bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Quantifying the evolutionary potential and constraints of a drug-targeted viral protein

Lei Dai[1, 2]*, Yushen Du[1]*, Hangfei Qi[1], Nicholas C. Wu[1,3], James O. Lloyd-Smith[2], Ren Sun[1]

[1]Department of Molecular and Medical Pharmacology

[2]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, United States

[3]Current address: Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, United States

*: these authors contributed equally to this study

Abstract

RNA viruses are most notorious for their ability to evolve rapidly under novel environments. It is known that the high mutation rate of RNA viruses can generate huge genetic diversity to facilitate viral adaptation. However, less attention has been paid to the underlying fitness landscape that represents the selection forces on viral genomes. Here we systematically quantified the distribution of fitness effects (DFE) of single amino acid substitutions (86 amino acids total) in the drug-targeted region of NS5A protein of Hepatitis C Virus (HCV). We found that the majority of non-synonymous substitutions incur large fitness costs, suggesting that NS5A protein is highly optimized in natural conditions. Furthermore, we characterized the evolutionary potential of HCV by subjecting the mutant viruses to varying concentrations of an NS5A inhibitor Daclatasvir. As the selection pressure increases, the DFE of beneficial mutations shifts from an exponential distribution to a heavy-tailed distribution with a disproportionate number of exceptionally fit mutants. The number of available beneficial mutations and the selection coefficient are both found to increase at higher levels of antiviral drug concentration, as predicted by a pharmacodynamics model describing viral fitness as a function of drug concentration. Our large-scale fitness data of mutant viruses also provide insights into the biophysical basis of evolutionary constraints and the role of the genetic code in protein evolution.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In our evolutionary battles with microbial pathogens, RNA viruses are among the most formidable foes. HIV-1 and Hepatitis C Virus acquire drug resistance in patients under antiviral therapy. Influenza and Ebola virus cross the species barrier to infect human hosts. Understanding the evolution of RNA viruses is therefore of paramount importance for developing antivirals and vaccines and assessing the risk of future emergence events (Domingo et al. 2012; Goldberg et al. 2012; Metcalf et al. 2015; Geoghegan et al. 2016). Comprehensive characterization of viral fitness landscapes, and the principles underpinning them, will provide us with a map of evolutionary pathways accessible to RNA viruses and guide our design of effective strategies to limit antiviral resistance, immune escape and cross-species transmission.

Although the concept of fitness landscapes has been around for a long time (Wright 1932), we still know little about their properties in real biological systems. Previous empirical studies of fitness landscapes have been constrained by very limited sampling of sequence space. In a typical study, mutants are generated by site-directed mutagenesis and assayed for growth rate individually. We and others have recently developed a high-throughput technique, often referred to as “deep mutational scanning”, to profile the fitness effect of mutations by integrating deep sequencing with selection experiments (Wu et al. 2013; Fowler and Fields 2014; Thyagarajan and Bloom 2014). This novel application of next generation sequencing has raised an exciting prospect of large-scale fitness measurements (Jacquier and Birgy 2013; Qi et al. 2014; Wu et al. 2014; Li et al. 2016; Puchta et al. 2016) and a revolution in our understanding of molecular evolution (He et al. 2016).

The distribution of fitness effects (DFE) of mutations is a fundamental entity in genetics and reveals the local structure of a fitness landscape (Burch and Chao 2000; Eyre-Walker and Keightley 2007; Hietpas et al. 2011; Desai et al. 2013; Jacquier et al. 2013; Bataillon and Bailey 2014; Bank et al. 2015; Chevereau et al. 2015). The deleterious mutations are usually abundant and impose severe constraints on the accessibility of fitness landscapes. In contrast, beneficial mutations are rare and provide the raw materials of adaptation. Quantifying the DFE of RNA viruses is crucial for understanding how the pathogens evolve to acquire drug resistance, evade the host immune system, cross species barriers, and surmount other evolutionary challenges.

The model system used in our study is Hepatitis C Virus (HCV), a positive sense single-stranded RNA virus with a genome of ~9.6 kb. The biology of HCV has been studied extensively in the past two decades and provides an excellent model system of human RNA viruses. We applied high-throughput fitness assays

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to map the fitness effects of all single amino acid substitutions in domain IA (amino acid 18-103) of HCV NS5A protein (Methods). This domain is the target of several directly-acting antiviral drugs, including Daclatasvir (DCV) (Gao et al. 2010; Chayama et al. 2012; Wang et al. 2012; Everson et al. 2014; Kumada et al. 2014; Wang et al. 2014; Hézode et al. 2015; Poordad et al. 2015; Yoshimi and Imamura 2015). We profiled the DFE of HCV NS5A protein under varying levels of positive selection by tuning the concentration of the antiviral drug DCV. In addition, we studied how viral evolution is constrained by deleterious mutations that impact protein stability. Finally, we analyzed the shape of DFE in nucleotide sequence space and analyzed how the structure of genetic code influences protein evolution.

Results

Profiling the fitness landscape of HCV NS5A protein

To study the DFE of mutations of HCV NS5A protein, we constructed a library of mutant viruses using saturation mutagenesis (Qi et al. 2014). Briefly, each codon in the mutated region was randomized to cover all possible single amino acid substitutions. We observed 2520 non-synonymous mutations in the plasmid library, which covered 99.6% (1628 out of 1634) of all possible single amino acid substitutions, as well as 105 synonymous mutations. After transfection, we performed selection on the mutant viruses in an HCV cell culture system (Lindenbach et al. 2005). Mutants with frequency below a certain cutoff after transfection were assigned as lethal mutations (Methods). The relative fitness of a mutant virus to the wild-type virus was calculated based on the changes in frequency of the mutant virus and the wild-type virus after one round of selection in cell culture (Supplementary Figure 1).

Our experiment provides a comprehensive profiling of the local fitness landscape of all single amino acid mutations. As expected, the fitness effects of synonymous mutations were nearly neutral, while most non-synonymous mutations were deleterious (Figure 1). After grouping together non-synonymous mutations leading to the same amino acid substitution, we found that around 90% of single amino acid mutations had fitness costs and almost half of them were found to be lethal (Supplementary Figure 2). This high sensitivity to mutation in HCV NS5A, which is an essential viral protein, is generally consistent with previous mutagenesis studies of RNA viruses (Sanjuán et al. 2004). Our data support the view that RNA viruses are very sensitive to the effect of deleterious mutations, possibly due to the compactness of their genomes (Elena et al. 2006; Rihn et al. 2013).

Using the distribution of fitness effects of synonymous mutations as a benchmark, we identified that only 3.4% of single amino acid mutations are beneficial (Methods). The estimated fraction of beneficial mutations

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is consistent with previous small-scale mutagenesis studies in viruses including bacteriophages, vesticular stomatitis virus, etc. (Ickinson 1998; Sanjuán et al. 2004; Burch et al. 2007; Silander et al. 2007). Our results indicate that HCV NS5A protein is under strong purifying selection, suggesting that viral proteins are highly optimized in their natural conditions.

Evolutionary potential as a function of positive selection

Beneficial mutations are the raw materials of protein adaptation (Eyre-Walker and Keightley 2007; Rodríguez-Verdugo et al. 2013). Previous studies have found that the evolvability of proteins is a function of purifying selection (Stiffler et al. 2015). In this study, we aim to study the role of positive selection in modulating the evolutionary potential of drug-targeted viral proteins. In addition to growing viruses in the natural condition without drugs, we selected the mutant library in 10, 40 and 100pM of a potent HCV NS5A inhibitor Daclatasvir (DCV). The drug concentrations were chosen based on in vitro IC_{50} of wild type HCV virus (~20pM) to represent different levels of positive selection (mild, intermediate and strong).

By tuning the concentration of DCV, we observed a shift in the DFE of beneficial mutations (Figure 2A, Supplementary Figure 3). At higher drug concentrations, we observed an increase in the average selection coefficient as well as the total number of beneficial mutations (Table 1). We further tested if the shape of this distribution changed under drug selection. Previous empirical studies supported the hypothesis that the DFE of beneficial mutations is exponential (Imhof and Schlötterer 2001; Sanjuán et al. 2004; Cowperthwaite et al. 2005; Rokyta et al. 2005; Kassen and Bataillon 2006; Burch et al. 2007; Carrasco et al. 2007; MacLean and Buckling 2009; Peris et al. 2010; Bataillon et al. 2011; Gutierrez et al. 2013). Following a maximum likelihood approach, we fit the DFE of beneficial mutations to the Generalized Pareto Distribution (Supplementary Figure 4, Methods). The fitted distribution (Table 1) is described by two parameters: a scale parameter (τ), and a shape parameter (κ) that determines the tail behavior of the distribution. Using a likelihood-ratio test (Beisel et al. 2007), we found that the distribution was exponential ($\kappa=0$) in the natural condition without drug selection, but shifted to a heavy-tailed distribution ($\kappa>0$) in the presence of DCV, where the wild type virus was poorly adapted. Our observation suggests that the DFE of beneficial mutations is dependent on how well adapted the wild type is in a certain environment (Eyre-Walker and Keightley 2007). When individuals are subject to novel environments (i.e. novel forms of selection pressure) (Rokyta et al. 2008; Schenk et al. 2012), the fitness of the wild type is no longer top-ranking and the DFE is expected to deviate from the exponential distribution (MacLean et al. 2010).

A simple pharmacodynamics model describing viral fitness as a function of drug concentration (i.e. phenotype-fitness mapping) can explain the changing spectra of beneficial mutations upon drug treatment

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 2B). For example, mutations that reduce the protein's binding affinity to drug molecules (i.e. with a higher inhibitory concentration than wild-type) may come with a fitness cost, i.e. fitness trade-offs (Wu et al. 2013). Thus, a drug-resistant mutant that is deleterious in the absence of drug may become beneficial under drug selection, leading to an increase in the number of beneficial mutations. Moreover, the relative fitness of the drug-resistant mutant is expected to increase with stronger selection pressure (Figure 2B, dashed line).

We measured the dose response curves for a set of mutants constructed by site-directed mutagenesis (Supplementary Figure 5). Indeed, we found that the relative fitness of drug-resistant mutants increased at higher drug concentration (Figure 2C); in contrast, the drug-sensitive mutant became less fit under drug selection. Furthermore, based on this set of mutants with validated dose response curves, we were able to use the fitness measurements to estimate the IC_{50} of all mutants in our library (Supplementary Figure 6, Methods). In particular, we found that a small group of mutations were highly resistant to DCV and this could explain the heavy-tailed DFE of beneficial mutations under drug selection. Overall, our results suggest that the evolutionary potential of proteins is modulated by the strength of positive selection, in addition to the previous findings on the role of purifying selection (Stiffler et al. 2015).

Deleterious mutations as evolutionary constraints

While beneficial mutations open up adaptive pathways to genotypes with higher fitness, mutations that reduce fitness impose constraints on the evolution of viruses. To understand the biophysical basis of mutational effects (Liberles et al. 2012), we took advantage of the available structural information. The crystal structure of NS5A domain I is available excluding the amphipathic helix at N-terminus (Tellinghuisen et al. 2005; Love et al. 2009).

