Cell Reports, Volume *7*

Supplemental Information

HIV-1 Adaptation to Antigen Processing Results in Population-Level Immune Evasion and Affects Subtype Diversification

Stefan Tenzer, Hayley Crawford, Phillip Pymm, Robert Gifford, Vattipally B. Sreenu, Mirjana Weimershaus, Tulio de Oliveira, Anne Burgevin, Jan Gerstoft, Nadja Akkad, Daniel Lunn, Lars Fugger, John Bell, Hansjörg Schild, Peter van Endert, and Astrid K.N. Iversen

“HIV-1 adaptation to antigen processing results in population-level immune evasion and affects subtype diversification”,

Tenzer et al.

Figure S1: Related to figure 2C. ERAP-trimming analyses of long epitope precursors with the protective IW9, KF11, KI8 or TW10 epitope motif. HIV subtype is noted next to the precursor sequence.

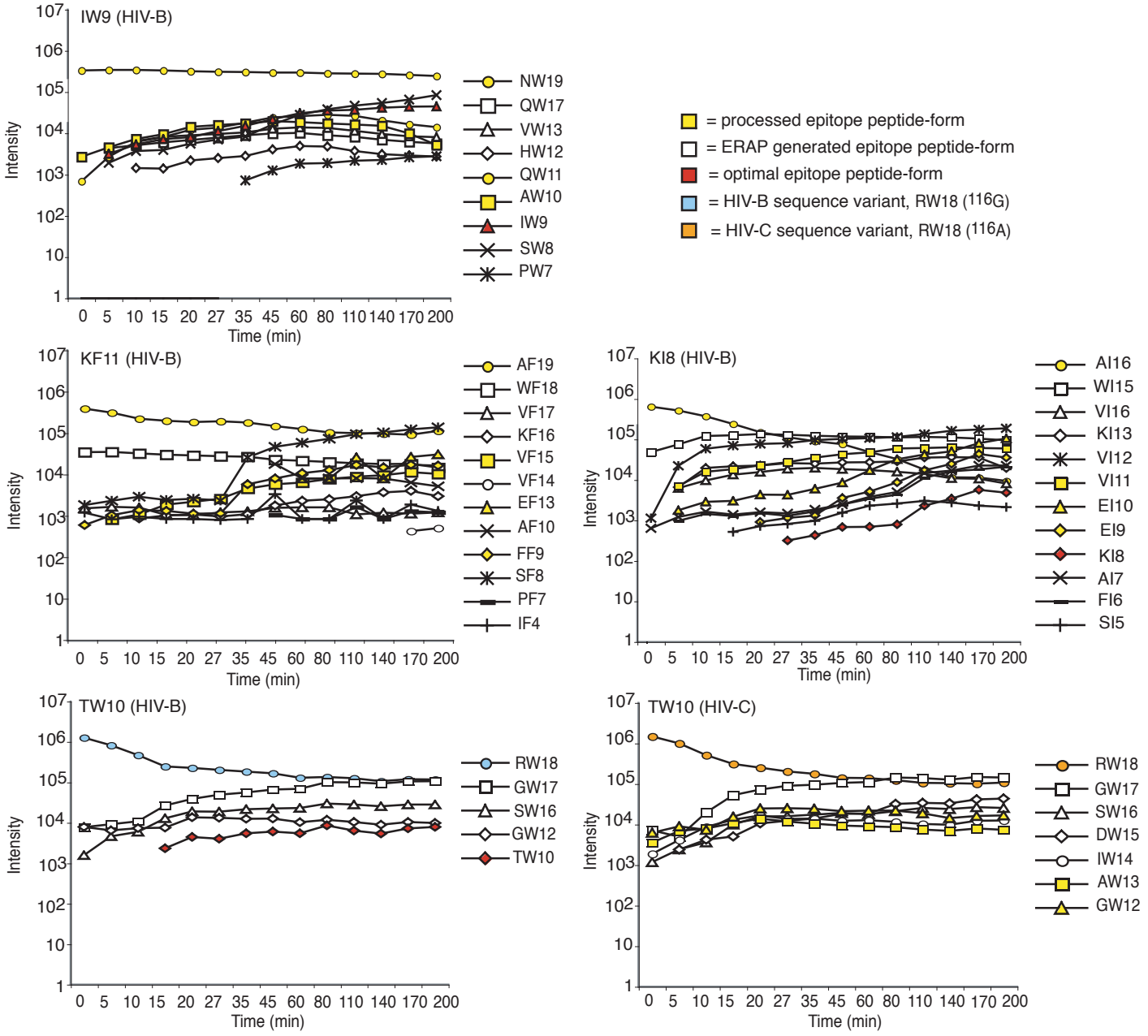
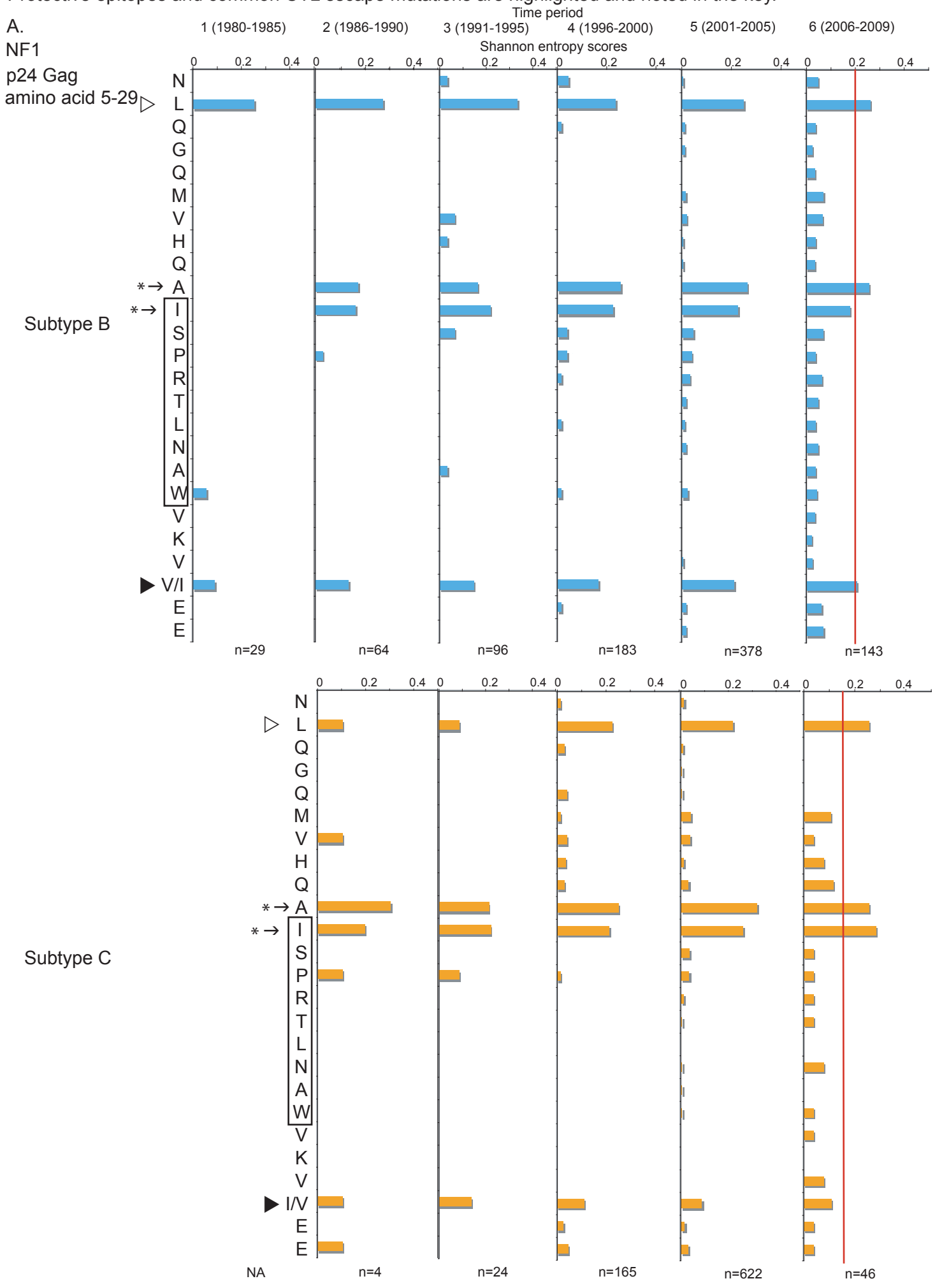


Figure S2: Shannon entropy scores of HIV subtype B and C sequences over time divided according to analyzed regions; A = NF1, B = CF1, C = NF2, D = MF2 and CF2.
Protective epitopes and common CTL escape mutations are highlighted and noted in the key.

