Weaker HLA footprints on HIV in the unique and highly genetically admixed host population of Mexico

Maribel Soto-Nava[1], Santiago Avila-Ríos[1], Humberto Valenzuela-Ponce[1], Claudia García-Morales[1], Jonathan M. Carlson[2], Daniela Tapia-Trejo[1], Daniela Garrido-Rodriguez[1], Selma N. Alva-Hernández[1], Thalía A. García-Tellez[1*], Akio Murakami-Ogasawara[1], the International HIV Adaptation Collaborative, Simon A, Mallal[3,4], Mina John[4], Mark A. Brockman[5,6], Chanson J. Brumme[6], Zabrina L. Brumme[5,6#], and Gustavo Reyes-Teran[1#] and the HIV MexNet Group

[1]Center for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico

[2]Microsoft Research, Redmond, Washington, USA

[3]Vanderbilt University, Nashville Tennessee, USA

[4]Murdoch University, Perth, Australia

[5]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[6]British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada

Running Head: HLA-associated footprint on HIV in Mexico vs. Canada/USA

#Address correspondence to:

Gustavo Reyes-Terán, gustavo.reyesteran@gmail.com

Zabrina L. Brumme, zbrumme@sfu.ca

*Present address: Thalía Garcia-Tellez, Institut Pasteur, Unité HIV, Inflammation and Persistence, Paris, France.

M.S.N. and S.A.R. contributed equally to this work.

ABSTRACT

HIV circumvents HLA class I-restricted CD8+ T cell responses through selection of escape mutations that leave characteristic mutational "footprints" - also known as HLA-associated polymorphisms (HAPs) - on HIV sequences at the population level. While many HLA footprints are universal across HIV subtypes and human populations, others can be region-specific as a result of the unique immunogenetic background of each host population. Using a published probabilistic phylogenetically-informed model, we compared HAPs in HIV Gag and Pol (PR-RT) in 1,612 subtype B-infected, antiretroviral treatment-naïve individuals from Mexico and 1,641 from Canada/USA. A total of 252 HLA class I allele subtypes were represented, including 140 observed in both cohorts, 67 unique to Mexico and 45 unique to Canada/USA. At the predefined statistical threshold of $q<0.2$, 358 HAPs (201 in Gag; 157 in PR-RT) were identified in Mexico, while 905 (534 in Gag and 371 in PR-RT) were identified in Canada/USA. HAP identified in Mexico included both "canonical" HLA-associated escape pathways and novel associations, in particular with HLA alleles enriched in Amerindian and mestizo populations.

Remarkably, HLA footprints on HIV in Mexico were not only fewer but also on average significantly weaker than those in Canada/USA, though some exceptions were noted. Moreover, exploratory analyses suggested that the weaker HLA footprint on HIV in Mexico may be due, at least in part, to weaker and/or less reproducible HLA-mediated immune pressures on HIV in this population. The implications of these differences for natural and vaccine-induced anti-HIV immunity merit further investigation.

IMPORTANCE

HLA footprints on HIV identify viral regions under intense and consistent pressure by HLA-restricted immune responses and the common mutational pathways that HIV uses to evade them. In particular, HLA footprints can identify novel immunogenic regions and/or epitopes targeted by understudied HLA alleles; moreover, comparative analyses across immunogenetically distinct populations can illuminate the extent to which HIV immunogenic regions and escape pathways are shared versus population-specific, information which can in turn inform the design of universal or geographically-tailored HIV vaccines. We compared HLA-associated footprints on HIV in two immunogenetically distinct North American populations - Mexico and Canada/USA. We identify both shared and population-specific pathways of HIV adaptation, but also make the surprising observation that HLA footprints on HIV in Mexico are overall fewer and weaker than in Canada/USA, raising the possibility that HLA-restricted antiviral immune responses in Mexico may be weaker, and/or escape pathways somewhat less consistent, than in other populations.

INTRODUCTION

CD8+ cytotoxic T lymphocytes (CTLs) recognize short, HIV-derived peptide epitopes presented by Human Leukocyte Antigen (HLA) class I molecules on the surface of infected cells, thereby modulating early viremia control (1, 2) and the establishment of the viral set-point (3). HLA-restricted CTL also exert strong evolutionary pressure on HIV *in vivo,* promoting viral adaptation through the selection of escape mutations (4) that interfere with epitope processing (5), prevent binding of the viral peptide to HLA (6, 7), or affect HLA-peptide recognition by the T cell receptor (8, 9). Early observations that CTL escape in HIV tended to occur along predictable mutational pathways in persons responding to a given HLA-restricted viral epitope (10-13) led to the development of statistical approaches to systematically identify HLA-associated polymorphisms (HAPs), also known as HLA-associated "footprints" on HIV, using large population-based datasets of viral sequences linked to HLA types (13). These analyses, which identify amino acids that are statistically over- (or under-) represented among persons expressing a given HLA allele while correcting for host and viral genetic confounders (14, 15) confirmed the broadly reproducible nature of CTL escape in HIV (16-20). These studies also revealed that certain HLA-associated footprints can be host population-specific, due to substantial immunogenetic variation across human populations. For example, even though HIV subtype B predominates in Japan, Canada, USA and Australia, two-thirds of HAPs in Japan are not observed in the latter epidemics (17) due to the unique HLA distribution of the Japanese population.

89 Identification of HLA-associated footprints is relevant to HIV vaccine design. An
90 effective vaccine will need to elicit sustained immune responses capable of
91 recognizing genetically diverse viral strains, from which HIV cannot escape (ideal)
92 or can only escape at substantial fitness cost (15). One promising strategy is to
93 select immunogenic yet mutationally constrained viral regions as vaccine antigens
94 [e.g. (21, 22)], which can be further optimized for natural sequence coverage [*e.g.*
95 using mosaic designs (23, 24)]. HLA footprints are vaccine-relevant because they
96 identify HIV regions under significant and consistent immune pressure by particular
97 HLA-restricted CTL (*i.e.* immunogenic regions) and the common mutational
98 pathways that HIV uses to evade them. As such, evaluation of HLA footprints in
99 concert with information on sequence conservation, mutational fitness costs and
100 escape mechanisms can be used to identify immunogenic yet constrained viral
101 regions and immune-relevant natural HIV sequence variation within them. For
102 example, conserved epitopes and their common variants that retain intracellular
103 processing and HLA binding ability might be considered as immunogens, albeit
104 with some caution (25). In particular, comparative analyses of HLA footprints in
105 immunogenetically distinct host populations wherein the same HIV subtype
106 circulates can illuminate the extent to which viral immunogenic regions and escape
107 pathways are universal versus host population-specific, thus potentially informing
108 the design of universal and geographically-tailored vaccine strategies.

109 Towards this goal, we compare HLA-associated HIV footprints in Gag and Pol in
110 two large North American populations: the immunogenetically distinctive Mexican
111 mestizo population, which features a mixture of Caucasian and Amerindian HLA

alleles (20, 26-29), and Canada/USA. Our study thus represents a unique opportunity to investigate the impact of host immunogenetics on HLA-associated adaptation in geographically proximal HIV subtype B epidemics. The present study significantly extends a preliminary study of HIV Pol adaptation by our group (20) by increasing cohort size by more than fivefold, performing all adaptation analyses at HLA subtype-level resolution, and additionally analyzing Gag. As such, it represents the largest comparative study to date of differential HIV adaptation to HLA across human populations. Gag and Pol were studied because these proteins are rich in conserved epitopes where escape can be fitness-costly (30-32) and where responses to these epitopes are associated with superior viremia control (7, 12, 14, 33-35). Overall, our results confirm that adaptation of HIV to HLA in Mexico, like in other global populations, occurs along broadly predictable pathways. HLA footprints observed in Mexico include canonical adaptation pathways described in many other populations (e.g. B*57 Gag-T242N (36) and B*51 RT-135X (11), as well as novel pathways attributable to the unique HLA distribution of Mesoamerican peoples (e.g. Gag A*02:06–F44Y, B*39:02-E319D, and A*68:03-K436R; PR B*39:06-V15I, B*39:02-K70R, and RT B*39:02-E79D, A*68:03-R103K, and B*35:12-P294S/T). Of note however, HLA-associated HIV footprints in Mexico were overall fewer, and their strengths of selection significantly weaker, than in Canada/USA, raising the intriguing hypothesis that HLA-restricted immune responses to HIV in Mexico may be less potent, and/or HIV mutational escape pathways somewhat less consistent, than in other populations.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Ethics Committee of the National Institute of Respiratory Diseases (INER) in Mexico City (codes E02-05, E10-10), the institution leading and coordinating the study, and was conducted according to the principles of the Declaration of Helsinki. All participants gave written informed consent before blood sample donation.

Mexican cohort

Antiretroviral-naïve, chronically HIV-1 subtype B-infected Mexican individuals were enrolled from 2000 to 2014 as part of a national project to assess HIV molecular epidemiology, drug resistance surveillance and HLA adaptation. Participants were enrolled by convenience sampling in HIV clinics and reference hospitals in Mexico City and the states of Baja California, Campeche, Chiapas, Chihuahua, Colima, Guerrero, Hidalgo, Jalisco, Michoacan, Morelos, Nuevo Leon, Oaxaca, Puebla, Queretaro, Quintana Roo, Sinaloa, Sonora, State of Mexico, Tabasco, Tlaxcala, Veracruz, and Yucatan. Each participant donated a single blood sample from which plasma and buffy coat/peripheral blood mononuclear cells were isolated and cryopreserved. All blood samples were processed at the Center for Research in Infection diseases (CIENI) of INER in Mexico City. HIV plasma viral load was determined with the m2000 system (Abbott, Abbott Park, IL, USA). CD4+ T cell counts were determined by flow cytometry using the TruCount Kit in a FACSCanto II instrument (BD Bioscience, San Jose, CA, USA).

Reference Canada/USA cohort

A reference population, comprising two published cohorts of antiretroviral treatment-naïve, HIV-1 subtype B infected individuals from Canada (the British Columbia Observational Medical Evaluation and Research [HOMER] cohort; n=1,103) (16, 37) and the USA (AIDS Clinical Trials Group [ACTG] protocol 5142 participants who also provided human DNA under ACTG protocol 5128; n=538) (38, 39), for whom HIV sequences linked to HLA class I types were available, was used as a comparison group. The Canada/USA cohort was chosen as a reference because the epidemics in these two countries and in Mexico are geographically linked, concentrated in persons with similar risk factors and predominantly HIV subtype B. The Canada/USA cohorts, along with another from Australia, were previously used to identify HLA-associated polymorphisms in HIV subtype B (7); here, the Canada/USA cohorts were re-analyzed for HLA footprints specific to North America. As described in (16, 37) the majority of HLA class I types were defined at subtype-level resolution; missing or intermediate-resolution data were imputed to subtype-level using a machine learning algorithm trained on HLA-A, B and C subtypes from >13,000 individuals with known ethnicity (40). Extensive validations of method robustness to HLA imputations are provided in (7), as are instructions for access to paired HIV/HLA data from this cohort.

HIV *gag* and *pol* amplification and sequencing in the Mexican cohort

Viral RNA was isolated from cryopreserved plasma (1 mL) using the QIAamp Viral RNA kit, (QIAGEN, Valencia, CA, USA). For *gag* amplification, primers 623Fi AAATCTCTAGCAGTGGCGCCCGAACAG (HXB2 genomic nucleotide positions

623-649) and 2cRx (2826-2849) were used for the first round RT-PCR (41) with Super Script III OneStep RT PCR kit (Invitrogen, Carlsbad, CA, US) and the following PCR conditions: 30 min at 55 ºC and 2 min at 94°C, followed by 35 cycles of (15 s at 94 ºC, 30 s at 55 ºC and 2 min at 68 ºC), and finishing with 5 min at 68 ºC. Second-round products were obtained with primers G1 GCAGGACTCGGCTTGCTGAA (691-710) and G10 TATCATCTGCTCCTGTATC (2,343-2,325) using Platinum Taq DNA polymerase (Invitrogen) and the following PCR conditions: 3 min at 94 ºC, followed by 35 cycles of (30 s at 94 ºC, 30 s at 56 ºC, 2 min at 72 ºC), and finishing with 5 min at 72 ºC. All positive *gag* products confirmed by agarose gel electrophoresis were purified using QIAquick PCR Purification Kit (QIAGEN). Sequences were obtained with eight primers (G2F, GCGGCGACTGGTGAGTA (734-750); GS1R, TTATCTAAAGCTTCCTTGGTGTCT (1074-1097); GAS3F, CATCAATGAGGAAGCTGCAG (1401-1420) GAS4R, GGTTCTCTCATCTGGCCTGG (1462-1481); GAS5F, CTCTAAGAGCCGAGCAAGCT (1697-1716); GAS6R, AAAATAGTCTTACAATCTGG (1771-1790); HPR1977F, GTTAAGTGTTTCAATTGTGG (1957-1976) and GA2274R TCTTTATTGTGACGAGGGGTCG (2274-2295) using the BigDye v3.1 chemistry on a 3730xl Genetic Analyzer (Thermo Fisher, Waltham, MA, USA). Sequences were assembled and manually edited using Geneious v5.6.7 (Biomatters, Auckland, NZ), then aligned using MEGA 7 software (42).

For *pol* (PR-RT) sequences, amplification of HIV protease (99 amino acids) and the first 335 amino acids of the reverse transcriptase (RT) was performed using a

Journal of Virology

previously described in-house protocol (43). Sequences were obtained with a 3730xl Genetic Analyzer (Thermo Fisher) and were assembled using the automated basecalling software RECall (44). Negative controls were included in all amplification runs and monthly phylogenetic controls were performed, including laboratory HIV strains, to detect possible contamination.

HIV subtyping

HIV subtypes were determined using REGA HIV Subtyping Tool (3.0) (http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/) and confirmed with the Recombination Identification Program (45) (RIP, https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). All non-subtype B sequences were removed prior to analysis.

Phylogenetic analyses and cluster identification

HIV *gag* and *pol* sequences were aligned to the HIV HXB2 reference strain using an in-house alignment algorithm based on HyPhy (46), and columns where HXB2 was gapped were stripped out. Shannon entropy of amino acid alignments was computed using the Los Alamos HIV sequence database (https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html) with 500 randomizations. Maximum likelihood phylogenies were inferred using FastTree (http://www.microbesonline.org/fasttree) using the generalized time-reversible (GTR) model (47, 48). Phylogenies were colored using Rainbow Tree (49) (https://www.hiv.lanl.gov/content/sequence/RAINBOWTREE/rainbowtree.html).

Patristic distances were extracted from cohort-specific phylogenies using

PATRISTIC (50). G*ag* and *pol* sequence clusters, defined by within-cluster patristic distances ≤1.5% and bootstrap support values ≥90%, were identified using Cluster Picker (The University of Edinburgh, UK) (51). This genetic distance threshold has been used previously for inferring transmission clusters in chronic cohorts (52). HIV genetic compartmentalization between cohorts was assessed using the Fixation index (F_{ST}) score (53) implemented in HyPhy (46).

