An evolutionary-based approach to quantify the genetic barrier to drug resistance in fast-evolving viruses: an application to HIV-1 subtypes and integrase inhibitors

Kristof Theys[1,*], Pieter Libin[1,2], Kristel Van Laethem[1], Ana B Abecasis[3]

1 Department of Microbiology and Immunology, Laboratory Clinical and Evolutionary Virology, Rega Institute for Medical Research - KU Leuven, 3000, Leuven, Belgium
2 Artifical Intelligence lab, Department of computer science, Vrije Universiteit Brussel, 1040, Brussels, Belgium
3 Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal.

* ktheyss@gmail.com

Abstract

Viral pathogens causing global disease burdens are often characterised by high rates of evolutionary changes, facilitating escape from therapeutic or immune selective pressure. Extensive viral diversity at baseline can shorten the time to resistance emergence and alter mutational pathways, but the impact of genotypic background on the genetic barrier can be difficult to capture, in particular for antivirals in experimental stages, recently approved or expanded into new settings. We developed an evolutionary-based counting method to quantify the population genetic potential to resistance and assess differences between populations. We demonstrate its applicability to HIV-1 integrase inhibitors, as their increasing use globally contrasts with limited availability of non-B subtype resistant sequences and corresponding knowledge gap on drug resistance. A large sequence dataset encompassing most prevailing subtypes and resistance mutations of first- and second-generation inhibitors were investigated. A varying genetic potential for resistance across HIV-1 subtypes was detected for 15 mutations at 12 positions, with notably 140S in subtype B, while 140C was discarded to vary across subtypes. An additional analysis for HIV-1 reverse transcriptase inhibitors identified a higher potential for 65R in subtype C, on the basis of a differential codon usage not reported before. The evolutionary interpretation of genomic differences for antiviral treatment remains challenging. Our framework advances existing counting methods with an increased sensitivity that identified novel subtype dependencies as well as rejected previous statements. Future applications include novel HIV-1 drug classes as well as other viral pathogens.

1 Introduction

The advent of antiviral treatment resulted globally in significant health gains and averted deaths by preventing viral infections and improving disease outcomes. Major human pathogens targeted by antivirals are fast-evolving, genetically diverse viruses [13, 19, 20], allowing for rapid adaptation through the emergence of

resistance-conferring mutations. A key concept in understanding the dynamics of resistance development is the genetic barrier, which ultimately quantifies the evolutionary time to viral escape from drug selective pressure [4, 14, 16, 39]. The virus make-up at baseline can shorten the evolutionary distance to antiviral resistance and, together with treatment- and patient-related factors, alter the mode and tempo of resistance emergence [20, 24]. The imprint of genotypic background on viral escape dynamics can be however difficult to capture. In particular for newer antivirals when resistance knowledge is initially limited to *in vitro* selection experiments informed on genetically-limited backbones and inferred from observations in well-controlled clinical studies or patient cohorts underrepresenting the full spectrum of viral diversity.

A notable example of success is the evolution of the management of Human Immunodeficiency Virus type-1 (HIV-1) infection in the last three decades, with antivirals available from multiple drug classes that drastically reduced morbidity and mortality related to HIV-1. The recent class of integrase strand transfer inhibitors (INSTIs), directed against the integrase enzyme by blocking the strand transfer step of viral DNA integration, has considerably expanded treatment options and reduced the probability of virological failure, predominantly in resource-rich settings where INSTI use is widespread. To date, first-generation INSTIs raltegravir (RTG) and elvitegravir (EVG) and second-generation INSTIs dolutegravir (DTG) and bictegravir (BIC) are approved for HIV-1 treatment and a preferred option for first-line regimens [27, 36]. INSTIs are anticipated to become also widespread in low- and middle-income countries (LMICs), and the use of INSTIs has been shown to be cost-effective [30], although rates of acquired drug resistance are increasing in these settings. An impact of HIV-1 genetic diversity, classified into groups and subtypes, on resistance development has been well documented for the historical classes of protease and reverse transcriptase (RT) inhibitors, mainly resulting from preferential codon usage across subtypes [1, 40, 42]. The evolutionary mechanisms underlying viral escape from INSTI selective pressure, despite the identification of mutational pathways, are still being unfold in particular for non-B subtypes which are prevalent in LMICs [12]. A low observed complexity of INSTI resistance profiles has consequences for the relevance of HIV-1 diversity for the genetic barrier to resistance [6, 17].

Extensive subtype mappings of integrase diversity in treatment-naive patients revealed amino acid polymorphisms at resistance-associated positions [20, 25, 35, 45]. Apart from a documented subtype B INSTI resistance pathway attributed to differential codon usage at position 140 [17, 23, 32, 38], the role of baseline nucleotide variation on INSTI resistance development has been less systematically investigated. The probabilistic models of resistance evolution that previously quantified the genetic barrier for the historical drug classes require sequence data of various subtypes from treatment-experienced patients [4, 14, 39], which are limited to date in the context of (second-generation) INSTIs [34, 35]. Alternatively, the genetic barrier can be estimated by the number and type of required nucleotide substitutions to evolve from wild-type virus to a resistance mutation [43]. While this modality was applied to HIV-1 integrase before [23, 32], these studies lacked resistance mutations of second-generation INSTIs, ignored subtype-specific effects or analyzed limited subtype distributions and relied on arbitrary substitution cost assignments.

We present a novel, optimized and evolutionary-based methodology to quantify the genetic potential to resistance in fast-evolving viral pathogens, facilitating a priori identification of an impact of varying genetic background on the emergence of resistance mutations. We demonstrate in this study our approach by the application to HIV-1 INSTIs. With the roll-out of INSTIs in LMICs, an increasing introduction of non-B subtypes in high-income countries [15, 18, 26, 28] and novel INSTIs anticipated, this study addressed the need for an in-depth understanding of the dynamics underlying

INSTI resistance development across HIV-1 subtypes. We derived an optimized genetic barrier score based on empirical substitution costs, which contrasts with previous approaches that used arbitrary costs differing from *in vivo* estimates [46]. Furthermore, we advanced existing studies by considering a population-based estimate, where we take into account the most prevailing subtypes globally and calculate subtype-tailored summary scores. The framework can be easily applied to novel HIV-1 drug classes (e.g. attachment or maturation inhibitors) and extended to pathogens (e.g. Zika, Influenza, RSV, Herpes, Varicella, Dengue) which are currently targeted by antiviral drug development efforts.

2 Methods

2.1 HIV-1 dataset and drug resistance mutations

A dataset of HIV-1 integrase sequences from INSTI-naive patients was obtained from the Stanford HIV drug resistance database [35], aligned codon-correct using VIRULIGN [22] and classified by REGA subtyping tool v3 [3, 21, 31]. Only 1 sequence per patient and sequences without a stop codon were retained. Additionally, sequences with mutations 143C/H/R, 148H/K/R or 155H/S were considered a proxy for incorrect treatment status and excluded, given that transmission of INSTI resistance is infrequently reported to date. Integrase positions and mutations important for INSTI resistance were defined by an association with reduced susceptibility and virological response [6], or inclusion in HIV-1 drug resistance interpretation systems (Rega v10, HIVdb v8.7, ANRS v29) [5, 9, 44]. A total of 41 codon positions and 77 amino acid mutations implicated in viral escape from first- or second-generation INSTIs were investigated. In addition, **Supplementary Material** presents the same approach applied to quantify the evolutionary potential to resistance of the RT enzyme.

