

RESEARCH ARTICLE

# Rare haplotype load as marker for lethal mutagenesis

Josep Gregori <sup>1,2,3</sup>, María Eugenia Soria <sup>1</sup>, Isabel Gallego<sup>2,4</sup>, Mercedes Guerrero-Murillo<sup>1</sup>, Juan Ignacio Esteban<sup>1,2,5</sup>, Josep Quer <sup>1,2,5\*</sup>, Celia Perales <sup>1,2,4\*</sup>, Esteban Domingo<sup>2,4</sup>

**1** Liver Unit, Liver Disease Laboratory-Viral Hepatitis, Internal Medicine Department, Vall d'Hebron Institut Recerca (VHIR)-Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Spain, **2** Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) del Instituto de Salud Carlos III, Madrid, Spain, **3** Roche Diagnostics, S.L., Sant Cugat del Vallés, Barcelona, Spain, **4** Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Campus de Cantoblanco, Madrid, Spain, **5** Universitat Autònoma de Barcelona, Barcelona, Spain

\* [celia.perales@vhir.org](mailto:celia.perales@vhir.org) (CP); [josep.quer@vhir.org](mailto:josep.quer@vhir.org) (JQ)



 OPEN ACCESS

**Citation:** Gregori J, Soria ME, Gallego I, Guerrero-Murillo M, Esteban JI, Quer J, et al. (2018) Rare haplotype load as marker for lethal mutagenesis. *PLoS ONE* 13(10): e0204877. <https://doi.org/10.1371/journal.pone.0204877>

**Editor:** Jean-Pierre Vartanian, Institut Pasteur, FRANCE

**Received:** July 18, 2018

**Accepted:** August 19, 2018

**Published:** October 3, 2018

**Copyright:** © 2018 Gregori et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Work in Barcelona was funded by Instituto de Salud Carlos III; by grants PI13/00456, PI15/00829, and PI16/00337 cofinanced by the European Regional Development Fund (ERDF); and by CDTI (Centro para el Desarrollo Tecnológico Industrial), Spanish Ministry of Economics and Competitiveness (MINECO), IDI-20151125. CP is supported by the Miguel Servet program of the Instituto de Salud Carlos III (CP14/00121)

## Abstract

RNA viruses replicate with a template-copying fidelity, which lies close to an extinction threshold. Increases of mutation rate by nucleotide analogues can drive viruses towards extinction. This transition is the basis of an antiviral strategy termed lethal mutagenesis. We have introduced a new diversity index, the rare haplotype load (RHL), to describe NS5B (polymerase) mutant spectra of hepatitis C virus (HCV) populations passaged in absence or presence of the mutagenic agents favipiravir or ribavirin. The increase in RHL is more prominent in mutant spectra whose expansions were due to nucleotide analogues than to multiple passages in absence of mutagens. Statistical tests for paired mutagenized versus non-mutagenized samples with 14 diversity indices show that RHL provides consistently the highest standardized effect of mutagenic treatment difference for ribavirin and favipiravir. The results indicate that the enrichment of viral quasispecies in very low frequency minority genomes can serve as a robust marker for lethal mutagenesis. The diagnostic value of RHL from deep sequencing data is relevant to experimental studies on enhanced mutagenesis of viruses, and to pharmacological evaluations of inhibitors suspected to have a mutagenic activity.

## Introduction

The mutant spectra of RNA viruses are a reflection of their evolutionary history, as well as important determinants of virus adaptability. Concerning control of viral diseases, mutant spectrum dynamics is an obstacle for the efficacy of therapeutic interventions due to selection of treatment-escape viral mutants. The antiviral agents to combat RNA viruses include those directed to specific viral targets [direct-acting antiviral agents (DAAs)], and those that inhibit cellular functions needed for the completion of the virus life cycle. The viral RNA-dependent RNA polymerase (RdRp) is the target of several effective antiviral agents. Some of them, notably base or nucleoside analogues, are intracellularly converted into their active nucleotide

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. The work in Madrid was supported by Spanish Ministry of Economics and Competitiveness (MINECO) grants number BFU-2011-23604, SAF2014-52400-R, SAF2017-87846-R and S2013/ABI-2906 (PLATESA from Comunidad de Madrid/FEDER). Institutional grants from the Fundación Ramón Areces and Banco Santander to the CBMSO are also acknowledged.

**Competing interests:** The funder “Roche Diagnostics S.L.” provided support in the form of salaries for author [JG], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. No other Competing Interests to declare. Thus, our adherence to PLOS ONE policies on sharing data and materials is not altered.

counterparts. The discovery that ribavirin (1- $\beta$ -D-ribofuranosyl-1-*H*-1,2,4-triazole-3-carboxamide) is mutagenic for poliovirus [1] introduced a new perspective in the antiviral mechanism of some nucleotide analogues. Three alternative —not mutually exclusive— mechanisms of anti-RdRp activity by nucleotide analogues have been described: RNA chain termination, inhibition of RNA synthesis without chain termination, and inhibition associated with viral genome mutagenesis.

Nucleotide analogue-induced mutagenesis is equivalent to a decrease of copying fidelity by the viral RdRp. Quasispecies theory predicts a maximum amount of genetic information that can be transmitted for a given average copying fidelity. This concept is mathematically formulated in the form of an error threshold relationship. An increase in mutation rate drives the population across the error threshold, into error catastrophe, equated with loss of inheritable information [2, 3]. The error threshold applies to finite populations in variable fitness landscapes, and its position in a fidelity scale depends also on the degree of adaptation of the mutant ensemble to the environment [3]. The error threshold concept has found experimental support in studies on the negative effects of chemical mutagenesis on the survival of RNA viruses ([4–8], among other studies). The convergence of theoretical and experimental results opened the way to lethal mutagenesis as an antiviral strategy [9].

The licensed antiviral nucleoside analogues favipiravir (T-705; 6-fluoro-3-hydroxy-2-pirazinocarboxamide) and ribavirin are mutagenic for several RNA viruses. Ribavirin has been used as antiviral agent for decades [10, 11], and only recently shown to be mutagenic for several RNA viruses [1, 12, 13]. Favipiravir has been licensed as an anti-influenza agent in Japan having potent antiviral activity against different influenza virus strains (types A, B and C) including those resistant to neuraminidase and M2 inhibitors [14]. Favipiravir has also been effective to inhibit the replication of other RNA viruses in vitro and in animal models, including flavivirus, norovirus, bunyavirus, arenavirus, filovirus and other RNA virus for which no antiviral therapy is currently available [reviewed in (15, 16)]. Favipiravir is converted intracellularly into the ribofuranosyl 5'-triphosphate metabolite (favipiravir-RTP) and in this form it can be recognized as a pseudopurine by the RdRp [14, 17]. Its selective inhibition of RdRp implicates a wider antiviral spectrum with a limited cell damage compared with other mutagens such as Ribavirin.

Incorporation of favipiravir-RTP in the nascent viral RNA could result in lethal mutagenesis, as has been proposed for influenza virus [18], norovirus [7], hepatitis C virus (HCV) [19]; foot-and-mouth disease virus [20], West Nile virus [21], Dengue virus [22] and Ebola virus [8]; coxsackievirus B3 [23]. It is not clear whether favipiravir acts as RNA chain terminator, inhibitor, mutagen or by a combination of these mechanisms; its dominant mode of action may depend on the virus-host system and concentration of the active form. In order to cause lethal mutagenesis, favipiravir-RTP needs to be incorporated into the RNA without causing immediate chain termination. It is possible that both lethal mutagenesis and chain termination occur depending on the available concentration of favipiravir-RTP. It has been hypothesized that incorporation of low levels of favipiravir-RTP could result in full-length extension of the viral RNA, leading to lethal mutagenesis and lower infectivity [24].

The standard way to distinguish an RNA virus mutagen from a non-mutagenic inhibitor, is that a mutagenic inhibitor promotes an increase of mutant spectrum diversity, and a decrease of the virus specific infectivity (defined as the ratio of the number of infectious units to the amount of viral RNA in the virus population) (reviewed in [25, 26]). The application of deep sequencing to the analysis of viral populations has introduced several new diversity indices that allow a more detailed description of mutant spectra [27–29]. Some diversity indices were adopted from ecology, and are classified in three groups: incidence (based on the count of entities in a multiple alignment of haplotypes), abundance (that considers both, counts of entities and their frequency), and functional (based on differences among the observed haplotypes)

[27]. The value of alternative diversity indices to diagnose the mechanism underlying the expansion of mutant spectra is an unsolved issue.