We found that the fitness effects of deleterious mutations at buried sites (i.e. with lower solvent accessibility) were more pronounced than those at surface exposed sites (Figure 3A, Supplementary Figure 7). Moreover, we performed simulations of protein stability for individual mutants using the Rosetta program (Das and Baker 2008). A mutation with $\Delta\Delta G>0$, i.e. shifting the free energy difference to favor the unfolded state, is expected to destabilize the protein. We found that mutations that destabilize protein stability reduced the viral fitness (Figure 3B). For example, mutations at a stretch of highly conserved residues (F88-N91) that run through the core of NS5A protein tend to destabilize the protein and significantly reduce the viral fitness (Supplementary Figure 8). Mutations that increase $\Delta\Delta G$ beyond a threshold were mostly lethal, which is consistent with the prediction that proteins become unfolded after using up the stability margin (Bloom et al. 2005; Wylie and Shakhnovich 2011; Olson et al. 2014). In contrast, mutations at some sites were highly

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deleterious despite having little impact on protein stability, suggesting that evolution at these sites may be under additional constraints to preserve protein function (Wu et al. 2015; Echave et al. 2016; Jack et al. 2016), such as RNA binding (Foster et al. 2010; Hwang et al. 2010).

We further tested whether the viral replication fitness in cell culture was predictive of evolutionary landscapes of viruses in patients (Ferguson et al. 2013). This is critical for the extrapolation of viral replication fitness in in vitro to in vivo (Hart and Ferguson 2015). We analyzed sequence diversity of HCV sequences in the database of Los Alamos National Lab (Methods). Indeed, we found that the within-patient sequence diversity was highly correlated to the replication fitness measured in cell culture (Spearman's ρ=0.82, Figure 3C), suggesting that fitness landscapes profiled in laboratory settings can provide insights on evolutionary pathways of viruses in nature (Gong et al. 2013).

The role of the genetic code in protein evolution

So far we have considered the spectrum of beneficial and deleterious mutations in the amino acid sequence space (Wu et al. 2016). In fact, the evolution of viral proteins in the nucleotide sequence space face additional constraints posed by the genetic code, as most mutations in RNA viruses occur as point mutations during genome replication.

Due to the degeneracy in the genetic code, many point mutations are synonymous and thus less likely to be deleterious than 2 or 3-nt substitutions. Even for non-synonymous mutations, we found that the deleterious impacts on fitness increased with the number of nucleotide substitutions (Figure 4A, B), supporting the hypothesis that the structure of the standard genetic code can buffer the mutational load (Firnberg and Ostermeier 2013). This observation is consistent with the facts that: 1) amino acids with similar biochemical properties tend to be adjacent in the genetic code (Yampolsky and Stoltzfus 2005); and 2) mutating to biochemically different amino acids is more likely to decrease fitness (Supplementary Figure 2).

In addition, we observed a slight (although not statistically significant) enrichment of beneficial mutations in point mutations under the natural condition (Figure 4C). Under drug selection, our conclusions on the shifting DFE of beneficial single amino acid mutations held true for point mutations (Supplementary Figure 9, Supplementary Table 1). Furthermore, we observed that beneficial mutations were significantly depleted in 3-nt substitutions (Supplementary Figure 10). However, this difference in the fraction of beneficial mutations could be confounded by the fact that 3-nt substitutions are more likely to be lethal. As discussed in previous studies (Firnberg and Ostermeier 2013), mutational robustness and the enrichment of beneficial mutations may actually be two sides of the same coin.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Mutation accumulation (Levy et al. 2015) and site-directed mutagenesis (Visher et al. 2016) are traditional approaches to examine the DFE. Both methods provide pivotal insights into the shape of DFE, however, both methods also have limitations. Under these methods, the sampling of mutations is sparse and/or biased (e.g. large-effect beneficial mutations are more likely to be observed in mutation accumulation). The "deep mutational scanning" approach (Wu et al. 2013; Fowler and Fields 2014), which utilizes next-generation sequencing to assay fitness or phenotype, allows for unbiased and large-scale sampling of fitness landscapes and is ideal for studying DFEs.

The shape of the DFE determines mutational robustness (Visser et al. 2003; Draghi et al. 2010; Visher et al. 2016), evolutionary accessibility and repeatability (de Visser and Krug 2014). One often overlooked aspect is that DFEs will vary as a function of selection pressure (Martin and Lenormand 2006; Lalić et al. 2011; Stiffler et al. 2015). For example, mutations that impair function would become more deleterious with increasing pressure of purifying selection, thus leading to reduced protein evolvability (Stiffler et al. 2015). In this study, we have focused on gain-of-function mutations in a novel environment. The pleiotropic effect of mutations would cause the spectra of beneficial mutations to shift between the natural condition and the condition with drug selection. Moreover, mutations enabling the new function (e.g. drug resistance) would become more beneficial with increasing pressure of positive selection. Although different systems have distinct protein-drug interactions (e.g. resistance profiles, genetic barrier of resistance) (Robinson et al. 2011), the results in our study provide a general framework to study DFEs of drug-targeted proteins across different environments. Future studies along this line will further our understanding of how proteins evolve new functions under the constraints to maintain the original function (Soskine and Tawfik 2010), as exemplified in the evolution of resistance to directly-acting antivirals (Rosenbloom et al. 2012).

Synonymous mutations are usually expected to have no or minimal influence on phenotype or fitness, but this view is being challenged as their effect on protein expression and folding becomes increasingly elucidated (e.g. via mRNA secondary structures or codon usage) (Yang et al. 2014; Agashe et al. 2016). One interesting observation in our selection experiments is that some synonymous mutations seemed to have phenotypic effects on drug sensitivity (Supplementary Figure 11). Although this is not the focus of our study, understanding the mechanism of natural selection at the RNA level and its implications for molecular evolution, particularly in the context of RNA viruses, will be a fruitful area for future studies.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusions

Many RNA viruses adapt rapidly to novel selection pressure, such as antiviral drugs. Understanding how pathogens evolve under drug selection is critical for the success of antiviral therapy against human pathogens. By combining deep sequencing with selection experiments in cell culture, we have quantified the distribution of fitness effects of mutations in the drug-targeted domain of Hepatitis C Virus NS5A protein. Our results indicate that the majority of single amino acid substitutions in NS5A protein incur large fitness costs. Combined with stability predictions based on protein structure, our fitness data reveal the biophysical constraints underlying the evolution of viral proteins. Furthermore, by subjecting the mutant viruses to positive selection under an antiviral drug, we find that the evolutionary potential of viral proteins in a novel environment is modulated by the strength of selection pressure.

Materials and Methods

Mutagenesis

The mutant library of HCV NS5A protein domain IA (86 amino acids) was constructed using saturation mutagenesis as previously described (Qi et al. 2014). In brief, the entire region was divided into five sub-libraries each containing 17-18 amino acids. NNK (N: A/T/C/G, K: T/G) was used to replace each amino acid. The oligos, each of which contains one random codon, were synthesized by IDT. The mutated region was ligated to the flanking constant regions, subcloned into the pFNX-HCV plasmid and then transformed into bacteria. The pFNX-HCV plasmid carrying the viral genome was synthesized in Dr. Ren Sun's lab based on the chimeric sequence of genotype 2a HCV strains J6/JFH1.

Cell culture

The human hepatoma cell line (Huh-7.5.1) was provided by Dr. Francis Chisari from the Scripps Research Institute, La Jolla. The cells were cultured in T-75 tissue culture flasks (Genesee Scientific) at 37 ^{o}C with 5% CO_2. The complete growth medium contained Dulbecco's Modified Eagle's Medium (Corning Cellgro), 10% heat inactivated Fetal Bovine Serum (Omega Scientific), 10mM HEPES (Life Technologies), 1x MEM Non-Essential Amino Acids Solution (Life Technologies) and 1x Penicillin-Streptomycin-Glutamine (Life Technologies).

Selection

Plasmid mutant library was transcribed in vitro using T7 RiboMAX Express Large Scale RNA Production System (Promega) and purified by PureLink RNA Mini Kit (Life Technologies). 10 µg of in vitro transcribed RNA was used to transfect 4 million Huh-7.5.1 cells via electroporation by Bio-Rad Gene Pulser. The

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

supernatant was collected 144 hours post transfection and virus titer was determined by immunofluorescence assay. The viruses collected after transfection were used to infect ~2 million Huh-7.5.1 cells at MOI around 0.1-0.2. The five mutant libraries were passaged for selection separately as previously described(Qi et al. 2015). For the three different levels of selection pressure, the growth media was supplemented with 10pM, 40pM and 100pM HCV NS5A inhibitor Daclatasvir (BMS-790052), respectively. The supernatant was collected 144 hours post infection.

Preparation of Illumina sequencing samples

For each sample, viral RNA was extracted from 700 µl supernatant collected after transfection and after selection using QIAamp Viral RNA Mini Kit (Qiagen). Extracted RNA was reverse transcribed into cDNA by SuperScript III Reverse Transcriptase Kit (Life Technologies). The targeted region in NS5A (51-54bp) was PCR amplified using KOD Hot Start DNA polymerase (Novagen). The Eppendorf thermocycler was set as following: 2 min at 95 °C; 25 to 35 three-step cycles of 20 s at 95 °C,15 s at 52-56 °C (sub-library #1, 52 °C; #2, 52 °C; #3, 52 °C; #4, 56 °C; #5, 54 °C) and 25s at 68 °C; 1 min at 68 °C. The number of PCR cycles are chosen based on the copy number of cDNA templates as determined by qPCR (Bio-Rad). The PCR primers are listed in Table S1. The PCR products were purified using PureLink PCR Purification Kit (Life Technologies) and prepared for Illumina Hiseq 2000 sequencing (paired-end 100 bp) following the standard protocol. Each sample was tagged with a unique 3-bp customized barcodes (Table S2).

Analysis of sequencing data

The sequencing data were parsed by SeqIO function of BioPython. The reads from different samples were de-multiplexed by the barcodes and mapped to the entire mutated region in NS5A by allowing at maximum 5 mismatches with the reference genome (Table S3)(Qi et al. 2014). Since both forward and reverse reads cover the whole amplicon, we used paired reads to correct for sequencing errors. A mutation was called only if it was observed in both reads and the quality score at the corresponding position was at least 30. Sequencing reads containing mutations not supposed to appear in the mutant library were excluded from downstream analysis. The sequencing depth for each sub-library is at least ~10^5 and two orders of magnitude higher than the library complexity.

