



Article

# Dissimilar Conservation Pattern in Hepatitis C Virus Mutant Spectra, Consensus Sequences, and Data Banks

Carlos García-Crespo <sup>1</sup>, María Eugenia Soria <sup>1,2</sup>, Isabel Gallego <sup>1,3</sup>, Ana Isabel de Ávila <sup>1</sup>, Brenda Martínez-González <sup>1,2</sup>, Lucía Vázquez-Sirvent <sup>1</sup>, Jordi Gómez <sup>3,4</sup>, Carlos Briones <sup>3,5</sup>, Josep Gregori <sup>3,6,7</sup>, Josep Quer <sup>3,6</sup>, Celia Perales <sup>1,2,3,\*</sup> and Esteban Domingo <sup>1,3,\*</sup>

<sup>1</sup> Department of Interactions with the environment, Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Campus de Cantoblanco, 28049 Madrid, Spain; carlos.garciac@cbm.csic.es (C.G.-C.); mesoria@cbm.csic.es (M.E.S.); igallego@cbm.csic.es (I.G.); aideavila@cbm.csic.es (A.I.d.Á.); brenda.vara65@gmail.com (B.M.-G.); luvazque@ucm.es (L.V.-S.)

<sup>2</sup> Department of Clinical Microbiology, IIS-Fundación Jiménez Díaz, UAM. Av. Reyes Católicos 2, 28040 Madrid, Spain

<sup>3</sup> Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) del Instituto de Salud Carlos III, 28029 Madrid, Spain; jgomez@ipb.csic.es (J.G.); cbriones@cab.inta-csic.es (C.B.); josep.gregori@vhir.org (J.G.); josep.quer@vhir.org (J.Q.)

<sup>4</sup> Department of Molecular Biology, Instituto de Parasitología y Biomedicina ‘López-Neyra’ (CSIC), Parque Tecnológico Ciencias de la Salud, Armilla, 18016 Granada, Spain

<sup>5</sup> Department of Molecular Evolution, Centro de Astrobiología (CAB, CSIC-INTA), Torrejón de Ardoz, 28850 Madrid, Spain

<sup>6</sup> Liver Unit, Liver Diseases—Viral Hepatitis, Vall d’Hebron Institut de Recerca (VHIR), Vall d’Hebron Hospital Universitari, Vall d’Hebron Barcelona Hospital Campus, Passeig Vall d’Hebron 119-129, 08035 Barcelona, Spain

<sup>7</sup> Roche Diagnostics, S.L., Sant Cugat del Vallés, 08174 Barcelona, Spain

\* Correspondence: cperales@cbm.csic.es or celia.perales@quironsalud.es (C.P.); edomingo@cbm.csic.es (E.D.)

Received: 8 September 2020; Accepted: 20 October 2020; Published: 27 October 2020



**Abstract:** The influence of quasispecies dynamics on long-term virus diversification in nature is a largely unexplored question. Specifically, whether intra-host nucleotide and amino acid variation in quasispecies fit the variation observed in consensus sequences or data bank alignments is unknown. Genome conservation and dynamics simulations are used for the computational design of universal vaccines, therapeutic antibodies and pan-genomic antiviral agents. The expectation is that selection of escape mutants will be limited when mutations at conserved residues are required. This strategy assumes long-term (epidemiologically relevant) conservation but, critically, does not consider short-term (quasispecies-dictated) residue conservation. We calculated mutant frequencies of individual loci from mutant spectra of hepatitis C virus (HCV) populations passaged in cell culture and from infected patients. Nucleotide or amino acid conservation in consensus sequences of the same populations, or in the Los Alamos HCV data bank did not match residue conservation in mutant spectra. The results relativize the concept of sequence conservation in viral genetics and suggest that residue invariance in data banks is an insufficient basis for the design of universal viral ligands for clinical purposes. Our calculations suggest relaxed mutational restrictions during quasispecies dynamics, which may contribute to higher calculated short-term than long-term viral evolutionary rates.

**Keywords:** antiviral intervention; consensus sequence; mutant spectrum; residue conservation; viral ligands; virus data banks

## 1. Introduction

The hepatitis C virus (HCV) is an important human pathogen that poses many global public health challenges, including the lack of a vaccine, and non-universal accessibility of effective treatments [1]. We are interested in its quasispecies dynamics since it has proven pertinent to the understanding of HCV-associated liver disease progression, and treatment efficacy [2–4], and its possible influence on virus evolution at the epidemiological level is an open question [5–7]. In particular, it is not known if long-term residue conservation as calculated from viral genome alignments of consensus sequences in data banks corresponds to the limited variation of the same residues in mutant spectra. A current aim in the control of diseases associated with RNA viruses is the computer- and dynamic simulations-guided design of universal vaccines, therapeutic antibodies or pan-genomic antiviral agents based on conserved viral residues, with conservation defined according to the alignment of nucleotide and amino acid sequences recorded in data banks. The expectation is to render reagents effective against all (or a majority of) circulating serotypes and genotypes of a virus [8–13]. This strategy rests on the assumption that viral genomic residues that are conserved among independent viral isolates tend to be also conserved within the mutant spectra displayed by the viruses during their intra-host multiplication. Yet, it will be during such stage of the infection cycle when the binding ligands should prevent viral progeny production, or when the newly assembled and released viral particles and newly infected cells should be controlled by a vaccine-induced immune response. The critical assumption of parallel residue conservation in mutant spectra and data banks has not been tested. This point is relevant because it interrogates a relationship between two ranks of operation of virus evolution: one which is associated with intra-host, short-term mutant cloud dynamics, and another that involves long-term evolution at the epidemiological level, generally evaluated with consensus sequences.

Here, we describe a two-step approach to this question with HCV. First, we compare the degree of individual residue conservation (percentage of residue identity) in the course of viral quasispecies dynamics—quantified by the mutant frequency of each variable residue in cell culture and infected patients—with the conservation deduced from an alignment of the consensus sequences of the same populations. Second, we compare conservation levels in mutant spectra with those of the same residues in the Los Alamos National Laboratory (LANL) data bank sequence alignment. Mutant frequencies in quasispecies have been obtained in cell culture thanks to the availability of infectious molecular clones of HCV that permit experimental evolution designs with this important pathogen [14–18]. In a recent study, we examined mutant spectrum variations upon subjecting a clonal HCV population to 200 serial infections of Huh-7.5 cells, using fresh cells at each passage, thereby preventing the co-evolution of the host cells [19]. Under these conditions, a variation of mutant frequency of several individual genomic residues even between successive passages was consistently observed in three biological replicas. Because mutant frequency variations were continuous, we termed them “mutational waves” to emphasize a difference with those residues whose frequency remained constant with passage number [17,19]. Mutational waves were characterized by deep sequencing analysis of 1005 genomic positions from the NS5A-NS5B-coding region, and they involved 145 different mutations. The dependability of the deep sequencing analyses was sustained on experimental and bioinformatics controls that yielded reliable cut-off mutant frequency values of 0.5% [19–21] (see also Section 2). Furthermore, Sanger sequencing of the HCV p0, HCV p100, HCV p200 and their derivatives at passage 4 identified 114 heterogeneous sites (two different nucleotides at the same genomic position) within the NS2- to NS5B-coding region, thus confirming mutant frequency variations by two independent procedures [19].

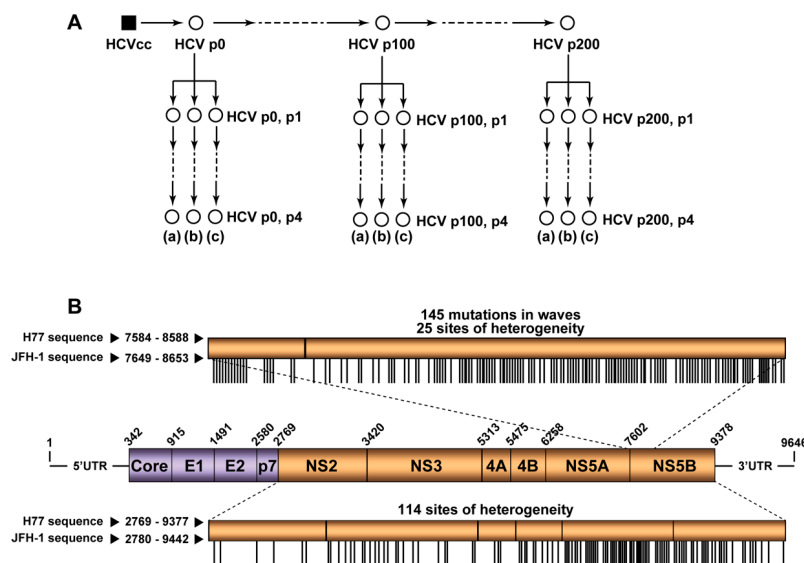
In addition, we included in our study 522 different amino acid substitutions in the NS5B mutant spectra of a cohort of 220 HCV-infected patients that failed antiviral therapy [22]. The regions analyzed include the entire non-structural protein-coding region and deduced amino acids, clinically important for being the target of direct-acting antiviral agents, and sites where multiple mutations accumulate during serial passage of the virus in human hepatoma cells [17–19]. This information on HCV quasispecies composition in cell culture and in vivo has documented that residue variation

at the quasispecies level does not match conservation in the corresponding consensus sequences or in the sequences deposited in the LANL data bank. The computations suggest a remarkably higher mutant tolerance in evolving HCV quasispecies than reflected in consensus sequences. This difference may contribute also to a higher calculated intra-host than inter-host evolutionary rates, an intriguing difference observed with several viral pathogens [6,7,23–27]. A new residue conservation score can be produced upon analysis of sequence alignments of the genomes in mutant spectra that may be more appropriate for the design of universal ligands or vaccines intended to control RNA viruses.

