Supporting Information

Mauger et al. 10.1073/pnas.1416266112

SI Methods

HCV Genomic Subclones and RNA Transcripts. DNA plasmids with HCV-JFH1 and HCV-Con1 sequences were obtained from Apath. DNA plasmids containing the HCV-H77c genome were a generous gift from Shelton Bradrick, Duke University, Durham, NC. Full-length genomic RNAs were prepared as described (1, 2). Briefly, plasmids were linearized with the appropriate restriction endonuclease, and RNA transcripts were synthesized in vitro (T7 MegaScript; Ambion). RNA was purified by size exclusion chromatography (ChromaSpin 1000 column; Clontech), and lengths were confirmed by denaturing agarose gel electrophoresis. Structure-disrupting mutations were created in the H77S.3 and JFH1-QL genomic clones and their /GLuc2A counterparts (3, 4) using PCR mutagenesis (Table S2).

SHAPE Modification of Genomic RNAs. Full-length HCV genomic RNAs were heated for 10 s at 90 °C and placed on ice. RNAs were folded in 1× folding buffer [50 mM Hepes (pH 8.0), 5 mM MgCl$_2$, 200 mM potassium acetate (pH 7.5)] at a final concentration of 45 ng/μL by incubating at 65 °C for 5 min followed by slow cooling to 37 °C. RNAs were treated with 0.1 volume of DMSO or with 100 mM 1M7 (5) in DMSO at 37 °C for 5 min. EDTA was added to 10-mM final concentration, reactions were chilled on ice, and the RNA was precipitated with ethanol. For the denaturing modification control, RNA was resuspended in 1× denaturing buffer [50 mM Hepes (pH 8), 4 mM EDTA, 50% (vol/vol) formamide] at a final concentration of 45 ng/μL, heated to 95 °C for 1 min, and modified with 0.1 volume of 100 mM 1M7 in DMSO. Reactions were chilled on ice for 2 min, and RNA was recovered by precipitation with ethanol.

SHAPE-MaP Library Construction. DNA libraries for massively parallel sequencing, consisting of plus-1M7 reagent, no-reagent, and denaturing control reactions, were prepared as described previously (6). Briefly, RNA was fragmented at 95 °C in 2.5× first-strand synthesis buffer (Invitrogen) for 4 min and purified through a G-50 Microspin Column (GE Lifesciences). SHAPE-MaP reverse transcription reactions (20 μL) contained 200 ng of random nonamer primers (New England Biolabs) in 1× SHAPE-MaP buffer [50 mM Tris·HCl (pH 8), 75 mM KCl, 6 mM MnCl$_2$] with 1 μL SuperScript II reverse transcriptase (Invitrogen). Reactions were incubated at 42 °C for 3 h, and cDNA products were purified using a G-50 Microspin Column (GE Lifesciences). Purified cDNA was then converted to PCR libraries (NEBNext sample preparation; New England Biolabs) beginning with the second-strand synthesis step. PCR libraries were quantified on a Qubit fluorometer (Life Technologies), analyzed with a Bioanalyzer DNA kit (Agilent), and sequenced on a MiSeq instrument (Illumina).

SHAPE-MaP Data Analysis. Each of the three HCV genomes were sequenced at a depth that allowed accurate recovery of the SHAPE structural information (6). All constructs showed notable decreases in read depth at two locations: (*i*) within the first 20 nt of each genome and (*ii*) within 30 nt of the poly-U stretch in the 3′ UTR. Excluding these two regions, read depth coverage for each of the HCV genomic RNAs (nucleotides 10–9400) is given in the chart below.

		Read depth	
RNA	Sample	(Median)	(Minimum)
H77	No reagent	13,459	2,318
	1M7	26,090	4,364
	Denatured	27,020	5,743
Con1b	No reagent	51,101	5,280
	1M7	37,997	3,533
	Denatured	35,857	3,976
JFH1	No reagent	35,998	2,254
	1M7	68,877	4,956
	Denatured	63,109	6,909

Reproducibility was assessed by performing full biological replicate SHAPE-MaP experiments on the H77 genomic RNA with matching read depths. SHAPE-MaP reactivities for the replicates gave a correlation value ($r = 0.88$), comparable to the known variability of SHAPE-MaP experiments.

SHAPE-MaP reactivities for each nucleotide within a given genome were generated from the raw FASTQ files using the SHAPE-MaP analysis pipeline (6). Because sequencing depth was low in the poly-U region and X-tail in the 3′ UTR, SHAPE reactivities are not reported for these regions. SHAPE reactivities for all three HCV RNA genomes are provided in Dataset S1.

Secondary Structure Models and Shannon Entropies. The minimum free-energy genome secondary structure models and base-pairing probabilities were generated using SHAPE reactivities as constraints using the SHAPE-MaP folding pipeline (6). Briefly, folding was performed using RNAstructure (v5.6) (7) and folding was performed in a stepwise manner. The genome primary sequence and SHAPE reactivities were divided into overlapping windows, 2,000 nt in length, and offset by 100 nt. These windows were folded using the *Partition* routine of RNAstructure with the following constraints: no maximum distance, SHAPEslope = 1.8, SHAPEintercept = –0.6, and the max pairing distance set to 500 nt. Shannon entropies were calculated for each window (6, 8) and averaged for overlapping regions to obtain genome-wide Shannon entropies and base pair probabilities. Base pairs predicted to form with >99% confidence were forced to form in calculating the minimum free-energy fold. The minimum free-energy structure was calculated using the *Fold* routine in RNAstructure for all possible 4,000-nt windows in 375-nt steps with the constraints listed above. Final minimum free-energy structures were assembled from the most common base pairs identified within overlapping regions. Secondary structure models in connect (.ct) format are provided in Dataset S2.

Comparison of Genome Structure Models Generated With and Without SHAPE Data. Secondary structure models were generated using the same procedures described above, except that experimental SHAPE-MaP reactivities were omitted (Fig. S6).

Helix Length and RNase L Cleavage Site Analyses. Minimum free-energy genome secondary structure models were analyzed to identify all continuous helices in each genome and to generate histograms of helix lengths. In a separate analysis, we identified the locations and reactivities of all potential high-affinity RNase L cleavage sites (UU/UA) dinucleotides within the H77 genome. The regions of RNase L cleavage sites reported by Han et al. (9) were compared with the entire population. The statistical significance of the observed differences was assigned by bootstrapping (with replacement) 10,000 equivalently sized populations of randomly selected cleavage sites from the total pool of potential RNase L sites. The distribution of mean SHAPE-MaP reactivities from the bootstrapped populations was compared with the mean SHAPE-MaP reactivity of the active RNase L cleavage sites.

