Article

Not peer-reviewed version

# Whole-Genome Sequencing–Based Characterization of HIV-1 in a Tertiary-Care Hospital in Mexico: Drug-Resistance Mutations and Genetic Diversity

Eduardo García-Moncada [†], Enoc Mariano Cortés-Malagón [†], Jesús Alejandro Pineda-Migranas, Montserrat Ruiz Santana, Iliana Alejandra Cortés-Ortíz, José Francisco Escutia Domínguez, Daniel Agustín Bravata-Alcántara, Gustavo Acosta-Altamirano, Saúl David Razo-González, Manuel Alberto Castillo Mendez, Mónica Sierra-Martínez [*], Juan Carlos Bravata-Alcántara [*]

Posted Date: 6 May 2026

doi: 10.20944/preprints202605.0280.v1

Keywords: HIV-1 molecular epidemiology; whole-genome sequencing; circulating recombinant forms; transmitted drug resistance; BF recombinants; genomic surveillance; Mexico

 doi:10.20944/preprints202605.0280.v1



*Article*

# Whole-Genome Sequencing–Based Characterization of HIV-1 in a Tertiary-Care Hospital in Mexico: Drug-Resistance Mutations and Genetic Diversity

**Eduardo García Moncada [1,†], Enoc Mariano Cortés-Malagón [2,3,†], Jesús Alejandro Pineda-Migranas [1], Montserrat Ruiz Santana [4], Iliana Alejandra Cortés-Ortíz [5], José Francisco Escutia Domínguez [6], Daniel Agustín Bravata-Alcántara [7], Gustavo Acosta-Altamirano [8], Saúl David Razo-González [9], Manuel Alberto Castillo Mendez [10], Mónica Sierra-Martínez [4,*] and Juan Carlos Bravata-Alcántara [4,*]**

[1] Medical School, National Polytechnic Institute, Mexico City 11340, Mexico
[2] Research Division, Hospital Juárez de México, Mexico City 07760, Mexico
[3] Genetics Laboratory, Hospital Nacional Homeopático, Mexico City 06800, Mexico
[4] Unidad de Investigación en Salud, Hospital de Alta Especialidad de Ixtapaluca, Servicios de Salud del Instituto Mexicano del Seguro Social para el Bienestar (IMSS-BIENESTAR), Ixtapaluca 56530, Mexico
[5] Laboratorio de Genética y Diagnóstico Molecular, Hospital Juárez de México, Mexico City 07760, Mexico
[6] Facultad de Estudios Superiores Cuautitlán, Campo 1, Universidad Nacional Autónoma de México (UNAM), Mexico 54740, Mexico
[7] National School of Medicine and Homeopathy, National Polytechnic Institute, Mexico City, Mexico (IPN), Av. Guillermo Massieu Helguera 239, Ticomán, Mexico City 07320, Mexico
[8] Hospital General de México, Mexico City 06726, Mexico
[9] Facultad de Estudios Superiores Zaragoza, Campus 2, Universidad Nacional Autónoma de México (UNAM), Iztapalapa, Mexico City 09230, Mexico
[10] Universidad Autónoma de la Ciudad de México, Plantel Cuautepec, Colegio de Ciencias y Humanidades, Academia de Nutrición, La Palma, Mexico City 07160, Mexico
* Correspondence: monica.sierram@hraei.gob.mx (M.S.-M.); juan.bravata@hraei.gob.mx (J.C.B.-A.)
† These authors contributed equally to this work.

## Abstract

Human immunodeficiency virus type 1 exhibits extensive genetic diversity, which has important implications for transmission dynamics, disease progression, and the effectiveness of antiretroviral therapy. In Mexico, molecular surveillance has largely relied on partial *pol* gene sequencing, limiting the detection of recombination events and resistance mutations outside canonical regions. In this study, we performed near-full-length whole-genome sequencing of HIV-1 from 40 treatment-naïve adults receiving care at a tertiary-care hospital in Mexico to characterize drug-resistance mutations, viral genetic diversity, and recombinant forms. Viral RNA was extracted from plasma and sequenced on an Illumina platform, followed by bioinformatic processing and interpretation using the DeepChek pipeline for subtype classification, recombinant profiling, and identification of drug-resistance mutations. Drug-resistance mutations were identified in 6/40 (15.0%) participants. NNRTI-associated DRMs were identified in 2/40 patients (5.0%), whereas NRTI- and protease inhibitor-associated DRMs were each identified in 1/40 patient (2.5%). In addition, accessory INSTI-associated substitutions were detected in 2/40 patients (5.0%). No statistically significant differences were observed between patients with and without DRMs with respect to age, sex, or plasma viral load. Furthermore, DRMs were distributed across all recombinant categories, with no significant association between recombinant profile and DRM presence ($p = 0.97$). Non-B subtypes and recombinant forms predominated (82.5%), while subtype B accounted for 17.5% of cases. Extensive intergenic recombination was observed, with discordant subtype assignments across *gag*, *pol*, and *env* regions, consistent with mosaic viral genomes. Multiple circulating recombinant forms, including CRF03_AB, CRF07_BC, CRF28_BF, and CRF39_BF, were identified, alongside a predominance of BF-

 doi:10.20944/preprints202605.0280.v1



related recombinants. In addition, several unique recombinant forms with complex mosaic structures were detected, reflecting ongoing recombination and viral evolution. These findings highlight the high genetic complexity of HIV-1 in this population, characterized by a predominance of recombinant forms and extensive genomic mosaicism. The detection of DRMs across diverse genetic backgrounds supports the value of baseline resistance testing and suggests that broader genomic surveillance may improve HIV-1 molecular epidemiology monitoring in Mexico.