## 2. Materials and Methods

### 2.1. Experimental Procedures for HCV in Cell Culture

The starting HCVcc population was obtained by transcription of plasmid Jc1FLAG2(p7-nsGluc2A) (abbreviated Jc1Luc) [28], and electroporation of the RNA into Huh-Lunet cells. Progeny virus was concentrated by filtration, and used to infect Huh-7.5 reporter cell monolayers at a multiplicity of infection (MOI) of about 0.5 tissue culture infectious doses 50 (TCID<sub>50</sub>) per cell; 3 days later the cells were subcultured and maintained a further 3 days, followed by a second round of infection and two additional cell passages. The virus in the cell culture supernatant (with aliquots stored at –80 °C until needed) was the initial population HCV p0, utilized to initiate the serial infections, as depicted in Figure 1; its titer was  $1.63 \times 10^5 \pm 1.03 \times 10^5$  TCID<sub>50</sub> per mL.



**Figure 1.** Experimental design, and hepatitis C virus (HCV) genome analysis. **(A)** Clonal HCV population HCVcc prepared by transcription of plasmid Jc1Luc [28], followed by RNA electroporation into Huh-7 Lunet cells, was amplified to produce HCV p0, and the population subjected to 200 passages in Huh-7.5 cells, as described [17–19]. The initial population HCV p0 and the populations at passages 100 (HCV p100) and 200 (HCV p200) were passaged four additional times in Huh-7.5 cells in triplicate (replicas (a), (b), (c)). The mutant spectrum of each of the populations (represented by empty circles) was analyzed. **(B)** Scheme of the HCV genome, encoded proteins and genomic regions analyzed; residue numbering is according to isolate H77. In the boxes depicted above and below the genome, the equivalence of nucleotide residue numbers for isolates H77 and JFH-1 is indicated. The upper box corresponding to the NS5A-NS5B-coding region indicates the positions involved in mutational waves within residues 7584 to 8588 (H77 numbering) (vertical lines). Frequency levels are described in Section 2 and indicated in Table S1 (Supplementary Material at the end of the manuscript). The bottom box corresponding to the NS2-NS5B-coding region indicates the sites of heterogeneity determined by Sanger sequencing (see also Section 2). Results are from [19], and the mutations under study are compiled in Tables S1 and S3.

For titration of infectivity, serial dilutions of cell culture supernatant were used to infect Huh-7.5 cells in 96-well plates, seeded with 6400 cells per well 16 h earlier; at 72 h post-infection the cells were washed with PBS, fixed with cold methanol, and stained to detect NS5A with monoclonal antibody 9E10 [16,18].

Huh-7 Lunet, Huh-7.5, and Huh-7.5 reporter cells were grown in Dulbecco's-modified Eagle's medium (DMEM) with 10% fetal calf serum, at 37 °C, in a 5% CO<sub>2</sub>, 95–97% humidity atmosphere. To initiate the serial viral infections  $4 \times 10^5$  Huh-7.5 reporter cells were infected with HCV p0 at an MOI of about 0.5 TCID<sub>50</sub> per cell; virus adsorption to the cells was allowed for 5 h under standard cell culture conditions; then the inoculum was removed, and 2 mL of cell culture medium was added, and the infection continued for 72 h to 96 h. For successive passages, 0.5 mL of the cell culture supernatant from the previous infection was used to infect  $4 \times 10^5$  Huh-7.5 reporter cells under the same conditions, with MOI ranging from 0.1 to 0.5 TCID<sub>50</sub> per cell. Mock-infected cells and cells infected with the replication-defective virus GNN [28] were used as negative infection controls. A scheme of the passages underwent by HCV p0 is depicted in Figure 1A, and further details are described in [17–19].

## 2.2. HCV Quasispecies from a Patient Cohort

HCV samples from a cohort of 220 patients assembled from 39 Spanish hospitals were analyzed by deep sequencing. The HCV genotype (G) and subtype distribution was the following: G1, 62.7%; G3a, 21.4%; G4d, 12.3%; G2, 1.8%; and 1.8% were mixed infections. The patients had failed DAA-based, interferon (IFN) $\alpha$ -free therapy. The participating hospitals adhered to current regulations for clinical research, and to PEAHC (Plan Estratégico para el Abordaje de la Hepatitis C; Strategic Plan to Approach Hepatitis C), and since samples were analyzed for diagnostic purposes, no further patient consent was required [22].

## 2.3. Experimental Data on HCV Quasispecies

The nucleotides in the HCV genome that participate in mutational waves or that belong to sites with composition heterogeneity in mutant spectra of virus evolving in cell culture (and deduced amino acid replacements) have been described [19]. For deep sequencing (Miseq platform,  $2 \times 300$  bp mode, v3 chemistry) basal sequencing errors, absence of PCR-mediated recombination, and haplotype frequency reproducibility were controlled experimentally and bioinformatically. Experimental controls involved deep sequencing of reconstructed mixtures of HCV RNAs containing different proportions of RNAs with specific mutations, comparison of results of mutant spectrum composition of the same viral sample subjected to two different amplifications and data processing cycles, and quantification of basal recombination rates upon amplification of marked RNAs [19,20,29]. Bioinformatically, controls were based on computations of binomial distributions for different coverages in the range of 500–10,000 reads [29]. The cut-off frequency of mutant detection was 0.5% [19].

HCV sequences from infected patients are those described in [22]. Amino acid substitutions belong to HCV from patients that failed direct-acting antiviral agents, and no distinction has been made among mutant frequency values in the populations analyzed. Deep sequencing procedures and bioinformatics processing of data were those applied to the HCV quasispecies in cell culture [22]. For the patient samples, the cut-off frequency of mutant detection was 1% [22,29]. Only mutations identified in the two DNA strands were included in the calculations.

Sites of nucleotide heterogeneity within the NS2-NS5B-coding region are those described in [19], without distinction of the frequency of the mutant nucleotide. Oligonucleotide primers, RT-PCR amplification conditions, and nucleotide sequencing (23 ABI 3730 XLS sequencer, Macrogen Inc., Seoul, Rep. of Korea) have been previously described [19].

## 2.4. Mutation Level and Nomenclature

Individual mutations that change in frequency (those that participate in mutational waves) have been previously divided into three levels according to the frequency that they reached in the HCV

population evolving in Huh-7.5 cells. Level L0 includes those mutations that at some passage reached a frequency above the 0.5% cut-off level and that never exceeded 1% frequency in the sample. Level L1 corresponds to mutations whose frequency was never above 10%. Level L2 is the one of mutations that at some virus passage attained a frequency higher than 10%. The level assignment of each mutation was previously described [19], but no distinction among levels has been made for the present study.

Some of the mutations identified were present in more than one population [for example, in populations from replicas (a), (b) or (c)] in cell culture which would qualify them as single nucleotide polymorphisms (SNPs) in general genetics nomenclature, while other mutations are present in only one lineage. The latter mutations deviate from the bonafide genetic polymorphism concept, and yet they contribute to the mutant swarm nature to the population. For the calculations performed in the present study, no distinction was made among mutations according to their L level or for being present in one or more lineages, and in consequence, the term SNP is not used.

In addition to mutations, heterogeneous sites refer to those positions where two nucleotides were detected in the consensus sequence determined by Sanger sequencing. The criteria to establish a position as heterogeneous is that the two coexisting nucleotides are detected upon sequencing of the two cDNA strands, and that the mutant nucleotide exceeds a detection limit value of 15% (further details in [19]). For the calculations of the present study, no distinction was made among heterogeneous sites according to the percentage of mutant nucleotide, provided its frequency was above 15%, set as the detection limit.

### 2.5. Sequences from the Los Alamos Database

The sequences were retrieved from LANL following previously described procedures [22,29]. Inclusion criteria were that the sequences had been confirmed, that they corresponded to full-length (or near-full length) genomes (without large insertions or deletions), and with no evidence of their being recombinants. Their HCV genotype/subtype distribution is: 553 sequences of genotype G1a; 427 of G1b; 3 of G1c; 33 of G2a; 81 of G2b; 8 of G2c; 5 of G2j; 4 of G2k; 49 of G3a; 17 of G4a; 5 of G4d; and 6 of G4f. They were aligned using the program BioEdit version 7.0.9.0. For the calculation of the conservation range of individual residues, no distinction has been made between HCV subtypes. As part of the controls to exclude a bias in the assignment of variable sites to the different conservation groups, calculations were made with conservation groups determined with an alignment of sequences that included equal representation of different subtypes.

### 2.6. Statistics

The statistical significance of differences in the distribution of variable sites among conservation groups was calculated with the Pearson's chi-square test using software R version 3.6.2, without or with Monte Carlo correction (based on 2000 replicates). Sample sizes are given for each comparison.

