



**HAL**  
open science

## Natural non-homologous recombination led to the emergence of a duplicated V3-NS5A region in HCV-1b strains associated with hepatocellular carcinoma

Hélène Le Guillou-Guillemette, Adeline Pivert, Elise Bouthry, Cécile Henquell, Odile Petsaris, Alexandra Ducancelle, Pascal Veillon, Sophie Vallet, Sophie Alain, Vincent Thibault, et al.

### ► To cite this version:

Hélène Le Guillou-Guillemette, Adeline Pivert, Elise Bouthry, Cécile Henquell, Odile Petsaris, et al.. Natural non-homologous recombination led to the emergence of a duplicated V3-NS5A region in HCV-1b strains associated with hepatocellular carcinoma. PLoS ONE, 2017, 12 (4), pp.e0174651. 10.1371/journal.pone.0174651 . hal-01882310

**HAL Id: hal-01882310**

**<https://hal.science/hal-01882310>**

Submitted on 26 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

# Natural non-homologous recombination led to the emergence of a duplicated V3-NS5A region in HCV-1b strains associated with hepatocellular carcinoma

Hélène Le Guillou-Guillemette<sup>1,2\*</sup>, Adeline Pivert<sup>1,2</sup>, Elise Bouthry<sup>1,2a</sup>, Cécile Henquell<sup>3</sup>, Odile Petsaris<sup>4</sup>, Alexandra Ducancelle<sup>1,2</sup>, Pascal Veillon<sup>1,2</sup>, Sophie Vallet<sup>4</sup>, Sophie Alain<sup>5</sup>, Vincent Thibault<sup>6</sup>, Florence Abravanel<sup>7</sup>, Arielle A. Rosenberg<sup>8</sup>, Elisabeth André-Garnier<sup>9</sup>, Jean-Baptiste Bour<sup>10</sup>, Yazid Baazia<sup>11</sup>, Pascale Trimoulet<sup>12</sup>, Patrice André<sup>13</sup>, Catherine Gaudy-Graffin<sup>14</sup>, Dominique Bettinger<sup>15</sup>, Sylvie Larrat<sup>16</sup>, Anne Signori-Schmuck<sup>16</sup>, Hénia Saoudin<sup>17</sup>, Bruno Pozzetto<sup>17</sup>, Gisèle Lagathu<sup>18</sup>, Sophie Minjolle-Cha<sup>18</sup>, Françoise Stoll-Keller<sup>19</sup>, Jean-Michel Pawlotsky<sup>20</sup>, Jacques Izopet<sup>7</sup>, Christopher Payan<sup>4</sup>, Françoise Lunel-Fabiani<sup>1,2</sup>, Christophe Lemaire<sup>21</sup>



**1** Laboratoire de Virologie, CHU Angers, France, **2** HIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France, **3** Laboratoire de Virologie, CHU Clermont-Ferrand, France, **4** Département de Bactériologie-Virologie-Hygiène Hospitalière et Parasitologie-Mycologie, CHRU, LUBEM, Brest, France, **5** Laboratoire de Virologie, CHU Limoges, France, **6** Laboratoire de Virologie, CHU Pitié-Salpêtrière, Paris, France, **7** Laboratoire de Virologie, CNR VHE, Inserm U1043, CHU Purpan, Toulouse, France, **8** AP-HP, GHU Cochin, Laboratoire de Virologie, Université Paris Descartes, Paris, France, **9** Laboratoire de Virologie, CHU Hôtel Dieu, Nantes, France, **10** Laboratoire de Virologie, CHU, Dijon, France, **11** Laboratoire de Virologie, CHU Avicenne, Bobigny, France, **12** Laboratoire de Virologie, Hôpital Pellegrin Tripode, CHU Bordeaux, France, **13** Laboratoire de Virologie, Centre de Biologie Nord, Hôpital de la Croix Rousse, Lyon, France, **14** Université François Rabelais, Inserm U966, CHU Tours, France, **15** Laboratoire de Virologie, CHU Besançon, France, **16** Laboratoire de Virologie, UMI 3265 UJF-EMBL-CNRS, CHU, Unit of Virus Host Cell Interactions, Grenoble, France, **17** Laboratoire de Bactériologie-Virologie, CHU Saint-Etienne, France, **18** Laboratoire de Virologie, CHU Rennes, France, **19** Institut de Virologie, CHU Strasbourg, Inserm U748, Strasbourg, France, **20** Laboratoire de Virologie-Bactériologie, CHU Henri-Mondor, Créteil, France, **21** IRHS, Université d'Angers, SFR4207 QUASAV, Angers, France

**OPEN ACCESS**

**Citation:** Le Guillou-Guillemette H, Pivert A, Bouthry E, Henquell C, Petsaris O, Ducancelle A, et al. (2017) Natural non-homologous recombination led to the emergence of a duplicated V3-NS5A region in HCV-1b strains associated with hepatocellular carcinoma. PLoS ONE 12(4): e0174651. <https://doi.org/10.1371/journal.pone.0174651>

**Editor:** Dimitrios Paraskevis, National and Kapodistrian University of Athens, GREECE

**Received:** November 16, 2016

**Accepted:** March 13, 2017

**Published:** April 10, 2017

**Copyright:** © 2017 Le Guillou-Guillemette 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:** HCV sequences obtained in this study have been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases: accession numbers KU879336 to ku880537.