Calculation of relative fitness

For each condition of selection experiments (i.e. different concentration of Daclatasvir [DCV]), the relative fitness (RF) of a mutant virus to the wild-type virus is calculated by the relative changes in frequency after selection,

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$RF_{mut}([DCV]) = \left(\frac{f_{mut}^{T=1}}{f_{mut}^{T=0}}\right) \Big/ \left(\frac{f_{WT}^{T=1}}{f_{WT}^{T=0}}\right)$$

where $f_{mut}^{T=round}$ and $f_{WT}^{T=round}$ is the frequency of the mutant virus and the wild-type virus at round 0 (after transfection) or round 1 (after selection). The fitness of wild-type virus is normalized to 1. The fitness values estimated from one round have been shown to be highly consistent to estimates from multiple rounds of selection (Qi et al. 2014).

Mutants with less than 10 read counts in the plasmid library were filtered. A mutation was considered lethal if one of the two criteria were met: 1) after transfection, the mutant had less than 10 read counts; 2) after transfection, the ratio between the mutant's frequency and the wild-type's frequency was smaller than 10^{-4}. The thresholds for beneficial and deleterious mutations are defined as $1+2\sigma_{silent}$ and $1-2\sigma_{silent}$, respectively. σ_{silent} is the standard deviation of the fitness effects of synonymous mutations under the natural condition (Figure 1). The fitness effect of non-synonymous mutations leading to the same amino acid substitution are averaged over to estimate the fitness effect of single amino acid substitutions.

Fitting the distribution of fitness effects of beneficial mutations

The distribution of selection coefficients of beneficial mutations were fitted to a Generalized Pareto Distribution following a maximum likelihood approach (Beisel et al. 2007),

$$F(x|\kappa,\tau) = \begin{cases} 1-(1+\frac{\kappa}{\tau}x)^{-1/\kappa}, x \geq 0, \ \ if\, \kappa > 0 & \text{(Frechet)} \\ 1-(1+\frac{\kappa}{\tau}x)^{-1/\kappa}, 0 \leq x < -\frac{\tau}{\kappa}, \ \ if\, \kappa < 0 & \text{(Weibull)} \\ 1-e^{-x/\tau}, x \geq 0, \ \text{if}\ \kappa = 0 & \text{(Gumbel)} \end{cases}$$

Only mutations with relative fitness higher than the beneficial threshold $1+2\sigma_{silent}$ were included in the distribution of beneficial mutations. The selection coefficients were normalized to the beneficial threshold. The shape parameter κ determines the tail behavior of the distribution, which can be divided into three domains of attraction: Gumbel domain (exponential tail, κ = 0), Weibull domain (truncated tail, κ < 0) and Fréchet domain (heavy tail, κ > 0). For each selection condition, a likelihood ratio test is performed to evaluate whether the null hypothesis κ=0 (exponential distribution) can be rejected.

Estimation of IC_{50} from fitness data

We can quantify the drug resistance of each mutant in the library by computing its fold change in relative

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fitness,

$$W([DCV]) = \frac{RF_{mut}([DCV])}{RF_{mut}}$$

Here RF_{mut} is the relative fitness of a mutant under the natural condition (i.e. no drug). W is the fold change in relative fitness and represents the level of drug resistance relative to the wild type. W>1 indicates drug resistance, and W<1 indicates drug sensitivity.

This empirical measure of drug resistance can be directly linked to a simple pharmacodynamics model (Rosenbloom et al. 2012), where the viral replicative fitness is modeled as a function of drug dose,

$$W([DCV]) = \frac{RF_{mut}([DCV])}{RF_{mut}} = \left(\frac{IC_{mut}}{[DCV]+IC_{mut}}\right) \Big/ \left(\frac{IC_{wt}}{[DCV]+IC_{wt}}\right)$$

Here IC denotes the half-inhibitory concentration. The Hill coefficient describing the sigmoidal shape of the dose response curve is fixed to 1, as used in fitting the dose response curves of wild-type virus and validated mutant viruses (Supplementary Figure 5). We can use the drug resistance score W to infer the dose response of each mutant. Because the dose response curve depends on the duration of drug treatment, we transformed $W_{observed}$ (144 hr drug treatment in selection experiments) to $W_{predicted}$ (48 hr drug treatment, Supplementary Figure 6A) and then calculated IC_{50} using the equation above.

Calculation of relative solvent accessibility

DSSP (http://www.cmbi.ru.nl/dssp.html) was used to compute the Solvent Accessible Surface Area (SASA) (Kabsch and Sander 1983) from the HCV NS5A protein structure (PDB: 3FQM) (Love et al. 2009). SASA was then normalized to Relative Solvent Accessibility (RSA) using the empirical scale reported in (Tien et al. 2013).

Predictions of protein stability

ΔΔG (in Rosetta Energy Units) of HCV NS5A mutants was predicted by PyRosetta (version: “monolith.ubuntu.release-104”) as the difference in scores between the monomer structure of mutants (single amino acid mutations from site 32 to 103) and the reference (PDB: 3FQM). The score is designed to capture the change in thermodynamic stability caused by the mutation (ΔΔG) (Das and Baker 2008). The sequence of the reference protein was different from the sequence of the wild-type virus used in this study. Thus instead of directly comparing ΔΔG to fitness effects, we used the median ΔΔG caused by amino acid substitutions at each site.

The PDB file of NS5A dimer was cleaned and trimmed to a monomer (chain A). Next, all side chains were

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

repacked (sampling from the 2010 Dunbrack rotamer library (Shapovalov and Dunbrack 2011)) and minimized for the reference structure using the talaris2014 scoring function. After an amino acid mutation was introduced, all residues within a 10 Å distance of the mutated residue's center were repacked, followed by quasi-Newton minimization (algorithm: "lbfgs_armijo_nonmonotone") of the backbone and all side chains. This procedure was performed 50 times, and the predicted ΔG of a mutant structure is the average of the three lowest scoring structures.

We note that stability predictions based on NS5A monomer structure have ignored possible effects on residues buried at the dimer interface, which is further complicated by the observations of two different NS5A dimer structures (Tellinghuisen et al. 2005; Love et al. 2009).

Within-patient sequence diversity

Aligned nucleotide sequences of HCV NS5A protein were downloaded from Los Alamos National Lab database (all HCV genotypes, ~2600 sequences total) and clipped to the region of interest (amino acid 18-103 of NS5A). Sequences that caused gaps in the alignment of H77 reference genome were manually removed. After translation to amino acid sequences, sequences with ambiguous amino acids were removed (~2300 amino acid sequences after filtering). The sequence diversity at each amino acid site was quantified by Shannon entropy.

Acknowledgements

L.D. was supported by HHMI Postdoctoral Fellowship from Jane Coffin Childs Memorial Fund for Medical Research. N.C.W was supported by Croucher Foundation Fellowship. R.S. was supported by NSFC 81172314, NIH DE023591 and NIH CA177322.

Author contributions

L.D., Y.S., H.Q. and R.S. designed the experiments. L.D., H.Q. and Y.S. performed the experiments. L.D. and Y.S. analyzed the experimental data. L.D. performed the bioinformatics analyses and simulations. L.D. wrote the first draft of the manuscript. All authors discussed the results and commented on the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Agashe D, Sane M, Phalnikar K, Diwan GD, Habibullah A, Martinez-Gomez NC, Sahasrabuddhe V, Polachek W, Wang J, Chubiz LM, et al. 2016. Large-Effect Beneficial Synonymous Mutations Mediate Rapid and Parallel Adaptation in a Bacterium. Mol. Biol. Evol. 33:1542–1553.

Bank C, Hietpas RT, Jensen JD, Bolon DNA. 2015. A systematic survey of an intragenic epistatic landscape. Mol. Biol. Evol. 32:229–238.

Bataillon T, Bailey S. 2014. Effects of new mutations on fitness: insights from models and data. Ann. New York Acad. … 1320:76–92.

Bataillon T, Zhang T, Kassen R. 2011. Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens. Genetics 189:939–949.

Beisel CJ, Rokyta DR, Wichman HA, Joyce P. 2007. Testing the extreme value domain of attraction for distributions of beneficial fitness effects. Genetics 176:2441–2449.

Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. 2005. Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. U. S. A. 102:606–611.

Burch C, Guyader S, Samarov D, Shen H. 2007. Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus ϕ6. Genetics 476:467–476.

Burch CL, Chao L. 2000. Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature 406:625–628.

Carrasco P, Iglesia F de la, Elena S. 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus. J. Virol. 81:12979–12984.

Chayama K, Takahashi S, Toyota J, Karino Y, Ikeda K, Ishikawa H, Watanabe H, McPhee F, Hughes E, Kumada H. 2012. Dual therapy with the nonstructural protein 5A inhibitor, daclatasvir, and the nonstructural protein 3 protease inhibitor, asunaprevir, in hepatitis C virus genotype 1b-infected null responders. Hepatology 55:742–748.

Chevereau G, Dravecká M, Batur T, Guvenek A, Ayhan DH, Toprak E, Bollenbach T. 2015. Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. PLoS Biol. 13:e1002299.

Cowperthwaite MC, Bull JJ, Meyers LA. 2005. Distributions of beneficial fitness effects in RNA. Genetics 170:1449–1457.

Das R, Baker D. 2008. Macromolecular Modeling with Rosetta. Annu. Rev. Biochem. 77:363–382.

Desai MM, H DW, O DT and P, R RM, J WMJ and GC, E ES and LR, Buckling A MRCBMA and CN, R ESF and S, Garland J R RMR, Lenski R RMSS and TS, et al. 2013. Statistical questions in experimental

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolution. J. Stat. Mech. Theory Exp. 2013:P01003.

Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. Microbiol. Mol. Biol. Rev. 76:159–216.

Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature 463:353–355.

Echave J, Spielman SJ, Wilke CO. 2016. Causes of evolutionary rate variation among protein sites. Nat. Rev. Genet. 17:109–121.

Elena SF, Carrasco P, Daròs J-A, Sanjuán R. 2006. Mechanisms of genetic robustness in RNA viruses. EMBO Rep. 7:168–173.

Everson GT, Sims KD, Rodriguez-Torres M, Hézode C, Lawitz E, Bourlière M, Loustaud-Ratti V, Rustgi V, Schwartz H, Tatum H, et al. 2014. Efficacy of an interferon- and ribavirin-free regimen of daclatasvir, asunaprevir, and BMS-791325 in treatment-naive patients with HCV genotype 1 infection. Gastroenterology 146:420–429.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8:610–618.

Ferguson AL, Mann JK, Omarjee S, Ndung'u T, Walker BD, Chakraborty AK. 2013. Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. Immunity 38:606–617.

Firnberg E, Ostermeier M. 2013. The genetic code constrains yet facilitates Darwinian evolution. Nucleic Acids Res. 41:7420–7428.

Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR, Harris M. 2010. All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. J. Virol. 84:9267–9277.

Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. Nat. Methods 11:801–807.

Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell R a, Serrano-Wu MH, Langley DR, Sun J-H, O'Boyle DR, et al. 2010. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature 465:96–100.

Geoghegan JL, Senior AM, Holmes EC. 2016. Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. Proc. R. Soc. London B Biol. Sci. 283.

Goldberg DE, Siliciano RF, Jacobs WR. 2012. Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. Cell 148:1271–1283.