Comparative Analysis of Structural Models. Median SHAPE reactivities and Shannon entropies were calculated over centered, 55-nt windows for each position in the genome. For each of the subtypes tested, we identified genomic regions where the median SHAPE reactivities and median Shannon entropy were below the global median values for at least 40 nt. Positions in the genome having median SHAPE reactivity or median Shannon entropy below the global medians within all three subtypes tested were defined as mutually structured and mutually low-entropy, respectively. Regions of conserved base pairing were computationally identified using a genomic alignment of the primary nucleotide sequences for H77, JFH1, and Con1. Regions of conserved base pairing (Table S1) were required to be longer than 75 nt in length and the predicted secondary structures had to have ≥75% secondary structure similarity, measured using the mountain similarity metric (10).

Synonymous Substitution Rate Analysis. Synonymous substitution rates were analyzed for conserved base pairings over 12 regions located within the ORF, using seven different alignments of clinically isolated HCV genomic sequences downloaded from publically accessible databases (Dataset S3). Each alignment contained 47–100 sequences and was aligned using MAFFT v6.952 with default settings (11). The first alignment contained a representative selection of HCV sequences across all seven HCV genotypes. The other six alignments contained segregated populations of sequences from genotypes 1a, 1b, 2, 3, 4, and 6. From the perspective of synonymous substitution rate inference, these six alignments represent independent HCV replicates. We were able to split genotype 1 into 1a and 1b for the analyses because of the large number of publicly available sequences. Genotypes 5 and 7 were excluded from the subtype-specific analysis because too few sequences are available to afford sufficient statistical power for synonymous substitution rate and coevolution analyses. Synonymous substitution rates at individual codons were inferred using FUBAR (12). A Mann–Whitney *U* test was then used to assess whether synonymous substitution rates within a region were significantly higher or lower than expected, where the expectation was based on synonymous substitution rates throughout the remainder of the genome. The *z*-scores lower than −1.96 were considered statistically significant evidence ($P < 0.05$) that the synonymous substitution rate of codons within a given region was lower than that of the remainder of the coding region.

Complementary Coevolution Analysis. Complementary coevolution was defined as nucleotide pairs that detectably covary to maintain Watson–Crick or GU-wobble base-pairing and thus base-pairing potential. A representative alignment of 250 HCV sequences (Datasets S1 and S2) was analyzed for evidence of complementarily coevolving nucleotide sites by examining every pair of polymorphic nucleotide sites within 100 nt of one another using HyPhy (13), as described previously (14). The maximum distance between potentially coevolving sites was restricted to 100 nt because of the long running time of this analysis. This restriction did not significantly impact subsequent tests for associations between base pairing and complementary coevolution because of the vast majority of the SHAPE-identified base pairings fell within this range. This analysis was used to generate a matrix of coevolution *P* values, where a *P* value < 0.05 was taken as statistically significant evidence that the evolution of a given pair of nucleotides was better supported by a model of complementary coevolution than a model of independent evolution. Using the matrix of coevolution *P* values, a *z*-score was calculated for the entire HCV genome and each of the 15 regions of conserved base pairing. For each region, a Mann–Whitney *U* test was used to calculate a *z*-score by comparing the coevolution *P* values corresponding to SHAPE-identified base-pairings within a given region to coevolution *P* values corresponding to all other pairs of nucleotides within the same region that were not base-paired within the SHAPE-informed structures. If *z*-scores were lower than −1.96, it was taken as significant evidence ($P < 0.05$) of better than expected correspondence between coevolving site pairs and SHAPE-inferred base-paired sites.

Comparison with Prior Bioinformatic Analysis of HCV Genome Sequences. Two prior studies noted that synonymous substitution rates within HCV coding regions are lowest in regions that were predicted to have the greatest degree of secondary structure (15, 16). Similarly, these studies also found that at least some of the nucleotides, predicted to be base-paired within secondary structures, have coevolved to maintain complementarity (15). Here we have explicitly tested for and statistically confirmed these associations using a maximum-likelihood synonymous substitution rate and coevolution inference methods and SHAPE-derived structural data (as opposed to computationally inferred structures). In addition, our coevolution and synonymous substitution rate analyses were performed on HCV datasets that were >10-fold larger than those that could be examined previously, which enabled us to statistically test—within individual structural elements—for associations between nucleotide coevolution and synonymous substitution rates.

HCV RNA Transfection. RNA was synthesized in vitro (T7 MEGAScript reagents; Ambion) after linearizing plasmids with XbaI. Following treatment with RNase-free DNase to remove template DNA, RNA was purified using affinity chromatograph (RNAeasy columns; Qiagen). RNA transfection was carried out by Trans-IT mRNA transfection (Mirus). The replication capacity of GLuc-containing constructs was determined by transfecting 250 ng RNA into 7.5×10^4 Huh-7.5 cells in 24-well plates. The capacity of the HCV mutants (no GLuc insertion) to produce infectious virus particles was determined by transfecting 1.25 µg RNA into 6×10^5 cells in six-well plates.

***Gaussia* Luciferase Activity Assay.** Structure-disrupting mutations were created in the H77S.3/GLuc2a variant (17) and in a cell culture-adapted JFH1 (JFH1-QL/GLuc2A, genotype 2a) (18) variant that produce *Gaussia princeps* luciferase (GLuc) from an in-frame insertion of the GLuc sequence between p7 and NS2A in the viral polyprotein. Following RNA transfection into human hepatoma Huh-7.5 cells, cell-culture supernatant fluids were collected and fresh medium added at 24-h intervals. Secreted GLuc activity was measured in 25 µL aliquots of the supernatant fluids in the Biolux *Gaussia* luciferase assay (New England Biolabs). The luminescent signal was measured on a multimode microplate reader (Synergy 2; Bio-Tek) (17).