**Funding:** This work was funded by the Agence Nationale de Recherche sur le VIH et les hépatites (ANRS, AO Mars 2006). The funders had no role in

<sup>a</sup> Current address: Laboratoire Cerba, Ivry sur Seine, France  
\* [heleguilou@chu-angers.fr](mailto:heleguilou@chu-angers.fr)

## Abstract

### Background

The emergence of new strains in RNA viruses is mainly due to mutations or intra and inter-genotype homologous recombination. Non-homologous recombinations may be deleterious and are rarely detected. In previous studies, we identified HCV-1b strains bearing two tandemly repeated V3 regions in the NS5A gene without ORF disruption. This polymorphism may be associated with an unfavorable course of liver disease and possibly involved in liver carcinogenesis. Here we aimed at characterizing the origin of these mutant strains and identifying the evolutionary mechanism on which the V3 duplication relies.

### Methods

Direct sequencing of the entire NS5A and E1 genes was performed on 27 mutant strains. Quasispecies analyses in consecutive samples were also performed by cloning and

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

sequencing the NS5A gene for all mutant and wild strains. We analyzed the mutant and wild-type sequence polymorphisms using Bayesian methods to infer the evolutionary history of and the molecular mechanism leading to the duplication-like event.

## Results

Quasispecies were entirely composed of exclusively mutant or wild-type strains respectively. Mutant quasispecies were found to have been present since contamination and had persisted for at least 10 years. This V3 duplication-like event appears to have resulted from non-homologous recombination between HCV-1b wild-type strains around 100 years ago. The association between increased liver disease severity and these HCV-1b mutants may explain their persistence in chronically infected patients.

## Conclusions

These results emphasize the possible consequences of non-homologous recombination in the emergence and severity of new viral diseases.

## Introduction

Evolution by genome recombination has been widely described in numerous RNA viruses and shown to be relevant for epidemiology. To date, two models of RNA recombination have been described. The most frequently described and best-characterized model consists in a replicative mechanism with RNA template switching during viral RNA replication [1, 2]. The second model involves a non-replicative joining of RNA fragments [3]. Evolution by recombination is associated with both immune evasion and therapeutic resistance as illustrated by the emergence of drug resistant variants in HIV-infected patients receiving antiretroviral therapy [4]. Viral recombination may enhance viral pathogenicity, modify the host range or even contribute to the emergence of new viral strains, for example the Western equine encephalitis virus [5]. Also, recombination can have important consequences in immunization, for example the apparition of vaccine-derived poliovirus with regained virulence and contagiousness [6].

However, the frequency of recombination in the *Flaviviridae* family is low. Among the *Flaviviridae*, hepatitis C virus (HCV) is a major cause of liver disease. Recent estimates suggest that worldwide, at least 180 million people are positive for anti-HCV antibodies [7]. Chronically infected patients are at risk of developing liver cirrhosis and hepatocellular carcinoma, making HCV a major public health issue [8, 9]. Like other RNA viruses such as HIV, HCV is characterized by its wide genetic diversity: currently, seven genotypes and more than 67 subtypes are described [10]. Both of the recombination models discussed above have been described in HCV [11]. Also, rare, naturally recombinant strains have been identified at the inter- and intra-genotypic levels [12–15] with no change in gene length (i.e. no insertion); thus, none of these works described duplicated regions. However, HCV recombinations may be underestimated due to a lack of accurate, routine screening tools [16, 17]. A recent work using ultra-deep pyrosequencing to analyze HCV subpopulations during intra-host evolution confirmed that recombinants are rare ( $p < 0.001$ ) and non-persistent [18].

While adaptive mutation is known to provide RNA viruses with increased resistance to the immune system or therapeutics and to enhance their virulence [19–21], little is known as to what impact recombination may have in these same evolutionary mechanisms. Interestingly,

gain of virulence may occur by acquisition of new genes or portions of genes during a recombination event, and represent a selective advantage. Gene duplication provide opportunities for species diversification, since genetic redundancy permits co-evolution of the duplicated genes. This in turn permits the acquisition of new biological functions or the modification of virulence. Although viral genomic duplications have been largely described in DNA viruses [22], they appear infrequent in RNA viruses due to biological constraints [23, 24], notably as concerns genome inflation. The few partial gene duplications reported in RNA viruses were located in untranslated regions of *Flaviviridae* [25, 26] or in short intragenic regions. Large insertions or duplications had never been reported in HCV before our recent identification of a duplicated NS5A-V3 region in genotype 1b [27].

The pleiotropic nature of NS5A in the HCV viral life cycle has sparked numerous analyses. The NS5A protein may be involved in resistance to interferon-based therapy, in hepatitis chronicity and in liver carcinogenesis [28]. NS5A also appears to be involved in viral particle assembly, giving it a critical role in the viral life cycle [29].

The recent use of anti-NS5A drugs, such as daclatasvir, in combination with other antiviral therapies is bringing even more attention to the protein [30, 31]. The V3 region (NS5A<sub>2356-2379</sub>) in the NS5A domain III was identified in 1991 [32]. Thereafter, our team and others went on to demonstrate [33] and confirm [34, 35] a correlation between the level of mutation in V3 and the response to interferon therapy.

In the present work, we analyzed the evolutionary history and dynamics of a previously reported set of HCV-1b strains bearing two tandemly repeated V3 regions in the NS5A gene (labelled R1 respectively and R2 as shown in the S1 Fig). In that previous work, we had identified a wide duplication event in the NS5A protein of a HCV-1b strain in a large prospective multicenter study in France including more than 800 patients (noted NS5A-dup hereafter) and demonstrated unprecedentedly that the duplication may be associated with a higher risk of liver complications [27]. Using phylogenomic analyses on two datasets: a direct sequence set (one sequence per patient) and a quasispecies (intra-host viral population) sequence set, we tested whether if the second V3 region arose via a duplication event or by another process like non-homologous recombination. Using molecular epidemiology tools, we compared mutant and wild-type strains within and between hosts.

We show here that the extra copy of the V3 region in NS5A did not arise by simple duplication. It is more likely that a double recombination event involving another HCV-1b is responsible for the doubling of the V3 region. Unfortunately, we could not identify a strain carrying a V3 region similar to the one found in the new type of NS5A protein. Finally, the relationships between a higher mutation rate and severity in hepatitis are discussed.