2.2 Codon diversity

We define the genetic barrier in terms of a particular sub-epidemic, e.g. with respect to all viruses that belong to a HIV-1 subtype in our study, and of a resistance mutation defined by an amino acid and a given position. An overview of the procedure to calculate the genetic barrier is given in Figure 1, using resistance mutation 140S as example. At every position, variability at the nucleotide level was assessed by the prevalence of nucleotide triplets (codon) in each sub-epidemic (Figure 1A). Codons with ambiguities consisting of >2 bases per nucleotide position or of two or more ambiguities per codon were not considered. A nucleotide ambiguity of exactly 2 bases was resolved in the two corresponding triplets, each counting for one half to their respective frequencies. All triplets with a prevalence above 1% in at least one sub-epidemic were retained, and triplets with a prevalence above 50% were defined as predominant for that sub-epidemic. The prevalence of resistance mutations and codon entropy were inferred from the triplet distribution.

2.3 Substitution cost

A penalty score is assigned to the different types of nucleotide substitution through the transformation of empirically estimated substitution matrices into corresponding cost matrices. For each codon index $ci \in \{1, 2, 3\}$, the cost matrix M_{ci}^c is given by the normalized and inverted substitution matrix M_{ci}^s following normalisation:

$$M_{ci}^c = \left(\frac{M_{ci}^s}{max_{(M_{ci}^s)}} \right)^{-1} \tag{1}$$

This transformation assigns the substitution with the highest rate (e.g. G→A) a cost of 1, and the costs of the other substitutions were proportionally adapted to this baseline cost. A cost of zero is assigned when no change occurs. For our study, a substitution matrix for each codon position was derived from HIV-1 integrase nucleotide sequences independently collected from the Los Alamos database, with codon-based substitution patterns and rates estimated under the General Time Reversible model (**Supplementary Material**).

2.4 Genetic barrier to resistance

Given a sub-epidemic $\mathcal{E}$, the genetic barrier to resistance mutation R is a cost function $GB_{\mathcal{E}}(R)$ that quantifies the number and type of nucleotide substitutions required for the virus to evolve from wild type diversity in the virus population of $\mathcal{E}$ to R (Figure 1). This baseline diversity is defined by the set of wild type type triplets $\{WT_{t,i}\}$, and we will denote their prevalences as prev $(WT_{t,i})$. Furthermore, as R is an amino acid, we enumerate all triplets that translate into R and will refer to this set as $\{R_{t,j}\}$. First, we determine a score for a given wild type triplet WT_t to R, while considering all triplets that translate into R (i.e., $\{R_{t,j}\}$):

$$\text{Score}_{WT_t}(R) = \sum_j \exp\left(-\text{Cost}\left(WT_t, R_{t,j}\right)\right), \tag{2}$$

where Cost $(.)$ quantifies the cumulative substitution cost to mutate from a WT_t into a particular resistance triplet $R_{t,a}$, as provided in Equation 3. This summation incorporates all possible evolution paths and subsequently assigns lower cumulative costs (i.e., more likely mutation pathways) a higher contribution to the total triplet score. Taking into account different but almost equally likely evolution paths, compared to only considering the minimum cumulative substitution cost, results in an elevated sensitivity to detect subtype-dependencies in the genetic barrier.

To compute the cumulative substitution cost for $R_{t,a}$ (i.e., subst $(WT_t, R_{t,a})$), we compute the sum of the substitution cost for each triplet position, using substitution matrices M_{ci}^c as defined in section 2.3:

$$\text{Cost}\left(WT_t, R_{t,a}\right) = M_1^c\left(WT_t, R_{t,a}\right) + M_2^c\left(WT_t, R_{t,a}\right) + M_3^c\left(WT_t, R_{t,a}\right) \tag{3}$$

Finally, the genetic barrier $GB_{\mathcal{E}}$ is defined as the sum of the scores (Equation 2) of the different wild type triplets in sub-epidemic $\mathcal{E}$, weighted according to their prevalence:

$$GB_{\mathcal{E}}(R) = \sum_{WT_{t,i}} \text{Score}_{WT_t}(R) \cdot \text{prev}\left(WT_{t,i}\right) \tag{4}$$

A $GB_{\mathcal{E}}(R)$ is obtained for each resistance-associated mutation and subtype. In order to visualise most pronounced differences between sub-epidemics, we calculated the average distance in population genetic barrier for each sub-epidemic against other sub-epidemics.

2.5 Statistics

The non-parametric Mann-Whitney test was used to identify differences in the genetic barrier between HIV-1 subtypes, corrected for multiple hypothesis testing using the Benjamini-Hochberg method [43]. Each subtype was compared against a weighted sample (N=500) of the other subtypes, repeated one thousand times. In addition, pairwise comparisons of subtype B with each subtype separately were performed given that most knowledge on integrase resistance development is available for subtype B. Significant comparisons with a difference in the barrier score above 0.1 are retained. All analyses were done using the statistical software package R [33].
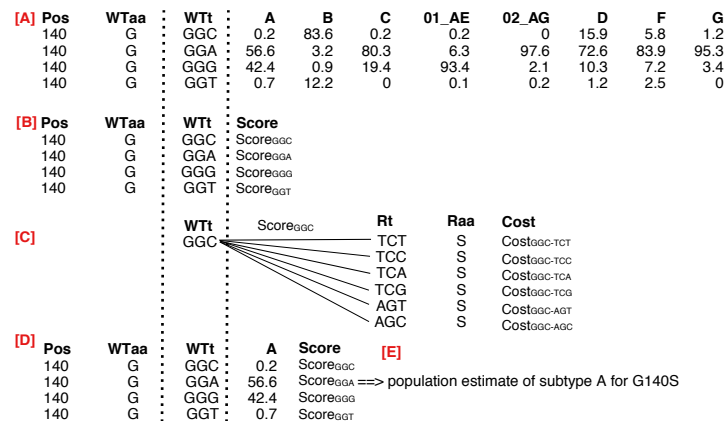
[A]

Pos	WTaa	WTt	A	B	C	01_AE	02_AG	D	F	G
140	G	GGC	0.2	83.6	0.2	0.2	0	15.9	5.8	1.2
140	G	GGA	56.6	3.2	80.3	6.3	97.6	72.6	83.9	95.3
140	G	GGG	42.4	0.9	19.4	93.4	2.1	10.3	7.2	3.4
140	G	GGT	0.7	12.2	0	0.1	0.2	1.2	2.5	0

[B]

Pos	WTaa	WTt	Score
140	G	GGC	$Score_{GGC}$
140	G	GGA	$Score_{GGA}$
140	G	GGG	$Score_{GGG}$
140	G	GGT	$Score_{GGT}$

[C]

WTt	$Score_{GGC}$	Rt	Raa	Cost
GGC		TCT	S	$Cost_{GGC\text{-}TCT}$
		TCC	S	$Cost_{GGC\text{-}TCC}$
		TCA	S	$Cost_{GGC\text{-}TCA}$
		TCG	S	$Cost_{GGC\text{-}TCG}$
		AGT	S	$Cost_{GGC\text{-}AGT}$
		AGC	S	$Cost_{GGC\text{-}AGC}$

[D]