**Keywords:** HIV-1 molecular epidemiology; whole-genome sequencing; circulating recombinant forms; transmitted drug resistance; BF recombinants; genomic surveillance; Mexico

## 1. Introduction

Human immunodeficiency virus type 1 (HIV-1) exhibits marked genetic diversity, which has contributed to the complexity and persistence of the epidemic. This diversity results from error-prone reverse transcription, rapid viral replication, frequent recombination, and selective pressures from host immune responses and antiretroviral therapy [1]. As a result, HIV-1 has diversified into multiple genetic lineages that differ in their geographic distribution, biological properties, and clinical implications [2].

HIV-1 is divided into four groups: M (major), N (new), O (outlier), and P (putative), with group M causing the global pandemic. Group M is further classified into subtypes (A-D, F-H, J, and K), sub-subtypes, circulating recombinant forms (CRFs), and unique recombinant forms (URFs) [3,4]. Recombination occurs when a single cell is coinfected with different viral strains, producing mosaic genomes that can efficiently disseminate within populations. Over one hundred CRFs have been identified worldwide, demonstrating that HIV-1 continues to evolve [5].

HIV-1 shows significant genetic diversity in Latin America. While subtype B remains common in many countries, reports of non-B subtypes and recombinant forms are rising, especially in South America and the Caribbean [6,7].

Next-generation sequencing (NGS) has improved HIV-1 research by enabling more detailed characterization of the viral genome. In contrast, conventional approaches based on partial genomic regions, such as *pol* or *env*, have important limitations. Whole-genome sequencing provides a clearer picture of the virus, helps detect complex recombination events, and can identify minor mutations that may be clinically relevant [8]. Although studying the *pol* region remains important for tracking drug resistance, this approach can miss mutations outside the canonical regions, underestimate the complexity of recombinants, and overlook other genomic regions that affect viral fitness and the emergence of resistance [9]. It can also miss resistance-related changes across the whole genome and accessory regions, fail to detect complex recombinant structures, and overlook the full range of resistance pathways that develop under treatment pressure [10].

In Mexico, most HIV-1 molecular surveillance has used partial *pol* sequencing [11], and limited recent data combines whole-genome resistance analysis with detailed studies of genetic diversity. This represents a major knowledge gap, especially given the genetic diversity of the Mexican epidemic and the growing use of integrase strand transfer inhibitors [12], which require broader genomic monitoring. Tertiary-care hospitals are important in this context because they treat complex cases and generate valuable information about how resistance develops in real-life clinical settings. No previous Mexican study has combined near full-length whole-genome sequencing (WGS), recombinant burden analysis and resistance profiling in treatment-naïve individuals.

In this study, WGS was used to characterize HIV-1 strains from patients at a regional tertiary care hospital in Mexico, with a focus on identifying and interpreting drug-resistance mutations (DRMs) affecting the main classes of antiretroviral drugs. By combining whole-genome resistance profiling with analyses of subtype distribution, recombination, and sequence diversity, we offer an updated, clinically useful view of the regional HIV-1 genomic landscape. Our results are intended to improve resistance monitoring, guide personalized patient care, and support regional and national efforts to control HIV-1 drug resistance as ART becomes more widely used.

 doi:10.20944/preprints202605.0280.v1



## 2. Results

### *2.1. Clinical and Virological Characteristics*

This study included 40 treatment-naïve adults. Their mean age was 36.1 years, with a median of 34.5 years (IQR: 30.5-41.5). Most participants were male (32 out of 40, or 80%). The median plasma viral load was 780,000 copies/ml (IQR: 600,000-1,800,000), or 5.89 $\log_{10}$ copies/ml (IQR: 5.78–6.26), indicating high levels of viral replication in this cohort. Median plasma viral load did not differ significantly between patients with and without DRMs.

There were no statistically significant differences between patients with and without DRMs regarding age, sex, viral load, or HIV-1 subtype (Table 1). While all DRMs in this cohort were identified in male participants, this finding was not statistically significant (p=0.318) and likely reflects the overall male predominance (80%) of the study population rather than a sex-specific biological difference in transmission.

**Table 1. Comparison of demographic and virological characteristics between patients with and without HIV-1 drug-resistance mutations (DRMs).**

| Variable | DRM (+) (n=6) | DRM (-) (n=34) | p-value |
|---|---|---|---|
| Age, years, mean ± SD | 36.0 ± 4.4 | 36.1 ± 7.6 | 0.955 |
| Age, years, median (IQR) | 35.0 (33.3-38.3) | 34.5 (30.0-41.8) | 0.955 |
| Sex, male, n (%) | 6 (100%) | 26 (76.5%) | 0.318 |
| Plasma viral load, copies/mL, median (IQR) | 810,000 (705,000-1,275,000) | 780,000 (525,000-1,800,000) | 0.970 |
| Plasma viral load, $\log_{10}$ copies/mL, median (IQR) | 5.91 (5.85-6.10) | 5.89 (5.72-6.26) | 0.970 |

SD, standard deviation; IQR, Interquartile range. Continuous variables were compared using the Mann-Whitney U test, and categorical variables using Fisher´s exact test.

### *2.2. Drug-Resistance Mutations*

A total of 6/40 patients (15.0%) harbored at least one DRM. NNRTI-associated DRMs were identified in 2/40 patients (5.0%), whereas NRTI- and PI-associated DRMs were each detected in 1/40 patients (2.5%). In the integrase region, accessory INSTI-associated substitutions were identified in 2/40 patients (5.0%). Overall, the frequency of major resistance-associated findings was low in this cohort. No mutations associated with resistance to capsid inhibitors were detected (Table 2).

