

# Protease and Gag Diversity and Drug Resistance Mutations Among Treatment-Naive Mexican People Living with HIV

## **Samantha Climaco-Arvizu**

Unidad de Investigación Médica en Inmunología e Infectología, Hospital de Infectología "Dr Daniel Méndez Hernández", CMN "La Raza" Instituto Mexicano del Seguro Social (IMSS).

## **Víctor Flores-Lopez**

Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

## **Carolina González-Torres**

División de Desarrollo de la Investigación, Instituto Mexicano del Seguro Social, México City, México.

## **Francisco Javier Gaytán-Cervantes**

División de Desarrollo de la Investigación, Instituto Mexicano del Seguro Social, México City, México.

## **María Concepción Hernández-García**

Hospital de Infectología "Dr Daniel Méndez Hernández", Centro Médico Nacional (CMN) "La Raza", Instituto Mexicano del Seguro Social (IMSS)

## **Paola Berenice Zárate-Segura**

Laboratorio de Medicina Traslacional, Instituto Politécnico Nacional, México City, México

## **Montserrat Chávez-Torres**

Centro de Investigación en Enfermedades Infecciosas, Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, México City, México.

## **Emiliano Tesoro-Cruz**

Unidad de Investigación Médica en Inmunología e Infectología, Hospital de Infectología "Dr Daniel Méndez Hernández", CMN "La Raza" Instituto Mexicano del Seguro Social (IMSS).

## **Sandra María Pinto-Cardoso**

Centro de Investigación en Enfermedades Infecciosas, Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, México City, México.

## **Vilma Carolina Bekker-Méndez (✉ [bekkermendez@yahoo.com](mailto:bekkermendez@yahoo.com))**

Unidad de Investigación Médica en Inmunología e Infectología, Hospital de Infectología "Dr Daniel Méndez Hernández", CMN "La Raza" Instituto Mexicano del Seguro Social (IMSS).

---

## Research Article

**Keywords:** Human Immunodeficiency Virus, Antiretroviral therapy, HIV Drug Resistance Mutations, HIV genotyping, Gag, Protease, Next Generation Sequencing

**Posted Date:** March 11th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1416680/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Introduction:** In Mexico, HIV genotyping is performed in people living with HIV (PLWH) failing their first-line antiretroviral (ARV) regimen; it is not routinely done for all treatment-naive PLWH before ARV initiation. The first nationally representative survey published in 2016 reported that the prevalence of pretreatment drug mutations in treatment-naive Mexican PLWH was 15.5% to any antiretroviral drug and 10.6% to non-nucleoside reverse transcriptase inhibitors (NNRTIs) using conventional Sanger sequencing. Most reports in Mexico focus on HIV *pol* gene and nucleoside and non-nucleoside reverse transcriptase inhibitor (NRTI and NNRTI) drug resistance mutations (DRMs) prevalence, using Sanger sequencing, next-generation sequencing (NGS) or both. To our knowledge, NGS has not been used to detect pretreatment drug resistance mutations (DRMs) in the HIV protease (PR) gene and its substrate the Gag polyprotein.

**Methods:** Treatment-naive adult Mexican PLWH were recruited between 2016 and 2019. HIV Gag and protease sequences were obtained by NGS and DRMs were identified using the WHO surveillance drug resistance mutation (SDRM) list.

**Results:** One hundred PLWH attending a public national reference hospital were included. The median age was 28 years-old, and most were male 1 female (1%). The median HIV viral load was 4.99 [4.39-5.40] log copies/mL and median CD4 cell count was 150 [68.0-355.78] cells/mm<sup>3</sup>. As expected, most sequences clustered with HIV-1 subtype B (97.9%). Major PI resistance mutations were detected: 8 (8.3%) of 96 patients at a detection threshold of 1% and 3 (3.1%) at a detection threshold of 20%. A total of 1184 mutations in Gag were detected, of which 51 have been associated with resistance to PI, most of them were detected at a threshold of 20%. Follow-up clinical data was available for 79 PLWH at 6 months post-ARV initiation, seven PLWH failed their first ARV regimen; however no major PI mutations were identified in these individuals at baseline.

**Conclusions:** The frequency of DRM in the HIV protease was 7.3% at a detection threshold of 1% and 3.1% at a detection threshold of 20%. NGS-based HIV drug resistance genotyping provide improved detection of DRMs. Viral load was used to monitor ARV response and treatment failure was 8.9%.

## Introduction

In 2021, the National Centre for AIDS Prevention and Control (CENSIDA) reported that 195,860 people were living with Human Immunodeficiency Virus (PLWH) in Mexico. Of these, 113,788 (58.1%) were prescribed antiretroviral therapy (ART) and of those, around 20% did not achieve sustained viral suppression (1)(2). According to the Mexican National Guidelines, the recommended first-line ART regimen for treatment-naive adults consists of two nucleoside reverse-transcriptase inhibitors (NRTIs) and an integrase inhibitor (INSTI) (3). As an alternative, the use of a non-nucleoside reverse-transcriptase inhibitor (NNRTI) or boosted protease inhibitor (PI) is recommended (3). Despite the effectiveness of ART and considerable efforts to help control the HIV/AIDS epidemic by 2030 (95-95-95), ART failure due to drug resistance mutations (DRMs) is proving a challenge for ART provision and HIV care (4). In 2017, the