Virus Yield Assays. Human hepatoma Huh-7.5 cells were split at a 1:2 ratio at 24 h after transfection of viral RNAs (with no GLuc insertion), and the medium replaced with fresh media containing 100 mM Hepes every 24 h thereafter. Cell culture supernatant fluids were collected at 72-h posttransfection for virus titration. Infectious virus was quantified using a fluorescent focus virus titration assay (17, 19). Briefly, cells were seeded in 48-well plates at a density of 1×10^5 cells per well 24 h before inoculation with 100 µL of virus-containing medium. Cells were maintained at 37 °C in a 5% CO_2 environment and fed with 200 µL medium 24 h later. After 24 h (JFH1-QL and its mutants) or after 48 h (H77S.3 and its mutants) of additional incubation, cells were fixed in methanol-acetone (1:1) at −20 °C for 10 min and stained with a monoclonal antibody (C7-50; Thermo Scientific) to the HCV core protein (1:300). After extensive washing, the cells were stained with Alexa488-conjugated goat anti-mouse IgG antibody. Clusters of infected cells stained for Core antigen were considered to constitute a single infectious focus-forming unit.

1. Wakita T, et al. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* 11(7):791–796.
2. Yanagi M, Purcell RH, Emerson SU, Bukh J (1997) Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc Natl Acad Sci USA* 94(16):8738–8743.
3. Tuplin A, Evans DJ, Simmonds P (2004) Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* 85(Pt 10):3037–3047.
4. Vassilaki N, et al. (2008) Role of the hepatitis C virus core+1 open reading frame and core *cis*-acting RNA elements in viral RNA translation and replication. *J Virol* 82(23):11503–11515.
5. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129(14):4144–4145.
6. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11(9):959–965.
7. Reuter JS, Mathews DH (2010) RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129.
8. Huynen M, Gutell R, Konings D (1997) Assessing the reliability of RNA folding using statistical mechanics. *J Mol Biol* 267(5):1104–1112.
9. Han JQ, Wroblewski G, Xu Z, Silverman RH, Barton DJ (2004) Sensitivity of hepatitis C virus RNA to the antiviral enzyme ribonuclease L is determined by a subset of efficient cleavage sites. *J Interferon Cytokine Res* 24(11):664–676.
10. Moulton V, Zuker M, Steel M, Pointon R, Penny D (2000) Metrics on RNA secondary structures. *J Comput Biol* 7(1-2):277–292.
11. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9(4):286–298.
12. Murrell B, et al. (2013) FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol* 30(5):1196–1205.
13. Pond SL, Frost SD, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.
14. Muhire BM, et al. (2014) Evidence of pervasive biologically functional secondary structures within the genomes of eukaryotic single-stranded DNA viruses. *J Virol* 88(4):1972–1989.
15. Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P (2002) Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8(6):824–841.
16. Davis M, Sagan SM, Pezacki JP, Evans DJ, Simmonds P (2008) Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. *J Virol* 82(23):11824–11836.
17. Shimakami T, et al. (2011) Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus. *Gastroenterology* 140(2):667–675.
18. Ma Y, Yates J, Liang Y, Lemon SM, Yi M (2008) NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly. *J Virol* 82(15):7624–7639.
19. Yi M, Villanueva RA, Thomas DL, Wakita T, Lemon SM (2006) Production of infectious genotype 1a hepatitis C virus (Hutchinson strain) in cultured human hepatoma cells. *Proc Natl Acad Sci USA* 103(7):2310–2315.

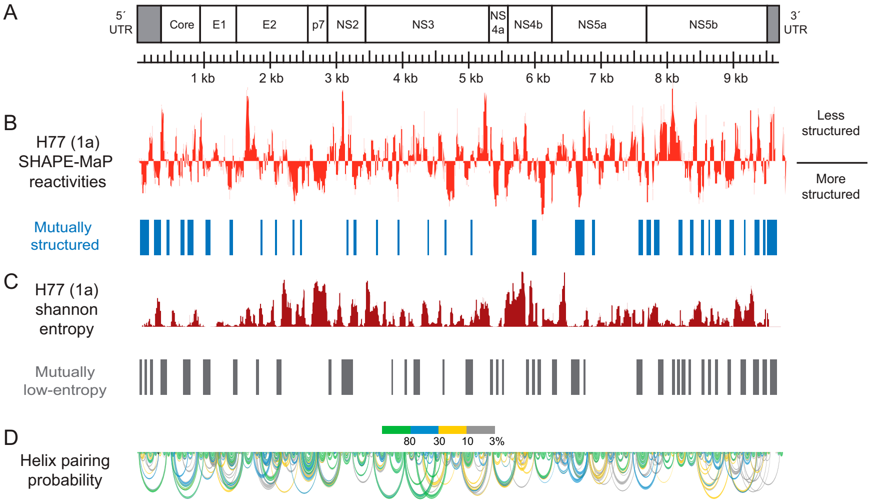


Fig. S1. Structural analysis of the HCV-H77 (1a) genome. (*A*) Diagram of the HCV-H77 genome. (*B*) Normalized median SHAPE reactivities (over 51-nt windows). The global median of the entire population is defined as 0.0. Valleys (low median SHAPE reactivity) represent highly structured regions, and peaks represent conformationally flexible regions. Regions identified as mutually structured across all three genomes are indicated with blue bars. (*C*) Median Shannon entropies (over 51-nt windows). Regions with low Shannon entropies are those likely to adopt single, well-defined conformations. Regions with low Shannon entropies across all three genomes are emphasized with gray bars. (*D*) Structural model for the entire H77 RNA genome. Helices are shown as arcs, colored according to base-pairing probabilities as calculated from the SHAPE-directed partition function (6). Helices shown with green arcs; regions with many green arcs have well-defined structures. Regions with overlapping blue, yellow, and gray arcs likely sample multiple conformations.

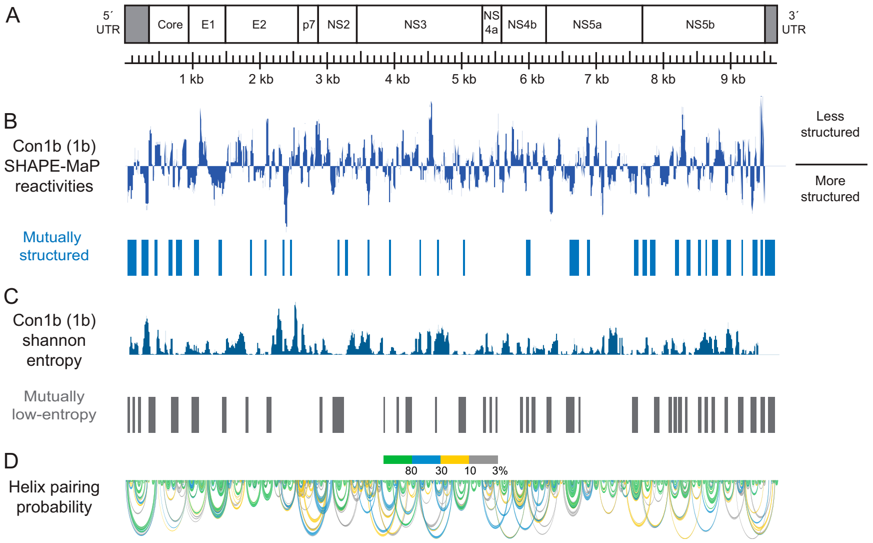


Fig. S2. Structural analysis of the HCV-Con1b (1b) Genome. Legend is given in Fig. S1.