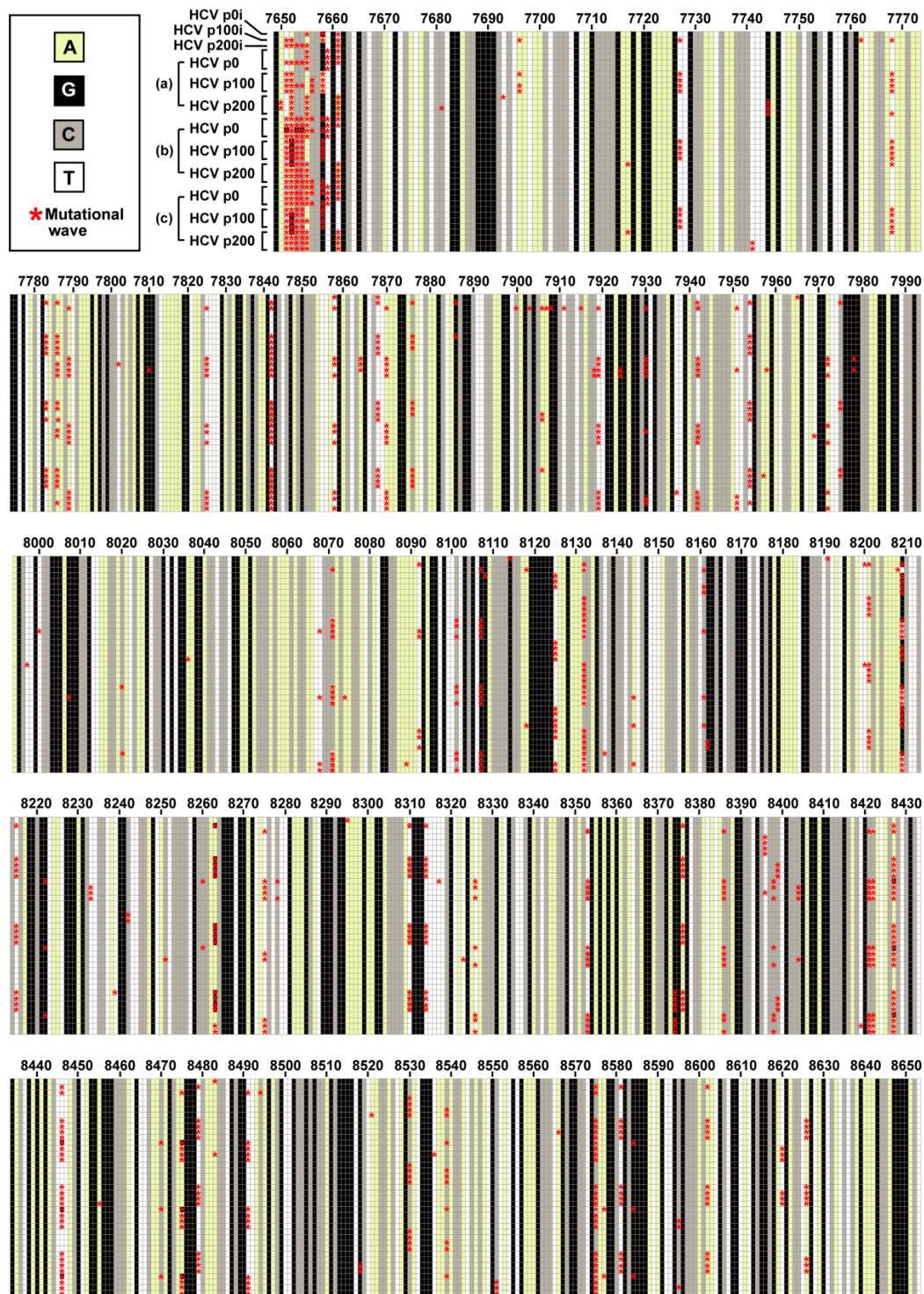
### 2.7. Sequence Accession Numbers and Data Availability

The reference accession numbers of sequences retrieved from LANL used to determine conservation groups are given in Table S2. Accession numbers for HCV samples included in the patient cohort are SAMN08741670 to SAMN08741673 [29]. Amino acid replacements in HCV from infected patients have been previously described [22], and they are compiled in Table S5. GenBank accession numbers for HCV p0, HCV p100 and HCV p200 are KC595606, KC595609 and KY123743, respectively. Illumina data can be retrieved from the NCBI BioSample database, with accession numbers SAMN13531332 to SAMN13531367 (Bio Project accession number PRJNA593382).

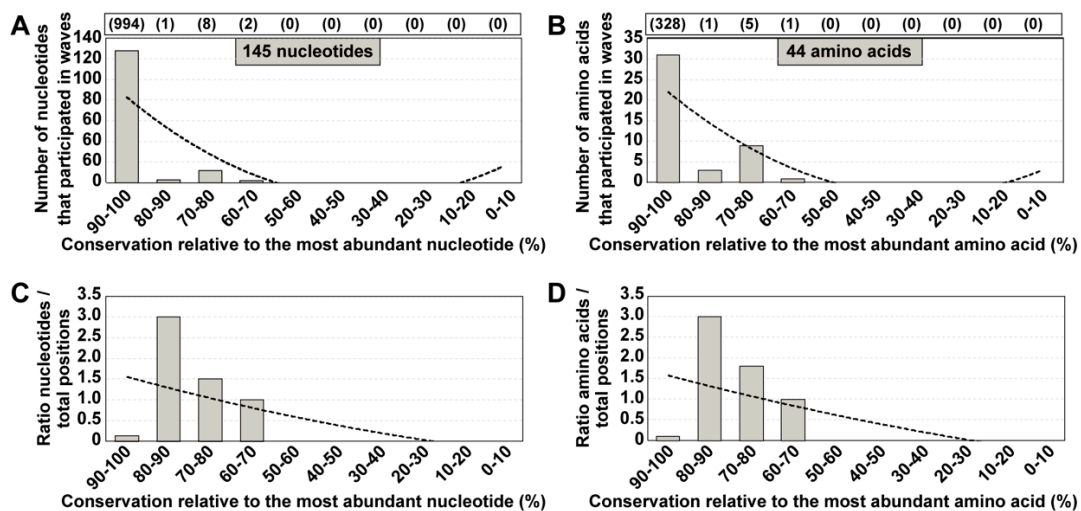
### 3. Results

#### 3.1. HCV Residues Involved in Mutational Waves and Their Conservation in the Consensus Sequences of the Populations

Two-hundred serial passages of a clonal (derived by transcription of plasmid Jc1Luc) HCV population [28] in Huh-7.5 cells identified 145 different mutations—located within genomic nucleotides 7649 to 8653—that participated in mutational waves based on deep sequencing data of 39 HCV populations [19] (Figure 1A) (residue numbers are according to HCV reference isolate JFH-1 (accession number #AB047639); mutations and deduced amino acid substitutions are given in Table S1. The alignment of the consensus nucleotide sequences for each of the 39 populations indicated a 98% sequence identity. The residues that participated in mutational waves were scattered along the region analyzed, with an accumulation within residues 7650 to 7661 (those encoding the C-terminal amino acids of NS5A and the N-terminal amino acids of NS5B) (Figure 2). A similar distribution was obtained at the amino acid level (Figure S1). To assign each variable residue in the quasispecies to a degree of conservation according to the consensus sequence alignment, the 1005 nucleotides comprised between positions 7649 and 8653 were divided into ten conservation categories calculated relative to the most abundant nucleotide at the corresponding position in the alignment. A residue that falls in the 90–100% conservation window means that the residue is present in at least 90% of the consensus sequences; this is followed by lower conservation categories, with 20–30% as the minimum possible conservation window at the nucleotide level. The 145 nucleotides (and deduced amino acids) that varied in frequency in the HCV quasispecies belonged to high conservation categories in the consensus sequence alignment (Figure 3A,B). The difference between the observed residue distributions and those in which an equal number of mutations was assigned uniformly among the possible conservation groups was statistically significant both for nucleotides and amino acids ( $p = 0.0004998$  and  $p = 0.0004998$ , respectively; Pearson's chi-squared test with Monte Carlo correction). To rectify a possible bias derived from the great abundance of residues in the 90–100% conservation window, values were normalized to the number of positions that fall in its conservation group. This correction shifted the maximum number of residues from the 90–100% to the 80–90% window (Figure 3C,D). In this case, the difference between the observed distribution and one in which the number of mutations was equally distributed among the possible conservation groups was not statistically significant ( $p = 1$  for nucleotides and amino acids; Pearson's chi-squared test with Monte Carlo correction). Conversion of complex mutant spectra into consensus sequences determines a ranking of residue conservation that does not correspond with conservation at the level of the mutant spectra that yielded the consensus sequences.



**Figure 2.** Heat map of consensus sequences of 39 HCV populations derived from HCV p0 upon passage in Huh-7.5 cells. The region analyzed by deep sequencing spans genomic residues 7649 to 8653 (residue numbering according to reference isolate JFH-1). The populations are those identified by empty circles in the HCV passage diagram of Figure 1A, and are indicated at the left of the top bloc of the alignment; (a), (b), and (c) refer to the three replicas of the four serial passages of HCV p0, HCV p100 and HCV p200. Each horizontal alignment of squares displays the consensus sequence (1005 nucleotide positions) of the population written on the left, with the nucleotide color code given in the upper box on the left. The red asterisks indicate the nucleotides that participated in mutational waves (that is, that changed in frequency among any of the populations analyzed). The complete set of mutations and deduced amino acid substitutions are given in Table S1.



