chronic stages of infection, in the absence of antiretroviral therapy. Ten men who have sex with men (MSM) were recruited during primary infection and yearly followed for five years. Even after a maximum of a five-year follow-up period, the phylogenetic analysis of HIV-1 pol-gene sequences showed a host-defined structured pattern, with a predominance of purifying selection forces during the follow-up. MSM had been acutely infected by different HIV-1 variants mainly ascribed to pure subtype B, or BF recombinant variants and showed different genetic mosaicism patterns that last until the chronic stage, representing a major challenge for prevention strategies.

1. Introduction

HIV-1 shows a highly diverse viral population at both levels, intra- and inter-host, essentially supported by low-fidelity reverse transcriptase, rapid turnover and increased virus production, and elevated recombination rate [1]. However, this genomic plasticity at pol-gene faces constraints due to the requirement to synthesize functional enzymes for viral replication. In antiretroviral (ART)-naïve hosts, this region is under weak pressure from the host-immune response [2, 3]. Consequently, the HIV-1 pol-gene based longitudinal analysis at fixed times is a valuable approach to study the viral-host interplay. At intra-host and mostly using the HIV-1 env-gene, viral variants heterogeneity at the chronic stage appears much greater than during the early/acute phase. However, men-who-have-sex-with-men (MSM), often exhibit more than one HIV-1 variant in the acute stage of infection [4] and a marked difference in the complex dynamics of the transmission bottleneck when compared to heterosexual individuals [5], being a particular host to observe HIV-1 evolution. In accordance with CDC statements [6], the Argentinean official data published in 2019 by the Ministry of Health has informed a prevalence between 12 and 15% in MSM [7]. The HIV-1 genomic characterization in Argentina is mostly represented by subtype B plus a wide variety of recombinant strains between subtypes B and F. Other less frequent HIV-1 subtypes (A, C, F) and recombinant forms (BC, BCF, A2D, and AG) have also been characterized. Particularly among Argentinean MSM, the genomic characterization of HIV-1 was performed by different cross-sectional studies involving phylogenetic inference [3, 8, 9, 10, 11, 12, 13].

* Corresponding author.
E-mail address: quarleri@fmed.uba.ar (J. Quarleri).

https://doi.org/10.1016/j.heliyon.2020.e05679
Received 22 September 2020; Received in revised form 9 November 2020; Accepted 3 December 2020

To study the HIV-1 complex evolutionary dynamics in vivo shaped by the interplay with the host, time fixed and genome wide data including multiple variants are essential. This longitudinal study aims to characterize HIV-1 variants, based on pol-gene sequences obtained by next-generation sequencing (NGS), at intra- and inter-host level, from acute/early to chronic stages of infection in ART-naïve MSM with emphasis in the dynamics of major and minor variants identified during transmission.

2. Material and methods

2.1. Patients, sampling, HIV-1 RNA isolation, and RT-PCR

Fifty-one blood samples were collected yearly for 5 years, between 2008 and 2014, from ten (P1 to P10) ART-naïve MSM belonging to the *Grupo Argentino de Seroconversion* [14]. As inclusion criteria, the participants did not receive antiretroviral treatment throughout the study following the Treatment Guidelines in force at that time, and at least five consecutive viral load and CD4 +T cell measurements were carried out yearly. The median [range] estimated time elapsed since infection was 3.9 [1.5–9.8] months (Fiebig V, n = 4; Fiebig VI, n = 6). All participants offered their informed consent, and the study was approved by the Ethics Committee of Huesped Foundation, Argentina.

Total RNA was extracted from 200 μl each patient's plasma sample and followed by cDNA synthesis. The QIAamp Viral RNA Mini Kit (Qiagen, Gaithersburg, MD) and the SuperScript™ IV First-Strand Synthesis System (Invitrogen, Carlsbad, CA) were used respectively and according to manufacturers' instructions. HIV-1 pol-gene (nucleotide 2253 to 3749 following HXB2 numbering) was amplified using specific primer pairs [15].

Absolute CD4+T lymphocytes count determination was performed by flow cytometry (FACS Calibur, BD Company, USA) within 24 h after blood sample processing. Plasma HIV-1 RNA viral load was measured using VERSANT HIV-1 RNA version 3.0 bDNA Assay, Siemens Diagnostics (range of 1.7–5.7 log copies/mL) according to the manufacturer's recommendations.

2.2. MiSeq (Illumina) next generation sequencing and viral bioinformatics

Samples were sequenced on Illumina MiSeq platform according to the manufacturer's instructions (Illumina). We performed 32 paired-end (2 × 150) sequencing runs. Overall, 1.39 Gbases were obtained (8,083,663 sequence reads). Median number of reads per amplicon was 91,994 (range 67,185 to 156,919). All DNA sequence files are available from the GenBank database under SRA accession: PRJNA591115. Each variant was validated by a minimal coverage of 3,000 reads, and a quality Phred scores above 30. Minority variant was defined as those present between 1% and 20% of the total viral population.

Raw reads were subjected to quality assessment using Casavav.1.8 software. The reads were analyzed by UGENE [8] with workflow analysis, including (i) Elimination of adapter sequences by Cutadapter [9]; (ii) Removal of all reads containing nucleotides with a Phred score <20 throughout FASTX toolkit (htpp://hannonlab.eshl.edu/fastx_toolkit); (iii) Mapping of reads onto the HIV-1 reference sequences from GenBank using Burrows-Wheeler (BWA MEM) [16], thus obtaining the BAM file; and (iv) Filtering and sorting of mapped reads with Samtools [17].