**Table 2. Antiretroviral resistance profile inferred from whole-genome HIV-1 sequencing.**

| Drug class | Patients with ≥1 DRM or resistance-associated substitution, n (%) | Drugs most frequently affected | Interpretation |
|---|---|---|---|
| NRTIs | 1 (2.5%) | Stavudine | Low prevalence; preserved class activity |
| NNRTIs | 2 (5.0%) | Doravirine, Efavirenz, Etravirine, Nevirapine, Rilpivirine | Low frequency; most frequently affected class within this cohort. |
| PIs | 1 (2.5%) | Atazanavir/r, Indinavir/r, Lopinavir/r, Nelfinavir, Saquinavir/r, Tipranavir/r | Low prevalence; largely preserved susceptibility. |
| INSTIs | 2 (5.0%) | Raltegravir, Elvitegravir | Accessory substitutions detected; interpret with caution |
| CIs | 0 (0%) | - | No resistance-associated mutations detected. |

NRTIs, nucleoside reverse transcriptase inhibitors; NNRTIs, non-nucleoside reverse transcriptase inhibitors; PIs, protease inhibitors; INSTIs, integrase strand transfer inhibitors; CIs, capsid inhibitors.

 doi:10.20944/preprints202605.0280.v1



*2.3. Mutation Profile*

Seven distinct resistance-associated mutations were identified in the reverse transcriptase (RT), integrase (IN), and protease (PR) regions. Four mutations were observed in RT. One patient carried two NNRTI-associated mutations (G190E and V179E); G190E is a major resistance mutation, whereas V179E is considered an accessory substitution. In two additional individuals, E138G and T215N/S were detected. E138G was classified as an NNRTI-associated mutation linked to reduced susceptibility to rilpivirine, whereas T215N/S was interpreted as a revertant mutation within the thymidine analogue resistance pathway. In the integrase region, S147G and G163K were identified in two different patients. Both are generally regarded as accessory or secondary INSTI-associated substitutions and, when detected individually, are not usually considered sufficient to confer clinically meaningful INSTI resistance. In the protease region, a single major mutation, I54T, associated with PI resistance, was detected. Although each mutation occurred at low frequency, 6 of 40 patients (15.0%) harbored at least one resistance-associated finding (Table 3), supporting the presence of baseline resistance-related variants in this cohort of treatment-naïve patients.

**Table 3. Profile of HIV-1 drug-resistance mutations identified by whole-genome sequencing.**

| Patient code | Region | Mutation | Patients, n (%) | Drug class | Comments |
|---|---|---|---|---|---|
| S021 | RT | G190E<br>V179E | 1 (2.5) | NNRTIs | Major resistance mutation<br>Accessory mutation |
| S003 | RT | E138G | 1 (2.5) | NNRTIs | Reduced susceptibility to rilpivirine |
| S009 | RT | T215N/S | 1 (2.5) | NRTIs | Revertant mutation (thymidine analog-associated) |
| S012 | IN | S147G | 1 (2.5) | INSTI | Accessory mutation; does not confer high-level DTG resistance alone |
| S039 | IN | G163K | 1 (2.5) | INSTI | Accessory mutation |
| S026 | PR | I54T | 1 (2.5) | PIs | Major resistance mutation |

RT, reverse transcriptase; IN, integrase; PR, protease; NRTIs, nucleoside reverse transcriptase inhibitors; NNRTIs, non-nucleoside reverse transcriptase inhibitors; PIs, protease inhibitors; INSTIs, integrase strand transfer inhibitors; DTG, dolutegravir.

*2.4. HIV-1 Subtype Diversity and Recombination Patterns by Genomic Segment*

To assess the extent of genetic diversity and recombination at high resolution, we analyzed eight genomic regions (P17, P24, P7, PROT, RT, INT, GP120, and GP41) using a heatmap visualization (Figure 1) and summarized the aggregate frequencies in Table 4.

Pure subtype B, defined as consistent B clustering across all segments, was identified in only 7 patients (17.5%) at the global level. However, the segment-specific analysis revealed a more complex picture: subtype B was most consistently preserved in RT. BF recombinants were the predominant lineage across all segments, with 42_BF/42_BF1 being the most frequent (11 detections, 27.5% of BF detections). The heatmap shows that BF lineages are distributed across most patients and segments, often in mosaic patterns where a patient has B in some segments and BF in others. For example, patients S002 and S003 show BF in P17 but B in RT and INT, while S011 show B in most segments but BF in INT and CRF03_AB in GP41. CRFs of non-BF origin were also detected, including CRF03_AB (5 detections, 12.5%), CRF07_BC (3 detections, 7.5%), and CRF28_BF, CRF39_BF, CRF47_BF1, CRF70_BF1, and CRF90_BF1 (collectively 17 detections). The heatmap reveals that these CRFs are not randomly distributed but tend to cluster in specific patients, such as S009 (CRF07_BC in PROT). A total of 18 URF detections and five additional non-B subtype fragments were identified across genomic regions, providing evidence of ongoing recombination and mosaic diversification. The heatmap shows that URF patterns are particularly common in the envelope genes, reflecting the high variability of these regions and ongoing recombination activity. Specific URFs such as 03_A6B (3 detections), 59_01B (3 detections), and D fragments (4 detections) were observed across multiple patients. Taken together, the heatmap and Table 4 demonstrate that pure subtype B represented a minority lineage in this cohort, while BF recombinants and URFs dominate, particularly in the

 doi:10.20944/preprints202605.0280.v1



envelope genes. The mosaic patterns visualized in Figure 1 underscore the limitations of partial pol sequencing and the added value of whole-genome approaches for accurate subtyping.