Pos	WTaa	WTt	A	Score	**[E]**
140	G	GGC	0.2	$Score_{GGC}$	
140	G	GGA	56.6	$Score_{GGA}$ ==> population estimate of subtype A for G140S
140	G	GGG	42.4	$Score_{GGG}$	
140	G	GGT	0.7	$Score_{GGT}$	

Figure 1. Overview of methodology for integrase resistance amino acid S (Raa) at position 140 (Pos). [A] Determine the distribution wild-type triplets (WTt) naturally present, the translated amino acid (WTaa) and their frequency across the 8 subtypes. [B] A score is assigned to each wild-type triplet to evolve into the resistance mutation S. [C] This score is obtained by iteratively determining the cumulative substitution cost for the change of WTt into each resistance triplet (Rt) translating into the amino acid S, and subsequently by the summation of the different cumulative costs. [D] Each score is iteratively weighted by the subtype frequency of the wild-type triplet, here shown for subtype A. [E] A population estimate of the genetic barrier is obtained for subtype A. This procedure is done for each subtype, resulting in a subtype-specific population genetic barrier for 140S. When all resistance-associated mutations are considered, a matrix of genetic barrier values is created where subtypes are shown as rows and mutations as columns. Figure 4 illustrates the visualisation of this matrix.

3 Results

3.1 Data

A total of 10235 viral sequences coding for the HIV-1 integrase enzyme fulfilled the inclusion criteria, resulting in 410 sequences classified as subtype A (4%), 5174 as subtype B (50.5%), 1837 as subtype C (17.9%), 1630 (15.9%) as CRF01_AE, 596 (5.8%) as CRF02_AG, 170 (1.7%) as subtype D, 257 (2.5%) as subtype F and 161 (1.6%) as subtype G. Nucleotide ambiguities only had a modest impact on data processing as the highest percentage of codons that did not fulfil the criteria was 7.6% for the highly polymorphic position 125 (**Supplementary Material**). The distributions of within-subtype pairwise diversity were unimodal and similar across subtypes (**Supplementary Material**).

3.2 Natural variability of integrase

Variation in genetic barrier across subtypes requires a combination of differences in WT_t frequencies and in associated cost scores. We first illustrate relevant integrase genetic variability using triplet entropy calculations to elucidate variation within a single subtype. Next, identifying predominant triplets can reveal major variation between subtypes. In particular triplets associated with INSTI resistance are of interest since they result in a minimal cost.

Figure 2 shows within-subtype entropy values for each resistance position. Some of

this within-subtype variation is translated into the presence of polymorphic mutations. Figure 3 shows the prevalence of resistance-associated mutations in the dataset **(Supplementary Material)**. Most prevalent mutations for subtype A were 50I (25.2%), 74I (22.4%), 203M (5.5%) and 97A (5.1%); for subtype B 156N (17.6%), 230N (10.4%), 50I (9.4%), 203M (6.4%), 119R (5.5%) and 151I (5.2%); for subtype C 50I (35.1%) and 74I (5.0%); for CRF01_AE 165I (17.2%); for CRF02_AG 74I (18.4%), 50I (10.6%), 74M (10.2%), 157Q (8.3%) and 97A (5.5%); for subtype D 203M (15.3%), 97A (6.2%) and 165I (6.2%); for subtype F 165I (30.2%), 50I (9.2%), 163R (6.2%) and 163K (5.4%); for subtype G 50I (15.1%), 74I (10.3%) and 165I (5.3%). These results illustrate both resistance-associated mutations prevalent across subtypes (e.g. 50I and 74I) as well as subtype-specific occurrences (e.g. 156N).

Figure 2 also provides a triplet entropy value of all subtypes combined for each position. High values, compared to within-subtype values, indicate between-subtype variation and suggest differences in predominant codon usage. Of the 41 integrase positions investigated, 16 positions (50, 51, 66, 75, 97, 121, 138, 142, 143, 145, 146, 149, 154, 155, 203 and 230) showed a similar predominant triplet across all subtypes. Of the remaining 25 positions, 11 positions showed a consensus in predominant WT_t among 7 subtypes (49, 54, 74, 92, 118, 148, 153, 157, 160, 170, 263), 9 positions showed a consensus among 6 subtypes observed (68, 95, 119, 128, 140, 147, 156, 165, 166), and finally 5 positions (114, 125, 151, 163, 232) showed a consensus at 5 subtypes. These results illustrate well the extent of relevant integrate diversity within and between subtypes, encompassing varying entropy, resistance mutations and codon predominance. Some positions (e.g. 114 and 140) fully conserved at the amino acid level showed high variability in codon usage, while other positions (e.g. 50 and 74) showed limited variation in codon predominance but resistance-associated mutations occurred at increased frequencies.
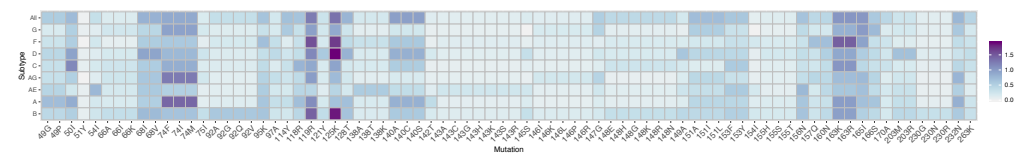

Figure 2. Triplet Entropy per position within each subtype, and also for all subtypes combined (All)
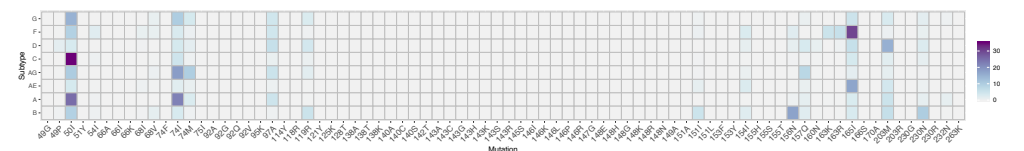

Figure 3. Prevalence (%) of INSTI resistance-associated mutations in treatment-naive patients. Mutations 49P, 50I, 54I, 68I/V, 74I/M, 97A, 119R, 151I, 154I, 156N, 157Q, 160N, 163K/R, 165I, 203M, 230N and 232N were observed with a frequency above 1% in at least one subtype **(Supplementary Material)**.

3.3 Calculated genetic barrier

The observed baseline differences in triplet frequency between subtypes will only impact the genetic barrier to resistance when they are also accompanied by variation in triplet

scores. Subtype frequencies of wild-type nucleotide triplets, translated amino acids and associated cumulative substitution costs are available as **Supplementary Material**.

Triplet score variability was subsequently investigated according to patterns in entropy and codon usage as described above. Limited variation in the evolutionary potential between subtypes is expected for the 35 resistance-associated mutations at 16 integrase positions with a similar predominant triplet and consequently similar scores. These positions are in addition generally characterised by a low frequency of resistance-associated mutations and a high level of consistent codon usage at baseline (**Supplementary Material**), although some exceptions exist with higher entropy values at positions 50, 138 and 203 (Figure 2) or resistance-associated mutations prevalent above 5% at low entropy positions 97 and 230 (Figure 3). Among the group of 25 positions, a larger heterogeneity in cost scores can be expected due to variable predominance complementing any variation in codon usage and mutation prevalence. The presence of resistance-associated mutations were 74I (subtypes A, G and CRF02_AG), 74M (CRF02_AG), 119R (subtype B), 151I (subtype B), 156N (subtype B), 157Q (CRF02_AG), 163K (subtype F), 163R (subtype F) an 165I (subtypes D, F, G and CRF01_AE) (Figure 3). In addition, subtype variation in predominant WT_t resulted most strongly in a different cost score for the following mutations: 54I, 68I, 74I/M, 92A/G, 95K, 119R, 125K, 128T, 140S, 147G, 148H/K/R, 151I, 156N, 160N, 163R, 165I, 166S, 170A, 232N and 263K (**Supplementary Material**).

