



# Genome-Wide Mutagenesis of Hepatitis C Virus Reveals Ability of Genome To Overcome Detrimental Mutations

Deepak Singh,<sup>a</sup> Shalini Soni,<sup>a</sup> Shaheen Khan,<sup>a</sup> Aditya N. Sarangi,<sup>b</sup> Ragothaman M. Yennamalli,<sup>c</sup> Rakesh Aggarwal,<sup>b</sup> Naga Suresh Veerapu<sup>a</sup>

<sup>a</sup>Virology Section, Department of Life Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh, India

<sup>b</sup>Gastroenterology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India

<sup>c</sup>Translational Genomics and Immunomics Unit, VivagenDx Labs, Chennai, Tamil Nadu, India

Deepak Singh and Shalini Soni contributed equally to this work. Deepak Singh contributed data for Fig. 2 to 4 and 5A, Shalini Soni contributed data for Fig. 1 and 6, and both Deepak Singh and Shalini Soni wrote the paper with input from all authors.

**ABSTRACT** To gain insight into the impact of mutations on the viability of the hepatitis C virus (HCV) genome, we created a set of full-genome mutant libraries, differing from the parent sequence as well as each other, by using a random mutagenesis approach; the proportion of mutations increased across these libraries with declining template amount or dATP concentration. The replication efficiencies of full-genome mutant libraries ranged between 71 and 329 focus-forming units (FFU) per 10<sup>5</sup> Huh7.5 cells. Mutant libraries with low proportions of mutations demonstrated low replication capabilities, whereas those with high proportions of mutations had their replication capabilities restored. Hepatoma cells transfected with selected mutant libraries, with low (4 mutations per 10,000 bp copied), moderate (33 mutations), and high (66 mutations) proportions of mutations, and their progeny were subjected to serial passage. Predominant virus variants (mutants) from these mutant libraries (Mutant<sup>l</sup>, Mutant<sup>m</sup>, and Mutant<sup>h</sup>, respectively) were evaluated for changes in growth kinetics and particle-to-FFU unit ratio, virus protein expression, and modulation of host cell protein synthesis. Mutant<sup>m</sup> and Mutant<sup>l</sup> variants produced >3.0-log-higher extracellular progeny per ml than the parent, and Mutant<sup>h</sup> produced progeny at a rate 1.0-log lower. More than 80% of the mutations were in a nonstructural part of the mutant genomes, the majority were nonsynonymous, and a moderate to large proportion were in the conserved regions. Our results suggest that the HCV genome has the ability to overcome lethal/deleterious mutations because of the high reproduction rate but highly selects for random, beneficial mutations.

**IMPORTANCE** Hepatitis C virus (HCV) *in vivo* displays high genetic heterogeneity, which is partly due to the high reproduction and random substitutions during error-prone genome replication. It is difficult to introduce random substitutions *in vitro* because of limitations in inducing mutagenesis from the 5' end to the 3' end of the genome. Our study has overcome this limitation. We synthesized full-length genomes with few to several random mutations in the background of an HCV clone that can recapitulate all steps of the life cycle. Our study provides evidence of the capability of the HCV genome to overcome deleterious mutations and remain viable. Mutants that emerged from the libraries had diverse phenotype profiles compared to the parent, and putative adaptive mutations mapped to segments of the conserved nonstructural genome. We demonstrate the potential utility of our system for the study of sequence variation that ensures the survival and adaptation of HCV.

**KEYWORDS** genome viability, genome-wide mutagenesis, hepatitis C virus, quasispecies, reverse genetic analysis, tolerance

**Citation** Singh D, Soni S, Khan S, Sarangi AN, Yennamalli RM, Aggarwal R, Veerapu NS. 2020. Genome-wide mutagenesis of hepatitis C virus reveals ability of genome to overcome detrimental mutations. *J Virol* 94:e01327-19. <https://doi.org/10.1128/JVI.01327-19>.

**Editor** Julie K. Pfeiffer, University of Texas Southwestern Medical Center

**Copyright** © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Naga Suresh Veerapu, [nagasuresh.veerapu@snu.edu.in](mailto:nagasuresh.veerapu@snu.edu.in).

**Received** 20 August 2019

**Accepted** 17 October 2019

**Accepted manuscript posted online** 13 November 2019

**Published** 17 January 2020

Hepatitis C virus (HCV) infection is a major global health concern, with approximately 71 million infected persons (1). Nearly 80% of individuals who acquire HCV infection are at risk of developing chronic hepatitis, 20% of those with chronic infection are at risk of developing cirrhosis, and 5% of those with HCV-related cirrhosis are at risk of progressing to hepatocellular carcinoma (2). HCV, an enveloped virus, contains an approximately 9,600-nucleotide (nt)-long, positive-sense single-strand RNA [ss(+) RNA] genome that translates into a polyprotein of ~3,000 amino acids. The polyprotein undergoes co- and posttranslational cleavage into three structural proteins (C, E1, and E2) and seven nonstructural (NS) proteins (P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B). The NS5B protein is an error-prone RNA-dependent RNA polymerase (RdRp), responsible for the generation of progeny HCV genomes (3).

HCV isolates exhibit high genetic heterogeneity, with 7 major genotypes that differ by 25 to 30% at the nucleotide level and 67 subtypes that differ by 15 to 20% (4). Furthermore, intrahost circulating HCV genomes differ from each other by as much as 5% at the nucleotide level; these circulating viral variants are referred to as quasispecies or mutant clouds (5–7). HCV replicates at very high levels, with generation of approximately  $10^{12}$  particles per day (8). A high replication rate combined with a high error rate of  $10^{-3}$  to  $10^{-4}$  per nucleotide of RdRp allows the virus to explore a vast niche of sequence space in the genome. Because of random substitutions, new viral variants continuously emerge during error-prone replication (9). The majority of substitutions are deleterious or impair viral fitness, and some are even lethal (10). In contrast, some variants with beneficial mutations are continuously selected on the basis of their replication capacity and adaptive potential to the ever-changing host environment in which the virus replicates. With the continuous emergence of new fit variants, it is conceivable that the HCV genome is capable of exploring its own sequence space for the need to conserve functions of viral structures and proteins.

Studies of the HCV life cycle, virus-host interactions, and development of antivirals have been made possible by the advent of HCV subgenomic replicon systems (11, 12), and subsequently, a robust full-genome HCV cell culture (HCVcc) system was established using the genotype 2a HCV cDNA clone JFH-1 (13) and the genotype 2a chimera J6/JFH-1 (14). The HCVcc system recapitulates all steps of the HCV life cycle. We believe that the JFH-1 HCVcc system is an ideal platform for studying the impact of random mutations in sequence space on the viability of the HCV genome.

In the present study, we tested whether a substantial increase in mutations can result in an increased sequence variation that is not associated with the complete loss of genetic information and thus results in the emergence of replication-competent genomes. This required a molecular technique that allows a massive increase in mutations during the replication of the HCV genome. Recently, we reported a molecular technique, full-length mutant RNA synthesis (FL-MRS) for random introduction of a few to several mutations in a controllable fashion in the hepatitis E virus (HEV) genome (15). The introduction of random mutations in the HCV genome may allow evaluation of the intrinsic mutational tolerance in the sequence space of the HCV genome (16).

In this study, we used FL-MRS to introduce random mutations in the HCV genome and created full-genome mutant libraries with various proportions of mutations. Mutant libraries representing full-genome HCV variants when expressed in hepatoma cells showed replicative capability and established persistent infections. Mutant libraries with high proportions of mutations were replication competent and survived extinction. In these cases, the surviving genomes showed differences in their phenotypic profiles, with respect to growth kinetics, multiplicity of infection, virus protein expression, and host protein synthesis shutoff, and showed an enrichment of putative adaptive mutations in the nonstructural region of HCV genome.

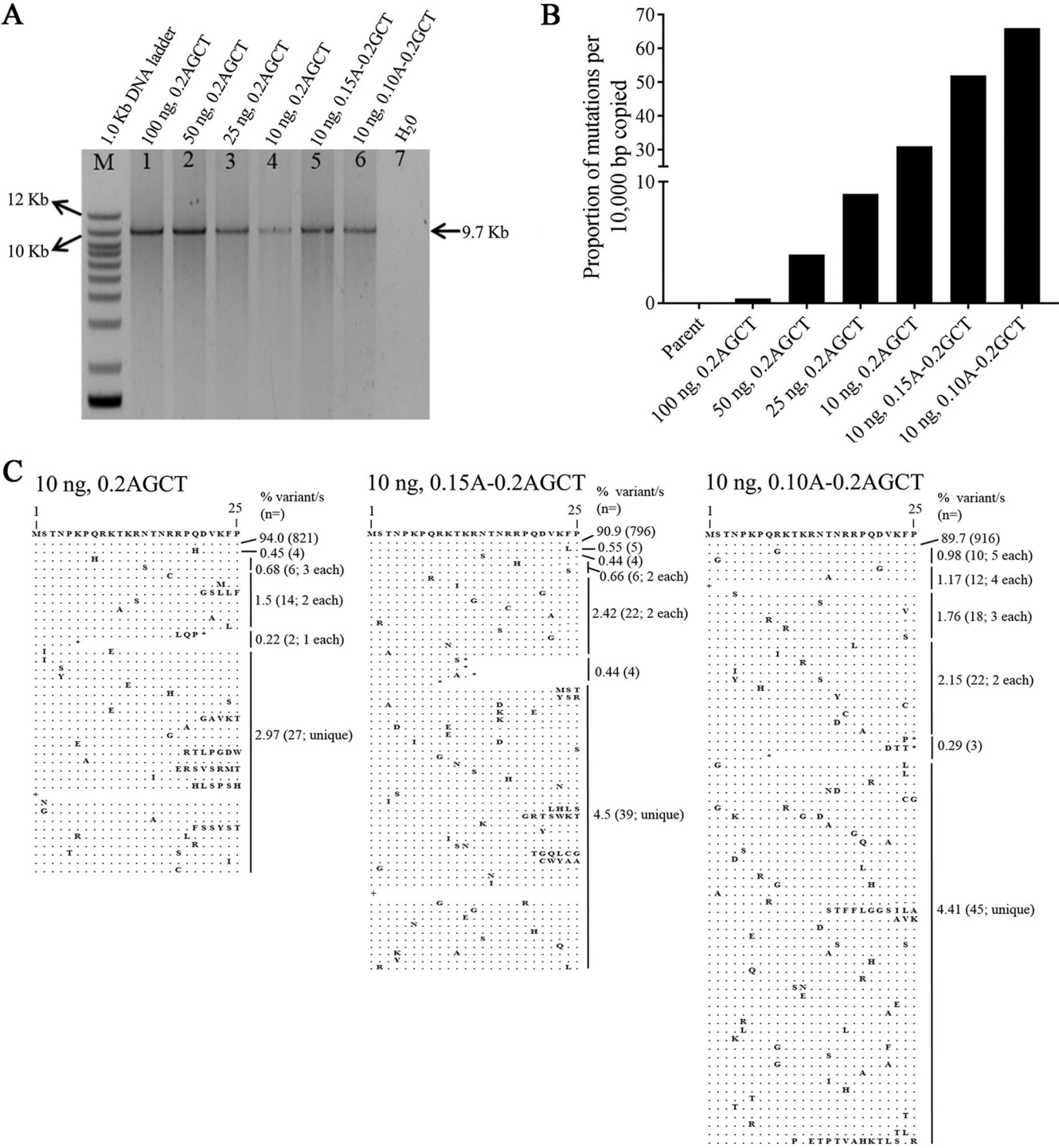
## RESULTS

**FL-MRS is a viable strategy that can contribute to the synthesis of full-genome mutant libraries.** First, we tested the possibility of inducing genome-wide mutagen-

esis by using FL-MRS. Error-prone PCRs (EP-PCRs) using decreasing template amounts of 100 ng, 50 ng, 25 ng, and 10 ng and all four deoxynucleoside triphosphates (dNTPs) at a 0.2 mM final concentration (100 ng 0.2AGCT, 50 ng 0.2AGCT, 25 ng 0.2AGCT, and 10 ng 0.2AGCT mutant libraries) were sustained and yielded 9.7-kb-long products (Fig. 1A, lanes 1 to 4). Previously, we were able to synthesize and diversify the HEV full-length genome of 7.2 kb by reducing the dATP and dGTP concentrations in EP-PCR (15). In this study, we first attempted to further induce genome-wide mutagenesis by reducing both dATP and dGTP concentrations; however, this failed to provide sustained yields. We then tried reducing only the dATP concentration. This yielded good results. EP-PCR of the HCV full-length genome using 10 ng template and final dATP concentrations of 0.15 and 0.10 mM with the other three dNTPs maintained at 0.2 mM (10 ng 0.15A-0.2GCT and 10 ng 0.10A-0.2GCT mutant libraries) provided sustained yields (Fig. 1A, lanes 5 and 6). Average yields of FL EP-PCR products obtained with balanced dNTPs, except for the 10 ng reaction template, and with imbalanced dNTP pools ranged between 8.4 ng/ $\mu$ l and 12.5 ng/ $\mu$ l ( $\pm$ 4.53). The average yield of EP-PCR products obtained with 10 ng template and a balanced dNTP pool was 4.2 ng/ $\mu$ l ( $\pm$ 0.8) (data not shown). There was no difference in amount of EP-PCR products generated using pJFH-1 or pJ6/JFH-1 as the templates.