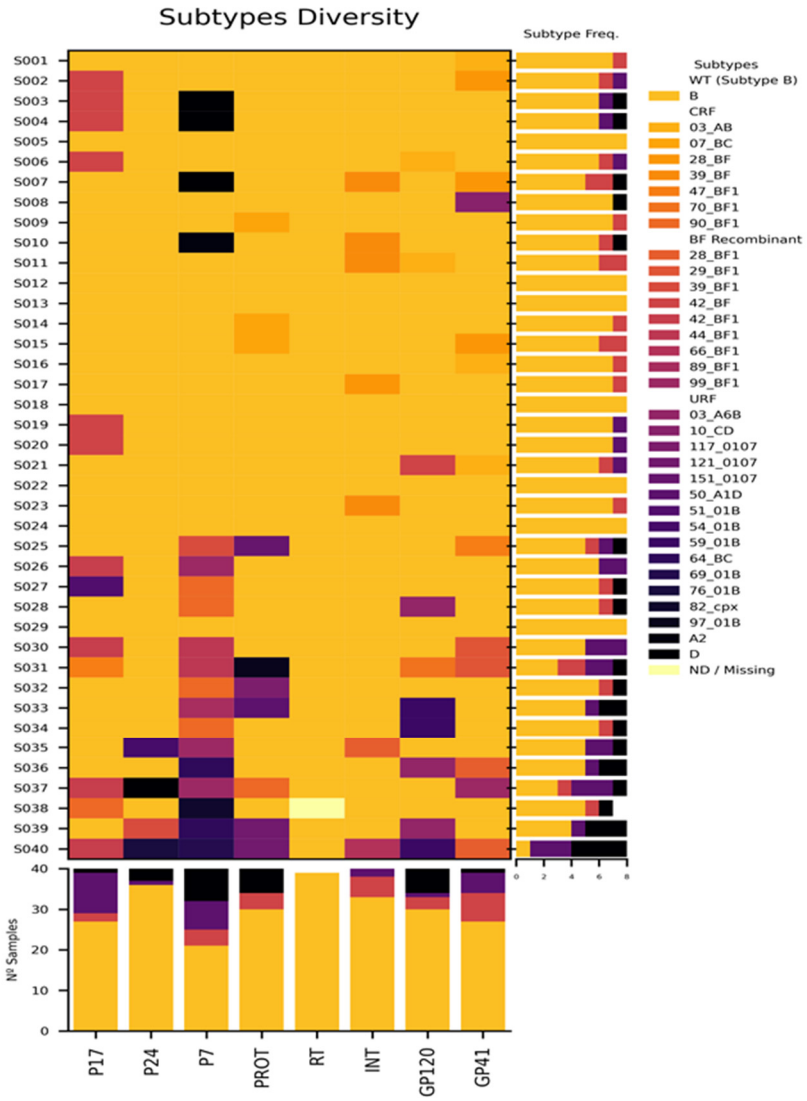


**Figure 1.** Heatmap of HIV-1 subtype diversity across eight genomic regions in 40 treatment-naïve patients from Mexico. Heatmap representing subtype assignments across gag (P17, P24, P7), pol (PROT, RT, INT), and env (GP120, GP41) regions. Color variation reflects subtype diversity and recombinant forms. Colors represent the inferred genetic lineage for each segment: pure subtype B, BF recombinants (including CRF28_BF, CRF39_BF, CRF42_BF, CRF47_BF, CRF90_BF, and other BF-derived lineages), CRF03_AB, CRF07_BC, URFs (unique recombinant forms, e.g., 03_A6B, 51_01B, 59_01B, 64_BC, 82_cpx, etc.). Gray for insufficient coverage or failed amplification. Right-side columns indicate the inferred genetic lineage or recombinant form detected in each patient across the genome, mosaic patterns (different colors across columns for the same patient) indicate complex recombination events. The bottom rows display the aggregate frequency of each subtype or recombinant form across the cohort, allow rapid visual comparison of which lineages are most common.

Multiple circulating recombinant forms (CRFs) were identified across different genomic regions, including CRF03_AB, CRF07_BC, CRF28_BF, and CRF39_BF. In addition, BF-related recombinants (including BF and BF1 variants) were the most frequently detected forms, representing the predominant recombinant lineage in the cohort.

Other recombinant patterns involving subtypes C, D, and A were also identified, including BC, CD, A1D, A6B profiles, as well as complex recombinant forms (cpx). These findings reflect the co-circulation of multiple HIV-1 lineages and the dynamic nature of viral evolution in this population.

 doi:10.20944/preprints202605.0280.v1



A substantial number of unique recombinant forms (URFs) were detected, characterized by complex mosaic genomes involving multiple subtype fragments. The presence of both CRFs and URFs confirms ongoing recombination processes and suggests active viral mixing within transmission networks.

A detailed classification of recombinant forms is presented in Table 4. CRFs such as CRF03_AB, CRF28_BF, and CRF39_BF were among the most frequently detected, while BF-derived recombinants constituted the dominant group. URFs, although less frequent, provided clear evidence of ongoing recombination and local viral evolution. Notably, frequencies reported for recombinant forms reflect their occurrence across genomic regions and may exceed the number of patients because multiple recombinant signals are detected within individual genomes.

**Table 4. Distribution of HIV-1 recombinant forms identified in the study cohort.**

| Category | Recombinant form | Frequency (n) | Percentage (%) | Interpretation |
|---|---|---|---|---|
| CRFs | CRF03_AB | 5 | 12.5 | Widely distributed intersubtype recombinant |
| | CRF07_BC | 3 | 7.5 | Common in Asia, emerging globally |
| | CRF28_BF | 4 | 10.0 | Latin American BF lineage |
| | CRF39_BF | 4 | 10.0 | Regional BF recombinant |
| | CRF47_BF1 | 2 | 5.0 | BF1-derived recombinant |
| | CRF70_BF1 | 1 | 2.5 | Rare BF1 recombinant |
| | CRF90_BF1 | 6 | 15.0 | Emerging BF1-related form |
| Subtotal CRFs | - | 25* | - | Multiple detections across genome regions |
| BF recombinants | 42BF/42_BF1 | 11 | 27.5 | Dominant regional recombinant |
| | 99_BF1 | 4 | 10.0 | BF1 lineage |
| | 89_BF1 | 1 | 2.5 | BF1 lineage |
| | 44_BF1 | 2 | 5.0 | BF1 lineage |
| | 39_BF1 | 2 | 5.0 | BF-1-related recombinant |
| | 28_BF1 | 3 | 7.5 | BF-related |
| | 29_BF1 | 2 | 5.0 | BF-related |
| | 66_BF1 | 1 | 2.5 | Minor BF variant |
| Subtotal BF recombinants | - | 26* | - | Predominant recombinant cluster |
| Unique Recombinant Forms (URFs) | 03_A6B | 3 | 7.5 | A6/B complex recombination |
| | 51_01B | 1 | 2.5 | CRF01-related mosaic |
| | 54_01B | 1 | 2.5 | CRF01-related mosaic |
| | 59_01B | 3 | 7.5 | CRF01-related mosaic |
| | 76_01B | 1 | 2.5 | CRF01-related mosaic |
| | A1D | 1 | 2.5 | A/D recombinant |
| | 01B-like | 2 | 5.0 | Hybrid recombinant |
| | 10_CD | 1 | 2.5 | C/D recombinant |
| | 151_0107 | 1 | 2.5 | Unique mosaic |
| | 117_0107 | 1 | 2.5 | Unique mosaic |
| | 121_0107 | 2 | 5.0 | Unique mosaic |
| | 50_A1D | 1 | 2.5 | A/D recombinant |
| | 82_cpx | 1 | 2.5 | Complex recombinant |
| | 64_BC | 2 | 5.0 | BC mosaic |
| | 69_01B | 1 | 2.5 | Hybrid recombinant |

