Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

HIV-1 Protease Evolvability Is Affected by Synonymous Nucleotide Recoding

Maria Nevot,[a] Ana Jordan-Paiz,[a] Glòria Martrus,[a]* Cristina Andrés,[a]* Damir García-Cehic,[b,c] Josep Gregori,[b,d] Sandra Franco,[a] Josep Quer,[b,c,e] Miguel Angel Martinez[a]

[a]IrsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona (UAB), Badalona, Spain

[b]Liver Unit, Liver Disease Laboratory-Viral Hepatitis, Internal Medicine Department, Vall d'Hebron Institut Recerca (VHIR)-Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Spain

[c]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain

[d]Roche Diagnostics SL, Sant Cugat del Vallès, Barcelona, Spain

[e]Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

ABSTRACT One unexplored aspect of HIV-1 genetic architecture is how codon choice influences population diversity and evolvability. Here we compared the levels of development of HIV-1 resistance to protease inhibitors (PIs) between wild-type (WT) virus and a synthetic virus (MAX) carrying a codon-pair-reengineered protease sequence including 38 (13%) synonymous mutations. The WT and MAX viruses showed indistinguishable replication in MT-4 cells or peripheral blood mononuclear cells (PBMCs). Both viruses were subjected to serial passages in MT-4 cells, with selective pressure from the PIs atazanavir (ATV) and darunavir (DRV). After 32 successive passages, both the WT and MAX viruses developed phenotypic resistance to PIs (50% inhibitory concentrations [IC_{50}s] of 14.6 ± 5.3 and 21.2 ± 9 nM, respectively, for ATV and 5.9 ± 1.0 and 9.3 ± 1.9, respectively, for DRV). Ultradeep sequence clonal analysis revealed that both viruses harbored previously described mutations conferring resistance to ATV and DRV. However, the WT and MAX virus proteases showed different resistance variant repertoires, with the G16E and V77I substitutions observed only in the WT and the L33F, S37P, G48L, Q58E/K, and L89I substitutions detected only in the MAX virus. Remarkably, the G48L and L89I substitutions are rarely found *in vivo* in PI-treated patients. The MAX virus showed significantly higher nucleotide and amino acid diversity of the propagated viruses with and without PIs ($P < 0.0001$), suggesting a higher selective pressure for change in this recoded virus. Our results indicate that the HIV-1 protease position in sequence space delineates the evolution of its mutant spectrum. Nevertheless, the investigated synonymously recoded variant showed mutational robustness and evolvability similar to those of the WT virus.

IMPORTANCE Large-scale synonymous recoding of virus genomes is a new tool for exploring various aspects of virus biology. Synonymous virus genome recoding can be used to investigate how a virus's position in sequence space defines its mutant spectrum, evolutionary trajectory, and pathogenesis. In this study, we evaluated how synonymous recoding of the human immunodeficiency virus type 1 (HIV-1) protease affects the development of protease inhibitor (PI) resistance. HIV-1 protease is a main target of current antiretroviral therapies. Our present results demonstrate that the wild-type (WT) virus and a virus with recoded protease exhibited different patterns of resistance mutations after PI treatment. Nevertheless, the developed PI resistance phenotypes were indistinguishable between the recoded virus and the WT virus, suggesting that the HIV-1 strain with synonymously recoded protease and the WT virus are equally robust and evolvable.

Received 4 May 2018 **Accepted** 29 May 2018

Accepted manuscript posted online 6 June 2018

Citation Nevot M, Jordan-Paiz A, Martrus G, Andrés C, García-Cehic D, Gregori J, Franco S, Quer J, Martinez MA. 2018. HIV-1 protease evolvability is affected by synonymous nucleotide recoding. J Virol 92:e00777-18. https://doi.org/10.1128/JVI.00777-18.

Editor Guido Silvestri, Emory University

Address correspondence to Miguel Angel Martinez, mmartinez@irsicaixa.es.

* Present address: Glòria Martrus, Heinrich-Pette-Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany; Cristina Andrés, Virology Unit, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

KEYWORDS evolutionary biology, human immunodeficiency virus, proteases, synonymous recoding

Alterations in a DNA or mRNA sequence that do not change the protein amino acid sequence are called synonymous mutations. Although they do not influence the resulting protein sequence, synonymous mutations can still substantially affect cellular processes (1, 2). Notably, synonymous virus genome recoding can have impacts on viral replication capacity and fitness (3), reportedly leading to attenuation of multiple RNA and DNA viruses, including poliovirus (4–7), influenza virus (8, 9), human immunodeficiency virus type 1 (HIV-1) (10–12), simian immunodeficiency virus (SIV) (13), chikungunya virus (14), human respiratory syncytial virus (15–17), porcine reproductive and respiratory syndrome virus (18), echovirus 7 (19, 20), tick-borne encephalitis virus (21), vesicular stomatitis virus, dengue virus (22), adeno-associated virus (23), and papillomavirus (24).

Synonymous virus genome recoding is being investigated as a new strategy for generating novel live attenuated vaccine candidates. This method is promising because the amino acid coding is completely unaffected, thereby avoiding the potential generation of new and undesirable biological properties. Moreover, synonymous virus genome recoding involves the introduction of hundreds or thousands of nucleotide substitutions, which minimizes the risk of phenotypic reversion via point mutations or through recombination with homologous sequences in circulating strains. This is particularly important with regard to RNA viruses, since viral RNA polymerases lack error correction mechanisms (25–27). The high genetic variability of RNA viruses is a critical limitation in designing novel antiviral strategies.

The usefulness of synonymous virus genome recoding goes beyond the generation of new live attenuated vaccines. This method has also been used to identify specific RNA structures required for virus replication (28), virus genome *cis*-inhibitory signal sequences important for complex viral functions (23), and novel antiviral mechanisms within the innate immune response (11, 29, 30), as well as to resolve the importance of codon usage in the temporal regulation of viral gene expression (31). In one interesting example, synonymous virus genome recoding was used to demonstrate that a synonymous position in sequence space can affect poliovirus evolvability and pathogenesis (32, 33).