Gong LI, Suchard MA, Bloom JD, Alexander J, Bilsel P, Guercio M del, Marinkovic-Petrovic A, Southwood S,

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stewart S, Assarsson E, et al. 2013. Stability-mediated epistasis constrains the evolution of an influenza protein. Hum Immunol 2:468–474.

Gutierrez a, Laureti L, Crussard S, Abida H, Rodríguez-Rojas a, Blázquez J, Baharoglu Z, Mazel D, Darfeuille F, Vogel J, et al. 2013. β-Lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. Nat. Commun. 4:1610.

Hart GR, Ferguson AL. 2015. Error catastrophe and phase transition in the empirical fitness landscape of HIV. Phys. Rev. E 91:32705.

He X, Liu L, Li C, Qian W, Maclean CJ, Zhang J, Puchta O, Pitt JN, Ferre-D'Amare AR, Pal C, et al. 2016. EVOLUTION. Toward a prospective molecular evolution. Science 352:769–770.

Hézode C, Hirschfield GM, Ghesquiere W, Sievert W, Rodriguez-Torres M, Shafran SD, Thuluvath PJ, Tatum H a, Waked I, Esmat G, et al. 2015. Daclatasvir plus peginterferon alfa and ribavirin for treatment-naive chronic hepatitis C genotype 1 or 4 infection: a randomised study. Gut 64:948–956.

Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. U. S. A. 108:7896–7901.

Hwang J, Huang L, Cordek DG, Vaughan R, Reynolds SL, Kihara G, Raney KD, Kao CC, Cameron CE. 2010. Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins. J. Virol. 84:12480–12491.

Ickinson WJD. 1998. Marginal fitness contributions of nonessential genes in yeast. 95:253–257.

Imhof M, Schlötterer C. 2001. Fitness effects of advantageous mutations in evolving Escherichia coli populations. Proc. Natl. Acad. Sci. U. S. A. 98:1113–1117.

Jack BR, Meyer AG, Echave J, Wilke CO, Bloom J, Labthavikul S, Otey C, Arnold F, Bloom J, Gong L, et al. 2016. Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes.Petsko GA, editor. PLOS Biol. 14:e1002452.

Jacquier H, Birgy A. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc. Natl. Acad. Sci. U. S. A. 110:13067–13072.

Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc. Natl. Acad. Sci. U. S. A. 110:13067–13072.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637.

Kassen R, Bataillon T. 2006. Distribution of fitness effects among beneficial mutations before selection in

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

experimental populations of bacteria. Nat. Genet. 38:484–488.

Kumada H, Suzuki Y, Ikeda K, Toyota J, Karino Y, Chayama K, Kawakami Y, Ido A, Yamamoto K, Takaguchi K, et al. 2014. Daclatasvir plus asunaprevir for chronic HCV genotype 1b infection. Hepatology 59:2083–2091.

Lalić J, Cuevas JM, Elena SF. 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. PLoS Genet. 7:e1002378.

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature advance on.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. Science (80-.).:aae0568.

Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, Bornberg-Bauer E, Colwell LJ, de Koning APJ, Dokholyan N V., Echave J, et al. 2012. The interface of protein structure, protein biophysics, and molecular evolution. Protein Sci. 21:769–785.

Lindenbach BD, Evans MJ, Syder AJ, Wölk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, et al. 2005. Complete replication of hepatitis C virus in cell culture. Science 309:623–626.

Love RA, Brodsky O, Hickey MJ, Wells PA, Cronin CN. 2009. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. J. Virol. 83:4395–4403.

MacLean RC, Buckling A. 2009. The distribution of fitness effects of beneficial mutations in Pseudomonas aeruginosa. PLoS Genet. 5:e1000406.

MacLean RC, Hall AR, Perron GG, Buckling A. 2010. The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. Nat. Rev. Genet. 11:405–414.

Martin G, Lenormand T. 2006. The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution (N. Y). 60:2413–2427.

Metcalf CJE, Birger RB, Funk S, Kouyos RD, Lloyd-Smith JO, Jansen VAA. 2015. Five challenges in evolution and infectious diseases. Epidemics 10:40–44.

Olson CA, Wu NC, Sun R. 2014. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr. Biol. 24:2643–2651.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. Genetics 185:603–609.

Poordad F, Sievert W, Mollison L, Bennett M, Tse E, Bräu N, Levin J, Sepe T, Lee SS, Angus P, et al. 2015. Fixed-dose combination therapy with daclatasvir, asunaprevir, and beclabuvir for noncirrhotic patients

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with HCV genotype 1 infection. Jama 313:1728–1735.

Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. Science (80-.). 352.

Qi H, Olson CA, Wu NC, Du Y, Sun R. 2015. Determining the Relative Fitness Score of Mutant Viruses in a Population Using Illumina Paired-end Sequencing and Regression Analysis. Bio-protocol 5:e1475.

Qi H, Olson CA, Wu NC, Ke R, Loverdo C, Chu V, Truong S, Remenyi R, Chen Z, Du Y, et al. 2014. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. PLoS Pathog. 10:e1004064.

Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme genetic fragility of the HIV-1 capsid. PLoS Pathog. 9:e1003461.

Robinson M, Tian Y, Delaney WE, Greenstein AE. 2011. Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals. Proc. Natl. Acad. Sci. U. S. A. 108:10290–10295.

Rodríguez-Verdugo A, Gaut BS, Tenaillon O, Orr H, Kimura M, Schrag S, Perrot V, Levin B, Weinreich D, Watson R, et al. 2013. Evolution of Escherichia coli rifampicin resistance in an antibiotic-free environment during thermal stress. BMC Evol. Biol. 13:50.

Rokyta D, Beisel C, Joyce P. 2008. Beneficial fitness effects are not exponential for two viruses. J. Mol. Evol. 67:368–376.

Rokyta D, Joyce P, Caudle S, Wichman H. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat. Genet. 37:441–444.

Rosenbloom DIS, Hill AL, Rabi SA, Siliciano RF, Nowak MA. 2012. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. Nat. Med. 18:1378–1385.

Sanjuán R, Moya A, Elena S. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. ….

Schenk MF, Szendro IG, Krug J, de Visser JAGM. 2012. Quantifying the adaptive potential of an antibiotic resistance enzyme. PLoS Genet. 8:e1002783.

Shapovalov M V, Dunbrack RL. 2011. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19:844–858.

Silander O, Tenaillon O, Chao L. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. PLoS Biol 5.

Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. Nat. Rev. Genet. 11:572–582.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. Cell 160:882–892.

Tellinghuisen TL, Marcotrigiano J, Rice CM. 2005. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. Nature 435:374–379.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife 3.

Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. 2013. Maximum allowed solvent accessibilites of residues in proteins. PLoS One 8:e80635.

Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS, Webster R, Bean W, Gorman O, Chambers T, Kawaoka Y, et al. 2016. The Mutational Robustness of Influenza A Virus.Ferguson NM, editor. PLOS Pathog. 12:e1005856.

de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. 15:480–490.

Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. PERSPECTIVE: EVOLUTION AND DETECTION OF GENETIC ROBUSTNESS. Evolution (N. Y). 57:1959–1972.

Wang C, Jia L, Huang H, Qiu D, Valera L, Huang X, Sun J-H, Nower PT, O'Boyle DR, Gao M, et al. 2012. In vitro activity of BMS-790052 on hepatitis C virus genotype 4 NS5A. Antimicrob. Agents Chemother. 56:1588–1590.

Wang C, Jia L, O'Boyle DR, Sun J-H, Rigat K, Valera L, Nower P, Huang X, Kienzle B, Roberts S, et al. 2014. Comparison of Daclatasvir Resistance Barriers on NS5A from Hepatitis C Virus Genotypes 1 to 6: Implications for Cross-Genotype Activity. Antimicrob. Agents Chemother. 58:5155–5163.

Wright S. 1932. The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. 1:356–366.

Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. Elife 5.

Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, et al. 2015. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality.Worobey M, editor. PLOS Genet. 11:e1005310.

Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen S-H, Lu I-H, Lin C-Y, Chin RG, et al. 2014. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Sci. Rep. 4:4942.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wu NC, Young AP, Dandekar S, Wijersuriya H, Al-Mawsawi LQ, Wu T-T, Sun R. 2013. Systematic identification of H274Y compensatory mutations in influenza A virus neuraminidase by high-throughput screening. J. Virol. 87:1193–1199.

Wylie CS, Shakhnovich EI. 2011. A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc. Natl. Acad. Sci. 108:9916–9921.

Yampolsky LY, Stoltzfus A. 2005. The exchangeability of amino acids in proteins. Genetics 170:1459–1472.

Yang J-R, Chen X, Zhang J, Warner J, Bulmer M, Qian W, Yang J, Pearson N, Maclean C, Zhang J, et al. 2014. Codon-by-Codon Modulation of Translational Speed and Accuracy Via mRNA Folding.Malik HS, editor. PLoS Biol. 12:e1001910.

Yoshimi S, Imamura M. 2015. Long term persistence of NS5A inhibitor-resistant hepatitis C virus in patients who failed daclatasvir and asunaprevir therapy. J. Med. Virol.:1–2.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

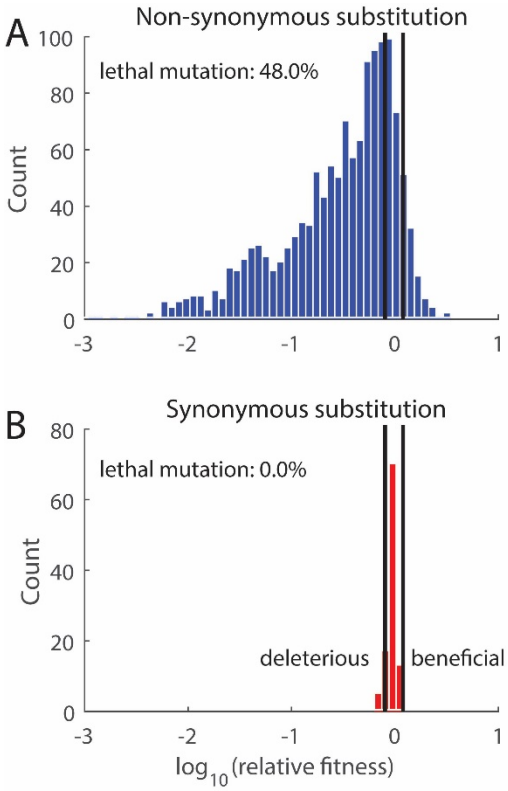


Figure 1. Distribution of fitness effects (DFE) of single codon substitutions of HCV NS5A protein. DFE of (A) non-synonymous substitutions and (B) synonymous substitutions. The thresholds (black lines) used for classifying beneficial, nearly neutral and deleterious mutations are determined by the variation of fitness values of synonymous substitutions (Methods). 88.7% of non-synonymous substitutions in NS5A protein are deleterious (among which 48.0% are lethal and not displayed in the histogram), 7.9% are nearly neutral and 3.4% are beneficial mutation.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