## Material and methods

### Samples

Two preexisting sample sets were used in the present study:

- a set of NS5A clonal quasispecies sequences taken from 10 HCV-1b infected patients included in the NAIF study. Nine of these patients had a wild NS5A gene and one a duplicated V3 gene. Two to four sequential samples were available for eight patients [34, 35].
- a set of direct NS5A sequences taken from the 27 patients who had been identified to be infected by a V3-duplicated HCV-1b in our previous studies [35].

We furthermore expanded our analyses using our French multicenter cohort: direct NS5A and E1 gene sequences were studied in 138 HCV strains (all the strains with the V3 duplication

and a mean of ten strains taken from each center where a mutant strain had been identified; (S1 Table). A quasispecies study was also performed on 25 newly mutant strains identified with a sequential analysis in two strains (four and five sequential samples respectively). HCV sequences obtained in this study have been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases: accession numbers KU879336 to ku880537.

We also downloaded all the complete NS5A genes from all HCV genotypes available from the Los Alamos HCV Sequence Database [36].

## Ethical approval

As required by the Ethics Committee, an information letter was sent to all patients; a written or oral consent was not required as no additional serum sample was taken. Sera from consecutive patients newly diagnosed with chronic HCV-1b infection were obtained during the clinical work at the routine genotyping step in each participating virology laboratory (French academic centers: Angers; Avicenne; Besançon; Bordeaux; Brest; Clermont-Ferrand; Dijon; Grenoble; Limoges; Lyon; Nantes; Paris-Cochin; Paris-Pitié Salpêtrière; Rennes; Saint-Etienne; Strasbourg; Toulouse; Tours). The research consisted in a non-interventional study, subsequently all data were analyzed anonymously. The study was approved by the Ethics Committee of the University Hospital of Brest (Avis CPP Ouest 6–15112006). The study was also registered with and approved by the national commission for information technology and civil liberties (CNIL, 907057–09032007).

## Experimental procedures

**Direct NS5A sequences and NS5A quasispecies study.** RNA extraction, entire NS5A gene reverse transcription, amplification, cloning and sequencing are explained in detail elsewhere [35]. Briefly, HCV RNA was extracted from 200  $\mu$ L of each serum sample using the EasyMag automated extraction system (BioMérieux, Craponne, France). A full-length NS5A gene amplification was performed using outer primers E1 and E2 for the RT-PCR (1665 bp) and inner primers I3 and I41b (1344 bp) for the nested PCR. NS5A PCR products were subjected to direct sequencing using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on the automated ABI3130xl, with the primers used previously in the nested-PCR rounds. NS5A PCR products were also cloned using the p GEM-T Easy Vector, the T4 DNA ligase and JM109 High Efficiency Competent Cells (Promega) following the manufacturer's instructions. Clones of interest, i.e., bearing the NS5A gene, were then sequenced as described above.

We sequenced a mean of 30 clones per sample and obtained 701 direct NS5A sequences and 1000 clonal NS5A sequences that were added to the previous sequence batches for analyses (37 direct NS5A sequences and 286 clonal NS5A sequences respectively).

**E1 gene amplification and direct nucleotide sequencing.** The E1 genomic segment (922 bp) was amplified with a two-step RT-PCR using an in-house protocol kindly provided by Professor C. Henquell. Briefly, an initial mixture containing 10  $\mu$ L of previously extracted HCV RNA, 0.4 pM of random primers (Life Technologies) and 0.5 mM of each dNTP (Life Technologies) was activated for 5 min at 65°C followed by 15 sec at 25°C. A second mixture containing 200U of SuperScript II Reverse Transcriptase (Invitrogen), 40U of RiboLock RNase Inhibitor (Fermentas), 1X of reverse transcriptase buffer and 8m M of DTT for a 10  $\mu$ L final volume was added to the previous reaction tube. Reverse transcription was carried out for 10 min at 25°C and 50 min at 50°C. For PCR amplification, a 10  $\mu$ L aliquot of cDNA was then added to a 50  $\mu$ L final reaction volume. The reaction mixture contained 0.4  $\mu$ M of Env-S (5'-TGG GYA ARG TCA TCG ATA CC-3') and Env-AS (5'-GGC 6GT 6CK RTT 6AT RTG CC-3', 6 = inosine),

forward and reverse primers respectively, 320 $\mu$ M of each dNTP, 1X of polymerase buffer and 1.25U of *Taq* DNA polymerase (Qiagen). PCR was then performed with a DNA denaturation for 2 min at 95°C, followed by 45 cycles of denaturation (94°C for 15 sec), annealing (49°C for 50 sec) and elongation (72°C for 50 sec) and a final elongation for 5 min at 72°C. Samples bearing a V3 insertion and a mean of ten samples originating from each laboratory (9 to 17 samples) in which an insertion was identified were studied (138 samples).

E1 PCR products were subjected to direct sequencing using the Big Dye Terminators v3.1 Cycle Sequencing kit (Life Technologies) on the automated ABI3130xl, with forward and reverse primers used in the previous two-PCR round. The E1 nucleotide sequences were assembled into a consensus sequence then aligned to the HCVJ-1b prototype sequence (SeqScape software).

**Sequences analysis.** 1286 sequences of the quasispecies NS5A gene were aligned according to their amino-acid sequence using the Muscle [37] alignment algorithm implemented in the MEGA v6.0.5 software [38]. Prior to any inference, the more likely substitution model was tested using jModelTest 2 [39]. The best model was found to be the HKY substitution model [40] and gamma heterogeneity and a proportion of invariant sites. This substitution model was employed when useful in all subsequent analyses. All genealogical and phylogenetic relationships between strains on NS5A and E1 genes were analyzed using the models implemented in the BEAST v1.8.2 software [41]. For each kind of analysis, we used empirical base frequencies, which prevent over-parameterization of the models.