World Health Organization (WHO) published a report on HIV drug resistance addressing the alarming increase in the prevalence of DRMs in individuals initiating their first-line ART regimen, linking DRMs to treatment failure (5). According to the 2018 Recommendations of the International Antiviral Society (IAS)–USA Panel, HIV resistance testing is recommended for all individuals with HIV infection who are newly diagnosed, before they initiate ART and in PLWH with ART failure (6). Genotyping DNA-based assays are the most widely used for HIV DRMs detection. Next-generation sequencing (NGS) improves the detection of DRMs, by yielding substantially more throughput with higher sensitivity and reproducibility compared to conventional Sanger sequencing (7)(8). In Mexico, HIV genotyping is performed in PLWH failing their first-line regimens; it is not routinely performed for all PLWH who are treatment-naive and starting their first-line ART (9). Several studies have reported the prevalence of HIV DRMs in treatment-naive Mexican PLWH (10) (11) (12). According to a nationally representative survey, in Mexico the prevalence of any antiretroviral (ARV) resistance drug in treatment-naive PLWH is greater than 10%. Also, this report concluded that PLWH who initiated with NNRTI-based regimens achieved significantly lower levels of viral suppression compared to those who initiated with PI-based regimens (11). To our knowledge, NGS has not been used to detect DRMs in the protease (PR) gene and its substrate the Gag polyprotein in ART-naive Mexican PLWH. Drug resistance to PI arises by the accumulation of major mutations in the PR, mainly in the active site pocket (V32I, L33F, M46I, I47V, I54L/M, V82A/F and I84V) (13). These DRMs confer resistance to most PI, among them the most commonly used: atazanavir/r (ATV/r), lopinavir/r (LPV/r) and darunavir/r (DRV/R) (14). Besides the DRMs directly affecting the PR, DRMs to PI have also been studied in the natural substrate, the Gag polyprotein (15). The PR recognizes and cleaves Gag polyproteins into their fully active forms required for the maturation of HIV particles at the final stage of the viral cycle (16). Gag cleavage (CS) and non-cleavage site (NCS) mutations are directly associated with drug resistance to PI and NRTI resulting in virologic failure (17) (18). Gag NCS mutations R76K, Y79F and T81A in the Matrix (MA), I389T and I401V in the Nucleocapsid (NC) have been associated with virologic failure to LPV/r and Nelfinavir (NFV) respectively (19)(20). Also, Gag mutations can directly confer resistance to PI, in the absence of detectable DRMs in the PR (15). Therefore, many studies have proposed that a complete study of PI resistance, should include the full length of Gag in addition to protease (21)(22). To our knowledge, this is the first study that uses NGS to detect the presence of PR and Gag DRMs in treatment-naive Mexican PLWH.

## Methods

**Study design and population.** A cross-sectional study was conducted at the Immunology and Infectology Research Unit of Hospital de Infectología, Centro Médico Nacional La Raza (CMNR) del Instituto Mexicano del Seguro Social (IMSS), Mexico City, Mexico. One Hundred HIV-positive ART-naive adult PLWH (over 18 years) were recruited between 2016 and 2019. This study was performed in accordance with guidelines of the Helsinki Declaration and was approved by the Bioethical Committee of the IMSS National Research (CNIC-R-2017-785-106). Written informed consent was obtained from all participants before blood sample donation and data were processed using unique identifiers to ensure confidentiality. Baseline and follow-up clinical data, including CD4 cell count, plasma viral load (pVL), HIV diagnosis

date, ART initiation date, first-line ART regimen, changes in the initial treatment regimen, time between HIV diagnosis and ART initiation were obtained from the individuals' medical records. Demographic characteristics, including gender, age and risk factors for HIV infection were obtained by means of a validated questionnaire at the time of blood-sample donation.

**Blood collection, and RNA extraction.** Approximately 4 ml of peripheral blood were collected from all PLWH in ethylenediaminetetraacetic acid (EDTA) Vacutainer tubes (BD, San Jose, CA), and immediately processed to separate plasma by centrifugation at 2,500 rpm for 10 minutes at room temperature. Plasma samples were apportioned into 1.0 mL aliquots and stored at -70°C until use. For RNA extraction, a 1mL plasma aliquot was thawed per subject and centrifuged at 14,000 rpm at 4°C for 2 hours to concentrate HIV particles. RNA was extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

**Complementary DNA synthesis (cDNA), and polymerase chain reaction (PCR).** The coding region of HIV-1 Gag and Protease (HXB2 nucleotides 790 to 2550) was amplified using primers and conditions as published elsewhere (23)(24). Briefly, extracted RNA was subjected to a one-step RT-PCR followed by a nested amplification. Primer sequences, reagents and cycling conditions are provided in (Additional file 1). Amplicons were visualized by electrophoresis on a 1% agarose gel stained with ethidium bromide under ultraviolet light to confirm expected amplicon size (1998 pb).

**Library construction and sequencing of HIV-1 Gag and Protease.** Amplicons were purified using the MinElute PCR purification kit (Qiagen, Hilden, Germany) and quantitated using Qubit ds DNA Assay Kit (ThermoFisher, Waltham, MA, USA) according to the manufacturers' instructions. DNA libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA), using 1 ng of input DNA, and indexed using Nextera XT Index Kit (Illumina, San Diego, CA, USA), according to manufacturers' instructions. Libraries were checked for size and molarity using the High sensitivity D1000 ScreenTape (Agilent, Santa Clara, CA, USA) on a 4200 TapeStation (Agilent, Santa Clara, CA, USA). DNA libraries were pooled at 4nM and further diluted to 9pM. Libraries were sequenced on a MiSeq platform (Illumina, Santa Clara, CA, USA) using 2x250 cycles (MiSeq Reagent Kits v.2) with 5% PhiX as a control.