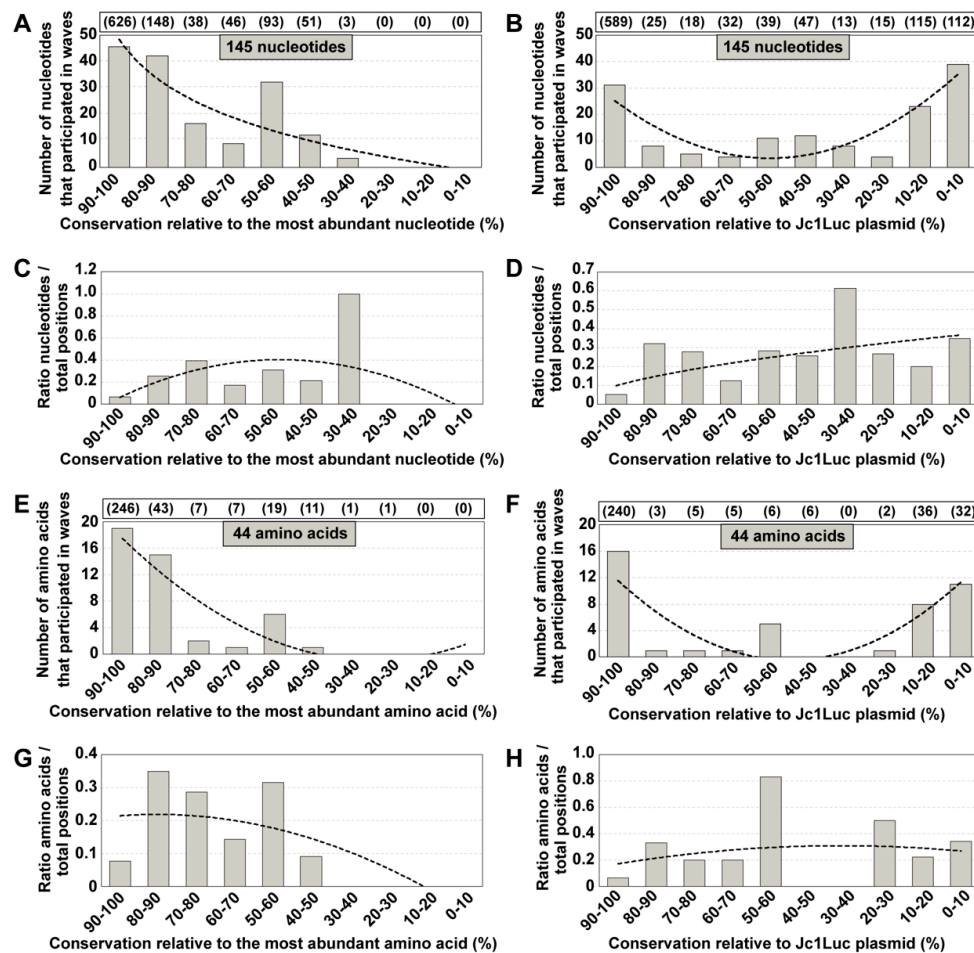
**Figure 3.** Degree of conservation of those residues that participated in mutational waves. (A) Number of nucleotides involved in mutational waves distributed among conservation groups, calculated relative to the most abundant nucleotide at the corresponding position in the alignment of 39 consensus sequences. Conservation groups are indicated in abscissa, and the number of nucleotides that participated in mutational waves in each group is given in ordinate. The total number of nucleotides within residues 7649 to 8653 from the alignment that falls in each conservation category is indicated in parenthesis in the upper box. The discontinuous line corresponds to function  $y = 2.8636x^2 - 39.009x + 118.8$  ( $R^2 = 0.6221$ ). (B) Same as A but at the amino acid level. The discontinuous line corresponds to function  $y = 0.6818x^2 - 9.6091x + 31$  ( $R^2 = 0.7135$ ). (C) Data of A normalized to the number of residues in each conservation group; normalization was performed by dividing the latter number by the total number of residues from the consensus sequence alignment that fell into the corresponding group. The discontinuous line corresponds to function  $y = 0.0086x^2 - 0.2926x + 1.8409$  ( $R^2 = 0.3595$ ). (D) Data of B normalized to the number of residues in each conservation group. The discontinuous line corresponds to function  $y = 0.0067x^2 - 0.2788x + 1.8651$  ( $R^2 = 0.3577$ ). The position of each mutation and amino acid substitution is given in Table S1.

### 3.2. Conservation in Mutant Spectra as Compared with Conservation in the Los Alamos Data Bank

Given that sites of variability in mutant spectra of the experimental populations did not match variability in consensus sequences, we explored whether the same discrepancy was observed when mutant spectra were confronted with residue conservation in the HCV sequence repository in the LANL data bank. The interest in this comparison resides in the fact that LANL alignments are often taken as a reference for residue conservation in viral genomes, and taken into consideration for the design of broad-spectrum viral ligands or vaccines [8–13]. To this aim, a total of 1191 HCV genomic sequences of LANL (<https://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html>) were used to define conservation groups, following the same procedure as for the consensus sequences of the HCV quasispecies in cell culture. Inclusion criteria and subtype distribution are given in Materials and Methods, and the accession numbers are listed in Table S2; 95.4% of the sequences are from infected patients (of which 26.1% are from antiviral-treated patients, 55.9% from untreated patients, and 17.8% without treatment information). The degree of conservation was calculated for residues 7584 to 8588 of sequence H77 (GenBank accession number AF009606, which is used as a reference in the LANL alignment [30]; residues 7584 to 8588 of H77 are equivalent to residues 7649 to 8653 of JFH-1). Conservation groups were calculated relative to the most abundant nucleotide and amino acid at the corresponding position in the LANL alignment. Then, the position of each of the 145 nucleotides (and of 44 deduced amino acids) that were involved in mutational waves was assigned to the conservation group of the same position in the alignment. The wave mutations were distributed between the 90–100% and 30–40% conservation range, with 54.5% of wave mutations falling into the



80–100% conservation groups (Figure 4A). The difference between this distribution and one in which an equal number of mutations was assigned uniformly among the possible conservation groups was statistically significant ( $p = 2.8113 \times 10^{-9}$ , chi-square test).



**Figure 4.** Degree of conservation of nucleotides and amino acids that participated in mutational waves, according to the Los Alamos National Laboratory (LANL) alignment. (A) Number of nucleotides involved in mutational waves distributed among conservation groups, calculated relative to the most abundant nucleotide at the corresponding position in the LANL alignment. Conservation groups are indicated in abscissa, and the number of nucleotides that participated in mutational waves in each group is given in ordinate. The total number of nucleotides within residues 7584 to 8588 (H77 numbering) from the alignment that falls in each conservation category is indicated in parenthesis in the upper box. The discontinuous line corresponds to function  $y = -19.21\ln(x) + 43.511$  ( $R^2 = 0.7816$ ). (B) Same as A but with nucleotide conservation in the LANL alignment calculated relative to the corresponding residues in plasmid Jc1Luc [28]. The discontinuous line corresponds to function  $y = 1.3068x^2 - 13.254x + 37.083$  ( $R^2 = 0.7401$ ). (C) Data of A normalized to the number of residues in each conservation group; normalization was performed by dividing the latter number by the total number of residues from the LANL alignment that fell into the corresponding group. The discontinuous line corresponds to function  $y = -0.0194x^2 + 0.2013x - 0.1189$  ( $R^2 = 0.2582$ ). (D) Data of B normalized to the number of residues in each conservation group. The discontinuous line corresponds to function  $y = 0.0989x^{0.5688}$  ( $R^2 = 0.3987$ ). (E–H) Same as (A–D) but at the amino acid level. The defining functions are E:  $y = 0.4205x^2 - 6.4068x + 23.45$  ( $R^2 = 0.8178$ ); F:  $y = 0.5871x^2 - 6.4826x + 17.45$  ( $R^2 = 0.6587$ ); G:  $y = -0.0045x^2 + 0.0175x + 0.2018$  ( $R^2 = 0.5213$ ); H:  $y = 0.0644\ln(x) + 0.1726$  ( $R^2 = 0.0347$ ). The position of each mutation and amino acid substitution is given in Table S1, and control calculations and simulations to assess the statistical relevance of the mutant distributions are described in Figure S4 and Table S6.

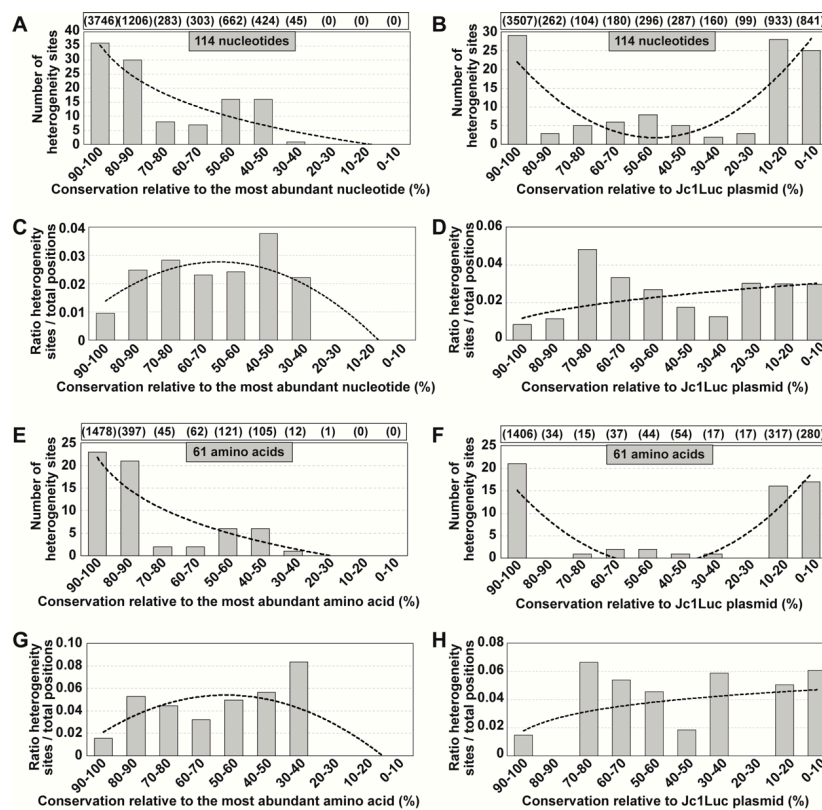
Given the absence of mutational wave residues in the low conservation categories, we wanted to exclude the possibility that the reference chosen for the calculation may have biased the mutation sites towards high conservation groups. Therefore, we also calculated the conservation ranges in the LANL alignment using as reference the corresponding residues in the HCV sequence in plasmid Jc1Luc. Since the sequence belongs to the specific HCV subtype 2a, and the LANL alignment includes HCV genomes of different subtypes, the reference modification was expected to increase the number of positions belonging to the low conservation windows. With this new reference, the plot of the number of nucleotides or amino acids in each conservation category yielded a U-shaped curve (Figure 4B). A considerable proportion (46%) of wave-involved residues belonged to the two extreme 90–100% or 0–10% conservation groups. Normalization to the total number of sites in each conservation category resulted in a broader distribution across the conservation spectrum (Figure 4C,D), with no significant difference with a uniform distribution among conservation groups ( $p > 0.999$ , chi-square test with Monte Carlo correction). The same calculations were repeated using the 33 sequences of genotype 2a from the LANL alignment. The results show the same nucleotide and amino acid distribution among conservation groups than for the entire LANL database (Figure S2). Thus, there is no correlation between sites of variation in mutational waves of HCV evolving in cell culture and sites of low residue conservation in the LANL alignment.