We have adapted a HCV serial passage design to study the genetic and phenotypic diversification of HCV in Huh-7.5 reporter cells in absence of cellular evolution [30–32]. The parental (plasmid-derived) HCV population was passaged in absence or presence of ribavirin or favipiravir. Populations whose mutant spectrum was expanded in absence of drugs were also subjected to mutagenesis. The design produced several HCV populations for comparative mutant spectrum analyses. NS5B amplicons were analyzed to quantify mutant spectrum complexity. We describe a new diversity index, the rare haplotype load (RHL), and show that its variation outstands among that of other diversity indices to characterize mutant spectra in their transition into error catastrophe. RHL may help in the understanding of quasispecies dynamics, and in the clarification of the mechanisms of action of antiviral agents.

## Results

### The rare haplotype load of hepatitis C virus populations

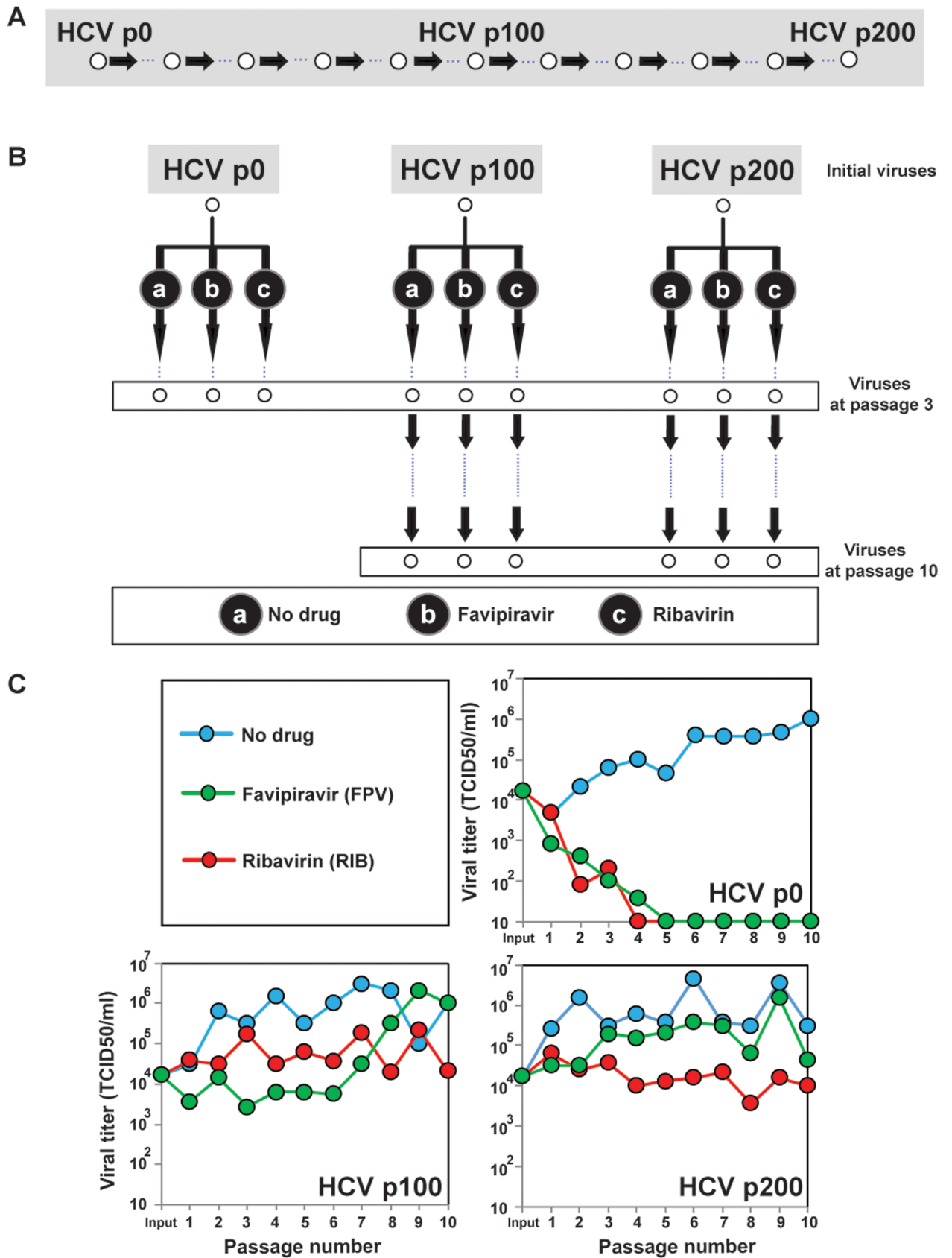
HCV RNA expressed from plasmid Jc1FLAG2(p7-nsGluc2A) (genotype 2a) [33] was transfected into Huh-7 Lunet cells and amplified in Huh-7.5 cells to produce the initial virus population HCV p0 [30]. HCV p0 was subjected to 200 serial passages in Huh-7.5 reporter cells in the absence of any drug. The populations at passage 100 (HCV p100) and at passage 200 (HCV p200) displayed increased replication in Huh-7.5 reported cells [31, 32]. HCV p0, HCV p100 and HCV p200 were further passaged either in the absence of any drug or in the presence of favipiravir or ribavirin (Fig 1A). Infectious progeny levels were those expected from previous quantifications of inhibition of HCV p0 by favipiravir [19] and ribavirin [34]; the sustained HCV p100 and HCV p200 production in the presence of the drugs is expected from the fitness-associated HCV resistance to antiviral agents [31, 32, 35] (Fig 1B).

Intracellular viral RNA was sequenced in MiSeq with 2x300 mode with v3 chemistry, and fastq files were analyzed as previously described [27, 29] to obtain forward and reverse consensus haplotypes with abundances not below 0.1%, median coverage 147,000 reads, interquartile range (IQR) 75570–226100. The fasta files obtained for each sample were further subjected to DSFT for diversity indices computation. The resulting median coverage was of 139200 with IQR 71,480–210,600 reads.

We introduce the rare haplotype load (RHL) as a new diversity index which may be considered intermediate between incidence and abundance indices. In the context of this work we define as rare those haplotypes with abundances below a given threshold (1%), and as load the fraction of molecules in the quasispecies belonging to these haplotypes. Translating this concept to next generation sequencing is not an easy task as we wish to take into account full reads in the range of abundances below the instrument noise level. Our approach consisted in taking all reads corresponding to haplotypes common to the forward and reverse strand with no previous abundance filtering and computing the RHL as the fraction of reads belonging to haplotypes with abundances below 1%. We suppose that technical noise affects equally all samples in the experiment and that the distinctive effect would be caused by the treatment. This index was not submitted to any sample size correction.

### Comparison of RHL with other diversity indices

A full set of diversity indices was computed for each sample: Hpl, number of haplotypes; PolySites, number of polymorphic sites; nMuts, number of mutations; Shannon, Shannon entropy; GiniS, Gini-Simpson index; q1D, Hill number of order 1; q2D, Hill number of order 2; qInfD, Hill number of order infinity; FAD, functional attribute diversity; Mf.e, mutation frequency by



**Fig 1. Experimental design and infectious HCV progeny production in absence or presence of favipiravir or ribavirin.** (A) Passage of HCV p0 in Huh-7.5 reporter cells to derive high fitness HCV p100 and HCV p200 populations. (B) Serial passages of HCV p0, HCV p100 and HCV p200 in absence of drugs (No drug) or the presence of 400  $\mu$ M favipiravir, or 100  $\mu$ M ribavirin. (C) Infectious progeny production during 10 serial passages in absence or presence of the drugs. Details on the infections are given in Materials and Methods.

<https://doi.org/10.1371/journal.pone.0204877.g001>

entity; Pi.e, nucleotide diversity by entity; Pi, nucleotide diversity, and Mf, mutation frequency. The correlation among all indices (including RHL) resulting from all samples in the experimental design, shows a structure with three groups (Fig 2) G1: RHL, Hpl, FAD, nMuts and PolySites; G2: Shannon, GiniS, q1D, q2D and qInfD; and G3: Mf.e, Pi.e, Mf and Pi. The three indices more correlated to RHL are Hpl 0.895, FAD 0.854 and Shannon 0.826. RHL falls within the group of incidence indices, but it results from the aggregation of abundances. Its high correlation with Shannon and q1D denotes properties of abundance-based indices, while its high correlation with FAD confers to RHL properties of functional incidence. These three properties were expected from the definition of RHL, and the correlations provide an empirical prove of the computations adequacy.