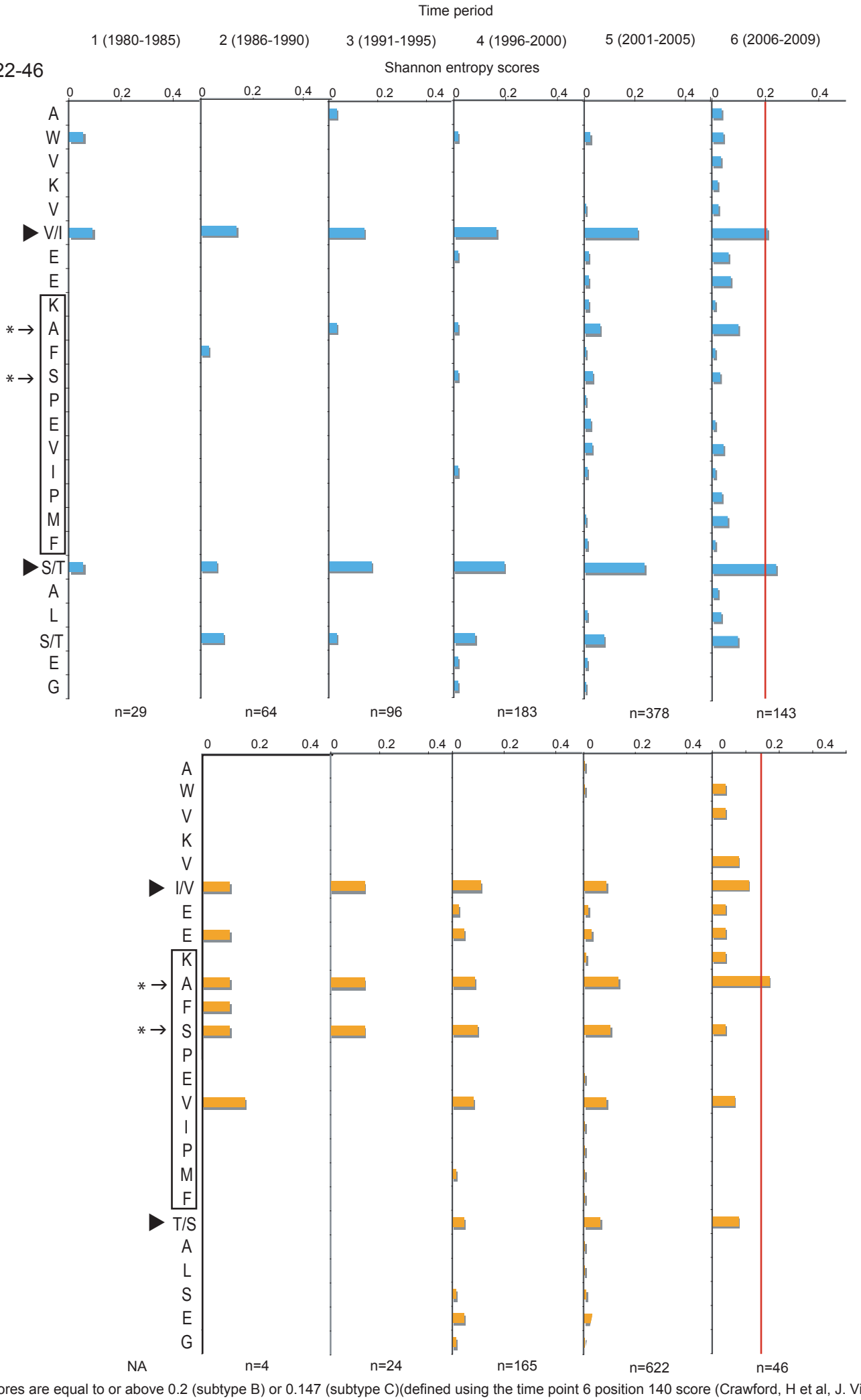


— High entropy scores are equal to or above 0.2 (subtype B) or 0.147 (subtype C)(defined using the time point 6 position 140 score (Crawford, H et al, J. Virol, 2010).

▷ Subtype specific motif between subtypes A and B/C. ▭ Box signify the HLA-A*25/B*5701/B*5703/B*63/Cw06 restricted IW9 epitope.

▶ Subtype specific motif between subtypes B and C. * → A14P is a CTL escape mutation that inhibit ERAP trimming and creation of an optimal IW9 epitope in patients with HLA-B57 and -B58 (Draenert, R, J.Exp Med, 199, 7, 905-915, 2004). I15X is an intra-epitopic escape mutation selected by various HLA-A25/B57/B58/63/Cw06 restricted IW9 respoonses.

Figure S2
B.
CF1
p24 Gag
amino acid 22-46



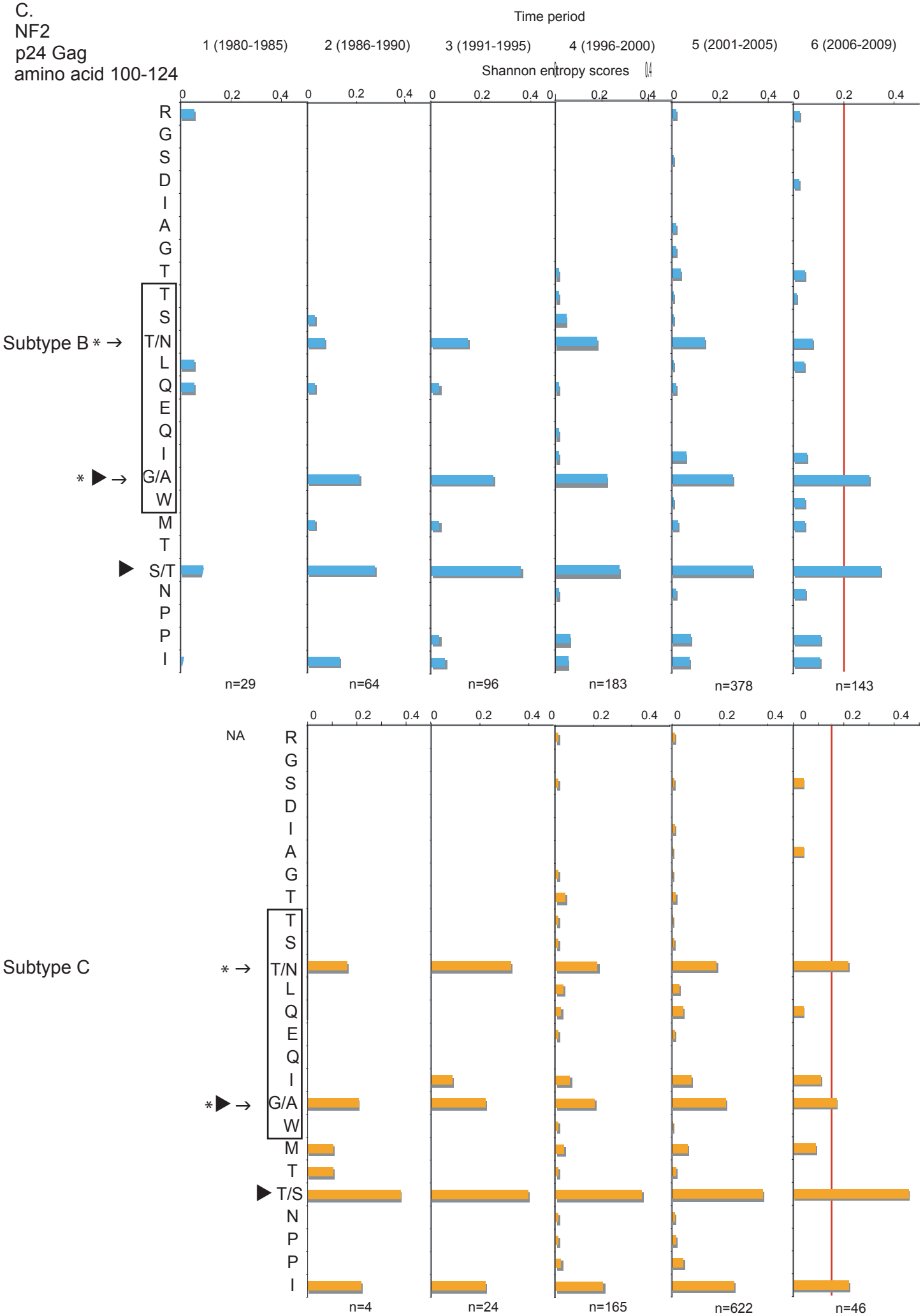
– High entropy scores are equal to or above 0.2 (subtype B) or 0.147 (subtype C)(defined using the time point 6 position 140 score (Crawford, H et al, J. Virol, 2010).