**Sequence data analysis.** In order to identify HIV Gag and Protease DRMs we took a three-fold approach, first we assembled full length sequences using automated Target Restricted Assembly Method (a TRAM, version 2.1.0) locus assembler, an iterative assembler that performs reference-guided local de novo assemblies using a variety of available methods (25), with Trinity (v2.9.0) (26) as the de novo assembler and the HIV-1 HXB2 sequence (GenBank: K03455) as the template. We also determined codon-specific variants as well as their relative frequencies using an alignment and variant-calling pipeline for Illumina deep sequencing of HIV-1, based on the probabilistic aligner profile Hidden Markov Model for HIV (HIVMMER, version 0.1.2) (27) with the HXB2 as reference sequence. Finally, we aligned the full length sequences to the HXB2 reference using Burrows-Wheeler Aligner (BWA, version 0.7.17), a software package for mapping low-divergent sequences against a large reference genome (28). Phylogenetic trees

were generated from sequences obtained from 96 treatment-naive Mexican PLWH. We identified two viral sequences originating from individual 37, located in different branches. HIV reference strains representing all HIV subtypes were used (Additional file 2). The nucleotide sequences were aligned using the multiple sequence alignment program Clustal-omega (version 1.2.1) (29) and edited manually using Seaview (version 5.0.4), a multiplatform program designed to facilitate multiple alignment and phylogenetic tree building from molecular sequence data through the use of a graphical user interface (30). A neighbor joining distance tree was constructed for each region separately: gag (1683 nucleotides) and protease (303 nucleotides). In order to provide statistical support to the constructed trees, they were subjected to 10,000 bootstrap replicas. Trees were custom-drawn using the ETE toolkit version 3 (31). To ensure accuracy, a detection threshold of 1% was used (based on the intrinsic error rate of the system (8)). To calculate sequencing depth we used Sequence Alignment/Map (SAM tools, version 1.10) (32). We achieved high level depth, with a median sequencing of 7616.84 reads (interquartile range (IQR), 7543.73-7793.41) for Gag and Protease respectively (Additional file 3).

**Drug resistance mutations (DRMs).** HIV Drug Resistance Database (HIVDB; version 8.8) and the International AIDS Society 2019 drug resistance mutations in HIV-1 list were used (33)(34). Gag sequences were analyzed for the presence of specific mutations previously reported to be associated with resistance to PI and treatment failure; these were: E12K, V35I, E40K, G62R, L75R, R76K, Y79F, T81A K112E, G123E, M200I, H219Q/P, Q369H, V370A/M, I389T, V390A/D, R409K, E468K, Q474L, P497L; as well as mutations in Gag CS: V128I/T/A, Y132F, A360V, V362I, L363M/F/C/N/Y, S368C/N, S373P/Q, A374P/S, T375N/S, I376V, G381S, E428G, Q430R, A431V, K436E/R I437T/V, L449F/P/V, S451T/G/R, P452S/K and P453A/L/T (20)(35). Different detection thresholds (1%, 2%, 5%, 10% and, 20%) were used to identify the prevalence of low abundance (or minority) mutations as these could be missed using standard clinically relevant threshold (20%, Sanger sequencing).

## Results

**Study cohort.** Demographic and clinical characteristics of our cohort are shown in Table 1. Hundred treatment-naive PWLH were enrolled. Their baseline median pVL was 4.99 [IQR 4.39–5.40] log RNA copies/mL and their baseline median CD4 cell count was 150 [IQR 68.0-355.78] cells/mm<sup>3</sup>. HIV acquisition risk factor was predominantly men who have sex with men (74.0%) (Table 1). HIV sequences were assigned predominantly to HIV-1 subtype B (Fig. 1).

Table 1  
Demographic and clinical characteristics of treatment-naive Mexican PLWH

<b>Variables</b>	
<b>Demographics</b>	
Gender, n (%)	
Male	99 (99.0%)
Female	1 (1.0%)
Age (years), median (min-max)	28 (18–66)
Marital status, n (%)	
Single	77 (77.0%)
Married	4 (4.0%)
Divorced	2 (2.0%)
Free union	16 (16.0%)
Widowed	1 (1.0%)
Literacy, n (%) <sup>1</sup>	
Primary school	3 (3.2%)
Junior High school	7 (7.4%)
High school	26 (27.4%)
Technical qualification	7 (7.4%)
Degree	50 (52.6%)
Postgraduate	2 (2.1%)

<sup>1</sup>Data on literacy was missing for 5 individuals

<sup>2</sup>Data was missing for 2 individuals

<sup>3</sup>Data on CD4 cell count (%) was missing for 27 individuals

<sup>4</sup>Data was missing for 1 individual

#### Abbreviations

ART = antiretroviral therapy, TDF = Tenofovir, FTC = Emtricitabine, EFV = Efavirenz, ATV/r = Atazanavir/ritonavir, ABC = Abacavir, 3TC = Lamivudine, LPV/r = Lopinavir/ritonavir, DTG = Dolutegravir, n = number, %= percentage, IQR = interquartile range, HIV = Human Immunodeficiency Virus, Dx = diagnosis

<b>Variables</b>	
Employment, n (%)	
Employed	86 (86.0%)
Unemployed	1 (1.0%)
Student	13 (13.0%)
<b>HIV</b>	
CD4 count (cell/mm <sup>3</sup> ), median [IQR] <sup>2</sup>	150 (68.0-355.78)
CD4 cell count (%), mean (SD) <sup>3</sup>	11.7 (7.8)
Plasma viral load (copies/mL) <sup>2</sup> , median [IQR]	99575 [24406–250117]
Plasma viral load (log copies/mL) <sup>2</sup> , median [IQR]	4.99 [4.39–5.40]
HIV acquisition risk factor, n (%) <sup>4</sup>	
Men who have sex with men	74 (74.7%)
Heterosexual	11 (11.1%)
Bisexual	14 (14.1%)
HIV-1 subtype, n (%)	
B	94 (97.9%)
Non-B	2 (2.1%)
<b>ART</b>	
Time Dx to ART initiation (days), median [IQR] <sup>2</sup>	38.5 [20.8–60.5]