A wide range of proportions of mutations, 0.4, 4, 9, 33, 52, and 66 per 10,000 bp, were copied in 100 ng 0.2AGCT, 50 ng 0.2AGCT, 25 ng 0.2AGCT, 10 ng 0.2AGCT, 10 ng 0.15A-0.2GCT, and 10 ng 0.10A-0.2GCT mutant libraries, respectively (Fig. 1B). A 3.6-fold increase in proportion of mutations was observed between 25 ng 0.2AGCT (9 per 10,000 bp copied) and 10 ng 0.2AGCT (33 per 10,000 bp copied). Overall, 100 ng 0.2 AGCT, 50 ng 0.2AGCT, and 25 ng 0.2AGCT libraries consisted of low proportions of mutations (0.4, 4, and 9 mutations per 10,000 bp, respectively; the combined average was 4.5 mutations per 10,000 bp), whereas 10 ng 0.2AGCT, 10 ng 0.15A-0.2GCT, and 10 ng 0.10A-0.2GCT libraries consisted of high proportions of mutations (33, 52, and 66 mutations per 10,000 bp, respectively; the combined average was 50 mutations per 10,000 bp) (Fig. 1B). Sanger sequencing of clones derived from 100 ng 0.2AGCT, 50 ng 0.2 AGCT, and 25 ng 0.2 AGCT libraries showed that the sequences had unique mutations. We had made a similar observation earlier: Sanger sequencing of EP-PCR products that were derived from an HEV cDNA clone had unique mutations (15). Table 1 shows a transition-transversion matrix for genome-wide substitutions for all six libraries. Transition-to-transversion ratios were between 1.6 and 5.8. Predicted nonsynonymous-to-synonymous substitution ratios ranged between 0.5 and 0.8 (10 full-length HCV coding regions were simulated to calculate the ratio based on transition and transversion data). Predicted polyprotein viability (without stops) was the lowest for 10 ng 0.10A-0.2GCT (20%) and the highest for 100 ng 0.2AGCT and 50 ng 0.2AGCT (100%) (Table 1). A total of 874 sequence reads from the 10 ng 0.2AGCT library, 881 from 10 ng 0.15A-0.2GCT, and 1,026 from 10 ng 0.10A-0.2GCT that encode amino acids 1 to 25 of core genes were extracted and aligned to assess amino acid sequence variation. Figure 1C demonstrates the robustness inducing random mutations by our experimental approach to the study of genome tolerance for mutations. Sequence alignments showed a decrease in percentage of parent sequences from 94 to 89.7 and an increase in number of unique variants from 3 to 4.5. Variants with stops (Fig. 1C, sequences indicated by an asterisk) and an altered start codon (indicated by a plus) were found.

**Genomes of mutant libraries are replicative.** Full-genome RNA variants of the same infectious parent generated using different degrees of mutagenesis are expected to have different *in vitro* growth characteristics. To determine the impact of genome-wide mutagenesis on replication capability of mutant libraries, we transfected each of the six libraries generated. NS5A-based immunostaining has been widely used for detection and quantitation of HCV infection in JFH-1 cell cultures. At 3 days posttransfection, cells were fixed and immunostained for NS5A, and the numbers of FFU per 100,000 cells were determined. Cells transfected with the parent were NS5A positive at



**FIG 1** Agarose gel electrophoresis of EP-PCR products and proportions of mutations in full-genome mutant libraries. (A) The HCV genome was amplified using various amounts of template and a balanced dNTP pool (lanes 1 to 4) or reduced dATP final concentrations (lanes 4 and 6). Amplification conditions are described in Materials and Methods. The expected size of the amplified genomes (EP-PCR products) was 9.7 kb. Equal volumes of EP-PCR products were loaded, resolved on an 0.8% agarose gel, and visualized by staining with ethidium bromide. Lane M, 1-kb DNA size ladder. (B) Proportions of mutations per 10,000 bp in amplicons determined by Sanger sequencing or next-generation sequencing. Sanger sequencing of 50,000 nt of the parent showed zero background errors. Next-generation sequencing of the parent showed 15 errors per 10,000 single-nucleotide reads, and these background errors were subtracted from the total proportions of mutations (corresponding proportion of transitions and transversions) found in 10 ng 0.2AGCT, 10 ng 0.15A-0.2GCT, and 10 ng 0.10A-0.2GCT libraries. (C) Alignments of amino acid reads for residues 1 to 25 of the core gene product. Amino acids are shown with the single-letter code for the parent to mutant. Amino acids are numbered consecutively relative to the JFH-1 polypeptide. A dot indicates that the amino acid within a sequence is identical to that in the parent. Sequences with a plus sign indicate an altered start codon, and an asterisk indicates nonsense mutations.

**TABLE 1** Sequencing analysis of mutant libraries of HCV synthesized by EP-PCR

Mutant library	Original base	No. of substitutions/10,000 bp copied at mutated base				Ts/Tv ratio <sup>d</sup>	Nonsynonymous/synonymous ratio <sup>a</sup>	% Polypeptide viability <sup>b</sup>
		A	G	C	T			
100 ng 0.2AGCT	A		0.20	0	0	1 (1/1)	0.67	100
	G	0		0	0			
	C	0	0		0			
	T	0	0.20	0				
50 ng 0.2AGCT	A		0	0	0	1.7 (2.5/1.5)	0.68	100
	G	0.35		0	0.70			
	C	0.30	0.70		0			
	T	0	1.40	0.70				
25 ng 0.2AGCT	A		0.95	0	0.32	1.6 (5.5/3.5)	0.55	90
	G	1.60		0	1.90			
	C	1.60	0.64		0.64			
	T	0	0.95	0.32				
10 ng 0.2AGCT <sup>c</sup>	A		9.23	0.49	1.37	4.5 (27/6)	0.83	60
	G	3.40		0.29	3.53			
	C	0.80	0.25		0.84			
	T	1.41	10.48	0.46				
10 ng 0.15A-0.2AGCT <sup>c</sup>	A		18.03	0.80	1.99	5.3 (44/8)	0.67	40
	G	3.25		0.35	3.18			
	C	0.85	0.28		0.89			
	T	2.36	19.62	0.78				
10 ng 0.10A-0.2GCT <sup>c</sup>	A		23.71	1.09	2.57	5.8 (56/10)	0.83	20
	G	3.11		0.37	3.07			
	C	0.56	0.31		0.57			
	T	3.05	25.96	1.05				

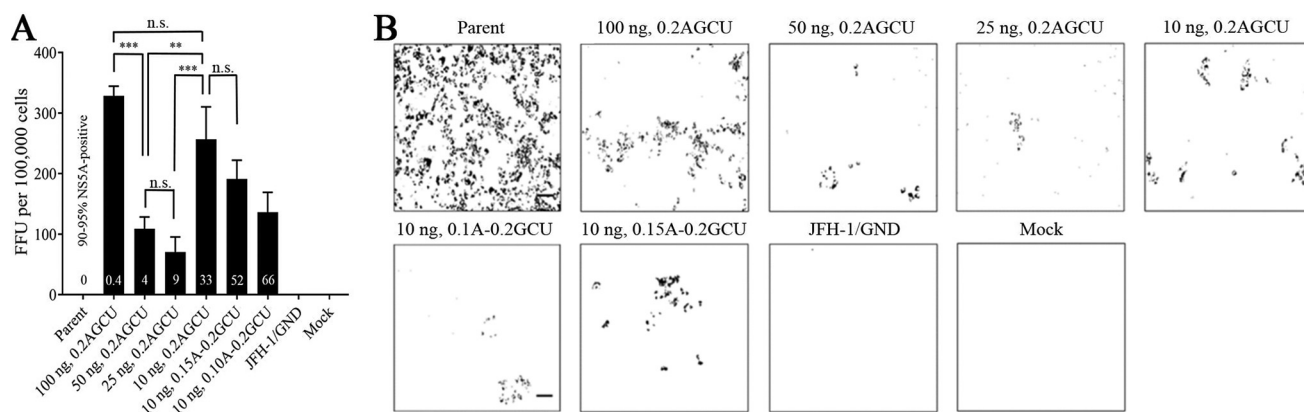
<sup>a</sup>Predicted nonsynonymous-to-synonymous substitution ratio. Ten full-genome sequences were simulated based on transition and transversion data.

<sup>b</sup>Predicted polypeptide viability. Ten full-genome sequences were simulated based on transition and transversion data.

<sup>c</sup>The numbers of nucleotide substitutions in the parent introduced by the Illumina platform were subtracted from the respective numbers of substitutions.

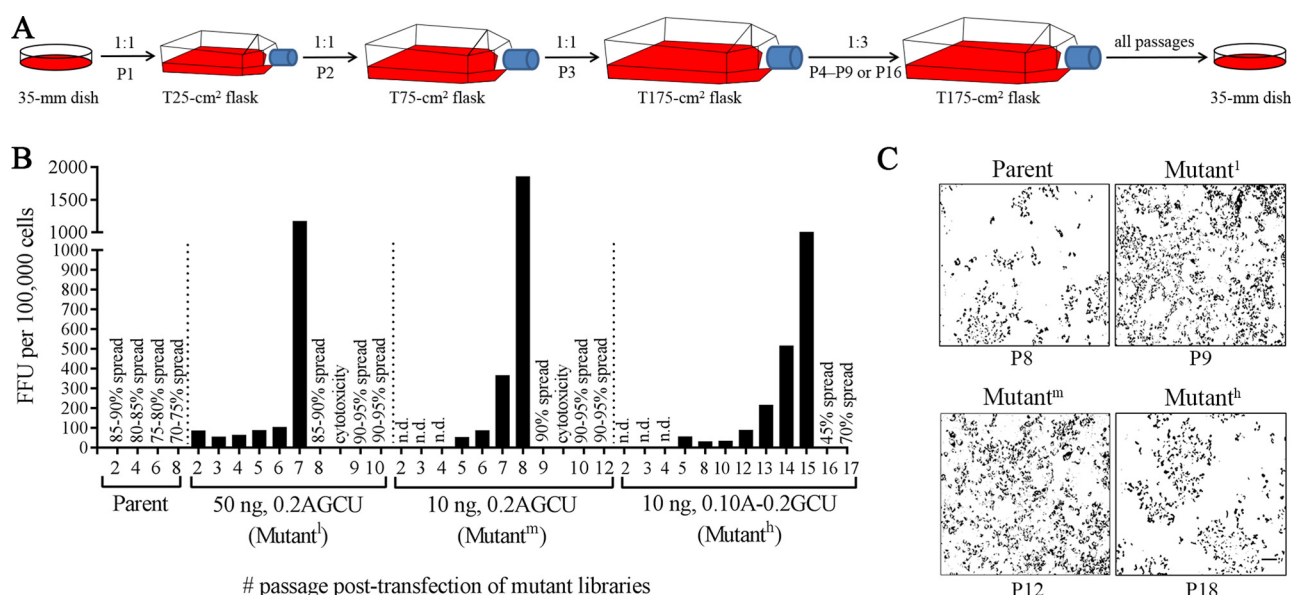
<sup>d</sup>Transition-to-transversion ratio.

about 90%. The average number of FFU observed for the six mutant libraries ranged between 71 and 329 (Fig. 2A). The number of FFU was inversely correlated with the proportion of mutations for libraries synthesized using a balanced dNTP pool and a template amount decreased in a stepwise manner, as follows: for 100 ng 0.2AGCU, 0.4



**FIG 2** Replicative capability of full-genome mutant libraries generated using full-length mutant RNA synthesis. (A) Huh7.5 cells were transfected with the indicated libraries. After 72 h, transfected cells were immunostained for NS5A and the numbers of FFU were determined visually under an optical microscope. Results are shown as means ( $\pm$  standard deviations) of three independent experiments, each performed in triplicate. Numbers indicating the proportion of mutations in the respective libraries are near the bottom of the columns. Significant differences in FFU between libraries are indicated with asterisks (\*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ). n.s., not significant. (B) Representative images show NS5A-positive cells (scale bar, 1  $\mu$ m). MAb 9E10 identifies HCV NS5A, a nonstructural protein found in the cytoplasm. Original magnification,  $\times 10$ .