Like those of other RNA viruses, HIV-1 populations comprise closely related mutant spectra or mutant clouds termed viral quasispecies (34, 35). Mutant cloud composition can affect virus evolvability, fitness, and virulence (25–27). One unexplored aspect of HIV-1 genetic architecture is how codon choice influences population diversity and evolvability. It is presently unclear whether HIV-1 sequences have evolved to optimize both protein coding and DNA/RNA sequences. The HIV-1 genome exhibits a particularly striking bias toward enrichment of A-rich codons, which may be a selectable trait (36) and may affect innate immune recognition (11, 29). Similarly, synonymous codon usage can temporally regulate the expression of structural gene products of SIV (31) and can regulate HIV-1 splicing and replication (12).

Here we aimed to explore whether the synonymous sequence space influences the development of protease inhibitor (PI) resistance and thus to determine whether HIV-1 evolvability is influenced by the natural position of a protease in sequence space. To this end, we compared the levels of development of HIV-1 resistance to the protease inhibitors atazanavir (ATV) and darunavir (DRV) between wild-type HIV-1 (WT) and a synthetic HIV-1 strain carrying a synonymously recoded protease sequence (MAX).

(This article was submitted to an online preprint archive [37].)

RESULTS

WT HIV-1 was compared to the MAX variant, which carried a protease gene including 38 synonymous mutations (13% of the protease sequence) (Fig. 1). These 38 synonymous substitutions were scattered throughout the protease coding region,

```
                *        20         *        40         *
WTp1   : CCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGG :  50
MAXp1  : ..............................................T.. :  50

               60         *        80         *       100
WTp1   : GCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAG : 100
MAXp1  : T........A......T.........................T..T.... : 100

                *       120         *       140         *
WTp1   : AAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATT : 150
MAXp1  : ........TCGC....G..CC.C.....G.................C... : 150

              160         *       180         *       200
WTp1   : GGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTG : 200
MAXp1  : .....A.....A...........A........A...T.A........A.. : 200

                *       220         *       240         *
WTp1   : TGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACA : 250
MAXp1  : ............A........T..............C..G..G....... : 250

              260         *       280         *
WTp1   : TAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTT : 297
MAXp1  : .C..A.........T.A..A..A..A..A..T............ : 297
```

FIG 1 HIV-1 protease nucleotide sequences of the wild-type (WT) virus, which corresponds to the HIV-1 HXB2 strain (http://www.hiv.lanl.gov), and the synthetic MAX variant, which was generated by a PCR combining three overlapping synthetic DNA oligonucleotides as previously described (10). No substitutions were introduced into the first 40 amino-terminal protease nucleotides overlapping the carboxy terminus of the *gag* p6 reading frame.

excluding the first 40 amino-terminal nucleotides overlapping the carboxy terminus of the *gag* p6 reading frame. These 38 substitutions were chosen in order to improve protease gene codon pair bias (6) without modifying its codon bias or folding free energy (10). The WT and MAX viruses have identical consensus amino acid sequences, and as we previously demonstrated (10), the MAX variant and the WT virus show indistinguishable replication in MT-4 cells (Fig. 2) or peripheral blood mononuclear cells (PBMCs) (10).

We subjected the WT and MAX viruses to selective pressure from two PIs: ATV and DRV. The viruses were propagated in duplicate in MT-4 cells over 32 serial passages (128 days of culture) without drugs or with increasing concentrations of ATV or DRV. The starting PI concentrations were near the half-maximal inhibitory concentration (IC_{50}) for the WT HXB2 virus (10). Before the passages, both viruses showed similar IC_{50}s for ATV and DRV (Table 1) (10). The viruses after 32 serial passages (WTp32 and MAXp32) still showed comparable IC_{50}s in the presence or absence of PIs (Table 1). WTp32 and MAXp32 showed 5-fold and 13-fold increases, respectively, in the IC_{50} for ATV and 6-fold and 10-fold increases, respectively, in the IC_{50} for DRV (Table 1). Although MAXp32 displayed more resistance to ATV and DRV than that of WTp32 (Table 1), these differences were not significant ($P = 0.4816$ and $P = 0.3451$, respectively). These assays demonstrated that the MAX variant virus did not show an impaired capacity to develop phenotypic resistance to PIs.

After 32 cell passages in the presence of ATV or DRV, virus RNA was recovered. This RNA was amplified by reverse transcription-PCR (RT-PCR) and ultradeep sequenced, and we compared the frequencies of resistance mutations. For each of the two studied viruses and the two tested drugs, we sequenced between 1.9×10^7 and 4.1×10^7 individual protease nucleotides (Table 2). Sequence clonal analysis revealed no