<sup>1</sup>Data on literacy was missing for 5 individuals

<sup>2</sup>Data was missing for 2 individuals

<sup>3</sup>Data on CD4 cell count (%) was missing for 27 individuals

<sup>4</sup>Data was missing for 1 individual

#### Abbreviations

ART = antiretroviral therapy, TDF = Tenofovir, FTC = Emtricitabine, EFV = Efavirenz, ATV/r = Atazanavir/ritonavir, ABC = Abacavir, 3TC = Lamivudine, LPV/r = Lopinavir/ritonavir, DTG = Dolutegravir, n = number, %= percentage, IQR = interquartile range, HIV = Human Immunodeficiency Virus, Dx = diagnosis

Variables	
First-line regimen, n (%) <sup>2</sup>	
TDF-FTC-EFV	49 (50.0%)
TDF-FTC-ATVr	18 (18.4%)
ABC-3TC-ATVr	15 (15.3%)
ABC-3TC-EFV	12 (12.2%)
ABC-TDF-ATVr	1 (1.0%)
TDF-FTC-LPVr	1 (1.0%)
TDF-3TC-EFV	1 (1.0%)
ABC-3TC-DTG	1 (1.0%)
<sup>1</sup> Data on literacy was missing for 5 individuals	
<sup>2</sup> Data was missing for 2 individuals	
<sup>3</sup> Data on CD4 cell count (%) was missing for 27 individuals	
<sup>4</sup> Data was missing for 1 individual	
Abbreviations	
ART = antiretroviral therapy, TDF = Tenofovir, FTC = Emtricitabine, EFV = Efavirenz, ATV/r = Atazanavir/ritonavir, ABC = Abacavir, 3TC = Lamivudine, LPV/r = Lopinavir/ritonavir, DTG = Dolutegravir, n = number, %= percentage, IQR = interquartile range, HIV = Human Immunodeficiency Virus, Dx = diagnosis	

**Residue variation of HIV Gag and protease.** Full-length Gag and Protease sequences analysis revealed high variation in both genes (Table 2). Mutations in both Gag CS and NCS were identified. The distribution of all the identified mutations in Gag and protease are shown in (Additional file 4). We observe 1186 amino acid (aa) changes at 346 positions at full-length Gag, and 175 aa changes in 74 positions throughout protease. The capsid protein (51.9%) contained the lowest number of polymorphic positions followed by nucleocapsid (78.2%), matrix (84.8%), and p6 (92.3%).

Table 2

HIV Gag and Protease identified mutations in patients naive to antiretroviral treatment (n = 96).

Protein	Length (aa)	Residues with at least one mutation n (%)	Total mutations observed
Gag	500	346 (69.2)	1186
Matrix	132	112 (84.8)	433
Capsid	231	120 (51.9)	221
p2	14	11 (78.6)	72
Nucleocapsid	55	43 (78.2)	153
p1	16	12 (75.0)	35
p6	52	48 (92.3)	272
Protease	99	74 (74.7)	175
Abbreviations			
aa= amino acids, n= number, %= percentage			

**Drug resistance mutations (DRMs).** Major and minor PI drug resistance mutations were detected (Figure 2). Presence of major PI resistance mutations in 8 (8.3%) of 96 patients at a detection threshold of 1% and 3 (3.1%) at a detection threshold of 20%. The most common major resistance mutation was M46I at 3.6% (3/96) frequency at 1% threshold and I54V with 2.4% (2/96) frequency at 20% threshold (Figure 2a). Several minor PI resistance mutations were identified, M36I was the most frequent with 50.0% (48/96) and 19.8% (19/96) frequency at 1% and 20% detection threshold, respectively. Inferred levels of resistance to PI were determined, intermediate resistance was found in 2.08% (2/96) for both atazanavir and lopinavir (Figure 2b).

**Gag mutations associated with resistance to PI.** Mutations associated with resistance to PI were found in Gag CS and NCS. We identified a total of 1184 mutations in Gag, of which 51 have been associated with resistance to PI, most of them were detected at the 20% threshold (Figure 3). The mutation with the highest frequency at 1% detection threshold was R76K with 65.1% (54/83) in the MA protein, H219Q 44.6% (37/83) in CA protein, I389T 24.1% (20/83) in NC protein and P453T 8.4% (7/83) in p6 protein.

**Follow-up data.** The median time to ART initiation was 38.5 days [IQR 20.8-60.5]. Data on the first-line ART regimen was available for 98 PLWH: 62 (63.3%) initiated with 2 NRTI + 1 NNRTI, 34 (34.7%) with 2 NRTI + 1 PI and 1 (1.0%) with 2 NRTI + 1 INSTI (Table 1). A total of 12 (12.2%) PLWH switch to second line ART, mainly due to side effects (Additional file 5). Follow-up HIV clinical variables (pVL and CD4 cell count) were available for 79 PLWH 6 months after ART initiation (Additional file 6). Sixty-three (80.8%)

PLWH achieved complete viral suppression (defined as pVL<50 copies/mL) while 72 (92.3%) achieved viral suppression (defined by pVL below 200 copies/mL after 6 months on ART). The median CD4 cell count was 437 cells/mm<sup>3</sup> [IQR 230-614], with 31 (39.7%) PLWH achieving CD4 cell counts above 500 cells/mm<sup>3</sup>. Seven PLWH failed to suppress plasma viremia below 200 copies/mL (Additional file 7) and might suspicion of virologic failure, which would need to be confirmed. No major PI mutations were identified in these individuals at baseline. Treatment outcomes according to the presence of baseline Gag and protease drug resistance-associated mutations in PLWH are visualized in table 3.