HLA typing in the Mexican cohort

Genomic DNA was extracted from a minimum of 6 million PBMC or 200 μL of buffy coat using QIAmp DNA Blood Mini Kit (QIAGEN). HLA class I HLA-A, -B, and -C typing was performed to subtype-level (4 digit) resolution using a modified in-house sequence-based method (54). Briefly, 1 kb fragments including exon 2 and 3 of HLA-A, -B and -C were amplified using universal, locus-specific primers and Roche Expand High Fidelity PCR system (Roche Applied Science, Laval, PQ, Canada). PCR products were cleaned up with ExoSAP-IT (Affimetrix, Cleveland, OH, USA) and sequenced on a 3730xl Genetic Analyzer using BigDye 3.1 chemistry (Thermo Fisher). HLA allele assignment was done using uTYPEv6 (Thermo Fisher) by comparison to the IMGT/HLA database (55, 56). Using this method, a total of 92 HLA-A, 91 HLA-B, and 39 HLA-C allele pairs within the Common and Well-Documented Catalogue (57) present polymorphism phase ambiguities at the resolution level of the first (i.e. allele-level) or second (i.e. subtype-level) HLA fields (**Table S1**). These ambiguities were resolved by assigning the most frequent allele combination according to linkage disequilibrium data obtained from our Mexican mestizo population. Ambiguous HLA pairs due to polymorphic differences outside

exons 2 and 3 were managed as G groups, including A*74:01:01G (A*74:01 in the analysis, encompassing A*74:01/A*74:02), C*18:01:01G (C*18:01 in the analysis, encompassing C*18:01/C*18:02), C*17:01:01G (C*17:01 in the analysis, encompassing C*17:01/C*17:02/C*17:03), and C*04:01:01G (C*04:01 in the analysis, encompassing C*04:01/C*04:09N) among others. All HLA haplotypes were confirmed using the HLA completion web tool (40) (available at http://boson.research.microsoft.com/hla/). Additionally, a total of 33 HLA-A or HLA-C types that failed amplification or sequencing were imputed using the same tool. HLA haplotypes with unresolved HLA-B loci were not imputed and were considered missing data (this included 8 individuals with both HLA-B alleles and 8 with one HLA-B allele missing). Raw HLA typing data are available via direct request to the authors.

Further validation of our HLA typing method in the context of a Mexican mestizo population was performed analyzing HLA data from 323 individuals from Mexico City for whom HLA typing had been performed by amplifying exons 1 to 8 for HLA-A and HLA-C, and exons 1 to 7 for HLA-B followed by next generation sequencing (TruSight HLA Kit, Illumina), thereby resolving gametic phase and achieving the highest possible typing resolution. In a blinded manner, we extracted the exon 2 and 3 consensus sequences (i.e. without gametic phase resolution) from these patients and re-interpreted them as above. Accuracy was 99.89% when comparing HLA subtypes assigned by sequencing all exons with gametic phase resolution versus exons 2 and 3 without gametic phase resolution when comparing HLA

subtypes at four-digit resolution (only 1 out of 969 HLA loci was inaccurate). The results of this validation are shown in **Table S2**.

HLA frequency comparison

HLA allelic frequencies in Mexico and Canada/USA were compared using the Los Alamos HIV Database HLA Comparison tool (https://www.hiv.lanl.gov/content/immunology/hla/hla_compare.html) which computes two-sided exact Fisher's test p-values corrected for multiple comparisons using Storey's q-value, which estimates the false discovery rate (58). Results with $p<0.05$ and $q<0.2$ were deemed statistically significant.

Identification and comparison of HLA-associated polymorphisms, including formal tests for differential escape between populations

HLA-associated polymorphisms in HIV subtype B Gag and PR-RT were identified in the Mexico and Canada/USA cohorts separately, using a published phylogenetically informed statistical model that corrects for potential host and viral genetic confounders including HLA linkage disequilibrium, the HIV phylogeny and HIV codon covariation (14). The model identifies two types of associations: "adapted" (viral amino acids over-represented in individuals expressing the HLA allele; representing the inferred escape form) and "non-adapted" (viral amino acids under-represented in individuals expressing the HLA allele, representing the inferred susceptible form). All associations with $q<0.2$ were organized into immune escape maps. We also wished to compare the strengths of association of individual HAPs across the two cohorts. To do this, we took the union of all HAPs identified in

Mexico and/or Canada/USA that were restricted by HLA class I alleles observed in a minimum of 10 individuals in both cohorts and applied a published phylogenetically-corrected logistic regression to test whether their strengths of selection by the restricting HLA allele differed significantly between the cohorts (17, 59). Briefly, for each HAP of interest, the model computes a p-value testing whether HLA-mediated selection is the same in Mexico compared to Canada/USA (null hypothesis) or whether selection differs between cohorts (alternative hypothesis) (17). As before, $q<0.2$ was defined as the significance threshold for the differential escape analysis. Finally, to demonstrate that our HLA imputation/ambiguity resolution method does not significantly affect HAP recovery or association strength in the Mexican cohort, we repeat all analyses excluding all ambiguous and imputed HLA loci for this cohort and verify that all original observations still hold (**Table S3**, **Figure S1**).

RESULTS

Cohort description

We studied two cohorts of antiretroviral-naïve, HIV-1 subtype B chronically infected individuals from Mexico (n=1,612) and Canada/USA (n=1,641). Both cohorts were predominantly male (Mexico: 78.5%; Canada/USA 85.1%). Median age at enrolment was 30 [IQR 24-38] years in Mexico and 37 [32-44] years in Canada/USA. The median pVL was 4.75 [IQR 4.18-5.27] Log_{10} RNA copies/ml in Mexico and 4.98 [4.55-5.46] Log_{10} RNA copies/ml in Canada/USA. Median CD4+ T-cell counts were 311 [IQR 121-519] cells/μl in Mexico and 260 [110-400] in

313 Canada/USA**.** Calendar year of enrolment was 2000-2014 in Mexico and 1996-
314 2004 for Canada/USA.

315 **Gag and PR-RT sequence diversity in Mexico and Canada/USA**

316 We first assessed HIV subtype B diversity and phylogenetic relationships between
317 our cohorts. Gag and PR-RT sequences were available for 1,450 and 1,529
318 individuals respectively in Mexico, and 1,320 and 1,555 individuals respectively in
319 Canada/USA. Cohort-specific consensus amino acid sequences differed at only 5
320 (of 500, 1%) Gag codons (positions 30, 312, 389, 403, 490) and 2 (of 434, 0.5%)
321 PR-RT codons (PR 93 and RT 272). Overall, Gag amino acid entropy was
322 significantly higher in Mexico compared to Canada/USA (median 0.056 versus
323 0.026 respectively; $p<0.0001$**)**: in particular 38.2% (191/500) of Gag codons
324 showed significantly higher entropy in the Mexican cohort, while only 4% (20/500)
325 showed higher entropy in the Canada/USA cohort (**Figure 1, Table S4)**. In
326 contrast, PR-RT entropy in Mexico was comparable to Canada/USA both overall
327 (median 0.022 versus 0.031 respectively; $p=0.08$) and in terms of the proportion of
328 codons with significantly higher entropy in one cohort versus the other (~17-18%)
329 (**Figure 1 and Table S5**). Next, we inferred phylogenies from Gag and PR-RT
330 nucleotide alignments **(Figure 2)**. As expected, overall Gag sequence diversity
331 exceeded that of PR-RT. Also, as expected, given their proximity on the North
332 American continent, Mexico and Canada/USA sequences were quite intermixed in
333 the phylogenies (fixation indices [F_{ST}] were very low for both Gag [0] and PR-RT
334 [0.006]). Moreover, while no statistically supported clusters containing sequences
335 from both cohorts were found at genetic distance $\leq$1.5% and bootstrap support

≥90%, increasing the distance threshold to 4.5% yielded 3 clusters for Gag and 4 in PR-RT containing sequences from both cohorts with 90% bootstrap support. On average, Mexican Gag and PR-RT sequences exhibited higher median patristic distances compared to Canada/USA sequences (Gag: 0.1612 and 0.1132; PR-RT: 0.1145 and 0.0912 for Mexico and Canada/USA respectively; $p<0.0001$ in both cases). Overall, results support an interlinked North American HIV-1 subtype B epidemic where overall nucleotide diversity is higher in Mexico.

HLA allelic frequency comparison between Mexico and Canada/USA

A total of 252 HLA class I alleles, defined at subtype-level resolution, were observed **(Figure 3 and Table S6)**. Of these, 140 were observed in both cohorts, 67 were observed exclusively in Mexico and 45 exclusively in Canada/USA. In Mexico, the most frequent HLA alleles were A*02:01, A*24:02, and A*02:06 for the A locus; B*35:01, B*39:05, and B*40:02 for the B locus; and C*04:01, C*07:02, and C*03:04 for the C locus. In Canada/USA, these were A*02:01, A*03:01, and A*01:01 for the A locus; B*07:02, B*35:01, and B*08:01 for the B locus; and C*07:02, C*07:01, and C*04:01 for the C locus **(Figure 3)**. Of the 252 HLA alleles observed, 86 (22 HLA-A, 46 HLA-B, and 18 HLA-C) differed significantly ($p<0.05$ and $q<0.2$) in frequency between Mexico and Canada/USA **(Figure 3)** (note: when HLA frequencies were computed separately by cohort, 81 alleles differed significantly in frequency between Canada and Mexico, 77 between USA and Mexico, but only 57 between Canada and USA, **Table S6**). Of these 86 HLA alleles, 41 were significantly more frequent in Mexico compared to Canada/USA, these included A*24:02, A*02:06, A*68:01, A*31:01, A*68:03, B*39:05, B*40:02,

B*39:06, C*04:01, C*07:02, C*01:02, alleles which are enriched in mestizo and Amerindian populations (26, 27, 29). Consistent with previous reports, (20, 60), most canonical protective HLA alleles (60) were enriched in the Canada/USA cohort compared to Mexico (e.g. B*57:01, B*58:01, B*27:05, B*13:02, B*42:01, B*44:03, A*25:01, A*32:01). Overall, results reveal marked immunogenetic differences between Mexico and the Canada/USA cohorts.

Differential HLA footprints on HIV Gag and PR-RT in Mexico and Canada/USA

Given the marked immunogenetic differences in neighboring North American populations, we hypothesized that HLA-associated polymorphisms would also differ between them. We identified HAPs in the Mexican and Canada/USA datasets using established methods (14) and constructed HIV immune escape maps showing HAPs identified in one or both cohorts at q<0.2 **(Figures 4 and 5; Tables S7 and S8)**. In the Mexican dataset, we identified a total of 201 HAPs (108 adapted; 93 non-adapted) that occurred at 95 (of 500, 19%) Gag codons, that were restricted by 66 HLA alleles. In the Canada/USA dataset, we identified a total of 534 HAPs, significantly more than in the Mexican dataset (p<0.0001), at 166 (32.3%) Gag codons, that were restricted by 77 HLA alleles. Overall, these summed to 662 unique HAPs identified in Gag, of which 73 (11.02%) (35 adapted and 38 non-adapted, occurring at 26 Gag codons), were identified in both cohorts, 128 were identified only in Mexico, and 461 were identified only in Canada/USA at the predefined statistical threshold of q<0.2 **(Figure 4)**. Consistent with previous reports (7) the total proportion of $p24^{Gag}$ codons harboring HAPs was lower than that of the rest of Gag, both in Mexico (12.1%, 28/231 for p24, vs. 24.5%, 66/269

for other Gag proteins) and in Canada/USA (19.0%, 44/231 vs. 45.4%, 122/269). This is expected given $p24^{Gag}$'s high overall sequence conservation (>50% of codons are 99.5%-100% conserved, which precludes identification of HLA associations at these positions). However, if one instead uses the total number of variable codons as the denominator, p24 ranks among the richest areas in the HIV proteome for HLA associations (7, 61); which is true also for Mexico.

In PR-RT, we identified 157 HAPs (78 adapted and 79 non-adapted) restricted by 58 HLA alleles, occurring at 70 codons, in the Mexican dataset. In the Canada/USA dataset we found 371 HAPs (201 adapted and 170 non-adapted) restricted by 78 HLA alleles, occurring at 105 codons, (again significantly more than in Mexico, p=0.0039) (**Figure 5)**. Overall, these summed to 470 unique HAPs identified in PR-RT, of which 58 (12.3%) were identified in both cohorts, 99 were identified only in Mexico, and 313 were identified only in Canada/USA at the predefined statistical threshold of q<0.2. Of note, all 7 codons (5 in Gag and 2 in PR-RT) where the consensus amino acid differed between cohorts showed evidence of HLA selection.

As noted above, a substantial fraction of HAPs were observed in both cohorts at q<0.2, further supporting the existence of "universal" HLA-associated escape pathways across human populations globally. These "shared" associations included canonical CTL escape pathways within epitopes restricted by protective HLA class I alleles - including B*57:01/B*57:03-Gag-T242N (within the TW10 epitope restricted by these alleles), B*27:05-Gag-R264K and L268M (within the B*27-restricted KK10 epitope) (62-66), and B*51:01-RT-I135T (within the B*51-

restricted TI8 epitope) (66, 67) as well as previously described HAPs within optimal epitopes (7) (including A*03:01-K28Q/R, A*24:02-K30R, B*57:03-A146P, B*14:01-K302R (68), B*07:02-S357G (69) and B*40:02-R429K (70) in Gag; B*44:03-E35D (71) in Protease; and A*11:01-K166R (72), B*35:01-D177E (73), and A*03:01-K277R (74) in RT) (**Figures 4 and 5**).

However, we also observed a substantial number of novel HAPs in Mexico that were not within previously described optimal epitopes. Moreover, these novel HAPs tended to be associated with Amerindian HLA alleles. Examples include A*02:06-F44Y, B*35:16-X82I, A*26:01-D230E, B*39:02-R286X, B*39:02-N315X, B*39:02-E319D, B*35:12-X357G, A*02:06-P386X, and A*68:03-K436R in Gag; B*39:06-V15I in protease; and A*02:06-I274V, and A*02:06-V276I in RT.

Taken together, HLA footprints in Mexico include both "canonical" HIV escape pathways shared across global populations as well as novel HAPs restricted by HLA alleles typically found in Amerindian or mestizo populations.