▶ Subtype specific motif between subtypes B and C.

☐ Box signify the HLA-B*5701/B*5703/B5801/B5802/B*63 restricted KF11 epitope.

* → Common intra-epitope CTL escape mutations in persons with HLA B*5703, not B*5701.

Figure S2
C.
NF2
p24 Gag
amino acid 100-124

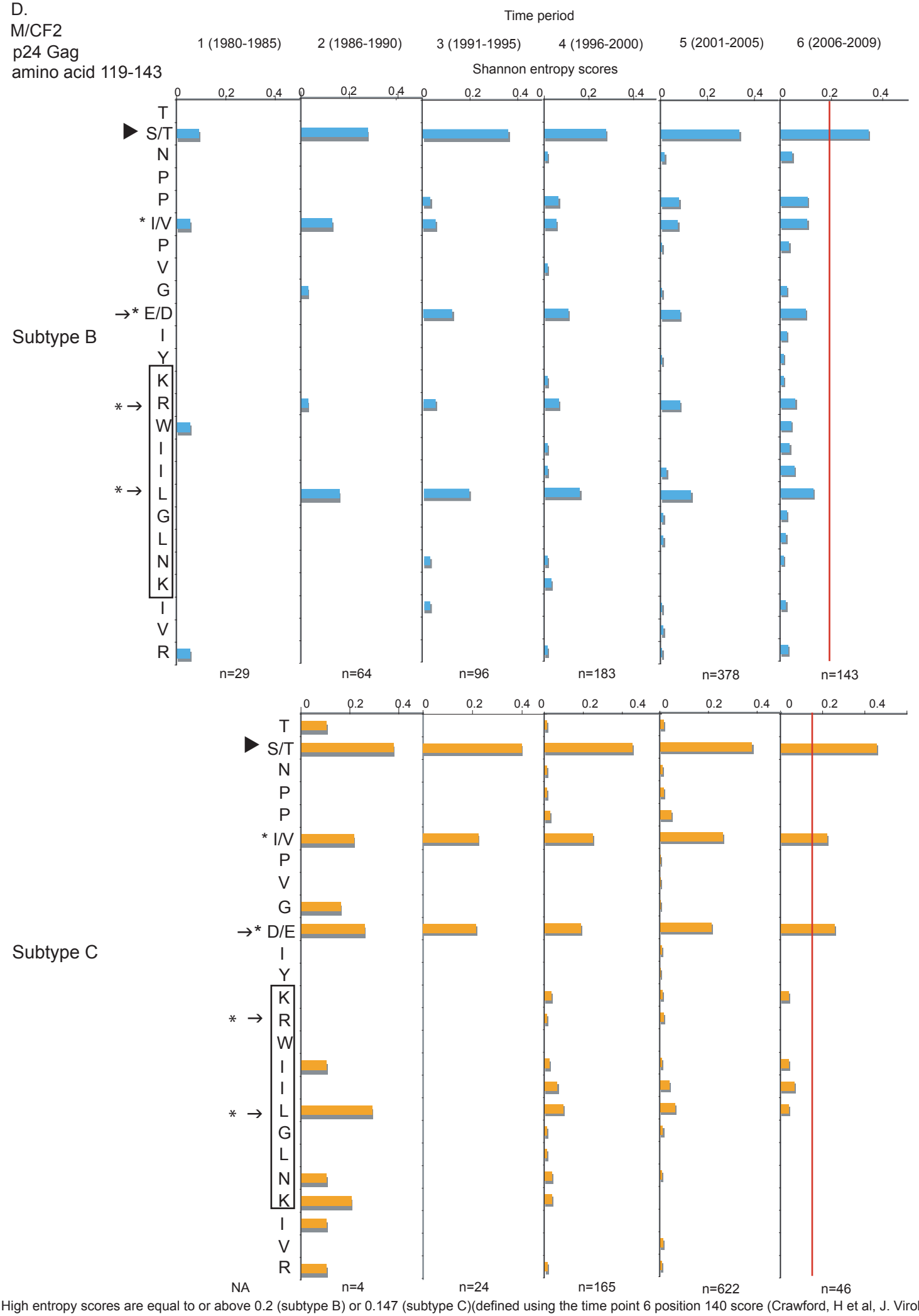


– High entropy scores are equal to or above 0.2 (subtype B) or 0.147 (subtype C)(defined using the time point 6 position 140 score (Crawford, H et al, J. Virol, 2010).

▶ Subtype specific motif between subtypes B and C. ☐ Box signify the HLA-B*57/58/B*63 restricted TW10 epitope.

* → Intra-epitope CTL escape mutation selected in patients with HLA B57 and B58 (Leslie,A, Nat. Med, 2004)

Figure S2
D.
M/CF2
p24 Gag
amino acid 119-143



– High entropy scores are equal to or above 0.2 (subtype B) or 0.147 (subtype C)(defined using the time point 6 position 140 score (Crawford, H et al, J. Virol, 2010).
▶ Subtype specific motif between subtypes B and C.
* Very common variation especially in subtype C
→* Intra-epitope CTL escape mutations in persons with HLA B*35 or B*53.
□ Box signify the optimal HLA-B*2705 restricted KK10 epitope.
* → Intra-epitope CTL escape mutations in persons with HLA B*2705.