The BAM files were converted to base calling file (BCF) using the *mpileup* command (Samtools) and processed with an AWK program (bcv2csv.awk) to summarize by nucleotide position the read depth and the number of A, C, G, and T.

2.3. Analysis of the codon position specific sequence diversity and dynamics

An additional AWK program (codon-pos-filter) was created to automate the analysis of the different codon positions (1st, 2nd, and 3rd). The program computed the per site mean pairwise distance of each nucleotide position with a minimum depth (arbitrarily set as 100 reads), then summed up all distances that belonged to the first, the second and the third codon positions, and finally each sum was divided by the number of first, second and third positions. The result of this program indicates the mean pairwise distance at each nucleotide position and the number of times that this position was read for each reference sequence. The diversity at each codon position was calculated by the codon-pos-filter.awk. For nucleotide, sequences of each individual, a regression analysis was carried out with the specific codon position mean pairwise distances and the putative time of infection. The regression analysis and plots were carried out with GraphPad Prism 7. The AWK files are available at github (https://git.io/fj4OX).

2.4. Phylogenetic, and inter-subtype recombination analysis based on assembled HIV-1 pol-gene sequences

The HIV-1 variants based on pol-gene datasets were reconstructed employing QuRe v0.9994 [18]. Thus, 214 curated pol-gene assembled sequences corresponding to the entire protease (codons 1–99) and partial RT (codons 1–252) were generated. Using these assembled pol-gene sequences datasets, phylogenetic analysis was carried-out by the maximum-likelihood method implemented in PhyML v3.0 [19]. A general time reversible (GTR+Γ+I) model of nucleotide evolution and gamma-distributed rate variation among sites was set according to the SMS software [20]. Branch support was assessed by bootstrapping using 1000 replications (≥70% were considered significant).

Recombination analysis was done using REGA, RIP, RDP4, and SCUEAL programs [21, 22, 23, 24, 25]. A recombination event was considered when at least three methods used matched. Sankey plots were used for each host to represent the dynamic profile of the HIV-1 variants over time (http://sankeymatic.com/).

2.5. Selection pressure analysis and viral diversity

Region-wide natural selection was evaluated by Z-Test carried-out with MEGA-X software [26, 27], after the estimation of dN-dS statistic with the Nei-Gojobori method. Variance was estimated by bootstrapping (1000 replicates) [28].

The dN/dS ratios of every codon site were calculated by SLAC, FEL, FUBAR, and MEME [29, 30, 31] methods implemented in Datamonkey server (HyPhy program version 2.5) [32,33,34].

Diversity was calculated as the mean pairwise genetic differences among the full dataset of pol sequences at a defined sampling time from each individual. For this, the average distance between base pairs was obtained through the calculation of the frequency of the bases for each nucleotide position, implementing the formula:

$$P_i = \frac{\left(1 - p_A^2 + p_C^2 + p_G^2 + p_T^2\right) \times N_i}{(N_i - 1)}$$

where p_A, p_C, p_G y p_T are the proportions of A, C, G, T respectively, N_i is the total number of readings for a nucleotide position. For diversity, only sites with minor allele frequency greater than 0.001% were considered.

3. Results

3.1. Longitudinal HIV-1 pol-gene characterization by inter and intra-host phylogenetic, and inter-subtype recombination analysis

Most individuals showed significant raises in their viral load level during the follow-up, which may reflect viral fitness optimization (data not shown). The level of plasma viral load (mean ± SD) among participants infected by HIV subtype B and BF_recombinants forms was 4.9 ± 2.6 and 5.1 ± 2.2 (log copies/ml), respectively. Their median (range) CD4+ T cell counts were 498 cells/μl (738-6) and 680 cell/μl (1185-