HLA footprints on HIV in Mexico are scarcer and weaker than in Canada/USA

A particularly striking observation from our analysis was the overall lower number of HLA footprints in Mexico compared to Canada/USA, despite cohorts being of comparable sizes. For example, considering only HIV codons at which adapted associations were identified with one or more HLA alleles, not only did the Mexican cohort exhibit fewer such codons in Gag compared to Canada/USA (108 at 75 Gag codons vs. 273 at 133 Gag codons respectively, $p<0.0001$) but Mexico also exhibited a lower number of adapted associations per codon (up to three HLA

alleles per codon vs. up to 7 HLA alleles per codon in Canada/USA) **(Figure 6)**. The same was true when all HLA-associated codons (adapted and non-adapted) were analyzed (data not shown). Specifically, of all Gag codons harboring HLA-adapted associations, fewer than 10% were identified exclusively in Mexico, while the remainder were observed in both cohorts (41%) or in Canada/USA only (49%) (**Figures 6A-6C**). Similar results were observed for PR-RT (78 adapted associations at 49 positions in Mexico vs. 201 adapted associations at 87 positions in Canada/USA, $p<0.0001$), where 10% were exclusively observed in Mexico, 40% were observed in both cohorts and 50% were observed in Canada/USA only ($p<0.0001$) **(Figure 6D-6F).** As a result, the number of "immunogenic zones" (defined as consecutive HIV amino acids harboring an adapted HLA association) also differed markedly between cohorts: whereas stretches of up to 11 positions under HLA pressure were observed in Canada/USA (e.g. Gag 118-128), the longest such zone was only 3 amino acids for Mexico. Furthermore, where immunogenic zones did occur in Mexico, these tended to coincide with zones also identified in Canada/USA (e.g. Gag 146-148; PR 35-37).

The scarcer HLA footprint in Mexico is likely to be at least partially attributable to the higher HIV and HLA diversity in Mexico compared to Canada/USA (**Figures 1-3**): this increases the total number of HLA-HIV pairwise comparisons required for Mexico, yielding a more stringent p-value cutoff mapping to $q<0.2$ for this cohort. Indeed, HAP identification required 726,206 HLA-HIV comparisons for Mexico compared to 592,677 for Canada/USA, such that $q<0.2$ mapped to $p<10^{-4}$ in Mexico, but $p<10^{-3}$ in Canada/USA (**Tables S7-S8**). However, our observations are

not solely explained by multiple comparisons correction. This is because, in addition to HLA footprints being overall scarcer, the statistical strengths of association between HLA alleles and HIV codons in Mexico are also overall weaker than those observed in Canada/USA. For example, a comparison of ranked $-\log_{10}$ transformed p-values for the top 201 Gag and 157 PR-RT associations between cohorts (201 and 157 because these represented the total number of HAPs identified at $q<0.2$ in Gag and PR-RT in Mexico) reveals that the Canada/USA one was always higher (*i.e.* more significant) than its corresponding Mexican one of the same ranking (**Figure 7A-7B**). This indicates that the strongest HLA footprints in Mexico are overall far weaker than the strongest HLA footprints in Canada/USA. Moreover, when analyzing only the $-\log_{10}$ p-value distribution of HAPs identified in Mexico, we observed lower (less significant) overall values for those identified exclusively in Mexico compared to those shared with Canada/USA, for both Gag ($p<0.0001$) and PR-RT ($p=0.0435$) (**Figure 7C-7D**). Thus, not only does the strength of association between HLA alleles and HIV codons appear to be inherently weaker in Mexico compared to Canada/USA, but of the HLA footprints that are detectable in Mexico, the strongest tend to be ones that are already known, whereas the novel HAPs restricted by unique mestizo HLA alleles tend to be even weaker.

We extended this analysis by comparing HAP selection strength across cohorts in a pairwise fashion. To do this, we took the union of all HAPs identified in either Mexico and/or Canada/USA that were restricted by HLA alleles observed in a minimum of 10 individuals in both cohorts (it is not possible to compare strengths

of selection of HAPs restricted by HLA alleles that are not observed, or only very rarely observed, in a given cohort). This yielded a total of 995 HAP for analysis (**Table S9**). Pairwise comparison of the absolute log-transformed odds ratios (Abs lnOR) of selection for each HAP across the two cohorts revealed statistically significantly higher values for Canada/USA (median 1.1; IQR 0.57-1.8) compared to Mexico (median 0.67; IQR 0.32-1.4), Wilcoxon matched pairs test ($p<0.0001$) (**Figure 8A)**. These results remained consistent upon stratification by HIV protein and when analyses were restricted to unique HLA-HIV codon pairs (to avoid double-counting of adapted and non-adapted associations at the same codon) ($p<0.0001$, data not shown). Similarly, results remained consistent when the analysis was restricted to shared HAPs (a total of 73 HAPs in Gag and 58 in PR-RT were observed in both Mexico and Canada/USA and were restricted by HLA alleles observed in at least 10 individuals in both cohorts): again, the absolute log-transformed odds ratios of selection of these HAPs were significantly higher in Canada/USA (median 1.8 IQR 1.3-2.5) compared to Mexico (median 1.7 IQR 1.1-2.0) overall (Wilcoxon matched pairs test; $p<0.0001$) (**Figure 8B)**. These results remained consistent upon stratification by HIV protein and when analysis was restricted to unique HLA-HIV codon pairs (all $p<0.05$, data not shown). Our observations thus indicate that, on a per-HAP basis, HLA footprints on HIV in Mexico are on average significantly weaker than in Canada/USA.

Scarcer and weaker HLA footprints on HIV in Mexico are not explained by challenges associated with HLA typing in this population

HLA class I typing of highly admixed human populations can be challenging due to elevated genetic diversity. To rule out ambiguous and/or imputed HLA calls as possible contributors to our observation of scarcer and weaker HLA footprints on HIV in Mexico, we repeated all analyses excluding N=255 HLA loci for which the original types were ambiguous in the Mexican cohort (these included 222 [92 HLA-A, 92 HLA-B, and 39 HLA-C] loci with phase ambiguities and 33 HLA-A or -C types that had been imputed due to failed amplification/sequencing. Results were entirely consistent with those of the original manuscript (**Figure S1**). Firstly, the number and location of HLA-associated polymorphisms identified Mexico were >80% consistent with those reported in the original manuscript (~20% discordance is expected given our use of a q-value correction for multiple testing; at q<0.2 we expect ~20% of identified associations to be false-positives; **Figure S1A**). Secondly, the p-values of HLA-associated polymorphisms identified in the original and revised analyses are highly concordant (Spearman's R=0.825, p<0.0001), **Figure S1**, panel **B**). Most importantly, results of the re-analysis fully corroborate our original observations of significantly fewer and weaker HLA-associated footprints in Mexico compared to Canada/USA (**Figure S1**, panels **C-H**). Results indicate that the scarcer and weaker HLA footprints on HIV in Mexico are not explained by challenges associated with HLA typing in this population.

Exploring reasons for weaker HLA selection on HIV in Mexico

Two possibilities, that are not necessarily mutually exclusive, could explain the scarcer and weaker HLA footprints on HIV in Mexico. The first is that HLA-restricted CTL responses on a given HIV codon are weaker, and/or the virus

preferred escape pathways less predictable, in Mexico than elsewhere. Therefore, for each shared *adapted* HAP restricted by an HLA allele observed in a minimum of 10 individuals in both cohorts, we compared its prevalence in persons expressing the restricting HLA with the hypothesis that, if HLA-mediated selection was weaker or less predictable in Mexico, polymorphism prevalence in HLA-expressing persons would be overall lower in Mexico compared to Canada/USA. The second possibility is that HIV sequences circulating in Mexico already harbor a high burden of HLA-adapted mutations, thus reducing power to detect further enrichment of these variants in persons expressing the restricting HLA. We therefore also compared the prevalence of each shared adapted HAP in persons *lacking* the restricting HLA with the hypothesis that, if circulating adaptation was higher in Mexico, these values would be overall higher in Mexico compared to Canada/USA.

We take the well-described B*51:01-RT I135T mutation as an example. While it is identified in both cohorts, its lnOR of selection is 1.23 in Mexico versus 2.40 in Canada/USA, a statistically significant difference (phylogenetically-informed logistic regression test $p=8.4x10^{-7}$). Computing the frequencies of RT I135T in HLA-B*51:01 and non-B*51:01 expressing individuals across cohorts, we note that less than 50% (68/143) of B*51:01-expressing Mexican individuals harbor 135T compared to nearly two thirds of B*51:01-expressing individuals in Canada/USA (105/144) (Fisher's exact test $p<0.0001$). This suggests that the weaker association between B*51:01 and RT-135T in Mexico is because B*51-restricted CTL in this population do not respond as strongly or frequently to the TI8 epitope

(or that HIV does not escape as reproducibly via selection of T at this position in response to this pressure) compared to HIV-subtype B infected populations to the north. On the other hand, the prevalence of RT-I135T in persons lacking B*51:01 is approximately 20% in both cohorts (282/1350 and 256/1330 for Mexico and Canada/USA respectively; Fisher's exact test p=0.3), suggesting that the weaker association between B*51:01 and RT-135T in Mexico is not attributable to elevated frequencies of circulating HIV harboring this mutation.

When we applied these analyses to all 61 adapted shared HAPs we observed that, overall, the proportion of individuals expressing the restricting HLA and harboring the adapted HIV variant was a median of 2.9% *lower* in Mexico compared to in Canada/USA (IQR -11.25 – 0.65%; p=0.0020, Wilcoxon matched pairs test), (**Figure 8C**), supporting weaker HLA-mediated selection in the latter region. There was nevertheless a wide distribution in the data, with certain polymorphisms observed more frequently in one cohort compared to the other. For example, among the HAPs that were observed more frequently among HLA-expressing persons in Canada/USA compared to Mexico were well-characterized escape mutations restricted by protective HLA alleles, including Gag B*57:03-242N (with 45.5% (5/11) of Mexican B*57:03s selecting for N in comparison to 75% (18/24) of Canada/USA B*57:03s), Gag B*58:01-242N (60%, 12/20 vs. 86.5%, 44/51, for Mexico and Canada/USA respectively), and B*51:01-RT135T (see above). By contrast, a minority of HAPs were observed more frequently among HLA-expressing persons in Mexico, including Gag B*57:03-146P (observed in 81.8% vs. 54.8% of B*57:03-expressing persons in Mexico compared to Canada/USA).

564 On the other hand, the frequencies of HAPs among individuals *lacking* the relevant
565 HLA allele were not overall significantly different between cohorts (p=0.9, Wilcoxon
566 matched pairs test), though we did note examples of specific HIV polymorphisms,
567 restricted by relatively common HLA alleles in Mexico that were significantly more
568 prevalent in circulation in Mexico compared to Canada/USA (*e.g*. Gag A*24:02-
569 30R: circulating frequency 54% in Mexico compared to 33% in Canada/USA; and
570 A*31:01-403K: 58% in Mexico vs. 37% in Canada/USA) (Fisher's exact test
571 p<0.0001 for both HAPs) (**Figure 8D**). Taken together, our observations suggest
572 that, even though pre-adaptation of HIV to certain common HLA alleles is observed
573 in Mexico, the sparser and weaker HLA footprints on HIV in Mexico may overall be
574 more attributable to weaker CTL pressure (and/or less reproducible escape) in this
575 population compared to those to the north.

576 **Exceptions: HLA footprints that are stronger in Mexico than in Canada/USA**

577 Although our results reveal an overall weaker HLA footprint on HIV in Mexico than
578 Canada/USA, there are nevertheless some exceptions. To identify these, we took
579 all HAPs identified in Mexico that were restricted by HLA alleles observed in a
580 minimum of 10 individuals in Canada/USA, and applied a phylogenetically-
581 corrected logistic regression test to compare their strengths of association across
582 cohorts. Of the 233 HAPs analyzed (137 in Gag and 96 in PR-RT), 45 (19.31%)
583 exhibited significantly stronger selection, as measured by higher absolute lnOR, in
584 Mexico compared to Canada/USA (all p<0.05, q<0.2) **(Figure 9)**. Among these
585 were A*24:02-374G, A*02:06-386P, B*15:01-126S, B*08:01-398Q in Gag,
586 B*39:01-15 in PR, and C*04:01-324D in RT, suggesting that these HLA alleles

mount stronger and/or more consistent immune pressure on these HIV sites in the Mexican population compared to those farther north. Of note, we found no examples of Mexican HAPs that exhibited diametrically opposed selection in Canada/USA (that is, where the significant HIV adapted form for a given HLA allele in Mexico represented the significant non-adapted form in Canada/USA, or vice-versa).

DISCUSSION

Although HLA-associated polymorphisms in HIV are being increasingly elucidated in global populations (14, 16, 17, 20, 37-39), our study is notable because it compares HLA footprints identified in Mexico, which comprises a highly genetically admixed and thus immunogenetically unique mestizo population that includes mainly Amerindian and European, but also African and East Asian ancestry components (26, 27), to those in HIV subtype B-infected populations to the north, (16, 37-39, 75), allowing us to investigate the impact of host immunogenetics on HIV adaptation in neighboring epidemics. We observed that HLA footprints on HIV in Mexico include well-known associations such as those restricted by "protective" HLA class I alleles (*e.g.* Gag B*57:01-T242N, Gag B*27:05-R264K/L268M (62-66), and RT B*51:01-I135T (66, 67)), as well as novel associations restricted by HLA alleles enriched in mestizo populations (*e.g.* B*39:02, B*39:05, B*35:12, B*35:14, A*02:06, A*68:03). Our results strengthen the growing body of evidence supporting both "universal" and region-specific immune escape pathways attributable to host population immunogenetic composition (17, 20).

609 An unanticipated observation was that HLA footprints in Mexico were overall
610 sparser (we observed 61% fewer HAPs in Mexico compared to Canada/USA) and
611 on average weaker (in terms of lower odds ratios and higher p-values) compared
612 to those in Canada/USA. While the higher HLA and HIV diversity in Mexico
613 reduces statistical power to identify associations to some extent, in part because of
614 the need to correct for a larger number of HLA/HIV comparisons, this is not the
615 sole explanator. Similarly, challenges associated with HLA typing of the highly
616 genetically admixed Mexican population was also ruled out as an explanator in
617 detailed sensitivity analyses (**Figure S1**). Rather, exploration of our data suggested
618 that the sparse HLA footprint on HIV in Mexico is *not* due to widespread viral pre-
619 adaptation (25) to HLA class I alleles (though individual exceptions were noted),
620 but rather due to weaker or less frequent HLA-restricted CTL responses on HIV,
621 and/or less reproducible viral escape from these responses, in the Mexican
622 population. The canonical B*51:01-RT-I135T association provides an example.
623 Despite similar HLA-B*51:01 and RT codon 135 frequencies across cohorts, this
624 association is significantly weaker in Mexico compared to Canada/USA. The
625 observation that RT-135T is not as prevalent among B*51:01-expressing persons
626 in Mexico (47.6%) compared to those in Canada/USA (72.4%) (note: the same is
627 true when one considers all RT codon 135 variants, i.e. RT-I135X, which occur in
628 72% of B*51:01-expressing persons in Mexico compared to 94% in Canada/USA),
629 but the frequency of RT-135T is comparable (~20%) in individuals lacking B*51:01
630 across cohorts suggests that the weaker B*51 footprint on this HIV codon in
631 Mexico is due to weaker B*51:01-mediated immune pressure (and/or less
632 reproducible viral escape) in Mexico compared to Canada/USA, and not due to

accumulation of this variant in circulation. This observation contrasts with Japan, where a similarly weak association between B*51:01-and RT-I135T at the population level in this region is instead attributable to the accumulation of this variant in circulation to the point that it has become consensus (11, 17)..