Finally, combining into a complex interplay, reported results on genetic diversity and evolutionary costs can be captured into a single value denoting a population-based genetic potential for an INSTI mutation and subtype (Figure 4 and Figure 4). Significant differences in calculated genetic barrier were observed for 15 mutations at 12 positions (Figure 4), with a lower resistance potential for 50I (subtype D), 54I (CRF01_AE), 119R (subtypes C and F), 148H (subtype C), 151I (subtypes A, G and CRF02_AG) and 156N (subtype A and CRF01_AE) and 166S (CRF02_AG), while a higher potential for 68I (subtype A and CRF01_AE), 119R (CRF01_AE), 125K (subtype B), 140S (subtype B), 148R (subtype C), 151I (subtype B) and 156N (subtype B). Pairwise comparisons with subtype B revealed additional differences for 50I (subtypes A and C), 74I (subtypes A and AG), 163R (CRF01_AE) and 165I (subtype F and CRF01_AE) (**Supplementary Material**). A similar analysis applied to the RT enzyme revealed respectively a higher potential for 8 and a lower potential for 7 RTI resistance-associated mutations, including mutations at positions 106 and 65.

The calculation of a wild type triplet score is based on the summation of all possible triplet cumulative costs, rather than only considering the minimum triplet cumulative cost. To illustrate the effect of this strategy, Figure 5 provides a comparison of the evolutionary potential calculated by our strategy and a simplified version which only considers the lowest cost for determine the cost. A modest impact on resistance potential was observed for a large number of resistance mutations, although for some mutations a large increase can be detected when almost equally likely paths were accounted for. Positions with high level of triplet variability (e.g. 119, 163, 230) are primarily expected to be affected across all subtypes, but important subtype-specific effects were also observed (e.g. 140S in subtype B, 148R and 263K in subtype C). **Supplementary Material** provides more comparative analyses on the impact of using empirically versus arbitrary defined substitution matrices.
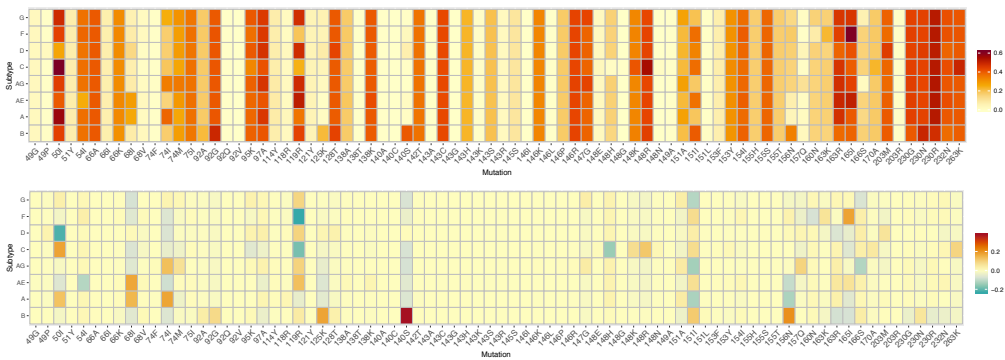
Figure 4. The upper panel shows the estimated population genetic barrier for each combination of a resistance-associated mutation and a subtype. A higher value (red) represents an increased potential for adaptation and is indicative for a lower genetic barrier to resistance. A subtype-specific value of this population genetic barrier is calculated by first assigning each wild-type triplet with a score that indicates the ease to evolve into the resistance amino acid (See the Methods section for a detailed description how this cost score for a wild-type triplet is derived). Next, the sum of all triplet scores is weighted by the triplet prevalence so that most frequent triplets have a larger impact on the population genetic barrier estimate. The lower panel shows for each subtype the average difference in population genetic barrier from the other subtypes. A lower genetic barrier to resistance for that subtype compared to the other subtypes is shown in red while a higher genetic barrier is shown in green.

4   Discussion:

We developed an intuitive and novel methodology to quantify the evolutionary potential for viral escape from selective pressure in fast-evolving viruses, and presented an application evaluating the impact of genetic background in HIV-1 integrase on INSTI resistance development. The integrase enzyme has proven a successful drug target, and the objective of maximising the benefits of INSTIs is being translated into their global roll-out. However, as for other drug classes, the effectiveness of INSTIs can be challenged by HIV-1's high rates of evolution when treatment options are non-optimal [2, 11, 29]. Viral adaptation to INSTI pressure through mutational pathways has been well reported, but the role of natural nucleotide variation on resistance development is less well understood [10, 17]. Additionally to resistance mutations present at baseline, differential codon usage can affect the mode and tempo of INSTI resistance pathways [23, 38]. This study presented here provided new insights into the processes underlying INSTI resistance development by quantifying the genetic barrier to resistance.

A large dataset of globally prevailing HIV-1 subtypes was investigated for their evolutionary potential for viral escape from all INSTIs available for HIV-1 treatment. This study confirms the non-polymorphic amino acid nature of most INSTI resistance positions [25, 35], as primarily single subtype-specific occurrences of resistance-associated mutations at low frequencies were detected. By contrast, variable codon predominance across subtypes was more pronounced and consequently we detected 15 resistance-associated mutations included in this study that significantly varied in estimated barrier in one or more subtypes. Modest variability in genetic barrier was observed for major INSTI resistance mutations except for 148H/R in subtype C which causes high-level resistance to all INSTIs [37]. More minor INSTI mutations, i.e. having an impact on drug activity in presence of a major mutation or
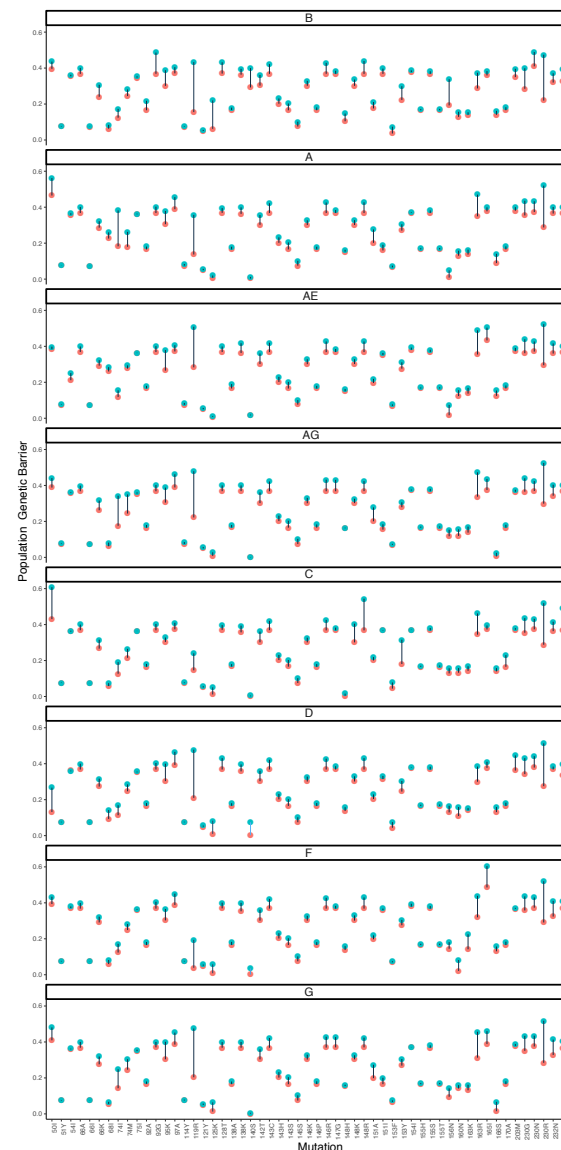
Figure 5. Each of 8 panels shows the estimated population genetic barrier for a single subtype, with the estimate either calculated as described in the Methods section (blue) or calculated by a simpler but similar methodology (red). Instead of obtaining the cost score for each wild-type triplet by summing all possible cumulative costs (blue), thereby taking into account almost equally likely substitution pathways of a wild-type triplet to the resistance amino acid, a cost score can also be calculated by only using the minimum cumulative cost (red) and therefore only considering the shortest substitution pathway to the resistance amino acid and ignoring other but almost equally likely mutational pathways to resistance. To increase the comparability of the two measures, we also took the negative exponential of the minimum score. Only increases in values are possible due to the summation, and we restricted the figure to the subset of mutations with a difference larger than 0.01.