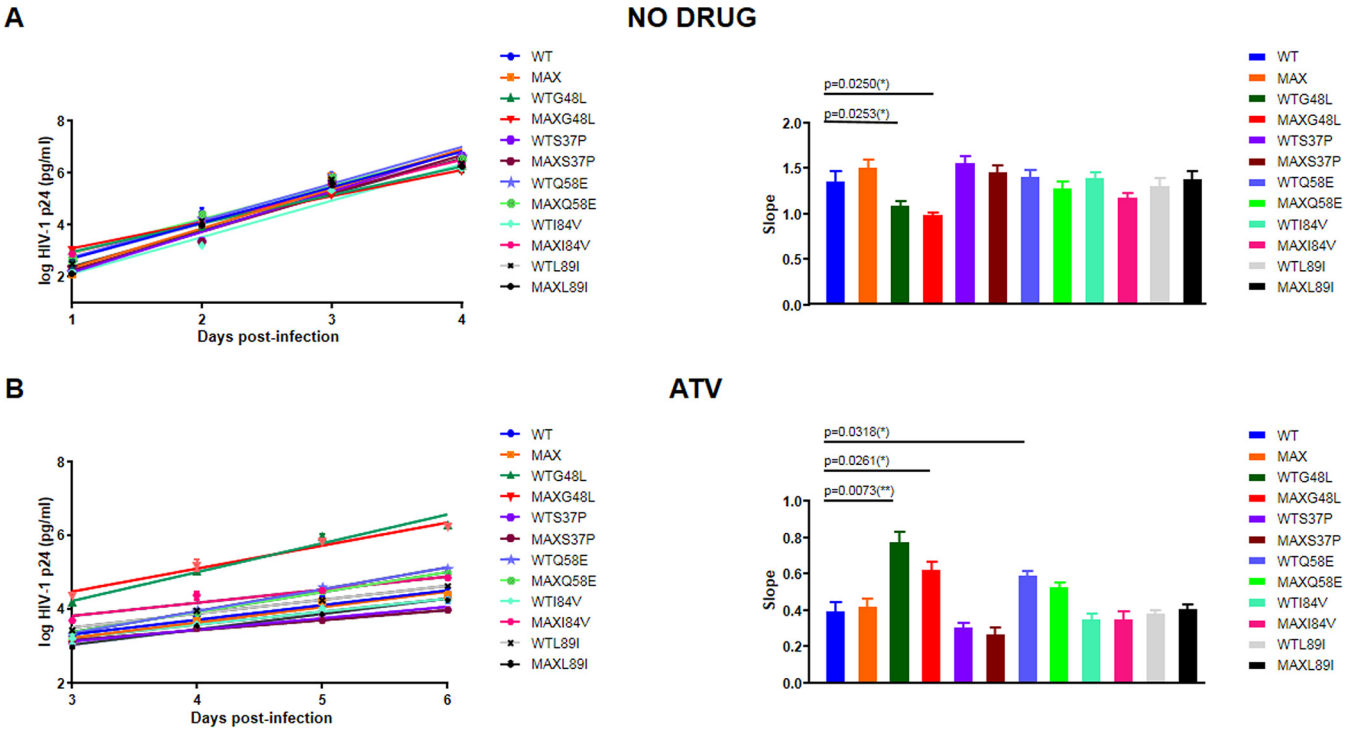

FIG 2 Replication kinetic assay of wild-type HIV-1 (WT) and the recoded MAX protease variant (MAX) in MT-4 cells. HIV-1 antigen p24 concentrations in culture supernatants were measured on days 0 to 4 in the absence of drug and days 0 to 6 in the presence of drug. For each virus, the slope of the plot provides an estimate of the viral replication capacity. Bars show the slope for p24 antigen production from each virus after infection of MT-4 cells. Comparisons of the WT (HXB2) and recoded mutant MAX viruses are shown, as well as comparisons of the corresponding WT and MAX virus variants (i.e., S37P, G48L, Q58E, I84V, and I89L). The significance of the differences between slopes was calculated using an unpaired *t* test with Welch's correction in GraphPrism v. 7 software. (A) Kinetic assays performed in the absence of drug. All slopes values were statistically tested against the WT value. Only WTG48L and MAXG48L displayed a lower replication capacity than that of the WT. (B) Kinetic assays performed in the presence of 20 nM atazanavir (ATV). All slopes values were statistically tested against the WT value. WTG48L, MAXG48L, and WTQ58E displayed a higher replication capacity than that of the WT in the presence of ATV. Values represent the means ± standard deviations (SD) for at least three independent experiments.

resistance-associated substitutions in viruses propagated without drugs (Table 2). On the other hand, both the WT and MAX viruses propagated in the presence of PIs developed previously described resistance mutations to ATV and DRV (Table 2). Moreover, the resistance variant repertoires differed between the MAX and WT viruses (Table 2). Specifically, the G16E substitution was observed only in the WT virus propagated with ATV or DRV. Notably, the WT protease required only a transition to develop this substitution, whereas the MAX protease would need two substitutions: a transition and a transversion. Additionally, the L33F, G48L, Q58E/K, and I89L substitutions were detected only in the recoded MAX protease. Other accompanying substitutions were also detected only in the MAX virus (e.g., E21K, H69Y, and T91S), mainly in the MAX virus (e.g., L10F), or only in the WT virus (e.g., L23I, P39Q, and V77I). Interestingly, some resistance mutations selected by the MAX virus (i.e., G48L and I89L) are extremely rare nonpolymorphic substitutions *in vivo* (38). This finding indicates that the MAX protease may explore a sequence space different from that of the WT protease. Similar to the G16E mutation, the I89L mutation requires two substitutions in the WT background and

TABLE 1 Susceptibilities of HIV-1 strains carrying WT or MAX protease to ATV or DRV

	IC_{50} (nM) (fold change)	
Protease	**ATV**	**DRV**
WTp1	3.1 ± 1.2 (1)	0.9 ± 0.3 (1)
WTp32	14.6 ± 5.3 (5)	5.7 ± 1.0 (6)
MAXp1	1.6 ± 0.1 (1)	0.9 ± 0.2 (1)
MAXp32	21.2 ± 9.0 (13)	9.3 ± 1.9 (10)

TABLE 2 Substitutions associated with protease inhibitor resistance detected after MT-4 cell passages in the presence of ATV, DRV, or no drug

	% of sequences			
	Expt 1		Expt 2	
Drug and substitution	WTp1	MAXp1	WTp1	MAXp1
ATV				
L10F	1.0	28.0	0	0
G16E	1.7	0	1.8	0
L23I	0.8	0	0	0
V32A	76.3	0	53.3	59.9
V32I	10.1	24.0	2.9	23.8
V32T	0.65	0	0	0
L33F	0	0.8	0	0
P39Q	0.9	0	0.9	0
K45I	0	10.2	6.0	0
K45R	0	0.8	0	0
M46I	12.1	0	1.1	10.4
G48L	0	0	0	22.4
I50L	0	28.6	27.3	0
Q58E	0	20.2	0	0
Q58K	0	0	0	0.8
A71V	88.2	23.2	62.8	71.1
V77I	0	0	11.4	0
V82I	0	0	9.8	0
V82D	0	0	0	1.2
I84V	0	43.0	0	0
N88S	14.0	28.2	19.7	3.4
I89L	0	0	0	59.4
T91S	0	3.6	0	0
DRV				
L10F	0	19.2	0	0
G16E	16.5	0	55.5	0
E21K	0	1.4	0	0
A28S	46.1	45.4	6.6	0
L33F	0	94.2	0	89.2
P39Q	0.6	0	0.6	0
K45I	0	0	0	9.8
M46I	6.0	3.9	94.0	14.6
I50L	49.7	52.3	39.1	100
Q58K	0	0	0	1.2
H69Y	0	0.6	0	0
A71V	0	25.2	0	0
V82I	85.2	0.6	34.8	19.7
I84V	3.4	0	0	0
No drug				
D30N	1.4	2.3		
S37P	0	11.7		
P39Q	2.1	0		
P39T	0.8	0		
R41K	0.8	0		
K43R	0	0.5		
Q58K	0	1.1		
I72T	1.8	0		
V82I	1.3	0		

only one substitution in the MAX background. However, there were no obvious reasons for the preferential emergence of the K45I, G48L, Q58E, and I84V substitutions in the MAX background. The L10F, G16E, L33F, Q58E/K, V77I, I84V, and I89L substitutions have previously been associated with PI resistance (39). In contrast, the E21K, L23I, P39Q, G48L, H69Y, and T91S substitutions have not been described as associated with PI resistance.