Furthermore, a much larger fraction of HAPs identified in Mexico constituted associations "shared" with Canada/USA, than vice-versa. For example, 36.6% (131/358) of HAPs identified in Mexico were shared with Canada/USA (that is, only 63.4% were specific to Mexico) whereas, of the 905 associations identified in Canada/USA, 774 (85.5%) were exclusive to this region and only 131 (14.5%) were shared with Mexico. In other words, the "footprints" left on HIV by typical "Mexican" (i.*e.* mestizo) alleles were fewer than expected given the size of the cohort. Moreover, absolute $-\log_{10}$ p-values, and lnOR of HAP unique to Mexico were significantly weaker than those shared with Canada/USA. Finally, it is worthwhile to note that the stronger HLA footprint in Canada/USA compared to Mexico is not likely to be driven by the lower frequencies of canonical protective alleles in the latter region. Support for this is provided by our analyses comparing the proportion of individuals expressing the restricting HLA and harboring the escape variant of interest, which are agnostic to HLA frequency. For example, Gag-242N was observed in 75% (18/24) of B*57:03-expressing persons in Canada/USA but only 45.5% (5/11) in Mexico and Canada/USA respectively, suggesting weaker selection strength in Mexico independent of B*57:03 prevalence. Together, our observations suggest that population-level HLA

pressures on HIV, in particular those attributable to HLAs enriched among mestizos, are inherently weaker in Mexico than in populations to the north.

Before proposing possible underlying mechanisms, some limitations and potential confounders merit mention. First, cohort CD4 count distributions suggest more advanced infection in the Canada/USA compared to the Mexico cohort; we therefore cannot rule out a longer time for within-host escape mutations to accumulate (and thus enhanced ability to detect them) in the former. However, given that the majority of escape occurs in the initial year or two following infection (76-79), and that escape is sufficiently frequent and reproducible to be detected at the population level as early as 6 months post-infection (80), and that both study cohorts are well into chronic infection, this is unlikely to fully account for the weaker HLA footprints on HIV in Mexico. Second, the cohort enrolment period was later for Mexico (2000-2014) compared to Canada/USA (1996-2004), and HIV sequence diversity was higher, raising the possibility that the Mexican epidemic may have been "older" at time of sampling than the Canada/USA one (81), and thus more pre-adapted to its host population (11, 82, 83). If so, this could reduce our overall ability to identify HAPs; however, we observed no strong evidence to support widespread pre-adaptation to all HLA alleles in Mexico (though evidence of HIV adaptation to certain common HLA alleles was indeed noted (**Figure 8D**). Furthermore, despite both epidemics being HIV subtype B, we cannot rule out the possibility that regional differences in viral backbone may influence adaptation pathways. However, the phylogenetic intermixing of study HIV sequences, and the observation of cohort consensus differences at only 7/934 (0.75%) HIV codons

Journal of Virology

argues against this as a major confounder. It is also important to note that, when designating a particular HAP as "shared" vs. "unique" to a given cohort, we are referring to HAPs identified at q<0.2 in both vs. only one cohort respectively. HAPs "unique" to a given cohort may still be present the other cohort above this significance threshold. Finally, we have not measured HLA-associated immune responses directly in this study; rather, we are using HLA footprint data to make inferences regarding the strength and reproducibility HLA-restricted antiviral cellular immune responses in given host populations (7).

We propose some hypotheses as to why HLA-mediated pressures on HIV may be weaker in Mexico. Firstly, it is possible that targeting of specific HLA-restricted CTL epitopes, and/or immunodominance hierarchies, are not as consistent in Mexico as in other populations. Host immunogenetic differences in genes encoding proteins that interact with HLA - in particular the T-cell receptor repertoire - could also explain differential recognition and/or escape within a given HLA-restricted CTL epitope across human populations (17). Indeed, our observation of substantial differential selection of HIV polymorphisms by HLA alleles present in both cohorts (**Figures 8, 9**) supports host factors beyond HLA in mediating these differences. Marked differences in HLA subtype distributions (*e.g.* the vast diversity of HLA-B*35 subtypes in Mexico compared to Canada/USA) may also play a role, as closely-related HLA alleles with similar or identical epitope binding motifs may nevertheless target epitopes at different frequencies, with different functional avidities, and elicit differential escape pathways (17, 59). The possibility of HLA locus-specific differences is also intriguing. Consistent with a dominant influence of

HLA-B in mediating anti-HIV immune responses (84), >50% of HAPs identified in Mexico and Canada/USA were HLA-B restricted; however, whereas an average of 12 HAPs were identified per HLA-B allele in Canada/USA, only 2.8 HAP were identified per HLA-B allele in Mexico. By contrast, the average number of HAPs per HLA-A and -C allele were only twofold lower in Mexico compared to Canada/USA (*e.g.* 5 and 3.4 per HLA-A and HLA-C allele in Mexico compared to 11 and 7.8 in Canada/USA, respectively), raising the intriguing possibility that individual HLA-B alleles may not restrict as broad or potent anti-HIV immune responses in Mexico as elsewhere. Also intriguing was our observation of relatively strong positive relationships between HLA frequency and the number of HLA-restricted adapted HAPs in Canada/USA (Spearman's rho=0.4098 and p=0.003 for Gag; Spearman's rho=0.3358 and p=0.0062 for PR-RT) but far less so in Mexico (p=0.0954 and rho=0.2137 for Gag and p=0.0539 and rho=0.2893 for PR-RT) (data not shown). Notable examples include B*35:01 (that restricts 3 adapted associations in Mexico and 12 in Canada/USA despite being present at comparable allele frequency) and A*02:01 (that restricts 2 adapted HIV associations in Mexico and 8 in Canada/USA despite being present at comparable frequency across cohorts. Converging selection pressures by different HLA alleles on the same HIV codon may also play a role: RT codon 135 for example harbors diametrically opposed HAPs restricted by different HLA alleles (B*51:01-135T and B*15:03-135I); it is intriguing that the latter HAP is among the few that are significantly stronger in Mexico than in Canada/USA, which could conceivably influence the strength of the B*51:01-135T association in Mexico. Overall, our observations highlight the need for detailed assessments of HLA-restricted CTL responses, possibly supplemented with the

characterization of T-cell receptor genetic and functional diversity in the Mexican mestizo population for select HLA-restricted HIV epitopes. Our observations also support extension of our analyses to other immunogenic HIV proteins such as Nef, which exhibit high HAP densities (7, 14, 17, 39).

CONCLUSION

Comparative HLA footprint studies are relevant to HIV vaccine design because they illuminate the extent to which viral immunogenic regions - and their associated escape pathways - are universal versus population-specific. Combined with information on sequence conservation, fitness costs and escape mechanisms, HLA footprints can be used to identify immunogenic yet constrained viral regions, and their common sequence variants, for potential vaccine inclusion. In particular, HLA footprints can guide the discovery of novel epitopes and/or immunogenic regions (85, 86), which may be of particular importance in understudied populations with unique HLA distributions. HLA footprints may similarly prove useful in the context of therapeutic vaccinations for reservoir eradication (87) - for example, by analyzing autologous HIV reservoir sequences to assess the burden of escape therein. Our study extends a growing body of evidence supporting both universal and population-specific HLA-associated footprints on HIV, even among neighboring epidemics where the same HIV subtype circulates. While the identification of shared immunogenic regions in Gag and Pol could support the notion of an HIV subtype B vaccine tailored to North American sequence diversity, the identification of novel HIV adaptation pathways restricted by typical "mestizo" HLA alleles, and more importantly the unexpected observation of a significantly

scarcer and weaker HLA "footprint" on HIV in Mexico, raises intriguing questions regarding the strength and quality of HLA-restricted antiviral immunity in the Mexican mestizo population and what implications this might have for vaccine-induced immune responses. Detailed characterization of HLA-restricted CTL responses in this unique population are thus merited.

ACKNOWLEDGMENTS

This work was supported by grants from the Mexican Government (Comisión de Equidad y Género de las Legislaturas LX-LXI y Comisión de Igualdad de Género de la Legislatura LXII de la H. Cámara de Diputados de la República Mexicana), received by GRT, and Consejo Nacional de Ciencia y Tecnología (CONACyT SALUD-2013-01-202475; http://www.conacyt.mx), received by SAR. This work was also supported in part by a grant from the Canadian Institutes for Health Research to ZLB, SAR, GRT, MAB, SAM and MJ (PJT-148621). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

MSN is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received a fellowship 317556 from the Consejo Nacional de Ciencia y Tecnología (CONACYT). MAB is a Tier II Canada Research Chair in Viral Pathogenesis and Immunity. ZLB is supported by a Scholar Award from the Michael Smith Foundation for Health Research.

The HIVDR MexNet Group incudes: Karla A. Romero-Mora, María Gómez Palacio, Verónica S. Quiroz-Morales, Ramón Hernández-Juan, Edna Rodríguez (National

Journal of Virology

JVI

Journal of Virology

Institute of Respiratory Diseases, Mexico City); María L. Méndez, David de los Santos Cebrero (CAPASITS Acapulco, Guerrero), César Rivera-Benítez (General Hospital, Mexico City), Juan Sierra-Madero, Audelia Alanis-Vega (National Institute of Medical Sciences and Nutrition, Mexico City), Luz A. González-Hernández, Jaime Andrade-Villanueva (Civil Hospital Fray Antonio Alcalde, Guadalajara, Jalisco), Jaime Álvarez-Zayas (CAPASITS Puerto Vallarta, Jalisco), Héctor Carrillo-Martínez (CAPASITS Nezahualcóyotl, Estado de México), José L. Centeno (CAPASITS Ecatepec, Estado de México), Everardo Barreto, Tanya Campos (CAPASITS Tlalnepantla, Estado de México), Jesús Oaxaca-Navarro (CAPASITS Cuernavaca, Morelos), Ricardo Aya-de la Fuente (CAPASITS Monterrey, Nuevo León), César A. Carrasco-Ayala, Lesvia M. Rivera-Abarca, Gabriela Velázquez (SEAI Oaxaca), Elizabeth Papaqui-Limón, Indiana Torres-Escobar (CAPASITS Puebla), María J. del Carmen-Ricalde, David Valenzo-Loaeza, Carlos A Barrera-Arellano (CAPASITS Cancún, Quintana Roo), Adrián Flores-Gaxiola (CAPASITS Culiacán, Sinaloa), Carlos A Avilez-Gaxiola (CAPASITS Hermosillo, Sonora), Adonay Jiménez-Jiménez (Dr. Juan Graham Casasus Hospital, Tabasco), Juan Beltrán-Saldaña (CAPASITS Tampico), Arturo Arteaga-Martínez (General Hospital, Veracruz), Elizabeth Domínguez-Ramírez (Subregional Hospital, Coatzacoalcos, Veracruz), Jorge M. de la Roca-Chiapas (Subregional Hospital, Rio Blanco, Veracruz), Miriam J. García-Collins, Hilda Basilio-Badillo (Subregional Hospital, Poza Rica, Veracruz), Dulce M. Cruz-Lavadores, Carlos R. González-Álvares (CAPASITS Mérida, Yucatán), Luis E Arias-Tlaculio (CAPASITS Valladolid, Yucatán), and Samuel Navarro-Álvarez (General Hospital Tijuana, Baja California).

794 **FIGURE LEGENDS**

795 **Figure 1. Entropy differences in Gag and PR-RT on HIV from Mexico and**
796 **Canada/USA.** Comparison of Shannon entropy scores at each HIV position
797 between cohorts. Each box represents an HIV codon. The top panel shows the 500
798 Gag positions; the bottom panel shows the 99 Protease (PR) and the first 335
799 reverse transcriptase (RT) codons. Positions with significantly different Shannon
800 entropy between Mexico and Canada/USA ($p<0.001$) are colored: blue for
801 positions with higher entropy in Mexico and gray for positions with higher entropy in
802 Canada/USA. Black dots denote positions with different consensus amino acids
803 between cohorts. The complete list of entropy values for Gag and PR-RT is
804 available in **Tables S4** and **S5** respectively.

805 **Figure 2. Genetic diversity of HIV-1 subtype B Gag and PR-RT from Mexico**
806 **and Canada/USA**. Unrooted maximum likelihood phylogenetic trees inferred from
807 (A) Gag (n=2,771) and (B) PR-RT sequences (n=3,084), drawn on the same
808 genetic distance scale. Branch color indicates the cohort to which the sequence
809 belongs to. Purple branches denote the HXB2 subtype B reference sequence. (C)
810 Number of clusters defined by within-cluster patristic distances ≤1.5% and
811 bootstrap support ≥90% in each tree. Median pairwise genetic distance
812 comparison between Mexico (blue box) and Canada/USA (gray box) for Gag (D)
813 and PR-RT (E).

814 **Figure 3. Differences in HLA frequencies between Mexico and Canada/USA.**
815 Comparisons of HLA frequencies between Mexico (n=1,612, blue bars) and
816 Canada/USA (n=1,641, gray bars). HLA alleles are ordered by descending

817 frequency in the Mexican cohort. HLA alleles with less than 0.1% frequency in the
818 Mexican cohort are not shown. The complete list of comparisons can be found in
819 **Table S6**. * p<0.05 and q<0.2.

820 **Figure 4. Comparative Gag immune escape map for Mexico and Canada/USA.**
821 Escape map showing the location of HLA-associated polymorphisms in HIV-1
822 subtype B *gag* sequences from Mexico (n=1,450) and Canada/USA (n=1,320). at
823 p<0.05 and q<0.2. The reference sequence represents the cohort-specific
824 consensus sequence: blue for Mexico and gray for Canada/USA. Black dots
825 denote codons where the consensus sequence differences between cohorts. Panel
826 A: Gag positions 1 - 300; Panel B: Gag positions 301 - 500. One hundred amino
827 acids are displayed per line; vertical bars separate blocks of 10 amino acids.
828 “Adapted” amino acids are shown in bold green letters and “Non-Adapted” amino
829 acids are shown in bold blue letters, along with their restricting HLA allele(s) (blue
830 for Mexico, gray for Canada/USA, and red for "shared" associations observed at
831 q<0.2 in both cohorts. Published optimal epitopes harboring HLA-polymorphism
832 associations are shown above the consensus sequences in black. The complete
833 list of associations can be found in **Tables S7** and **S8**.