 doi:10.20944/preprints202605.0280.v1



| | | | | |
|---|---|---|---|---|
| Subtotal URFs | - | 18* | - | Evidence of ongoing recombination |
| Non-B subtype fragments detected in mosaic genomes | A2 | 1 | 2.5 | Non-B subtype fragment |
| | D | 4 | 10.0 | Non-B subtype fragment |
| Subtotal fragments | - | 5* | - | Subtype fragments embedded within mosaic genomes |

*Frequencies represent segment-level detections and may exceed the number of patients due to the detection of multiple recombinant forms within individual genomes.

When stratified by recombinant category, DRMs were detected across all groups, including subtype B, CRFs, BF recombinants, and URFs (Table 5). No statistically significant association could be demonstrated between recombinant category and the presence of DRMs ($p = 0.97$), indicating that transmitted drug resistance is distributed across diverse genetic backgrounds within the cohort.

The combined analysis of subtyping and resistance profiles underscores the high complexity of HIV-1 molecular epidemiology in this setting and the importance of whole-genome approaches for accurate characterization of viral diversity.

Frequencies in Table 4 represent segment-level recombinant detections and therefore may exceed the number of patients, whereas Table 5 reflects mutually exclusive patient-level classification used for association analyses.

**Table 5. Association between patient-level HIV-1 recombinant categories and drug-resistance mutations (DRMs).**

| Category | Total (n=40) | DRM (+), n (%) | DRM (-), n (%) | p-value |
|---|---|---|---|---|
| Subtype B (no recombination) | 7 (17.5%) | 1 (14.3%) | 6 (85.7%) | - |
| CRFs (any detected) | 12 (30.0%) | 2 (16.7%) | 10 (83.3%) | - |
| BF recombinants (predominant) | 15 (37.5%) | 2 (13.3%) | 13 (86.7%) | - |
| URFs (unique recombinants) | 6 (15.0%) | 1 (16.7%) | 5 (83.3%) | - |
| TOTAL | 40 (100%) | 6 (15.0%) | 34 (85.0%) | 0.97 |

## 3. Discussion

The present study describes the baseline HIV-1 resistance-associated profile and genomic diversity of treatment-naïve individuals receiving care at a tertiary-care hospital in Mexico. Because the cohort was restricted to samples with high plasma viral load, the results should not be interpreted as a population-level estimate. Even so, the data provides a useful snapshot of two relevant features of the local HIV-1 landscape: resistance-associated mutations can be detected before treatment initiation, and recombinant viral genomes appear to be frequent in this group.

Resistance-associated mutations were found in 6 of 40 patients. This proportion should be interpreted with caution, mainly because of the small sample size and the sequencing-oriented inclusion criteria. However, the finding is still clinically relevant, as it shows that baseline viral populations may already carry variants with potential implications for antiretroviral susceptibility. The mutation profile was heterogeneous and involved the RT, IN, and PR regions, rather than being restricted to a single genomic region.

Among the detected mutations, G190E and I54T were the clearest findings with potential clinical relevance. G190E is a major NNRTI resistance mutation and has been associated with reduced susceptibility to first-generation NNRTIs such as nevirapine and efavirenz [13–15]. I54T, detected in the protease region, is also relevant because it is considered a major PI-associated mutation [16]. In contrast, V179E, S147G, and G163K should be interpreted more carefully. These substitutions are generally regarded as accessory or secondary mutations, and their individual clinical impact is likely to be limited unless they occur with other resistance-associated changes [15]. Accessory substitutions

 doi:10.20944/preprints202605.0280.v1



should therefore be interpreted primarily as markers of viral genetic background rather than direct predictors of treatment failure.

The RT mutation E138G was classified as an NNRTI-associated mutation linked to reduced rilpivirine susceptibility [15]. This is important because it avoids misclassifying the mutation as NRTI-related and places it in the correct therapeutic context. T215N/S was interpreted as a thymidine analogue mutation revertant [17]. Although revertant mutations do not necessarily confer the same level of resistance as classical thymidine analogue mutations, they may indicate prior evolutionary pathways related to NRTI resistance and can be useful for understanding viral history and transmission dynamics [18]. Taken together, these findings support the value of baseline genotypic resistance assessment, but they also show why interpretation should distinguish between major mutations and accessory substitutions.

No capsid inhibitor resistance-associated mutations were detected. This observation is consistent with the relatively recent introduction and limited clinical use of this drug class, which may limit selective pressure. However, given the small sample size, this finding should be interpreted as descriptive rather than evidence of absence in the broader untreated population.