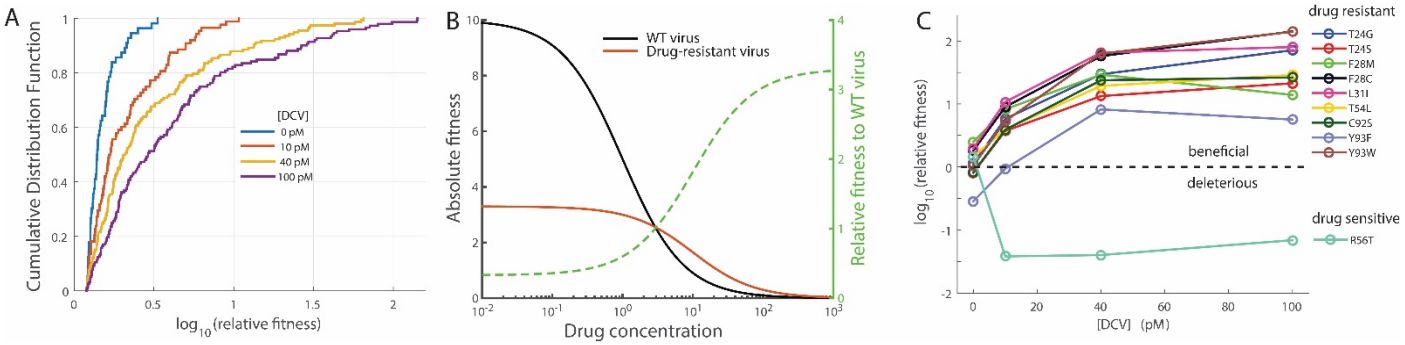


Figure 2. The spectrum of beneficial mutations shifts under the selection of an antiviral drug. (A) The cumulative distribution function of relative fitness of beneficial single amino acid substitutions under different selection conditions. (B) Hypothetical dose response curves of the wild-type virus and a drug-resistant mutant virus. Relative fitness of the drug-resistant mutant is expected to increase with drug concentration. (C) Relative fitness of validated drug resistant and sensitive mutations as a function of Daclatasvir concentration.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

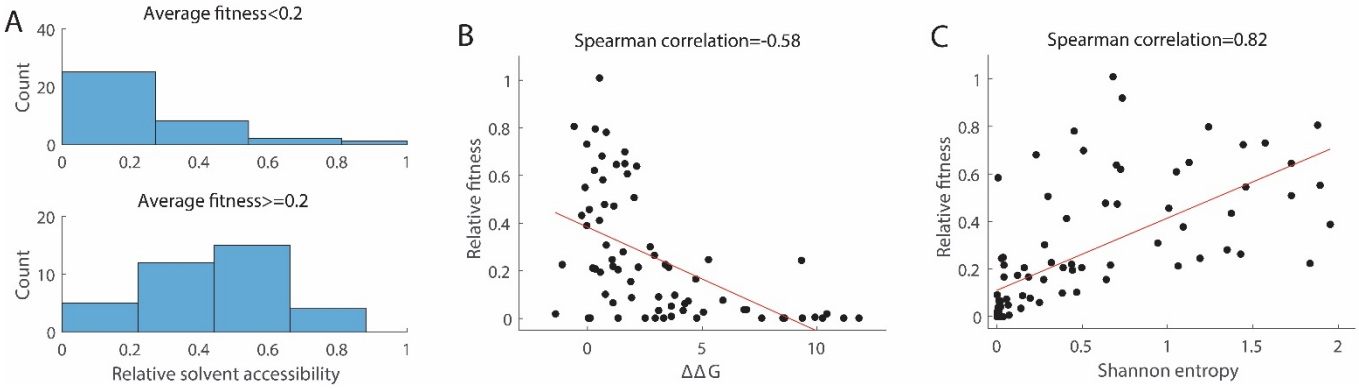


Figure 3. Deleterious mutations reveal constraints of protein evolution. (A) Amino acid sites that are less tolerant of mutations (average fitness of mutants <0.2) have lower relative solvent accessibility. The fitness is averaged over all observed mutants at each amino acid site (mutations that lead to the same amino acid change are grouped together). (B) Mutations that destabilize protein stability reduce the viral replicative fitness. Changes in folding free energy ΔΔG (Rosetta Energy Unit) of NS5A monomer are predicted by PyRosetta (Methods). The median of ΔΔG at each amino acid site is shown. (C) The within-patient sequence diversity of HCV NS5A protein is highly correlated to the replicative fitness measured in cell line, suggesting that evolutionary pathways of viral proteins are indeed constrained by mutations that reduce viral replicative fitness. Red lines represent the fits by linear regression and are only used to guide the eye.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

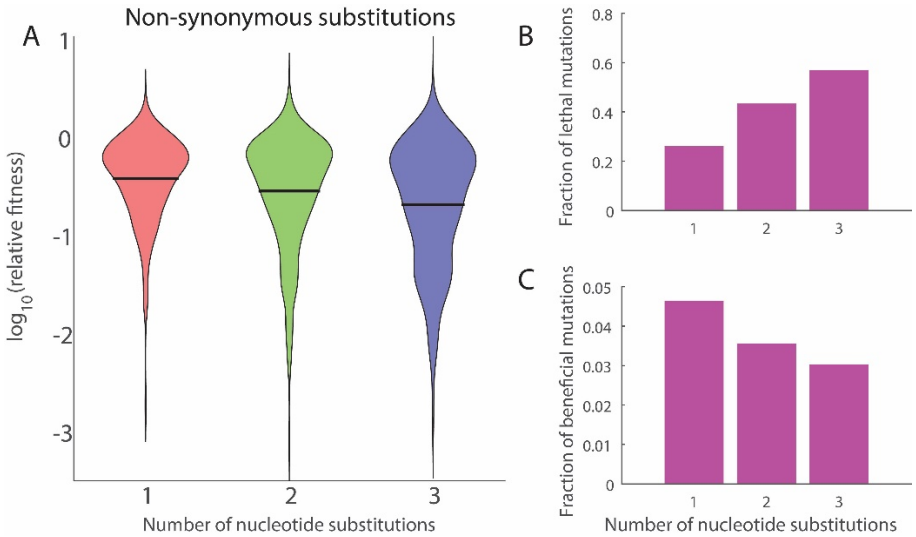


Figure 4. The role of the standard genetic code in viral evolution. (A) DFE of non-synonymous substitutions with 1, 2, and 3 nucleotide changes (lethal mutations are not displayed). Black lines indicates the mean. (B) The fraction of lethal mutations increases with the number of nucleotide changes, suggesting that the genetic code is optimized to buffer the mutational load (Chi-squared test, $p=1.4\times10^{-21}$). (C) The fraction of beneficial mutations is slightly enriched for point mutations (Chi-squared test, p=0.41). Only non-synonymous substitutions are included in the analysis.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[DCV]	Fraction of beneficial codon substitutions	Scale parameter τ	Shape parameter κ	p-value
0 pM	3.4% (56/1634)	0.26	$0.27^{n.s.}$	0.09
10 pM	5.4% (88/1634)	0.63	0.62	0.0001
40 pM	9.7% (158/1634)	0.88	1.11	<0.0001
100 pM	9.3% (152/1634)	1.57	1.18	<0.0001

Table 1. Statistics of the distribution of fitness effects of beneficial single amino acid substitutions under varying selection pressure. n.s.: cannot reject the null hypothesis that the distribution is exponential (p>0.05).