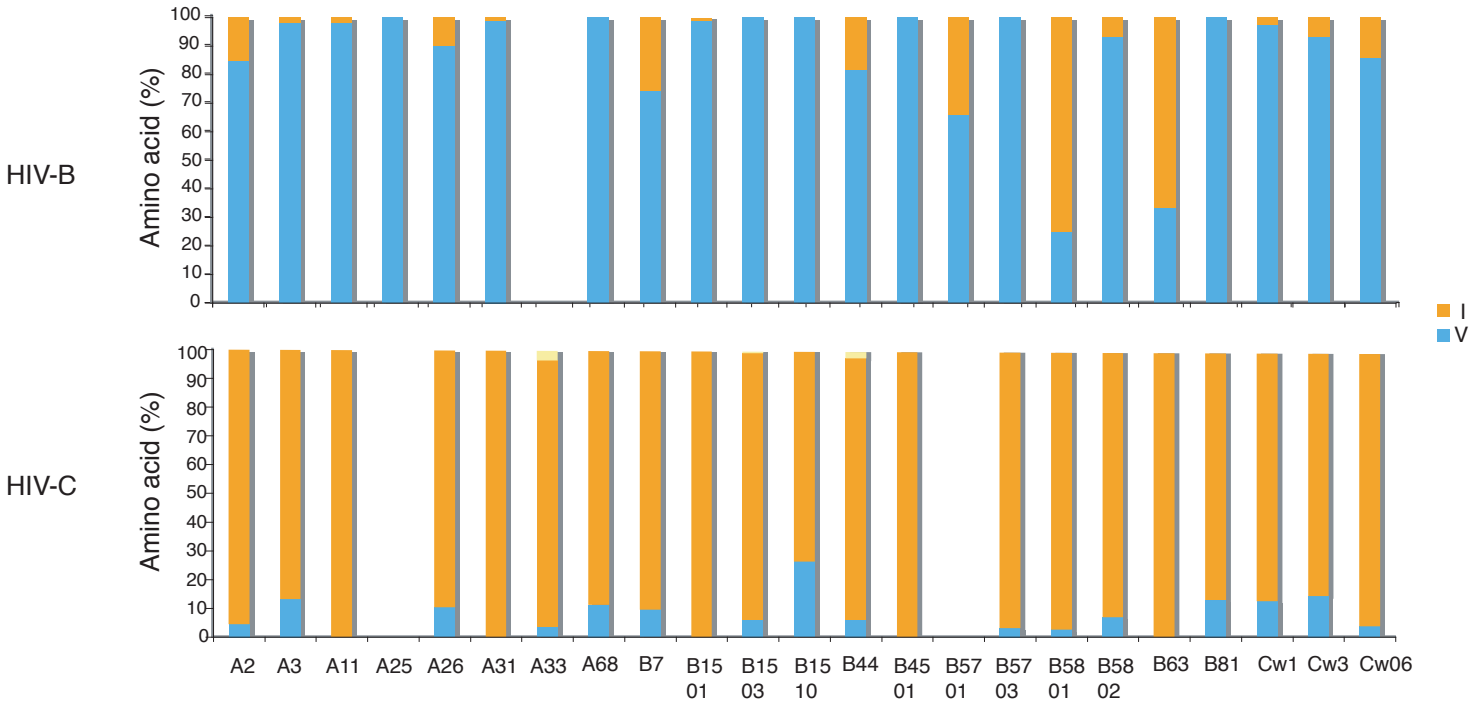
Figure S3

Amino acid polymorphisms at HIV p24 Gag positions 27 and 41 (F1) and 116, 120 and 128 (F2) in subtype B and C, respectively.

The predominant amino acid in HIV-B is blue and orange in HIV-C; HLA restriction is shown below each bar.
Fewer than three HIV-B patients carried HLA-B*1510, -B*63 and -B*81 (Table S2).

Fragment 1.

Gag position 27



Gag position 41

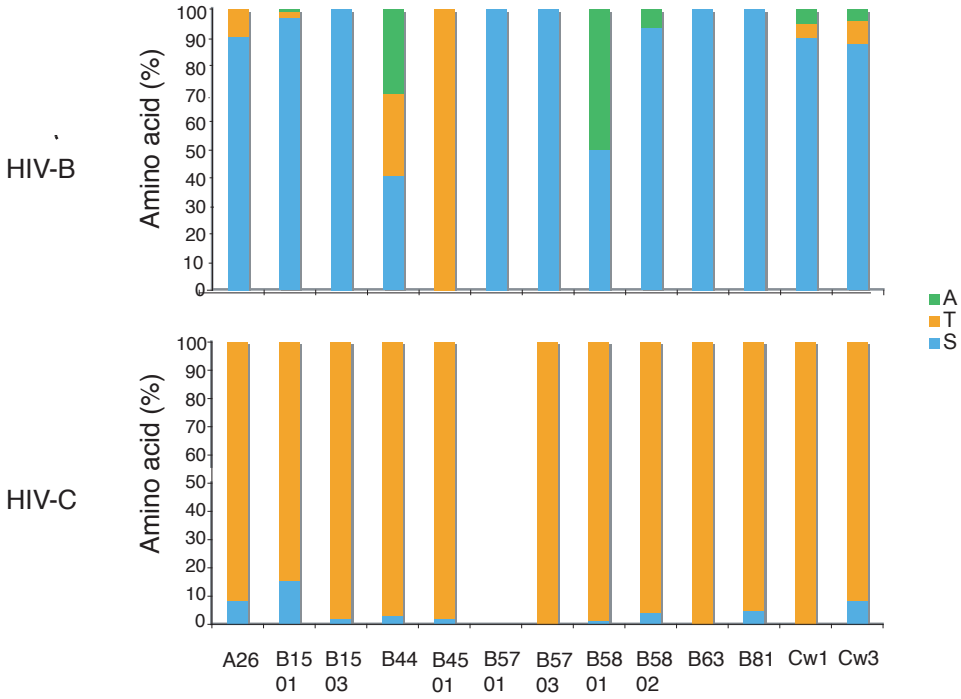


Figure S3

The predominant amino acid in HIV-B is blue and orange in HIV-C; HLA restriction is shown below each bar.

Fragment 2.