The recombinant profile was one of the most relevant findings of the study. Pure subtype B accounted for only a minority of cases, whereas BF-related recombinants, CRFs, and URFs were frequent. This pattern supports the use of near-full-length or whole-genome sequencing for HIV-1 molecular epidemiology, particularly in settings where mosaic genomes are expected. Importantly, non-B subtype fragments (A2 and D) were distinguished from bona fide URFs, emphasizing that not all discordant segmental subtype assignments necessarily represent established recombinant forms. This distinction refines the interpretation of mosaic genomes and highlights the complexity captured by segment-level whole genome analyses. Partial *pol*-based approaches remain useful for resistance surveillance, but they can miss discordant subtype assignments across genomic regions and may underestimate the real complexity of recombinant forms [8,9,11].

For the association analysis, each patient was assigned to a single predominant recombinant category based on the whole-genome profile. Under this patient-level classification, no significant association was observed between recombinant category and DRM presence. In practical terms, resistance-associated mutations were found across different viral backgrounds rather than clustering within one specific recombinant group. This result should not be overinterpreted, since only six patients harbored DRMs and the study was not powered to detect lineage-specific differences.

These findings are relevant in the Mexican context. Pretreatment resistance to NNRTIs has been documented previously in Mexico, and national treatment strategies have progressively shifted toward second-generation INSTI-based regimens [19,20]. In that therapeutic landscape, baseline resistance testing and genomic surveillance remain useful. Their value is not that every accessory mutation should change treatment selection, but that they help place resistance findings and viral diversity within a broader epidemiological framework.

This study has limitations. It was conducted at a single center, including only 40 patients, and selected samples with high viral loads to increase the probability of successful near-full-length sequencing. The ≥ 20% variant-calling threshold may have underestimated minority resistant variants. As a result, the findings may not be generalizable to all treatment-naïve individuals in Mexico. In addition, several detected mutations were accessory substitutions with uncertain clinical relevance when present alone. Despite these limitations, the study shows that whole-genome sequencing can provide a more complete view of HIV-1 diversity and baseline resistance-associated variation in this setting.

## 4. Materials and Methods

*Study Design and Population*

A cross-sectional study was conducted at the Regional High Specialty Hospital of Ixtapaluca, Mexico, between 2024 and 2025. The study included 40 adult patients with confirmed HIV-1 infection who had not received antiretroviral therapy before sample collection.

 doi:10.20944/preprints202605.0280.v1



The study protocol was approved by the institutional ethics and research committees (approval number NR-027-2024), and all participants provided written informed consent in accordance with the Declaration of Helsinki.

*Sample Collection and Processing*

Peripheral blood samples were collected in EDTA-containing tubes and processed within 2 hours of collection. Plasma was separated by centrifugation at 1,300 × g for 10 minutes at room temperature and stored at −80 °C until analysis. Only samples with plasma viral load >100,000 copies/mL were included to ensure sufficient viral RNA for whole-genome sequencing.

*RNA Extraction and Viral Load Quantification*

Viral RNA was extracted from 140 µL of plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany), following the manufacturer's instructions.

Briefly, plasma samples were lysed with Buffer AVL containing carrier RNA and incubated at room temperature for 10 minutes. Ethanol (96–100%) was added, and the mixture was loaded onto QIAamp Mini spin columns. After washing with Buffer AW1 and AW2, RNA was eluted in 60 µL of Buffer AVE and stored at −80 °C until use.

An aliquot of the extracted RNA was used for HIV-1 viral load quantification using the Artus ® HI Virus-1 RG RT-PCR Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Real-time RT-PCR amplification was performed on a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA), and viral load values were expressed as copies/mL using the assay-provided calibration standards. RNA concentration was determined using the Qubit RNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). Only samples with a minimum RNA concentration of 50 ng/ µL and a viral load >100,000 copies/mL were considered suitable for downstream sequencing.

*Whole-Genome Sequencing*

Whole-genome HIV-1 sequencing was performed using the DeepChek® Whole Genome HIV-1 Genotyping Assay (ABL Diagnostics, Luxembourg), according to the manufacturer's protocol.

Briefly, extracted RNA was reverse-transcribed and amplified spanning the near-full-length HIV genome. Reverse transcription and amplification steps were carried out using a VeritiPro Thermal Cycler (Applied Biosystems, Foster City, CA, USA).

Sequencing libraries were prepared following the kit protocol and sequenced on the Illumina MiniSeq platform (Illumina, San Diego, CA, USA) using a Mid Output flow cell (300 cycles) with paired end reads (2 × 150 bp).

Sequencing data were considered of acceptable quality according to the criteria established by the assay and analysis pipeline, including adequate read quality, successful mapping of reads to the HIV-1 reference genome, and a minimum coverage threshold of ≥100 reads per position for regions included in the analysis.

*Bioinformatic Analysis*

Raw sequencing data were processed using the DeepChek® bioinformatics pipeline (ABL Diagnostics), which includes read quality control, trimming, genome assembly, and consensus sequence generation. Reads were aligned to the HIV-1 HXB2 reference genome, and variants were called using the algorithms implemented within the platform. Only variants detected at a frequency of ≥20% within the viral population were considered for downstream analysis. Drug-resistance mutations were interpreted using the resistance interpretation workflow integrated into the platform.

Subtype classification and recombinant analysis were performed using the automated subtyping tools available in DeepChek. To improve interpretability, subtype assignments were reviewed by genomic region, including P17, P24, P7, PROT, RT, INT, GP120, and GP41. Recombinant forms were

 doi:10.20944/preprints202605.0280.v1



categorized as CRFs when the genomic pattern matched previously described circulating recombinant forms recognized by the platform database. BF recombinants were defined as genomes showing consistent evidence of B/F mosaicism without meeting criteria for another named CRF. URFs were defined as mosaic genomes with discordant subtype assignments across genomic regions that did not match a recognized CRF pattern. For patient-level categorization, each genome was assigned to a single final category according to the predominant whole-genome recombinant profile used for downstream association analysis. Heatmap of HIV-1 subtype diversity was generated to visualize subtype assignments and recombination patterns across eight genomic segments (Gag: P17, P24, P7; Pol: Protease [PROT], Reverse Transcriptase [RT], Integrase [INT]; Env: GP120, GP41) for all 40 study participants. The visualization was implemented in Python (version 3.12; Python Software Foundation, Wilmington, DE, USA) using the matplotlib and seaborn libraries.