### 3.3. Extension of the Calculations to Mutations Associated with Sites of Heterogeneity in Consensus Sequences

An independent evaluation of individual mutant spectrum residues that vary in frequency in HCV evolving in cell culture was provided by the genomic positions that contain more than one nucleotide, as quantified in consensus sequences determined by Sanger sequencing for HCV p0, HCV p100, HCV p200, and these populations at passage 4. A total of 25 points of heterogeneity were scored within residues 7584 to 8588 (H77 numbering) of populations HCV p0, HCV p100 and HCV p200 [19] (Figure 1B and Table S3). A plot of the nucleotides at sites displaying heterogeneity versus the degree of conservation of those nucleotide and amino acid sites in the consensus sequences of the cell culture HCV populations or in the LANL alignment, confirmed the results obtained with the wave mutations (Figure S3). Again, residues that varied in mutant frequency—as determined by a procedure which is independent of deep sequencing—during quasispecies dynamics were not the most variable in consensus sequences from progeny HCV p0 populations or those reported in the LANL alignment.

### 3.4. Extension to Other HCV Genomic Regions

To exclude that the NS5A-NS5B residues analyzed might display behavior which is not representative of other HCV genomic regions, we extended the calculations to a total of the 114 genomic sites that display composition heterogeneity, located between nucleotides 2769 (beginning of the NS2-coding region) and 9377 (end of the NS5B-coding region) (H77 numbering) of populations HCV p0, HCV p100 and HCV p200 [19] (Figure 1B and Table S3). The results (Figure 5) are very similar to those obtained with the NS5A-NS5B region. The difference between the distribution of the 114 nucleotides and one in which the distribution included an equal number of sites across the different possible conservation groups, was statistically significant ( $p = 1.0371 \times 10^{-7}$ ; chi-square test). In conclusion, residue invariance (undetectable mutant frequencies by current deep sequencing and Sanger sequencing methodology) in the mutant spectra of HCV evolving in Huh-7.5 cells does not correlate with the conservation of the same residues in consensus sequences or among epidemiologically distant isolates as represented in LANL.

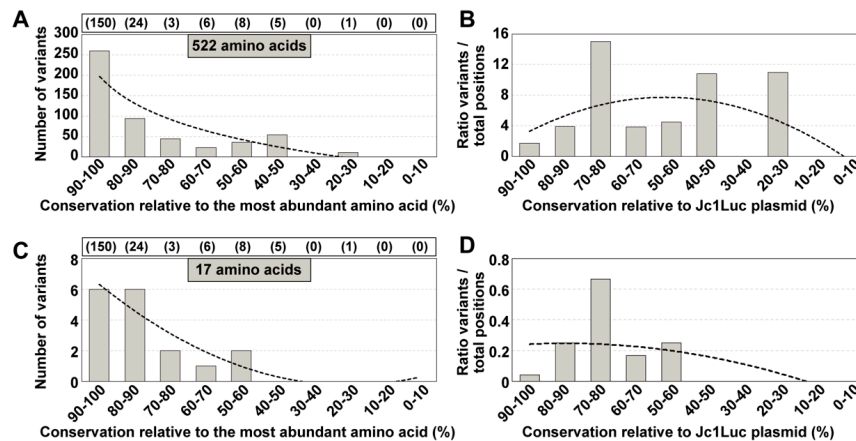


**Figure 5.** Distribution of sites of heterogeneity in HCV quasispecies among residue conservation groups defined according to LANL. Sites of heterogeneity are those identified within HCV genomic residues 2769 and 9377 (H77 numbering). In this case, the analysis was performed separately with 983 sequences of genotype G1 and 129 sequences of genotype G2 (excluding one G2c and one G2k sequence due to the presence of an insertion (<https://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html>)). (A) The genomic position at which each heterogeneity site is found is assigned to the conservation range calculated relative to the most abundant nucleotide in each position according to the LANL alignment. Conservation groups are indicated in abscissa, and the number of heterogeneity sites in each group is given in ordinate. The total number of nucleotides that fall in each conservation category is indicated in parenthesis in the upper box. The discontinuous line corresponds to function  $y = -15.83\ln(x) + 35.306$  ( $R^2 = 0.7957$ ). (B) Same as A except that the conservation groups in LANL alignment were calculated using the HCV sequence in plasmid Jc1Luc as reference. The discontinuous line corresponds to function  $y = 1.1439x^2 - 11.892x + 32.767$  ( $R^2 = 0.6507$ ). (C) Data of A normalized to the number of residues in each conservation group. The discontinuous line corresponds to function  $y = -0.0012x^2 + 0.0103x + 0.0047$  ( $R^2 = 0.7036$ ). (D) Data of B normalized to the number of residues in each conservation group. The discontinuous line corresponds to function  $y = 0.0117x^{0.4096}$  ( $R^2 = 0.2824$ ). (E–H) Same as (A–D) but at the amino acid level. In this case, the possible conservation windows covered 0% to 100%. The defining functions are E:  $y = -10.39\ln(x) + 21.797$  ( $R^2 = 0.7694$ ); F:  $y = 0.9015x^2 - 9.5106x + 23.7$  ( $R^2 = 0.7085$ ); G:  $y = -0.0023x^2 + 0.0222x + 0.0011$  ( $R^2 = 0.5193$ ); H:  $y = 0.0127\ln(x) + 0.0178$  ( $R^2 = 0.1291$ ). The position of each heterogeneity site is given in Table S3.

### 3.5. Correlation between Quasispecies Residue Variation and Mutations Recorded in HCV from Infected Patients

Each HCV mutant distribution found in an infected patient is the result of virus diversification under unique physiological conditions and immunological responses, thereby contributing to disparate mutational repertoires. Sites of variation in HCV replicating within individual patients may not segregate into the same conservation groups as the sites that mutate in cell culture. To explore this possibility, we examined the conservation category of the 197 amino acids (comprised between amino

acids 124 and 320 of NS5B), which included 177 variable positions in HCV from a cohort of 220 HCV-infected patients [22] (Tables S4 and S5). (The amino acid level is the one for which we had developed the informatics processing algorithms for HCVs of infected patients [22,29]). The results show again a spread among conservation groups (Figure 6). Thus, the HCV amino acid conservation pattern deduced from the LANL alignment does not fit the conservation observed in mutant spectra evolving in a cohort of infected patients.



**Figure 6.** Distribution of positions with variant amino acids identified in HCV from infected patients among amino acid conservation groups according to the LANL amino acid sequence alignment. The residues under study span amino acid 124 to amino acid 320 of protein NS5B (which correspond to genomic nucleotides 7971 to 8561; H77 numbering). (A) Assignment of amino acid position where substitutions were located (listed in Table S5) to amino acid conservation groups calculated from the amino acid sequence alignment of LANL. The 522 variant amino acids (within the 197 amino acids stretch) result from multiple substitutions at a given site due to the several HCV subtypes that infected the patients of the cohort under study [22] (shown also in Table S5). Conservation groups are indicated in abscissa, and the number of variants is indicated in ordinate. The total number of amino acids that fall in each conservation category is indicated in parenthesis in the upper box. The discontinuous line corresponds to function  $y = -96.03\ln(x) + 197.25$  ( $R^2 = 0.8028$ ). (B) Same as A, but with values normalized to the total number of amino acids that falls within each conservation group. The defining function is  $y = -0.3049x^2 + 2.9414x + 0.6407$  ( $R^2 = 0.2448$ ). (C,D) Same as A, B but with the 17 positions where a variant amino acid at a given position is common to infected patients and the HCV cell culture quasispecies. The defining functions are: C:  $y = 0.1326x^2 - 2.1311x + 8.3167$  ( $R^2 = 0.8948$ ); D:  $y = -0.0054x^2 + 0.0219x + 0.2249$  ( $R^2 = 0.3209$ ). In this case, a calculation of conservation groups using as reference the HCV amino acid sequence encoded in plasmid Jc1Luc was not possible because the patients under study were not infected with HCV of the same subtype. See Table S5 for the position of each amino acid substitution.