To explore the favored emergence in the presence of ATV of the G48L, Q58E, I84V, and I89L substitutions in the MAX background and to determine the effects of these

TABLE 3 Susceptibilities of HIV-1 strains carrying WT or MAX protease variants to ATV

Protease	ATV IC_{50} (nM) (fold change)
WT	1.8 ± 1.2 (1)
MAX	1.6 ± 0.4 (0.9)
WTS37P	2.2 ± 1.0 (1.2)
MAXS37P	2.2 ± 1.4 (1.2)
WTG48L	14.4 ± 0.5 (8.1)
MAXG48L	14.4 ± 1.4 (8.1)
WTQ58E	4.6 ± 2.5 (2.6)
MAXQ58E	5.6 ± 3.8 (3.2)
WTI84V	1.3 ± 0.9 (0.7)
MAXI84V	1.5 ± 0.8 (0.8)
WTL89I	1.5 ± 0.5 (0.8)
MAXL89I	2.1 ± 0.6 (1.2)

substitutions in the WT background, we used site-directed mutagenesis to introduce these mutations into both the WT and MAX protease backgrounds. The S37P substitution, preferentially selected in the MAX background when the virus was propagated in the absence of drug (Table 2), was also included in this analysis. We found that the five generated mutants displayed comparable ATV IC_{50}s in both backgrounds (WT and MAX) (Table 3). We also determined the replication capacities of these 10 mutant viruses in MT-4 cells with and without ATV (Fig. 2). Similar to the IC_{50} results, the five tested variants exhibited similar replication capacities in both backgrounds (WT and MAX) in either the absence or presence of 20 nM ATV (Fig. 2). Only the WTG48L and MAXG48L variants showed a lower replication capacity in the absence of drug. Likewise, only the G48L and Q58E substitutions conferred an advantage when viruses were propagated in the presence of 20 nM ATV. These results demonstrated that the G48L, Q58E, I84V, and I89L substitutions are not intrinsically prohibited in the WT protease background and that other factors must explain their low *in vivo* frequency.

The amino acid mutant repertoires also differed between the WT and MAX viruses when they were propagated without drugs (Table 2). Only one variant, the D30N variant, was detected in both virus populations. We do not know whether the observed variants carry adaptive or neutral mutations. Regardless, completely different mutant spectra were detected in these two viruses.

We performed a maximum likelihood phylogenetic reconstruction of all WT and MAX unique amino acid variants that were recovered after 32 MT-4 cell passages in the presence of ATV or DRV. Remarkably, the results showed that the two viruses, which shared an identical starting amino acid sequence, followed different evolutionary trajectories (Fig. 3). Upon visual inspection of these phylogenetic trees, it was also apparent that the MAX protease generated higher amino acid variant diversity (see below).

We next compared the overall population nucleotide diversities of the WT and MAX proteases after 32 passages in the absence or presence of ATV or DRV (Table 4). Overall, nucleotide sequence diversity was significantly higher in MAX populations propagated with ATV (0.00946 ± 0.00005 versus 0.00517 ± 0.00004) ($P < 0.0001$), but no meaningful differences were observed in the presence of DRV (0.00490 ± 0.00004 versus 0.00482 ± 0.00003). As expected, in the presence of a PI, we detected a larger number of nonsynonymous substitutions than synonymous substitutions in both the WT and MAX virus proteases. However, the MAX populations always displayed significantly higher diversity ($P < 0.0001$) in either synonymous or nonsynonymous mutations (Table 4). Diversity was also strikingly higher in MAX populations when viruses were propagated in the absence of drug (0.00095 ± 0.00001 versus 0.00057 ± 0.00001) ($P < 0.0001$) (Table 4). Interestingly, in the absence of drugs, the MAX population showed significantly higher nonsynonymous diversity (0.00117 ± 0.00001 versus 0.00061 ± 0.00001) ($P < 0.0001$) but not synonymous diversity (0.000310 ± 0.00002 versus 0.00053 ± 0.00002) ($P < 0.0001$). This suggested that the WT and MAX viruses are subjected to different selective forces in the absence of pressure from a PI. Compared

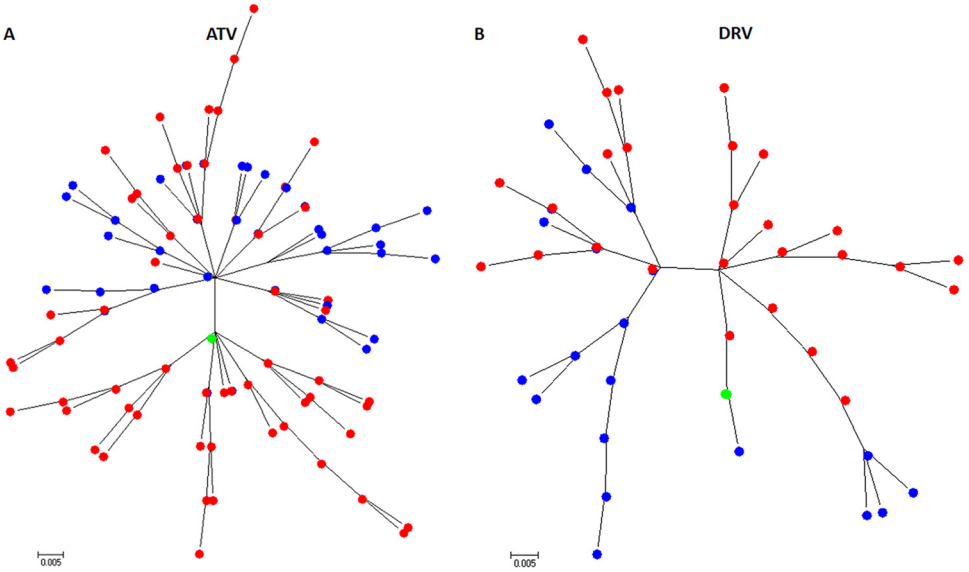

FIG 3 Maximum likelihood phylogram of wild-type (WT) and MAX unique HIV-1 protease amino acid variants selected after 32 passages in MT-4 cells and in the presence of atazanavir (ATV) (A) or darunavir (DRV) (B). Phylogenetic reconstruction was generated using a Jones-Taylor Thornton (JTT) model as implemented in the MEGA6 software package. Both phylogenetic trees show that the WT and MAX viruses, which shared an identical starting amino acid sequence, followed different evolutionary trajectories. Blue and red labels correspond to WT and MAX variants, respectively. Green labels represent the starting HXB2 protease amino acid sequence.