Samples with insufficient coverage in one or more regions were retained only for the genomic segments that passed quality thresholds. The minimum coverage threshold required for interpretation was ≥100 reads per analyzed position.

*Ethical Considerations*

All procedures were conducted in accordance with institutional guidelines and international ethical standards. Patient data were anonymized before analysis to ensure confidentiality.

*Statistical Analysis*

Descriptive statistics were used to summarize demographic, clinical, and virological data. Plasma viral load values were $\log_{10}$-transformed prior to statistical analysis.

Statistical analyses were performed using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). Statistical significance was set at $p < 0.05$. Continuous variables were compared using the Mann-Whitney U test due to non-normal distribution, while categorical variables were analyzed using Fisher´s exact test.

## 5. Conclusions

This study demonstrates a high burden of recombinant HIV-1 forms and a low-frequency but clinically relevant burden of baseline resistance-associated mutations in treatment-naïve individuals from a tertiary-care center in Mexico. These findings highlight the added value of whole-genome sequencing for improving genomic surveillance, refining recombinant characterization, and supporting resistance-informed HIV care in this setting.

**Author Contributions:** Conceptualization, J.C.B.A. and M.S.M.; methodology, E.G.M., J.A.P.M., E.M.C.M., I.A.C.O. and J.F.E.D.; software, E.G.M. and J.A.P.M.; validation, E.M.C.M., I.A.C.O., M.R.S. and M.A.C.M.; formal analysis, E.G.M., J.A.P.M., E.M.C.M., I.A.C.O. and M.A.C.M.; investigation, E.G.M., J.F.E.D., E.M.C.M., I.A.C.O., M.R.S., D.A.B.A., G.A.A., S.D.R.G. and M.A.C.M.; resources, J.C.B.A., M.S.M., G.A.A. and D.A.B.A.; data curation, E.G.M., J.A.P.M., J.F.E.D. and M.R.S.; writing—original draft preparation, E.M.C.M. and I.A.C.O.; writing—review and editing, E.M.C.M., I.A.C.O., M.R.S., D.A.B.A., M.A.C.M., M.S.M. and J.C.B.A.; visualization, E.G.M., J.A.P.M. and M.A.C.M.; supervision, J.C.B.A., M.S.M. and G.A.A.; project administration, J.C.B.A. and M.S.M.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Research and Ethics Committees of the Hospital de Alta Especialidad de Ixtapaluca, Servicios de Salud del Instituto Mexicano del Seguro Social para el Bienestar (IMSS-BIENESTAR), under protocol NR-027-2024, approved on 2 June 2025, for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available from the corresponding author upon reasonable request.

 doi:10.20944/preprints202605.0280.v1



**Acknowledgments:** The authors thank the clinical and laboratory staff of the Hospital de Alta Especialidad de Ixtapaluca for their support in patient recruitment, sample collection, and processing. The authors also thank Dra. Minerva Mata Rocha from the Human Genetics Research Unit, Pediatric Hospital, Siglo XXI National Medical Center, Mexican Social Security Institute, for her technical support in sequencing. Finally, the authors acknowledge the technical support provided during bioinformatic analysis and data interpretation.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| ART | Antiretroviral therapy |
| BF | Recombinant forms derived from subtypes B and F |
| CI | Capsid inhibitors |
| CRFs | Circulating recombinant forms |
| DRMs | Drug-resistance mutations |
| env | Envelope gene region of HIV-1 |
| gag | Group-specific antigen gene region of HIV-1 |
| HIV-1 | Human immunodeficiency virus type 1 |
| IQR | Interquartile range |
| IN | Integrase gene |
| INSTIs | Integrase strand transfer inhibitors |
| NRTIs | Nucleoside reverse transcriptase inhibitors |
| NNRTIs | Non-nucleoside reverse transcriptase inhibitors |
| NGS | Next-generation sequencing |
| PI | Protease inhibitors |
| PR | Protease gene |
| pol | Polymerase gene region of HIV-1 |
| RNA | Ribonucleic acid |
| RT | Reverse transcriptase gene |
| SD | Standard deviation |
| URFs | Unique recombinant forms |
| WGS | Whole-genome sequencing |

## References

1. Smyth, R.P.; Davenport, M.P.; Mak, J. The origin of genetic diversity in HIV-1. *Virus Res* **2012**, *169*, 415-429, doi:10.1016/j.virusres.2012.06.015.
2. Taylor, B.S.; Sobieszczyk, M.E.; McCutchan, F.E.; Hammer, S.M. The challenge of HIV-1 subtype diversity. *N Engl J Med* **2008**, *358*, 1590-1602, doi:10.1056/NEJMra0706737.
3. Osmanov, S.; Pattou, C.; Walker, N.; Schwardlander, B.; Esparza, J.; Isolation, W.-U.N.f.H.; Characterization. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr* **2002**, *29*, 184-190, doi:10.1097/00042560-200202010-00013.
4. Udeze, A.O.; Olaleye, D.O.; Odaibo, G.N. Phylogeny of partial gag, pol and env genes show predominance of HIV-1G and CRF02_AG with emerging recombinants in south-eastern Nigeria. *Heliyon* **2020**, *6*, e04310, doi:10.1016/j.heliyon.2020.e04310.
5. Rashid, A.; Li, K.; Feng, Y.; Ahmad, T.; Getaneh, Y.; Yu, Y.; Hu, X.; Abidi, S.H.; Shao, Y. HIV-1 genetic diversity a challenge for AIDS vaccine development: a retrospective bibliometric analysis. *Hum Vaccin Immunother* **2022**, *18*, 2014733, doi:10.1080/21645515.2021.2014733.
6. Crispim, M.A.E.; Reis, M.; Stefani, M.M.A. HIV-1 genetic diversity and drug resistance mutations in the northern Brazilian region. *Braz J Infect Dis* **2021**, *25*, 101596, doi:10.1016/j.bjid.2021.101596.
7. Williams, A.; Menon, S.; Crowe, M.; Agarwal, N.; Biccler, J.; Bbosa, N.; Ssemwanga, D.; Adungo, F.; Moecklinghoff, C.; Macartney, M.; et al. Geographic and Population Distributions of Human Immunodeficiency Virus (HIV)-1 and HIV-2 Circulating Subtypes: A Systematic Literature Review and Meta-analysis (2010-2021). *J Infect Dis* **2023**, *228*, 1583-1591, doi:10.1093/infdis/jiad327.