Each analysis using BEASTv1.8 was run using both a strict and uncorrelated lognormal relaxed clock [42]. The number of iterations and sampling intervals were tuned for each analysis to obtain good ESS values (>200). Thus, the number of iterations varied from  $2 \times 10^7$  to  $4 \times 10^7$  and sampling intervals from 1,000 to 10,000. Estimated trees were analyzed using the TreeAnnotator software included in the BEAST package using a burn-in of 1000 and a posterior probability of 0.9. Several datasets were analyzed: intra-host analyses for NS5A carrying duplications or not, complete NS5A and E1 genes among wild and duplicated strains, NS5A V3 regions and NS5A without V3.

**Intra-host dynamics.** Clones of NS5A were sequenced in 55 quasispecies. First, we looked for occurrences of V3 region duplication to evaluate the homogeneity of NS5A types within a quasispecies. This was done because quasispecies composed only of NS5A-dup strains would favor the hypothesis of a clearly differentiated population over a recurrent mutational effect in the wild-type NS5A locus. Demographic dynamics of each quasispecies were inferred using the Bayesian Skyline plot [43] model implemented in BEAST. This model permits to plot the variation of effective size against evolutionary times given here in coalescent units (2 times the effective size  $N_e$ ). Temporal samplings of a same quasispecies were grouped, decreasing the number of wild types from 18 to 9.

**Differences in evolutionary rates between wild and NS5A-dup loci.** To test for differences in the rates of evolution between NS5A-dup and wild loci, analyses were performed under the hierarchical phylogenetic model [44] using three strains with duplicated V3 regions and seven wild-type strains. HPM incorporates fixed effects to test for differences in evolutionary dynamics across host populations in a formal statistical framework employing stochastic search variable selection and model averaging [45]. Compared to the classical strategy of independent analyses of sequences from each patient, HPM provides a better estimation of evolutionary parameters such as nucleotide substitution rates. For the ten HCV strains, a temporal sampling was available, thus allowing for the estimation of the mutation rate of the locus [46] at the quasispecies level. According to hosts, the heterochronous sampling covers periods of several months (AG5sf) to 12 years (AG2) (see S1 Table). For patient AG2, we had serological and virological follow up around the renal graft period, which allowed us to demonstrate HCV

seroconversion. Among the four temporal samples studied in the present work, we had the first positive HCV RNA just after the graft.

Statistical inferences of mutation rates and demography were performed using the extended Bayesian skyline plot (EBSP) model under the HPM framework. This model assumes population fluctuation, a common situation in viruses, especially in humans where treatments can strongly influence HCV demography.

**Origins of NS5A-dup and the duplicated V3 region.** The origins of the NS5A-dup were inferred from the analysis of their phylogenetic relationships with and divergence time from wild strains. We performed direct sequencing of the NS5A locus from 27 patients where the V3 duplication was found plus 70 patients carrying the wild type NS5A strains. We first used the complete sequences of the NS5A locus and thereafter the NS5A locus without the V3 region(s). For comparison purposes, the same analysis was performed for the E1 locus.

A second set of phylogenetic analyses were performed only on the V3 region. For the NS5A-dup locus, we compared the first copy of V3, labelled R1, with the second, labelled R2, treating them as different loci. We then tested the hypothesis of a separate origin of the extra copy of V3 in NS5A-dup, as opposed to a duplication.

The hypothesis of an origin of the supplementary V3 from outside of France was tested using 204 sequences of this locus in the Los Alamos database [36]. We performed all the analyses described in this section using the constant size coalescent model with random relaxed log-normal clock rate, which was found to better fit our data.

## Results

### Intra-host dynamics

A total of 55 quasispecies were studied. We found no patients carrying a mix of NS5A-dup and wild-type at any time, confirming the results of our recent clinical study [27] where patients were found to be infected by only one type of HCV strain, harboring NS5A-dup or not. However certain strains had two incomplete regions (RE1, LM1 and TR3), i.e., one complete V3 region and a piece of the second one, or in the case of TR3, the 5-prime part of the first region and the 3-prime part of the second.