834 **Figure 5. Comparative PR-RT immune escape map for Mexico and**
835 **Canada/USA.** Escape map showing the location of HLA-associated
836 polymorphisms in HIV-1 subtype B *PR-RT* sequences from Mexico (n=1,529) and
837 Canada/USA (n=1,555). Panel A PR positions 1 - 99 and RT positions 1 - 200;
838 Panel B. RT positions 201 - 335. Features of this map are the same as in **Figure 4**.
839 The complete list of associations can be found in **Tables S7** and **S8.**

840 **Figure 6. Distribution of Gag and PR-RT HIV codons harboring adapted HLA**
841 **associations in Mexico and Canada/USA.** Panels A-C depict Gag while D-F
842 depict PR-RT. Panels A and D show HIV codons harboring adapted associations in
843 Mexico, panels B and E show HIV codons harboring adapted associations in
844 Canada/USA. The number in each box corresponds to the number of HLA-adapted
845 associations observed at that specific position. Panels C and F provide a merged
846 map, showing HLA-adapted associations present in Mexico only (blue),
847 Canada/USA only (gray), and in both cohorts (red).

848 **Figure 7. Comparison of HAP p-value distributions between Mexico and**
849 **Canada/USA.** -Log_{10} p-value transformations for the top 201 Gag HAPs found in
850 Mexico and Canada/USA (A) and the top 157 PR-RT HAPs found in Mexico
851 Canada/USA (B) are shown. Transformed p-values are ranked from smallest (least
852 significant) to largest (most significant) in each cohort and plotted as paired
853 observations. The red line represents the null expectation. Panels C and D show
854 the -Log_{10} p-value distribution of shared vs. unique HAPs observed in Mexico in
855 Gag (C) and PR-RT (D).

856 **Figure 8. Weaker HLA-associated footprint in Mexico compared to**
857 **Canada/USA**. Panel A: Pairwise comparisons of the absolute log-transformed
858 odds ratios (Absolute lnOR) for all HAPs identified in Mexico and/or Canada/USA,
859 that were restricted by HLA alleles observed in a minimum of 10 individuals in both
860 cohorts (n=995). Panel B: same as A, but restricted to "shared" HAPs (i.e. those
861 identified in both cohorts at q<0.2 in the original analysis; n=131). Results support
862 a significantly weaker HLA-associated footprint in Mexico compared to

Canada/USA. Panel C: The difference in the percentage of persons *expressing* the restricting HLA and harboring the relevant adapted HIV variant in Mexico versus Canada/USA, where a negative value indicates the variant is *less frequently* found among HLA-expressing persons in Mexico. The horizontal line denotes the median, box edges denote the 25 and 75 percentiles, whiskers denote 10-90 percentiles and individual outliers are labeled. The p-value is derived from a Wilcoxon matched pairs test applied to the corresponding variant frequencies between cohorts. Panel D: The difference in the percentage of persons *lacking* the restricting HLA and harboring the relevant adapted HIV variant in Mexico versus Canada/USA. The p-value is derived from a Wilcoxon matched pairs test applied to the corresponding variant frequencies between cohorts.

Figure 9. HLA-associated HIV polymorphisms showing stronger HLA-associated selection in Mexico than in Canada/USA. We took all HLA-associated HIV polymorphisms (HAPs) identified in Mexico that were restricted by HLA alleles observed in a minimum of 10 individuals in Canada/USA, and applied a phylogenetically-corrected logistic regression test to compare their strengths of association across cohorts. HAP displaying significantly stronger HLA-associated selection in Mexico ($p<0.05$, $q<0.2$) in Gag (A) and PR-RT (B) are shown. The complete list of comparisons is available **in Table S9**.

REFERENCES

1. **Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB.** 1994. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. J Virol **68:**6103-6110.

2. **Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD.** 1994. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J Virol **68:**4650-4655.
3. **Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, Weinhold KJ, Moore S, B CCC, Letvin N, Haynes BF, Cohen MS, Hraber P, Bhattacharya T, Borrow P, Perelson AS, Hahn BH, Shaw GM, Korber BT, McMichael AJ.** 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med **206:**1253-1272.
4. **Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR, McMichael AJ.** 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature **354:**453-459.
5. **Yokomaku Y, Miura H, Tomiyama H, Kawana-Tachikawa A, Takiguchi M, Kojima A, Nagai Y, Iwamoto A, Matsuda Z, Ariyoshi K.** 2004. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. J Virol **78:**1324-1332.
6. **Bronke C, Almeida CA, McKinnon E, Roberts SG, Keane NM, Chopra A, Carlson JM, Heckerman D, Mallal S, John M.** 2013. HIV escape mutations occur preferentially at HLA-binding sites of CD8 T-cell epitopes. AIDS **27:**899-905.
7. **Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, Le AQ, Chui CK, Cotton LA, Knapp DJ, Riddler SA, Haubrich R, Nelson G, Pfeifer N, Deziel CE, Heckerman D, Apps R, Carrington M, Mallal S, Harrigan PR, John M, Brumme ZL, International HIVAC.** 2012. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. J Virol **86:**13202-13216.
8. **Price DA, West SM, Betts MR, Ruff LE, Brenchley JM, Ambrozak DR, Edghill-Smith Y, Kuroda MJ, Bogdan D, Kunstman K, Letvin NL, Franchini G, Wolinsky SM, Koup RA, Douek DC.** 2004. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. Immunity **21:**793-803.
9. **Reid SW, McAdam S, Smith KJ, Klenerman P, O'Callaghan CA, Harlos K, Jakobsen BK, McMichael AJ, Bell JI, Stuart DI, Jones EY.** 1996. Antagonist HIV-1 Gag peptides induce structural changes in HLA B8. J Exp Med **184:**2279-2286.
10. **Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan K M, Desouza I, Feeney ME, Eldridge RL, Maier EL, Kaufmann DE, Lahaie MP, Reyor L, Tanzi G, Johnston MN, Brander C, Draenert R, Rockstroh JK, Jessen H, Rosenberg ES, Mallal SA, Walker BD.** 2005. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J Virol **79:**13239-13249.
11. **Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, Kuse N, Oka S, Duda A, Prendergast A, Crawford H, Leslie A, Brumme Z, Brumme C, Allen T, Brander C, Kaslow R, Tang J, Hunter E, Allen S, Mulenga J, Branch S, Roach T, John M, Mallal S, Ogwu A, Shapiro R, Prado JG, Fidler S, Weber J, Pybus OG, Klenerman P, Ndung'u T, Phillips R, Heckerman D, Harrigan PR, Walker BD, Takiguchi M, Goulder P.** 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature **458:**641-645.
12. **Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, Brander C, Bishop K, Mlotshwa N, Mullins JI, Coovadia H, Ndung'u T, Walker BD, Heckerman D, Goulder PJ.** 2008. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. J Virol **82:**8548-8559.

13. **Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA.** 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science **296:**1439-1443.
14. **Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D.** 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol **4:**e1000225.
15. **Carlson JM, Le AQ, Shahid A, Brumme ZL.** 2015. HIV-1 adaptation to HLA: a window into virus-host immune interactions. Trends Microbiol **23:**212-224.
16. **Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, Swenson LC, Tao I, Szeto S, Rosato P, Sela J, Kadie CM, Frahm N, Brander C, Haas DW, Riddler SA, Haubrich R, Walker BD, Harrigan PR, Heckerman D, Mallal S.** 2009. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. PLoS One **4:**e6687.
17. **Chikata T, Carlson JM, Tamura Y, Borghan MA, Naruto T, Hashimoto M, Murakoshi H, Le AQ, Mallal S, John M, Gatanaga H, Oka S, Brumme ZL, Takiguchi M.** 2014. Host-specific adaptation of HIV-1 subtype B in the Japanese population. J Virol **88:**4764-4775.
18. **Huang KH, Goedhals D, Carlson JM, Brockman MA, Mishra S, Brumme ZL, Hickling S, Tang CS, Miura T, Seebregts C, Heckerman D, Ndung'u T, Walker B, Klenerman P, Steyn D, Goulder P, Phillips R, Bloemfontein-Oxford Collaborative G, van Vuuren C, Frater J.** 2011. Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. PLoS One **6:**e19018.
19. **Rousseau CM, Lockhart DW, Listgarten J, Maley SN, Kadie C, Learn GH, Nickle DC, Heckerman DE, Deng W, Brander C, Ndung'u T, Coovadia H, Goulder PJ, Korber BT, Walker BD, Mullins JI.** 2009. Rare HLA drive additional HIV evolution compared to more frequent alleles. AIDS Res Hum Retroviruses **25:**297-303.
20. **Avila-Rios S, Ormsby CE, Carlson JM, Valenzuela-Ponce H, Blanco-Heredia J, Garrido-Rodriguez D, Garcia-Morales C, Heckerman D, Brumme ZL, Mallal S, John M, Espinosa E, Reyes-Teran G.** 2009. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. Retrovirology **6:**72.
21. **Rolland M, Nickle DC, Mullins JI.** 2007. HIV-1 group M conserved elements vaccine. PLoS Pathog **3:**e157.
22. **Borthwick N, Ahmed T, Ondondo B, Hayes P, Rose A, Ebrahimsa U, Hayton EJ, Black A, Bridgeman A, Rosario M, Hill AV, Berrie E, Moyle S, Frahm N, Cox J, Colloca S, Nicosia A, Gilmour J, McMichael AJ, Dorrell L, Hanke T.** 2014. Vaccine-elicited human T cells recognizing conserved protein regions inhibit HIV-1. Mol Ther **22:**464-475.
23. **Ondondo B, Murakoshi H, Clutton G, Abdul-Jawad S, Wee EG, Gatanaga H, Oka S, McMichael AJ, Takiguchi M, Korber B, Hanke T.** 2016. Novel Conserved-region T-cell Mosaic Vaccine With High Global HIV-1 Coverage Is Recognized by Protective Responses in Untreated Infection. Mol Ther **24:**832-842.
24. **Abdul-Jawad S, Ondondo B, van Hateren A, Gardner A, Elliott T, Korber B, Hanke T.** 2016. Increased Valency of Conserved-mosaic Vaccines Enhances the Breadth and Depth of Epitope Recognition. Mol Ther **24:**375-384.
25. **Carlson JM, Du VY, Pfeifer N, Bansal A, Tan VY, Power K, Brumme CJ, Kreimer A, DeZiel CE, Fusi N, Schaefer M, Brockman MA, Gilmour J, Price MA, Kilembe W, Haubrich R, John M, Mallal S, Shapiro R, Frater J, Harrigan PR, Ndung'u T, Allen S, Heckerman D, Sidney J, Allen TM, Goulder PJ, Brumme ZL, Hunter E, Goepfert PA.** 2016. Impact of pre-adapted HIV transmission. Nat Med **22:**606-613.

26. **Barquera R, Zuniga J, Hernandez-Diaz R, Acuna-Alonzo V, Montoya-Gama K, Moscoso J, Torres-Garcia D, Garcia-Salas C, Silva B, Cruz-Robles D, Arnaiz-Villena A, Vargas-Alarcon G, Granados J.** 2008. HLA class I and class II haplotypes in admixed families from several regions of Mexico. Mol Immunol **45:**1171-1178.

27. **Gorodezky C, Alaez C, Vazquez-Garcia MN, de la Rosa G, Infante E, Balladares S, Toribio R, Perez-Luque E, Munoz L.** 2001. The genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites. Hum Immunol **62:**979-991.

28. **Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, Acuna-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S, Ortiz-Tello P, Robles V, Kenny EE, Nuno-Arana I, Barquera-Lozano R, Macin-Perez G, Granados-Arriola J, Huntsman S, Galanter JM, Via M, Ford JG, Chapela R, Rodriguez-Cintron W, Rodriguez-Santana JR, Romieu I, Sienra-Monge JJ, del Rio Navarro B, London SJ, Ruiz-Linares A, Garcia-Herrera R, Estrada K, Hidalgo-Miranda A, Jimenez-Sanchez G, Carnevale A, Soberon X, Canizales-Quinteros S, Rangel-Villalobos H, Silva-Zolezzi I, Burchard EG, Bustamante CD.** 2014. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science **344:**1280-1285.

29. **Zuniga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, Acuna-Alonzo V, Yunis M, Granados-Montiel J, Cruz-Lagunas A, Vargas-Alarcon G, Rodriguez-Reyna TS, Fernandez-Vina M, Granados J, Yunis EJ.** 2013. HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations. PLoS One **8:**e74442.

30. **Kozisek M, Henke S, Saskova KG, Jacobs GB, Schuch A, Buchholz B, Muller V, Krausslich HG, Rezacova P, Konvalinka J, Bodem J.** 2012. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. Antimicrob Agents Chemother **56:**4320-4330.

31. **Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, Thobakgale C, Honeyborne I, Crawford H, Matthews P, Pillay T, Rousseau C, Mullins JI, Brander C, Walker BD, Stuart DI, Kiepiela P, Goulder P.** 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J Virol **80:**3617-3623.

32. **Murakoshi H, Koyanagi M, Chikata T, Rahman MA, Kuse N, Sakai K, Gatanaga H, Oka S, Takiguchi M.** 2017. Accumulation of Pol Mutations Selected by HLA-B*52:01-C*12:02 Protective Haplotype-Restricted Cytotoxic T Lymphocytes Causes Low Plasma Viral Load Due to Low Viral Fitness of Mutant Viruses. J Virol **91**.

33. **Singh AR, Hill RL, Lingappa JR.** 2001. Effect of mutations in Gag on assembly of immature human immunodeficiency virus type 1 capsids in a cell-free system. Virology **279:**257-270.

34. **Boutwell CL, Carlson JM, Lin TH, Seese A, Power KA, Peng J, Tang Y, Brumme ZL, Heckerman D, Schneidewind A, Allen TM.** 2013. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. J Virol **87:**3952-3965.

35. **Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, Schneidewind A, Power KA, Toth I, Frahm N, Alter G, Brander C, Carrington M, Walker BD, Altfeld M, Heckerman D, Allen TM.** 2009. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. J Virol **83:**1845-1855.

36. **Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, St**

John A, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJ. 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat Med **10:**282-289.

37. **Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, Mo T, Hogg RS, Montaner JS, Frahm N, Brander C, Walker BD, Harrigan PR.** 2007. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. PLoS Pathog **3:**e94.

38. **Almeida CA, Bronke C, Roberts SG, McKinnon E, Keane NM, Chopra A, Kadie C, Carlson J, Haas DW, Riddler SA, Haubrich R, Heckerman D, Mallal S, John M.** 2011. Translation of HLA-HIV associations to the cellular level: HIV adapts to inflate CD8 T cell responses against Nef and HLA-adapted variant epitopes. J Immunol **187:**2502-2513.