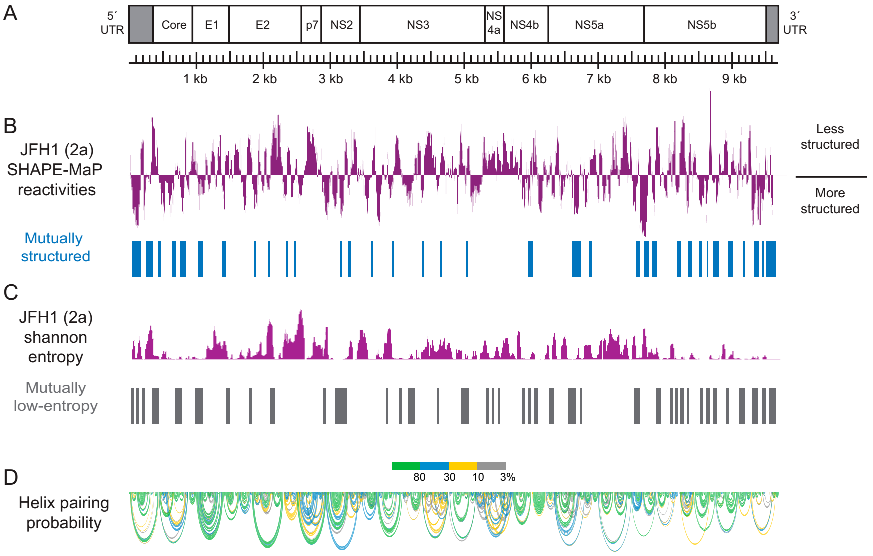


Fig. S3. Structural analysis of the HCV-JFH1 (2a) Genome. Legend is given in Fig. S1.

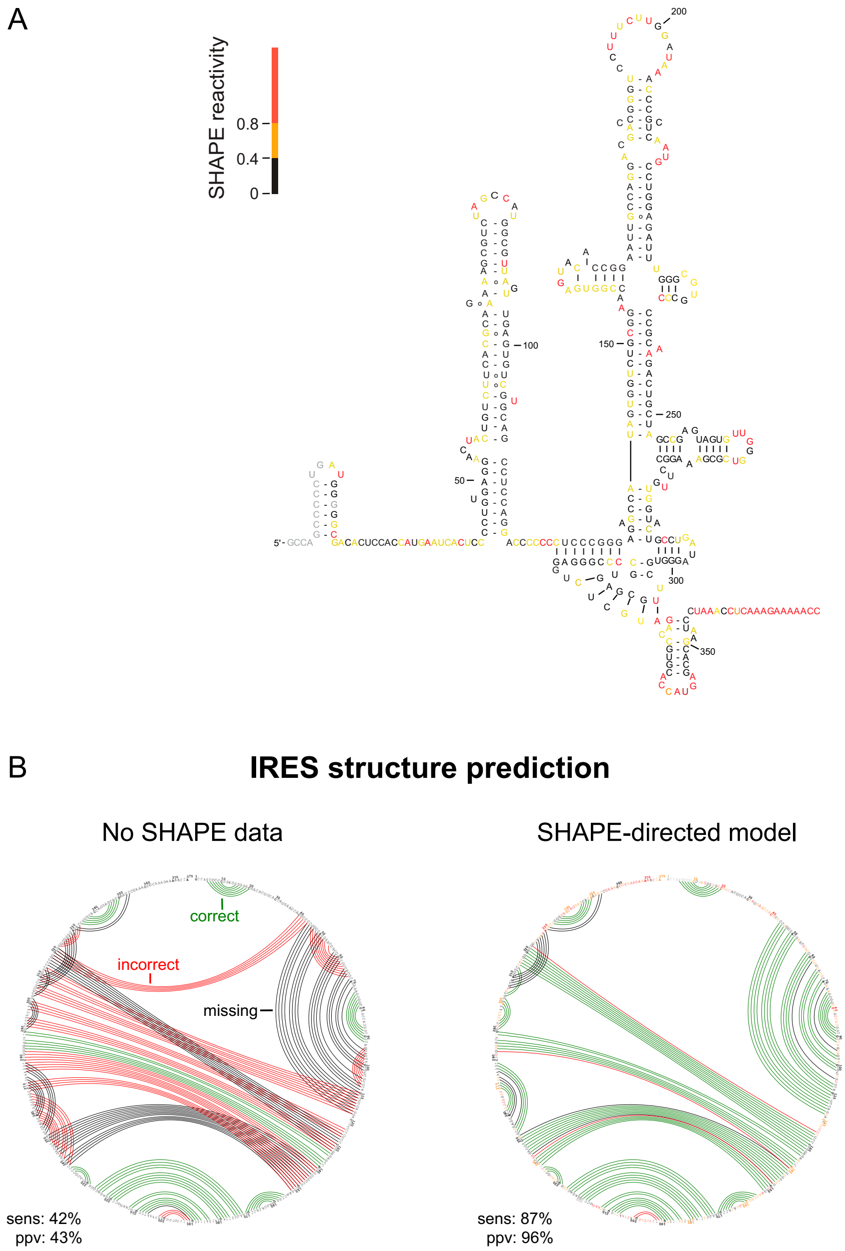
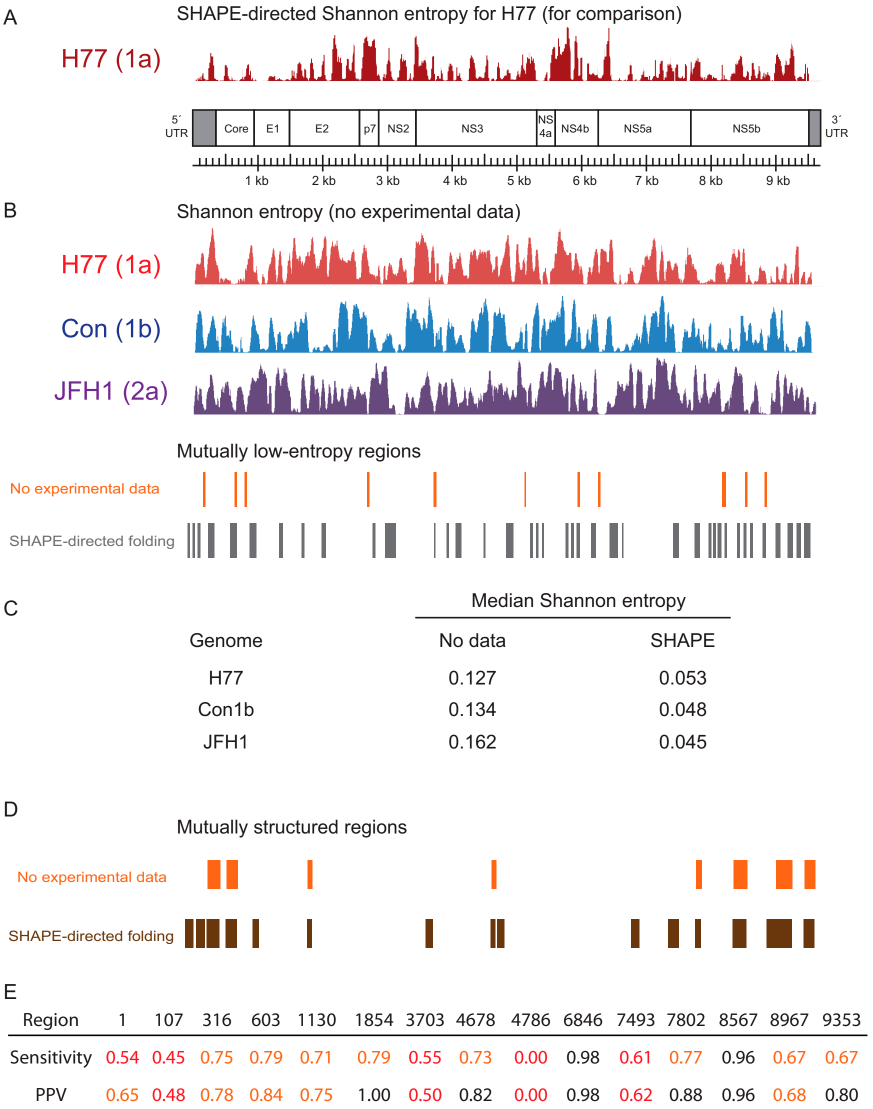