Table 3  
Presence of drug resistance associated mutations on patient treatment outcomes

Regimen ART follow up n = 79	Parameters	*Without Baseline protease DRMs n (%)	*With Baseline protease DRMs n (%)	With < 5 Baseline Gag drug resistance-associated mutations n (%)	With ≥ 5 Baseline Gag drug resistance-associated mutations n (%)
2INRT + 1IP 28 (35.4)	Viral load				
	< 50	4 (14.3)	17 (60.7)	4 (14.3)	17 (60.7)
	50–200	-	3 (10.7)	1 (3.6)	2 (7.1)
	> 200	1 (3.6)	2 (7.1)	-	3 (7.9)
2INRT + 1INNRT 50 (63.3)	Viral load				
	< 50	7 (14)	32 (64)	15 (30)	24 (48)
	50–200	-	5 (10)	-	5 (10)
	> 200	-	4 (8)	1 (2)	3 (6)
*Minor and major protease resistance mutations					
Abbreviations					
ART = antiretroviral therapy, 2INRT + IP = two nucleoside reverse transcriptase inhibitors plus one protease inhibitor, 2INRT + 1INNRT = two nucleoside reverse transcriptase inhibitors plus one non- two nucleoside reverse transcriptase inhibitor, DRMs = drug resistance mutations, n = number, %= percentage.					

## Discussion

HIV genotyping is a critical tool in HIV care and ART management, as it provides clinicians with relevant information to guide the choice for the most suited ART regimen, this can lead to ART success, together with treatment adherence. Previous studies in Mexico have analyzed DRMs in PLWH, but mainly in the HIV *pol* gene. In the present study we analyzed the presence of DRMs by NGS in the HIV protease and HIV Gag in ART-naive PLWH from the Hospital de Infectología, Instituto Mexicano del Seguro Social (IMSS), Mexico City. The Hospital de Infectología is a public national reference center that provides care to nearly

1,000 PLWH a month. Therefore, our cohort reflects a key population of Mexican PLWH, even though the study is not nationally representative. Demographic and clinical characteristics from our cohort were similar to those previously reported in Mexican PLWH including late stages of HIV disease and CD4 cell counts below 200 cells/ml at HIV diagnosis (36). The prevalence of MSM was higher in the study compared to national reports (74.0% vs. 53.0%; 2016 national statistics).

Most sequences clustered with HIV-1 subtype B, as expected (10)(11), with 2 PLWH identified with HIV-1 non-B subtypes. In Mexico, the presence of inter-subtype recombinants (IRs) such as BG, BF, CRF12-BF and the CRF06-cpx have been reported (37)(10). This can be explained by immigration (both temporary and permanent). In this context, Mexico is the second country of destination for migrants from South America (38) and also, Mexicans traveling abroad has increased (39). The application of NGS techniques for HIV genotyping has been increasing; however, it is advisable to validate with Sanger sequencing (40). Unfortunately, one of the limitations of this study was not performing Sanger sequencing, so we cannot validate our results (20% detection threshold). NGS techniques improve the detection of DRMs (8); however, it has been suggested that low abundance variants detected at 5% threshold have little clinical relevance in treatment outcome (41). In this study, we decided to use different detection thresholds from 1% to 20% with the purpose detecting a wide range of variants. Comparing to two recent studies (42)(43), we found higher rates of variation in the protease and in each Gag protein. Previous analysis of protease DRMs also reported a 3% of PI resistance with a threshold of 20% and 9.5% of resistance with a lower threshold (2%), this value was higher compared to our own results (4.2% of PI resistance) (11). This difference may be due to NGS and bioinformatics pipelines. The most common major resistance mutation was M46I with 3.6% (threshold 1%) and 2.1% (threshold 2%) and I54V with 2.1% (20% threshold), frequencies that were higher compared to 1.5% for M46I (2% threshold) and 0.4% (20% threshold) detected in a previous study (11). A study of amino acid variations of 777 Mexican protease sequences (12) showed similar presence of major and minor resistance mutations to protease inhibitors except for V111L, K43T, M46L, G48R, N88D, that were not detected in our study, however we were able to identify other mutations like M46I, L10F, K20T, M36I and I85V. We detect intermediate level of resistance for ATV and LPV, compared to a study reporting low level of resistance to PI (11). This result is relevant since these PI are still being used in the Hospital de Infectología del IMSS. Several Gag mutations in CS and NCS have been associated with resistance to PI (35). We identified 51 Gag mutations in CS and NCS, with R76K being the most predominant at 65.1%. Resistance to PI and viral fitness, are often associated with a signature of mutations in Gag with the presence or absence of mutations in the protease (44). A previous study found the R76K mutation was present in 57% in seven cases of virologic failure to boosted PI therapy (43) These mutations: R76K, Y79F and T81A enhance protease accessibility to Gag CS, viral replication and reduced susceptibility to PI (45). Therefore, the high frequency of R76K mutation found in our study cohort is subject of concern. The CA Gag protein NCS mutation H219Q was found at a frequency of 44.6%. It has been shown that the H219Q mutation increases viral replication fitness carrying T242N or T242S mutations (46). In a previous study multiple Gag NCS mutations including L75R, H219, V390A/D, R409K, and E468K were indispensable for HIV-1 replication in the presence of PI (47). Our study showed that the p6 protein had the highest percentage of variation compared to the other

Gag proteins and the mutation with the highest frequency was P453L/T. Mutations in P453 residue were found to enhance resistance to PI in the presence of PR mutations I50V and I84V (48)(49). Six months after ART initiation, seven of the 79 PLWH for whom we had follow-up data, failed their first ART regimen. Although no major PI mutations were observed, several minor mutations in the protease and Gag associated resistance mutations were identified on these individuals. In this study, we identified several mutations in Gag proteins that to our knowledge have not been previously reported, such as H89Q, K418Q, I479M with a 17%, 18% and 21% frequency respectively. It would be interesting to perform phenotypic assays to determine if these mutations whether they have a role in PI drug resistance.