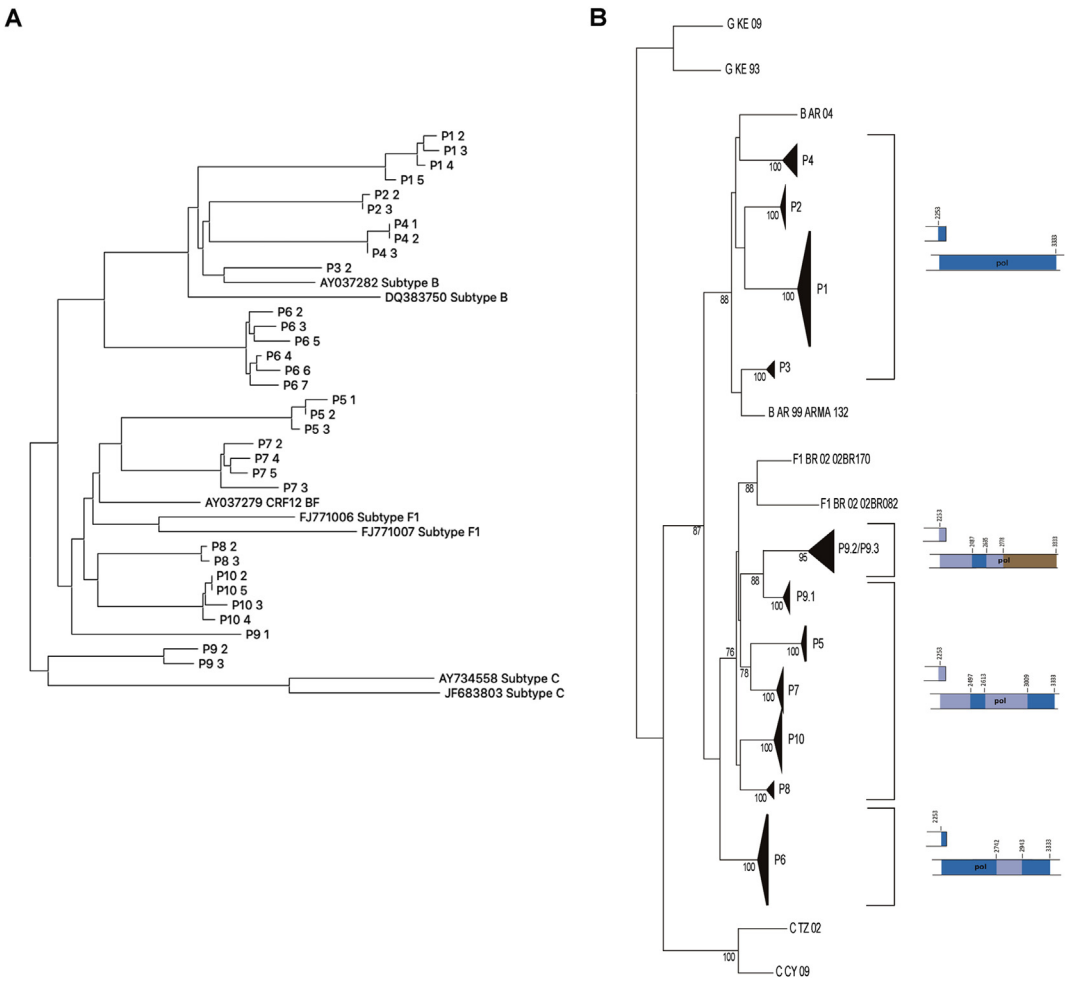


Figure 1. A. Evolutionary history inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Maximum Composite Likelihood method. This analysis involved consensus pol-gene nucleotide sequences obtained from each participant (P) at different sampling times during the follow-up. HIV-1 reference sequences (named by their Gene Bank accession number) for different subtype and CRF12_BF recombinant form were included. **B. Inter-host phylogenetic and recombination analysis of HIV-1.** Maximum likelihood tree for the HIV-1 pol-gene sequences obtained from all hosts (P) along the follow-up. Branch lengths are proportional to the number of substitutions per aligned position. Numbers close to branch correspond to bootstrap support values. The tree was midpoint rooted. Pol-gene mosaicism illustration was designed using the Recombinant HIV-1 Drawing Tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html).

263), respectively. In most individuals, the decrease in CD4 T cell level was gradual and slight, except for participant P2 who exhibited a steeper descent during the follow-up. The amplification of pol-gene fragments from plasma samples with no measurable HIV viral load was precluded.

Based on assembled pol-gene sequences (n = 214), and in accordance with our previous report [10], the recombination analysis revealed the presence of CRF12_BF-like genetic mosaicism (P5, P7, P8, P9 .1 and P10) and pure B subtype (P1, P2, P3, and P4) variants. In addition, two other genetic mosaicisms were found: the CRF89_BF-like pattern (P6), and a recombinant FBC complex (P9.2 and P9.3) (Figure 1). Moreover, the inter-host HIV-1 phylogenetic analysis showed a strongly supported (bootstrap ≥98) host-defined viral genetic structuring, appearing as monophyletic groups (Figure 1).

Phylogenetic analysis at intra-host, showed multivariant HIV-1 infection in all cases. Major and minor viral variants (>20%, and ≤20% of abundance, respectively) coexist regardless of the infection stage, appearing intermingled, and closely related among each other. Interestingly, for the P9 a phylogenetic shift was observed between the viral variants of the early/acute and those of the chronic stage (Figure 2). In this case, a plausible superinfection event may have occurred favoring new stochastic forces such as recombination events. Further support for the superinfection event was obtained after the estimation of evolutionary divergence between sequences (P9.1 vs. P9.2 and P9.3) using the Maximum Composite Likelihood model in MEGA X [26, 27]. The mean number of base substitutions per site from between sequences was 0.055 (0.043–0.066).

3.2. Intra-host longitudinal analysis of HIV-1 variants based on pol-gene short reads sequencing data

The intra-host HIV-1 genomic characterization, using UGENE software as an additional platform, was performed, defining the relative abundance at every nucleotide position by short read-based analysis (n = 8,083,663 reads). The output graphics reveal that viral variants identified in all ascribed to pure subtype B cases (P1, P2, P3, and P4) showed a high degree of sequence homology at intra-host (Figure 3). Likewise, the short read-based analysis of each viral variant characterized as inter-subtype recombinant forms (P5, P6, P7, P8, P9, and P10) clearly support the predominant pol-gene mosaicism pattern. Except for P6, where the coexistence of validated (≥1000 reads) minor viral variants with alternative genetic mosaicism, was inferred. These plausible minor variants have appeared when at defined breakpoints, the recombination