Fig. S4. Structural analysis of the HCV IRES. (*A*) Accepted RNA secondary structure for the H77 IRES region colored by SHAPE reactivities (see scale). (*B*) Base-pairing predictions for the IRES are represented as colored lines indicating correctly predicted (green lines) and incorrectly predicted (red and black lines) base pairs relative to the accepted structure (1). Predictions without structural data (*Left*) and using SHAPE reactivities (*Right*) are shown. Sensitivity (sens) and positive predictive value (ppv) are listed for each model. Pseudoknots were not modeled in this work.

1. Honda M, et al. (1996) Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology* 222(1):31–42.



	Median Shannon entropy	
Genome	No data	SHAPE
H77	0.127	0.053
Con1b	0.134	0.048
JFH1	0.162	0.045

Region	1	107	316	603	1130	1854	3703	4678	4786	6846	7493	7802	8567	8967	9353
Sensitivity	0.54	0.45	0.75	0.79	0.71	0.79	0.55	0.73	0.00	0.98	0.61	0.77	0.96	0.67	0.67
PPV	0.65	0.48	0.78	0.84	0.75	1.00	0.50	0.82	0.00	0.98	0.62	0.88	0.96	0.68	0.80

Fig. S5. Performance of structure-first genome-wide analysis depends on SHAPE data. (*A*) Median Shannon entropies across the H77 (1a) genome. These are the same data as presented in Fig. 1*C* and are included to illustrate the large differences in genome-wide Shannon entropies calculated without and with experimental SHAPE data, shown in *B*. (*B*) Median Shannon entropies (51-nt windows) for each of three HCV genomes. Regions with low Shannon entropies in all three genomes are indicated below the histograms with orange bars. For comparison, regions with low Shannon entropies calculated using SHAPE data (reproduced from Fig. 1*C*) are shown with gray bars. (*C*) Table showing the global median Shannon entropies for three HCV genomes without and with SHAPE data. (*D*) Locations of eight regions (orange boxes) that have conserved base pairing in the no-experimental data models across all three HCV genotypes. These are compared with the 15 regions (brown boxes) of conserved base pairing identified in models generated using experimental SHAPE data. (*E*) Sensitivities (sens) and positive predictive values (ppv) for the comparison of the SHAPE-directed model and the no data model for the H77 genome within the 15 regions of conserved base pairing identified by the SHAPE-directed structure-first approach. Values 0.85–0.65 and <0.65 are emphasized in orange and red, respectively. Critically, even when the no-experimental data models identified mutually structured regions that overlapped with those predicted using SHAPE information, the underlying secondary structure models often differed substantially.