**FIG 3** Scaling-up transfected Huh7.5 cells for virus recovery and early FFU kinetics posttransfection. (A) Schematic illustration of the passage cycles used to recover viruses from cells transfected with 10 ng 0.2AGCU and 10 ng 0.10A-2GCU libraries. Cell culture vessel growth areas are not shown to scale. (B) Numbers of FFU per 100,000 cells at each indicated passage for cells transfected with the respective libraries. These data are a representation from cell passages that resulted in successful virus recovery. (C) N55A stainings of cells at the earliest passages when virus spread had reached the maximum (scale bar, 1  $\mu$ m). Parallel transfections were carried out with JFH-1/GND RNA as a negative control.

mutations and 329 FFU; for 50 ng 0.2AGCU, 4 mutations and 109 FFU; and for 25 ng 0.2AGCU, 9 mutations and 71 FFU. Genomes of 100 ng 0.2 AGCU produced significantly higher numbers of FFU ( $P < 0.001$ ) than genomes of 50 ng 0.2AGCU. The FFU number was also correlated inversely with the proportion of mutations for libraries synthesized using 10 ng template and differing levels of dATP, as follows: for 10 ng 0.20AGCU, 33 mutations and 257 FFU; for 10 ng 0.15A-0.2GCU, 52 mutations and 191 FFU; and for 10 ng 0.10A-0.2GCU, 66 mutations and 136 FFU. These negative correlations between proportions of mutations and numbers of FFU are expected. However, despite a >2-fold increase in proportion of mutations, no such inverse correlation was observed between 50 ng 0.2AGCU (4 mutations and 109 FFU) or 25 ng 0.2AGCU (9 mutations and 71 FFU) and 10 ng 0.2AGCU (33 mutations and 257 FFU); the difference was statistically significant at  $P$  values of  $<0.001$  and  $<0.01$ , respectively (Fig. 2A). Moreover, despite an 80-fold increase in proportion of mutations, there was no significant difference in number of FFU between 100 ng 0.2AGCU (0.4 mutations and 329 FFU) and 10 ng 0.2AGCU (33 mutations and 257 FFU). These comparative measures of numbers of FFU convey the relative capacities to initiate replication of the genomes of mutant libraries. NS5A staining patterns of cells transfected with libraries were similar to those of J6/JFH-1-infected cells (14).

**Replication-competent viruses emerged and recovered.** Upon transfection, the majority of parent transcripts are expected to establish a replication cycle and provide easy recovery of the parent virus when the cells are routinely passaged at a 1:3 split. However, for libraries harboring high proportions of mutations, a decrease in the recovery of replication-competent viruses is expected. Mutants of 50 ng 0.2AGCU were recovered easily at a 1:3 split; however, our attempts to recover mutants from cells transfected with 10 ng 0.10A-0.2GCU and 10 ng 0.2AGCU at a 1:3 split were unsuccessful. We therefore repeated the experiment by scaling up the transfected cells transferring to new, large growth areas during the initial virus passages, as shown in Fig. 3A, which resulted in the successful recovery of mutants; otherwise, genomes of 10 ng 0.10A-0.2GCU and 10 ng 0.2AGCU might be construed as replication incompetent.

In initial experiments, cells transfected with 10 ng 0.2AGUC and 50 ng 0.2AGUC did not survive due to cytopathy (detachment of cells from the monolayer; no signs of

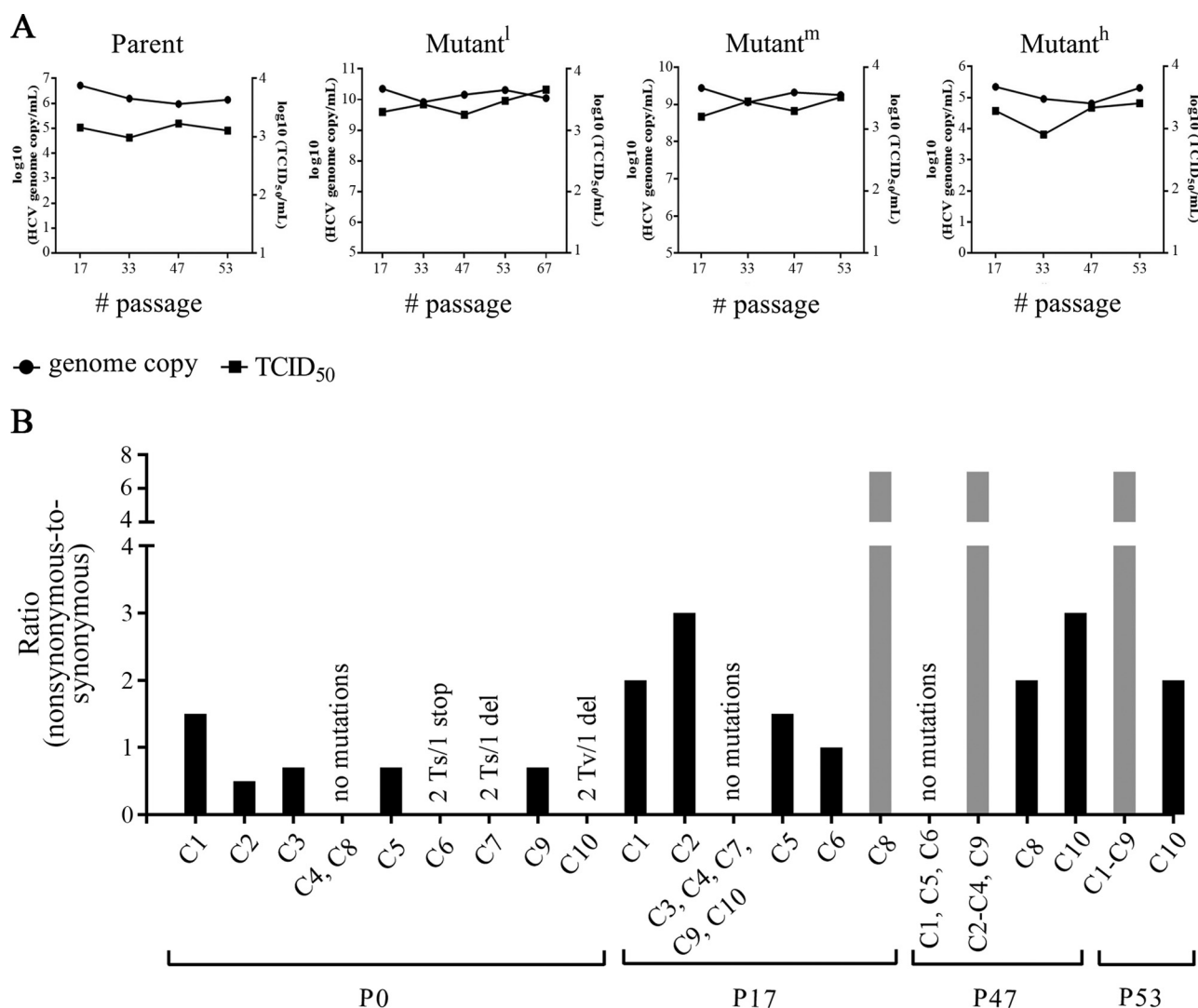
bacterial or fungal contamination) induced by the rapid virus spread that was observed mostly between passages 7 and 10. Figure 3B shows data from the experiments in which cells supported sustained replication of the virus population. After a transient decline during initial passages, the virus infection spread rapidly to 90 to 95% of cells by passage 9 (P9; about day 27 of intracellular replication) or P12 (about day 36 of intracellular replication) of cells transfected with 50 ng 0.2AGCU and 10 ng 0.2AGCU, respectively. Infection in cells transfected with 10 ng 0.10A-0.2GCU spread slowly, reaching a maximum of 70% by P17 (about day 50 of intracellular replication) (Fig. 3B). A transient decline in number of FFU during initial virus passages indicates elimination of a large replication-incompetent virus subpopulation (Fig. 3B). Based on the numbers of mutations in 10 ng 0.10A-0.2GCU, 10 ng 0.2AGCU, and 50 ng 0.2AGCU (Table 1) libraries, the populations or the mutants derived from these libraries were designated as being high (Mutant<sup>h</sup>), medium (Mutant<sup>m</sup>), or low (Mutant<sup>l</sup>), respectively. We extended virus passages to study the phenotypic traits of emerged virus variants and also whether these variants represented the mutant library from which they emerged. Figure 3C shows that virus spread reached the maximum at the earliest passage number for Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup> populations.

Virus genome copy numbers and infectious titers were measured at P17, P33, P47, and P53 for Mutant<sup>h</sup> and Mutant<sup>m</sup> populations and at P67 as well for Mutant<sup>l</sup> population. Detection of genomic RNA and infectious progeny released into the culture medium indicated a sustained infection over passages (Fig. 4A). Lack of variation in average virus RNA copy number and infectious titer indicates that the outcome of infection kinetics was determined by the input libraries (Fig. 4A). Virus populations of Mutant<sup>h</sup> at passages 0, 17, 47, and 53 were analyzed for adaptive changes in the NS3 gene. Figure 4B shows the ratio of nonsynonymous to synonymous substitutions across the input genomes (passage 0) and the mutant viruses as determined from the Sanger sequencing. The Sanger sequencing effectively selected for viable genomes, as evidenced by a purge of genomes with stop codons and deletions. Predominant Mutant<sup>l</sup> populations circulating at passage 53 (9 of 10 NS3 clones were identical) first appeared at passage 17 (1 out of 10 sequences). NS3 sequences without putative adaptive mutations appeared frequently at passages 17 and 47, showing selection against mutations for the sequence analyzed.

All virus passages were monitored for clonal diversity at 3 weeks prior to the last passage. Ten positive bacterial clones were sequenced, and the presence of a majority of identical sequences was taken as evidence of the presence of a predominant variant. All NS3 sequences were identical at 3 weeks prior to the last passage of Mutant<sup>l</sup> (last passage was P67), and for Mutant<sup>h</sup> (last passage was P53), 9 of 10 clones were identical at this time.

**Enhanced growth of mutant viruses.** Two mutant populations, Mutant<sup>m</sup> and Mutant<sup>l</sup>, showed replication with production of progeny at rates of  $1.8 \times 10^9$  and  $1.1 \times 10^{10}$  per ml, respectively, i.e., about 3-log and 4-log (1,307-fold and 7,852-fold, respectively) higher than the parent. Mutant<sup>h</sup> produced <6-fold less progeny than the parent (Table 2). No significant changes in 50% tissue culture infectious dose (TCID<sub>50</sub>) of mutant viruses were found in comparison to the parent TCID<sub>50</sub>.

**The majority of mutations were in nonstructural genes, and nonsynonymous mutations were predominant in mutant viruses.** Experimental studies with serial passage of replicative HCV clones in the presence of interferon have demonstrated that selection pressures do not apply to a single gene region but are genome wide (17). We compared genome-wide mutation counts in the predominant mutants circulating at passage 53 for Mutant<sup>h</sup> and Mutant<sup>m</sup> and at passage 67 for Mutant<sup>l</sup> to those of their respective input genomes (Fig. 5A) (Table 1). Though the counts in mutants did not directly reflect the mutation counts in the input genomes, they did show a general relation. Mutant<sup>h</sup> had a 2.3-fold lower mutation count ( $n = 29$ ) than the count ( $n = 66$ ) in the input, 10 ng 0.10A-0.2GCU. Mutant<sup>m</sup> had a 1.8-fold lower mutation count ( $n = 18$ ) than the count ( $n = 33$ ) in the input, 10 ng 0.2AGCU. These data are consistent



**FIG 4** HCV infectivity and genome copy numbers during serial passages in cell culture and amino acid changes during serial passage of Mutant<sup>h</sup>. (A) Huh7.5 cells were transfected with transcripts of the parent, Mutant<sup>l</sup>, Mutant<sup>m</sup>, and Mutant<sup>h</sup>. Extracellular HCV genome copy levels (quantified by real-time RT-PCR) and infectious viral titers (TCID<sub>50</sub>) of the supernatants as a function of passage number are shown. (B) The ratios of nonsynonymous-to-synonymous changes across the input genomes at passage 0 and the mutants at passages 17, 47, and 53 as determined by Sanger sequencing of 9 or 10 NS3 gene-positive clones (C1 to C10) are shown. Each bar represents a unique sequence or identical sequences. The gray bars represent the sequence that was dominant at passage 53. Ts, transition; Tv, transversion; del, deletion.

with the notion that mutation selection was against mutations that are not required for virus life cycle processes. In contrast, Mutant<sup>l</sup> had a 2.7-fold-higher mutation count ( $n = 11$ ) than the count ( $n = 4$ ) in the input, 50 ng 0.2AGCU. We do not rule out the possibility that adaptive changes are gained during virus passages.