### Association tests

Table 1 summarizes the results of the association tests of each diversity index considered, including RHL, to mutagenicity, for each drug (favipiravir and ribavirin) sorted by decreasing order of standardized effect. No distinction has been made of amplicon or treatment length. Hence the results represent averaging over the full NS5B region sequenced, and over the two treatment lengths. The RHL is the index with the highest standardized effect among all in both mutagenic treatments, with adjusted p-values of the order of  $10^{-4}$ . Top indices are also FAD, Hpl, Shannon, nMuts, and PolySites. No significant association is found in neither treatment, for Mf and Pi, both at the abundance and at the entity level. RHL is still the top indicator when distinguishing among amplicons or treatment length, followed by FAD and Hpl.

### Logistic regression

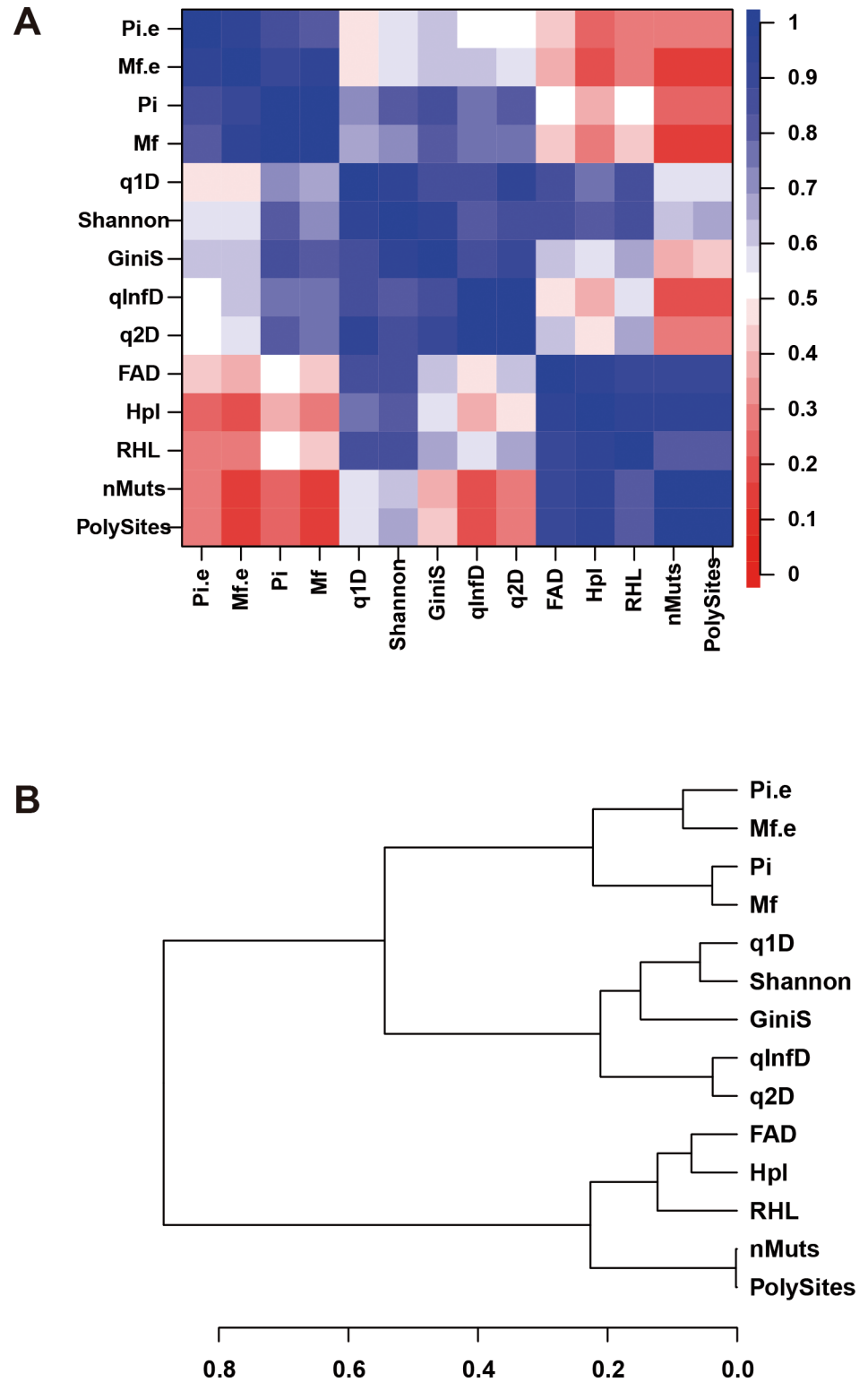
As a further step towards characterizing RHL as mutagenicity marker, samples were relabeled as under mutagenicity (Mut) if treated with favipiravir or ribavirin, or not mutagenized (control) for passages in absence of drug. Each group includes the variability due to the factors amplicon, drug, and treatment length. Then a univariate logistic regression was fit with each of the diversity indices, including RHL (Figs 3 and S1). The regression to RHL resulted in the best value of the Aikake Information content (AIC), the highest area under the ROC curve (AUC), and the lowest leave-one out cross-validation (LOOCV) error rate to classification (Table 2). According to the LOOCV error rate the predictive capacity of these indices follows the order RHL, Hpl, PolySites, nMuts and FAD (S2 Fig).

When this analysis was performed separately for each treatment length, for 10 passes both RHL, Shannon and q1D resulted in the lowest values of AIC and LOOCV, for 3 passes RHL, FAD and Hpl resulted in the lowest values of AIC and LOOCV (S3 and S4 Figs). No multivariate logistic regression model resulted better than RHL as a single predictor in the whole dataset.

These results prove that within the experimental design of this study RHL is the most sensitive diversity index for predicting mutagenic effects, with independence of factors such as amplicon, base line passage, drug, and treatment length.

### In-silico study to test the robustness and unbiasedness of RHL

To study the robustness of RHL and its possible dependence on sample size, an in-silico study on the full set of fasta files with no exclusion was performed. Robustness was evaluated by



**Fig 2. Correlation among diversity.** (A) Plot illustrating the correlation between diversity indices in this study. The correspondences between colors and correlation values is shown on the right bar. (B) Hierarchical clustering of diversity indices computed with the square root of one minus the correlation matrix, as measure of dissimilarity.

<https://doi.org/10.1371/journal.pone.0204877.g002>



**Table 1. Wilcoxon signed rank test of paired samples, treatment vs control.** Rows sorted in descending order of standardized effect. Estimate: median of treatment difference, SD: standard deviation, StdEffect: standardized effect (median divided by sd), p.value: Wilcoxon test p-value, adj.pv: multiple test adjusted p-value by the Bonferroni method.