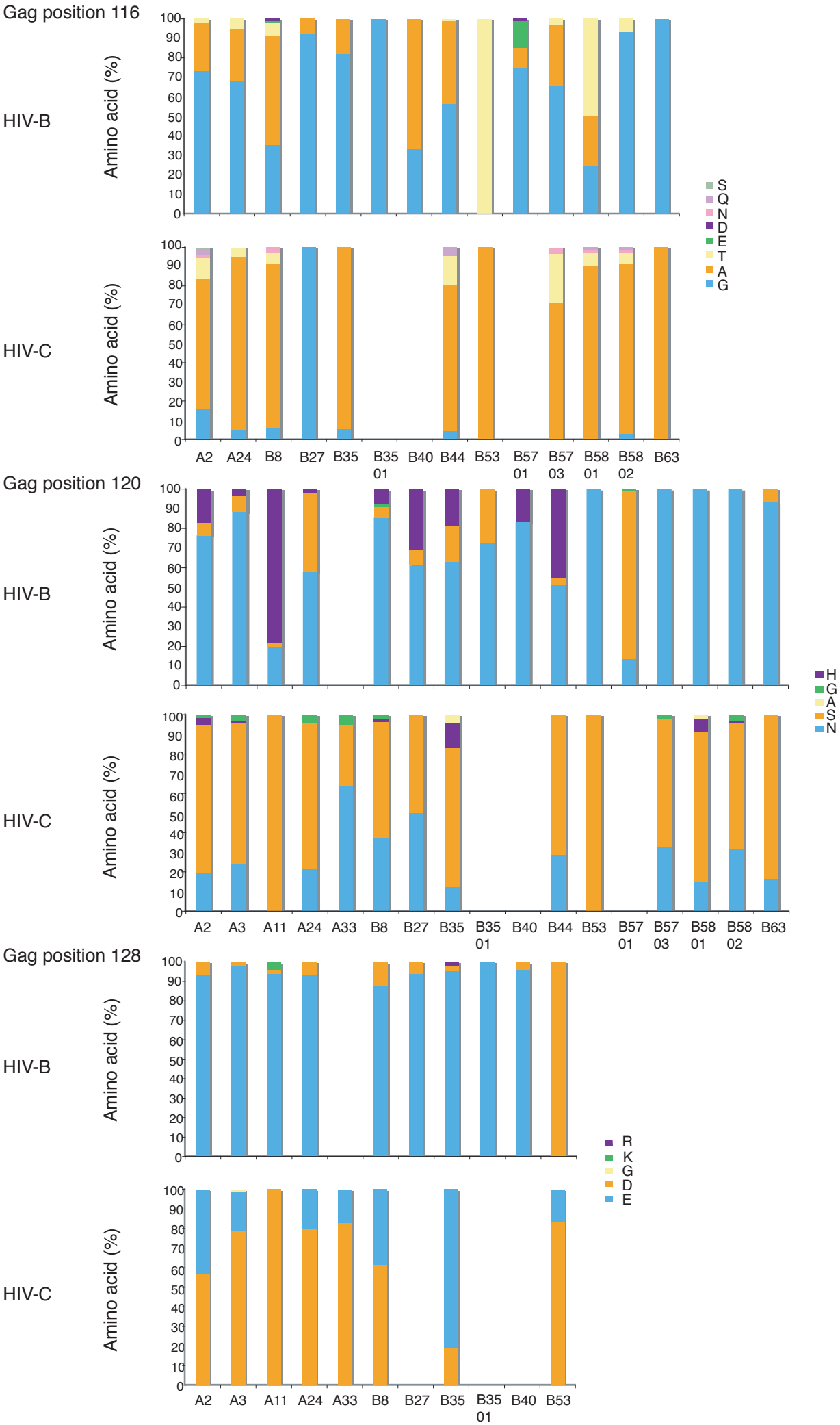


Figure S4. Related to figure 4. Subtype-specific HIV substitutions in patients with and without specific HLA alleles.

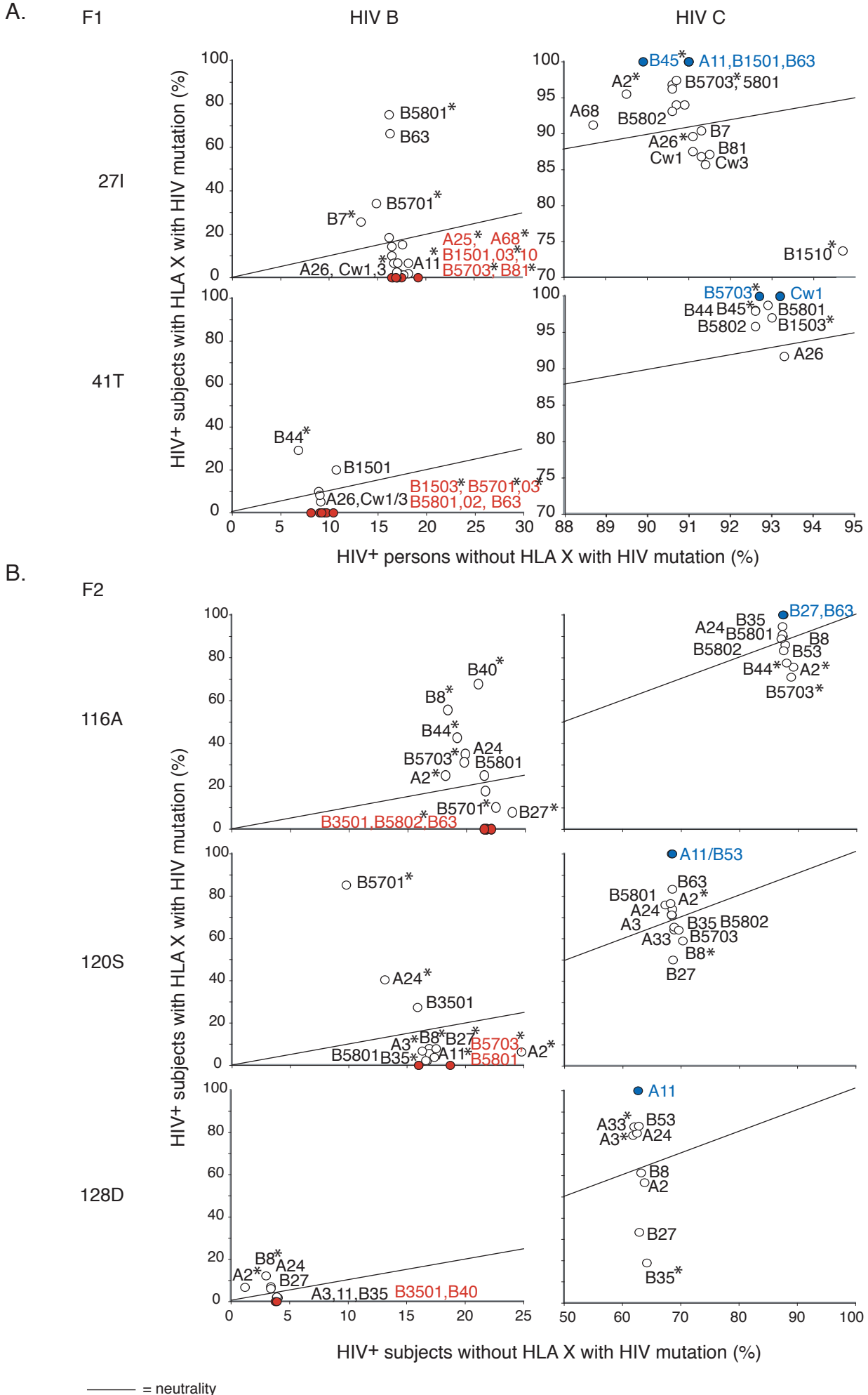


Figure S4: Subtype-specific HIV substitutions in patients with and without specific HLA alleles.

A. Graphic representation of intra-host selection for subtype-specific amino acid substitutions (^{27}V→I, ^{41}S→T) in HIV-B- and HIV-C-infected subjects carrying HLA-variants that restrict CD8-epitopes in F1 (please see details of all HLA groups in **Table S4**). Circles represent the percentage of a given amino acid (e.g. ^{27}I) in HIV from subjects with a specific HLA-allele versus its percentage in those without it; thus the data show to what extent a specific HLA-allele affects HIV evolution at a given position. The line symbolizes equal amounts of change in subjects with and without specific HLA alleles, i.e. neutrality. The specific HLA-allele (X) is shown next to the circle; blue signifies that all HIV sequences from subjects with that allele contain the substitution, red that none do. Whether or not a significant difference existed between the groups with and without given HLA variants was tested using Fisher's exact test. Less than 3 HIV-B infected patients carried HLA-B*1510, HLA-B*63 and HLA-B*81 and too few sequences were available from patients with HLA-B*5801 and HLA-B*1501 and from HIV-C infected patients with HLA-A*11, HLA-B*1501, HLA-B*53, HLA-B*63 and HLA-B*27 to test for significance. Asterisks indicate that the two-tailed p value is less than 0.05.

^{27}V→I. **HIV-B**: HLA-A*11, HLA-A*25, HLA-A*68, HLA-B*07, HLA-B*1503, HLA-B*5701, HLA-B*5703, HLA-B*5801, HLA-B*8101, HLA-Cw*01, and HLA-Cw*03 (HLA-A*03 and HLA-A*11, not highlighted due to space constraints) ($p<0.05$). **HIV-C**: HLA-A*02, HLA-A*26, HLA-B*1510, HLA-B*45, and HLA-B*5703 ($p<0.05$). The frequencies of ^{27}I in HIVB and HIV-C from patients without HLA variants that could present epitopes from F1 were 8% and 97%, respectively.

^{41}S→T. **HIV-B**: HLA-B*1503, HLA-B*44, HLA-B*5701, and HLA-B*5703 ($p<0.05$). **HIV-C**: HLA-B*1503, HLA-B*45, HLA-B*5703 ($p<0.05$). The frequencies of ^{41}T in HIVB and HIV-C from patients without HLA variants that could present epitopes from F1 were 10% and 94%, respectively.

B. Similar plot for subtype-specific amino acid substitutions in region F2 (^{116}G→A, ^{120}N→S, ^{128}E→D). Asterisks indicate $p<0.05$.