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

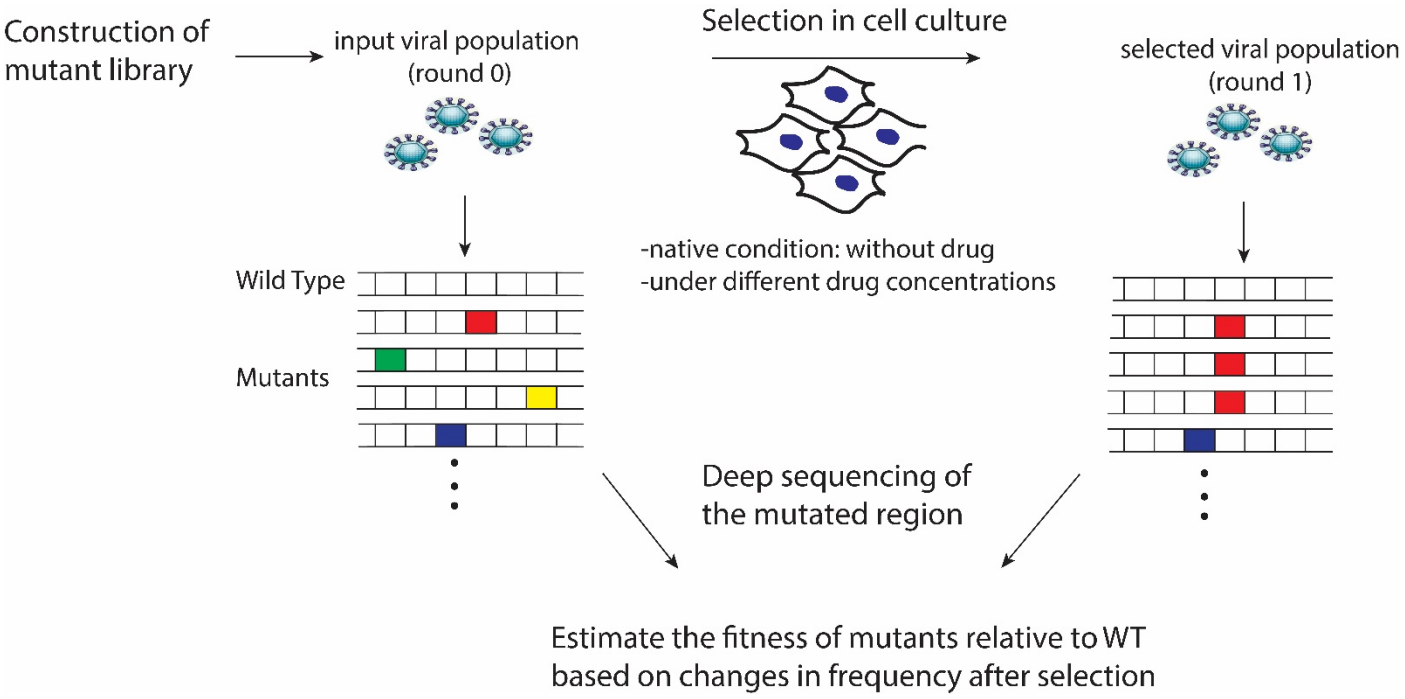
Supplementary Materials

Supplementary Figures 1-11

Supplementary Tables 1-3

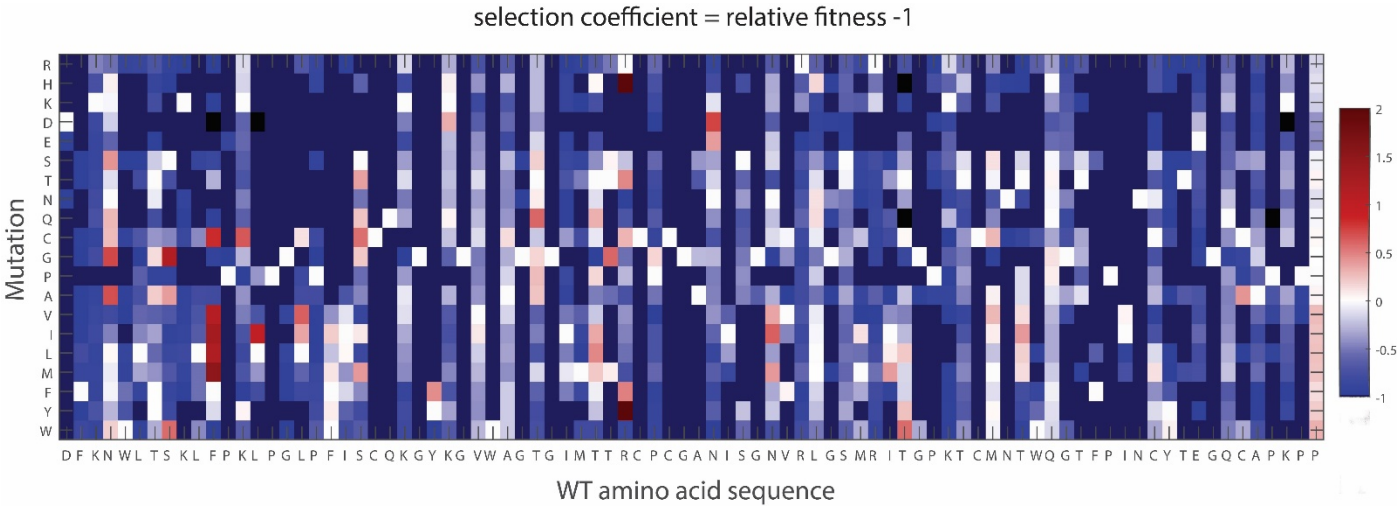
Supplementary Datasets 1-3

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



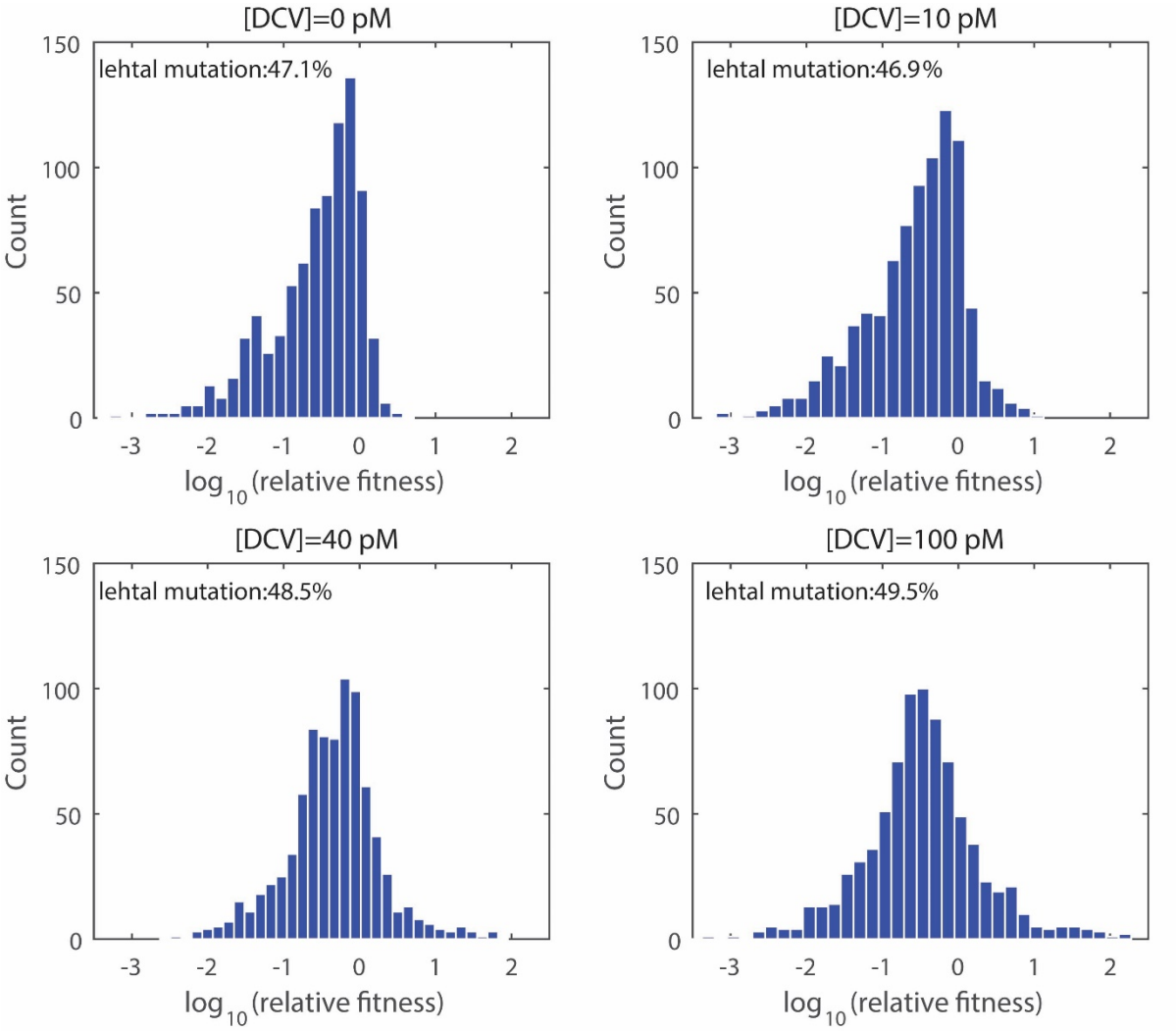
Supplementary Figure 1. Experimental workflow of high-throughput fitness assays. We performed the selection of the mutant virus library using HCV cell culture system. Viral RNA was extracted after transfection or after selection, and reverse transcribed into cDNA. The mutated region in NS5A protein was amplified by PCR and sequenced by Illumina HiSeq. The relative fitness of a mutant virus to the wild-type virus is calculated based on the frequency of the mutant virus and the wild-type virus at round 0 (after transfection) and round 1 (after selection). See Methods for more details.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



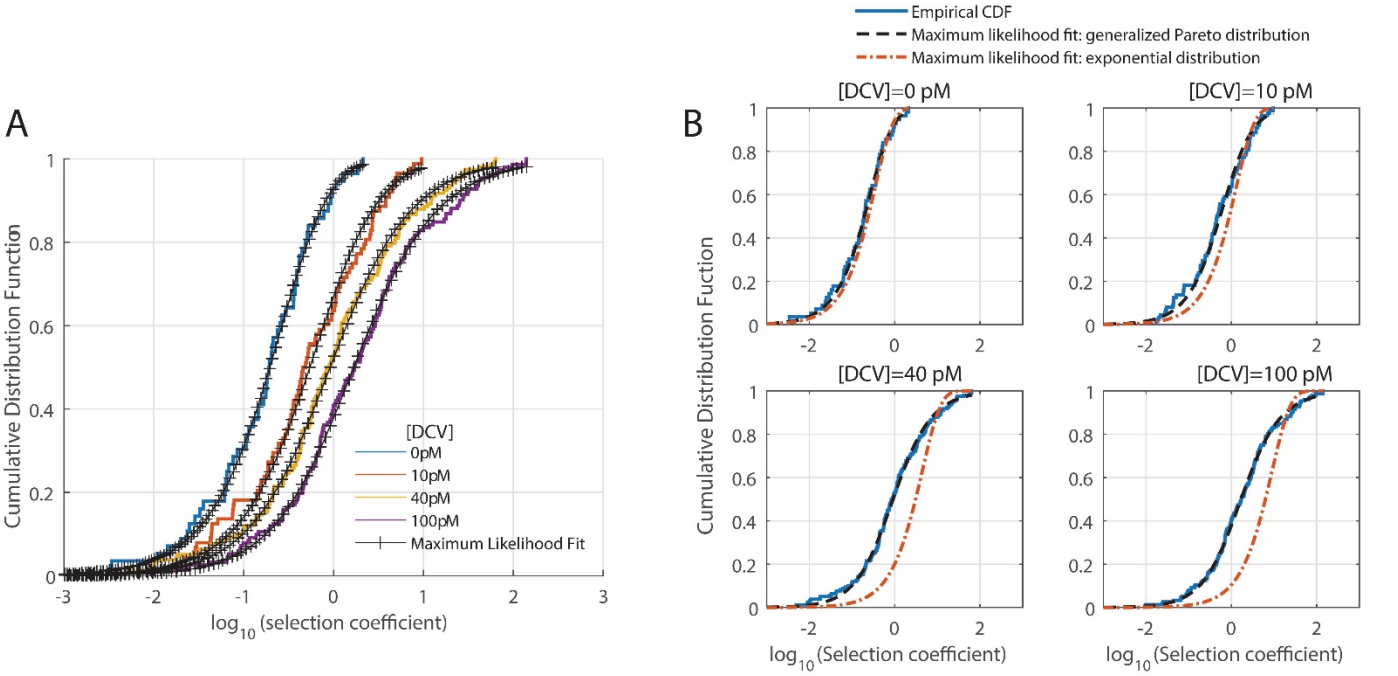
Supplementary Figure 2. Fitness effects of single amino acid substitutions in HCV NS5A protein under native condition. Different amino acid substitutions at the same site can have different fitness effects. The missing variants are colored black; lethal variants are colored dark blue.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