Mapping of putative adaptive mutations showed a distribution across the genomes of the mutants and were predominant ( $\geq 80\%$ ) in the nonstructural part (Fig. 5A). We identified putative adaptive mutations mostly in regions of the NS2 to NS5A genes, which represent  $\sim 60\%$  of the entire genome. Mutant<sup>h</sup> had 26 out of 29 (90%), Mutant<sup>m</sup> had 15 out of 18 (83%), and Mutant<sup>l</sup> had 9 out of 11 (82%) mutations in the nonstructural part of the genome. No nucleotide changes were found in the 5' untranslated region (UTR) and core gene of the three mutants; this may be due to larger functional and structural constraints in these regions (Fig. 5A).

One of the bases that drive genetic heterogeneity *in vivo* is selection of beneficial mutations with structural and functional advantages (17). The nonsynonymous-to-synonymous substitution ratio has been used as an indicator of selection pressure in



**TABLE 2** Particle-to-FFU ratios for replication-competent mutants<sup>a</sup>

Virus	No. of genome copies/ml	Fold change in no. of genome copies/ml relative to parent	No. of genome copies/TCID <sub>50</sub>
Parent <sup>b</sup>	1.4 × 10 <sup>6</sup>		1.2 × 10 <sup>3</sup>
Mutant <sup>hc</sup>	2.1 × 10 <sup>5</sup>	−6.80	2.5 × 10 <sup>3</sup>
Mutant <sup>md</sup>	1.8 × 10 <sup>9</sup>	1,307	3.3 × 10 <sup>3</sup>
Mutant <sup>le</sup>	1.1 × 10 <sup>10</sup>	7,852	4.6 × 10 <sup>3</sup>

<sup>a</sup>Huh7.5 cells were transfected with transcripts of parent and mutant libraries. Replication-competent viruses were serially passaged. Infectious titers were determined by the TCID<sub>50</sub> method. Genome copies were measured by quantitative RT-PCR.

<sup>b</sup>Passage 53 virus.

<sup>c</sup>Passage 53 virus, derived from 10 ng 0.1A-0.2GCU library.

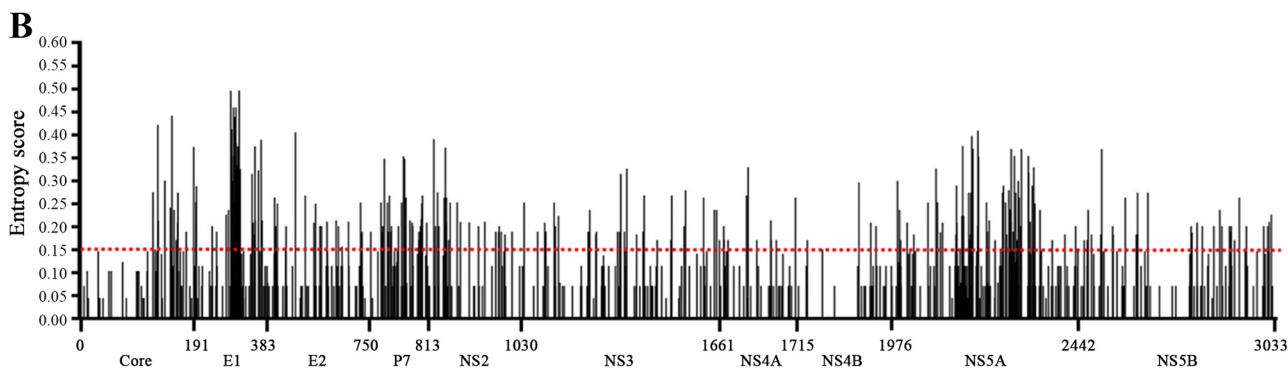
<sup>d</sup>Passage 53 virus, derived from 10 ng 0.2AGCU library.

<sup>e</sup>Passage 67 virus, derived from 50 ng 0.2AGCU library.

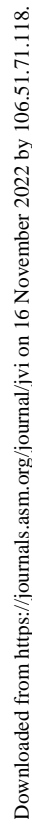
RNA viruses (18). Nonsynonymous-to-synonymous substitution ratios for the three mutants, Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup>, were 3.1 (22:7), 3.5 (14:4), and 1.8 (7:4), respectively (Fig. 5A), higher than those predicted (Table 1). This provides evidence of selection pressure on input genomes.

**Moderate to high nonsynonymous mutations mapped to conserved regions of the NS gene.** Sixteen subtype 2a full-genome sequences were obtained from HCV sequence databases, and sequence conservation was measured using Shannon en-

Mutant	Core	E1	E2	P7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	% mutations in NS gene (NS/St)	Ratio (nonsynonymous-to-synonymous)	Ratio (nonsynonymous-c-to-nonsynonymous-nc)
Mutant <sup>h</sup>		K370R	L427S, T62S		T836I, Y852C, I953T	I1047T, D1055N, F1268L, I1316T, D1320G, D1326G, I1330, K1403R, D1426G, V1597A	A1672, A1680, F1709S	Q1755R, D1882, T1949, L1951I	S1977P, V2043, T2063A, V2097A, C2118Y	T2782A	90 (26/3)	3.1 (22/7)	2.7 (16/6)
Mutant <sup>m</sup>		S273P, S294P		G784		S1218G, C1322Y, V1325A, I1456T, T1535A, R1544G		M1731T, K1760, N1771S, A1790	D2147N, F2281L, E2299, W2301R, S2336G		83 (15/3)	3.5 (14/4)	0.75 (6/8)
Mutant <sup>l</sup>		V203A			V824M	A1108T, P1145, K1152R, C1319R, V1541A		V1851, S1879	L2266, S2357P		82 (9/2)	1.8 (7/4)	0.75 (3/4)



**FIG 5** Analysis of sequence heterogeneity across the HCV subtype 2a genome identified several conserved regions, and amino acid changes are distributed across the nonstructural part of the mutants' genomes. (A) Amino acid changes (relative to the parent) in the polyprotein sequences of Mutant<sup>h</sup> (passage 53 virus), Mutant<sup>m</sup> (passage 53 virus), and Mutant<sup>l</sup> (passage 67 virus). Nonsynonymous mutations identified in conserved regions are colored red, and nonconserved regions are colored blue. Synonymous mutations are colored green. Amino acids are shown with the numerical position and single-letter code for the change from the parent to the mutant. Amino acids are numbered consecutively relative to the JFH-1 polypeptide. NS/St, nonstructural/structural; nonsynonymous-c to nonsynonymous-nc, nonsynonymous conserved to nonsynonymous nonconserved. (B) An entropy plot obtained from the analysis of 16 complete polyprotein sequences is shown. Each bar represents sequence heterogeneity at a single amino acid site. A polypeptide region of a minimum of 10 consecutive amino acids and an entropy score of <0.15 for a single amino acid site are considered conserved (dotted red line). (C and D) Sequence alignments for Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup> corresponding to conserved (C) and nonconserved (D) coding regions were identified in the parent by using Shannon entropy. Amino acids are numbered relative to the JFH-1 polypeptide (47). Numbered nonsynonymous changes are in bold. A dot indicates that the amino acid within a sequence is identical to that of the parent.



**FIG 5 (Continued)**

tropy. The topography of the measurements (Fig. 5B) was consistent with a previous report for subtype 3a sequences (19). We defined coding regions with a minimum stretch of 10 amino acids with an entropy score of  $<0.15$  as conserved, and those with an entropy score of  $\geq 0.15$  as nonconserved. In all, 96 distinct conserved regions with an average length of 19 amino acids were identified (Table 3) (the entropy score measured with the above-described cutoff criteria identified 92 conserved regions for 18 HCV subtype 3a sequences reported by Humphreys et al. [19]; data not shown). The longest of these was in the NS4B region and was 89 amino acids long. NS3 and NS5B proteins consisted of 23 and 18 conserved regions, respectively. Across all conserved genome regions, NS3 conserved regions had more putative adaptive mutations. Mutant<sup>h</sup> had 6 out of 9, Mutant<sup>m</sup> had 2 out of 6, and Mutant<sup>l</sup> had 3 out of 4 nonsynonymous changes. Mutant<sup>h</sup> NS5B had a single nonsynonymous change (T2782A) in conserved regions; the other two mutants had no mutations in the NS5B region, which indicates structural and functional constraints. The ratios of nonsynonymous mutations in conserved regions to nonsynonymous mutations in nonconserved regions for the three mutants Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup> were 2.7 (16:6), 0.75 (6:8), and 0.75 (3:4), respectively (Fig. 5A). Nonsynonymous changes identified in conserved and nonconserved regions along with parent and 16 subtype 2a protein alignments are shown in Fig. 5C and D, respectively. These data indicate that the HCV genome can tolerate mutations in conserved regions of the nonstructural part of the HCV genome. Our BLAST search found 2 of 22 nonsynonymous mutations of Mutant<sup>h</sup> in subtype 2a natural isolates, K370R (E1 conserved region; sequence deposited in GenBank under accession no. [GQ290832](#)) (data not shown) and T2782A (NS5B conserved region; GenBank accession no. [AY746460](#)) (Fig. 5A and C) and [MG891162](#) (data not shown). N1771S (NS4B conserved region), 1 of 18 nonsynonymous mutations of Mutant<sup>m</sup>, was found in one natural isolate deposited in GenBank under accession no. [KY620321](#) (data not shown). Two of 11 nonsynonymous mutations of Mutant<sup>l</sup>, K1152R (NS3 conserved region; GenBank accession no. [HQ639943](#)) (Fig. 5A and C) and S2357P (NS5A nonconserved region), were found in 11 natural isolates (Fig. 5A and D).

**Enhanced virus protein synthesis correlates with alternations of host translation initiation factors.** The replicative capability of viruses is often determined by comparative quantitation of virus expression products. On day 5 postinfection, Mutant<sup>m</sup> and Mutant<sup>l</sup> produced significantly higher numbers of genome copies ( $P < 0.01$  and  $P < 0.001$ , respectively) than the parent. The number of genome copies produced by Mutant<sup>h</sup> was comparable to that of the parent (Fig. 6A). The immunofluorescence signal in Mutant<sup>l</sup> was higher than that of the parent, while signal intensities were similar for Mutant<sup>h</sup> and the parent (Fig. 6B). On densitometry analysis, the immunofluorescence signal (anti-NS5A) from infection with Mutant<sup>l</sup> was significantly higher than that from Mutant<sup>h</sup> and parent infections ( $P < 0.001$ ) (data not shown). Cells infected with Mutant<sup>l</sup> and Mutant<sup>m</sup> showed increased levels of core and NS3 protein expression compared to cells infected with the parent (Fig. 6C). Mutant<sup>h</sup> expressed virus proteins at levels similar to those for the parent. In previous studies, HCV genome translation has been shown to shut off host protein synthesis, which is 5' cap dependent without affecting internal ribosome entry site (IRES)-dependent viral protein synthesis (20). In the case of Mutant<sup>l</sup> and Mutant<sup>m</sup> virus infections, expression levels of eIF2 $\alpha$  and eIF3 $\beta$  and host translation initiation factors decreased (Fig. 6D). In agreement with previous studies, a decrease in eIF2 $\alpha$  level was associated with an increase in phosphorylated eIF2 $\alpha$ . No difference in eIF2 $\alpha$  or eIF3 $\beta$  levels was found for Mutant<sup>h</sup> or the parent virus at day 5 postinfection (Fig. 6C). The enhanced growth kinetics of mutant viruses in hepatoma cells was associated with decreased host cell protein synthesis and increased eIF2 $\alpha$  phosphorylation, as reported by other studies (20).