39. **John M, Heckerman D, James I, Park LP, Carlson JM, Chopra A, Gaudieri S, Nolan D, Haas DW, Riddler SA, Haubrich R, Mallal S.** 2010. Adaptive interactions between HLA and HIV-1: highly divergent selection imposed by HLA class I molecules with common supertype motifs. J Immunol **184:**4368-4377.

40. **Listgarten J, Brumme Z, Kadie C, Xiaojiang G, Walker B, Carrington M, Goulder P, Heckerman D.** 2008. Statistical resolution of ambiguous HLA typing data. PLoS Comput Biol **4:**e1000016.

41. **Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, Trocha A, Block BL, Schneidewind A, Allen TM, Heckerman D, Walker BD.** 2009. HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J Virol **83:**140-149.

42. **Kumar S, Stecher G, Tamura K.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **33:**1870-1874.

43. **Avila-Rios S, Garcia-Morales C, Garrido-Rodriguez D, Tapia-Trejo D, Giron-Callejas AC, Mendizabal-Burastero R, Escobar-Urias IY, Garcia-Gonzalez BL, Navas-Castillo S, Pinzon-Meza R, Mejia-Villatoro CR, Reyes-Teran G.** 2015. HIV-1 drug resistance surveillance in antiretroviral treatment-naive individuals from a reference hospital in Guatemala, 2010-2013. AIDS Res Hum Retroviruses **31:**401-411.

44. **Woods CK, Brumme CJ, Liu TF, Chui CK, Chu AL, Wynhoven B, Hall TA, Trevino C, Shafer RW, Harrigan PR.** 2012. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol **50:**1936-1942.

45. **Siepel AC, Halpern AL, Macken C, Korber BT.** 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. AIDS Res Hum Retroviruses **11:**1413-1416.

46. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676-679.

47. **Price MN, Dehal PS, Arkin AP.** 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol **26:**1641-1650.

48. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One **5:**e9490.

49. **Paradis E, Claude J, Strimmer K.** 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20:**289-290.

50. **Fourment M, Gibbs MJ.** 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol Biol **6:**1.

51. **Ragonnet-Cronin M, Hodcroft E, Hue S, Fearnhill E, Delpech V, Brown AJ, Lycett S, Database UHDR.** 2013. Automated analysis of phylogenetic clusters. BMC Bioinformatics **14:**317.

52. **Chaillon A, Avila-Rios S, Wertheim JO, Dennis A, Garcia-Morales C, Tapia-Trejo D, Mejia-Villatoro C, Pascale JM, Porras-Cortes G, Quant-Duran CJ, Lorenzana I, Meza RI, Palou EY, Manzanero M, Cedillos RA, Reyes-Teran G, Mehta SR, Mesoamerican Project G.** 2017. Identification of major routes of HIV transmission throughout Mesoamerica. Infect Genet Evol **54:**98-107.

53. **Hudson RR, Slatkin M, Maddison WP.** 1992. Estimation of levels of gene flow from DNA sequence data. Genetics **132:**583-589.

54. **Cotton LA, Abdur Rahman M, Ng C, Le AQ, Milloy MJ, Mo T, Brumme ZL.** 2012. HLA class I sequence-based typing using DNA recovered from frozen plasma. J Immunol Methods **382:**40-47.

55. **Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG.** 2013. IPD--the Immuno Polymorphism Database. Nucleic Acids Res **41:**D1234-1240.

56. **Robinson J, Waller MJ, Stoehr P, Marsh SG.** 2005. IPD--the Immuno Polymorphism Database. Nucleic Acids Res **33:**D523-526.

57. **Mack SJ, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, Moraes ME, Pereira SE, Kempenich JH, Reed EF, Setterholm M, Smith AG, Tilanus MG, Torres M, Varney MD, Voorter CE, Fischer GF, Fleischhauer K, Goodridge D, Klitz W, Little AM, Maiers M, Marsh SG, Muller CR, Noreen H, Rozemuller EH, Sanchez-Mazas A, Senitzer D, Trachtenberg E, Fernandez-Vina M.** 2013. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. Tissue Antigens **81:**194-203.

58. **Storey JD.** 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Serie B (Statistical Methodology) **64:**479 - 498.

59. **Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, Ndung'u T, Shapiro R, Frater J, Brumme ZL, Goulder PJ, Heckerman D.** 2012. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. J Virol **86:**5230-5243.

60. **Goulder PJ, Walker BD.** 2012. HIV and HLA class I: an evolving relationship. Immunity **37:**426-440.

61. **Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P.** 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat Med **13:**46-53.

62. **Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB.** 2007. A whole-genome association study of major determinants for host control of HIV-1. Science **317:**944-947.

63. **International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, Carrington M, Kadie CM, Carlson JM, Heckerman D, Graham RR, Plenge RM, Deeks SG, Gianniny L, Crawford G, Sullivan J, Gonzalez E, Davies L, Camargo A, Moore JM, Beattie N, Gupta S, Crenshaw A, Burtt NP, Guiducci C, Gupta N, Gao X, Qi Y, Yuki Y, Piechocka-Trocha A, Cutrell E, Rosenberg R, Moss KL, Lemay P, O'Leary J, Schaefer T, Verma P, Toth I, Block B, Baker B, Rothchild A, Lian J, Proudfoot J, Alvino DM, Vine S, Addo MM, et al.** 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science **330:**1551-1557.

64. **Kaslow RA, McNicholl JM.** 1999. Genetic determinants of HIV-1 infection and its manifestations. Proc Assoc Am Physicians **111:**299-307.

65. **Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M.** 2000. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci U S A **97:**2709-2714.

66. **O'Brien SJ, Gao X, Carrington M.** 2001. HLA and AIDS: a cautionary tale. Trends Mol Med **7:**379-381.

67. **Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL.** 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med **2:**405-411.

68. **Rowland-Jones SL, Dong T, Fowke KR, Kimani J, Krausa P, Newell H, Blanchard T, Ariyoshi K, Oyugi J, Ngugi E, Bwayo J, MacDonald KS, McMichael AJ, Plummer FA.** 1998. Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. J Clin Invest **102:**1758-1765.

69. **Kunwar P, Hawkins N, Dinges WL, Liu Y, Gabriel EE, Swan DA, Stevens CE, Maenza J, Collier AC, Mullins JI, Hertz T, Yu X, Horton H.** 2013. Superior control of HIV-1 replication by CD8+ T cells targeting conserved epitopes: implications for HIV vaccine design. PLoS One **8:**e64405.

70. **Sabbaj S, Bansal A, Ritter GD, Perkins C, Edwards BH, Gough E, Tang J, Szinger JJ, Korber B, Wilson CM, Kaslow RA, Mulligan MJ, Goepfert PA.** 2003. Cross-reactive CD8+ T cell epitopes identified in US adolescent minorities. J Acquir Immune Defic Syndr **33:**426-438.

71. **Pereyra F, Heckerman D, Carlson JM, Kadie C, Soghoian DZ, Karel D, Goldenthal A, Davis OB, DeZiel CE, Lin T, Peng J, Piechocka A, Carrington M, Walker BD.** 2014. HIV control is mediated in part by CD8+ T-cell targeting of specific epitopes. J Virol **88:**12937-12948.

72. **Streeck H, Lu R, Beckwith N, Milazzo M, Liu M, Routy JP, Little S, Jessen H, Kelleher AD, Hecht F, Sekaly RP, Alter G, Heckerman D, Carrington M, Rosenberg ES, Altfeld M.** 2014. Emergence of individual HIV-specific CD8 T cell responses during primary HIV-1 infection can determine long-term disease outcome. J Virol **88:**12793-12801.

73. **Rowland-Jones S, Sutton J, Ariyoshi K, Dong T, Gotch F, McAdam S, Whitby D, Sabally S, Gallimore A, Corrah T, Takiguchi M, Schultz T, McMichael AJ, Whittle H.** 1995. HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. Nat Med **1:**59-64.

74. **Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker B, Kadie C, Markowitz M, Jessen H, Kelleher AD, Rosenberg E, Kaldor J, Yuki Y, Carrington M, Allen TM, Mallal S, Altfeld M, Heckerman D, Walker BD.** 2008. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. J Virol **82:**9216-9227.

75. **Carlson JM, Brumme ZL.** 2008. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. Microbes Infect **10:**455-461.

76. **Payne RP, Kloverpris H, Sacha JB, Brumme Z, Brumme C, Buus S, Sims S, Hickling S, Riddell L, Chen F, Luzzi G, Edwards A, Phillips R, Prado JG, Goulder PJ.** 2010. Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. J Virol **84:**10543-10557.

77. **Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT.** 2010. Transmission of

single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One **5:**e12303.

78. **Lemey P, Rambaut A, Pybus OG.** 2006. HIV evolutionary dynamics within and among hosts. AIDS Rev **8:**125-140.

79. **Koibuchi T, Allen TM, Lichterfeld M, Mui SK, O'Sullivan KM, Trocha A, Kalams SA, Johnson RP, Walker BD.** 2005. Limited sequence evolution within persistently targeted CD8 epitopes in chronic human immunodeficiency virus type 1 infection. J Virol **79:**8171-8181.

80. **Martin E, Carlson JM, Le AQ, Chopera DR, McGovern R, Rahman MA, Ng C, Jessen H, Kelleher AD, Markowitz M, Allen TM, Milloy MJ, Carrington M, Wainberg MA, Brumme ZL.** 2014. Early immune adaptation in HIV-1 revealed by population-level approaches. Retrovirology **11:**64.

81. **Kinloch NN, MacMillan DR, Le AQ, Cotton LA, Bangsberg DR, Buchbinder S, Carrington M, Fuchs J, Harrigan PR, Koblin B, Kushel M, Markowitz M, Mayer K, Milloy MJ, Schechter MT, Wagner T, Walker BD, Carlson JM, Poon AF, Brumme ZL.** 2016. Population-Level Immune-Mediated Adaptation in HIV-1 Polymerase during the North American Epidemic. J Virol **90:**1244-1258.

82. **Leslie A, Kavanagh D, Honeyborne I, Pfafferott K, Edwards C, Pillay T, Hilton L, Thobakgale C, Ramduth D, Draenert R, Le Gall S, Luzzi G, Edwards A, Brander C, Sewell AK, Moore S, Mullins J, Moore C, Mallal S, Bhardwaj N, Yusim K, Phillips R, Klenerman P, Korber B, Kiepiela P, Walker B, Goulder P.** 2005. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. J Exp Med **201:**891-902.

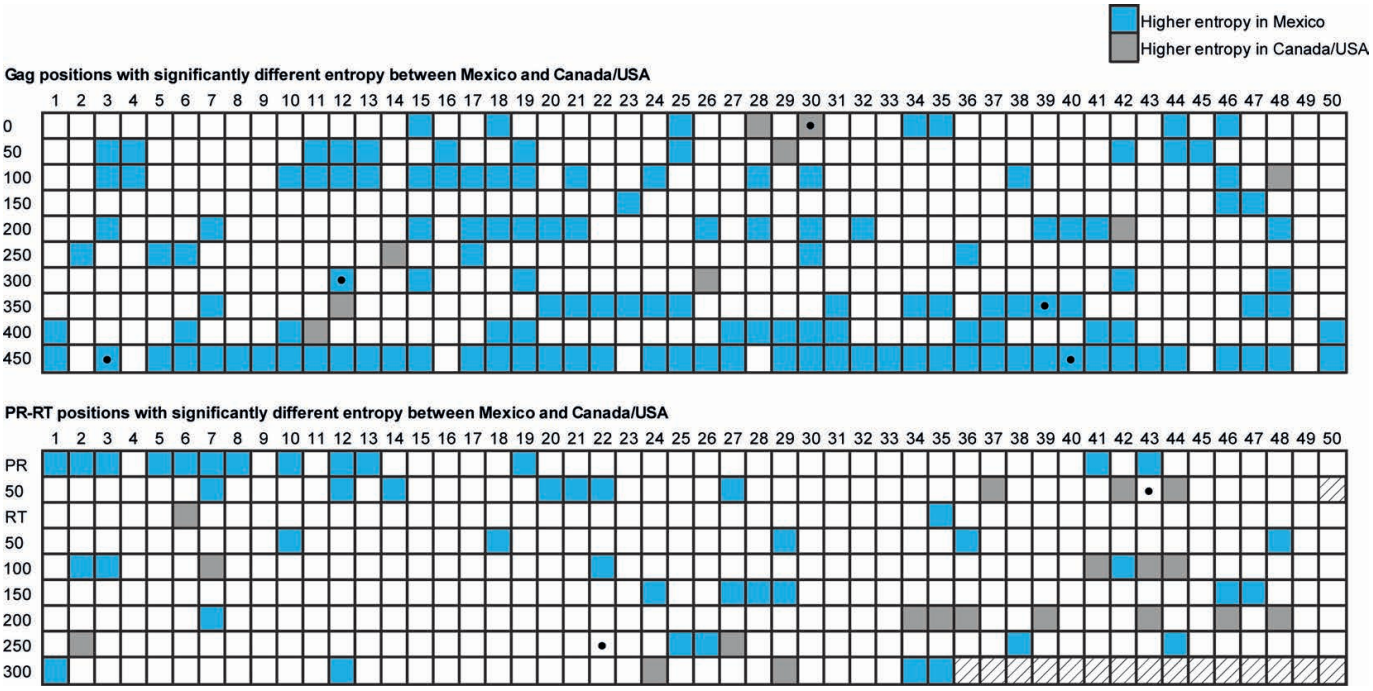
83. **Payne RP, Branch S, Kloverpris H, Matthews PC, Koofhethile CK, Strong T, Adland E, Leitman E, Frater J, Ndung'u T, Hunter E, Haubrich R, Mothe B, Edwards A, Riddell L, Chen F, Harrigan PR, Brumme ZL, Mallal S, John M, Jooste JP, Shapiro R, Deeks SG, Walker BD, Brander C, Landis C, Carlson JM, Prado JG, Goulder PJ.** 2014. Differential escape patterns within the dominant HLA-B*57:03-restricted HIV Gag epitope reflect distinct clade-specific functional constraints. J Virol **88:**4668-4678.