^{116}G→A. **HIV-B**: HLA-A*02, HLA-B*08, HLA-B*27, HLA-B*40, HLA-B*44, HLA-B*5701, HLA-B*5703, and HLA-B*5802 ($p<0.05$). **HIV-C**: HLA-A*02, HLA-B*44, HLA-B*5703 ($p<0.05$). The frequencies of ^{116}A in HIVB and HIV-C from patients without HLA variants that could present epitopes from F2 were 10% and 90%, respectively.

^{120}N→S. **HIV-B**: HLA-A*02, HLA-A*03, HLA-A*11, HLA-A*24, HLA-B*08, HLA-B*27, HLA-B*35, HLA-B*5701, and HLA-B*5703 ($p<0.05$). **HIV-C**: HLA-A*02, HLA-B*08 ($p<0.05$). The frequencies of ^{120}S in HIVB and HIV-C from patients without HLA variants that could present epitopes from F2 were 6% and 72%, respectively.

^{128}E→D. **HIV-B**: HLA-A*02 and HLA-B*08 ($p<0.05$). **HIV-C**: HLA-A*03, HLA-A*33, and HLA-A*35 ($p<0.05$). The frequencies of ^{128}D in HIVB and HIV-C from patients without HLA variants that could present epitopes from F2 were 1.2% and 55%, respectively.

Figure S5.Related to figure 6B. HLA frequency in BT/USW versus HIV B mutation frequency in people with/without a given HLA variant.

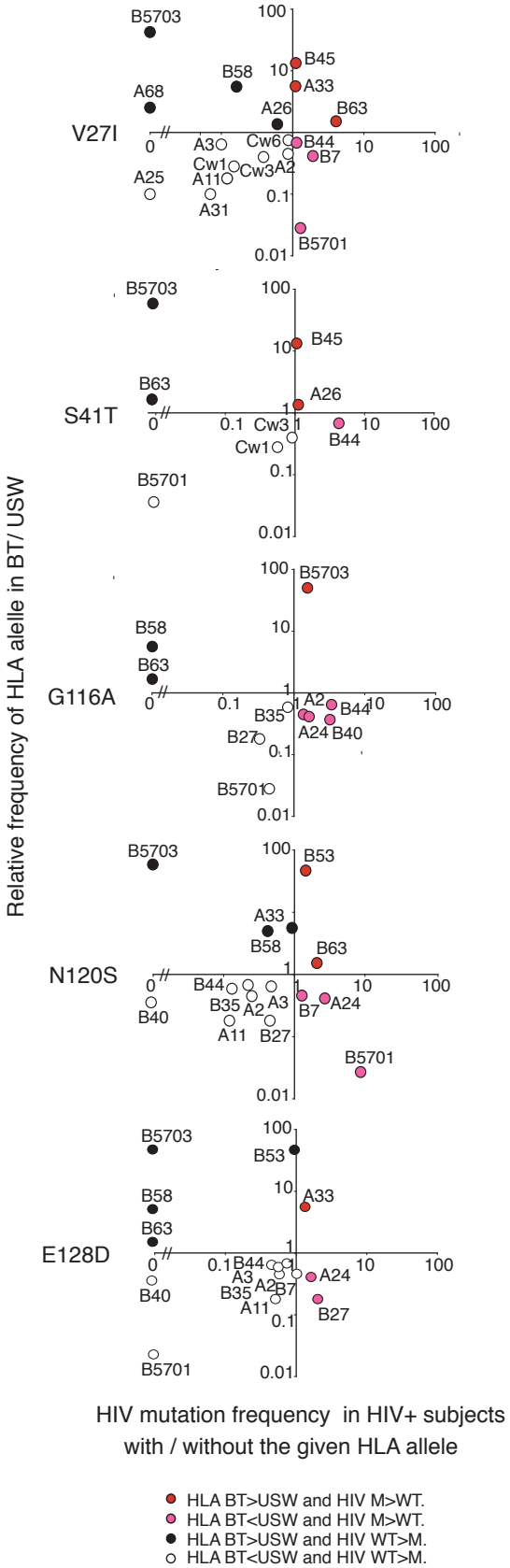


Figure S5: HIV-B adaptation to host populations' HLA profile; HLA frequency versus HIV substitution frequency.

The frequency of a presenting HLA-variant in Black Trinidad Caribbean's (BT)[2]/ US White[2] (USW) versus the HIV substitution frequency in subjects with/without that HLA-allele is shown using log scales; only data from Black Trinidadians covered all the necessary HLA-alleles, so they were used to represent all Caribbeans; likewise those from USW for all Whites. Circles represent the relative frequency of different HLAs in BT/USW and the relative selective force on HIV associated with carrying an HLA-allele/not carrying that HLA-allele for each substitution.

References:

1. *HIV Molecular Immunology*, (Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico, Los Alamos, 2009).
2. Gonzalez-Galarza, F.F., Christmas, S., Middleton, D. & Jones, A.R. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res* **39**, D913-9 (2011).

Table S1: Related to figure 1. Overview of all reported CD8-epitopes in the analyzed HIV p24 Gag regions. HLA class I restrictions are shown and the naturally-processed HIV-B and HIV-C peptide-forms of all epitopes are listed. The percentage of each peptide-form of each epitope is furthermore indicated. Attached as an Excel file.

Table S2. Related to figure 1C and 1D. Result of the multilevel analysis of the three-way interaction of fragment (NF1, CF1 or MF2&CF2), the presenting HLA allele frequency and ethnic group on epitope yield following constitutive proteasomal and immunoproteasomal digestion, respectively.

Immunoproteasomal data

Response: Epitope Yield

Coefficient	Symbol	Value	Std.Error	p-value	
Intercept	α	0.6639	0.0939	0.0000	***
Ethnic group (EG)	$\beta 1$	0.7992	0.1635	0.0006	**
HLA frequency squared (HLA sfreq)	β_2	-0.6920	0.2765	0.0130	*
Fragment MF2&CF2 (FrMF2&CF2)	β_3	0.2289	0.1149	0.0476	*
Fragment NF1 (FrNF1)	β_4	0.1471	0.1469	0.3178	
EGC:HLA sfreq	γ_{12}	-2.7595	0.4970	0.0000	***
EGC:FrCF2	γ_{13}	-0.6791	0.1923	0.0005	**
EGC:FrNF1	γ_{14}	-0.8086	0.2157	0.0002	**
HLA sfreq:FrCF2	γ_{23}	-0.8801	0.3627	0.0160	*
HLA sfreq:FrNF1	γ_{24}	-0.3853	0.3995	0.3358	
EGC:HLA sfreq:FrCF2	δ_{123}	3.0581	0.5713	0.0000	***
EGC:HLA sfreq:FrNF1	δ_{124}	2.9339	0.6121	0.0000	***

*** = p≤0.0001, ** = p<0.001, * = p<0.05

Constitutive proteasomal data

Response: Epitope Yield

Coefficient	Symbol	Value	Std.Error	p-value		**Constitutive and immunoproteasomal coefficient testing** p-values from two-tail t-test
Intercept	α	0.5693	0.0908	0		0.42
Ethnic group (EG)	$\beta 1$	0.7490	0.1598	0.0009	**	0.84
HLA frequency squared (HLA sfreq)	β_2	-0.4471	0.2703	0.0994		0.40
Fragment MF2&CF2 (FrMF2&CF2)	β_3	0.2540	0.1138	0.0265	*	0.81
Fragment NF1 (FrNF1)	β_4	0.1165	0.1455	0.4242		0.84
EGC:HLA sfreq	γ_{12}	-2.6716	0.4894	0		0.93
EGC:FrCF2	γ_{13}	-0.5635	0.1902	0.0034	*	0.57
EGC:FrNF1	γ_{14}	-0.7162	0.2134	0.0009	**	0.69
HLA sfreq:FrCF2	γ_{23}	-0.9202	0.3589	0.011	*	0.89
HLA sfreq:FrNF1	γ_{24}	-0.3723	0.3953	0.3473		0.98
EGC:HLA sfreq:FrCF2	δ_{123}	2.8037	0.5649	0		0.70
EGC:HLA sfreq:FrNF1	δ_{124}	2.6221	0.6054	0		0.64

*** = p≤0.0001, ** = p<0.001, * = p<0.05

The references for fragments and ethnic groups are fragment CF1 and Africans, respectively.