to the WT populations, the MAX populations also showed higher values for Shannon's entropy, another parameter for measuring genetic population diversity (Table 4). Notably, after 32 passages in cell culture, we observed no significant reversions of the starting synonymous substitutions introduced into the MAX protease in either the presence or absence of PIs. Overall, our results demonstrated that MAX viruses displayed higher population genetic diversity after 32 passages in cell culture in the presence or absence of PIs.

DISCUSSION

Previous research shows that codon usage can determine the mutational robustness, evolutionary capacity, and virulence of poliovirus (32). Earlier results indicated that polioviruses with synonymously mutated capsids were less mutationally robust and displayed an attenuated phenotype in an animal model. However, that study did not focus on how synonymous mutations might affect the development of escape mutations to overcome specific selection pressure targeting a precise virus gene. Here we tested the extent to which a synonymously recoded HIV-1 protease reacted to the specific selective pressure of a PI. Our present study also explored the evolvability of a retrovirus which, in contrast to other RNA viruses,

TABLE 4 Summary of population metrics

Protease	No. of bases sequenced	No. (%) of mutations detected	Diversity (*p* distance)			Sn value[a]
			Sequence	Synonymous	Nonsynonymous	
WTp1	8.8×10^6	2,691 (0.03)	0.00047 ± 0.00000	0	0.00062 ± 0.00001	0.04
MAXp1	8.5×10^6	4,257 (0.05)	0.00085 ± 0.00004	0	0.00113 ± 0.00006	0.05
WTp32	7.2×10^6	2,465 (0.03)	0.00057 ± 0.00001	0.00053 ± 0.00002	0.00061 ± 0.00001	0.05
MAXp32	5.2×10^6	2,924 (0.06)	0.00095 ± 0.00001	0.000310 ± 0.00002	0.00117 ± 0.00001	0.06
WTp32 ATV	1.9×10^7	71,206 (0.37)	0.00517 ± 0.00004	0.00016 ± 0.00001	0.0067 ± 0.00003	0.2
MAXp32 ATV	2.5×10^7	180,514 (0.73)	0.00946 ± 0.00005	0.000356 ± 0.00001	0.01252 ± 0.00005	0.27
WTp32 DRV	4.1×10^7	148,066 (0.36)	0.00482 ± 0.00003	0.00016 ± 0.00001	0.00624 ± 0.00003	0.16
MAXp32 DRV	1.9×10^7	95,202 (0.73)	0.00490 ± 0.00004	0.00026 ± 0.00001	0.00756 ± 0.00004	0.17

[a]Shannon's entropy value.

integrates into the host cell genome such that viral proteins are translated from mRNAs by use of the host cellular machinery.

We found that the WT and MAX protease viruses displayed different patterns of resistance mutations after PI treatment. These findings extend those of Lauring et al. (32), confirming that synonymously recoded and WT HIV-1 proteases occupy different sequence spaces. We further demonstrated that although the MAX and WT proteases occupied different sequence spaces, they still showed similar levels of development of phenotypic resistance to PIs. These findings indicate that the recoded protease did not attenuate the virus's capability to develop PI resistance, strongly suggesting that the MAX protease was as robust as the WT protease with regard to this trait. To our knowledge, this is the first study to investigate the evolvability of a synonymously recoded virus enzyme. Even if the resistance selection was conducted in MT-4 cells and the result probably would be qualitatively different in primary cells, our results build on and augment the convincing evidence that recoded proteins occupy a different sequence space.

In some instances, the different mutant repertoire within the MAX protease background can easily be explained by proximity within the corresponding sequence space (e.g., for the G16E and L89I mutations). However, with regard to other mutations (e.g., S37P, G48L, Q58E, and I84V), the explanation for the difference is not readily apparent. In particular, the G48L substitution is very rarely selected *in vivo* in patients undergoing PI therapy (38). However, when the above substitutions (S37P, G48L, Q58E, I84V, and L89I) were introduced into the WT sequence, the resultant variants showed replication capacities parallel to those observed with the MAX background. We can speculate that the introduced synonymous substitutions affected neighboring residues (e.g., RNA structure). However, it must be noted that the MAX and WT proteases have similar RNA folding free energies (10). Since some of these substitutions were observed in only one of the two independent experiments, it cannot be discarded that, in some cases, a mutational stochastic process was involved in the emergence of PI-associated resistance.

One plausible explanation is based on epistatic interactions between protease amino acid substitutions. Epistasis is a phenomenon by which a mutation's impact on protein stability or fitness depends on the genetic background in which it is acquired (40). Complex mutational patterns often arise during the development of resistance to HIV-1 protease inhibitors. More therapy-associated mutations accumulate under PI therapy than under all other types of antiretroviral therapy. Moreover, among patients experiencing therapy failure, the majority of *in vivo* drug-experienced protease sequences include more than four mutations associated with PI therapy (41). Recent findings suggest that the consequences of acquiring primary HIV-1 protease resistance mutations depend on epistatic interactions with the sequence background (42). In our study, over 80% of the MAX protease clones harboring the G48L mutation also had the I89L substitution. As mentioned above, the I89L mutation requires two nucleotide substitutions in the WT background and only one in the MAX background. In either case, our results strongly suggest that the MAX protease's sequence position affects its genotypic PI resistance profile. Synonymous codons differ in their propensity to mutate, and as previously suggested (25, 32, 43), this differential access to the protein sequence space may affect adaptive pathways.