**Rearrangement of noncovalent interactions in mutant structures compensates for loss of short-range interactions.** Predictive structural analysis of NS3 protein was used as a proof of concept to understand how selected, putative adaptive mutations benefitted the virus and how mutant viruses remained viable and productive. The predicted structures of NS3 of three mutants obtained from Robetta were of high

**TABLE 3** Conserved regions identified in JFH-1 polyprotein<sup>a</sup>

Protein and sequence no.	Amino acid sequence	Location in polyprotein <sup>a</sup>	Length (aa)
<b>Core</b>			
1	MSTNPKPQRKT	1–11	11
2	FPGGGQIVGGVYLLPRRGPRLGVR	24–47	24
3	RRQPIPKDRRSTGK	61–74	14
4	PWPLYGNEGLGWAGWLLSPRGSRPSWGP	82–109	28
5	GKVIDTLTCGFADLMGYIPVVGAPL	120–144	25
6	FSIFLLALLSCIT	174–186	13
<b>E1</b>			
7	AAVLHVPGCVPCE	218–230	13
8	CSALYVGDLCCGGVMLAAQ	272–289	18
9	ITGHRMAWDMMMNWSPT	313–329	17
10	MRVPEVIDI	338–347	10
11	GAHWGVMFGLAYFSMQAWAKV	350–371	22
<b>E2</b>			
12	NGSWHINRTALNC	417–429	13
13	VTNPEDMRPYCWHYPP	478–493	16
14	SVCGPVYCFTPSPVVGTDD	503–522	20
15	GVPTYTWGENETDVF	525–539	15
16	GSWFGCTWMNSTGFTKTCGAPPCR	549–572	24
17	RADFNASTDLLCPTDCFRKHP	574–594	21
18	KCGSGPWLT	600–609	10
19	YPYRLWHYPCT	615–625	11
20	PLLHSTTEWAILPC	668–681	14
21	SDLPALSTGLLHLHQNIVDVQYMYGL	684–709	26
22	LLFLLADARVCAC	725–738	14
<b>E2 and P7 junction</b>			
23	WMLILLGQAEAALEKLV	740–756	17
<b>NS2</b>			
24	LWWLCYLLTL	847–856	10
25	FCPGVVFDTKWLLA	884–898	15
26	VPYFVRAHAL	913–922	10
27	ALLALGRWTGTIYDHL	941–957	17
28	GLRDLAVAVEPIIFSPMEKKVI	967–988	22
29	WGAETAACGDILHGLPVSARLG	990–1011	22
30	LLGPADGYTSKGW	1015–1027	13
<b>NS2 and NS3 junction</b>			
31	LLAPITAYAQQTRGLLG	1029–1045	17
<b>NS3</b>			
32	GVLWTVYHGAGNKTLAG	1080–1096	17
33	RGPVTQMYSSAEGDLVGW	1098–1115	18
34	CGAVDLYLVTRNADVIPARRRGDKRGALLSPRP	1129–1161	33
35	STLKGSSGGPVLCPRGH	1163–1179	17
36	CSRGVAKSIDFIPVE	1189–1203	15
37	PPAVPQTYQVGYLHAPTSGSGKSTKVPVAYA	1220–1249	30
38	QGYKVLVLNPSVAATLGFGAYL	1251–1272	22
39	KAHGINPNIRTGVRTV	1274–1289	16
40	ITYSTYGFADGGC	1295–1309	15
41	GAYDIIICDECH	1312–1323	12
42	LGIGTVLDQAE	1331–1341	11
43	AGVRLTVLATATPPGS	1343–1358	16
44	IKGGRHLIFCHSKKCDLAAALR	1389–1412	24
45	MGLNAVAYYRGLDVS	1414–1428	15
46	QGDVVVATDALMTG	1433–1447	15
47	TGDFDSVIDCNVAV	1449–1462	14
48	VDFSLDPTFTI	1466–1476	11
49	TQTVPQDAVSRQRGRGRGR	1478–1499	22
50	YRYVSTGERASGMFDSV	1503–1519	17
51	NTPGLPVCQDHLEFWEAV	1548–1565	18

(Continued on next page)

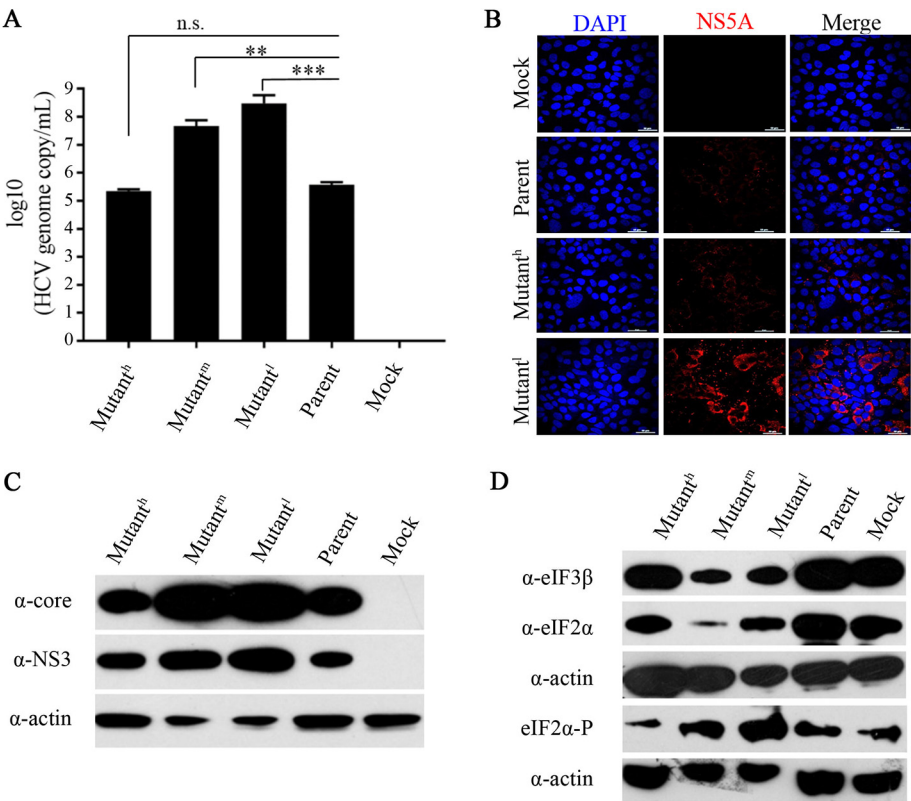
**TABLE 3** (Continued)

Protein and sequence no.	Amino acid sequence	Location in polyprotein <sup>a</sup>	Length (aa)
52	AYQATVCARAKAPPPSWD	1592–1609	18
53	GPTPLLYRLG	1624–1633	10
54	TKYIATCMQADLE	1646–1658	13
NS3 and NS4a junction			
55	MTSTWVLAGGVLA	1660–1673	14
NS4a			
56	AAYCLATGCVSIIGRLH	1675–1691	17
NS4a and NS4b junction			
57	VVAPDKEVLYEAFDEMEECAS	1697–1717	21
NS4b			
58	QRIAEMLKSKIQGLLQQASKQAQD	1726–1749	24
59	QASWPKVEQFWA	1755–1766	12
60	HMWNFISGIQYLAGLS	1768–1783	16
61	LPGNPAVASMMAFSAA	1785–1800	16
62	TSPLSTSTILLNI	1802–1815	14
63	GGWLASQIAPPAGATGFVVSGLVGAA	1817–1842	26
64	GLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDV	1847–1883	37
65	NLLPGILSPGALVVGICAAILRRHVGPGEAVQWMNRLIAFASRGN HVAPHYVTESDASQSRVTQLLGSLLTITSLRLHNNWITEDCP	1885–1973	89
NS5			
66	GSWLRDWDWVCTILTDFKNWLTSLKLPK	1978–2006	29
67	TRPCGANISGNVR	2031–2044	14
68	GSMRITGPKTCMN	2046–2058	13
69	WQGTFFPINCYTEGQC	2060–2074	15
70	TAIWRVAASEYAEVT	2084–2098	15
71	PSPEFFSWVDGVQIHRFAPTPKPFRRDEVSF	2121–2151	31
72	AARRLARGSPPEASSS	2190–2206	17
73	SQLSAPSLRATC	2208–2219	12
74	WARPDYNPPLVE	2288–2299	12
75	VAGCALPPPK	2311–2320	10
76	SSMPPLEGEPGDPDLE	2390–2405	16
77	SDSGSWSTCSEE	2424–2435	12
NS5a and NS5b junction			
78	VCCSMSYSWTGALITPC	2440–2456	17
NS5b			
79	LSNSLLRYHNKVY	2468–2480	13
80	YDSVLKDIKLAASKVSAR	2506–2523	18
81	LTPPHSARSKYGFGA	2533–2547	15
82	QTPIPTIMAKNEVFVCV	2573–2589	17
83	ARLIVYPDLGVRVCEKMALYD	2599–2619	21
84	SYGFQYSPAQRVE	2632–2644	13
85	DPMGFSYDTRCFDSTVTERDIRTEE	2655–2679	25
86	CGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGDDLWVISESQG	2716–2770	55
87	EEDERNLRAFTEAMTRYSAAPPDPPRPEYDLELITSCSSNVSA	2772–2815	44
88	RYYLTRDPTTP	2822–2833	12
89	ARAAWETVRHSP	2835–2846	12
90	NSWLGNIQYAPT	2848–2860	13
91	QDTLDQNLNFEMYG	2878–2891	14
92	ALRKLGAAPLR	2930–2940	11
93	WKSRAVARASLIS	2942–2955	14
94	CGRYLFNWAVKTKLKLTPPEAR	2963–2985	23
95	LDLSSWFTVGAGGGDI	2987–3002	16
96	VGVGFLLLPAR	3023–3033	11

<sup>a</sup>Amino acid numbering according to JFH-1 polyprotein (49).

quality, as the sequence identity between the parent sequence and mutants was 97.94%. Among the five models provided by Robetta, the first model (Model1) was used for comparison, as the error per residue (in angstroms) was below 0.5 Å. The quality of the model was reassessed by plotting Ramachandran plots, and the models were found

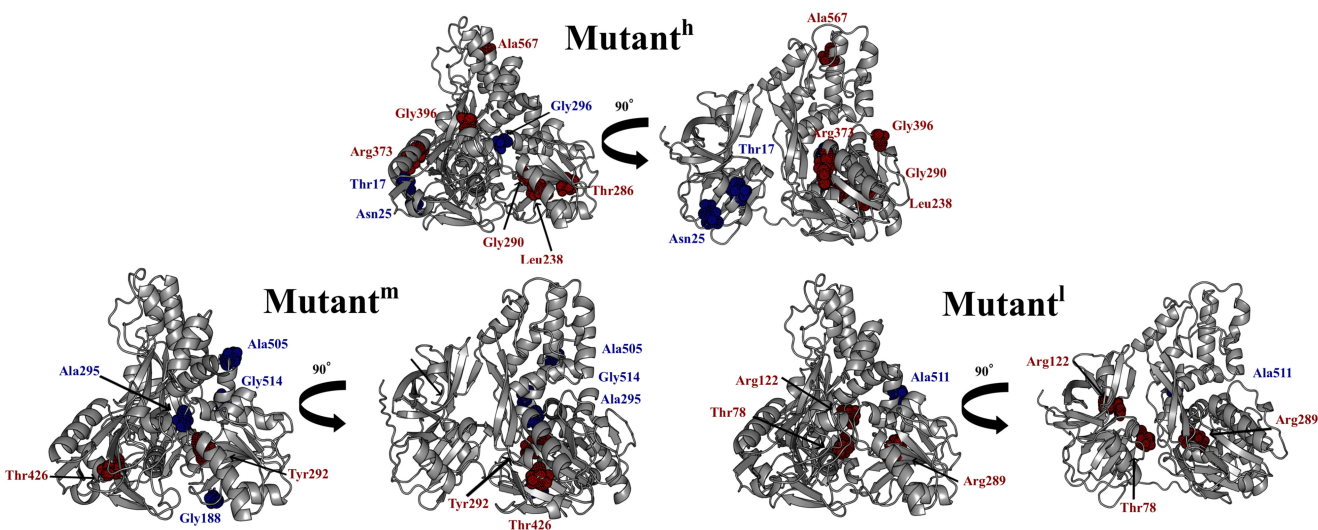




**FIG 6** Mutant<sup>h</sup> (passage 53 virus), Mutant<sup>m</sup> (passage 53 virus), Mutant<sup>l</sup> (passage 67 virus), and parent RNA (passage 53 virus) expression product profiles and their effect on host translation initiation factors upon infection. Huh7.5 cells were mock infected or infected with mutants and parent at an MOI of 0.04. (A) Extracellular HCV genome copy levels (quantified by real-time RT-PCR). Significant differences in genome copy levels between mutants and parent are indicated with asterisks (\*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ). n.s., not significant. (B) Cells were fixed with 4% PFA and stained with mouse monoclonal NSSA antibody, followed by secondary staining with goat anti-mouse antibody 568 and with DAPI. Scale bar, 50  $\mu$ m (all panels). (C) Cell lysates normalized to the amount of cellular protein were resolved on SDS-PAGE gels and stained by Western blotting using anti-core and anti-NS3 antibodies. (D) Cell lysates on Western blots were stained with anti-eIF3 $\beta$ , anti-eIF2 $\alpha$ , and anti-eIF2 $\alpha$ -phospho antibodies.

to have 99.8% of the residues generally allowed in regions of the plot. PROCHECK analysis was performed for all the predicted models, and the quality of the model obtained was comparable to that in the range of a medium- to high-resolution X-ray crystal structure. The resultant structures (Fig. 7) show the positions of mutations in the conserved regions (residues colored red) and in the nonconserved regions (residues colored blue) of NS3 (Table 4).