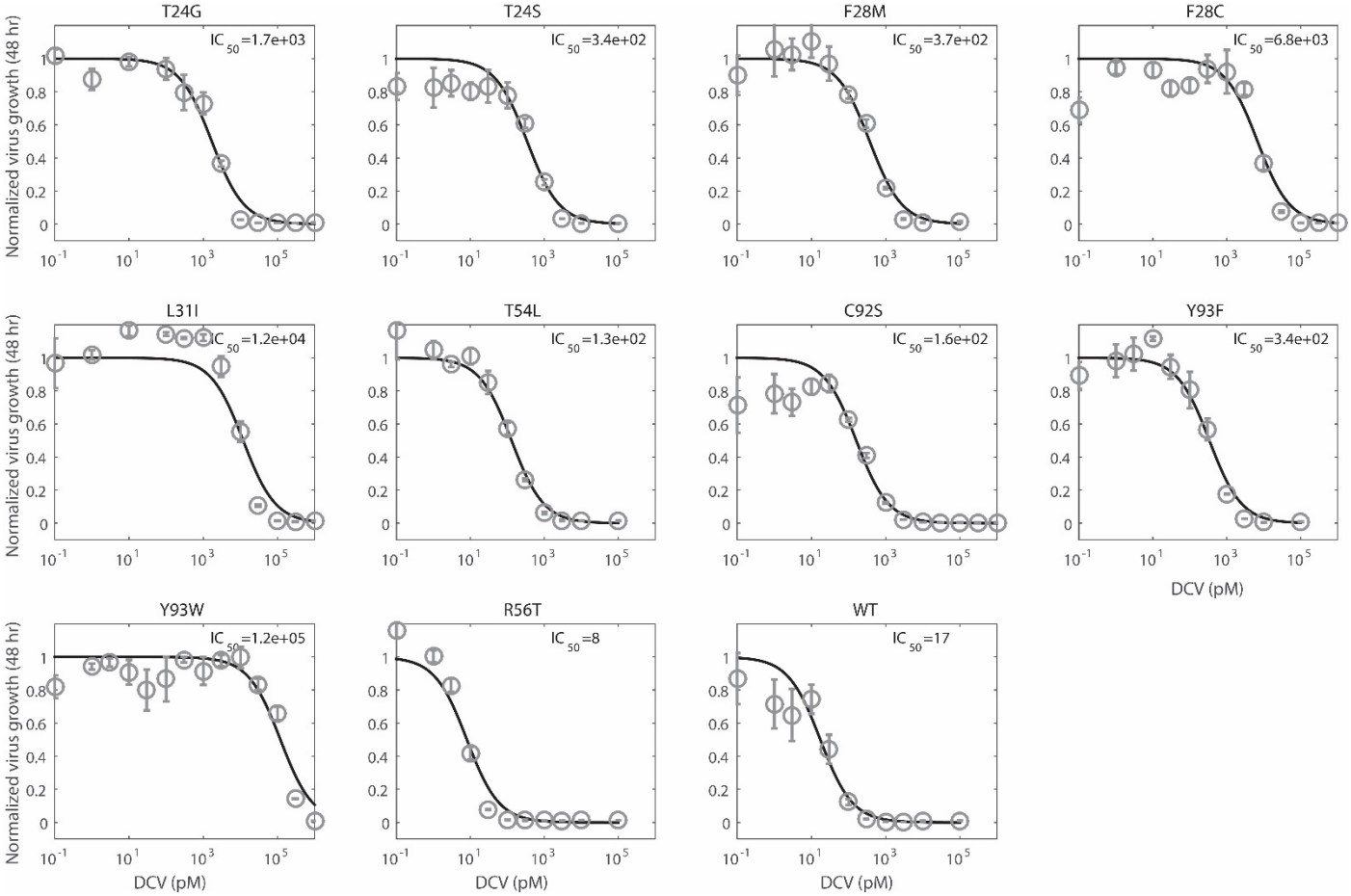
Supplementary Figure 3. DFE of non-synonymous substitutions under drug selection.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



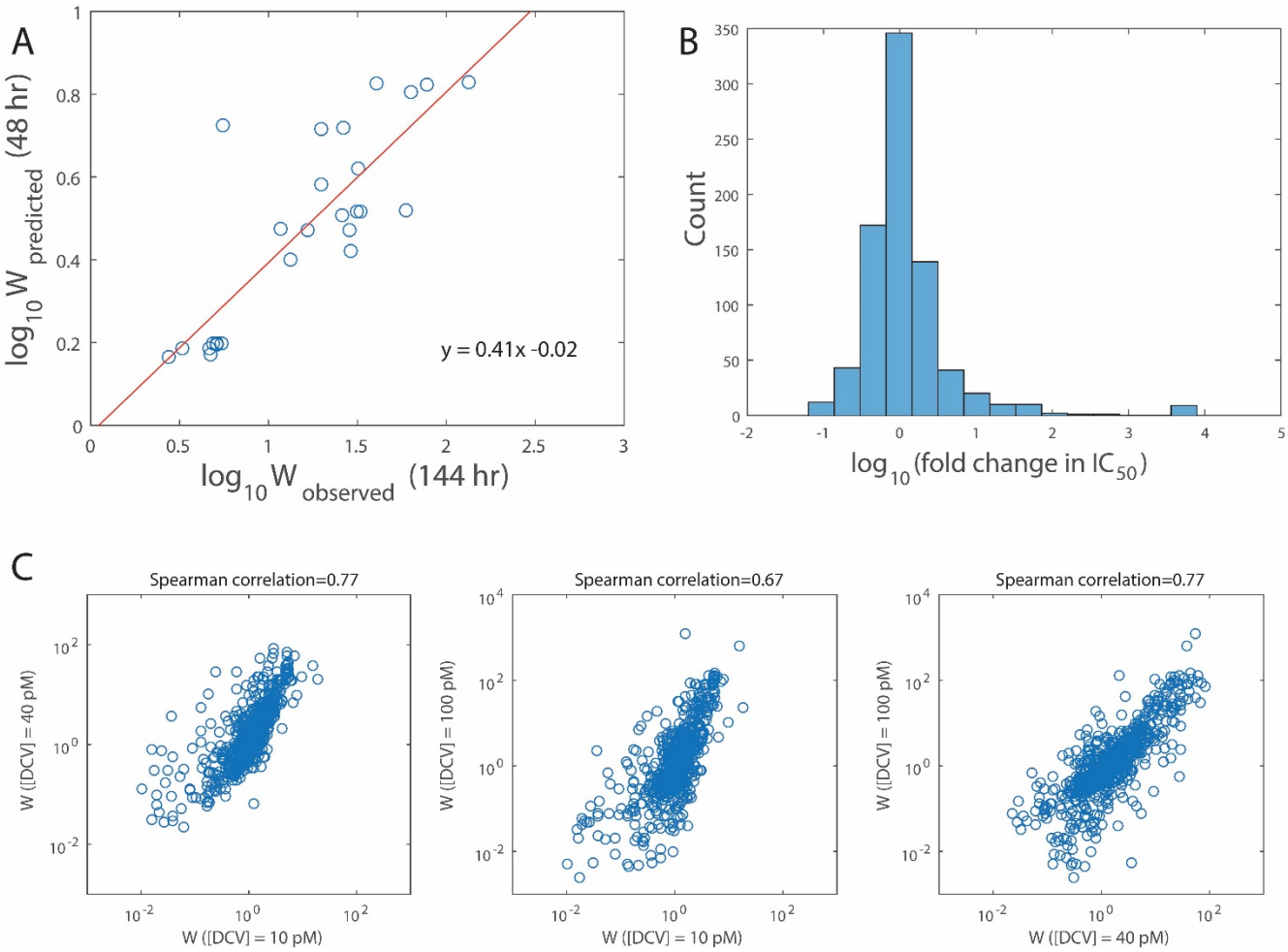
Supplementary Figure 4. Fitted distribution of fitness effects of beneficial single amino acid mutations. (A) Comparison of the fitted distribution to data. (B)The exponential distribution fails to fit the spectrum of beneficial mutations under conditions with drug selection.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



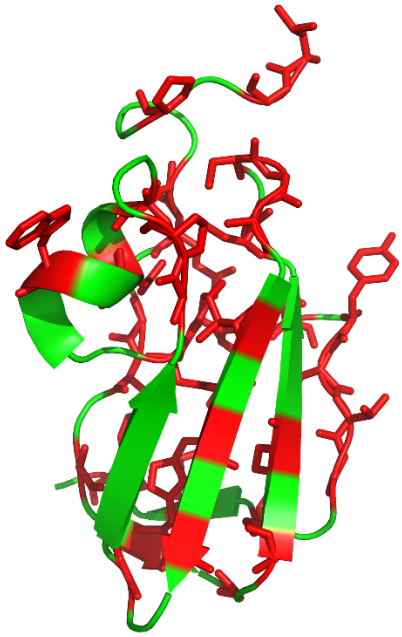
Supplementary Figure 5. Dose response curve of validated mutants (10 drug-resistant mutant, 1 drug-sensitive mutant) and WT virus. The Hill coefficient is fixed at 1 in fitting the dose response curves (Methods). The unit of IC_{50} is pM. The virus titer was measured after 48 hr of growth under drug treatment.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



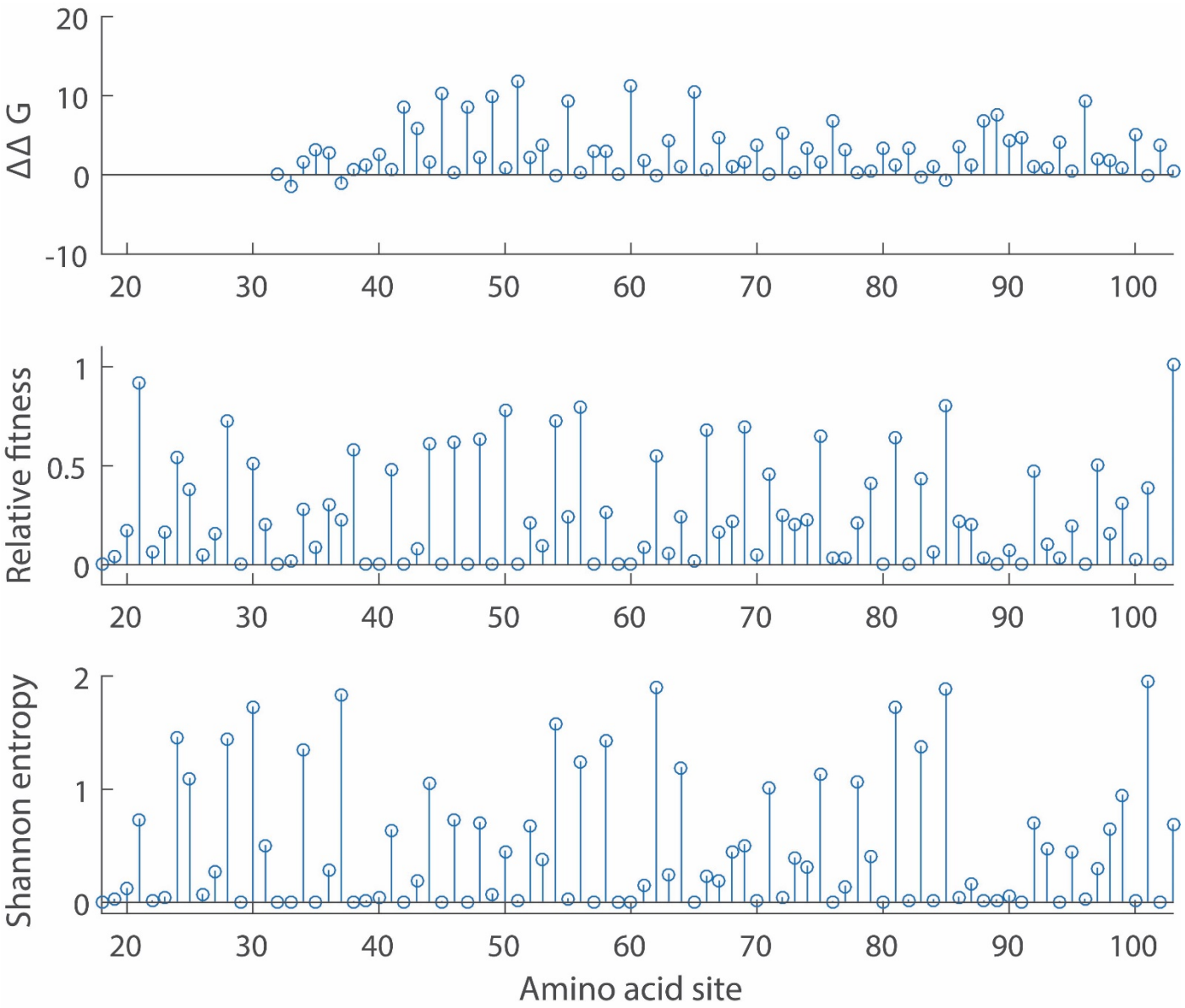
Supplementary Figure 6. Infer IC_{50} from fitness data under drug selection. (A) W is the fold change in relative fitness and represents the level of drug resistance relative to the wild type (Methods). Because the dose response curve depends on the duration of drug treatment, we normalized $W_{observed}$ (144 hr drug treatment in selection experiments) to $W_{predicted}$ (48 hr drug treatment, as used in the measurement of dose response curves of validated mutants). If viral growth is always exponential, the exponent is expected to be $\frac{48}{144}=\frac{1}{3}$. The fitted exponent is larger than $\frac{1}{3}$, suggesting that virus titer starts to saturate in 144 hr. (B) The fold change in IC_{50} caused by a single amino acid substitution is inferred from the measured fitness profiles under native condition and under drug selection (100 pM [DCV]). The existence of a group of highly resistant mutants (>10 fold change in IC_{50}) can explain why DFE of beneficial mutations shifts to a heavy-tailed distribution under drug selection. The resistance score of 9 single amino acid substitutions exceeds the maximum (Methods) and is manually set to 10^4 pM, which can be seen by the small peak in the right tail of histogram. (C) The measurement of drug resistance is consistent across different conditions of drug selection.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