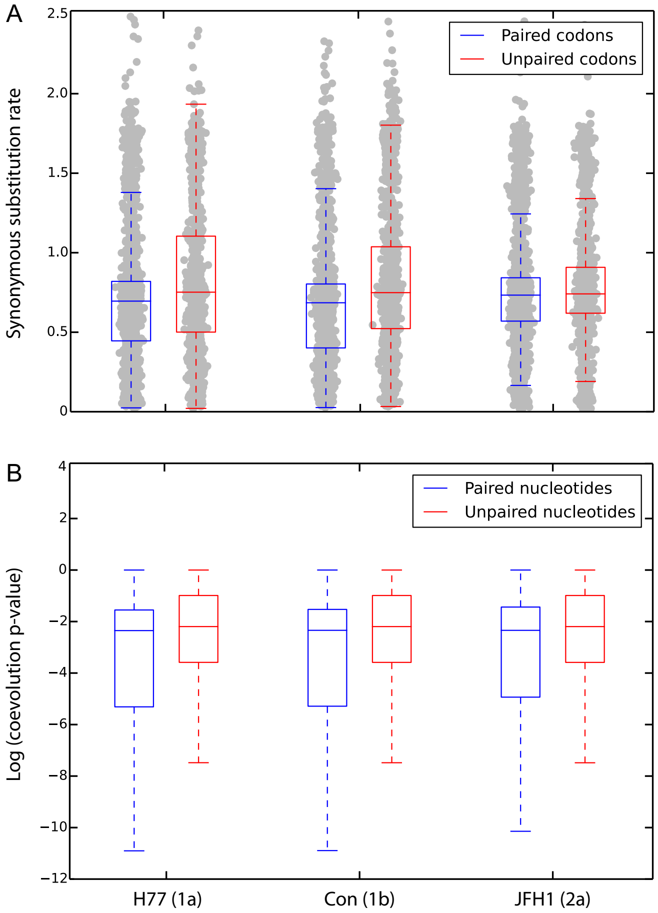


Fig. S6. Genome-wide distributions of synonymous substitution rates and degrees of coevolution across the HCV coding region identified in the SHAPE-directed minimum free-energy analyses. (*A*) Synonymous codon substitution rates. Base-paired codons were defined as those with two or three base-paired nucleotides, and unpaired codons were defined as containing 0 or 1 base-paired nucleotides. Horizontal lines represent median synonymous substitution rates; the bounds of the box and whiskers reflect the 50% and 95% ranges of these values, respectively. For H77, Con, and JFH1, the median synonymous substitution rates of the paired codons (respectively 0.696, 0.685, and 0.733) were significantly lower (respective Mann–Whitney *U* test *P* values are 1.046×10^{-8}, 4.270×10^{-14}, and 3.552×10^{-2}) than those of the unpaired codons (respectively 0.752, 0.749, and 0.741). (*B*) Degrees of coevolution at base-paired versus unpaired nucleotides. Median log(likelihood ratio test *P* values) for each site-pair across paired nucleotides in H77, Con, and JFH1 (respectively 0.095, 0.096, and 0.096) are significantly lower (respective Mann–Whitney *U* test, *P* values are 3.45×10^{-43}, 7.43×10^{-35}, and 2.02×10^{-30}) than those of the unpaired nucleotides (respectively 0.111, 0.111, and 0.111). Critically, despite their high degrees of statistical significance, the ranges of the synonymous substitution rates (*A*) and distributions of these *P* values (*B*) broadly overlap between the base-paired and unpaired sites, emphasizing the need for additional structure-based criteria to accurately identify evolutionarily conserved structural motifs.

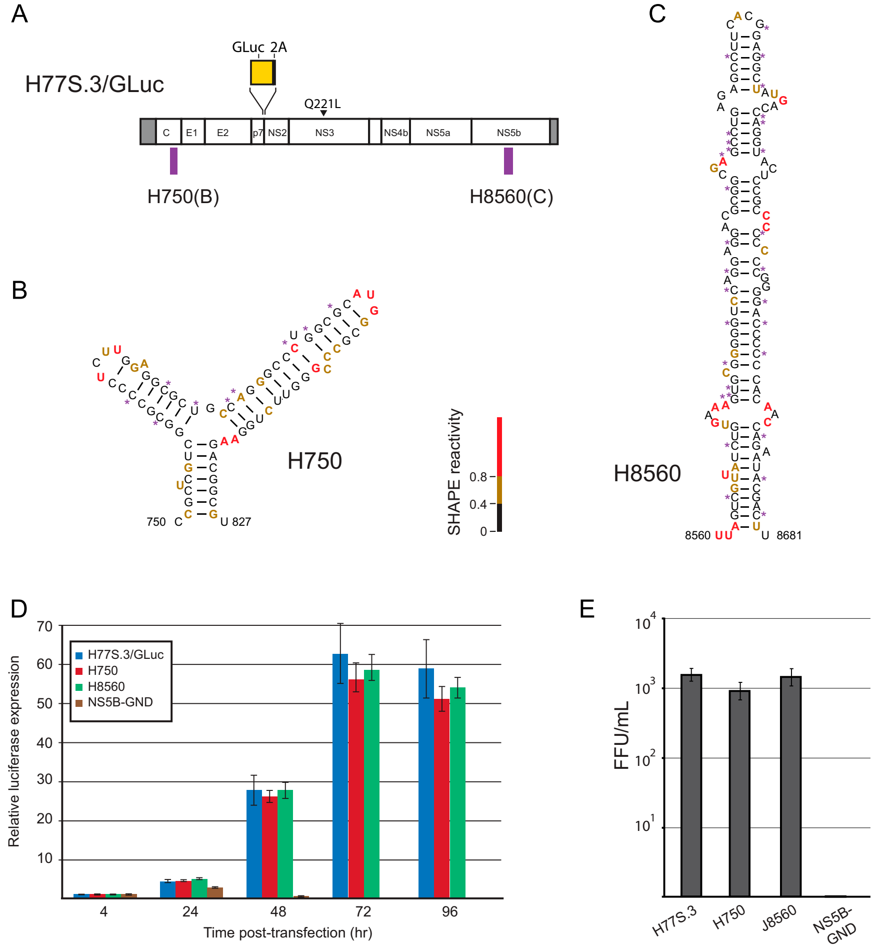


Fig. S7. Analysis of regulatory structures within H77. (*A*) H77S.3 /GLuc2A expression construct. The two elements, identified by SHAPE-directed genome modeling and tested in the GLuc assay, are shown. RNA secondary structure models for the (*B*) H750 and (*C*) H8560 elements. Nucleotides are colored by SHAPE reactivities (see scale). For each structure, the positions of engineered, structure-disrupting, silent mutations are shown with asterisks. (*D*) Relative levels of HCV-encoded luciferase protein, normalized to the 4-h time point, expressed from H77S.3/GLuc2A compared with structure disrupting mutants. The bars show the average of triplicate measurements; error bars indicate SDs. NS5B-GND is a lethal (negative) control. (*E*) Titers of infectious virus generated 72-h post-transfection of H77S.3 RNA, the lethal mutant (NS5B-GND), and the two structure-disrupted mutants. Histograms show the average of triplicate measurements; error bars indicate SDs.