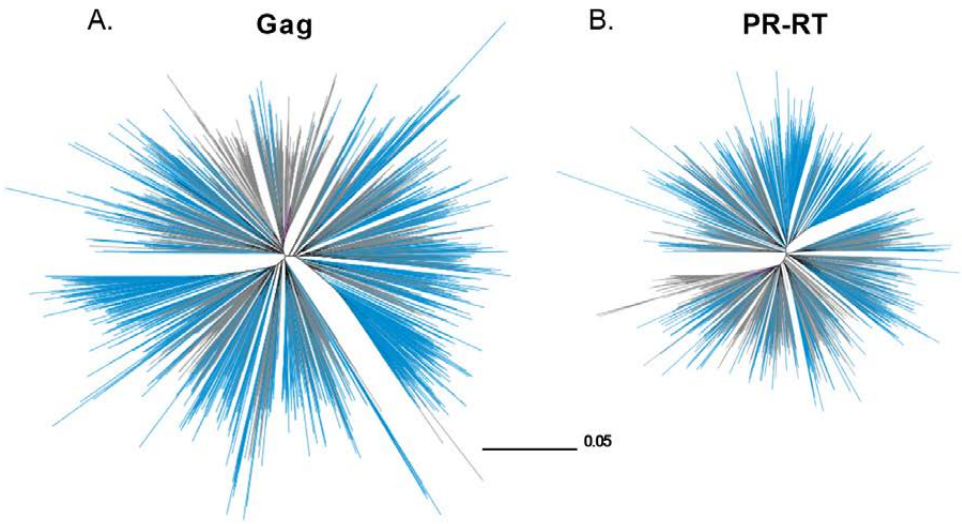
84. **Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature **432:**769-775.

85. **Almeida CA, Roberts SG, Laird R, McKinnon E, Ahmad I, Keane NM, Chopra A, Kadie C, Heckerman D, Mallal S, John M.** 2010. Exploiting knowledge of immune selection in HIV-1 to detect HIV-specific CD8 T-cell responses. Vaccine **28:**6052-6057.

86. **Shahid A, Olvera A, Anmole G, Kuang XT, Cotton LA, Plana M, Brander C, Brockman MA, Brumme ZL.** 2015. Consequences of HLA-B*13-Associated Escape Mutations on HIV-1 Replication and Nef Function. J Virol **89:**11557-11571.

87. **Jones RB, Walker BD.** 2016. HIV-specific CD8(+) T cells and HIV eradication. J Clin Invest **126:**455-463.

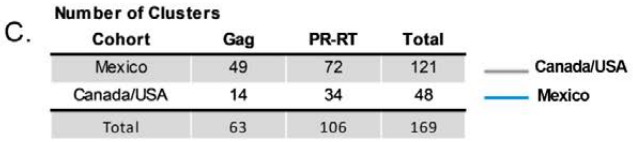


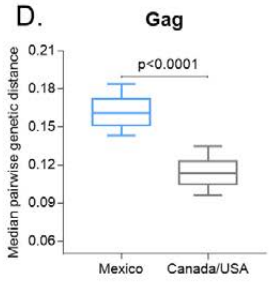


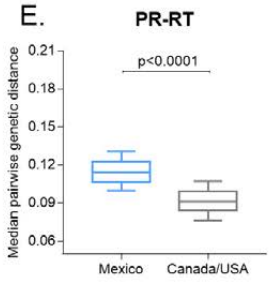
C.

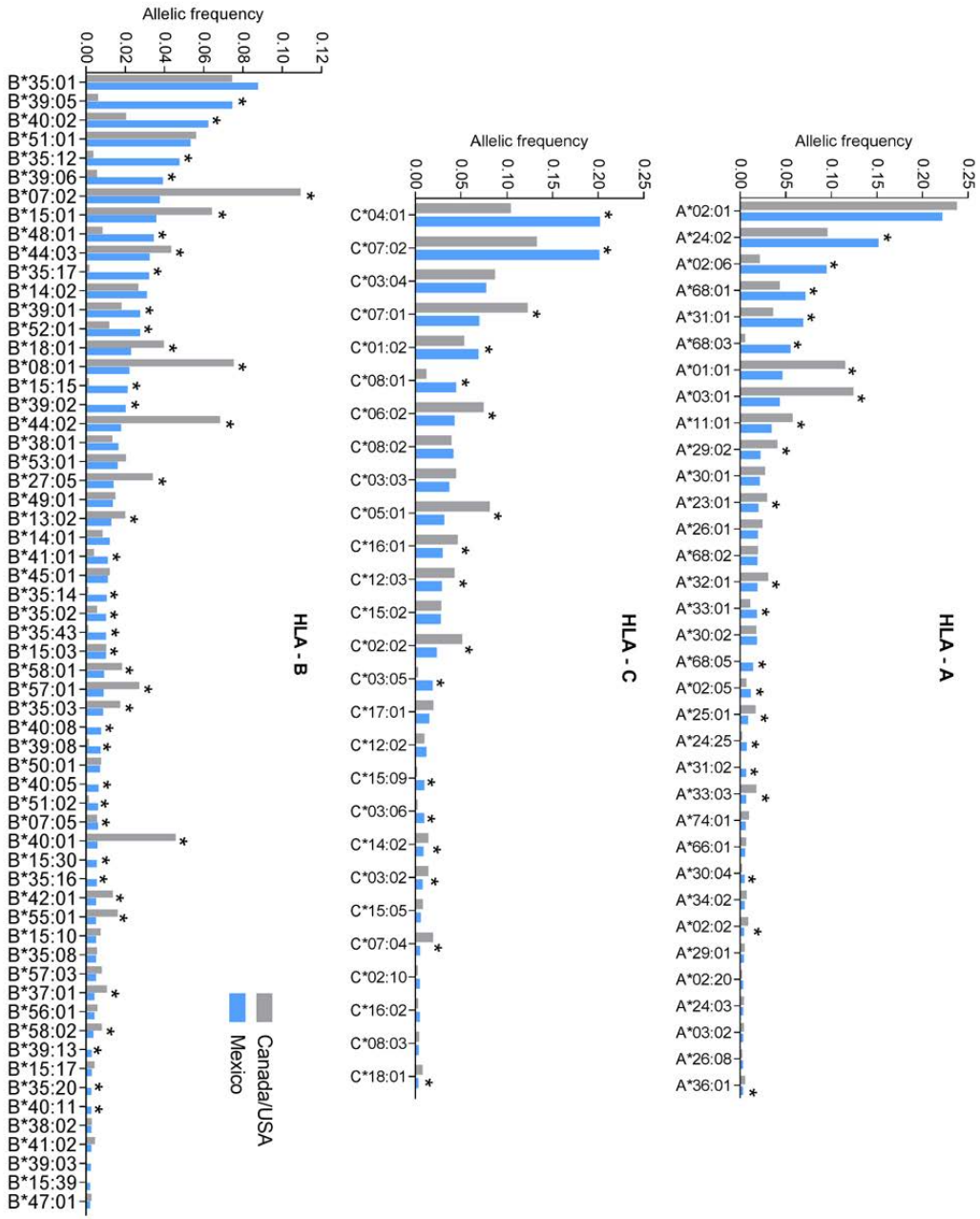
Number of Clusters

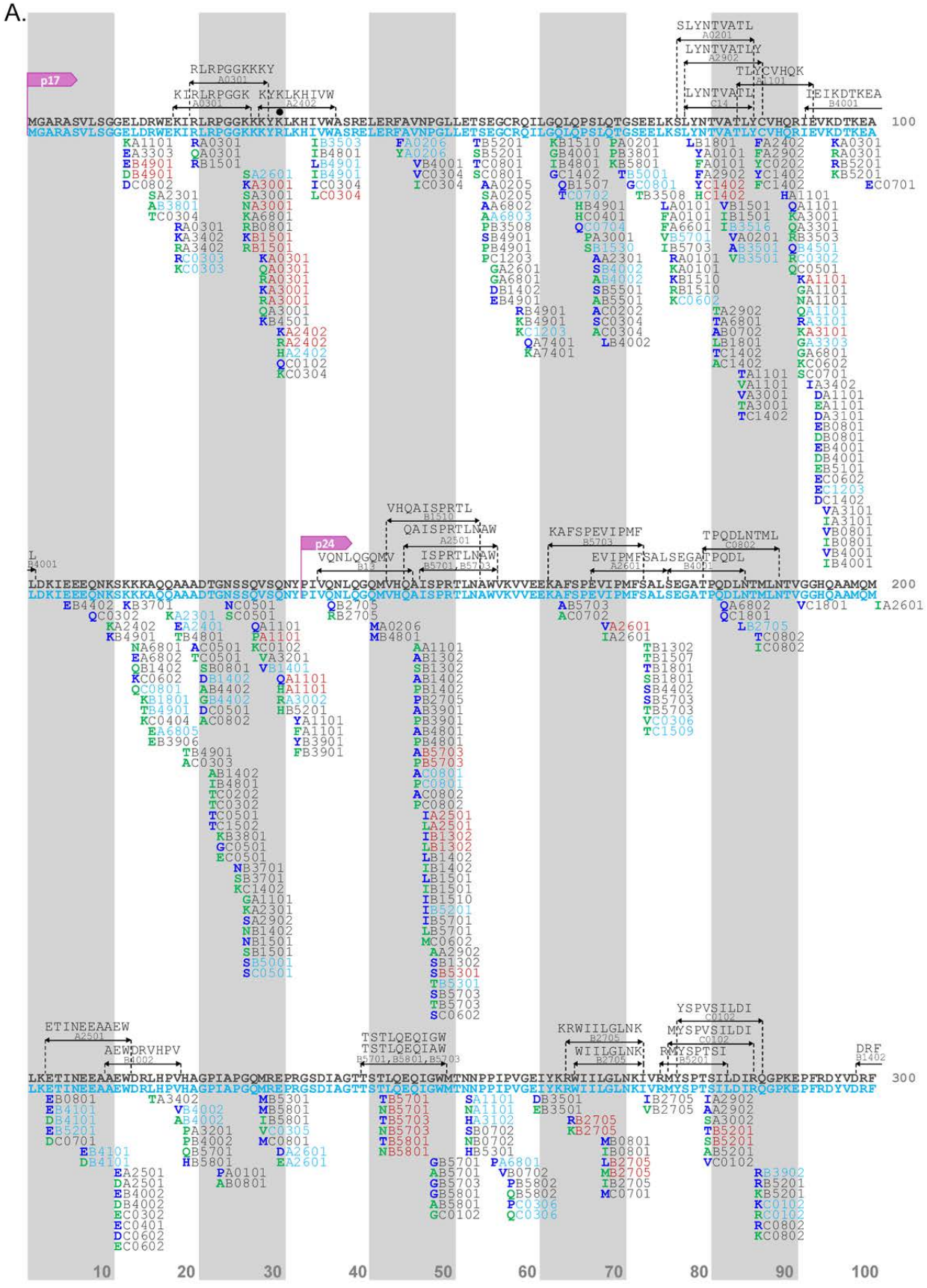
Cohort	Gag	PR-RT	Total
Mexico	49	72	121
Canada/USA	14	34	48
Total	63	106	169