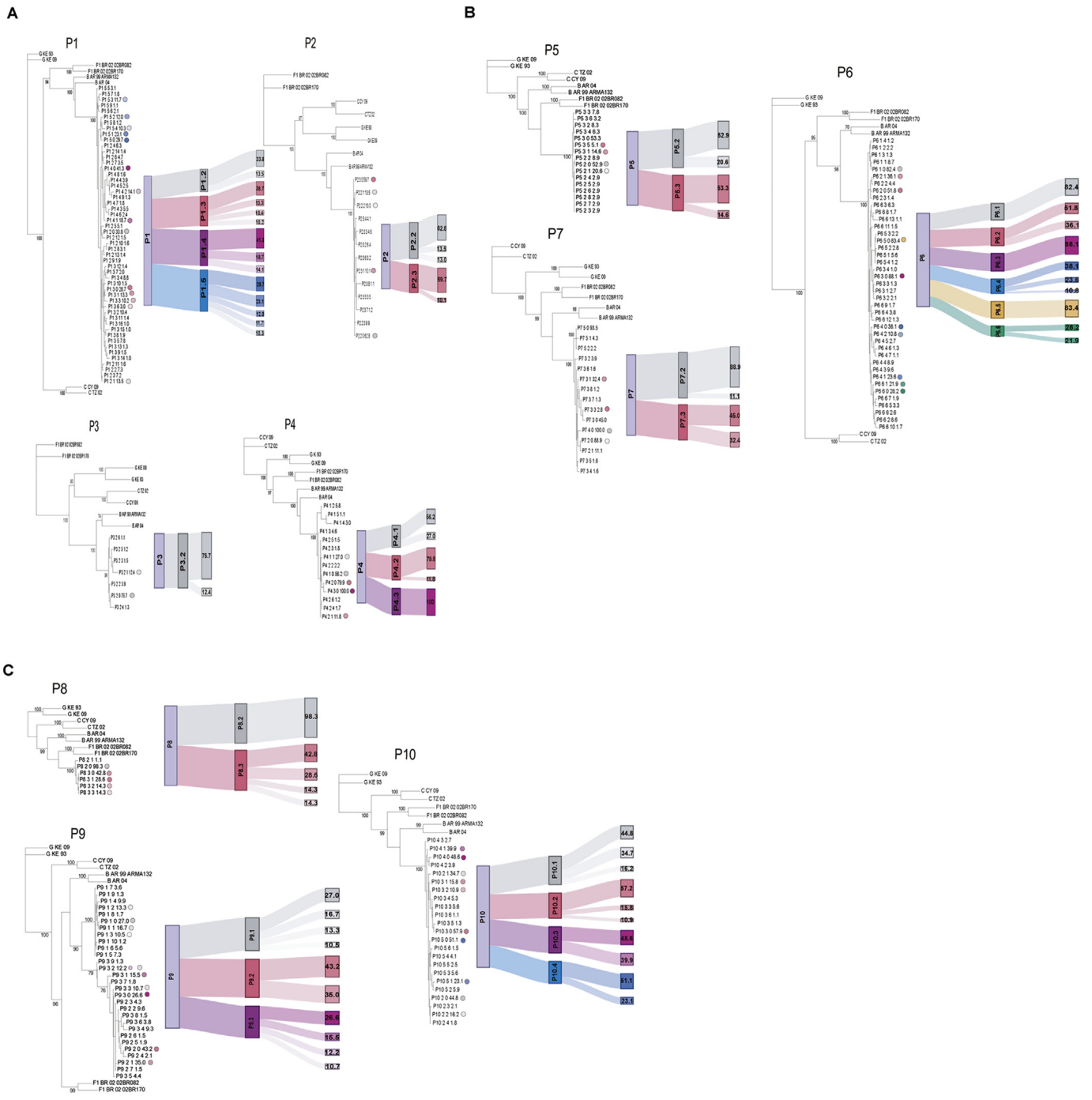


Figure 2. (A, B, and C). Individual phylogenetic analysis of HIV-1. Maximum likelihood trees for the HIV-1 pol-gene sequences obtained from each host (P) along the follow-up. Branch lengths are proportional to the number of substitutions per aligned position. Numbers close to branch correspond to bootstrap support values. The tree was midpoint rooted. Each viral variant is named according to its sampling time, and relative abundance (as percentage). The viral variants more abundant at each time (above 10%) are identified by a color circle that corresponds with the Sankey plots, that represent the relative abundance of viral variants.

may have not occurred (Figure 3). In P5 and P8, although the predominant variant did not show a CRF_12BF-like pattern, its presence could be deduced as a minor variant (Figure 3).

3.3. Natural selection and viral diversity analysis

The natural selection and viral diversity analysis using the Z-test were performed separately for pure subtype B, and BF recombinants. The dS–dN statistics were 5.67 (subtype B) and 8.57 (BF recombinant), indicating a higher rate of synonymous than non-synonymous substitutions. Negative selection predominates when tested for purifying selection and the null hypothesis of neutrality was rejected ($p < 0.001$). The 3rd codon position had significantly higher diversity compared to the 1st ($p < 0.001$) and the 2nd ($p < 0.001$) (Figure 4). Besides, the dN/dS ratio at intra-host denoted positive selection ($p < 0.05$) only in codon 22 (P6) and 74 (P1), both at protease gene.