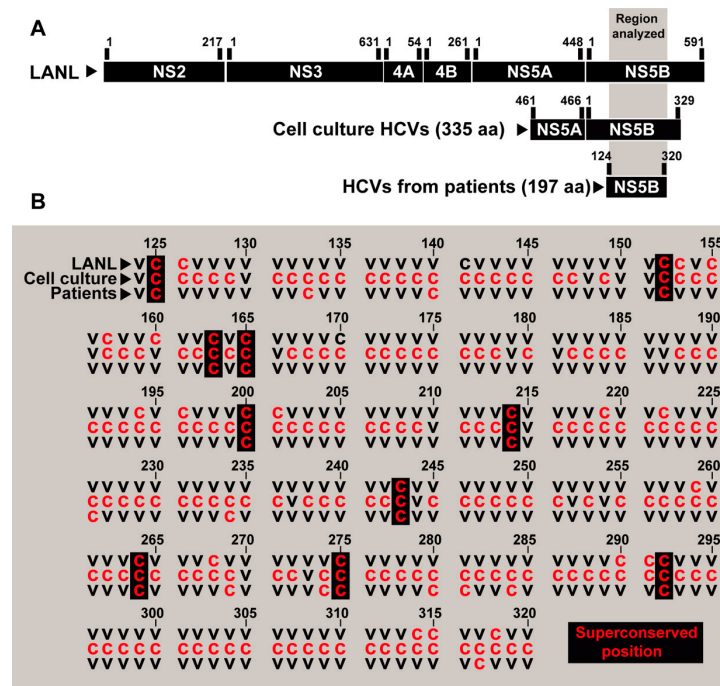
### 3.6. Evidence of Relaxed Mutational Acceptance in HCV Quasispecies

One possibility to explain that nucleotides and amino acids that are conserved in consensus sequences or in the LANL alignment vary in mutant spectra is a higher tolerance for mutations during quasispecies dynamics than reflected in the consensus sequences or in data bank repositories. To address this point, we performed additional calculations, including conservation group resampling with some specific HCV genotypes, and simulations with randomly chosen genomic positions (Figure S4). We compared the conservation score of each position between residues 2769 and 9377 for selected HCV genotypes in the LANL alignment (H77 numbering) (Figure S4A–D), and that of the heterogeneity sites. Both distributions were similar but with a larger accumulation of sites in the highest conservation window for the all-residue LANL calculation (compare Figure 4A and Figure S4A–H); the difference between them was statistically significant ( $p = 0.0125$ , chi-square test with Monte Carlo correction; calculation made after reducing the 6669 residues proportionally to 114 to equate both datasets).

The difference was also significant when the conservation groups were calculated using as reference the HCV sequence in Jc1Luc ( $p = 0.0065$ , chi-square test with Monte Carlo correction; Figure 4B and Figure S4B). As a simulation, the distribution of conservation groups of 114 positions randomly chosen between residues 2769 and 9377 (H77 numbering) carried out in triplicate was also similar to the experimental distribution, but again with a significant accumulation of sites in the highest conservation window (Tables S6 and S7, Figure S4K–N). Thus, the conservation groups of the variable sites in the HCV mutant spectra suggest a relaxed acceptance of mutations but significantly different from the prediction of a chance distribution of mutations.

### 3.7. A Possible Alternative Conservation Criterion

The relaxed (but not unlimited) acceptance of mutations in evolving quasispecies hints at the possibility that a new criterion for conservation might result from sequence alignments in which all mutations found in mutant spectra are included, independently of their origin or frequency level. We derived such alignment for NS5B amino acids 124 to 320, with the inclusion of all available cell culture and in vivo sequences used in the present study. The results (Figure 7) indicate that 5.1% of the amino acid positions are conserved in this new alignment. The proportion of conserved residues is dramatically reduced relative to the classical alignment with LANL sequences. Obviously, as additional mutant spectra of HCV populations are characterized by broad implementation of deep sequencing procedures in the clinic, the number of conserved residues is expected to diminish. Although absolute conservation cannot be foreseen as an attribute of the great majority of residues, the more stringent assessment of conservation attained with mutant spectrum alignments may provide a more realistic criterion on which to base the design of broad-spectrum antiviral ligands and vaccines.



**Figure 7.** Distribution of superconserved positions among the amino acids 124 to 320 of NS5B (corresponding to genomic residues 7971 to 8561; numbering according to HCV reference isolate H77). (A) Scheme of the different regions and the number of amino acids analyzed from patients, cell culture or LANL. (B) Sequence alignment to indicate conserved positions. All amino acid substitutions identified in quasispecies of HCV evolving in cell culture and in infected patients, analyzed in the present study, were considered to determine that a site is variable. Totally conserved positions are depicted by a red C, and variable positions (independently of the origin and extent of the variation) by a black V. Superconserved positions are indicated by a black rectangle.

#### 4. Discussion

The characterization of viral quasispecies *in vivo* is based mainly on intra-host mutant spectrum groupings of related sequences, time-dynamics analyses, and quantification of diversity indices [20,31–34]. These approaches have progressed quite independently from those employed to address long-term viral evolution, which are centered on phylogenetic and phylodynamic analyses of consensus sequences of independent viral isolates [6,7,35]. Attempts to connect the two domains of virus evolution have been limited [5]. The main objective of the present study has been to evaluate if conservation of HCV residues in consensus sequences or in data banks reflects also conservation at the level of mutant spectra of the populations. The comparisons have involved mutant spectra from a clonal HCV population passaged in cell culture, and viral populations from infected patients. HCVcc virus and Huh-7.5 cells were used because this system allows sustained viral progeny production during at least 200 serial passages with sufficient viral yields for reliable quantification of mutant frequencies [19]; *in vitro* systems to quantify extrahepatic HCV replication do not fulfill the requirements for our quantifications of mutant frequency variations [36,37].

In all cases examined, residue conservation defined in consensus sequences was ampler than that in mutant spectra. A similar discordance was quantified when mutant spectra were compared with HCV sequences in the LANL data bank. This quantitatively important lack of correspondence between conservation scores is probably influenced by a relaxed negative selection acting on the newly arising mutations that do not reach dominance in the populations (frequencies below 50% in the mutant spectrum). In addition to the loss of resolution inherent to the conversion of a mutant spectrum into a consensus sequence, the great majority of sequences deposited in LANL belong to viruses that have undergone selective filters and bottleneck events in natural environments. Specifically, immune responses are evoked in most infected individuals and population bottlenecks accompany host-to-host transmissions, and probably also intra-host events [38]. The results of the conservation scores of HCV sampled from infected patients are worth emphasizing. While HCV that replicates in cell culture broadens its mutant spectrum to a remarkable degree [7], diversification *in vivo* is expected to be even more accentuated when viruses from different patients are used for the comparisons. Despite this likely difference, the conclusions regarding the blurring of conservation records at the quasispecies level are strikingly similar in HCV from cell culture and *in vivo*. A possible criticism of our observations is that the level of mutation analysis is so different between consensus sequences and mutant spectra that no matter which comparisons we perform, the result will always be the same. This is not correct. In a recent study [39], we defined a subset of amino acid substitutions in HCV-infected patients, termed highly represented substitutions (HRS), identified in the same cohort of the present study [22]; in contrast with the complete set of substitutions, HRSs are distributed among intermediate conservation groups from the LANL alignment [39].

Differences in conservation between mutant spectra and consensus sequences or sequences in data banks emphasize two general points: (i) the relative (contingent) nature of residue conservation in viral genomes, and (ii) the simplification implied in the conversion of a mutant cloud (the real biological entity) into a single consensus sequence which is a weighted average of many different sequences. Concerning (i), in addition to the results reported here, there are other lines of evidence that support the contingency of residue conservation or variation. For example, upon passage of HCV in Huh-7.5 cells, the NS5A- and NS5B-coding regions accumulated many mutations while the hypervariable regions 1 and 2 (HVR1 and HVR2)—defined as hypervariable on the basis of comparison of clinical isolates—remained largely invariant [17]. In this case, variation was probably conditional upon the virus having responded to the immune response in infected patients [17]. Another example was provided by foot-and-mouth disease virus (FMDV), which includes an Arg-Gly-Asp triplet exposed on its capsid, and that serves both as a receptor recognition and antigenic site. The triplet is highly conserved among natural isolates, and was considered essential, yet it varied and proved dispensable for infection in cell culture [40], and in some cases also for infection *in vivo* [41]. Concerning (ii), we previously underlined that exclusion of mutant spectrum information in data banks represents

a limitation for many biological studies with viruses that exhibit quasispecies dynamics [38,42]. Such limitation has also been expressed on theoretical grounds when reducing information conveyed by individual sequences into a consensus sequence that may not even exist in the population it intends to represent [43].

As a practical consequence of our conclusions, long-term residue constancy, as calculated from sequence alignments in data banks, does not guarantee a limitation in the selection of escape mutants during intra-host evolution. It may be considered that some residues (i.e., those that are part of the catalytic site of viral enzymes) should be perfect ligands for inhibitors, no matter what the type of conservation correlations here described might indicate. However, ligand design is often based on several residues (or on structural elements that depend on several residues), and not all of them necessarily belong to a catalytic site; in addition, alteration of the conformation of an RNA structure or a protein catalytic domain by distant residues has been reported [44,45].

When a ligand is directed towards a viral population harboring a wide variant repertoire, the selection and survival of ligand-escape mutants will be critically dependent on the viral population size and the number of different sequences endowed with comparable fitness level. The broad sequence repertoire may guide minority subpopulations towards dominance despite the transiency of the initial mutations that mediated the escape. Therefore, residue conservation in data banks should not endorse the design of viral ligands intended to control viral quasispecies. A new, more stringent criterion for conservation based on meta-quasispecies analyses with minority mutant spectra included in the alignments would seem more adequate. Even with this new tool, the random nature of mutations, their rate of occurrence, the positive and negative intra-genome interactions among mutations (epistasis), and inter genome interactions (complementation, cooperation, interference) render evolutionary pathways unpredictable and changeable [38,46]. The design of universal vaccines and antiviral agents (with the potential of broad coverage to confront viral infections) is likely to remain an important challenge.