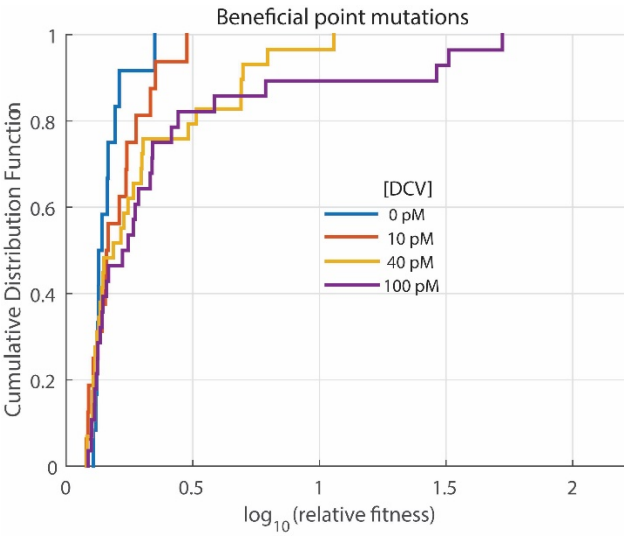
Supplementary Figure 7. Mutations at buried sites are highly deleterious. The structure of HCV NS5A monomer is visualized by Pymol (PDB: 3FQM). Amino acid sites with average fitness less than 0.2 are colored red and the corresponding side chains are shown in sticks.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



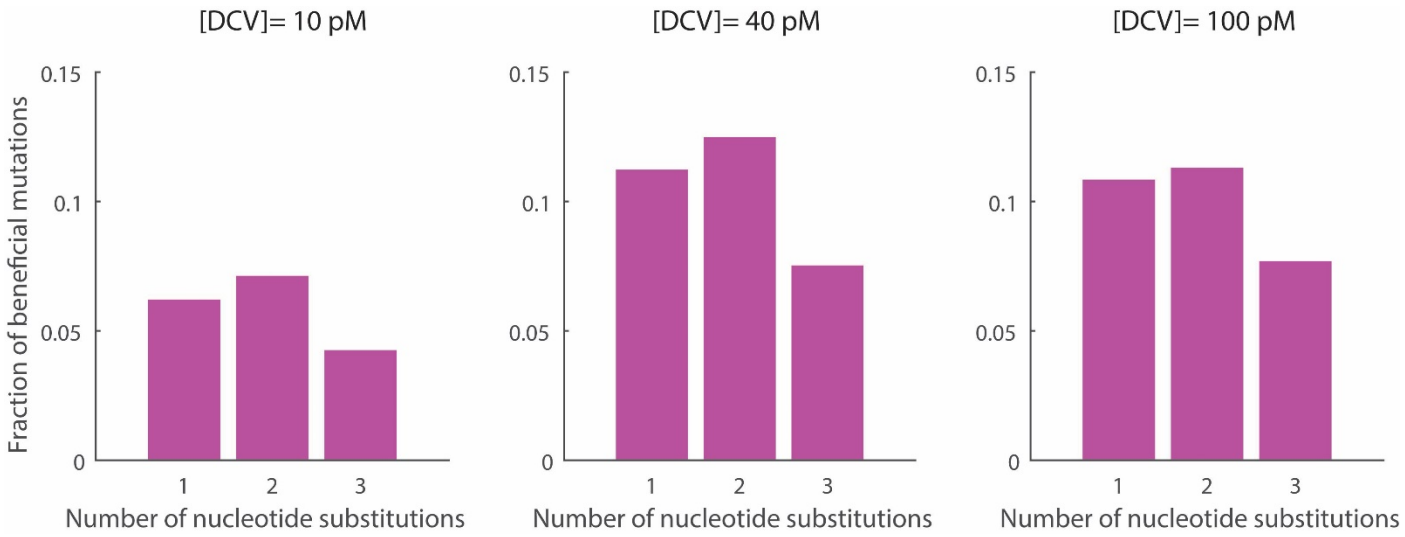
Supplementary Figure 8. Effects of mutations on protein stability, viral fitness and sequence diversity at each amino acid site. The fitness is averaged over all observed mutants at each amino acid site. The medium of ddG at each amino acid site is shown.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



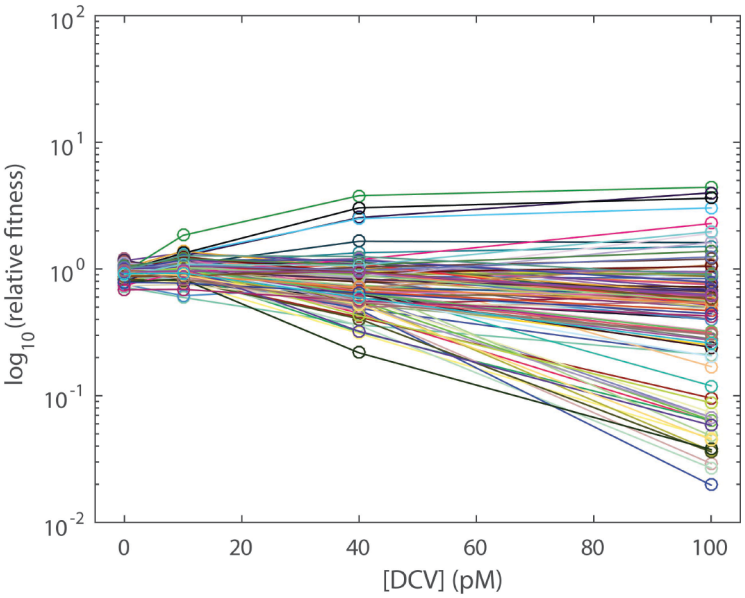
Supplementary Figure 9. Distribution of fitness effects of beneficial point mutations. Only non-synonymous mutations are included.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 10. The potential role of genetic code in adaptive mutations. For single codon substitutions that lead to amino acid changes, the fraction of beneficial mutations is markedly smaller for 3-bp substitutions. Pearson's chi-squared test: 10pM, $p=1.2\times10^{-2}$, 40pM: $p=3.7\times10^{-4}$, 100pM: $p=1.1\times10^{-2}$.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 11. Fitness effects of synonymous mutations under drug selection. Relative fitness of some synonymous mutations increase or decrease with drug concentration, suggesting that these mutations may have phenotypic effects on drug sensitivity.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[DCV]	Fraction of beneficial mutations	Scale parameter	Shape parameter	p-value
0 pM	4.6% (12/259)	0.13	$0.35^{n.s.}$	0.42
10 pM	6.2% (16/259)	0.41	$0.08^{n.s.}$	0.85
40 pM	11.2% (29/259)	0.32	1.06	0.0003
100 pM	10.8% (28/259)	0.32	1.50	<0.0001

Supplementary Table 1. Fitted parameters of the distribution of beneficial point mutations. Only non-synonymous mutations are included in the analysis. n.s.: p>0.05.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Library (amino acid site)	Forward primer	Reverse primer
1 (18-34)	5'-GTT TGC ACC ATC TTG ACA-3'	5'-TTG ACA AGA GAT GAA GGG-3'
2 (35-51)	5'-CAA GCT GCC CGG CCT C-3'	5'-GCA GCG CGT GGT CAT GAT-3'
3 (52-68)	5'-TGG GCC GGC ACT GGC-3'	5'-GAT CCT CAT AGA GCC CAG-3'
4 (69-85)	5'-CAT CTC TGG CA A TGT CCG C-3'	5'-AGC AAT TGA TAG GAA AGG CCC-3'
5 (86-103)	5'-TGC ATG AAC ACC TGG CAG-3'	5'-ATG GCG GTC TTG TAG TTC GT-3'

Supplementary Table 2. PCR primers used in preparation of sequencing samples.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sample	Barcode
Plasmid	ATG
Transfection (round 0)	CCC
No drug (round 1)	CGG
[DCV]=10 pM (round 1)	CTT
[DCV]=40 pM (round 1)	ACT
[DCV]=100 pM (round 1)	AAC

Supplementary Table 3. Barcodes used in multiplexing Illumina sequencing samples.

bioRxiv preprint doi: https://doi.org/10.1101/078428; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Dataset 1. Sequence of HCV NS5A protein.

Nucleotide sequence (amino acid site 18-103, “wild-type” in this study)

GACTTCAAAAATTGGCTGACCTCTAAATTGTTCCCCAAGCTGCCCGGCCTCCCCTTCATCTCTTGTCAA

AAGGGGTACAAGGGTGTGTGGGCCGGCACTGGCATCATGACCACGCGCTGCCCTTGCGGCGCCAAC

ATCTCTGGCAATGTCCGCCTGGGCTCTATGAGGATCACAGGGCCTAAAACCTGCATGAACACCTGGCA

GGGGACCTTTCCTATCAATTGCTACACGGAGGGCCAGTGCGCGCCG

Amino acid sequence

DFKNWLTSKLFPKLPGLPFISCQKGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWQGT

FPINCYTEGQCAP

Supplementary Dataset 2. Fitness values of all nucleotide mutations.

Supplementary Dataset 3. Fitness values of all single amino acid mutations.