compensating for defects in replication capacity [37], were characterised by increased   262
variability in diversity and associated scores. Most notably was the varying potential for   263

140S in subtype B viruses, confirming previous findings [23,32], which is an accessory mutation often occurring in combination with 148H. We and others have indeed previously reported the preference of the 148H pathway in treatment-experienced patients infected with subtype B viruses [17,38].

Our framework improves related counting studies at several stages. Previous studies calculated the genetic barrier by only distinguishing between transitions and transversions, following an approach developed for the HIV-1 protease and reverse transcriptase enzymes [43]. We applied an empirically-derived and evolutionary-based cost for each possible substitution and triplet position, which resulted in an increased detection sensitivity, in particular transversions displayed substantial variation in costs compared to the arbitrary defined values in other studies. Furthermore, we take into account cost scores from all possible pathways, compared to restricting to the minimum score. This approach allowed to detect relevant increases in the estimated evolutionary potential (e.g. 74I), in particular for positions with high level of codon variability. To illustrate the improved sensitivity, the suggested subtype effect on the potential for 140C could not be confirmed in our study, as this mutation requires a substitution at the first triplet position, known to be evolutionary costly. In contrast to 140S, the prevalence of 140C in patients failing INSTI-based treatment does not differ across subtypes in the Stanford HIV drug resistance database [35], which strengthens our finding. A complete in-depth comparison of our findings with these studies is hindered by their limited number of HIV-1 subtypes and INSTI resistance mutations included. As most genetic barrier information to date has been established for the historical drug class of RTIs, we also applied our approach to the RT enzyme. Subtype-dependencies in genetic barrier both previously predicted and observed in clinical practice could be confirmed (e.g. 106M in subtype C). However, compared to related RTI counting studies [43], the increased sensitivity of our method also resulted in the identification of a higher potential for 65R in subtype C. A higher selection rate of 65R in this subtype has been well established [40] and attributed to a mechanistic basis of template factors [7]. However, an explanation by codon usage has been discarded [7], although we suggest an additional impact of codon usage when all possible paths are taken into account.

While our framework provides a population-based cost for a single position, it only represents a simplified estimate of the actual *in vivo* genetic barrier, which results from a multi-modal interplay that eventually defines the evolutionary time to drug resistance. Selection of a mutation and its rate of fixation heavily depend on factors such as patient adherence, treatment potency, impact on viral fitness and (non-additive) epistatic mutational interactions. Positions implicated in immune escape (e.g. 125) can also further influence the rate of resistance accumulation [8,41]. It can be difficult to obtain information on all these influencing factors in a timely manner and hence to construct an accurate model capturing resistance evolutionary dynamics. When such an adequate model is lacking, our distance-based method offers a valuable alternative to assess the genetic barrier to resistance. Our methodology however can not predict novel resistance-associated mutations or subtype-specific pathways. A cost matrix was used for all subtypes assuming that substitution rates are equal between subtypes, however, the use of group-specific cost matrices can be accommodated by our framework.

In conclusion, the findings presented in this study are important in the context of up-scaled introduction of DTG and novel INSTIs in LMICs, where non-B subtypes prevail, although more studies are needed to further validate our results. Future applications of this reproducible framework do not only relate to novel HIV-1 drug classes or subtypes but, as principles of HIV-1 drug resistance are generally shared with other pathogens known to escape selective pressure, our methodology can be easily transferred to identify a role of genetic diversity of these pathogens, in particular for antivirals which are being evaluated in early development or clinical trials.

5 Competing interests

Acknowledgements

References

1. A. B. Abecasis, K. Deforche, L. T. Bacheler, P. McKenna, A. P. Carvalho, P. Gomes, A. M. Vandamme, and R. J. Camacho. Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C, F, G and CRF02AG. *Antivir. Ther. (Lond.)*, 11(5):581–589, 2006.

2. L. Achieng and D. J. Riedel. Dolutegravir Resistance and Failure in a Kenyan Patient. *J. Infect. Dis.*, 219(1):165–167, Jan 2019.

3. L. C. Alcantara, S. Cassol, P. Libin, K. Deforche, O. G. Pybus, M. Van Ranst, B. Galvao-Castro, A. M. Vandamme, and T. de Oliveira. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res.*, 37(Web Server issue):W634–642, Jul 2009.

4. N. Beerenwinkel, H. Montazeri, H. Schuhmacher, P. Knupfer, V. von Wyl, H. Furrer, M. Battegay, B. Hirschel, M. Cavassini, P. Vernazza, E. Bernasconi, S. Yerly, J. Boni, T. Klimkait, C. Cellerai, H. F. Gunthard, V. Aubert, J. Barth, M. Bateau, E. Bernasconi, J. Boni, H. C. Bucher, C. Burton-Jeangros, A. Calmy, M. Cavassini, M. Egger, L. Elzi, J. Fellay, P. Francioli, H. Furrer, C. A. Fux, M. Gorgievski, H. Gunthard, D. Haerry, B. Hasse, H. H. Hirsch, B. Hirschel, I. Hosli, C. Kahlert, L. Kaiser, O. Keiser, C. Kind, T. Klimkait, H. Kovari, B. Ledergerber, G. Martinetti, B. Martinez de Tejada, K. Metzner, N. Muller, D. Nadal, G. Pantaleo, A. Rauch, S. Regenass, M. Rickenbach, C. Rudin, P. Schmid, D. Schultze, F. Schoni-Affolter, J. Schupbach, R. Speck, P. Taffe, A. Telenti, A. Trkola, P. Vernazza, R. Weber, and S. Yerly. The individualized genetic barrier predicts treatment response in a large cohort of HIV-1 infected patients. *PLoS Comput. Biol.*, 9(8):e1003203, 2013.

5. B. J. Betts and R. W. Shafer. Algorithm specification interface for human immunodeficiency virus type 1 genotypic interpretation. *J. Clin. Microbiol.*, 41(6):2792–2794, Jun 2003.