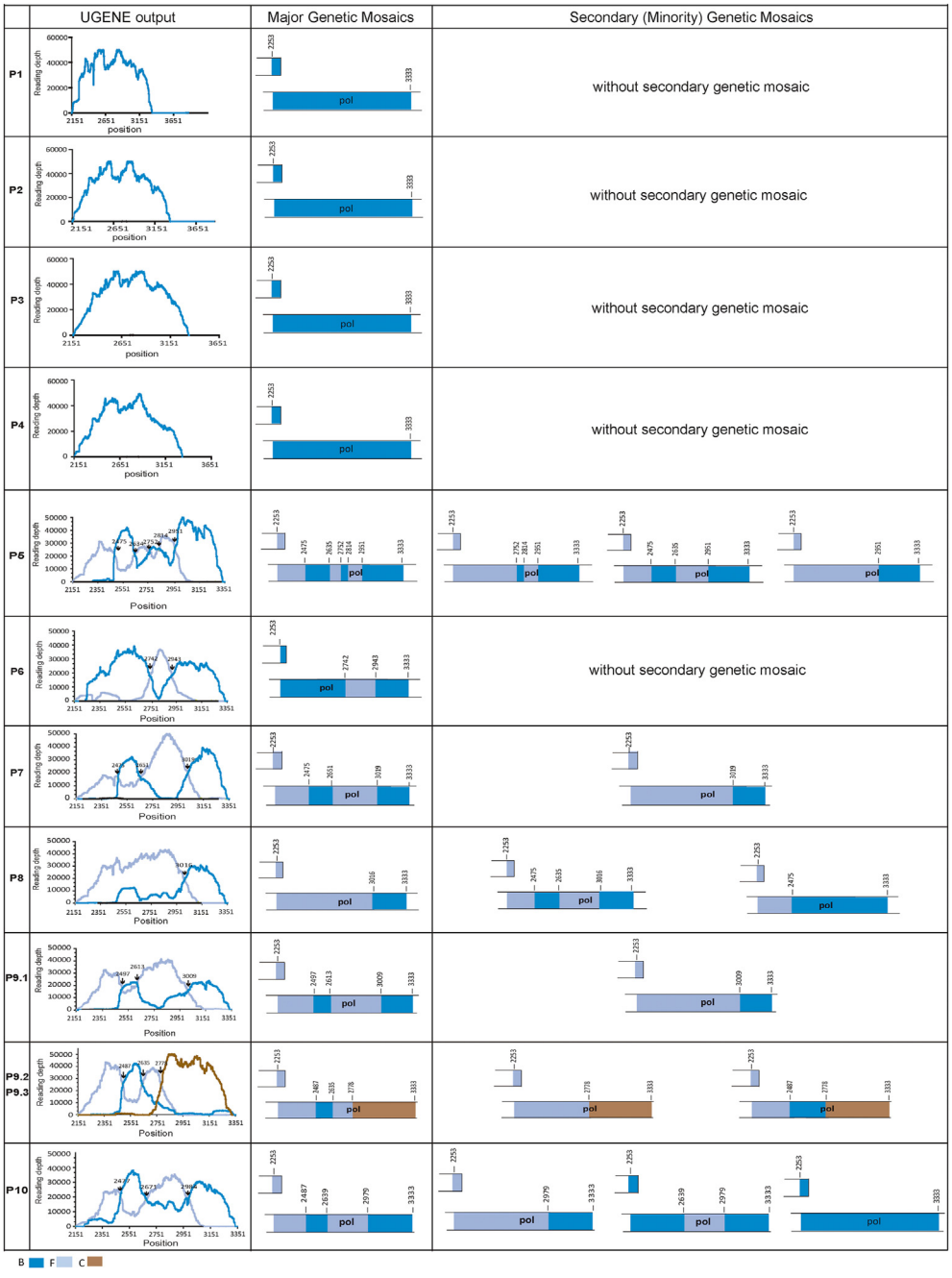


Figure 3. Recombinant structure of pol gene from all recombinant samples along the course of infection. Representative graphics obtained from UGENE are shown on the left based on the number of reads for each position on the pol-gene sequence. HIV-1 pol-gene mosaicism of the majority (center), and minority (right) variants, are represented. Mosaics were constructed implementing the Recombinant HIV-1 Drawing Tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html).

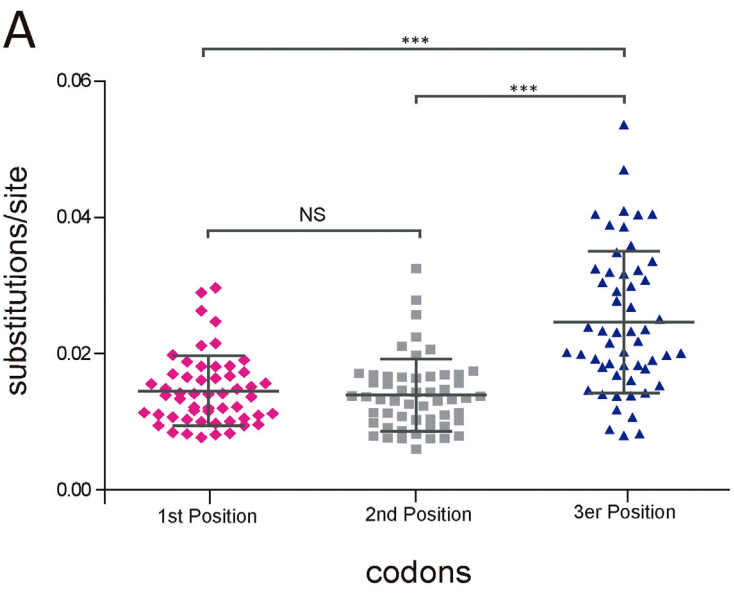
4. Discussion

HIV-1 infection is characterized by an early acute phase, occasionally symptomatic, followed by an asymptomatic period of inconstant duration that ends in clinically overt immunodeficiency. Here, having samples obtained from the early stage and, from there, for several years to chronicity, we were able to analyze HIV-1 variants based on pol-gene sequences obtained by next-generation sequencing, at the intra- and inter-host level., in the absence of antiretroviral therapy among MSM. It allowed inferring the presence and dynamics of minor variants coexisting with the more abundant ones through a bottleneck influenced by the host (immune response) and viral restrictions of polymerase factors (enzymatic functions) in the absence of antiretroviral therapy pressure.