Plaxco and coworkers showed that contact order (CO) is a first-principle-based method to quantify the global protein topology, where the values obtained are correlated with folding rates in proteins (21). In other words, CO measures the stability of a protein by capturing long-range and short-range interactions. It has been shown that CO values in mutant structures are significantly higher than those in native structures (22). In the case of NS3 proteins of the native and mutants, we used a similar approach to quantify the changes in the structure by analyzing short-range interactions such as main chain-main chain hydrogen bonds and ionic interactions and long-range interactions such as hydrophobic interactions (Table 5). CO values of the NS3 protein of all three mutants were similar to (or slightly higher than) that of native NS3. This correlated with existing reports (23, 24) that to compensate for the loss of short-range covalent interactions, there is rearrangement of noncovalent interactions (short range and long range) to maintain the stability of the protein to function optimally.



**FIG 7** Mutant structures and locations of the putative adaptive mutations in NS3 protein. The predicted model structures of Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup> are in cartoon representation showing the conserved (residues colored red) and nonconserved (residues colored blue), where the mutations are located in D1, D2, D3, and protease domain. Some mutations in Mutant<sup>m</sup> and Mutant<sup>l</sup> are located in the D1-D3 connecting region. Details of mutations are given in Table 4.

DISCUSSION

The experimental genome-wide mutation rate for HCV has been shown to range between 10<sup>-3</sup> and 10<sup>-4</sup> nucleotide substitutions per base per replication (25–27). The genetic heterogeneity and complexity of the resultant mutant clouds of HCV are dynamic, though they collectively contribute to traits of the virus, such as chronic infection and adaptation to the evolving host immune response and other adverse conditions, such as the presence of an antiviral drug (9, 28, 29). Mutant libraries, which contain a heterogeneous mixture of viral genomes, such as those generated in this study, are useful tools to understand the role of the sequence space available to HCV in adaptation and evolution. Using our chosen method, FL-MRS, through a stepwise reduction in template amount or one nucleoside triphosphate, we induced mutations

**TABLE 4** Spatial location of putative adaptive mutations in NS3 protein

Mutant	Residue change in conserved regions (aa position in polyprotein) <sup>a</sup>	Residue change in nonconserved regions (aa position in polyprotein) <sup>b</sup>	Location
Mutant <sup>h</sup>	F238L (1268)		D1 domain
	I286T (1316)		D1 domain
	D290G (1320)		D1 domain
	K373R (1403)		RNA entry site
	D396G (1426)		RNA entry site
	V567A (1597)		D3 domain
		I17T (1047)	D1 domain
Mutant <sup>m</sup>	C292Y (1322)		D1 domain
	I426T (1456)		D2 domain
		S188G (1218)	D1-D3 connecting region
		V295A (1325)	D1-D3 connecting region
		T505A (1535)	D3 domain
		R514G (1544)	D1-D3 connecting region
Mutant <sup>l</sup>	A78T (1108)		Protease domain
	K122R (1152)		Protease domain
	C289R (1319)		D1 domain
		V511A (1541)	D1-D3 connecting region

<sup>a</sup>Colored red in Fig. 5A and 7.

<sup>b</sup>Colored blue in Fig. 5A and 7.

**TABLE 5** Contact order of short-range and long-range interactions in NS3 native and mutant structures

Protein	CO value for:		
	Main chain-main chain hydrogen bond interactions	Ionic interactions	Hydrophobic interactions
Native	16.70365	5.05713	35.81775
Mutant <sup>h</sup>	18.01585	5.538827	37.85895
Mutant <sup>m</sup>	18.03011	6.022187	38.99208
Mutant <sup>l</sup>	18.02536	7.019017	37.22979

at proportions of 0.4 to 66 per 10,000 bp. The mutant libraries provided us an overview of genetic flexibility in the HCV genome and its encoded proteins for virus replication. Our findings suggest that the HCV genome may readily tolerate mutations even at a high proportion and that the majority of putative adaptive mutations are in the nonstructural region of the viral genome. The mutants allowed us study their phenotypic characteristics, such as replication rate, *in vitro* infectivity, and host protein synthesis.

The FL-MRS technique reduces the fidelity of *Taq* DNA polymerase in a controllable fashion to induce random mutations during the synthesis of full-genome templates and virus transcripts (15). The products of EP-PCR are ideal for the synthesis of transcripts of a defined length, since EP-PCR uses a forward primer located upstream of the T7 promoter and a reverse primer that ends precisely at the last nucleotide of the HCV genome in the cDNA clone. Genome-wide insertional mutagenesis has been widely applied in recent years to understand genome tolerance in ss(+) RNA viruses, such as HCV (30), dengue virus (31), and Venezuelan equine encephalitis virus (32). These studies revealed insights of functional profiling of HCV as a demonstration of genome tolerance (30) and discovered tolerant sites for insertion of a reporter tag (31). However, insertional mutagenesis leads to a net increase in length of the gene of interest, a sequence heterogeneity alien to ss(+) RNA virus genomes. Insertional mutagenesis is also laborious and time-consuming, involves bacterial transformation and ligation of plasmids, and is difficult to iterate. In comparison, our FL-MRS method provides a simple, rapid, and bacterium-free approach, except that it requires careful primer design and initial optimization of EP-PCR conditions for amplification of the target genome of interest.

The generated full-genome mutant libraries allowed us decipher the relationship between the proportions of mutations and viability. Parent virus established infection in >90% of cells upon transfection (day 3), whereas the few FFU yielded by the genomes of libraries that were closest to the parent sequence suggest that most new mutations impose structural and functional constraints on the genome (33, 34). The correlation between the increase in proportions of mutations among libraries that are genetically closer to the parent (i.e., libraries with low proportions of mutations) and the decrease in their viable-genome fractions suggests increased deleterious/lethal effects of mutations on genome viability. The fact that genomes of libraries not closely similar to the parent genome (i.e., libraries with high proportions of mutations) are replication competent despite high proportions of mutations indicates that the HCV genome (i) is capable of overcoming lethal/deleterious mutational effects, (ii) can tolerate a high mutation count, and (iii) can retain viability. Initial failures to rescue viable Mutant<sup>h</sup> and Mutant<sup>m</sup> variants reflect deep structural and functional constraints imposed by the mutations in these hypermutated genomes.

HCV causes chronic infection and is constantly exposed to immune pressure in the infected individual. Therefore, the virus genome needs to fix beneficial amino acids in order to escape from the host immune pressure and antiviral exposure but also remain viable and infectious. We identified putative adaptive mutations mostly in regions of the NS2 to NS5A genes, which represent ~60% of the entire genome. High nonsynonymous-to-synonymous substitution ratios for the mutants signify the strong

selection pressure for nonsynonymous mutations over synonymous mutations. Earlier, serially passaged wild-type murine hepatitis virus (WT-MHV) acquired 17 nonsynonymous changes out of 23 mutations (35). In our study, the experimental determination of the nonsynonymous-to-synonymous substitution ratios was possible due to not only the extreme supply of mutations but also the mutational tolerance in protein regions, which may impart a large adaptation potential along with conservation of phenotypic traits. Viruses develop mutations, termed second-site compensatory mutations, to compensate for and restore phenotypic traits such as infectivity and replication (36). The putative adaptive mutations found in mutants might be compensatory and conferred phenotypic traits. To identify conserved and nonconserved regions in the sequence space of the HCV genome, we considered 16 genotype 2a full-genome sequences, which may not have captured sufficient sequence diversity. The absence of a majority (90%) of putative adaptive mutations (found in genomes of all three mutants at the last passage) in natural subtype 2a isolates indicates that these mutations are unlikely to be generated in a replication cycle with the parental sequence. Genomes with artificially increased mutations away from parental sequence may remain viable by the selection of random, beneficial mutations.

Mutant<sup>m</sup> and Mutant<sup>l</sup> variants were associated with increased replicative capability evidenced by enhanced levels of genome copy and viral proteins, while Mutant<sup>h</sup> expression levels were comparable to those of the parent. HCV proteins are multifunctional and interact with each other and multiple host proteins (37). Enhanced replication in association with beneficial mutations may benefit IRES-dependent viral RNA translation and substantially subverted cap-dependent host RNA translation, and this subversion has been associated with increased phosphorylation of the  $\alpha$  subunit of eukaryotic initiation factor 2 (eIF2 $\alpha$ ) (17, 37) and decreased levels of eIF2 $\alpha$  and eIF3 $\beta$ . Random, beneficial mutations might contribute to differences in core and NS3 expression levels of mutants and modulation of host protein synthesis. We do not rule out the possibility of gaining adaptive changes and translational efficiency during passages of the three mutants. Mutants were blocked from entry by anti-CD81 and anti-Claudin-1 (at 5.0  $\mu$ g each; data not shown), which indicates that the mutant viruses followed the classical HCV entry pathway.

Mutants from the libraries of proviral HIV plasmids with random mutations in the envelope (Env) gene when subjected to viral growth in cell culture were shown to undergo strong purifying selection against stop codons and many nonsynonymous mutations; portions of the Env protein shown to gain beneficial mutations and virus remained infectious (33). We also made similar observations, purging out stop codons and deletion-containing genomes and nonsynonymous changes that may not have functional benefits during serial passages of the Mutant<sup>h</sup> population. Insertional mutagenesis revealed that genes coding for hemagglutinin (HA) and NS1 proteins of influenza virus (38) and the NS1 protein of dengue virus (31) are more tolerant than other parts of their genomes. We identified mutations in the genomes of mutants other than those in the 5' UTR and the core and NS5B genes (except one mutation in Mutant<sup>h</sup>). However, in complete contrast, these three regions were shown to have high tolerance for 15-nucleotide (nt) insertions (30).

To understand the impact of putative adaptive mutations on virus protein structure and function, we chose the NS3 protein and performed predictive structural analysis. NS3 contains a helicase domain and an appended serine protease domain, playing a critical role in virus replication that makes it an important drug target (3). Proteins upon mutation undergo subtle rearrangements of interactions, which can be short range (main chain-main chain interactions and ionic interactions) or long range (hydrophobic interactions), which may impose functional and structural constraints (39, 40). To remain functional, these new interactions must compensate for the loss of interactions due to the mutation. Similar quantitated data of long-range and short-range interactions for native (parent) and mutant structures suggest that selected random, beneficial mutations in the NS3 sequence space restored NS3 protein function and stability.

Similarly, deep mutational scanning effectively selected for beneficial mutations with long-range changes that led to functional Env (33, 41).

The full-genome mutant libraries used in this study represent artificial genetic tools; nevertheless, libraries enabled cross dissection of spatially resolved sequence information for comprehensive analysis of mutational tolerance in the HCV genome. The data presented in this study were derived from a limited number of viable variants but certainly reflect variant selection strategy for adaptation. Future studies involving endpoint dilution and infection strategy (42) for the isolation of multiple single variants may help us better understand the mutational tolerance in the HCV genome; however, that is unlikely to alter the overall stance of our study findings.

In summary, this paper documents the ability of the HCV genome to overcome deleterious/lethal mutations by selection of random, beneficial mutations and to remain viable because of a high reproduction rate. Our ability to experimentally increase the mutations in the HCV genome and infect cells *in vitro* with the genomes so generated, combined with advanced sequencing technologies, has given rise to new opportunities to study the evolution of HCV and direct antiviral resistance.

## MATERIALS AND METHODS

**Plasmids and Huh7.5 cell culture.** Plasmid pJFH-1, encoding an HCV genotype 2a strain (13), and pJFH-1/GND, a replication-deficient GND control (GDD catalytic motif of NS5B mutated to GND) were kind gifts from Takaji Wakita, University of Tokyo. Huh7.5 cells and pJ6/JFH-1 were obtained as a kind gift from Charles Rice, Rockefeller University (14). Huh7.5 cells were cultured at 37°C in 5% CO<sub>2</sub> using Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Invitrogen), 100 U/ml penicillin, and 100 µg/ml streptomycin.