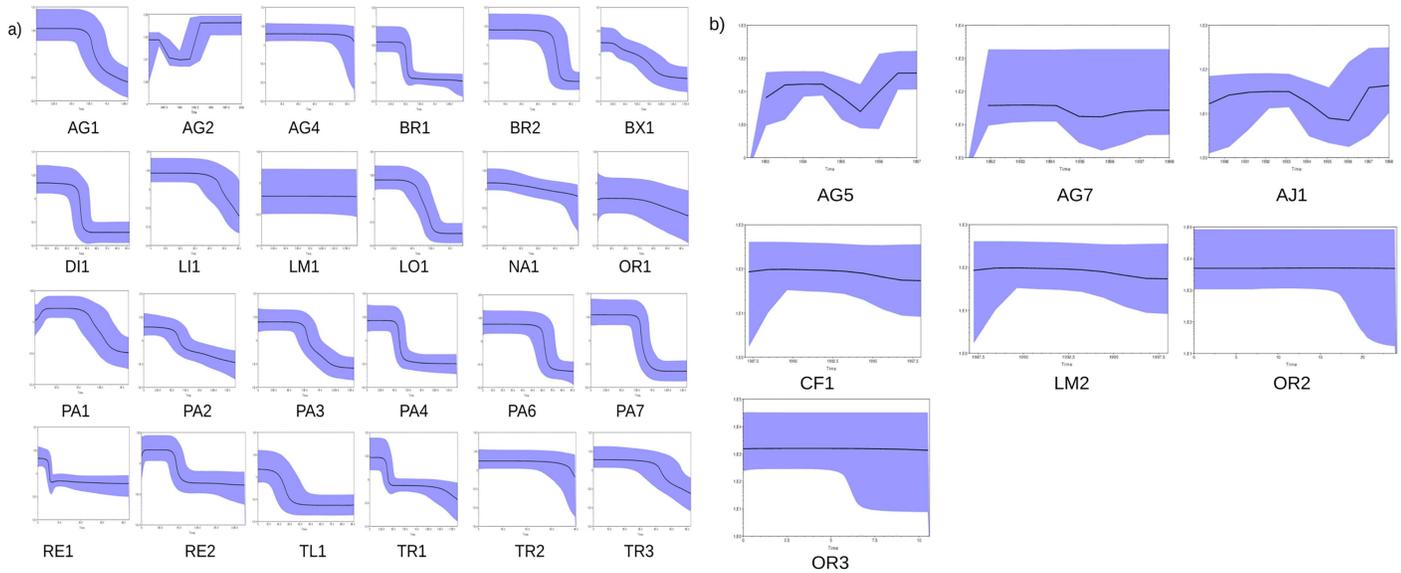
Except for strains AG3, AG6, BR3 and PA5 for which polymorphism was not sufficient to allow such inference, estimates of demographic dynamics using the Bayesian skyline model indicated that NS5A-dup strains had significantly greater demographic growth than wild ones (Mann-Whitney U,  $p = 0.001$ ) (Fig 1). Bayesian skyline plots presented different shapes among the 26 NS5a-dup and the 9 wild strains. For every strain, demographic dynamics showed a plateau effect. Clear growth dynamics were nonetheless observed for NS5A-dup strains, whereas more constant population sizes were observed for wild types.

### NS5A-dup loci evolve faster than wild ones

Hierarchical phylogenetic modelling (HPM) indicated that the NS5A-dup haplotypes had a greater mean mutation rate ( $5.74E-4 \pm 7.2E-5$ ) than the wild-type haplotypes ( $3.94E-4 \pm 8.2E-5$ ) (Mann-Whitney U,  $p < 0.01$ ) (Fig 2). Similarly, the tree heights for wild-type NS5A quasispecies were greater ( $42.71 \pm 5.4$ ) than those of NS5A-dup quasispecies ( $33.05 \pm 2.4$ ) (Mann-Whitney U,  $p < 0.01$ ).

### Origin of the extra V3 copy

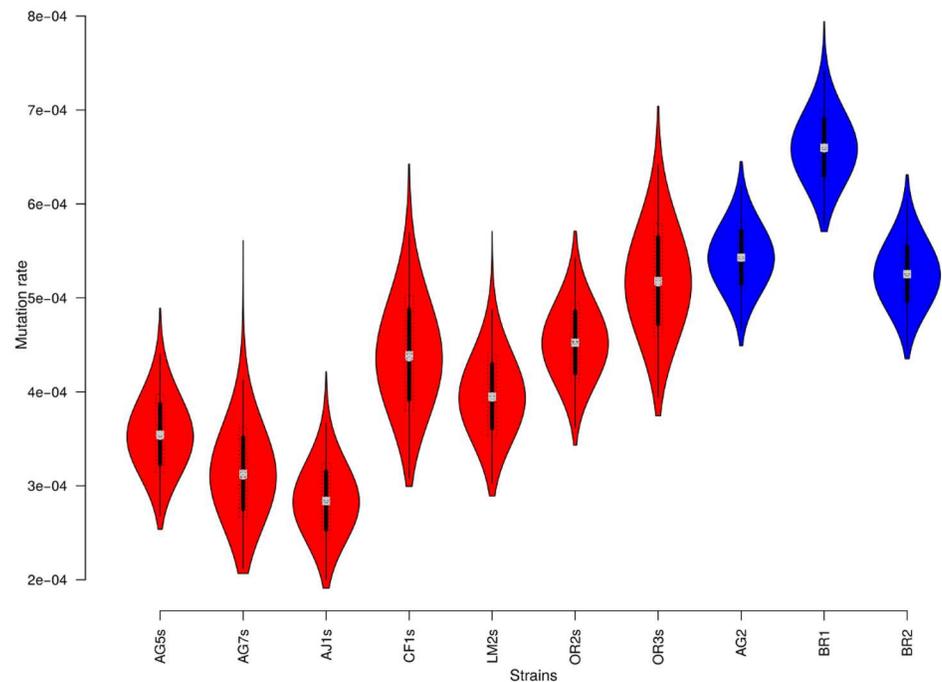
Fig 3 indicates that strains carrying NS5A-dup locus evolved separately from wild strains. To eliminate the differences induced by the presence of an extra V3 in NS5A-dup viruses, we



**Fig 1. Intra-host demographic dynamics for NS5A-dup (A) and wild (B) strains of HCV1-b derived from Bayesian Skyline Plotting at the NS5A locus.** When temporal sampling was available, time was scaled in years; otherwise time is given in coalescence time.