The elevated mutant repertoire that participates in quasispecies dynamics and its concealment in consensus sequences, together with evolutionary constraints at the epidemiological levels, may contribute to lower rates calculated for inter-host than intra-host evolution, one of the open questions in virus evolution [6,23–27]. Explorations of sequence space and opportunities for rapid short-term RNA virus evolution are far more abundant at the quasispecies level than reflected in data banks.

**Supplementary Materials:** The following are available online at <http://babia.cbm.uam.es/~lab121/SupplMatGarcia-Crespo>, Figure S1: Heat map of an alignment of the 39 consensus amino acid sequences determined for the NS5A-NS5B region [encoded by nucleotides 7649 to 8653; numbering according to isolate JFH-1 (GenBank accession number #AB047639)]. Figure S2: Degree of conservation of nucleotides and amino acids that participated in mutational waves, according to the LANL alignment using only sequences from genotype 2a (33 sequences). Figure S3: Number of nucleotides within HCV genomic residues 7584–8588 (H77 numbering) and deduced amino acid mixtures that belonged to heterogeneity sites, distributed among conservation groups according to the Los Alamos database alignment. Figure S4: Resampling with specific genotypes, and simulations of the distribution of HCV genomic nucleotides among conservation groups according to the LANL sequence alignment. Table S1: Mutations and corresponding amino acid substitutions of the NS5A-NS5B-coding region in the mutant spectra of viral populations analyzed by ultra-deep sequencing. Table S2: Reference accession numbers of sequences retrieved from Los Alamos database. Table S3: Mutations and corresponding amino acid substitutions deduced from the genomic sites with composition heterogeneity of the beginning of the NS2-coding region to the end of the NS5B-coding region in the genomic consensus sequence analysed by Sanger sequencing. Table S4: Number of patients infected by each HCV subtype. Table S5: Location within NS5B amino acids 124 to 320 of the positions where amino acid substitutions were identified in infected patients. Table S6: Random distribution of positions in the HCV genome. Table S7: Statistical analysis of the different distributions identified between HCV genomic residues 2769 (beginning of the NS2-coding region) and 9377 (end of the NS5B-coding region) (H77 numbering).

**Author Contributions:** Conceptualization, E.D. and C.P.; methodology, M.E.S., I.G. and A.I.d.Á.; software, J.G. (Josep Gregori); validation, J.G. (Jordi Gómez) and C.B.; formal analysis, C.G.-C., M.E.S. and J.Q.; investigation, C.G.-C., M.E.S.; resources, J.Q., E.D. and C.P.; data curation, B.M.-G., L.V.-S.; writing—original draft preparation, E.D.; writing—review and editing, E.D., C.P., J.G. (Jordi Gómez), C.B. and C.G.-C.; visualization, M.E.S.; supervision, E.D. and C.P.; project administration, A.I.d.Á., I.G.; funding acquisition, J.Q., E.D. and C.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research at Centro de Biología Molecular Severo Ochoa (CSIC-UAM) (CBMSO) was funded by Ministerio de Economía y Competitividad (MINECO), grant numbers SAF2014-52400-R and SAF2017-87846-R, and by Ministerio de Ciencia, Innovación y Universidades (MCIU), grant number BFU2017-91384-EXP, and by Instituto de Salud Carlos III, grant number PI18/00210, and by PLATESA from Comunidad de Madrid/FEDER, grant numbers S2013/ABI-2906 and S2018/BAA-4370. C.P. is supported by the Miguel Servet program of the Instituto de Salud Carlos III (CP14/00121 and CPII19/00001) cofinanced by the European Regional Development Fund (ERDF). CIBERehd (Centro de Investigación en Red de Enfermedades Hepáticas y Digestivas) is funded by Instituto de Salud Carlos III. Institutional grants from the Fundación Ramón Areces and Banco Santander to the CBMSO are also acknowledged. The research at Vall d’Hebron Institut de Recerca (VHIR) was funded by Instituto de Salud Carlos III, cofinanced by the European Regional Development Fund (ERDF), grant number PI19/00301 and by the Centro para el Desarrollo Tecnológico Industrial (CDTI) from the MCIU, grant number IDI-20151125. The research at Centro de Astrobiología (CAB) was supported by MINECO, grant numbers BIO2016-79618R and PID2019-104903RB-I00 (funded by EU under the FEDER program) and by the Spanish State research agency (AEI), grant number MDM-2017-0737 (Unidad de Excelencia “María de Maeztu”-Centro de Astrobiología (CSIC-INTA). C.G.-C. is supported by predoctoral contract PRE2018-083422 from MCIU. B.M.-G. is supported by predoctoral contract PFIS FI19/00119 from Instituto de Salud Carlos III (Ministerio de Sanidad y Consumo) cofinanced by Fondo Social Europeo (FSE).

**Acknowledgments:** We are indebted to J.C. de la Torre and Matthias Pauthner for the critical reading of the manuscript and valuable comments. The team at CBMSO belongs to the Global Virus Network (GVN).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Bartenschlager, R.; Baumert, T.F.; Bukh, J.; Houghton, M.; Lemon, S.M.; Lindenbach, B.D.; Lohmann, V.; Moradpour, D.; Pietschmann, T.; Rice, C.M.; et al. Critical challenges and emerging opportunities in hepatitis C virus research in an era of potent antiviral therapy: Considerations for scientists and funding agencies. *Virus Res.* **2018**, *248*, 53–62. [[CrossRef](#)]
2. Farci, P. New insights into the HCV quasispecies and compartmentalization. *Semin. Liver Dis.* **2011**, *31*, 356–374. [[CrossRef](#)] [[PubMed](#)]
3. Mercuri, L.; Thomson, E.C.; Hughes, J.; Karayiannis, P. Quasispecies Changes with Distinctive Point Mutations in the Hepatitis C Virus Internal Ribosome Entry Site (IRES) Derived from PBMCs and Plasma. *Adv. Virol.* **2018**, *2018*, 4835252. [[CrossRef](#)] [[PubMed](#)]
4. Marascio, N.; Quirino, A.; Barreca, G.S.; Galati, L.; Costa, C.; Pisani, V.; Mazzitelli, M.; Matera, G.; Liberto, M.C.; Foca, A.; et al. Discussion on critical points for a tailored therapy to cure hepatitis C virus infection. *Clin. Mol. Hepatol.* **2019**, *25*, 30–36. [[CrossRef](#)] [[PubMed](#)]
5. Glebova, O.; Knyazev, S.; Melnyk, A.; Artyomenko, A.; Khudyakov, Y.; Zelikovsky, A.; Skums, P. Inference of genetic relatedness between viral quasispecies from sequencing data. *BMC Genomics* **2017**, *18*, 918. [[CrossRef](#)] [[PubMed](#)]
6. Geoghegan, J.L.; Holmes, E.C. Evolutionary Virology at 40. *Genetics* **2018**, *210*, 1151–1162. [[CrossRef](#)] [[PubMed](#)]
7. Domingo, E.; Soria, M.E.; Gallego, I.; de Ávila, A.I.; García-Crespo, C.; Martínez-González, B.; Gómez, J.; Briones, C.; Gregori, J.; Quer, J.; et al. A new implication of quasispecies dynamics: Broad virus diversification in absence of external perturbations. *Infect. Genet. Evol.* **2020**, *82*, 104278. [[CrossRef](#)]
8. Nachbagauer, R.; Krammer, F. Universal influenza virus vaccines and therapeutic antibodies. *Clin. Microbiol. Infect.* **2017**, *23*, 222–228. [[CrossRef](#)]
9. Wijesundara, D.K.; Gummow, J.; Li, Y.; Yu, W.; Quah, B.J.; Ranasinghe, C.; Torresi, J.; Gowans, E.J.; Grubor-Bauk, B. Induction of Genotype Cross-Reactive, Hepatitis C Virus-Specific, Cell-Mediated Immunity in DNA-Vaccinated Mice. *J. Virol.* **2018**, *92*, e02133-17. [[CrossRef](#)]
10. Hart, G.R.; Ferguson, A.L. Computational design of hepatitis C virus immunogens from host-pathogen dynamics over empirical viral fitness landscapes. *Phys. Biol.* **2018**, *16*, 16004. [[CrossRef](#)]
11. McLean, G.R. Vaccine strategies to induce broadly protective immunity to rhinoviruses. *Hum. Vaccin. Immunother.* **2020**, *16*, 684–686. [[CrossRef](#)] [[PubMed](#)]
12. Miller, M.M. Sofosbuvir-velpatasvir: A single-tablet treatment for hepatitis C infection of all genotypes. *Am. J. Health Syst. Pharm.* **2017**, *74*, 1045–1052. [[CrossRef](#)]