Another intriguing finding of our study is that the MAX virus showed higher population diversity in the recoded and targeted gene following propagation in both the absence and presence of PIs. Again, we speculate that although the MAX virus shows high fitness in tissue culture, it is subjected to greater pressure to change or revert to a WT-synonymous background. However, remarkably, we detected almost no reversions of the synonymous substitutions introduced into the MAX protease following propagation in the presence or absence of PIs. One limitation of our study is that we investigated only one virus enzyme or protein. Further studies should include other virus proteins and other selective pressures (e.g., neutralizing antibodies and cellular virus restriction factors).

It has been suggested that RNA virus synonymous recoding can be used to push a virus to a sequence space region with a low density of neutral mutations (32). Such a lack of access to neutral substitutions may potentially reduce the virus's capacity to generate fit progeny and adaptability to the host's selective pressures, such that this method might serve as a new strategy for development of attenuated vaccines. Our present data suggest that this approach must be developed cautiously and support a need to evaluate the long-term stability of synonymously recoded viruses and to carefully test individual candidates.

MATERIALS AND METHODS

Cell line and viruses. MT-4 cells were obtained from the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program and were grown in RPMI 1640 L-glutamine medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco). The utilized WT virus corresponded to the HIV-1 HXB2 strain (http://www.hiv.lanl.gov) (GenBank accession number K03455). The synthetic MAX HIV-1 protease was generated by PCR amplification with a combination of three overlapping synthetic DNA oligonucleotides, as previously described (10). In MT-4 cells, the MAX protease PCR product was recombined with a protease-deleted HXB2 infectious clone that had previously been linearized with BstEII (44). The protease PCR oligonucleotides used to reconstruct the full-length protease have been described previously (44). WT and MAX S37P, G48L, Q58E, I84V, and I89L mutants were generated by site-directed mutagenesis using overlap extension PCR with mutated oligonucleotides, as previously described (45). Again, the mutant protease PCR products were recombined with the protease-deleted HXB2 infectious clone in MT-4 cells. Cell culture supernatants were harvested at 3, 5, and 7 days posttransfection when the HIV-1 p24 antigen concentration surpassed 500 ng/ml as measured by the Genscreen HIV-1 Ag assay (Bio-Rad). Virus titration was performed in MT-4 cells, and values were expressed as 50% tissue culture infective doses ($TCID_{50}$) as previously described (46).

Replication capacity assays. Viral replication kinetics were analyzed by infecting 1×10^6 MT-4 cells with 200 $TCID_{50}$ (multiplicity of infection [MOI] of 0.0002) of virus. The infected cells were incubated for 4 h at 37°C and 5% CO_2, washed twice with phosphate-buffered saline (PBS), and then resuspended in RPMI medium supplemented with 10% FBS. To quantify viral replication, we measured the HIV-1 capsid p24 antigen concentration in 200-μl aliquots of supernatant collected every 24 h for 4 to 6 days. Growth kinetics were analyzed by fitting a linear model to the log-transformed p24 data during the exponential growth phase, using maximum likelihood methods as previously described (10).

HIV-1 drug susceptibility tests. ATV and DRV were obtained from the NIH AIDS Research and Reference Reagent Program. Following virus propagation and titration, we used a tetrazolium-based colorimetric method to determine the HIV-1 drug susceptibility (IC_{50}) to ATV and DRV in MT-4 cells, using an MOI of 0.003, as previously described (10, 47).

Selection of ATV- and DRV-resistant viruses. WT and MAX viruses were added at an MOI of 0.01 to 1×10^6 MT-4 cells, and the cells were maintained as described above. After 4 days, we transferred 1/10 of the culture, including cells and supernatant, into 1×10^6 fresh MT-4 cells. All virus passages were performed in duplicate. Virus production was monitored by measurements of p24 antigen. The starting concentrations were 4 nM ATV and 3 nM DRV. Throughout the passages, the drug concentrations were increased until they reached 40 nM ATV and 25 nM DRV. In parallel, both viruses were also propagated without either drug.

At passages 1 and 32, 140-μl aliquots of culture supernatant were collected, from which we isolated WT and MAX viral genomic RNAs by using QIAamp viral RNA kits (Qiagen). The purified viral RNAs were then reverse transcribed and PCR amplified using the SuperScript III first-strand synthesis system for RT-PCR (Invitrogen) and 10 pmol (each) of the corresponding protease oligonucleotides, which are described elsewhere (44). Details of this protocol were reported previously (48, 49). These PCR products were the starting material for ultradeep sequencing.

Ultradeep sequencing. Massive parallel sequencing was performed with the MiSeq (Illumina) platform. Libraries of 558-nucleotide (nt) DNA fragments were ligated to the Illumina adapters by use of a Kapa HyperPrep kit (Roche), a SeqCap adapter kit A (Roche), and a SeqCap adapter kit B (Roche). The products were purified using Kapa Pure beads (Roche). All libraries were quantified using a Qubit dsDNA HS assay kit (ThermoFisher) and a Qubit fluorometer (ThermoFisher) and were qualified using an Agilent DNA 1000 kit (Agilent) and a bioanalyzer (Agilent). Sequencing was performed using an Illumina MiSeq reagent kit v3 (600 cycles) (Illumina) following the manufacturer's protocol. Sequencing and paired-end analysis were performed to obtain robust fastq data for bioinformatics analysis. We obtained a mean of 50,000 sequences (reads) per amplicon and per patient sample.

fastq files were obtained from the MiSeq system by index and pool and submitted to FLASH (50). The 2×300 paired-end reads were overlapped to reconstruct the amplicons, with the minimum number of overlapping nucleotides set to 20 and the maximum number of overlapping mismatches set to 10%. The subsequent analysis was performed as previously described (51). Briefly, fastq files were demultiplexed using amplicon oligonucleotides, and oligonucleotides were trimmed at both ends. Each amplicon and strand read was pairwise aligned with respect to the reference WT sequence, insertions were removed, and deletions were repaired if fewer than three gaps were produced. Reads with multiple indeterminations were removed, while reads having a single indetermination were repaired per the reference sequence. Filtered and repaired reads were collapsed into haplotypes with corresponding frequencies. Haplotypes with abundances below 0.1% or that were unique to the forward or reverse strand were

removed. Haplotypes common to the forward and reverse strands and with abundances of ≥0.1% were considered consensus haplotypes, and their frequencies were summed. In the final step to remove artifacts, consensus haplotypes with abundances below 0.5% were filtered out. All computations were performed in the R language and platform, using in-house scripts as well as the packages Biostrings (R package, 2.24.1), Ape (52), and Seqinr (53).