6. J. L. Blanco, V. Varghese, S. Y. Rhee, J. M. Gatell, and R. W. Shafer. HIV-1 integrase inhibitor resistance and its clinical implications. *J. Infect. Dis.*, 203(9):1204–1214, May 2011.

7. B. G. Brenner and D. Coutsinos. The K65R mutation in HIV-1 reverse transcriptase: genetic barriers, resistance profile and clinical implications. *HIV Ther*, 3(6):583–594, Nov 2009.
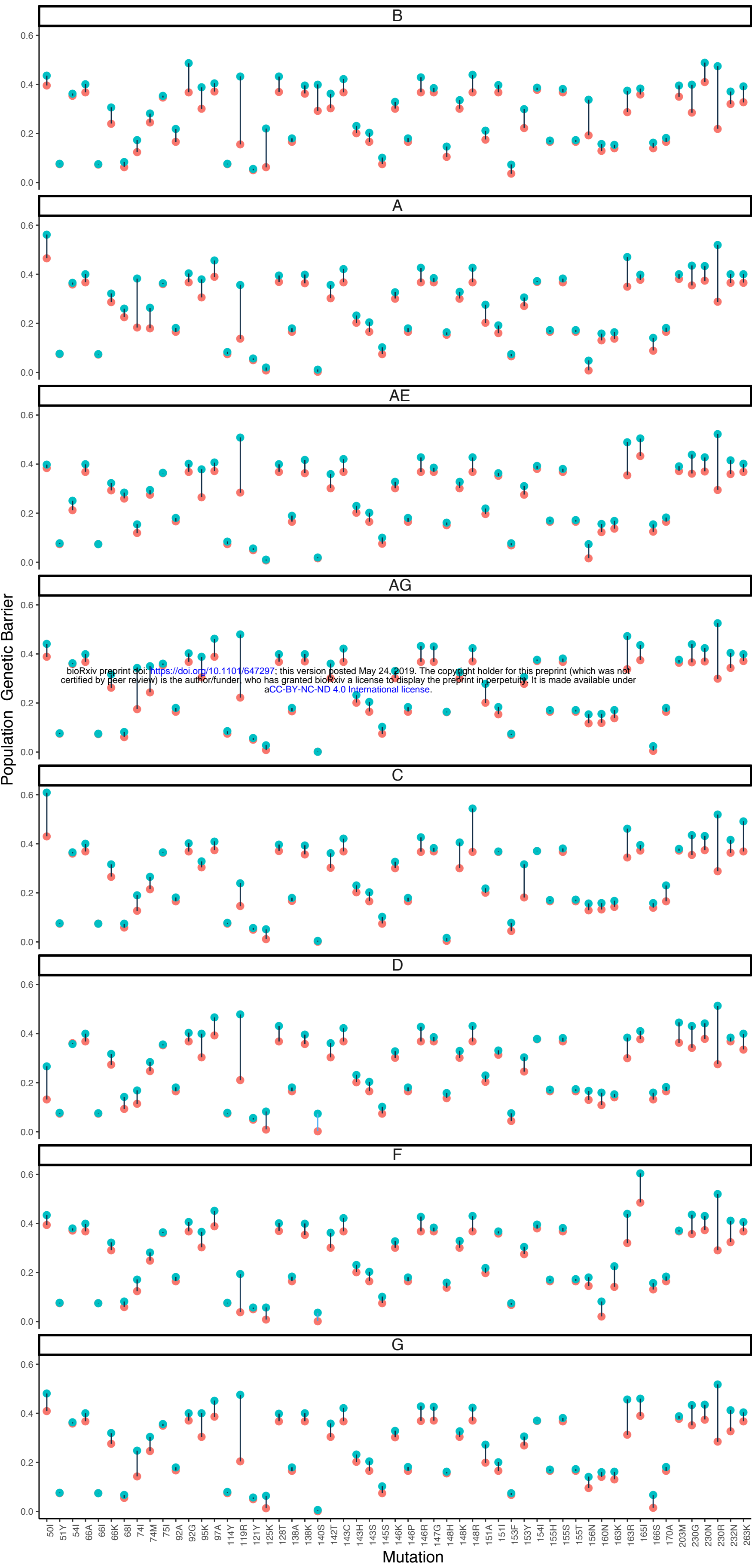
8. M. A. Brockman, D. R. Chopera, A. Olvera, C. J. Brumme, J. Sela, T. J. Markle, E. Martin, J. M. Carlson, A. Q. Le, R. McGovern, P. K. Cheung, A. D. Kelleher, H. Jessen, M. Markowitz, E. Rosenberg, N. Frahm, J. Sanchez, S. Mallal, M. John, P. R. Harrigan, D. Heckerman, C. Brander, B. D. Walker, and Z. L. Brumme. Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. *J. Virol.*, 86(12):6913–6923, Jun 2012.

9. F. Brun-Vezinet, D. Descamps, A. Ruffault, B. Masquelier, V. Calvez, G. Peytavin, F. Telles, L. Morand-Joubert, J. L. Meynard, M. Vray, and D. Costagliola. Clinically relevant interpretation of genotype for resistance to abacavir. *AIDS*, 17(12):1795–1802, Aug 2003.

10. P. Cahn, A. L. Pozniak, H. Mingrone, A. Shuldyakov, C. Brites, J. F. Andrade-Villanueva, G. Richmond, C. B. Buendia, J. Fourie, M. Ramgopal, D. Hagins, F. Felizarta, J. Madruga, T. Reuter, T. Newman, C. B. Small, J. Lombaard, B. Grinsztejn, D. Dorey, M. Underwood, S. Griffith, S. Min, A. E. Saud, F. Ajana, B. Akil, J. F. Villanueva, G. M. Anstead, D. Banhegyi, C. Barros, J. C. Bartczak, W. H. Belloso, C. J. Buendia, J. A. Benetucci, G. Blick, M. Bloch, V. B. Martinez, I. Brar, C. R. Alves, J. H. Burack, P. E. Cahn, A. C. Sanchez, M. M. Caseiro, K. K. Casey, L. I. Cassetti, M. A. Iglesia, O. E. Chernova, B. C. Sala, N. Clumeck, B. Conway, J. L. De Andrade Neto, I. d. e. L. Gil, P. De Truchis, E. D. Comerma, J. F. Delfraissy, J. G. Den Hollander, G. Di Perri, R. S. Diaz, D. Duiculescu, J. H. Elliott, J. J. Eron, J. Feinberg, F. A. Felizarta, T. M. File, E. Florence, J. F. Cid, J. Fourie, M. J. Galindo, P. Gargalianos-Kakolyris, J. C. Gathe, J. Ghosn, C. Gogos, J. G. Garcia, M. G. Hernandez-Mora, A. Gori, C. Goulston, E. R. Granados-Reyes, B. G. Grinsztejn, P. G. Gulick, S. K. Gupta, D. P. Hagins, W. D. Hardy, P. Hay, W. K. Henry, C. B. Hicks, S. L. Hodder, J. Horton, G. Huhn, T. T. Jefferson, M. A. Johnson, P. C. Johnson, C. Katlama, S. Kegg, K. M. Khafizov, H. Khanlou, S. H. Khoo, M. J. Kozal, O. A. Kozyrev, V. V. Kulagin, P. N. Kumar, M. Lasso, G. Latiff, A. Lazzarin, J. C. Legrand, H. H. Lin, Y. H. Lin, K. J. Logue, J. J. Lombaard, S. H. Lupo, J. V. Madruga, M. d. e. l. M. Masia, L. H. McCurdy, C. K. McDonald, J. L. Meier, E. V. Mikhaylov, D. Mildvan, H. Mingrone, J. M. Molina, G. F. Moshkovich, J. L. Mosqueda-Gomez, M. Moutschen, M. D. Murphy, F. I. Nagimova, R. G. Nahass, T. N. Newman, C. L. Newman-Whitlow, T. C. Nguyen, J. P. O'Keefe, O. O. Osiyemi, V. Paparizos, D. A. Parks, C. Perez, G. Pialoux, S. Piga, I. Poizot-Martin, V. V. Pokrovsky, A. Pronin, Y. Quertainmont, A. G. Rakhmanova, M. N. Ramgopal, T. Q. Reuter, E. Ribera, G. J. Richmond, W. J. Robbins, J. R. Bano, J. E. Rodriguez, M. M. Rosas-Dossetti, P. J. Ruane, S. Rugina, S. Rusconi, E. P. Ryamova, M. S. Saag, R. Samuel, U. S. Sandkovsky, A. J. Scarsella, S. Schneider, S. R. Schrader, A. R. Scribner, P. Shalit, A. A. Shuldyakov, D. S. Siraj, J. Slim, L. M. Sloan, F. M. Smaill, C. B. Small, D. B. Sonin, A. Streinu-Cercel, K. T. Tashima, P. Tebas, E. Teicher, C. Tremblay, H. C. Tsai, U. Tupinambas, T. J. Vanig, F. J. Mendez, M. A. De Miguel, J. H. Wang, D. J. Ward, D. A. Warner, D. A. Wheeler, E. G. Wilkins, M. J. Wolff, W. W. Wong, Y. Yazdanpanah, G. P. Yeni, and B. S. Zingman. Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. *Lancet*, 382(9893):700–708, Aug 2013.