The characterization of HIV-1 variants revealed the presence of pure subtype B and BF_inter-subtype recombinant forms. Among the BF recombinant forms, the most frequently identified genetic mosaicism at pol-gene was CRF12_BF-like pattern, and -to a lesser extent-a CRF89_BF like pattern. Other genetic mosaicism compatible with unique recombinant forms were also inferred. In line with previous reports, we characterized recombination hot spots throughout the pol-gene among subtypes B, C, and F reference sequences [35].

Throughout the pol-gene mosaicism, we have identified recombination hot spots, suggesting that recombination *in vivo* may occur preferentially during reverse transcription through the strand displacement-assimilation model [36].

The inter-host phylogenetic relatedness of HIV-1 revealed a robust phylogenetic host-dependent clustering, either among HIV-1 variants characterized as pure subtype B, or BF recombinants. However, at intra-host level, no temporally structured phylogenies were observed, and both major and minor viral variants characterized at acute/early or chronic stage of infection appear intermingled. Thus, in the absence of



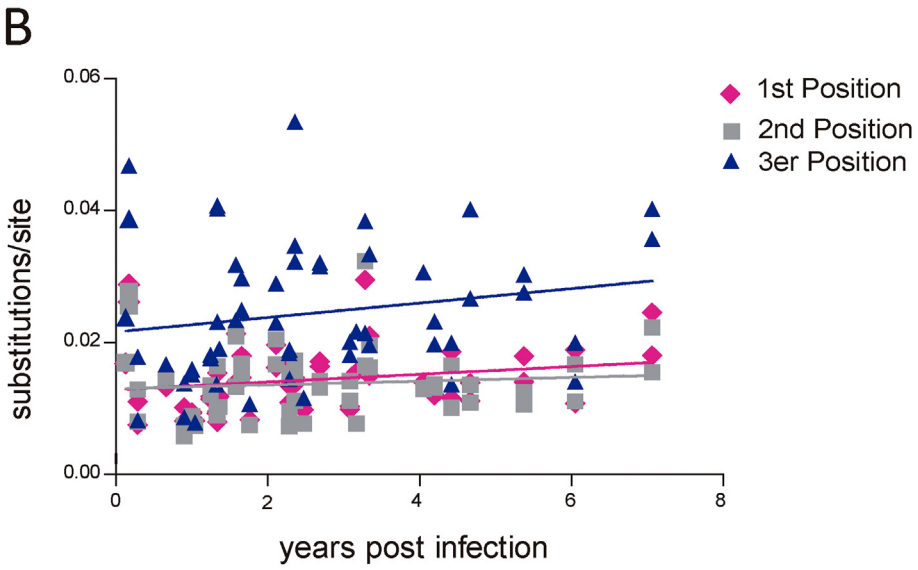


Figure 4. A. Global diversity determined on the 1st, 2nd and 3rd codons positions as the mean pairwise distance on the pol gen. The points represent the substitution/sites on the 1st, 2nd, and 3rd codon positions respectively. The mean ± standard deviation for n = 54. The significant differences ($p < 0.001$) between 1st and 3rd positions as well as between 2nd and 3rd positions are indicated with an asterisk (Friedman test followed by Dunn test). **B. Mean pairwise diversity at 1st, 2nd, and 3rd codon positions over time for all individuals (highlighted in pink, gray and violet).** The average change in mean pairwise diversity over time was inferred by linear regression.

ART the natural selection on pol-gene is mainly driven by purifying than diversifying forces [37, 38, 39]. In accordance, we have observed that viral diversity increases as infection progressed consistent with a strong purifying selection and constraints related to fitness cost at pol-gene. Hence, diversity in pol-gene was mostly synonymous and accumulates at 3rd codon position, while at 1st and 2nd codon positions it remained low during follow-up and were mostly non-synonymous changes. This finding reflects a steady state of the pol-gene at the population level during acute and chronic infection stages. Furthermore, it may be related with fitness cost associated with the mutation in pol-gene that constrains its occurrence limiting the HIV-1 evolutionary dynamics [40]. Interestingly, a recently published longitudinal study among patients who underwent ART at the time of primary HIV-1 infection with subsequent effective long-term response (undetectable HIV viral load during at least 5 years) showed tree topologies depicting an absence of segregation between sequences by phylogenetic analysis (based on reverse transcriptase gene and partial env-gene sequences) since viral populations from all time-points were intermingled. Comparison of the average pairwise distance showed the absence of significant viral diversity at the time of primary infection and afterwards during 5 years of full virological control under ART. There was no clear evidence of viral evolution during a prolonged period of time in this population of highly controlled adult patients treated at time of primary infection [41].

At intra-host level, the presence of minor variants displaying different mosaicisms at pol-gene were observed, which supports, according to previous reports, multivariant HIV-1 infections in MSM [4, 42]. In most individuals, the intra-host dynamics of the HIV-1 variants showed no detectable shifts in major variants, suggesting that no changes have occurred in the fitter viral variant during adaptive evolution. However, in P9 a drastic change in the predominant sequence of the pol-gene, that appeared as a phylogenetic shift, was observed and could reflect a superinfection event, which could imply new stochastic forces such as recombination events. The in-depth characterization of the virus

population in plasma at acute/early stage of infection (freely circulating HIV-1 variants) did not showed any evidence of viral variants containing pol-gene sequences needed for the newer mosaic. Moreover, after estimating the genetic distance of pol-gene sequences obtained at different time points, we also have a further support in the occurrence of this event since 5% genetic distance is a hallmark of superinfection [43]. Nevertheless, it cannot not rule out its eventual prior existence as latency stage able to emerge after reactivation and to recombine generating adapted and persistent variants.