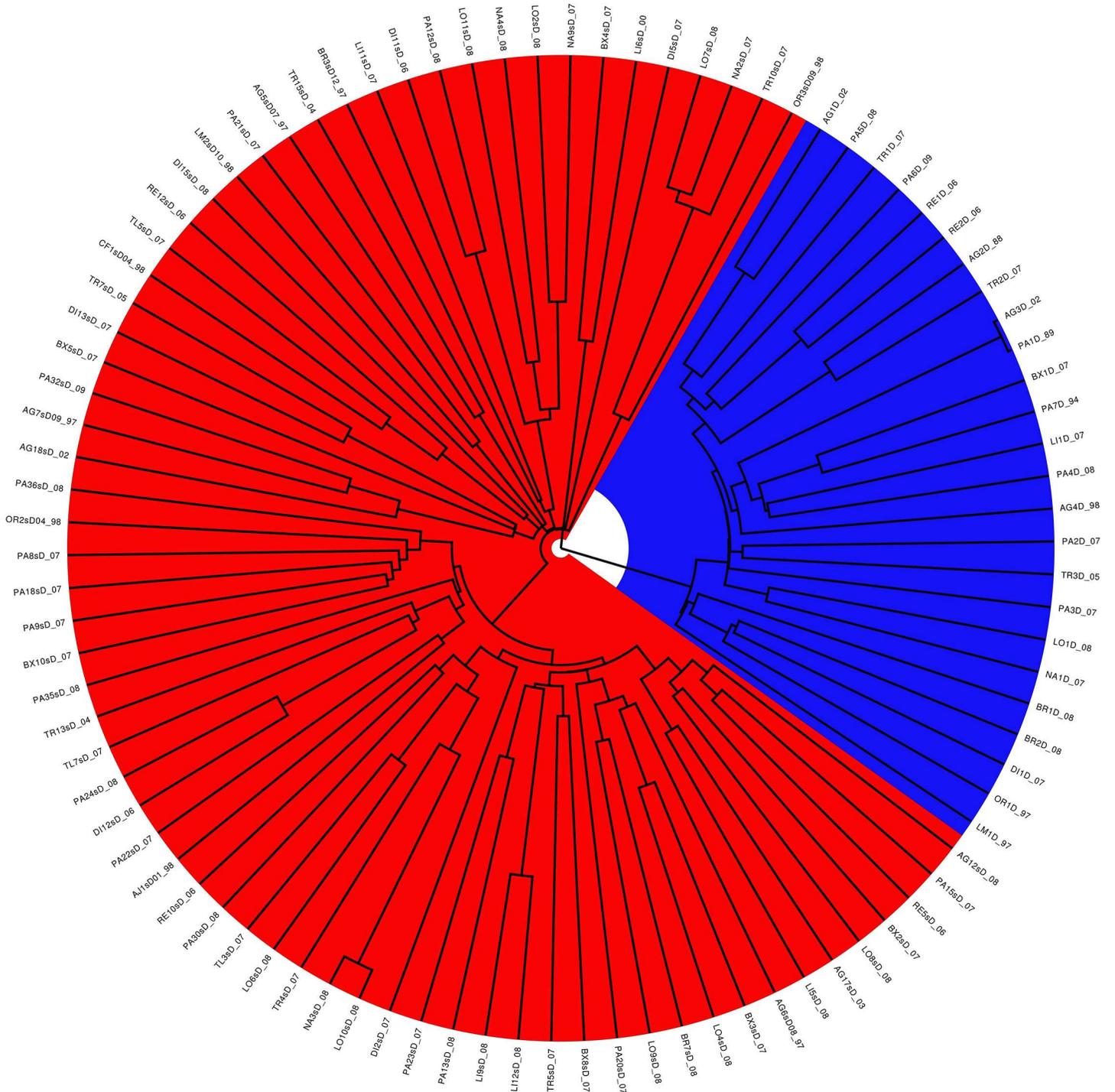
<https://doi.org/10.1371/journal.pone.0174651.g001>

performed the same analysis using sequences without the V3 region. This second analysis gave the same result (Fig 4), indicating that the clustering is genuinely due to a real divergence between strains. Strains carrying NS5A-dup appeared to have derived from the wild type in the last century. Finally, phylogeny of the E1 locus was congruent with that of the NS5A locus (S2 Fig).



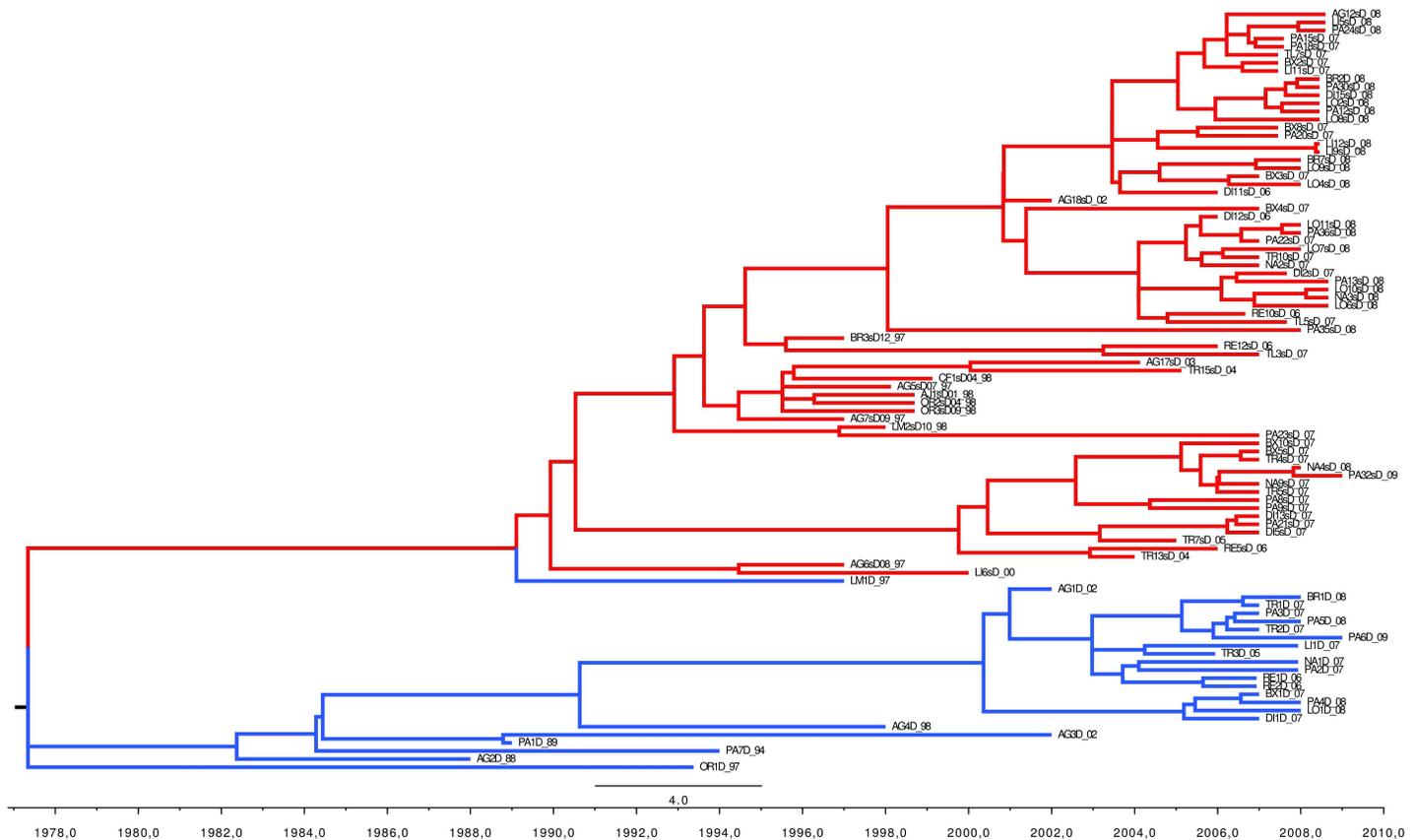
**Fig 2. Posterior distributions of mean substitution rates of NS5a for each quasispecies of HCV1-b.** Quasispecies carrying only one V3 domain (wild type) are represented in red; quasispecies carrying two V3 domains are represented in blue.