11. M. Cardoso, T. Baptista, I. Diogo, M. J. Aleixo, N. Marques, K. Mansinho, and P. Gomes. Two cases of dolutegravir failure with R263K mutation. *AIDS*, 32(17):2639–2640, Nov 2018.

12. C. Charpentier and D. Descamps. Resistance to HIV Integrase Inhibitors: About R263K and E157Q Mutations. *Viruses*, 10(1), 01 2018.

13. L. Cuypers, G. Li, P. Libin, S. Piampongsant, A. M. Vandamme, and K. Theys. Genetic Diversity and Selective Pressure in Hepatitis C Virus Genotypes 1-6: Significance for Direct-Acting Antiviral Treatment and Drug Resistance. *Viruses*, 7(9):5018–5039, Sep 2015.

14. K. Deforche, A. Cozzi-Lepri, K. Theys, B. Clotet, R. J. Camacho, J. Kjaer, K. Van Laethem, A. Phillips, Y. Moreau, J. D. Lundgren, and A. M. Vandamme. Modelled in vivo HIV fitness under drug selective pressure and estimated genetic barrier towards resistance are predictive for virological response. *Antivir. Ther. (Lond.)*, 13(3):399–407, 2008.

15. A. M. Dennis, S. Hue, E. Learner, J. Sebastian, W. C. Miller, and J. J. Eron. Rising prevalence of non-B HIV-1 subtypes in North Carolina and evidence for local onward transmission. *Virus Evol*, 3(1):vex013, Jan 2017.

16. M. Gotte. The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. *Curr Opin Virol*, 2(5):644–650, Oct 2012.

17. Y. S. Han, T. Mesplede, and M. A. Wainberg. Differences among HIV-1 subtypes in drug resistance against integrase inhibitors. *Infect. Genet. Evol.*, 46:286–291, 12 2016.

18. A. Hauser, A. Hofmann, K. Meixenberger, B. Altmann, K. Hanke, V. Bremer, B. Bartmeyer, and N. Bannert. Increasing proportions of HIV-1 non-B subtypes and of NNRTI resistance between 2013 and 2016 in Germany: Results from the national molecular surveillance of new HIV-diagnoses. *PLoS ONE*, 13(11):e0206234, 2018.

19. A. Kramvis. Genotypes and genetic variability of hepatitis B virus. *Intervirology*, 57(3-4):141–150, 2014.

20. G. Li, S. Piampongsant, N. R. Faria, A. Voet, A. C. Pineda-Pena, R. Khouri, P. Lemey, A. M. Vandamme, and K. Theys. An integrated map of HIV genome-wide variation from a population perspective. *Retrovirology*, 12:18, Feb 2015.

21. P. Libin, G. Beheydt, K. Deforche, S. Imbrechts, F. Ferreira, K. Van Laethem, K. Theys, A. P. Carvalho, J. Cavaco-Silva, G. Lapadula, C. Torti, M. Assel, S. Wesner, J. Snoeck, J. Ruelle, A. De Bel, P. Lacor, P. De Munter, E. Van Wijngaerden, M. Zazzi, R. Kaiser, A. Ayouba, M. Peeters, T. de Oliveira, L. C. Alcantara, Z. Grossman, P. Sloot, D. Otelea, S. Paraschiv, C. Boucher, R. J. Camacho, and A. M. Vandamme. RegaDB: community-driven data management and analysis for infectious diseases. *Bioinformatics*, 29(11):1477–1480, Jun 2013.

22. P. Libin, K. Deforche, A. B. Abecasis, and K. Theys. VIRULIGN: fast codon-correct alignment and annotation of viral genomes. *Bioinformatics*, Oct 2018.

23. A. I. Maiga, I. Malet, C. Soulie, A. Derache, V. Koita, B. Amellal, L. Tchertanov, O. Delelis, L. Morand-Joubert, J. F. Mouscadet, R. Murphy, M. Cisse, C. Katlama, V. Calvez, and A. G. Marcelin. Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02_AG subtypes. *Antivir. Ther. (Lond.)*, 14(1):123–129, 2009.

24. S. Megens and K. V. Laethem. HIV-1 genetic variation and drug resistance development. *Expert Rev Anti Infect Ther*, 11(11):1159–1178, Nov 2013.

25. R. E. Myers and D. Pillay. Analysis of natural sequence variation and covariation in human immunodeficiency virus type 1 integrase. *J. Virol.*, 82(18):9228–9235, Sep 2008.

26. U. Neogi, A. Haggblom, M. Santacatterina, G. Bratt, M. Gisslen, J. Albert, and A. Sonnerborg. Temporal trends in the Swedish HIV-1 epidemic: increase in non-B subtypes and recombinant forms over three decades. *PLoS ONE*, 9(6):e99390, 2014.

27. P. on Antiretroviral Guidelines for Adults and P. Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV. *Department of Health and Human Services*, -(-):–, Oct 2018.

28. A. M. Oster, W. M. Switzer, A. L. Hernandez, N. Saduvala, J. O. Wertheim, N. Nwangwu-Ike, M. C. Ocfemia, E. Campbell, and H. I. Hall. Increasing HIV-1 subtype diversity in seven states, United States, 2006-2013. *Ann Epidemiol*, 27(4):244–251, 04 2017.

29. M. J. Pena, N. Chueca, A. D'Avolio, J. M. Zarzalejos, and F. Garcia. Virological Failure in HIV to Triple Therapy With Dolutegravir-Based Firstline Treatment: Rare but Possible. *Open Forum Infect Dis*, 6(1):ofy332, Jan 2019.