	Ribavirin						Favipiravir				
	Estimate	SD	StdEffect	p.value	adj.pv		Estimate	SD	StdEffect	p.value	adj.pv
<b>RHL</b>	1,49E+01	5,09E+00	2,920	3,05E-05	4,27E-04	<b>RHL</b>	1,81E+01	3,86E+00	4,680	3,05E-05	4,27E-04
<b>nMuts</b>	2,05E+01	1,00E+01	2,040	3,58E-04	5,01E-03	<b>Shannon</b>	8,76E-01	4,22E-01	2,080	3,05E-05	4,27E-04
<b>Hpl</b>	2,25E+01	1,11E+01	2,020	3,61E-04	5,06E-03	<b>FAD</b>	1,74E+01	9,33E+00	1,860	6,10E-05	8,54E-04
<b>PolySites</b>	2,00E+01	9,92E+00	2,010	3,58E-04	5,01E-03	<b>Hpl</b>	2,75E+01	1,54E+01	1,790	5,45E-04	7,62E-03
<b>FAD</b>	1,21E+01	6,55E+00	1,850	3,05E-05	4,27E-04	<b>GiniS</b>	2,15E-01	1,39E-01	1,550	2,14E-04	2,99E-03
<b>Shannon</b>	6,99E-01	3,89E-01	1,800	3,05E-05	4,27E-04	<b>q1D</b>	5,18E+00	3,33E+00	1,550	3,05E-05	4,27E-04
<b>GiniS</b>	1,62E-01	1,27E-01	1,280	3,05E-04	4,27E-03	<b>nMuts</b>	2,60E+01	1,77E+01	1,470	1,59E-03	2,22E-02
<b>q1D</b>	4,04E+00	3,95E+00	1,020	3,05E-05	4,27E-04	<b>PolySites</b>	2,40E+01	1,71E+01	1,400	1,18E-03	1,65E-02
<b>qInfD</b>	7,74E-01	7,64E-01	1,010	5,80E-04	8,12E-03	<b>q2D</b>	1,56E+00	1,74E+00	0,901	3,05E-04	4,27E-03
<b>q2D</b>	1,58E+00	2,06E+00	0,766	3,05E-04	4,27E-03	<b>qInfD</b>	6,71E-01	9,05E-01	0,741	1,31E-03	1,84E-02
<b>Mf.e</b>	6,52E-04	1,55E-03	0,420	5,35E-02	7,49E-01	<b>Pi</b>	1,22E-03	2,04E-03	0,596	1,77E-02	2,47E-01
<b>Pi</b>	5,44E-04	1,34E-03	0,407	4,73E-02	6,62E-01	<b>Pi.e</b>	5,73E-04	1,40E-03	0,410	8,44E-02	1,00E+00
<b>Mf</b>	5,38E-04	1,37E-03	0,393	2,77E-02	3,88E-01	<b>Mf</b>	7,04E-04	2,32E-03	0,304	3,19E-02	4,46E-01
<b>Pi.e</b>	3,32E-04	1,17E-03	0,284	1,51E-01	1,00E+00	<b>Mf.e</b>	3,85E-04	1,72E-03	0,223	2,11E-01	1,00E+00

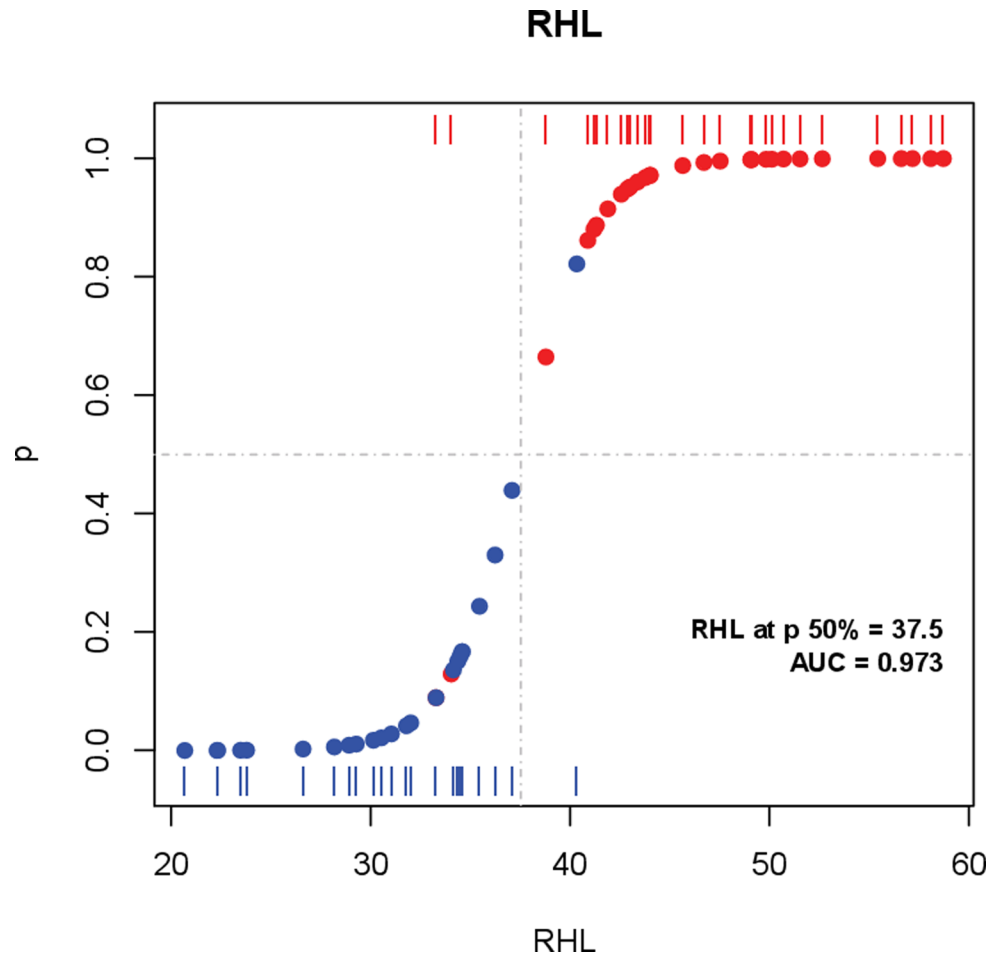
<https://doi.org/10.1371/journal.pone.0204877.t001>

comparing RHL with the number of haplotypes, Hpl. According to the main statistics of the distribution of median values obtained after 2000 simulation cycles on each fasta file, RHL is on average five times less variable than Hpl, in terms of the coefficient of variation (CV) and of interquartile dispersion (QD) (Table 3). The main statistics of the relative error of the median value of RHL in the sample replicates indicate a median error of 0.04% with a maximum of 1.99%. This confirms that the observed value of RHL in a sample is unbiased and not influenced by the sample size. No sample size correction was needed for RHL, contrary to most diversity indices [27, 29]. Thus, RHL is far more stable and robust than Hpl and other diversity indices to characterize a virus transition into error catastrophe with deep sequencing data.

## Discussion

Comparison of diversity indices for complexity evaluation of mutant spectra of HCV populations has unveiled that RHL is a reliable index to diagnose mutant spectrum expansions associated with a mutagenic treatment. Previous studies with HCV quantified average 5.1-fold (range 3.8–6.6) increases of mutation frequency following 200 serial large population passages in Huh-7.5 cells, and 3.5-fold (range 1.6–5.6) increases as a result of up to 5 serial passages in the presence of favipiravir or ribavirin [19, 32, 34]. Despite comparable or even larger mutation frequency increases associated with multiple passages compared with mutagenic treatments, RHL stood as a reliable, robust and unbiased marker for lethal mutagenesis. RHL is less influenced by standard serial passages.

A salient informative role of RHL can be interpreted in the light of current evidence of the molecular mechanisms that underlie the transition of RNA viruses towards an extinction threshold. One event is suppression of viable genome replication by defective genomes that are produced as a result of mutagenesis [36–38]. This is an extension to mutagenized populations of the capacity of mutant spectra to suppress replication of high fitness cognate populations [39]. Mutant-dependent interference was formulated as the lethal defection model of virus extinction [40] that has as one of its consequences that during mutagenesis infectivity is lost earlier than the capacity of viral RNA to replicate, thus leading to decrease of specific infectivity



**Fig 3. Logistic regression plot of mutagenicity over RHL.** Red bars at the top depict the values of RHL for HCV samples subject to mutagenesis. Blue bars at the bottom depict RHL values for control samples. Corresponding blue and red dots on the fitted logistic curve show the predicted probability of mutagenesis for each sample, with a predicted 50% probability at a RHL value of 37.5%. Area under the ROC curve 0.973. Only two samples under mutagenic drug treatment and one control sample were mistakenly classified.

<https://doi.org/10.1371/journal.pone.0204877.g003>

[19, 20, 34, 41, 42]. Examination of individual biological clones of viruses that remain viable amidst a mutagenic treatment evidenced 200-fold reduction in infectivity, with 8-fold increase in mutation frequency [43]. Therefore, the high RHL value in mutagenized populations is likely to reflect a fundamental property of mutant spectra subjected to continuous mutagenesis in which many low fitness genomes are generated. Such genomes, because of the continuous input of new mutations, do not have the opportunity of fitness recovery thus replenishing a low fitness sub-swarm captured by the RHL value. An increase in the proportion of minority (low frequency) mutations has been observed in lethal mutagenesis experiments both in cell culture [19, 20] and *in vivo* [8]. In mutant spectrum expansions that occur under basal mutation rate, RHL is expected to be less abundant because no enhanced mutagenesis jeopardizes opportunities for fitness gain, a tendency documented for RNA viruses when allowed unrestricted replication in a constant environment [31, 32, 44, 45].