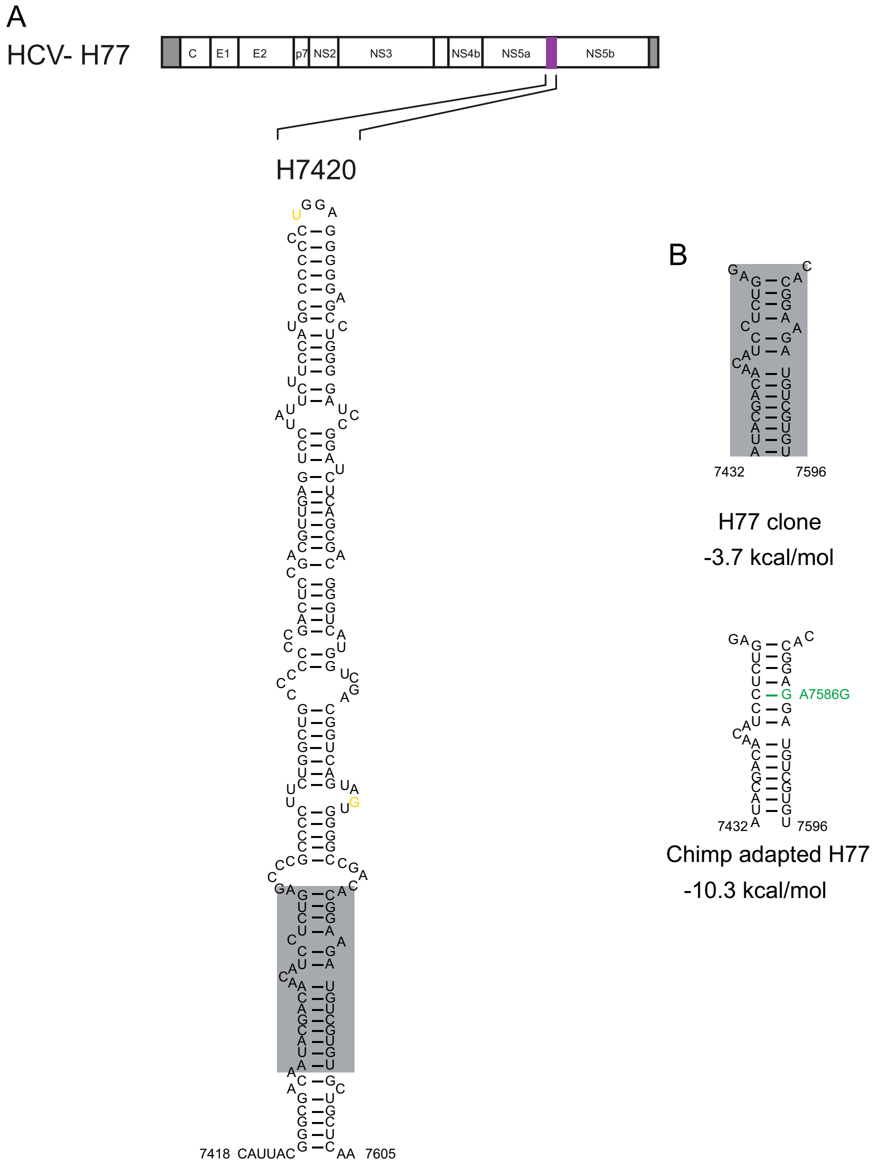


Fig. S8. Structure and chimp-selected mutation in the H7420 element in H77. (*A*) Genome location and secondary structure model of element H7420 within the HCV-H77 genome. The model was defined by SHAPE-directed probing. (*B*) The H7420 element and likely stable helix adopted by the chimpanzee-adapted H77 virus with the A7586G mutation (1).

1. Yi M, et al. (2014) Evolution of a cell culture-derived genotype 1a hepatitis C virus (H77S.2) during persistent infection with chronic hepatitis in a chimpanzee. *J Virol* 88(7):3678–3694.

Table S1. Conserved regions

Identified region	Elements within region	H77 position	Con1b position	JFH1 position	Previous biochemical data	Previous structural model	Model agreement: Previous model vs. SHAPE	Publications
1	IRES Domains I and II	1–98	1–98	1–97	Yes	Yes	Yes	Tuskiyama-Kohara et al. (1) and Brown et al. (2)
107	IRES Domain III	108–270	108–270	107–269	Yes	Yes	Yes	Tuskiyama-Kohara et al. (1) and Brown et al. (2)
316	IRES Domains IV and V	314–469	314–469	313–468	Yes	Yes	Yes	Tuplin et al. (3)
603	Structure 700	601–838	601–838	600–837	Partial	Yes	Partially	Tuplin et al. (3)
1130		1130–1206	1130–1206	1129–1205	No	No		
1854		1854–1951	1854–1951	1858–1955	No	No		
3703		3703–3785	3703–3785	3712–3794	No	No		
4678		4678–4757	4678–4757	4687–4766	No	No		
4786		4786–4861	4786–4861	4795–4870	No	No		
6848		6846–6921	6846–6921	6855–6930	No	No		
7493	Structure 7500	7493–7641	7493–7638	7502–7704	No	No		
7802	Structure 7900	7802–7954	7799–7951	7865–8017	No	Yes	Partially	Tuplin et al. (3), Chu et al. (4), and You et al. (5)
8567	Structure 8600	8567–8719	8564–8716	8630–8782	No	Yes	Partially	Tuplin et al. (3), Chu et al. (4), and You et al. (5)
8967	SL 9098, 9198, 9283	8967–9299	8964–9296	9030–9362	Yes	Yes	Yes	Tuplin et al. (3), Chu et al. (4), You et al. (5), and Diviney et al. (6)
9353	SL 9332, 9389	9353–9499	9350–9605	9416–9676	Yes	Yes	Yes	Tuplin et al. (3), Chu et al. (4), You et al. (5), and Diviney et al. (6)

1. Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A (1992) Internal ribosome entry site within hepatitis C virus RNA. *J Virol* 66(3):1476–1483.
2. Brown EA, Zhang H, Ping LH, Lemon SM (1992) Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* 20(19):5041–5045.
3. Tuplin A, Evans DJ, Simmonds P (2004) Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* 85(Pt 10):3037–3047.
4. Chu D, et al. (2013) Systematic analysis of enhancer and critical cis-acting RNA elements in the protein-encoding region of the hepatitis C virus genome. *J Virol* 87(10):5678–5696.
5. You S, Stump DD, Branch AD, Rice CM (2004) A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J Virol* 78(3):1352–1366.
6. Diviney S, et al. (2008) A hepatitis C virus cis-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B. *J Virol* 82(18):9008–9022.