30. A. N. Phillips, F. Venter, D. Havlir, A. Pozniak, D. Kuritzkes, A. Wensing, J. D. Lundgren, A. De Luca, D. Pillay, J. Mellors, V. Cambiano, L. Bansi-Matharu, F. Nakagawa, T. Kalua, A. Jahn, T. Apollo, O. Mugurungi, P. Clayden, R. K. Gupta, R. Barnabas, P. Revill, J. Cohn, S. Bertagnolio, and A. Calmy. Risks and benefits of dolutegravir-based antiretroviral drug regimens in sub-Saharan Africa: a modelling study. *Lancet HIV*, 6(2):e116–e127, Feb 2019.

31. A. C. Pineda-Pena, N. R. Faria, S. Imbrechts, P. Libin, A. B. Abecasis, K. Deforche, A. Gomez-Lopez, R. J. Camacho, T. de Oliveira, and A. M. Vandamme. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.*, 19:337–348, Oct 2013.

32. A. Piralla, S. Paolucci, R. Gulminetti, G. Comolli, and F. Baldanti. HIV integrase variability and genetic barrier in antiretroviral naïve and experienced patients. *Virol. J.*, 8:149, 2011.

33. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

34. S. Y. Rhee, M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, and R. W. Shafer. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.*, 31(1):298–303, Jan 2003.

35. S. Y. Rhee, T. F. Liu, M. Kiuchi, R. Zioni, R. J. Gifford, S. P. Holmes, and R. W. Shafer. Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology*, 5:74, 2008.

36. M. S. Saag, C. A. Benson, R. T. Gandhi, J. F. Hoy, R. J. Landovitz, M. J. Mugavero, P. E. Sax, D. M. Smith, M. A. Thompson, S. P. Buchbinder, C. Del Rio, J. J. Eron, G. Fatkenheuer, H. F. Gunthard, J. M. Molina, D. M.

Jacobsen, and P. A. Volberding. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2018 Recommendations of the International Antiviral Society-USA Panel. *JAMA*, 320(4):379–396, 07 2018.

37. R. W. Shafer and J. M. Schapiro. HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev*, 10(2):67–84, 2008.

38. K. Theys, A. Abecasis, P. Libin, P. T. Gomes, J. Cabanas, R. J. Camacho, and K. Van Laethem. Discordant predictions of residual activity could impact dolutegravir prescription upon raltegravir failure. *J. Clin. Virol.*, 70:120–127, Sep 2015.

39. K. Theys, K. Deforche, G. Beheydt, Y. Moreau, K. van Laethem, P. Lemey, R. J. Camacho, S. Y. Rhee, R. W. Shafer, E. Van Wijngaerden, and A. M. Vandamme. Estimating the individualized HIV-1 genetic barrier to resistance using a nelfinavir fitness landscape. *BMC Bioinformatics*, 11:409, 2010.

40. K. Theys, J. Vercauteren, J. Snoeck, M. Zazzi, R. J. Camacho, C. Torti, E. Schulter, B. Clotet, A. Sonnerborg, A. De Luca, Z. Grossman, D. Struck, A. M. Vandamme, and A. B. Abecasis. HIV-1 subtype is an independent predictor of reverse transcriptase mutation K65R in HIV-1 patients treated with combination antiretroviral therapy including tenofovir. *Antimicrob. Agents Chemother.*, 57(2):1053–1056, Feb 2013.

41. M. Tschochner, A. Chopra, T. M. Maiden, I. F. Ahmad, I. James, H. Furrer, H. F. Gunthard, S. Mallal, A. Rauch, and M. John. Effects of HIV type-1 immune selection on susceptability to integrase inhibitor resistance. *Antivir. Ther. (Lond.)*, 14(7):953–964, 2009.

42. D. Turner, B. Brenner, D. Moisi, M. Detorio, R. Cesaire, T. Kurimura, H. Mori, M. Essex, S. Maayan, and M. A. Wainberg. Nucleotide and amino acid polymorphisms at drug resistance sites in non-B-subtype variants of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, 48(8):2993–2998, Aug 2004.

43. D. A. van de Vijver, A. M. Wensing, G. Angarano, B. Asjo, C. Balotta, E. Boeri, R. Camacho, M. L. Chaix, D. Costagliola, A. De Luca, I. Derdelinckx, Z. Grossman, O. Hamouda, A. Hatzakis, R. Hemmer, A. Hoepelman, A. Horban, K. Korn, C. Kucherer, T. Leitner, C. Loveday, E. MacRae, I. Maljkovic, C. de Mendoza, L. Meyer, C. Nielsen, E. L. Op de Coul, V. Ormaasen, D. Paraskevis, L. Perrin, E. Puchhammer-Stockl, L. Ruiz, M. Salminen, J. C. Schmit, F. Schneider, R. Schuurman, V. Soriano, G. Stanczak, M. Stanojevic, A. M. Vandamme, K. Van Laethem, M. Violin, K. Wilbe, S. Yerly, M. Zazzi, and C. A. Boucher. The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes. *J. Acquir. Immune Defic. Syndr.*, 41(3):352–360, Mar 2006.

44. K. Van Laethem, A. De Luca, A. Antinori, A. Cingolani, C. F. Perna, and A. M. Vandamme. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. *Antivir. Ther. (Lond.)*, 7(2):123–129, Jun 2002.

45. V. Varghese, T. F. Liu, S. Y. Rhee, P. Libiran, C. Trevino, W. J. Fessel, and R. W. Shafer. HIV-1 integrase sequence variability in antiretroviral naïve patients and in triple-class experienced patients subsequently treated with raltegravir. *AIDS Res. Hum. Retroviruses*, 26(12):1323–1326, Dec 2010.

46. F. Zanini, V. Puller, J. Brodin, J. Albert, and R. A. Neher. In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evol*, 3(1):vex003, Jan 2017.

[A]

Pos	WTaa	WTt	A	B	C	01_AE	02_AG	D	F	G
140	G	GGC	0.2	83.6	0.2	0.2	0	15.9	5.8	1.2
140	G	GGA	56.6	3.2	80.3	6.3	97.6	72.6	83.9	95.3
140	G	GGG	42.4	0.9	19.4	93.4	2.1	10.3	7.2	3.4
140	G	GGT	0.7	12.2	0	0.1	0.2	1.2	2.5	0

[B]

Pos	WTaa	WTt	Score
140	G	GGC	$Score_{GGC}$
140	G	GGA	$Score_{GGA}$
140	G	GGG	$Score_{GGG}$
140	G	GGT	$Score_{GGT}$

[C]

WTt	$Score_{GGC}$	Rt	Raa	Cost
GGC		TCT	S	$Cost_{GGC\text{-}TCT}$
		TCC	S	$Cost_{GGC\text{-}TCC}$
		TCA	S	$Cost_{GGC\text{-}TCA}$
		TCG	S	$Cost_{GGC\text{-}TCG}$
		AGT	S	$Cost_{GGC\text{-}AGT}$
		AGC	S	$Cost_{GGC\text{-}AGC}$

[D]

Pos	WTaa	WTt	A	Score	**[E]**
140	G	GGC	0.2	$Score_{GGC}$	
140	G	GGA	56.6	$Score_{GGA}$	==> population estimate of subtype A for G140S
140	G	GGG	42.4	$Score_{GGG}$	
140	G	GGT	0.7	$Score_{GGT}$	