Additional p values quoted in the main text were calculated using the tabulated coefficients and standard errors in the tables along with their correlation matrix.

Table S3. Related to figure 2D. HIV positive patient data; years infected, CD4 counts, viral load and HLA profiles.

Clinical data.

Patient ID	First positive test	Sample time point			HLA class I	HLA class II	B*5701	Treatment
		ears infecte	Viral load copies/ml	CD4 count cells/µl				
1	2007	4	9593	790	A*01/23, B*44/57, Cw*04/06	DRB1*07/07,	yes	no
2	2006	5 to 6*	20	420	A*24/30, B*18/57, Cw*05/06	DRB1*07/12,	yes	no
3	1991	20	39	330	A*02/02, B*44/57, Cw*05/06	DRB1*07/13,	yes	no
4	1985	26	19	1000	A*01/02, B*15/57, Cw*03/06	DRB1*04/07,	yes	no

* The patient estimates that he was infected about one year prior to the first positive test

Table S4: Related to figure 4 and figure S4. Selection of specific HIV substitutions in Gag positions 27, 41, 116, 120 and 128 in patients infected by either HIV-B or HIV-C, who are divided according to the HLA variant they carry. The number of HIV positive subjects and the percentage and number (in brackets) of HIV sequences with and without a given substitution is shown in groups with and without a given HLA-variant. The numbers of HIV sequences with and without a given substitution in the two groups are compared using Fishers exact test. The effect of the given substitution on processing of each epitope in the region is indicated. We hypothesized that selection in HIV positive subjects with a given HLA molecule will favor viral mutations that result in decreased production of CD8-epitopes presented by that HLA variant, unless the epitope contained a CTL-escape mutation or was know to elicit very weak CTL responses. We indicate if this hypothesis is correct and add comments if appropriate; final analyses was done using a sign test. Attached as an Excel file.

Table S5. Related to figure 6C and figure 6D. Results of a GLM test of the distribution of the sums of the weighted HLA allele frequencies of HLA variants selecting for HIV B consensus and HIV-C-like amino acids in BT and USW, respectively, and results of a binomial model to test whether the ratio of HLA frequencies selecting for HIV-C-like over HIV B consensus amino acids were different between BT and USW.

Data for figure 6C

Coefficient	Estimate S	Std. Error	Pr(>|z|)	
(Intercept)	-3.0628	0.0062	< 2e-16	***
Ethnic group (Eth) USW	-0.4352	0.0088	< 2e-16	***
Position (Pos) 116	0.7488	0.0092	< 2e-16	***
Pos120	-0.2881	0.0091	< 2e-16	***
Pos128	-1.0853	0.0108	< 2e-16	***
Pos41	-0.0015	0.0114	0.895	
EthUSW:Pos116	1.0242	0.0121	< 2e-16	***
EthUSW:Pos120	-0.0694	0.0129	7.14E-08	***
EthUSW:Pos128	0.8983	0.0139	< 2e-16	***
EthUSW:Pos41	0.2214	0.0161	< 2e-16	***

*** = p≤0.0001, ** = p<0.001, * = p<0.05

The references for positions and ethnic groups are position 27 and Africans, respectively.

Additional p values quoted in the main text were calculated using the tabulated coefficients and standard errors in the tables along with their correlation matrix.

Data for figure 6D

Position	OR	L 95% CI	U 95% CI
27	0.6472	0.6361	0.6584
41	0.8075	0.7864	0.8292
116	1.8022	1.7730	1.8319
120	0.6038	0.5927	0.6150
128	1.5891	1.5558	1.6231

Supplemental Experimental Procedures

Analysis of peptide digests by mass spectrometry.

We performed capillary liquid chromatography of the peptide digests using a Waters NanoAcquity UPLC system with a Waters NanoEase BEH-C18, 75 micron × 15 cm reverse-phase column as described in (Tenzer et al., 2009). For the mass spectrometry analysis of the peptide fragments, we used a Waters Q-Tof Premier in positive Vmode equipped with a nano-ESI source after calibration with a [Glu1]-fibrinopeptide solution (500 fmol/µl at 300 nl/min) through the reference sprayer of the NanoLockSpray source. The instrument was run in MS^E-mode for fragment identification and simultaneous relative peptide fragment quantification. Each sample was analyzed in triplicate. We used the ProteinLynx Global Server (PLGS) version 2.2 for the processing, fragment identification and quantification of the LC-MS^E data. The mass error tolerance values were typically less than 5 ppm. For data analyses, we used mass spectrometric fragment intensity as a surrogate marker for quantity, because we previously observed a highly significant correlation between these two parameters (Tenzer et al., 2009).

TAP-peptide binding assays

TAP-peptide binding of epitope-precursors and optimal epitopes was determined as described elsewhere (Burgevin et al., 2008; Tenzer et al., 2009). Briefly, the relative TAP-affinities of the test peptides were measured in a competitive binding assay as the concentration required to reduce the specific binding of the high-affinity reporter peptide RL9 (RRYNACTEL) by one-half (IC_{50}) to Sf9 insect cell microsomes over-expressing human TAP1-TAP2 complexes. The reporter peptide was labeled by coupling a fluorescein-isothiocyanate (FITC)

group to a Cys substituted for the Ser at position 6 and was used at 20 nM in this assay. The specific binding was measured by fluorescence polarization using a Mithras LB940 reader (Berthold) and MikroWin 2000 software (Mikrotek). Each peptide was tested in at least two independent experiments using competitor concentrations ranging from 1 to 10,000-fold molar excess relative to the FITC-conjugated reporter peptide. The unlabeled, non-substituted reporter peptide, RRYNASTEL, was included in each assay for normalization.

***In vitro* peptide digestions with ERAP enzymes.**

To assess the ERAP trimming of of epitope-precursor peptide mixtures, 10 µmol of each peptide was incubated with 12.5 µg of human ERAP1/2 complexes as described previously (Tenzer et al., 2009). The digestions were performed at 37°C in PBS supplemented with 50 µM $ZnCl_2$, 20 mM NaCl and 1 mM DTT. Aliquots were removed for analysis at the indicated time points, and the reactions were stopped by the addition of formic acid to a final concentration of 2% and snap-freezing.

Epitope designation and HLA analyses

The HLA phenotype and allele frequencies were obtained from the Allele Frequency Database (Gonzalez-Galarza et al., 2011) and sources found in the literature (Assane et al., 2010; Novitsky et al., 2001). When more than one HLA frequency dataset was available from one country, we used weighted mean values unless otherwise specified. For the analyses shown in figure 1, we used HLA frequency information from two UK cities (UK1 = Leeds (n=5024) and UK2 = Sheffield (n=4755)) and from cities in the US (US1 = Los Angeles, Seattle and Detroit (n=1070), US2 = Bethesda (n=307))(Gonzalez-Galarza et al., 2011). We used two White cohorts from the UK and two from the US because most HIV B sequences in