<https://doi.org/10.1371/journal.pone.0174651.g002>



**Fig 3. Divergence of the NS5A-dup strain from wild-type strains.** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with a random relaxed uncorrelated lognormal clock. NS5A-dup strains are indicated in blue. All strains carrying two V3-domains show common ancestry.

<https://doi.org/10.1371/journal.pone.0174651.g003>



**Fig 4. Divergence of NS5A-dup strains (in blue) from wild strains (in black).** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with a random relaxed uncorrelated lognormal clock. Age of nodes is mentioned on time-axis. The divergence between wild and duplicated type sequences is estimated to have occurred at the end of 70's.

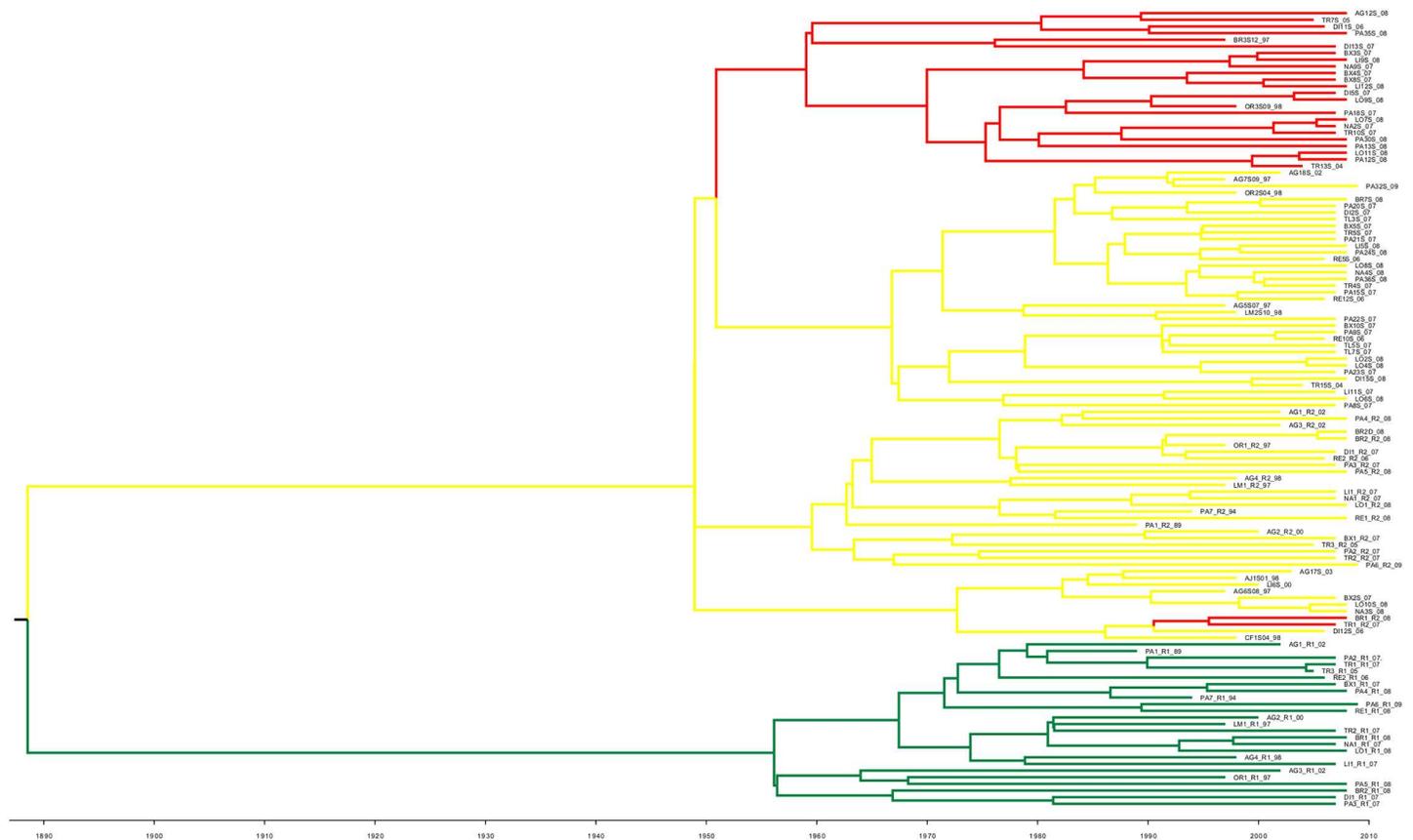
<https://doi.org/10.1371/journal.pone.0174651.g004>