Regarding diversity index adequacy to characterize lethal mutagenesis, RHL is followed by the highly correlated incidence-based indices RHL, Hpl, nMuts, PolySites and FAD, the latter probably because its entity level quality prevails under the conditions of our study. In contrast,



**Table 2. Results of the logistic regressions.** Dev: residual deviance. AIC: Aikake Information Content. Sensit: Sensitivity. Specif: Specificity. AUC: area under the ROC curve. Err: classification error rate. LOOCV: Leave-one-out cross-validation error rate.

Index	Dev	AIC	Sensit	Specif	AUC	Err	TenFoldCV	LOOCV
RHL	19,6	23,6	0,933	0,958	0,973	0,056	0,057	0,055
Hpl	24,2	28,2	0,867	0,917	0,968	0,111	0,123	0,111
PolySites	44,2	48,2	0,800	0,875	0,899	0,167	0,176	0,165
nMuts	45,3	49,3	0,800	0,875	0,897	0,167	0,176	0,172
FAD	36,5	40,5	0,867	0,792	0,928	0,167	0,182	0,181
Shannon	49,7	53,7	0,833	0,750	0,863	0,204	0,208	0,204
q1D	50,4	54,4	0,800	0,792	0,863	0,204	0,215	0,221
q2D	62,5	66,5	0,633	0,833	0,738	0,278	0,290	0,278
qInfD	62,8	66,8	0,633	0,833	0,724	0,278	0,298	0,312
GiniS	62,4	66,4	0,767	0,542	0,738	0,333	0,345	0,367
Pi.e	74,1	78,1	1,000	0,000	0,557	0,444	0,485	0,465
Pi	72,3	76,3	0,700	0,292	0,597	0,482	0,539	0,516
Mf	71,4	75,4	0,633	0,417	0,589	0,463	0,520	0,529
Mf.e	73,4	77,4	0,933	0,208	0,562	0,389	0,494	0,571

<https://doi.org/10.1371/journal.pone.0204877.t002>

Mf and Pi, despite being widely used in the description of mutant spectra, exhibit poor correlation with mutagenesis treatment. We also examined by logistic regression the capacity of each index to discriminate between a history of mutagenesis and non-mutagenesis accompanying a mutant spectrum expansion. Sorting of indices by LOOCV error rate placed RHL on top, followed by Hpl, PolySites, nMuts, FAD and Shannon. No discriminating capacity is observed for Mf, Pi, Mf.e and Pi.e, in agreement with the poor results of these indices in the association tests. The performed logistic regression has aimed at a more complex scenario, recognizing a mutagenic state independently of population history, where the signal could be blurred and affected by different phases of quasispecies dynamics, either of expansion or contraction of its mutant spectrum.

It could be anticipated that multivariate models such as logistic PCLR and PLSLR might describe the mutagenic effects more accurately than individual indices by adding the contributed predictive capacity of different indices despite its high correlation, in the sense that they could have a higher incidence with samples under mutagenic effect. But no logistic multivariate model beats RHL as a single predictor.

The information we seek to be captured with RHL lies below technical noise. Our approach has consisted in supposing that technical noise affects equally all samples in the experiment

**Table 3. Results of the in-silico study to characterize the biasedness and robustness of RHL with respect to Hpla.** The table provides a summary of distributional values for the % error with respect to the true value of RHL (top row); the coefficient of variation (CV) observed for RHL and Hpl, rows second and third; the ratio of both CVs in row fourth; the interquartile deviation (QD) observed for RHL and Hpl, rows fifth and sixth; and the ratio of both QDs in the last row.

	Min.	Q1	Median	Mean	Q3	Max.
RHL.err %	0,00	0,02	0,04	0,25	0,32	1,99
RHL.cv	0,0048	0,0077	0,0098	0,0109	0,0134	0,0211
Hpl.cv	0,0279	0,0399	0,0443	0,0505	0,0548	0,1086
CV.ratio	2,1	3,2	4,7	5,3	6,9	13,1
RHL.qd	0,0057	0,0097	0,0120	0,0147	0,0184	0,0378
Hpl.qd	0,0000	0,0488	0,0606	0,0637	0,0741	0,1538
QD.ratio	0,0	2,9	5,1	5,2	7,0	13,7

<https://doi.org/10.1371/journal.pone.0204877.t003>

and that the distinctive effect would be caused by mutagenesis; that level will include both authentic rare haplotypes and those that are introduced by technical noise.

Deep sequencing has become an important tool to analyze viral populations subjected to mutagenic treatments [8, 22, 46]. The ranking of diversity indices to best characterize mutant spectra subjected to lethal mutagenesis is relevant to a growing body of fundamental and applied studies in virology and microbial genetics in general. Multiple high and low fidelity RNA virus mutants have been characterized [47], and how such mutants modify diversity indices is an open question that may shed light on the biological consequences of altered polymerase fidelity. Also, a large RHL questions the meaning of lethality of mutations in viral and microbial populations [41, 48–51]. Specifically, it is not clear whether the genomes that contribute to the RHL are slow replicators that can participate in evolutionary events, or are dead-end products transiently kept in viral populations by complementation [41, 52]. Ranking of diversity indices may provide also relevant information on the adaptative dynamics under enhanced mutagenesis [53], or the action of mutagenic agents on plant viruses [54]. From the perspective of pharmacology, RHL offers the means to distinguish whether base or nucleoside analogues that display antiviral activities affect viral RNA replication by direct inhibition of polymerase function or by enhanced mutagenesis, thus contributing to clarify uncertainties of drug action. The mechanism of activity of nucleotide analogues has consequences for the types of drug combinations that used together or sequentially can exert a more suppressive antiviral effect [55–58]. Studies with additional viruses, fidelity mutants, and nucleotide analogues are needed to provide a clearer picture of the relevance of different diversity indices to characterize mutant spectra with alternative evolutionary histories. Cellular heterogeneity in important clinical disorders such as cancer parallels the population dynamics and the collective behavior of RNA viruses. Evaluation of a possible applicability of RHL determination as a marker in the mutagenic spectra generated during different cellular tumorigenic processes could also be considered. The present study, however, points towards RHL as a valuable marker for lethal mutagenesis of virus, and emphasizes that the choice of diversity indices to describe mutant spectra is not trivial.

## Materials and methods

### Cells, viruses, infections, and drugs

Huh-7.5 and Huh-7.5 reporter human hepatoma cell lines were grown in Dulbecco's modified Eagle's medium, and controlled as previously described [30, 32, 59, 60]. HCV p0 is the parental viral population obtained by electroporation into Huh-7.5-Lunet cells of a transcript of plasmid Jc1FLAG2(p7-nsGluc2A) (a chimera of J6 and JFH-1, genotype 2a) [33], and amplified in Huh-7.5 cells [30]. HCV p100 and HCV p200 resulted from population HCV p0 passaged 100 and 200 times, respectively, in Huh-7.5 reporter cells, as described [32]. Fitness of HCV p100 and HCV p200 relative to HCV p0 was measured in different growth-competition experiments between virus pairs. In initial determinations at a total MOI of 0.03 TCID<sub>50</sub>/cell, HCV p100 fitness was  $2.2 \pm 0.4$  that of HCV p0 [31]. Subsequent determinations gave  $1.28 \pm 0.34$  at a MOI of 0.03 TCID<sub>50</sub>/cell, and  $1.10 \pm 0.02$  TCID<sub>50</sub>/cell at a MOI of 1 TCID<sub>50</sub>/cell; the corresponding values for HCV p200 relative to HCV p0 were  $1.33 \pm 0.46$  and  $1.17 \pm 0.02$ , respectively [32]. Infectious HCV was titrated as previously described [32]: serially diluted samples were applied to Huh-7.5 cells in 96-well plates (6,400 cells/well seeded 16 h earlier), and three days post-infection, cells were washed with PBS, fixed with ice-cold methanol, and stained to detect anti-NS5A monoclonal antibody 9E10 [61]. Titrations were performed in triplicate, and titers expressed as TCID<sub>50</sub>/ml. Favipiravir (T-705) (Atomax Chemicals Co. Ltd) and ribavirin

(Sigma) were prepared and used as previously described [19, 34, 35]. Their concentrations were chosen to produce comparable inhibition of HCV p0 progeny production.