the HIV database derive from the UK or the US; the HLA distributions in these countries would therefore be likely to add a greater footprint on HIV B evolution that White HLA frequencies from other countries. We did, however, also include HLA distributions from Germany (N = 11407) and France (n = 130). For the HIV C analyses, we used HLA frequency data from both the Zulu (n = 100) and the Xhosa (n = 50) tribes in South Africa to cover the predominant HLA variability because most HIV C sequences derive from this country (Gonzalez-Galarza et al., 2011; 2009). We did, however, also include HLA data from other African countries in which HIV C infection predominates; Zambia (n = 256) (Gonzalez-Galarza et al., 2011), Zimbabwe (n = 108) (Gonzalez-Galarza et al., 2011), Botswana (n = 161) (Novitsky et al., 2001) and Mozambique (n = 250) (Assane et al., 2010). Overall, White HLA frequencies were similar to each other regardless of country of origin, while the differences in HLA frequencies in the Africa populations were huge. The Caribbean HIV-B data originate from Barbados (BB), Trinidad and Tobago (TT), The Dominican Republic (DO), Haiti (HT) and Jamaica (JM). The HIV-B sequence contribution from the various Caribbean countries differ somewhat between 2001-2005 (BB, 75%; TT, 3%; DO, 4%; HT, 10% and JM, 7%) and 2006-2009 (BB, 42%; TT, 0%; DO, 0%; HT, 29% and JM, 29%). However, because 1) the HLA profiles of Black Caribbean are highly related regardless of which Caribbean country they originate from (Gonzalez-Galarza et al., 2011)and 2) Black Caribbean compose over 90% of the population in the three countries that contributes most of the sequences (BB, HT and JM, combined for 2001-2005: 92% and combined for 2006-2009: 100%)(www.indexmundi.com/factbook), it is highly likely that very similar selective pressures are imposed on HIV-B by the HLA distribution in these Caribbean countries. Because of this, the variation in country of origin of the Caribbean HIV-B sequences is unlikely to confound the conclusion of temporal population evolution. We

designated peptide fragments as CD8 epitopes based on the HIV immunology database epitope summary tables (2009). Because studies disagree as to whether HLA Cw*03 can bind the VL10 epitope (Zappacosta et al., 1997; Zarling and Lee, 1998), we did not include a selection pressure from HLA Cw*03 in our analyses of this epitope.

Statistical analyses.

We used multilevel modeling to examine the effects of HLA frequency, HIV region (NF1, CF1, MF/CF2) and ethnic group (Africans, Whites) on epitope yield (EY) (shown in **Figure 1 D**). To achieve normality and homoscedasticity in the residuals, it was necessary to transform the response variable by raising the epitope yield data to the power 0.2; this was the optimal value produced by the Box-Cox method (Box and Cox, 1964). Similarly, we used the square root of the HLA frequency data to eliminate right skew. The explanatory variable x_1 took the value 0 for Africans, 1 for Whites, x_2 was the square root of the HLA frequency, x_3 took the value 0 if the region was NF1 and 1 if the region was CF2, x_4 took the value 0 if the region was CF2 and 1 if the region was NF1. Thus CF1 was defined by $x_3 = 0$ and $x_4 = 0$. The model was defined as:

$$y_{rs}^{0.2} = \alpha + \sum_{i=1}^{4} \beta_i x_{i,rs} + \sum_{i<j} \gamma_{ij} x_{i,rs} x_{j,rs} + \sum_{i<j<k} \delta_{ijk} x_{i,rs} x_{j,rs} x_{k,rs} + \nu_s + \varepsilon_{rs},$$

where the coefficients are as referred to in **Table 2**. ν_s is a normally distributed random variable with zero mean which represents the departure of the s^{th} population's intercept from the overall value and ε_{rs} is a normally distributed random variable with zero mean which represents the variation from epitope to epitope within a population group; in terms of multilevel modelling ε_{rs} is a level one residual and ν_s is a level 2 residual.

For both Black Trinidadians (BT) and US Whites (USW), we summed the weighted allele frequencies of HLA variants selecting for HIV B consensus (wild type) and HIV-C-like (mutant) amino acids, respectively, at each of the five subtype-specific positions (**Fig. 6C**). A binomial generalised linear model (GLM) (McCullagh and Nelder, 1989) was used to examine the relationship between the sums of HLA frequencies of mutant and wild type and the different ethic groups in each subtype-specific position; the response variable was the sum as a fraction of the total with population size being used as a weighting factor. The model was fitted using the canonical logit link function. We examined whether the distributions of the sums were similar at each position within and between the two populations (i.e. the distribution of the light and dark orange (BT) and the light and dark blue (USW) bar-heights, (**Fig. 6C**). We also fitted a binomial model to test whether the ratio of HLA frequencies selecting for HIV-C-like over HIV B consensus amino acids were different between BT and USW (**Fig. 6D** and **table S5**, odds ratios and 95% CI). We next used a binomial model to fit the ratios of HLA frequencies selecting for HIV-C-like over HIV B consensus amino acids to the frequency of an HIV-C-like amino acid at positions 27, 41, 116, 120 and 128 in USW and BT, respectively (**Fig. 6E**). Lastly, a binomial model was used to examine the association between population HLA B*2705 frequencies and the percentage of ^{120}S and ^{120}N in HIV from infected individuals in each population (**Figure 6F**).

All of the models were fitted using the R statistical package (R version 3.0.0 (2013-04-03)); for all of the fitted models the diagnostics were satisfactory. Additional statistical methods include Fisher's exact test to examine changes in the number of HIV-C-like amino acids in HIV-B from different populations at two specific times (**Fig. 6A**), t-tests for comparison of coefficients of the modeling of immunoproteasomal and constitutive proteasomal epitope

yields (shown in **Fig. 1D** and **Table S2**) and Fisher's exact test and a non-parametric sign test to examine the relationship between epitope production and HIV selection in groups of patients with the same HLA variant (**Table S4**). These analyses were performed using GraphPad Prism 5 (GraphPad Software Inc.), R, Microsoft Excel 2007 (Microsoft Corp.) and Stat Trek (Stattrek.com) software (**Table S4**). A value of $p < 0.05$ was considered significant. Graphing was done using GraphPad Prism 5 (GraphPad Software Inc.), R and Microsoft Excel 2007.

References:

Assane, A. A., Fabricio-Silva, G. M., Cardoso-Oliveira, J., Mabunda, N. E., Sousa, A. M., Jani, I. V., Ferreira, O. C., Jr., and Porto, L. C. (2010). Human leukocyte antigen-A, -B, and -DRB1 allele and haplotype frequencies in the Mozambican population: a blood donor-based population study. Hum Immunol *71*, 1027-1032.

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformation. Journal of the Royal Statistical Society, *Series B*, 211-252.

Burgevin, A., Saveanu, L., Kim, Y., Barilleau, E., Kotturi, M., Sette, A., van Endert, P., and Peters, B. (2008). A detailed analysis of the murine TAP transporter substrate specificity. PLoS One *3*, e2402.

Gonzalez-Galarza, F. F., Christmas, S., Middleton, D., and Jones, A. R. (2011). Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res *39*, D913-919.

McCullough, P., and Nelder, J. A. (1989). Generalized Linear Models: Chapman & Hall).
Novitsky, V., Flores-Villanueva, P. O., Chigwedere, P., Gaolekwe, S., Bussman, H., Sebetso, G., Marlink, R., Yunis, E. J., and Essex, M. (2001). Identification of most frequent HLA class I antigen specificities in Botswana: relevance for HIV vaccine design. Hum Immunol *62*, 146-156.

Tenzer, S., Wee, E., Burgevin, A., Stewart-Jones, G., Friis, L., Lamberth, K., Chang, C. H., Harndahl, M., Weimershaus, M., Gerstoft, J.*, et al.* (2009). Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. Nat Immunol *10*, 636-646.

www.hiv.lanl.gov (2009). HIV Molecular Immunology (Los Alamos: Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.).

Zappacosta, F., Borrego, F., Brooks, A. G., Parker, K. C., and Coligan, J. E. (1997). Peptides isolated from HLA-Cw*0304 confer different degrees of protection from natural killer cell-mediated lysis. Proc Natl Acad Sci U S A *94*, 6313-6318.

Zarling, A. L., and Lee, D. R. (1998). Conversion of a human immunodeficiency virus cytotoxic T lymphocyte epitope into a high affinity HLA-Cw3 ligand. Hum Immunol *59*, 472-482.