 doi:10.20944/preprints202605.0280.v1




8. Smoquina, F.; Berno, G.; Forbici, F.; Sberna, G.; Rozera, G.; Abbate, I.; Lazzari, E.; Amendola, A.; Mazzotta, V.; Gagliardini, R.; et al. Improved HIV-1 Subtyping Accuracy Using near Full-Length Sequencing: A Comparison of Common Tools. *Int J Mol Sci* **2025**, *26*, doi:10.3390/ijms262311666.
9. Faraci, G.; Park, S.Y.; Dube, M.P.; Lee, H.Y. Full-spectrum HIV drug resistance mutation detection by high-resolution complete pol gene sequencing. *J Clin Virol* **2023**, *164*, 105491, doi:10.1016/j.jcv.2023.105491.
10. Gaitan, N.C.; D'Antoni, M.L.; Acosta, R.K.; Gianella, S.; Little, S.J.; Chaillon, A. Brief Report: Comparative Analysis of Pre-existing HIV Drug Resistance Mutations in Proviral DNA Using Next-Generation Sequencing and Routine HIV RNA Genotyping. *J Acquir Immune Defic Syndr* **2023**, *93*, 213-218, doi:10.1097/QAI.0000000000003195.
11. Vazquez-Valls, E.; Escoto-Delgadillo, M.; Lopez-Marquez, F.C.; Castillero-Manzano, M.; Echegaray-Guerrero, E.; Bitzer-Quintero, O.K.; Kobayashi-Gutierrez, A.; Torres-Mendoza, B.M. Molecular epidemiology of HIV type 1 in Mexico: emergence of BG and BF intersubtype recombinants. *AIDS Res Hum Retroviruses* **2010**, *26*, 777-781, doi:10.1089/aid.2009.0195.
12. Orta-Resendiz, A.; Rodriguez-Diaz, R.A.; Angulo-Medina, L.A.; Hernandez-Flores, M.; Soto-Ramirez, L.E. HIV-1 acquired drug resistance to integrase inhibitors in a cohort of antiretroviral therapy multi-experienced Mexican patients failing to raltegravir: a cross-sectional study. *AIDS Res Ther* **2020**, *17*, 6, doi:10.1186/s12981-020-0262-y.
13. Melikian, G.L.; Rhee, S.Y.; Varghese, V.; Porter, D.; White, K.; Taylor, J.; Towner, W.; Troia, P.; Burack, J.; Dejesus, E.; et al. Non-nucleoside reverse transcriptase inhibitor (NNRTI) cross-resistance: implications for preclinical evaluation of novel NNRTIs and clinical genotypic resistance testing. *J Antimicrob Chemother* **2014**, *69*, 12-20, doi:10.1093/jac/dkt316.
14. Database, S.U.H.D.R. NNRTI Resistance Notes. **2026**.
15. Wensing, A.M.; Calvez, V.; Ceccherini-Silberstein, F.; Charpentier, C.; Gunthard, H.F.; Jacobsen, D.M.; Paredes, R.; Shafer, R.W.; Richman, D.D. 2025 update of the drug resistance mutations in HIV-1. *Top Antivir Med* **2025**, *33*, 457-473.
16. de Medeiros, R.M.; Junqueira, D.M.; Matte, M.C.; Barcellos, N.T.; Chies, J.A.; Matos Almeida, S.E. Co-circulation HIV-1 subtypes B, C, and CRF31_BC in a drug-naive population from Southernmost Brazil: analysis of primary resistance mutations. *J Med Virol* **2011**, *83*, 1682-1688, doi:10.1002/jmv.22188.
17. Cilento, M.E.; Kirby, K.A.; Sarafianos, S.G. Avoiding Drug Resistance in HIV Reverse Transcriptase. *Chemical Reviews* **2021**, *121*, 3271-3296, doi:10.1021/acs.chemrev.0c00967.
18. Pingen, M.; Nijhuis, M.; de Bruijn, J.A.; Boucher, C.A.; Wensing, A.M. Evolutionary pathways of transmitted drug-resistant HIV-1. *J Antimicrob Chemother* **2011**, *66*, 1467-1480, doi:10.1093/jac/dkr157.
19. Matias-Florentino, M.; Chaillon, A.; Avila-Rios, S.; Mehta, S.R.; Paz-Juarez, H.E.; Becerril-Rodriguez, M.A.; Del Arenal-Sanchez, S.J.; Pineirua-Menendez, A.; Ruiz, V.; Iracheta-Hernandez, P.; et al. Pretreatment HIV drug resistance spread within transmission clusters in Mexico City. *J Antimicrob Chemother* **2020**, *75*, 656-667, doi:10.1093/jac/dkz502.
20. Pineirua-Menendez, A.; Caro-Vega, Y.; Mosqueda-Gomez, J.L.; Rios, S.A.; Badial-Hernandez, F.; la Torre-Rosas, A.; Sierra-Madero, J. Experience of a Nationwide implementation of single-tablet second-generation integrase inhibitors-based treatment in Mexico. *IJID Reg* **2025**, *15*, 100645, doi:10.1016/j.ijregi.2025.100645.