### RNA extraction, cDNA amplification and deep sequencing

Total intracellular viral RNA was extracted from infected cells using the Qiagen RNeasy kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. RT-PCR was carried out using AccuScript (Agilent Technologies), with specific oligonucleotide primers (S1 Table) The amplicons covered the following genomic regions: A1, spanning genomic residues 7626 to 7962; A2, residues 7941 to 8257; and A3, residues 8229 to 8653. Negative controls without template RNA were included in parallel to ascertain the absence of cross-contamination by template nucleic acids. PCR products were purified (QIAquick Gel Extraction kit), quantified (Pico Green assay), and analyzed for quality (Bioanalyzer) prior to Illumina MiSeq sequencing.

### Experimental design

The experiment is described schematically in Fig 1A and 1B, as described above. Four factors have been considered. (i) Amplicon, with three levels A1, A2 and A3. Different regions in the ORF are submitted to different functional restrictions. (ii) Base-line passage, with analyses at passes 0, 100 and 200, where starts quasispecies evolution in absence or presence of treatment. (iii) Treatment, with three levels: no drug, favipiravir and ribavirin. (iv) Treatment passages, with analyses at passes 3 and 10. (Fig 1).

### Bioinformatics and statistics

All computations were done in the R environment and language (Team R 2017).

### Fastq data treatment

The fastq files obtained from the MiSeq were subjected to the following treatment. A haplotype-centric data analysis pipeline was developed on targeted samples by amplicons following described procedures [62, 63] adapted to the Illumina MiSeq platform in a paired-end 2x300 mode. It involved the following steps:

- Quality control of fastq files by inspecting profiles of per-site quality, read length and general instrument parameters of quality.
- In paired-end experiments overlap paired reads by FLASH [64] imposing a minimum of 20bp overlapped with a maximum of 10% mismatches.
- Quality profiles of FLASH reads.
- Demultiplex reads by identifying oligonucleotide sequences within windows of expected positions in the sequenced reads.
  - By MID (10bp oligonucleotide) distinguishing samples from different patients/origins, only one mismatch is allowed.
  - By specific primer (20-30bp oligonucleotides) distinguishing different regions in the genome, and the two strands, up to three mismatches are allowed.
  - Trim MID and primers.
  - As a result, obtain a fasta file by each combination of MID, primer and strand in the run, where the reads were collapsed to haplotypes with corresponding observed frequencies.

- Align haplotypes in each fasta file to the wild type reference sequence or the master sequence in the file (most abundant haplotype) and quality filter.
  - Discard haplotypes not covering the full amplicon.
  - Discard haplotypes with more than two indeterminations, three gaps or more than 30% differences with respect to the reference.
  - Repair accepted indeterminations and gaps as per the reference sequence.
- Intersect haplotypes in both strands with a minimum abundance of 0.1%
  - Select haplotypes in both strands with abundances not below 0.1%
  - Discard haplotypes unique to one strand.
  - Take coverage of haplotypes passing the filter as the sum of reads in both strands.
- The final haplotypes are called consensus haplotypes, and are the basis of the downstream analysis, except for the rare haplotypes load. Final yield 15–25% with respect to raw reads.

The pipeline consists in a set of R [65] scripts using objects and functions in packages Biostrings [66], ShortRead [67], and ape [68].

### Rare haplotypes load (RHL)

It is computed as the fraction of reads in the sample belonging to haplotypes common to the forward and reverse strands with abundance below a given threshold. In the present work 1%, 0.1% and 0.01% have been studied as thresholds, finally taking 1% as the most informative and reliable.

### Down sampling and fringe trimming (DSFT)

To compensate for possible biases in diversity indices due to differences in sample size [27, 29] we used down-sampling followed by fringe trimming, which consists in the following steps: (i) start with fasta files collecting the set of consensus haplotypes with abundance not below 0.1%; (ii) compute the total number of reads in each fasta file in the analysis, and take the minimum as the reference size; (iii) re-size the read number of each haplotype in each fasta file to the reference size; (iv) filter out all haplotypes below 0.2% with 95% confidence. These are the haplotypes and frequencies used in the computation of all diversity indices, except for RHL.

### Association tests

Association tests of all indices with mutagenicity were computed by the non-parametric Wilcoxon signed rank test for paired samples, comparing the diversity values of the mutagenic treatment samples versus the paired control samples, and correcting the p-values for multi-test by the Bonferroni method; we did not distinguish among amplicons. Function `wilcox.test` with `alternate` as `greater` and in paired mode, and `p.adjust` in package 'stats' were used in the computations. Rather than declaring association at any p-value threshold, the models were ranked according to descending order of the standardized effect. The most associated index is considered to be that with the highest effect and still with a low p-value.

### Logistic regression

Logistic regression was used to fit a predictive model of mutagenic activity regardless of the other factors in the design. Computations were performed by function `glm` in package 'stats'.

To assess the predictive error rate of a fitted model, a sample was considered as under mutagenic effects if the fitted probability was above 0.5. The parameters used in the assessment were: residual deviance, Aikake Information Content (AIC), sensitivity, specificity, area under the ROC curve, and classification error rate. To minimize bias due to overfitting, the cross-validation error rate under 11-fold cross-validation and under leave-one-out cross validation (LOOCV) was used. The models were sorted by increasing order of LOOCV. Function `cv.glm` in package 'boot' was used for the computation of CV error rates [69].

### In silico study

An in-silico analysis was performed to assess the robustness of RHL when compared with Hpl. To this aim, a fasta file of a sample with all haplotypes common to the forward and reverse strands with no previous abundance filter was used to sample and compute its RHL. Then the following steps were repeated 2000 times: (i) take a sample of 40,000 reads from the population; (ii) compute the RHL of this sample, (iii) filter out all haplotypes with abundance below 0.1% and not common to both strands; (iv) DSFT to 20,000 reads with haplotypes not below 0.2% with 95% confidence; (v) count Hpl. In each cycle the number of final reads, haplotypes, and RHL was computed and registered. With the set of 2000 values computed for each fasta file, the mean, median, coefficient of variation (CV), and interquartile dispersion coefficient (QD) were determined. A robust index is that showing the lowest CV and QD.

**Free available R package.** An R package collecting all important functions, the package manual, and tutorial vignettes are freely available from GitHub (<https://github.com/VHIRHepatiques/QSutils>).

### Supporting information

**S1 Table. Oligonucleotides used to amplify and sequence HCV p0, HCV p100 and HCV p200 virus subjected to serial passages in the absence or presence of 400  $\mu$ M favipiravir and 100  $\mu$ M ribavirin.**

(DOCX)

**S1 Fig. Logistic regression plots of mutagenicity over each diversity index considered.** A plot for each diversity index. As in Fig 3, bars at the top and bottom depict the values of each diversity index for HCV samples subject to mutagenesis or control. Dots on the logistic curve represent the predicted probability of mutagenesis.

(PDF)

**S2 Fig. AUC and LOOCV error rates.** (top) Barplot with AUC values for each diversity index considered. (bottom) Barplot with LOOCV error rate values for the logistic regression to each single diversity index. All samples included.

(PDF)

**S3 Fig. AUC and LOOCV error rates.** (top) Barplot with AUC values for each diversity index considered. (bottom) Barplot with LOOCV error rate values for the logistic regression to each single diversity index. Samples with a control/treatment of three passes only.

(PDF)

**S4 Fig. AUC and LOOCV error rates.** (top) Barplot with AUC values for each diversity index considered. (bottom) Barplot with LOOCV error rate values for the logistic regression to each single diversity index. Samples with a control/treatment of ten passes only.

(PDF)

## Acknowledgments

We thank Dr. Charles M. Rice for the supply of plasmid Jc1FLAG2(p7-nsGluc2A) and helpful advice for the implement of HCV replicon in cell culture.