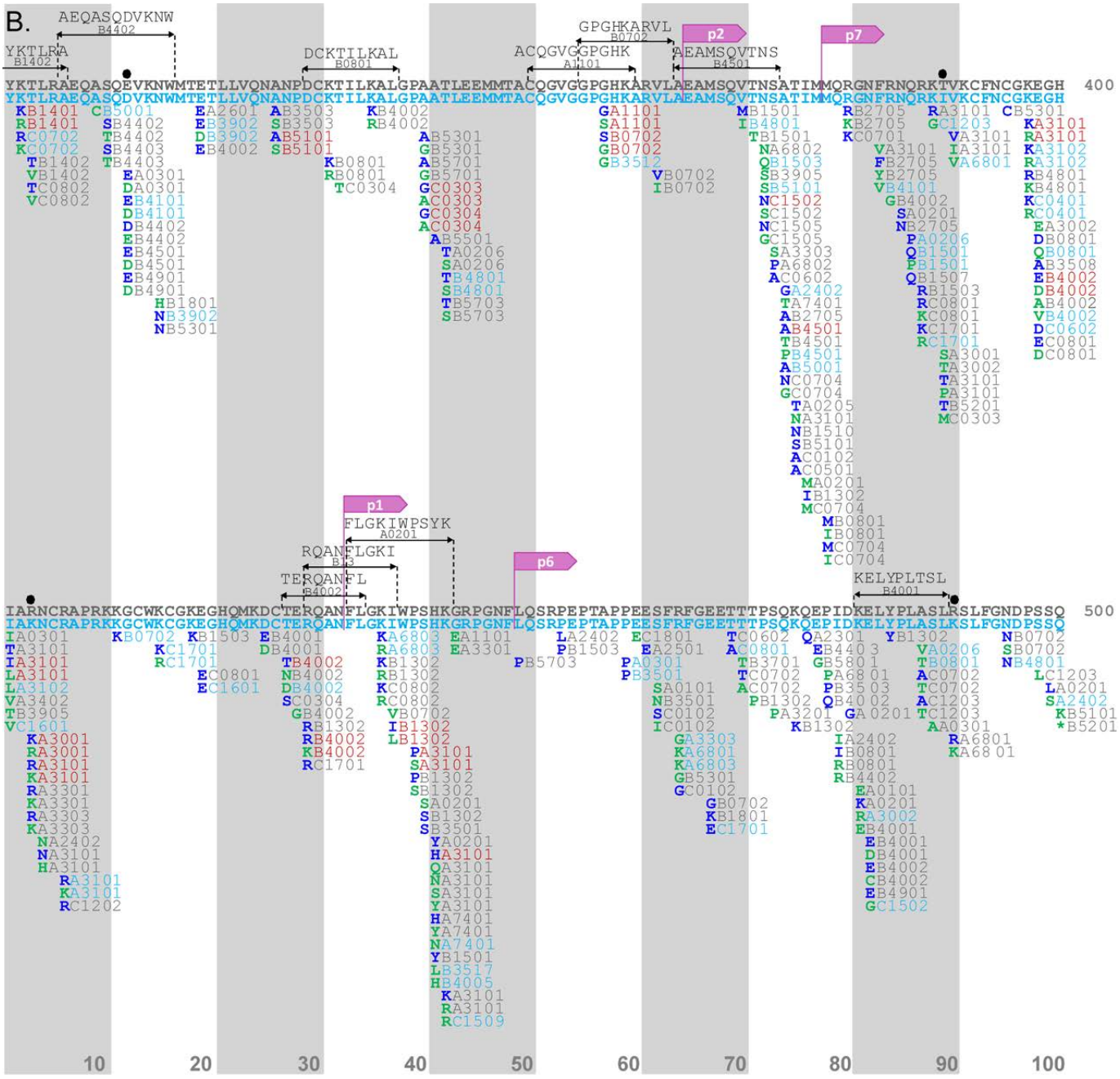


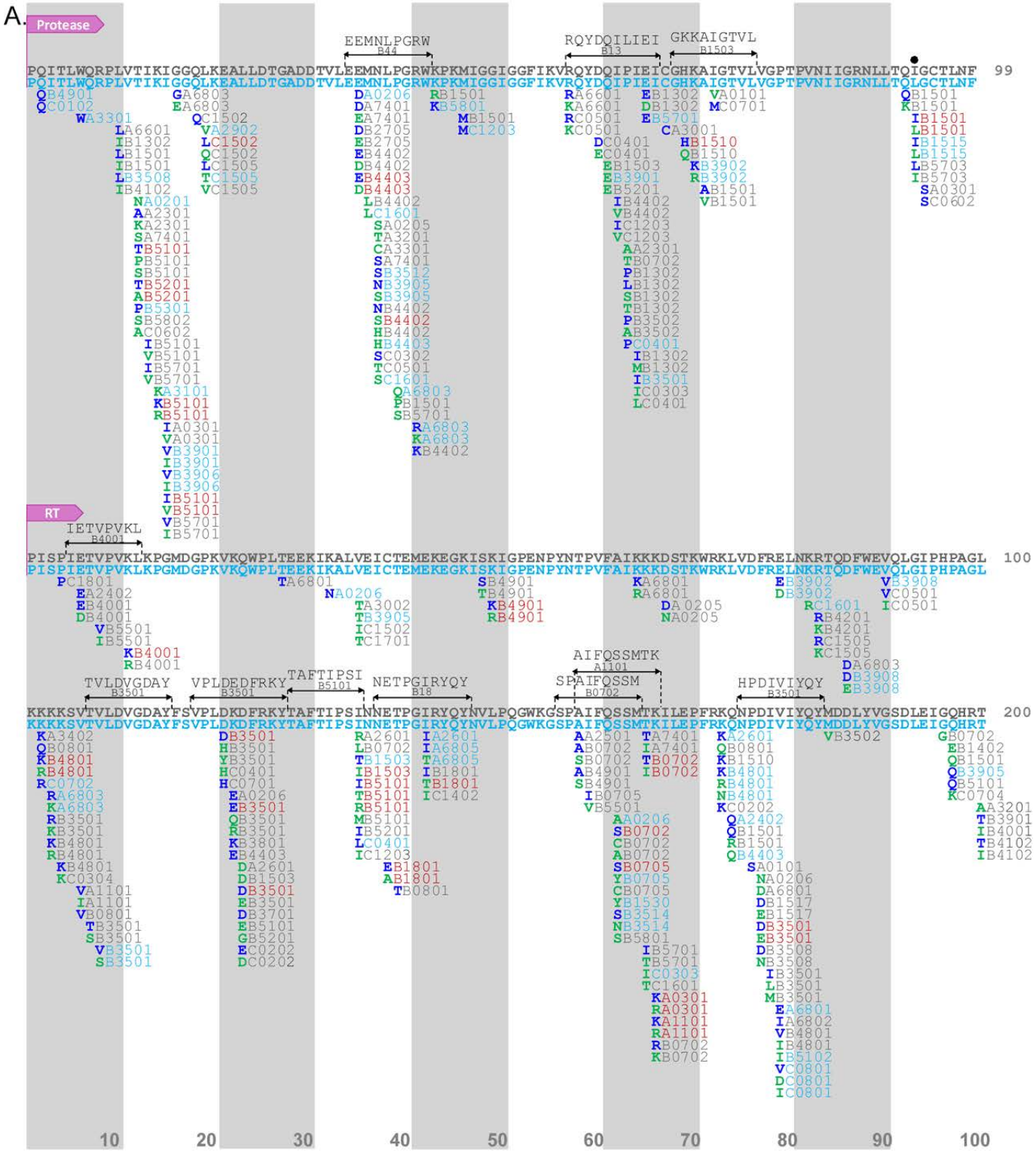


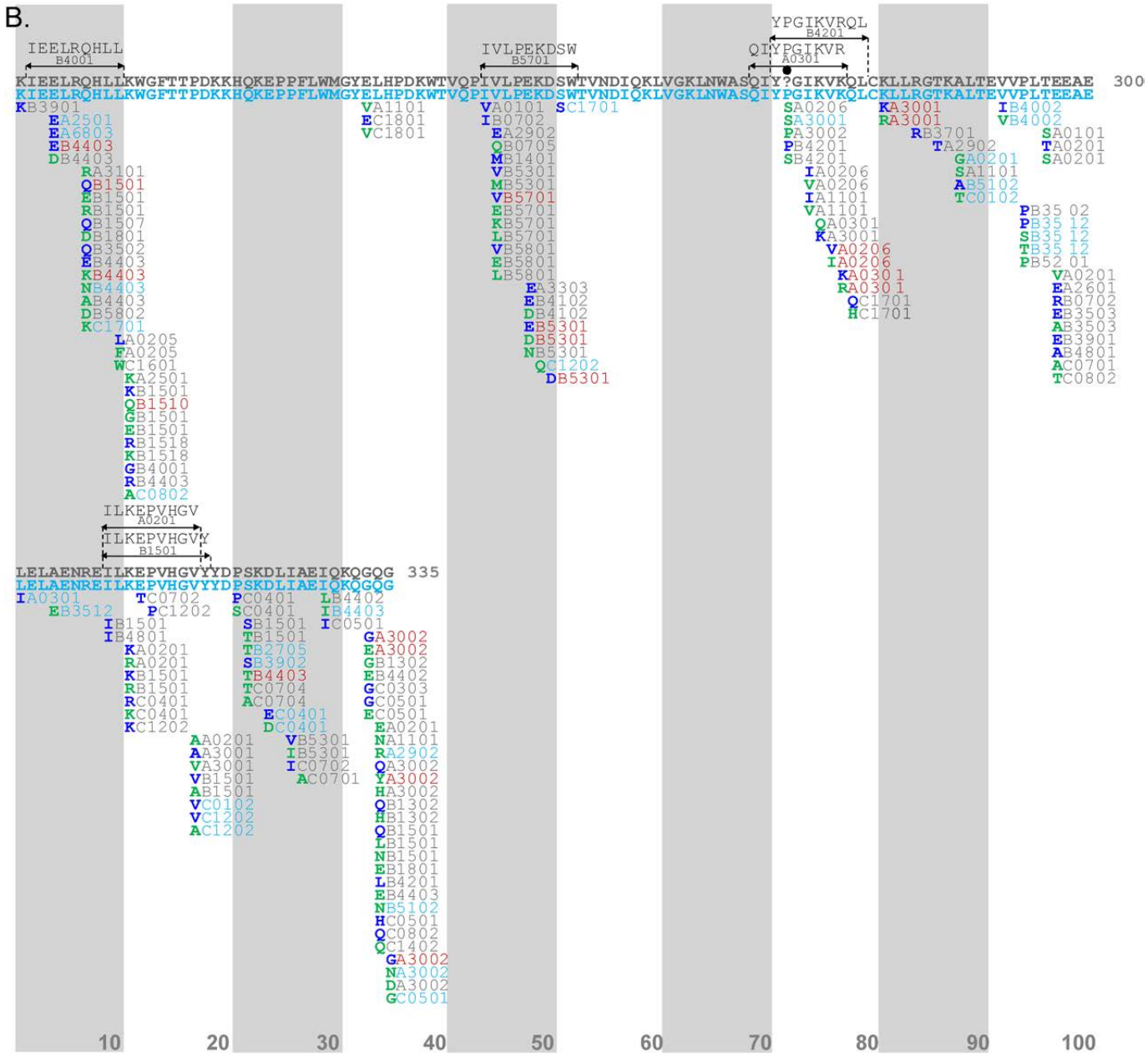


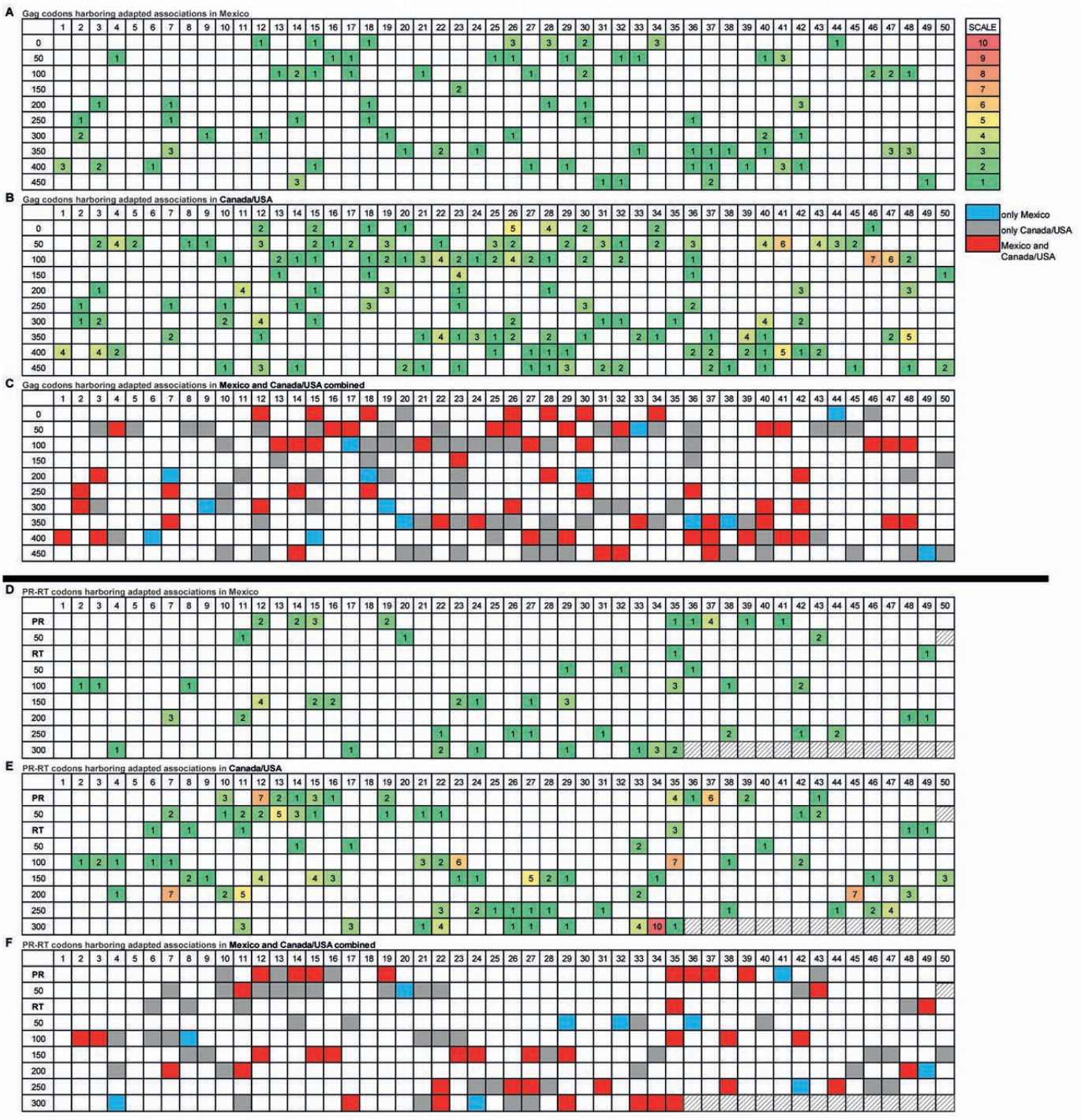








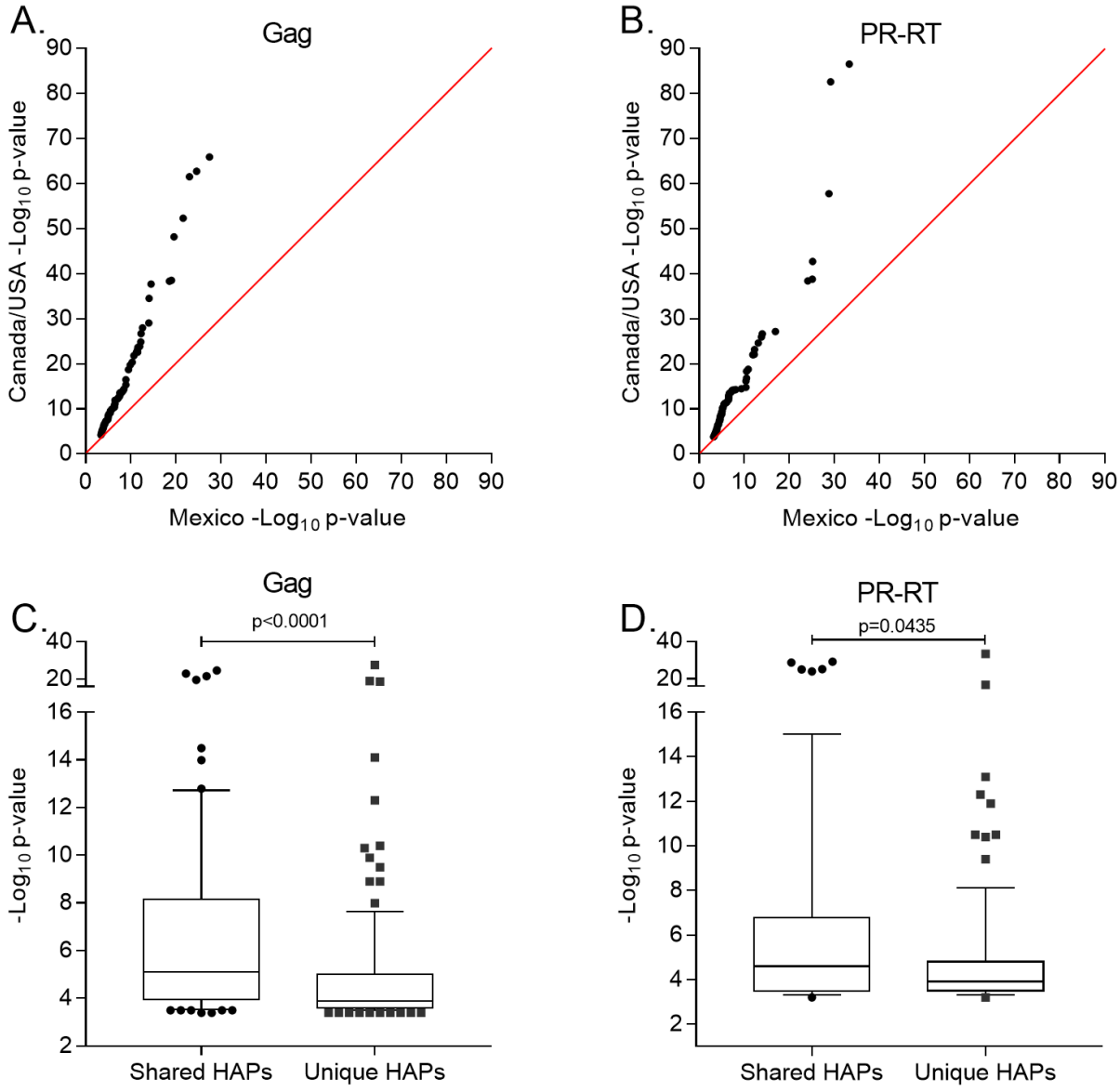




Journal of Virology

JVI

Journal of Virology



Journal of Virology
JVI
Journal of Virology