Virus population genetic diversity (*p* distance) was determined using the MEGA6 software package (54). To determine possible selective pressures, the MEGA6 software package was used to calculate the proportion of synonymous substitutions per potential synonymous site and the proportion of nonsynonymous substitutions per potential nonsynonymous site. Shannon's entropy values were calculated as follows: $Sn = -\sum_i (p_i \ln p_i)/\ln N$, where N is the total number of analyzed sequences and p_i is the frequency of each sequence in the viral quasispecies. Sn values vary from 0 (no complexity) to 1 (maximum complexity) (55). The phylogenetic reconstructions were also performed using the MEGA6 software package.

Statistical analysis. Virus population diversity was compared by the unpaired *t* test, using GraphPad Prism, version 7 for Windows. The significance of the differences between replication kinetic slopes and IC_{50}s was calculated using an unpaired *t* test with Welch's correction as implemented in GraphPad Prism.

ACKNOWLEDGMENTS

This study was supported by the Spanish Ministry of Economy and Competitiveness (grant SAF2016-75277-R), the Instituto de Salud Carlos III (grant PI16/00337), and the European Regional Development Fund (ERDF). M.N. was supported by the Instituto de Salud Carlos III through the Spanish AIDS Network (grant RD16/0025/0041). A.J.-P. was supported by a contract from the Spanish Ministry of Economy and Competitiveness (grant BES-2014-069931).

REFERENCES

1. Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. 2014. Exposing synonymous mutations. Trends Genet 30:308–321. https://doi.org/10.1016/j.tig.2014.04.006.
2. Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12:32–42. https://doi.org/10.1038/nrg2899.
3. Martinez MA, Jordan-Paiz A, Franco S, Nevot M. 2016. Synonymous virus genome recoding as a tool to impact viral fitness. Trends Microbiol 24:134–147. https://doi.org/10.1016/j.tim.2015.11.002.
4. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. 2009. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J Virol 83:9957–9969. https://doi.org/10.1128/JVI.00508-09.
5. Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A, Quay J, Kew O. 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol 80:3259–3272. https://doi.org/10.1128/JVI.80.7.3259-3272.2006.
6. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. Science 320:1784–1787. https://doi.org/10.1126/science.1155761.
7. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol 80:9687–9696. https://doi.org/10.1128/JVI.00738-06.
8. Mueller S, Coleman JR, Papamichail D, Ward CB, Nimnual A, Futcher B, Skiena S, Wimmer E. 2010. Live attenuated influenza virus vaccines by computer-aided rational design. Nat Biotechnol 28:723–726. https://doi.org/10.1038/nbt.1636.
9. Yang C, Skiena S, Futcher B, Mueller S, Wimmer E. 2013. Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. Proc Natl Acad Sci U S A 110:9481–9486. https://doi.org/10.1073/pnas.1307473110.
10. Martrus G, Nevot M, Andres C, Clotet B, Martinez MA. 2013. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology 10:78. https://doi.org/10.1186/1742-4690-10-78.
11. Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, Bieniasz PD. 2017. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature 550:124–127. https://doi.org/10.1038/nature24039.
12. Takata MA, Soll SJ, Emery A, Blanco-Melo D, Swanstrom R, Bieniasz PD. 2018. Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. PLoS Pathog 14:e1006824. https://doi.org/10.1371/journal.ppat.1006824.
13. Vabret N, Bailly-Bechet M, Lepelley A, Najburg V, Schwartz O, Verrier B, Tangy F. 2014. Large-scale nucleotide optimization of simian immunodeficiency virus reduces its capacity to stimulate type I interferon in vitro. J Virol 88:4161–4172. https://doi.org/10.1128/JVI.03223-13.
14. Nougairede A, De Fabritus L, Aubry F, Gould EA, Holmes EC, de Lamballerie X. 2013. Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. PLoS Pathog 9:e1003172. https://doi.org/10.1371/journal.ppat.1003172.
15. Le Nouen C, Brock LG, Luongo C, McCarty T, Yang L, Mehedi M, Wimmer E, Mueller S, Collins PL, Buchholz UJ, DiNapoli JM. 2014. Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. Proc Natl Acad Sci U S A 111:13169–13174. https://doi.org/10.1073/pnas.1411290111.
16. Le Nouen C, McCarty T, Brown M, Smith ML, Lleras R, Dolan MA, Mehedi M, Yang L, Luongo C, Liang B, Munir S, DiNapoli JM, Mueller S, Wimmer E, Collins PL, Buchholz UJ. 2017. Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure. Proc Natl Acad Sci U S A 114:E386–E395. https://doi.org/10.1073/pnas.1619242114.
17. Meng J, Lee S, Hotard AL, Moore ML. 2014. Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. mBio 5:e01704-14. https://doi.org/10.1128/mBio.01704-14.
18. Ni YY, Zhao Z, Opriessnig T, Subramaniam S, Zhou L, Cao D, Cao Q, Yang H, Meng XJ. 2014. Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. Virology 450–451:132–139. https://doi.org/10.1016/j.virol.2013.12.009.
19. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. 2014. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res 42:4527–4545. https://doi.org/10.1093/nar/gku075.
20. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. 2014. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife 3:e04531. https://doi.org/10.7554/eLife.04531.
21. de Fabritus L, Nougairede A, Aubry F, Gould EA, de Lamballerie X. 2015. Attenuation of tick-borne encephalitis virus using large-scale random codon re-encoding. PLoS Pathog 11:e1004738. https://doi.org/10.1371/journal.ppat.1004738.