## Author Contributions

**Conceptualization:** Josep Gregori, Esteban Domingo.

**Data curation:** Josep Gregori, Mercedes Guerrero-Murillo, Esteban Domingo.

**Formal analysis:** Josep Gregori, Celia Perales.

**Investigation:** María Eugenia Soria, Isabel Gallego, Josep Quer, Celia Perales.

**Methodology:** Josep Gregori, María Eugenia Soria, Isabel Gallego, Celia Perales.

**Project administration:** Josep Quer.

**Resources:** Juan Ignacio Esteban, Josep Quer.

**Software:** Josep Gregori, Mercedes Guerrero-Murillo.

**Supervision:** Juan Ignacio Esteban, Josep Quer, Celia Perales, Esteban Domingo.

**Validation:** Juan Ignacio Esteban, Josep Quer, Celia Perales, Esteban Domingo.

**Visualization:** Celia Perales.

**Writing – original draft:** Josep Gregori, Celia Perales, Esteban Domingo.

**Writing – review & editing:** Isabel Gallego, Juan Ignacio Esteban, Josep Quer, Esteban Domingo.

## References

1. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, et al. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat Med.* 2000; 6(12):1375–9. <https://doi.org/10.1038/82191> PMID: 11100123
2. Eigen M, Schuster P. The hypercycle. A principle of natural self-organization. Berlin: Springer; 1979.
3. Schuster P. Quasispecies on Fitness Landscapes. *Curr Top Microbiol Immunol.* 2016; 392:61–120. [https://doi.org/10.1007/82\\_2015\\_469](https://doi.org/10.1007/82_2015_469) PMID: 26597856
4. Holland JJ, Domingo E, de la Torre JC, Steinhauer DA. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J Virol.* 1990; 64:3960–2. PMID: 1695258
5. Domingo E, Schuster P. Quasispecies: From Theory to Experimental Systems. Switzerland: Springer; 2016.
6. Mullins JI, Heath L, Hughes JP, Kicha J, Styrchak S, Wong KG, et al. Mutation of HIV-1 genomes in a clinical population treated with the mutagenic nucleoside KP1461. *PLoS One.* 2011; 6(1):e15135. <https://doi.org/10.1371/journal.pone.0015135> PMID: 21264288
7. Arias A, Thorne L, Goodfellow I. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. *eLife.* 2014; 3:e03679. <https://doi.org/10.7554/eLife.03679> PMID: 25333492
8. Guedj J, Piorkowski G, Jacquot F, Madelain V, Nguyen THT, Rodallec A, et al. Antiviral efficacy of favipiravir against Ebola virus: A translational study in cynomolgus macaques. *PLoS medicine.* 2018; 15(3): e1002535. <https://doi.org/10.1371/journal.pmed.1002535> PMID: 29584730
9. Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc Natl Acad Sci U S A.* 1999; 96(4):1492–7. PMID: 9990051
10. Sidwell OW, Simon LN, Witkowski JT, Robins RK. Antiviral activity of virazole: review and structure-activity relationships. *Prog Chemotherapy.* 1974; 2:889–903.
11. Smith RA, Kirkpatrick W. Ribavirin: a broad spectrum antiviral agent. New York: Academic Press, Inc.; 1980.



12. Graci JD, Cameron CE. Mechanisms of action of ribavirin against distinct viruses. *Rev Med Virol*. 2006; 16(1):37–48. <https://doi.org/10.1002/rmv.483> PMID: 16287208
13. Beaucourt S, Vignuzzi M. Ribavirin: a drug active against many viruses with multiple effects on virus replication and propagation. Molecular basis of ribavirin resistance. *Curr Opin Virol*. 2014; 8:10–5. <https://doi.org/10.1016/j.coviro.2014.04.011> PMID: 24846716
14. Furuta Y, Takahashi K, Kuno-Maekawa M, Sangawa H, Uehara S, Kozaki K, et al. Mechanism of action of T-705 against influenza virus. *Antimicrob Agents Chemother*. 2005; 49(3):981–6. <https://doi.org/10.1128/AAC.49.3.981-986.2005> PMID: 15728892
15. Furuta Y, Komeno T, Nakamura T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci*. 2017; 93(7):449–63. <https://doi.org/10.2183/pjab.93.027> PMID: 28769016
16. Delang L, Abdelnabi R, Neyts J. Favipiravir as a potential countermeasure against neglected and emerging RNA viruses. *Antiviral Res*. 2018; 153:85–94. <https://doi.org/10.1016/j.antiviral.2018.03.003> PMID: 29524445
17. Delang L, Segura Guerrero N, Tas A, Querat G, Pastorino B, Froeyen M, et al. Mutations in the chikungunya virus non-structural proteins cause resistance to favipiravir (T-705), a broad-spectrum antiviral. *J Antimicrob Chemother*. 2014; 69(10):2770–84. <https://doi.org/10.1093/jac/dku209> PMID: 24951535
18. Baranovich T, Wong SS, Armstrong J, Marjuki H, Webby RJ, Webster RG, et al. T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses In Vitro. *J Virol*. 2013; 87(7):3741–51. <https://doi.org/10.1128/JVI.02346-12> PMID: 23325689
19. de Avila AI, Gallego I, Soria ME, Gregori J, Quer J, Esteban JI, et al. Lethal Mutagenesis of Hepatitis C Virus Induced by Favipiravir. *PLoS One*. 2016; 11(10):e0164691. <https://doi.org/10.1371/journal.pone.0164691> PMID: 27755573
20. de Avila AI, Moreno E, Perales C, Domingo E. Favipiravir can evoke lethal mutagenesis and extinction of foot-and-mouth disease virus. *Virus Res*. 2017; 233:105–12. <https://doi.org/10.1016/j.virusres.2017.03.014> PMID: 28322918
21. Escribano-Romero E, Jimenez de Oya N, Domingo E, Saiz JC. Extinction of West Nile Virus by Favipiravir through Lethal Mutagenesis. *Antimicrob Agents Chemother*. 2017; 61(11).
22. Qiu L, Patterson SE, Bonnac LF, Geraghty RJ. Nucleobases and corresponding nucleosides display potent antiviral activities against dengue virus possibly through viral lethal mutagenesis. *PLoS neglected tropical diseases*. 2018; 12(4):e0006421. <https://doi.org/10.1371/journal.pntd.0006421> PMID: 29672522
23. Abdelnabi R, Morais ATS, Leyssen P, Imbert I, Beaucourt S, Blanc H, et al. Understanding the Mechanism of the Broad-Spectrum Antiviral Activity of Favipiravir (T-705): Key Role of the F1 Motif of the Viral Polymerase. *J Virol*. 2017; 91(12).
24. Jin Z, Smith LK, Rajwanshi VK, Kim B, Deval J. The ambiguous base-pairing and high substrate efficiency of T-705 (Favipiravir) Ribofuranosyl 5'-triphosphate towards influenza A virus polymerase. *PLoS One*. 2013; 8(7):e68347. <https://doi.org/10.1371/journal.pone.0068347> PMID: 23874596
25. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev*. 2012; 76(2):159–216. <https://doi.org/10.1128/MMBR.05023-11> PMID: 22688811
26. Perales C, Domingo E. Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. *Curr Top Microbiol Immunol*. 2016; 392:323–39. [https://doi.org/10.1007/82\\_2015\\_459](https://doi.org/10.1007/82_2015_459) PMID: 26294225
27. Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, Domingo E. Viral quasispecies complexity measures. *Virology*. 2016; 493:227–37. <https://doi.org/10.1016/j.virol.2016.03.017> PMID: 27060566
28. Seifert D, Beerenwinkel N. Estimating Fitness of Viral Quasispecies from Next-Generation Sequencing Data. *Curr Top Microbiol Immunol*. 2016; 392:181–200. [https://doi.org/10.1007/82\\_2015\\_462](https://doi.org/10.1007/82_2015_462) PMID: 26318139
29. Gregori J, Salicru M, Domingo E, Sanchez A, Esteban JI, Rodriguez-Frias F, et al. Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics*. 2014; 30(8):1104–11. <https://doi.org/10.1093/bioinformatics/btt768> PMID: 24389655
30. Perales C, Beach NM, Gallego I, Soria ME, Quer J, Esteban JI, et al. 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(13):7593–607. <https://doi.org/10.1128/JVI.02824-12> PMID: 23637397
31. Sheldon J, Beach NM, Moreno E, Gallego I, Pineiro D, Martinez-Salas E, et al. Increased replicative fitness can lead to decreased drug sensitivity of hepatitis C virus. *J Virol*. 2014; 88(20):12098–111. <https://doi.org/10.1128/JVI.01860-14> PMID: 25122776
32. Moreno E, Gallego I, Gregori J, Lucia-Sanz A, Soria ME, Castro V, et al. Internal Disequilibria and Phenotypic Diversification during Replication of Hepatitis C Virus in a Noncoevolving Cellular Environment. *J Virol*. 2017; 91(10):e02505–16. <https://doi.org/10.1128/JVI.02505-16> PMID: 28275194