**Synthesis and sequencing of full-genome mutant libraries.** The full-length (FL) JFH-1 or J6/JFH-1 genome was randomly mutagenized using error-prone PCR (EP-PCR) to create full-genome mutant libraries. A 9,737-bp-long fragment consisting of the FL HCV genome was amplified in a 50-µl reaction mixture containing primers J-For (5'-GTTTCCAGTCAGCACGTTGTAACGACGGC-3') (33 mer; located 8 nt upstream of the T7 promoter sequence) and J-Rev (5'-CATGATCTGCAGAGACGTTACGGCACTCTC-3') (34 mer; located at the end of the 3' UTR of the HCV genome) at 0.2 µM (each), 1.5 mM MgCl<sub>2</sub>, 2 µl KB Extender, 2 U Platinum Taq DNA polymerase (Invitrogen), dNTPs at a final concentration of 0.2 mM, and 100-, 50-, 25-, and 10-ng amounts of template. The resultant full-genome libraries were designated 100 ng 0.2AGCT, 50 ng 0.2AGCT, 25 ng 0.2AGCT, and 10 ng 0.2AGCT. EP-PCRs with 10 ng template were further imbalanced with final concentrations of dATP at 0.15 and 0.10 mM and the other three dNTPs at a final concentration of 0.2 mM. The resultant full-genome libraries from these reactions were designated 10 ng 0.15A-0.2GCT and 10 ng 0.10A-0.2GCT. The cycling conditions were as follows: 1 cycle of 2 min at 94°C; 30 cycles of 30 s at 94°C, 25 s at 60°C, and 8 min at 72°C; and 1 cycle of 10 min at 72°C. The EP-PCR products were separated using a 0.8% agarose gel, and yields were measured against known amounts of a 1-kb DNA ladder. The 100 ng 0.2AGCT, 50 ng 0.2AGCT, and 25 ng 0.2AGCT libraries (from pJFH-1) synthesized from three independent EP-PCRs were pooled and digested with restriction enzymes EcoRI and NotI and released 2.96-kb fragments cloned into a compatible vector, and bacterial colonies positive for the insert were subjected to Sanger sequencing. A total of 50,000 nt of sequence data per library was obtained. The parent was also subjected to Sanger sequencing. The 10 ng 0.2AGCT, 10 ng 0.15A-0.2GCT, and 10 ng 0.10A-0.2GCT libraries (from pJ6/JFH-1) synthesized from three independent EP-PCRs were pooled and purified using a MinElute PCR purification kit (Qiagen) and submitted for Illumina-based next-generation sequencing (NGS) at the Xcelris Genomics, Ahmedabad, India. Insertion of pJ6/JFH-1 (released by digestion with EcoRI and XbaI) was subjected to NGS to assess background errors imparted by the Illumina platform. Trimmomatic-0.30 was used for quality filtration to generate high-quality data, and the quality value (QV) was set to ≥20. Raw reads from all four conditions were processed using an NGS QC toolkit (43) to exclude those with an average Phred quality score below 30. For each condition, 1,000,000 read pairs were extracted randomly 4 times, resulting in 16 subsets. Read pairs in each subset were aligned against the J6/JFH-1 sequence using a BWA aligner (44). Reads that did not align with the parent were filtered out. The resulting SAM alignment files were converted to BAM format and subsequently sorted and indexed using SAMtools (45). Base counts at individual positions were done using the bam-read count tool (<https://github.com/genome/bam-readcount>), with a minimum mapping quality of 20 and a minimum base quality of 20. Proportions of reads with A, G, C, and T at each location under each of the four conditions were tabulated, and the rates of misincorporation of individual bases were assessed.

**Viral RNA synthesis.** One microgram of column-purified (10 ng of the 0.2 AGCT library was vacuum concentrated) EP-PCR products and XbaI-linearized plasmid DNA (pJFH-1 and pJFH-1/GND) were subjected to *in vitro* transcription per the instructions of the manufacturer (RiboMAX large-scale RNA production system; T7) (Promega). RNase-free DNase-treated transcripts were then purified using an RNA cleanup procedure, and their integrity was analyzed in a 0.8% 3-(N-morpholino)propanesulfonic acid formaldehyde-denaturing agarose gel.

**Transfection of Huh7.5 cells.** To determine the relative replicative capabilities of six mutant libraries, Huh7.5 cells at 70% confluence (achieved 16 h after seeding) in 35-mm culture dishes were



transfected with 5  $\mu$ g of normalized viral FL transcripts using Lipofectamine 2000 (Invitrogen) and incubated in complete DMEM supplemented with 3% FBS. At 3 days posttransfection, virus-infected foci were determined using a focus-forming assay (FFA) and results were expressed in number of focus-forming units (FFU) per 100,000 target cells. Anti-NS5A monoclonal antibody (9E10) was used for detection of FFU as described below. Parent RNA (pJFH-1) was the positive control, and JFH-1/GND was the negative control of replication. One FFU was defined as 1 or more NS5A-positive cells, separated from other NS5A-positive cells by at least 2 NS5A-negative cells. Posttransfection, percentages of NS5A-positive cells were indicated in cases of virus spread reaching >50%.

**Isolation of mutants with altered HCV growth.** To isolate HCV mutants, serial passage of HCV was initiated on Huh7.5 cells at 70% confluence in a 35-mm cell culture dish. Cells were transfected with selected libraries, i.e., 10 ng 0.10A-0.2GCU, 10 ng 0.2AGCU, 50 ng 0.2AGCU, parent, or mock. Multiple transfections of mutant libraries were done; however, transfected cells did not survive due to cytotoxicity induced by rapid virus spread, especially between passages 7 and 10. All virus passages and sequence data shown are from a single transfection experiment where cells supported sustained virus replication. Cells transfected with 50 ng 0.2AGCU and parent were passaged at 1:3 split for 67 and 53 passages, respectively. Cells transfected with 10 ng 0.1A-0.2GCU and 10 ng 0.2AGCU libraries were scaled up as shown in Fig. 3A. Before cell split, virus supernatants were harvested and stored at  $-80^{\circ}\text{C}$  for further analysis. At 3 weeks prior to the last passage (passage 53 for 10 ng 0.1A-0.2GCU, 10 ng 0.2AGCU, and parent libraries, and passage 67 for 50 ng 0.2AGCU library), viral RNA was isolated using a QIAamp viral RNA minikit (Qiagen) from all virus passages, cDNAs were synthesized using Superscript III reverse transcriptase (Invitrogen), the NS3 gene was amplified (coding nt 3080 to 4796) using high-fidelity PCR, and products were column purified, with an "A" overhang added, followed by another round of purification, ligated with TA vector, and used to transform *Escherichia coli* DH5a. The NS3 gene was amplified and cloned as described above from virus passages 0 (input), 17, 47, and 53 of the 10 ng 0.10A-0.2GCU library. Before the last virus passage, transfected cells were expanded; large volumes of virus supernatants were collected and stored. Virus supernatants were concentrated, as indicated, using 100-kDa-molecular-size-cutoff centrifugal filters. Virus titers were determined using the 50% tissue culture infectious dose (TCID<sub>50</sub>) method, and the complete genome sequence of the predominant variant was determined by Sanger sequencing. In parallel, mock-transfected cells were maintained to monitor contamination of cells with virus.

**Infectious virus titer determination.** For titration of infectious HCV, an FFA was performed on Huh7.5 cells as described previously (14). Briefly, serially diluted virus supernatants applied to  $6.5 \times 10^3$  Huh7.5 cells were plated in a 96-well plate (seeded 16 h prior to infection). Three days postinfection, the cells were washed twice with phosphate-buffered saline (PBS), fixed, permeabilized with ice-cold methanol, washed three times again, and incubated in blocking solution (PBS-T containing 1% bovine serum albumin and 0.2% skim milk) for 30 min at room temperature. Endogenous peroxidase activity was blocked by incubating the cells with 3% H<sub>2</sub>O<sub>2</sub> in PBS for 5 min. The cells were then incubated for 1 h with 1:20,000-diluted anti-NS5A mouse monoclonal antibody (MAb) 9E10 (stock was 1 mg/ml; a gift of Charles Rice, Rockefeller University), washed three times in PBS-T, and incubated with horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (1:4,000 dilution), followed by incubation with 3,3'-diaminobenzidine (DAB) substrate. FFU were visualized using a light microscope, and the TCID<sub>50</sub> was determined (46).

**Huh7.5 cell infections.** Naïve Huh7.5 cells,  $0.4 \times 10^6$ , were mock infected or infected with mutant viruses or the parent at a fixed TCID<sub>50</sub> of 16,000 (multiplicity of infection [MOI] of 0.04). Cells and supernatants were collected on day 5 postinfection.

**Quantification of HCV genome using qRT-PCR.** Quantitative reverse transcription-PCR (qRT-PCR) was carried out to quantify genome copies in virus supernatants. Viral RNA was isolated using a QIAamp viral RNA minikit (Qiagen) and subjected to qRT-PCR using a TaqMan RNA-to-C<sub>T</sub> 1-step kit (Invitrogen), per the manufacturer's instructions, and previously described primer (R6-130-S17 and R6-290-R19) and probe (R6-148-S21FT) sequences (47). Run conditions were as follows: 48°C for 20 min, 95°C for 10 min, and then 40 cycles of 95°C for 15 s and 60°C for 1 min. A standard curve was obtained with known amounts of HCV transcripts used for quantification. Negative controls (without template RNA and RNA from mock-infected cells) were run in parallel with each reaction to ascertain the absence of contamination with undesired templates.

**Western blotting.** At day 5 postinfection, cells were pelleted by centrifugation and lysed in radioimmunoprecipitation assay buffer at 4°C for 15 min. The protein concentration was estimated with the Bradford assay reagent (Pierce), and the lysates were resolved by electrophoresis on 12% SDS polyacrylamide gels and then electroblotted onto polyvinylidene difluoride membranes (Millipore, Billerica, MA). Viral and cellular proteins were detected by overnight incubation at 4°C with one of the following primary antibodies: mouse MAb core (C7-50; diluted 1:3,000), mouse MAb NS3 (1:2,000), rabbit polyclonal Ab eIF3 $\beta$  (1:1,000), mouse MAb eIF2 $\alpha$  (1:200), and rabbit polyclonal Ab eIF2 $\alpha$  phospho specific (Invitrogen) (1:500). After washing, the blots were exposed to 1:3,000-diluted appropriate secondary antibodies conjugated with HRP for 1 h at room temperature. The reaction products were visualized using enhanced chemiluminescence reagent (Bio-Rad) according to the manufacturer's instructions. The loading control was  $\beta$ -actin.

**Immunofluorescence assay.** At day 5 postinfection, cells were fixed in 4% paraformaldehyde and permeabilized with PBS containing 0.1% Triton X-100 for 5 min and washed three times for 5 min each in PBS. Nonspecific antigens were blocked (3% bovine serum albumin and 10% normal goat serum in PBS), and the cells were incubated with anti-NS5A MAb (1:25,000) (9E10) at room temperature for 2 h and then washed three times for 5 min each in PBS. This was followed by incubation with secondary antibody conjugated to Alexa Fluor 568 for 30 min and counterstaining with 1  $\mu$ g/ml 4',6-diamidino-2-

phenylindole (DAPI). After washes, the coverslips were mounted on slides, which were then analyzed by fluorescence microscopy, using a Nikon A1 spinning-disk confocal microscope. Densitometry analysis of immunofluorescence staining was performed on digitized images using ImageJ.

**Blocking of HCV entry.** Anti-CD81 (clone M38) and anti-Claudin-1 (clone 2H10D10) MAbs were purchased from Invitrogen. Huh7.5 cells seeded in 48-well plates 16 h earlier were mock treated or treated with 0.5, 1.0, 2.5, and 5.0  $\mu\text{g}$  of anti-CD81 and anti-Claudin-1 for 30 min at 37°C and 5%  $\text{CO}_2$  with intermittent shaking (twice) for 2 min and then infected at an MOI of 0.04 with parent or mutant viruses. Infection was quantified 3 days later by an FFA.

**HCV genome sequencing.** Genomes of HCV mutants from the last passages of all libraries and the parent were isolated from 140  $\mu\text{l}$  of cell culture supernatant using a QIAamp viral RNA extraction kit (Qiagen). cDNAs were synthesized from 5  $\mu\text{l}$  of undiluted or 1:5-diluted eluates using Superscript III (Invitrogen), and the complete genome was PCR amplified in 12 overlapping fragments. Primer sequences have been described previously (48). Carryover contamination was monitored using negative controls. Chromatograms were processed, sequences were assembled, and contigs were generated using the CLC Main Workbench.

**Shannon entropy.** Sequence conservation among 16 full-length HCV subtype 2a sequences (derived from the HCV sequence databases of the LANL and euHCVdb) was measured using entropy. The entropy score for each amino acid site was calculated by using the Shannon Heterogeneity In Alignments Tool v1.0 (<http://evolve.zoo.ox.ac.uk/Evolve/SHiAT.html>) (19). Amino acids are numbered relative to the JFH-1 polyprotein (GenBank accession number [AB047639](#)) (49).

**Mutant structure prediction and CO calculations.** The three-dimensional (3D) structures of the NS3 proteins of Mutant<sup>h</sup>, Mutant<sup>m</sup>, and Mutant<sup>l</sup> were determined using the Robetta method (<http://new.rosetta.org>) to obtain high-quality structures (50). For the three mutants, the template used was the native structure, JFH-1, whose coordinates were retrieved from the Protein Data Bank (PDB ID 5WDX). By use of a comparative modeling approach, a sampling of 1,000 models was obtained. Contact order (CO) calculations were performed as described previously (21). Specifically, the total list of noncovalent interactions was taken and the distance between interacting residues (in terms of sequence numbers) was calculated. The sum of all distances was divided by the total length of the NS3 protein (631 residues) for normalization. The Protein Interactions Calculator (<http://pic.mbu.iisc.ernet.in/job.html>) tool (51) was used to identify main chain-main chain hydrogen bonds, ionic interactions, and hydrophobic interactions in the NS3 structures of the parent and mutants.