22. Shen SH, Stauft CB, Gorbatsevych O, Song Y, Ward CB, Yurovsky A, Mueller S, Futcher B, Wimmer E. 2015. Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc Natl Acad Sci U S A 112:4749–4754. https://doi.org/10.1073/pnas.1502864112.
23. Sitaraman V, Hearing P, Ward CB, Gnatenko DV, Wimmer E, Mueller S, Skiena S, Bahou WF. 2011. Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. Proc Natl Acad Sci U S A 108:14294–14299. https://doi.org/10.1073/pnas.1102883108.
24. Cladel NM, Budgeon LR, Hu J, Balogh KK, Christensen ND. 2013. Synonymous codon changes in the oncogenes of the cottontail rabbit papillomavirus lead to increased oncogenicity and immunogenicity of the virus. Virology 438:70–83. https://doi.org/10.1016/j.virol.2013.01.005.
25. Dolan PT, Whitfield ZJ, Andino R. 2018. Mapping the evolutionary potential of RNA viruses. Cell Host Microbe 23:435–446. https://doi.org/10.1016/j.chom.2018.03.012.
26. Andino R, Domingo E. 2015. Viral quasispecies. Virology 479–480:46–51. https://doi.org/10.1016/j.virol.2015.03.022.
27. Mas A, Lopez-Galindez C, Cacho I, Gomez J, Martinez MA. 2010. Unfinished stories on viral quasispecies and Darwinian views of evolution. J Mol Biol 397:865–877. https://doi.org/10.1016/j.jmb.2010.02.005.
28. Song Y, Liu Y, Ward CB, Mueller S, Futcher B, Skiena S, Paul AV, Wimmer E. 2012. Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. Proc Natl Acad Sci U S A 109:14301–14307. https://doi.org/10.1073/pnas.1211484109.
29. Li M, Kao E, Gao X, Sandig H, Limmer K, Pavon-Eternod M, Jones TE, Landry S, Pan T, Weitzman MD, David M. 2012. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature 491:125–128. https://doi.org/10.1038/nature11433.
30. Vabret N, Bhardwaj N, Greenbaum BD. 2017. Sequence-specific sensing of nucleic acids. Trends Immunol 38:53–65. https://doi.org/10.1016/j.it.2016.10.006.
31. Shin YC, Bischof GF, Lauer WA, Desrosiers RC. 2015. Importance of codon usage for the temporal regulation of viral gene expression. Proc Natl Acad Sci U S A 112:14030–14035. https://doi.org/10.1073/pnas.1515387112.
32. Lauring AS, Acevedo A, Cooper SB, Andino R. 2012. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe 12:623–632. https://doi.org/10.1016/j.chom.2012.10.008.
33. Lauring AS, Frydman J, Andino R. 2013. The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol 11:327–336. https://doi.org/10.1038/nrmicro3003.
34. Eigen M. 1971. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465–523. https://doi.org/10.1007/BF00623322.
35. Domingo E, Sabo D, Taniguchi T, Weissmann C. 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell 13:735–744. https://doi.org/10.1016/0092-8674(78)90223-4.
36. van der Kuyl AC, Berkhout B. 2012. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. Retrovirology 9:92. https://doi.org/10.1186/1742-4690-9-92.
37. Nevot M, Jordan-Paiz A, Martrus G, Andrés C, García-Cehic D, Gregori J, Franco S, Quer J, Martinez MA. 2018. HIV-1 protease evolvability is affected by synonymous nucleotide recoding. bioRxiv https://doi.org/10.1101/315366.
38. Gifford RJ, Liu TF, Rhee SY, Kiuchi M, Hue S, Pillay D, Shafer RW. 2009. The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. Bioinformatics 25:1197–1198. https://doi.org/10.1093/bioinformatics/btp134.
39. Wensing AM, Calvez V, Gunthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW, Richman DD. 2017. 2017 update of the drug resistance mutations in HIV-1. Top Antivir Med 24:132–133.
40. Parera M, Martinez MA. 2014. Strong epistatic interactions within a single protein. Mol Biol Evol 31:1546–1553. https://doi.org/10.1093/molbev/msu113.
41. Shafer RW, Schapiro JM. 2008. HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. AIDS Rev 10:67–84.
42. Flynn WF, Haldane A, Torbett BE, Levy RM. 2017. Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. Mol Biol Evol 34:1291–1306. https://doi.org/10.1093/molbev/msx095.
43. Cambray G, Mazel D. 2008. Synonymous genes explore different evolutionary landscapes. PLoS Genet 4:e1000256. https://doi.org/10.1371/journal.pgen.1000256.
44. Maschera B, Furfine E, Blair ED. 1995. Analysis of resistance to human immunodeficiency virus type 1 protease inhibitors by using matched bacterial expression and proviral infection vectors. J Virol 69:5431–5436.
45. Parera M, Clotet B, Martinez MA. 2004. Genetic screen for monitoring severe acute respiratory syndrome coronavirus 3C-like protease. J Virol 78:14057–14061. https://doi.org/10.1128/JVI.78.24.14057-14061.2004.
46. Pauwels R, Balzarini J, Baba M, Snoeck R, Schols D, Herdewijn P, Desmyter J, De Clercq E. 1988. Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds. J Virol Methods 20:309–321. https://doi.org/10.1016/0166-0934(88)90134-6.
47. Betancor G, Alvarez M, Marcelli B, Andres C, Martinez MA, Menendez-Arias L. 2015. Effects of HIV-1 reverse transcriptase connection subdomain mutations on polypurine tract removal and initiation of (+)-strand DNA synthesis. Nucleic Acids Res 43:2259–2270. https://doi.org/10.1093/nar/gkv077.
48. Capel E, Martrus G, Parera M, Clotet B, Martinez MA. 2012. Evolution of the human immunodeficiency virus type 1 protease: effects on viral replication capacity and protease robustness. J Gen Virol 93:2625–2634. https://doi.org/10.1099/vir.0.045492-0.
49. Fernandez G, Clotet B, Martinez MA. 2007. Fitness landscape of human immunodeficiency virus type 1 protease quasispecies. J Virol 81:2485–2496. https://doi.org/10.1128/JVI.01594-06.
50. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507.
51. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, Alvarez-Tejado M, Rodriguez-Frias F, Guardia J, Domingo E, Quer J. 2013. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. PLoS One 8:e83361. https://doi.org/10.1371/journal.pone.0083361.
52. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412.
53. Charif D, Thioulouse J, Lobry JR, Perriere G. 2005. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics 21:545–547. https://doi.org/10.1093/bioinformatics/bti037.
54. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
55. Wolinsky SM, Kunstman KJ, Safrit JT, Koup RA, Neumann AU, Korber BT. 1996. Response: HIV-1 evolution and disease progression. Science 274:1010–1011. https://doi.org/10.1126/science.274.5289.1010.