33. Marukian S, Jones CT, Andrus L, Evans MJ, Ritola KD, Charles ED, et al. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology*. 2008; 48(6):1843–50. <https://doi.org/10.1002/hep.22550> PMID: 19003912
34. Ortega-Prieto AM, Sheldon J, Grande-Perez A, Tejero H, Gregori J, Quer J, et al. Extinction of hepatitis C virus by ribavirin in hepatoma cells involves lethal mutagenesis. *PLoS ONE*. 2013; 8(8):e71039. <https://doi.org/10.1371/journal.pone.0071039> PMID: 23976977
35. Gallego I, Sheldon J, Moreno E, Gregori J, Quer J, Esteban JI, et al. Barrier-Independent, Fitness-Associated Differences in Sofosbuvir Efficacy against Hepatitis C Virus. *Antimicrob Agents Chemother*. 2016; 60(6):3786–93. <https://doi.org/10.1128/AAC.00581-16> PMID: 27067341
36. González-López C, Arias A, Pariente N, Gómez-Mariano G, Domingo E. Preextinction viral RNA can interfere with infectivity. *J Virol*. 2004; 78(7):3319–24. <https://doi.org/10.1128/JVI.78.7.3319-3324.2004> PMID: 15016853
37. Crowder S, Kirkegaard K. Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nat Genet*. 2005; 37(7):701–9. <https://doi.org/10.1038/ng1583> PMID: 15965477
38. Perales C, Mateo R, Mateo MG, Domingo E. Insights into RNA virus mutant spectrum and lethal mutagenesis events: replicative interference and complementation by multiple point mutants. *J Mol Biol*. 2007; 369(4):985–1000. <https://doi.org/10.1016/j.jmb.2007.03.074> PMID: 17481660
39. de la Torre JC, Holland JJ. RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J Virol*. 1990; 64(12):6278–81. PMID: 2173792
40. Grande-Pérez A, Lázaro E, Lowenstein P, Domingo E, Manrubia SC. Suppression of viral infectivity through lethal defection. *Proc Natl Acad Sci USA*. 2005; 102(12):4448–52. <https://doi.org/10.1073/pnas.0408871102> PMID: 15767582
41. Domingo E, Perales C. Quasispecies and virus. *Eur Biophys J*. 2018.
42. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci USA*. 2001; 98(12):6895–900. <https://doi.org/10.1073/pnas.111085598> PMID: 11371613
43. Arias A, Isabel de Avila A, Sanz-Ramos M, Agudo R, Escarmis C, Domingo E. Molecular dissection of a viral quasispecies under mutagenic treatment: positive correlation between fitness loss and mutational load. *J Gen Virol*. 2013; 94(Pt 4):817–30. <https://doi.org/10.1099/vir.0.049171-0> PMID: 23239576
44. Grande-Pérez A, Gómez-Mariano G, Lowenstein PR, Domingo E. Mutagenesis-induced, large fitness variations with an invariant arenavirus consensus genomic nucleotide sequence. *J Virol*. 2005; 79(16):10451–9. <https://doi.org/10.1128/JVI.79.16.10451-10459.2005> PMID: 16051837
45. Novella IS, Duarte EA, Elena SF, Moya A, Domingo E, Holland JJ. Exponential increases of RNA virus fitness during large population transmissions. *Proc Natl Acad Sci USA*. 1995; 92(13):5841–4. PMID: 7597039
46. Rawson JM, Landman SR, Reilly CS, Bonnac L, Patterson SE, Mansky LM. Lack of mutational hot spots during decitabine-mediated HIV-1 mutagenesis. *Antimicrob Agents Chemother*. 2015; 59(11):6834–43. <https://doi.org/10.1128/AAC.01644-15> PMID: 26282416
47. Borderia AV, Rozen-Gagnon K, Vignuzzi M. Fidelity Variants and RNA Quasispecies. *Curr Top Microbiol Immunol*. 2016; 392:303–22. [https://doi.org/10.1007/82\\_2015\\_483](https://doi.org/10.1007/82_2015_483) PMID: 26499340
48. Jee J, Rasouly A, Shamovsky I, Akivis Y, Steinman SR, Mishra B, et al. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature*. 2016; 534(7609):693–6. <https://doi.org/10.1038/nature18313> PMID: 27338792
49. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2011; 108(23):9530–5. <https://doi.org/10.1073/pnas.1105422108> PMID: 21586637
50. Kennedy SR, Schmitt MW, Fox EJ, Kohn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc*. 2014; 9(11):2586–606. <https://doi.org/10.1038/nprot.2014.170> PMID: 25299156
51. Gregory MT, Bertout JA, Ericson NG, Taylor SD, Mukherjee R, Robins HS, et al. Targeted single molecule mutation detection with massively parallel sequencing. *Nucleic Acids Res*. 2016; 44(3):e22. <https://doi.org/10.1093/nar/gkv915> PMID: 26384417
52. Agol VI, Gmyl AP. Emergency Services of Viral RNAs: Repair and Remodeling. *Microbiol Mol Biol Rev*. 2018; 82(2).
53. Arribas M, Cabanillas L, Kubota K, Lazaro E. Impact of increased mutagenesis on adaptation to high temperature in bacteriophage Qbeta. *Virology*. 2016; 497:163–70. <https://doi.org/10.1016/j.virol.2016.07.007> PMID: 27471955

54. Diaz-Martinez L, Brichette-Mieg I, Pineno-Ramos A, Dominguez-Huerta G, Grande-Perez A. Lethal mutagenesis of an RNA plant virus via lethal defection. *Sci Rep*. 2018; 8(1):1444. <https://doi.org/10.1038/s41598-018-19829-6> PMID: 29362502
55. Perales C, Agudo R, Tejero H, Manrubia SC, Domingo E. Potential benefits of sequential inhibitor-mutagen treatments of RNA virus infections. *PLoS Pathog*. 2009; 5(11):e1000658. <https://doi.org/10.1371/journal.ppat.1000658> PMID: 19911056
56. Moreno H, Grande-Pérez A, Domingo E, Martín V. Arenaviruses and lethal mutagenesis. Prospects for new ribavirin-based interventions. *Viruses*. 2012; 4(11):2786–805. <https://doi.org/10.3390/v4112786> PMID: 23202505
57. Iranzo J, Perales C, Domingo E, Manrubia SC. Tempo and mode of inhibitor-mutagen antiviral therapies: A multidisciplinary approach. *Proc Natl Acad Sci U S A*. 2011; 108(38):16008–13. <https://doi.org/10.1073/pnas.1110489108> PMID: 21911373
58. Perales C, Iranzo J, Manrubia SC, Domingo E. The impact of quasispecies dynamics on the use of therapeutics. *Trends in microbiology*. 2012; 20(12):595–603. <https://doi.org/10.1016/j.tim.2012.08.010> PMID: 22989762
59. Blight KJ, McKeating JA, Rice CM. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol*. 2002; 76(24):13001–14. <https://doi.org/10.1128/JVI.76.24.13001-13014.2002> PMID: 12438626
60. Jones CT, Catanese MT, Law LM, Khetani SR, Syder AJ, Ploss A, et al. Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nat Biotechnol*. 2010; 28(2):167–71. <https://doi.org/10.1038/nbt.1604> PMID: 20118917
61. Lindenbach BD, Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, Liu CC, et al. Complete replication of hepatitis C virus in cell culture. *Science*. 2005; 309(5734):623–6. <https://doi.org/10.1126/science.1114016> PMID: 15947137
62. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, et al. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One*. 2013; 8(12):e83361. <https://doi.org/10.1371/journal.pone.0083361> PMID: 24391758
63. Ramirez C, Gregori J, Buti M, Tabernero D, Camos S, Casillas R, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res*. 2013; 98(2):273–83. <https://doi.org/10.1016/j.antiviral.2013.03.007> PMID: 23523552
64. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27(21):2957–63. <https://doi.org/10.1093/bioinformatics/btr507> PMID: 21903629
65. Team R. R: A language and environment for statistical computing. Vienna, Austria: Available from: <https://www.r-project.org/>; 2017
66. Pages H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.38.4.
67. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*. 2009; 25(19):2607–8. <https://doi.org/10.1093/bioinformatics/btp450> PMID: 19654119
68. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004; 20(2):289–90. PMID: 14734327
69. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. 2nd ed. New York: Springer; 2009.