Table S2. Structure disrupting mutant inserts

Mutant element	Genome region	DNA start	DNA end	DNA sequence
JFH1 constructs				
J750	Core	741	833	TGGGGTACATACCGGTGGTAGGGGCGCCGCTTAGTGGGGCGGCGCGAGCAGTGGCCCACGGCGTGAGAGTCCTGGAGGATGGCGTTAATTATG
J7490	NS5A	7495	7646	CAGGTTCCGCGAGTAGTATGCCCCCGCTCGAGGGCGAACCTGGTGATCCTGACCTGGAAAGTGATCAAGTAGAACTTCAACCGCCGCCGCAGGGCGGCGGCGTTGCACCCGGTTCTGGCTCTGGCAGTTGGAGTACGTGCAGCGAGGAGGAC
J7880	NS5B	7764	8459	CATAACAAGGTGTACTGTACAACATCAAAGAGCGCCTCACAGAGGGCTAAAAAGGTAACTTTTGACAGGACGCAAGTGCTCGACGCCCATTATGACTCAGTCTTAAAGGACATCAAACTAGCAGCATCTAAAGTATCCGCTAGGTTGTTGACCTTGGAGGAGGCGTGTCAATTGACACCACCACATTCTGCAAGAAGTAAGTATGGATTCGGAGCCAAGGAGGTAAGATCCCTGAGTGGACGAGCTGTTAATCATATTAAATCCGTCTGGAAGGACCTCTTGGAAGACCCACAAACACCAATTCCCACAACCATCATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCGCCAAGGGGGGTAAGAAACCAGCTCGCCTCATCGTTTACCCTGACCTCGGCGTCCGGGTCTGCGAGAAAATGGCCCTCTATGACATTACACAAAAGCTTCCTCAGGCGGTAATGGGAGCTTCCTATGGCTTCCAGTACTCCCCTGCCCAACGGGTGGAGTATCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCCATGGGTTTTTCGTATGATACCCGATGCTTCGACTCAACCGTCACTGAGAGAGACATCAGGACCGAGGAGTCCATATACCAGGCCTGCTCCCTGCCCGAGGAGGCCCGCACTGCCATACACTCGCTGACTGAGAGACTTTACGTAGGAGGGCCC
J8200	NS5B	8004	8809	AGGGCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACCCACAAACACCAATTCCCACAACCATCATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCGCCAAGGGGGGTAAGAAACCAGCTCGCCTCATCGTTTACCCTGACCTCGGCGTCCGGGTCTGCGAGAAAATGGCCCTCTATGACATTACACAAAAGCTACCACAAGCGGTAATGGGCGCTTCCTATGGCTTCCAATACTCCCCAGCCCAAAGAGTCGAATACCTCTTGAAAGCCTGGGCCGAAAAGAAAGATCCCATGGGATTCTCGTATGATACCAGATGTTTCGACTCAACCGTGACAGAGAGAGACATCAGGACCGAGGAATCCATATACCAAGCCTGTTCTTTACCCGAAGAAGCCCGCACTGCCATACACTCGCTGACTGAGAGACTTTACGTAGGAGGGCCCATGTTCAACAGCAAGGGTCAAACCTGCGGTTACAGACGTTGCCGCGCCAGCGGGGTGCTAACCACTAGCATGGGTAACACCATCACATGCTATGTGAAAGCCCTAGCGGCCTGCAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGATGACCTAGTAGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCCTTCACGGAGGCCATGACCAGGTACTCTGCCCCTCCTGGTGATCCCCCCAGACCGGAATATGACCTGGAGCTAATAACATCCTGTTCCTCAAATGTGTCTGTGGCGTTGGGCCCGCGGGGCCGCCGCAGATA
J8640	NS5B	8635	8735	CAGAAAGCCAAGGCACTGAAGAAGACGAACGGAACCTCAGAGCGTTCACGGAGGCCATGACCAGATACAGTGCCCCACCAGGTGACCCGCCCAGACCGGA
J8877	NS5B	8784	9345	TTGGGCCCGCGGGGCCGCCGCAGATACTACCTGACCAGAGACCCAACCACTCCACTCGCCCGGGCTGCCTGGGAAACAGTTAGACACTCCCCTATCAACAGTTGGTTAGGCAACATAATCCAATACGCACCAACAATTTGGGTTCGCATGGTGCTAATGACACACTTCTTCTCCATTCTAATGGTGCAAGATACTCTTGATCAAAACCTGAACTTCGAAATGTACGGATCCGTATACTCCGTGAACCCTCTCGACCTGCCCGCAATAATAGAGAGGTTACACGGGCTTGACGCCTTTTCTATGCACACATACTCTCACCACGAACTGACGCGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGGGTGTGGAAGAGTCGGGCTCGCGCAGTCAGGGCGTCCCTCATCTCCCGTGGAGGGAAAGCGGCCGTTTGCGGCCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCATTGCCGGAGGCGCGCCTACTGGACTTATCCAGTTGGTTCACCGTCGGCGCCGGCGGGGGCGACATT
H77 constructs				
H750	Core	745	832	GGTACATACCCCTGGTGGGGGCGCCTCTTGGAGGGGCGGCGCGGGCGCTCGCCCATGGCGTCCGGGTGCTGGAAGATGGCGTGAACT
H7430	NS5A	7418	7617	GCATTACGGGGGACAATACCACAACAAGCAGTGAGCCCGCGCCTTCCGGGTGCCCGCCCGACAGCGACGTAGAAAGCTATAGTAGCATGCCGCCCCTGGAAGGGGAACCAGGGGATCCCGATCTGAGCGACGGATCATGGTCGACTGTGTCTAGTGGCGCCGACACCGAAGACGTCGTCTGCTGCAGAATGTCTTATTC

Table S2. Cont.

Mutant element	Genome region	DNA start	DNA end	DNA sequence
H8560	NS5B	7906	9167	GGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATAGACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCCCTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCGATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGCGGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTAGTGGTTATATGTGAATCAGCAGGAGTGCAAGAAGACGCAGCGTCATTGAGAGCATTCACAGAAGCAATGACACGCTACTCCGCCCCACCAGGAGATCCACCACAACCTGAATATGATTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGAGCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATAGCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATAGAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTCCCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGAGGAGGCAGGGCTGCCATATGTGG

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (CT)
Dataset S3 (FASTA)