Our study has several limitations, first the limited number of PWLH included in this study, and the lack of follow-up clinical data, which precludes analyzing the relationship between the presence of PR and Gag mutations and response to ART. We could not validate our results as Sanger sequencing wasn't performed.

## Conclusions

In conclusion, the present study reports the detection by NGS of drug resistant mutations in the Gag polyprotein and protease gene associated with protease inhibitor resistance in treatment-naïve Mexican PLWH from The Hospital de Infectología del IMSS, a public national reference center that provides care to nearly 1,000 PLWH a month. The frequency of DRM in the HIV protease was 7.3% at a detection threshold of 1% and 3.1% at a detection threshold of 20%. Most of the mutations associated with resistance in Gag were detected at a threshold of 20%. Seven PLWH failed their ART regimen, and no major PI DRMs were found at baseline. HIV genotyping is a critical tool in HIV care and ART management and efforts should be made to genotype all ART-naïve PLWH prior to ART initiation.

## Abbreviations

(3TC) Lamivudine, (a TRAM) Target Restricted Assembly Method, (ABC) Abacavir, (ART) antiretroviral therapy, (ARV) antiretroviral, (ATV/r) atazanavir/r, (ATV/r) Atazanavir/ritonavir, (BWA) Burrows-Wheeler Aligner, (cDNA) Complementary DNA synthesis, (CENSIDA) National Centre for AIDS Prevention and Control, (CMNR) Centro Médico Nacional La Raza, (CS) cleavage site, (DRMs) drug resistance mutations, (DRV/R) darunavir/r, (DTG) Dolutegravir, (Dx) diagnosis, (EDTA) ethylenediaminetetraacetic acid, (EFV) Efavirenz, (FIS) Fondo de Investigación en Salud, (FTC) Emtricitabine, (HIV) Human Immunodeficiency Virus, (HIVDB) HIV Drug Resistance Database, (HIVMMER) Hidden Markov Model for HIV, (IMSS) Instituto Mexicano del Seguro Social, (INSTI) integrase inhibitor, (IQR) interquartile range, (LPV/r) Lopinavir/ritonavir, (MA) Matrix, (n) number, (NC) Nucleocapsid, (NCS) non-cleavage site, (NFV) Nelfinavir, (NGS) next-generation sequencing, (NNRTI) non-nucleoside reverse-transcriptase inhibitor, (NRTIs) nucleoside reverse-transcriptase inhibitors, (PCR) polymerase chain reaction, (PI) protease inhibitor, (PLWH) people living with Human Immunodeficiency Virus, (PR) protease, (pVL) plasma viral load, (SAM) Sequence Alignment/Map, (SDRM) surveillance drug resistance mutation, (TDF) Tenofovir, (WHO) World Health Organization.

# Declarations

## Ethics approval and consent to participate

This study was performed in accordance with guidelines of the Helsinki Declaration and was approved by the Bioethical Committee of the IMSS National Research (CNIC-R-2017-785-106). All participants provided written informed consent to participate in the study.

## Consent for publication

Not applicable.

## Availability of data and materials

The datasets are available from the corresponding author on reasonable request.

## Competing interests

The authors declare no competing financial interests.

## Funding

This work was funded by grants from the Fondo de Investigación en Salud (FIS) Instituto Mexicano del Seguro Social (IMSS), Number: FIS/IMSS/PROT/PRI0/18/066. Instituto Politécnico Nacional Number SIP 20221517.

## Authors' contributions

SCA obtained demographic data from participating subjects, performed experiments, contributed to data analyses and interpretation, wrote manuscript. VFL analyzed sequencing data, performed phylogenetic analysis. CGT, FJGC and MCT performed gene sequencing. MCHG carried subject recruitment, obtained written consent, and supervised specimen collection. PBZS and ETC revised the manuscript. VCBM and SMPC designed and supervised the overall study and edited manuscript.

## Acknowledgements

We thank all personal from the Immunology and Infectology Research Unit of Hospital de Infectología, CMNR del IMSS. We thank Mónica Patricia Ortiz Maganda and Francisco Xavier Guerra for locating samples and generating the dataset for the samples used in this study.

# References

1. Dirección General de Epidemiología de Enfermedades Transmisibles. Secretaria de Salud. Informe Histórico De VIH 2Do Trimestre 2021. Sistema de Vigilancia Epidemiológica de VIH.