This study has limitations. The included cohort is relatively small. Moreover, the source HIV-1 was only plasma samples and peripheral blood mononuclear cells as a plausible source for retrieving stored latent variants were not analyzed. Although this is focused on a particular genomic region of the viral genome, methodologically, the bulk amplification step instead of limited dilution or single genome analysis (SGA) may introduce a bias during library construction prior to NGS for studying virus populations, and the pol-sequence variation could be underrepresented [44].

Nevertheless, in this longitudinal study using time fixed and HIV-1 pol gene wide data including multiple variants, we were able to compare early and late time points of the viral evolution at intra and inter-host level, which allowed to depict the HIV-1 dynamics in vivo shaped by the interplay with the host.

In conclusion, the HIV-1 transmission among MSM involves multiple pol-gene viral variants coexisting at the acute stage of infection. Then, up to the chronic phase and without antiretroviral therapy pressure, these viral variants display a phylogenetic structured-pattern in a host-dependent manner, without evidence of shifts in their dominance. This multivariant HIV-1 infection in MSM represents an important challenge for HIV-1 prevention strategies in this particular vulnerable population, particularly when considering pre-exposure prophylaxis to control the viral diversity.

Declarations

Author contribution statement

Cintia Cevallos, Andrés C. A. Culasso: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Carlos Modenutti: Performed the experiments; Analyzed and interpreted the data.

Ana Gun, Omar Sued: Analyzed and interpreted the data.

María M. Avila, Diego Flichman, M. Victoria Delpino: Analyzed and interpreted the data; Wrote the paper.

Jorge Quarleri: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (Argentina) from grant PICT-2015-1921 (to JQ) and by Universidad de Buenos Aires (to JQ grant Nº20020150100033BA).

Data availability statement

Data associated with this study has been deposited at Sequence Read Archive (SRA) under the accession number PRJNA591115.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] A. Rambaut, D. Posada, K.A. Crandall, E.C. Holmes, The causes and consequences of HIV evolution, Nat. Rev. Genet. 5 (1) (Jan 2004) 52–61.
[2] J. Snoeck, J. Fellay, I. Bartha, D.C. Douek, A. Telenti, Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints, Retrovirology 8 (Nov 1 2011) 87.
[3] J.M. Coffin, HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy, Science 267 (5197) (Jan 27 1995) 483–489.
[4] H. Li, K.J. Bar, S. Wang, et al., High multiplicity infection by HIV-1 in men who have sex with men, PLoS Pathog. 6 (5) (May 13 2010), e1000890.
[5] D.C. Tully, C.B. Ogilvie, R.E. Batorsky, et al., Differences in the selection bottleneck between modes of sexual transmission influence the genetic composition of the HIV-1 founder virus, PLoS Pathog. 12 (5) (May 2006), e1005619.
[6] CDC. https://www.cdc.gov/hiv/risk/analsex.html, 2019.
[7] A.N.H. Ministry, AIDS and STD Department. 36.Th Bulletin of HIV-AIDS in Argentina, 2019. Buenos Aires, https://www.aahitc.org.ar/wp-content/uploads/2019/12/0000001754cnt-boletin-epidemiologico-2019_vih-sida-its.pdf.
[8] K. Okonechnikov, O. Golosova, M. Fursov, U. team, Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics 28 (8) (Apr 15 2012) 1166–1167.
[9] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet J. 17 (2011) 1.
[10] C.G. Cevallos, L.R. Jones, M.A. Pando, J.K. Carr, M.M. Avila, J. Quarleri, Genomic characterization and molecular evolution analysis of subtype B and BF recombinant HIV-1 strains among Argentinean men who have sex with men reveal a complex scenario, PloS One 12 (12) (2017), e0189705.
[11] D.A. Dilernia, A.M. Gomez, L. Lourtau, et al., HIV type 1 genetic diversity surveillance among newly diagnosed individuals from 2003 to 2005 in Buenos Aires, Argentina, AIDS Res. Hum. Retrovir. 23 (10) (Oct 2007) 1201–1207.
[12] M. Segura, S. Sosa Estani, R. Marone, et al., Buenos Aires cohort of men who have sex with men: prevalence, incidence, risk factors, and molecular genotyping of HIV type 1, AIDS Res. Hum. Retrovir. 23 (11) (Nov 2007) 1322–1329.
[13] M.A. Pando, M. Gomez-Carrillo, M. Vignoles, 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. Retrovir. 27 (1) (Jan 2011) 17–23.
[14] M.E. Socias, O. Sued, N. Laufer, et al., Acute retroviral syndrome and high baseline viral load are predictors of rapid HIV progression among untreated Argentinean seroconverters, J. Int. AIDS Soc. 14 (Aug 10 2011) 40.
[15] M. Gomez-Carrillo, J.F. Quarleri, A.E. Rubio, et al., Drug resistance testing provides evidence of the globalization of HIV type 1: a new circulating recombinant form, AIDS Res. Hum. Retrovir. 20 (8) (Aug 2004) 885–888.
[16] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinf. 26 (5) (Mar 1 2010) 589–595.
[17] H. Li, B. Handsaker, A. Wysoker, et al., The sequence alignment/map format and SAMtools, Bioinf. 25 (16) (Aug 15 2009) 2078–2079.
[18] M.C. Prosperi, M. Salemi, QuRe: software for viral quasispecies reconstruction from next-generation sequencing data, Bioinf. 28 (1) (Jan 1 2012) 132–133.
[19] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (3) (May 2010) 307–321.
[20] V. Lefort, J.E. Longueville, O. Gascuel, SMS: smart model selection in PhyML, Mol. Biol. Evol. 34 (9) (Sep 01 2017) 2422–2424.
[21] D.P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: detection and analysis of recombination patterns in virus genomes, Virus Evol. 1 (1) (2015) vev003.
[22] D.P. Martin, B. Murrell, A. Khoosal, B. Muhire, Detecting and analyzing genetic recombination using RDP4, Methods Mol. Biol. 1525 (2017) 433–460.
[23] T. de Oliveira, K. Deforche, S. Cassol, et al., An automated genotyping system for analysis of HIV-1 and other microbial sequences, Bioinformatics 21 (19) (Oct 1 2005) 3797–3800.
[24] W. Delport, A.F. Poon, S.D. Frost, S.L. Kosakovsky Pond, Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology, Bioinf. 26 (19) (Oct 1 2010) 2455–2457.
[25] A.C. Siepel, A.L. Halpern, C. Macken, B.T. Korber, A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences, AIDS Res. Hum. Retrovir. 11 (11) (Nov 1995) 1413–1416.
[26] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (Jun 1 2018) 1547–1549.
[27] G. Stecher, K. Tamura, S. Kumar, Molecular evolutionary genetics analysis (MEGA) for macOS, Mol. Biol. Evol. 37 (4) (Apr 1 2020) 1237–1239.
[28] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (5) (Sep 1986) 418–426.
[29] B. Murrell, S. Moola, A. Mabona, et al., FUBAR: a fast, unconstrained bayesian approximation for inferring selection, Mol. Biol. Evol. 30 (5) (2013) 1196–1205.
[30] S.L. Kosakovsky Pond, S.D. Frost, Not so different after all: a comparison of methods for detecting amino acid sites under selection, Mol. Biol. Evol. 22 (5) (2005) 1208–1222.
[31] B. Murrell, J.O. Wertheim, S. Moola, T. Weighill, K. Scheffler, S.L. Kosakovsky Pond, Detecting individual sites subject to episodic diversifying selection, PLoS Genet. 8 (7) (2012), e1002764.
[32] S.L. Pond, S.D. Frost, S.V. Muse, HyPhy: hypothesis testing using phylogenies, Bioinf. 21 (5) (Mar 01 2005) 676–679.