**Statistical analysis.** All statistical analyses were performed by using GraphPad Prism7. The Kruskal-Wallis test followed by *post hoc* analysis by Dunn's test was used for multiple comparisons.

**Data availability.** The models are deposited in ModelArchive (Mutant<sup>h</sup>, <https://www.modelarchive.org/doi/10.5452/ma-o9v3l>; Mutant<sup>m</sup>, <https://www.modelarchive.org/doi/10.5452/ma-r30z5>; Mutant<sup>l</sup>, <https://www.modelarchive.org/doi/10.5452/ma-3ybvkl>).

## ACKNOWLEDGMENTS

We are grateful to Jordan Feld (Toronto Center for Liver Disease) for useful discussions.

This study was supported by grant BT/PR10906/MED/29/860/2014 to N.S.V. and R.A. from the Department of Biotechnology, Ministry of Science and Technology, Government of India. D.S., S.S., and S.K. were supported by graduate scholarships from Shiv Nadar University.

No author has any financial, professional, or personal potential conflicts that are relevant to the article.

## REFERENCES

- World Health Organization. 2017. Global hepatitis report. World Health Organization, Geneva, Switzerland.
- Westbrook RH, Dusheiko G. 2014. Natural history of hepatitis C. *J Hepatol* 61:s58–s68. <https://doi.org/10.1016/j.jhep.2014.07.012>.
- Manns MP, Buti M, Gane E, Pawlotsky JM, Razavi H, Terrault N, Younossi Z. 2017. Hepatitis C virus infection. *Nat Rev Dis Primers* 3:17006. <https://doi.org/10.1038/nrdp.2017.6>.
- Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P. 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes, updated criteria assignment web resource. *Hepatology* 59:318–327. <https://doi.org/10.1002/hep.26744>.
- Eigen M. 1971. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523. <https://doi.org/10.1007/bf00623322>.
- Eigen M, Schuster P. 1977. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64:541–565. <https://doi.org/10.1007/bf00450633>.
- Forns X, Purcell RH, Bukh J. 1999. Quasispecies in viral persistence and pathogenesis of hepatitis C virus. *Trends Microbiol* 7:402–410. [https://doi.org/10.1016/s0966-842x\(99\)01590-5](https://doi.org/10.1016/s0966-842x(99)01590-5).
- Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS. 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* 282:103–107. <https://doi.org/10.1126/science.282.5386.103>.
- Bowen DG, Walker CM. 2005. The origin of quasispecies: cause or consequence of chronic hepatitis C viral infection. *J Hepatol* 42:408–417. <https://doi.org/10.1016/j.jhep.2004.12.013>.
- Pellerin M, Lopez-Aguirre Y, Penin F, Dhumeaux D, Pawlotsky JM. 2004. Hepatitis C virus quasispecies variability modulates nonstructural protein 5A transcriptional activation, pointing to cellular compartmentalization of virus-host interactions. *J Virol* 78:4617–4627. <https://doi.org/10.1128/jvi.78.9.4617-4627.2004>.
- Blight KJ, Kolykhalov AA, Rice CM. 2000. Efficient initiation of HCV RNA replication in cell culture. *Science* 290:1972–1974. <https://doi.org/10.1126/science.290.5498.1972>.
- Lohmann V, Körner F, Koch J, Herian U, Theilmann L, Bartschlag R. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma

- cell line. *Science* 285:110–113. <https://doi.org/10.1126/science.285.5424.110>.
13. Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Kräusslich HG, Mizokami M, Bartenschlager R, Liang TJ. 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* 11:791–796. <https://doi.org/10.1038/nm1268>.
  14. Lindenbach BD, Evans MJ, Syder AJ, Wölk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, Rice CM. 2005. Complete replication of hepatitis C virus in cell culture. *Science* 309:623–626. <https://doi.org/10.1126/science.1114016>.
  15. Agarwal S, Baccam P, Aggarwal R, Veerapu NS. 2018. Novel synthesis and phenotypic analysis of mutant clouds for hepatitis E virus genotype 1. *J Virol* 92:e01932–17. <https://doi.org/10.1128/JVI.01932-17>.
  16. Qi H, Wu NC, Du Y, Wu TT, Sun R. 2015. High-resolution genetic profile of viral genomes: why it matters. *Curr Opin Virol* 14:62–70. <https://doi.org/10.1016/j.coviro.2015.08.005>.
  17. Elena SF, Sanjuán R. 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol* 79:11555–11558. <https://doi.org/10.1128/JVI.79.18.11555-11558.2005>.
  18. Contreras AM, Hiasa Y, He W, Terella A, Schmidt EV, Chung RT. 2002. Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system. *J Virol* 76:8505–8517. <https://doi.org/10.1128/jvi.76.17.8505-8517.2002>.
  19. Humphreys I, Fleming V, Fabris P, Parker J, Schulenberg B, Brown A, Demetriou C, Gaudieri S, Pfaffertott K, Lucas M, Collier J, Huang KH, Pybus OG, Klennerman P, Barnes E. 2009. Full-length characterization of hepatitis C virus subtype 3a reveals novel hypervariable regions under positive selection during acute infection. *J Virol* 83:11456–11466. <https://doi.org/10.1128/JVI.00884-09>.
  20. Perales C, Beach NM, Gallego I, Soria ME, Quer J, Esteban JI, Rice C, Domingo E, Sheldon JJ. 2013. Response of hepatitis C virus to long-term passage in the presence of alpha interferon: multiple mutations and a common phenotype. *J Virol* 87:7593–7607. <https://doi.org/10.1128/JVI.02824-12>.
  21. Plaxco KW, Simons KT, Baker D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 10:985–994. <https://doi.org/10.1006/jmbi.1998.1645>.
  22. Rader AJ, Yennamalli RM, Harter AK, Sen TZ. 2012. A rigid network of long-range contacts increases thermostability in a mutant endoglucanase. *J Biomol Struct Dyn* 30:628–637. <https://doi.org/10.1080/07391102.2012.689696>.
  23. Bigman LS, Levy Y. 2018. Stability effects of protein mutations: the role of long-range contacts. *J Phys Chem B* 122:11450–11459. <https://doi.org/10.1021/acs.jpcc.8b07379>.
  24. Bandyopadhyay B, Mondal T, Unger R, Horovitz A. 2019. Contact order is a determinant for the dependence of GFP folding on the chaperonin GroEL. *Biophys J* 116:42–48. <https://doi.org/10.1016/j.bpj.2018.11.019>.
  25. Abe K, Inchauspe G, Fujisawa K. 1992. Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan. *J Gen Virol* 73:2725–2729. <https://doi.org/10.1099/0022-1317-73-10-2725>.
  26. Ogata N, Alter HJ, Miller RH, Purcell RH. 1991. Nucleotide sequence and mutation rate of H strain of hepatitis C virus. *Proc Natl Acad Sci U S A* 88:3392–3396. <https://doi.org/10.1073/pnas.88.8.3392>.
  27. Geller R, Estada U, Peris JB, Andreu I, Bou JV, Garijo R, Cuevas JM, Sabariego R, Mas A, Sanjuán R. 2016. Highly heterogeneous mutation rates in the hepatitis C virus genome. *Nat Microbiol* 1:16045. <https://doi.org/10.1038/nmicrobiol.2016.45>.
  28. Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzer A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ. 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288:339–344. <https://doi.org/10.1126/science.288.5464.339>.
  29. Pawlotsky JM. 2011. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. *Hepatology* 53:1742–1751. <https://doi.org/10.1002/hep.24262>.
  30. Arumugaswami V, Remenyi R, Kanagavel V, Sue EY, Ho TN, Liu C, Fontanes V, Dasgupta A, Sun R. 2008. High-resolution functional profiling of hepatitis C virus genome. *PLoS Pathog* 4:e1000182. <https://doi.org/10.1371/journal.ppat.1000182>.
  31. Eyre NS, Johnson SM, Eltahla AA, Aloia AL, McDavitt CA, Bull RA, Beard MR. 2017. Genome-wide mutagenesis of dengue virus reveals plasticity of the NS1 protein and enables generation of infectious tagged reporter viruses. *J Virol* 91:e01455–17. <https://doi.org/10.1128/JVI.01455-17>.
  32. Beitzel BF, Bakken RR, Smith JM, Schmaljohn CS. 2010. High-resolution functional mapping of the Venezuelan equine encephalitis virus genome by insertional mutagenesis and massively parallel sequencing. *PLoS Pathog* 6:e1001146. <https://doi.org/10.1371/journal.ppat.1001146>.
  33. Haddox HK, Dings AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. *Elife* 28:e34420.
  34. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The mutational robustness of influenza A virus. *PLoS Pathog* 12:e1005856. <https://doi.org/10.1371/journal.ppat.1005856>.
  35. Graepel KW, Lu X, Case JB, Sexton NR, Smith EC, Denison MR. 2017. Proofreading-deficient coronaviruses adapt for increased fitness over long-term passage without reversion of exoribonuclease-inactivating mutations. *mBio* 7:e01503–17. <https://doi.org/10.1128/mBio.01503-17>.
  36. Noviello CM, López CS, Kukull B, McNett H, Still A, Eccles J, Sloan R, Barklis E. 2011. Second-site compensatory mutations of HIV-1 capsid mutations. *J Virol* 85:4730–4738. <https://doi.org/10.1128/JVI.00099-11>.
  37. Walsh D, Mohr I. 2011. Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol* 9:860–875. <https://doi.org/10.1038/nrmicro2655>.
  38. Heaton NS, Sachs D, Chen CJ, Hai R, Palese P. 2013. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. *Proc Natl Acad Sci U S A* 110:20248–20253. <https://doi.org/10.1073/pnas.1320524110>.
  39. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Imura S, Gajiwala K, Scholtz JM, Grimsley GR. 2011. Contribution of hydrophobic interactions to protein stability. *J Mol Biol* 408:514–528. <https://doi.org/10.1016/j.jmb.2011.02.053>.
  40. Pace CN, Fu H, Lee Fryar K, Landua J, Trevino SR, Schell D, Thurlkill RL, Imura S, Scholtz JM, Gajiwala K, Sevcik J, Urbanikova L, Myers JK, Takano K, Hebert EJ, Shirley BA, Grimsley GR. 2014. Contribution of hydrogen bonds to protein stability. *Protein Sci* 23:652–661. <https://doi.org/10.1002/pro.2449>.
  41. Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. 2000. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* 8:1329–1339. [https://doi.org/10.1016/s0969-2126\(00\)00547-5](https://doi.org/10.1016/s0969-2126(00)00547-5).
  42. Sugiyama N, Murayama A, Suzuki R, Watanabe N, Shiina M, Liang TJ, Wakita T, Kato T. 2014. Single strain isolation method for cell culture-adapted hepatitis C virus by end-point dilution and infection. *PLoS One* 9:e98168. <https://doi.org/10.1371/journal.pone.0098168>.
  43. Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. <https://doi.org/10.1371/journal.pone.0030619>.
  44. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
  45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
  46. Reed LJ, Muench H. 1938. A simple method of estimating fifty percent end points. *Am J Hyg* 27:493–497. <https://doi.org/10.1093/oxfordjournals.ajeh.a118408>.
  47. Takeuchi T, Katsume A, Tanaka T, Abe A, Inoue K, Tsukiyama-Kohara K, Kawaguchi R, Tanaka S, Kohara M. 1999. Real-time detection system for quantification of hepatitis C virus genome. *Gastroenterology* 116:636–642. [https://doi.org/10.1016/s0016-5085\(99\)70185-x](https://doi.org/10.1016/s0016-5085(99)70185-x).
  48. Russell RS, Meunier JC, Takikawa S, Faulk K, Engle RE, Bukh J, Purcell RH, Emerson SU. 2008. Advantages of a single-cycle production assay to study cell culture-adaptive mutations of hepatitis C virus. *Proc Natl Acad Sci U S A* 105:4370–4375. <https://doi.org/10.1073/pnas.0800422105>.
  49. Kato T, Furusaka A, Miyamoto M, Date T, Yasui K, Hiramoto T, Nagayama K, Tanaka T, Wakita T. 2001. Sequence analysis of hepatitis C virus isolated from a fulminant hepatitis patient. *J Med Virol* 64:334–339. <https://doi.org/10.1002/jmv.1055>.
  50. Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32:W526–W531. <https://doi.org/10.1093/nar/gkh468>.
  51. Tina KG, Bhadra R, Srinivasan N. 2007. PIC: protein interactions calculator. *Nucleic Acids Res* 35:W473–W476. <https://doi.org/10.1093/nar/gkm423>.