- 2021.[https://www.gob.mx/cms/uploads/attachment/file/667817/VIH-Sida\\_2doTrim\\_2021.pdf](https://www.gob.mx/cms/uploads/attachment/file/667817/VIH-Sida_2doTrim_2021.pdf). Accessed 13 Jul 2021.
2. CENSIDA, Secretaría de Salud. Boletín de atención integral de personas que viven con VIH. 2021.[https://www.gob.mx/cms/uploads/attachment/file/670963/Bol\\_DAI\\_Vol7\\_N3\\_SEP2021.pdf](https://www.gob.mx/cms/uploads/attachment/file/670963/Bol_DAI_Vol7_N3_SEP2021.pdf). Accessed 13 Jul 2021.
  3. Centro Nacional para la Prevención y Control del VIH y el sida. Guía de manejo antirretroviral de las personas con VIH. 2021. p. 1–279.  
[https://www.gob.mx/cms/uploads/attachment/file/670762/Guia\\_ARV\\_2021.pdf](https://www.gob.mx/cms/uploads/attachment/file/670762/Guia_ARV_2021.pdf). Accessed 13 Jul 2021.
  4. World Health Organization. HIV Drug Resistance Report 2019. WHO. 2019. p. 68.  
<http://www.who.int/hiv/pub/drugresistance/hivdr-report-2019/en/%0Ahttp://scholar.google.com/scholar/Who+hiv+drug+resistance+report+2012>. Accessed 16 Jul 2021.
  5. Haile-Selassie H. Hiv drug resistance report 2017. World Health Organization. 2017. 68 p.
  6. Günthard HF, Calvez V, Paredes R, Pillay D, Shafer RW, Wensing AM, et al. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel. *Clin Infect Dis*. 2019;68(2):177–87.
  7. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. Deep sequencing: Becoming a critical tool in clinical virology. *J Clin Virol*. 2014;61(1):9–19.
  8. Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, Ruiz-Mateos E, et al. Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. *Antimicrob Agents Chemother*. 2014;58(4):2167–85.
  9. Centro Nacional para la Prevención y Control del VIH y el sida. Guía de manejo antirretroviral de las personas con VIH. 2019. p. 14.  
[https://www.gob.mx/cms/uploads/attachment/file/470115/Fragmento\\_Gu\\_a\\_de\\_Manejo\\_ARV.pdf](https://www.gob.mx/cms/uploads/attachment/file/470115/Fragmento_Gu_a_de_Manejo_ARV.pdf). Accessed 16 Jul 2021.
  10. Avila-Ríos S, García-Morales C, Garrido-Rodríguez D, Ormsby CE, Hernández-Juan R, Andrade-Villanueva J, et al. National prevalence and trends of hiv transmitted drug resistance in mexico. *PLoS One*. 2011;6(11).
  11. Ávila-Ríos S, García-Morales C, Matías-Florentino M, Romero-Mora KA, Tapia-Trejo D, Quiroz-Morales VS, et al. Pretreatment HIV-drug resistance in Mexico and its impact on the effectiveness of first-line antiretroviral therapy: a nationally representative 2015 WHO survey. *Lancet HIV*. 2016;3(12):e579–91.
  12. Hernandez-Sanchez PG, Guerra-Palomares SE, Arguello JR, Noyola DE, Garcia-Sepulveda CA. Mexican HIV-1 Protease Sequence Diversity. *AIDS Res Hum Retroviruses*. 2020;36(2):161–6.
  13. Wensing AMJ, van Maarseveen NM, Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res*. 2010;85(1):59–74.

14. Stanford University. HIV DRUG RESISTANCE DATABASE. 2016. <https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/>. Accessed 7 Oct 2019.
15. Nijhuis M, Van Maarseveen NM, Lastere S, Schipper P, Coakley E, Glass B, et al. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med*. 2007;4(1):0152–63.
16. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology*. 2005;2:12–6.
17. Clavel F, Mammano F. Role of gag in HIV resistance to protease inhibitors. *Viruses*. 2010;2(7):1411–26.
18. Soriano V, de Mendoza C. Genetic mechanisms of resistance to NRTI and NNRTI. *HIV Clin Trials*. 2002;3(3):237–48.
19. Li G, De Clercq E. HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. *Microbiol Mol Biol Rev*. 2016;80(3):679–731.
20. Fun A, Wensing AMJ, Verheyen J, Nijhuis M. Human Immunodeficiency Virus gag and protease: Partners in resistance. *Retrovirology*. 2012;9(3):1–14.
21. Sutherland KA, Goodall RL, McCormick A, Kapaata A, Lyagoba F, Kaleebu P, et al. Gag-Protease Sequence Evolution Following Protease Inhibitor Monotherapy Treatment Failure in HIV-1 Viruses Circulating in East Africa. *AIDS Res Hum Retroviruses*. 2015;31(10):1032–7.
22. Giandhari J. Genetic Changes in HIV-1 Gag and Pol genes associated with Protease Inhibitor-Based Therapy Failure in Paediatric Patients PhD Virology in the. 2015;
23. Van Laethem K, Schrooten Y, Dedecker S, Van Heeswijck L, Deforche K, Van Wijngaerden E, et al. A genotypic assay for the amplification and sequencing of gag and protease from diverse human immunodeficiency virus type 1 group M subtypes. *J Virol Methods*. 2006;132(1–2):181–6.
24. Vrancken B, Trovão NS, Baele G, van Wijngaerden E, Vandamme AM, van Laethem K, et al. Quantifying next generation sequencing sample pre-processing bias in HIV-1 complete genome sequencing. *Viruses*. 2016;8(1):1–11.
25. Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP. aTRAM 2.0: An Improved, Flexible Locus Assembler for NGS Data. *Evol Bioinforma*. 2018;14:0–3.
26. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011 Jul;29(7):644–52.
27. Howison M, Coetzer M, Kantor R. Measurement error and variant-calling in deep Illumina sequencing of HIV. Berger B, editor. *Bioinformatics*. 2019 Jun;35(12):2029–35.
28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
29. Sievers F, Higgins DG. Clustal Omega. In: *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2014. p. 3.13.1-3.13.16.

30. Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol.* 2010 Feb;27(2):221–4.
31. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol.* 2016;33(6):1635–8.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
33. Stanford-HIVDB. Stanford University HIV Drug Resistance Database: A curated public database to represent, store, and analyze drug resistance data. <https://hivdb.stanford.edu/>. Accessed 7 Oct 2019.
34. Wensing AM, Calvez V, Günthard HF, Ceccherini-Silberstein F, Charpentier C, Günthard HF, et al. 2019 Update of the Drug Resistance Mutations in HIV-1. *Top Antivir Med.* 2019;27(3):1–11.
35. Su CT-T, Koh DW-S, Gan SK-E. Reviewing HIV-1 Gag Mutations in Protease Inhibitors Resistance: Insights for Possible Novel Gag Inhibitor Designs. *Molecules.* 2019;24(18):3243.
36. Secretaría de Salud. Sistema de Vigilancia Epidemiológica de VIH (3er. trimestre 2020). 2020. p. 19. [https://www.gob.mx/cms/uploads/attachment/file/588571/VIH-Sida\\_3erTrim\\_2020.pdf](https://www.gob.mx/cms/uploads/attachment/file/588571/VIH-Sida_3erTrim_2020.pdf). Accessed 08 Dec 2021.
37. Vázquez-Valls E, Escoto-Delgadillo M, López-Márquez FC, Castellero-Manzano M, Echegaray-Guerrero E, Bitzer-Quintero OK, et al. Molecular epidemiology of HIV type 1 in Mexico: Emergence of BG and BF intersubtype recombinants. *AIDS Res Hum Retroviruses.* 2010;26(7):777–81.
38. IOMs Global Migration Data Portal. Migration data in South America. 2020. <https://migrationdataportal.org/regional-data-overview/migration-data-south-america>. Accessed 8 Dec 2021.
39. SECTUR, DATATUR. Results of Tourism Activity 2018. 2017. p. 1–28. <http://www.datatur.sectur.gob.mx/SitePages/versionesRAT.aspx>. Accessed 8 Dec 2021.
40. Parkin NT, Avila-Rios S, Bibby DF, Brumme CJ, Eshleman SH, Harrigan PR, et al. Multi-laboratory comparison of next-generation to sanger-based sequencing for HIV-1 drug resistance genotyping. *Viruses.* 2020; 12(7): 694.
41. Ávila-Ríos S, Parkin N, Swanstrom R, Paredes R, Shafer R, Ji H, et al. Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations. *Viruses.* 2020;12(6):617.
42. Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz JF, et al. Deep Sequencing of Protease Inhibitor Resistant HIV Patient Isolates Reveals Patterns of Correlated Mutations in Gag and Protease. *PLoS Comput Biol.* 2015;11(4):1–27.
43. Li G, Verheyen J, Rhee SY, Voet A, Vandamme AM, Theys K. Functional conservation of HIV-1 Gag: Implications for rational drug design. *Retrovirology.* 2013;10(1):1–11.
44. Blanch-Lombarte O, Santos JR, Peña R, Jiménez-Moyano E, Clotet B, Paredes R, et al. HIV-1 Gag mutations alone are sufficient to reduce darunavir susceptibility during virological failure to boosted PI therapy. *J Antimicrob Chemother.* 2020;75(9):2535–46.

45. Codoñer FM, Peña R, Blanch-Lombarte O, Jimenez-Moyano E, Pino M, Vollbrecht T, et al. Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors. *Sci Rep.* 2017;7(1):1–10.
46. Gijbsbers EF, van Nuenen AC, Schuitemaker H, Kootstra NA. Gag sequence variation in a human immunodeficiency virus type 1 transmission cluster influences viral replication fitness. *J Gen Virol.* 2013;94(PART2):354–9.
47. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, et al. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem.* 2002;277(8):5952–61.
48. Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, Wolfram J, et al. Changes in Human Immunodeficiency Virus Type 1 Gag at Positions L449 and P453 Are Linked to I50V Protease Mutants In Vivo and Cause Reduction of Sensitivity to Amprenavir and Improved Viral Fitness In Vitro. *J Virol.* 2002;76(15):7398–406.
49. Kolli M, Stawiski E, Chappey C, Schiffer CA. Human Immunodeficiency Virus Type 1 Protease-Correlated Cleavage Site Mutations Enhance Inhibitor Resistance. *J Virol.* 2009;83(21):11027–42.

## Figures

### Figure 1

Phylogenetic tree analysis of HIV gag and protease from antiretroviral treatment-naive Mexican individuals. Neighbor joining trees were constructed from the aligned sequences of a) protease and b) Gag. HIV risk factors are denoted by MSM (men who have sex with men), MSW (men who have sex with women), and MBI (men bisexual). The corresponding subgroup for each reference strain is shown next to the accession number.

### Figure 2

HIV Protease drug resistance mutation frequency and levels of PI resistance. a) Frequency of minor and major PI resistance mutations identified with NGS at different detection sensitivity thresholds in patients without experience of antiretroviral treatment (n= 96). b) Levels of resistance to protease inhibitors calculated with the Stanford HIVdb program. ATV/r=ritonavir-boosted atazanavir. LPV/r=ritonavir-boosted lopinavir. FPV/r= ritonavir-boosted fosamprenavir. IDV/r=ritonavir-boosted indinavir. NFV=nelfinavir. SQV/r=ritonavir-boosted saquinavir. TPV/r=ritonavir-boosted tipranavir.

## Figure 3

HIV Gag drug resistance-associated mutation frequency at different detection sensitivity thresholds (n = 96). Gag cleavage site mutations are indicated with an asterisk.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1Table1.docx](#)
- [Additionalfile2Table2.docx](#)
- [Additionalfile3Fig1.SequencingdepthofGagandProtease.png](#)
- [Additionalfile4Fig2.DistributionofHIVGagandProtease.png](#)
- [Additionalfile5Table3.docx](#)
- [Additionalfile6Table4.docx](#)
- [Additionalfile7Table5.docx](#)