[33] S.L. Kosakovsky Pond, A.F.Y. Poon, R. Velazquez, et al., HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies, Mol. Biol. Evol. 37 (1) (Jan 1 2020) 295–299.
[34] S. Weaver, S.D. Shank, S.J. Spielman, M. Li, S.V. Muse, S.L. Kosakovsky Pond, Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes, Mol. Biol. Evol. 35 (3) (Mar 1 2018) 773–777.
[35] M.A. Pando, L.M. Eyzaguirre, M. Segura, et al., First report of an HIV-1 triple recombinant of subtypes B, C and F in Buenos Aires, Argentina, Retrovirology 3 (Sep 7 2006) 59.
[36] G. Magiorkinis, D. Paraskevis, A.M. Vandamme, E. Magiorkinis, V. Sypsa, A. Hatzakis, In vivo characteristics of human immunodeficiency virus type 1 intersubtype recombination: determination of hot spots and correlation with sequence similarity, J. Gen. Virol. 84 (Pt 10) (Oct 2003) 2715–2722.
[37] T. Leitner, J. Albert, The molecular clock of HIV-1 unveiled through analysis of a known transmission history, Proc. Natl. Acad. Sci. U. S. A. 96 (19) (Sep 14 1999) 10752–10757.
[38] W. Fu, J.M. Akey, Selection and adaptation in the human genome, Annu. Rev. Genom. Hum. Genet. 14 (2013) 467–489.
[39] P.C. Sabeti, S.F. Schaffner, B. Fry, et al., Positive natural selection in the human lineage, Science 312 (5780) (Jun 16 2006) 1614–1620.
[40] F. Zanini, V. Puller, J. Brodin, J. Albert, R.A. Neher, In vivo mutation rates and the landscape of fitness costs of HIV-1, Virus Evol. 3 (1) (Jan 2017) vex003.
[41] B. Abdi, T. Nguyen, S. Brouillet, et al., No HIV-1 molecular evolution on long-term antiretroviral therapy initiated during primary HIV-1 infection, AIDS 34 (12) (Oct 1 2020) 1745–1753.
[42] S.L. Lamers, M. Salemi, D.C. Galligan, et al., Extensive HIV-1 intra-host recombination is common in tissues with abnormal histopathology, PloS One 4 (3) (2009), e5065.
[43] M.V. Sivay, P.J. Palumbo, Y. Zhang, et al., HIV drug resistance, phylogenetic analysis, and superinfection among men who have sex with men and transgender women in sub-Saharan Africa: HPTN 075, Clin. Infect. Dis. (Aug 6 2020).
[44] V.F. Boltz, J. Rausch, W. Shao, et al., Ultrasensitive single-genome sequencing: accurate, targeted, next generation sequencing of HIV-1 RNA, Retrovirology 13 (1) (Dec 20 2016) 87.