13. Vogel, O.A.; Manicassamy, B. Broadly Protective Strategies Against Influenza Viruses: Universal Vaccines and Therapeutics. *Front. Microbiol.* **2020**, *11*, 135. [[CrossRef](#)]
14. Zhong, J.; Gastaminza, P.; Cheng, G.; Kapadia, S.; Kato, T.; Burton, D.R.; Wieland, S.F.; Uprichard, S.L.; Wakita, T.; Chisari, F. V Robust hepatitis C virus infection in vitro. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9294–9299. [[CrossRef](#)]
15. Wakita, T.; Pietschmann, T.; Kato, T.; Date, T.; Miyamoto, M.; Zhao, Z.; Murthy, K.; Habermann, A.; Krausslich, H.G.; Mizokami, M.; et al. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* **2005**, *11*, 791–796. [[CrossRef](#)]
16. Lindenbach, B.D.; Evans, M.J.; Syder, A.J.; Wolk, B.; Tellinghuisen, T.L.; Liu, C.C.; Maruyama, T.; Hynes, R.O.; Burton, D.R.; McKeating, J.A.; et al. Complete replication of hepatitis C virus in cell culture. *Science* **2005**, *309*, 623–626. [[CrossRef](#)]
17. Moreno, E.; Gallego, I.; Gregori, J.; Lucia-Sanz, A.; Soria, M.E.; Castro, V.; Beach, N.M.; Manrubia, S.; Quer, J.; Esteban, J.I.; et al. Internal Disequilibria and Phenotypic Diversification during Replication of Hepatitis C Virus in a Noncoevolving Cellular Environment. *J. Virol.* **2017**, *91*, e02505-16. [[CrossRef](#)]
18. Perales, C.; Beach, N.M.; Gallego, I.; Soria, M.E.; Quer, J.; Esteban, J.I.; Rice, C.; Domingo, E.; Sheldon, J. Response of hepatitis C virus to long-term passage in the presence of alpha interferon: Multiple mutations and a common phenotype. *J. Virol.* **2013**, *87*, 7593–7607. [[CrossRef](#)]
19. Gallego, I.; Soria, M.E.; Garcia-Crespo, C.; Chen, Q.; Martinez-Barragan, P.; Khalfaoui, S.; Martinez-Gonzalez, B.; Sanchez-Martin, I.; Palacios-Blanco, I.; de Avila, A.I.; et al. Broad and Dynamic Diversification of Infectious Hepatitis C Virus in a Cell Culture Environment. *J. Virol.* **2020**, *94*, e01856-19. [[CrossRef](#)]
20. Gregori, J.; Salicru, M.; Domingo, E.; Sanchez, A.; Esteban, J.I.; Rodriguez-Frias, F.; Quer, J. Inference with viral quasispecies diversity indices: Clonal and NGS approaches. *Bioinformatics* **2014**, *30*, 1104–1111. [[CrossRef](#)]
21. Gregori, J.; Perales, C.; Rodriguez-Frias, F.; Esteban, J.I.; Quer, J.; Domingo, E. Viral quasispecies complexity measures. *Virology* **2016**, *493*, 227–237. [[CrossRef](#)]
22. Chen, Q.; Perales, C.; Soria, M.E.; Garcia-Cehic, D.; Gregori, J.; Rodriguez-Frias, F.; Buti, M.; Crespo, J.; Calleja, J.L.; Tabernero, D.; et al. Deep-sequencing reveals broad subtype-specific HCV resistance mutations associated with treatment failure. *Antivir. Res.* **2020**, *174*, 104694. [[CrossRef](#)]
23. Ali, A.; Melcher, U. Modeling of Mutational Events in the Evolution of Viruses. *Viruses* **2019**, *11*, 418. [[CrossRef](#)] [[PubMed](#)]
24. Simmonds, P.; Aiewsakun, P.; Katzourakis, A. Prisoners of war—host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol.* **2019**, *17*, 321–328. [[CrossRef](#)] [[PubMed](#)]
25. Leslie, A.J.; Pfafferott, K.J.; Chetty, P.; Draenert, R.; Addo, M.M.; Feeney, M.; Tang, Y.; Holmes, E.C.; Allen, T.; Prado, J.G.; et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* **2004**, *10*, 282–289. [[CrossRef](#)]
26. Herbeck, J.T.; Nickle, D.C.; Learn, G.H.; Gottlieb, G.S.; Curlin, M.E.; Heath, L.; Mullins, J.I. Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *J. Virol.* **2006**, *80*, 1637–1644. [[CrossRef](#)] [[PubMed](#)]
27. Redd, A.D.; Collinson-Streng, A.N.; Chatziandreou, N.; Mullis, C.E.; Laeyendecker, O.; Martens, C.; Ricklefs, S.; Kiwanuka, N.; Nyein, P.H.; Lutalo, T.; et al. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. *J. Infect. Dis.* **2012**, *206*, 1433–1442. [[CrossRef](#)]
28. Marukian, S.; Jones, C.T.; Andrus, L.; Evans, M.J.; Ritola, K.D.; Charles, E.D.; Rice, C.M.; Dustin, L.B. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* **2008**, *48*, 1843–1850. [[CrossRef](#)]
29. Soria, M.E.; Gregori, J.; Chen, Q.; Garcia-Cehic, D.; Llorens, M.; de Avila, A.I.; Beach, N.M.; Domingo, E.; Rodriguez-Frias, F.; Buti, M.; et al. Pipeline for specific subtype amplification and drug resistance detection in hepatitis C virus. *BMC Infect. Dis.* **2018**, *18*, 446. [[CrossRef](#)]
30. Kuiken, C.; Combet, C.; Bukh, J.; Shin, I.T.; Deleage, G.; Mizokami, M.; Richardson, R.; Sablon, E.; Yusim, K.; Pawlotsky, J.M.; et al. A comprehensive system for consistent numbering of HCV sequences, proteins and epitopes. *Hepatology* **2006**, *44*, 1355–1361. [[CrossRef](#)]
31. Baccam, P.; Thompson, R.J.; Fedrigo, O.; Carpenter, S.; Cornette, J.L. PAQ: Partition Analysis of Quasispecies. *Bioinformatics* **2001**, *17*, 16–22. [[CrossRef](#)]

32. Skums, P.; Zelikovsky, A.; Singh, R.; Gussler, W.; Dimitrova, Z.; Knyazev, S.; Mandric, I.; Ramachandran, S.; Campo, D.; Jha, D.; et al. QUENTIN: Reconstruction of disease transmissions from viral quasispecies genomic data. *Bioinformatics* **2018**, *34*, 163–170. [[CrossRef](#)]
33. Ahn, S.; Ke, Z.; Vikalo, H. Viral quasispecies reconstruction via tensor factorization with successive read removal. *Bioinformatics* **2018**, *34*, i23–i31. [[CrossRef](#)] [[PubMed](#)]
34. Henningsson, R.; Moratorio, G.; Borderia, A.V.; Vignuzzi, M.; Fontes, M. DISSEQT-DIStribution-based modeling of SEquence space Time dynamics. *Virus Evol.* **2019**, *5*, vez028. [[CrossRef](#)]
35. Grenfell, B.T.; Pybus, O.G.; Gog, J.R.; Wood, J.L.; Daly, J.M.; Mumford, J.A.; Holmes, E.C. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **2004**, *303*, 327–332. [[CrossRef](#)]
36. Blackard, J.T.; Nyingi, K.; Sherman, K.E. Extrahepatic replication of HCV: Insights into clinical manifestations and biological consequences. *Hepatology* **2006**, *44*, 15–22. [[CrossRef](#)]
37. Skardasi, G.; Chen, A.Y.; Michalak, T.I. Authentic Patient-Derived Hepatitis C Virus Infects and Productively Replicates in Primary CD4 + and CD8 + T Lymphocytes In Vitro. *J. Virol.* **2018**, *92*, e01790-17. [[CrossRef](#)]
38. Domingo, E.; Perales, C. Viral quasispecies. *PLoS Genet.* **2019**, *15*, e1008271. [[CrossRef](#)]
39. Soria, M.E.; García-Crespo, C.; Martínez-González, B.; Vazquez-Sirvent, L.; Lobo-Vega, R.; de Ávila, A.I.; Gallego, I.; Ferrer-Orta, C.; Verdaguer, N.; Chen, Q.; et al. Amino acid substitution associated with treatment failure of hepatitis C virus infection. *J. Clin. Microbiol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
40. Martínez, M.A.; Verdaguer, N.; Mateu, M.G.; Domingo, E. Evolution subverting essentiality: Dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6798–6802. [[CrossRef](#)] [[PubMed](#)]
41. Tami, C.; Taboga, O.; Berinstein, A.; Nuñez, J.I.; Palma, E.L.; Domingo, E.; Sobrino, F.; Carrillo, E. Evidence of the coevolution of antigenicity and host cell tropism of foot-and-mouth disease virus in vivo. *J. Virol.* **2003**, *77*, 1219–1226. [[CrossRef](#)] [[PubMed](#)]
42. Domingo, E.; Brun, A.; Nuñez, J.I.; Cristina, J.; Briones, C.; Escarmís, C. Genomics of Viruses. In *Pathogenomics: Genome Analysis of Pathogenic Microbes*; Hacker, J., Dobrindt, U., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006; pp. 369–388.
43. Eigen, M. *From Strange Simplicity to Complex Familiarity*; Oxford University Press: Oxford, UK, 2013.
44. Van Slyke, G.A.; Arnold, J.J.; Lugo, A.J.; Griesemer, S.B.; Moustafa, I.M.; Kramer, L.D.; Cameron, C.E.; Ciota, A.T. Sequence-Specific Fidelity Alterations Associated with West Nile Virus Attenuation in Mosquitoes. *PLoS Pathog.* **2015**, *11*, e1005009. [[CrossRef](#)]
45. Shatoff, E.; Bundschuh, R. Single nucleotide polymorphisms affect RNA-protein interactions at a distance through modulation of RNA secondary structures. *PLoS Comput. Biol.* **2020**, *16*, e1007852. [[CrossRef](#)]
46. Holland, J.J.; de La Torre, J.C.; Steinhauer, D.A. RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* **1992**, *176*, 1–20.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).