We also analyzed the V3 region alone. In NS5A-dup, the first and second copies were labelled R1 and R2 respectively. Fig 5 shows that R1 appears to be the most divergent, whereas R2 keep a closer relation with wild-type NS5A. This result is at odds with the hypothesis of a recent apparition of R1 by duplication of R2. Indeed, we hypothesized that R1 had another origin. This analysis of the V3 region alone also showed that the R1 cluster likely diverged from the others about a century ago. Although R2 copies clustered with wild-type regions, they formed a distinct clade, confirming an independent evolution.

Finally, when performing the analysis with the Los Alamos HCV database, we found that R1 likely came from a HCV1-b genotype. However, no V3 region clearly clustered with the R1 copy (Fig 6). Thus, at this global scale, we confirmed the common origin of R2 copies with the wild-type V3 region.

## Discussion

Emergence of new pathogens is always of high interest, especially when it concerns human health. However, the biological mechanisms underlying emergence are poorly understood. In particular, the role of recombination in the creation of new viral populations is often underestimated as it needs co-infection events to occur. Even less well-described is the non-homologous recombination mechanism that can occur within or between genotypes of viruses. In the present work, we studied the evolutionary dynamics of new HCV1-b strains carrying two V3



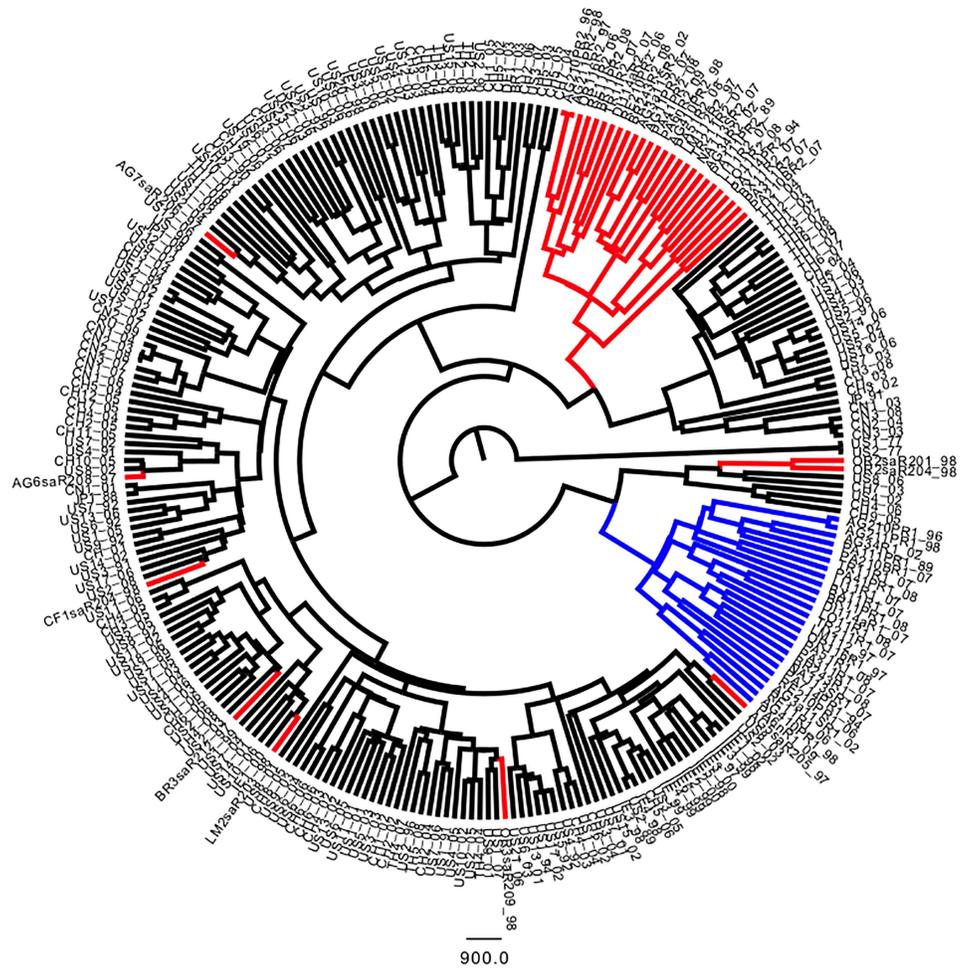
**Fig 5. The origin of the first V3 region (R1, in green) in NS5A-dup strains differs from that of the second V3 region (R2, in red), which is more closely related to the wild-type region (in yellow).** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with a random relaxed uncorrelated lognormal clock. Time is on the x-axis and represents the divergence time between clades. Divergence between R1 and R2 is estimated to have occurred around 1920.

<https://doi.org/10.1371/journal.pone.0174651.g005>

regions instead of one at the gene encoding the NS5A protein. Using clone sequences of quasispecies, direct sequencing and heterochronous sampling, we showed that the new V3 region arose from non-homologous recombination events involving an unknown HCV1-b strain rather than by duplication. The population of NS5A-dup strains then probably derived from wild-type strains about a century ago. Finally, the increased evolutionary capacity of the NS5A protein with two V3 regions may influence the severity of hepatitis C in humans, as our recent clinical work suggested [27, 47].

### Origins of the NS5A-dup strains

Our phylogenetics analyses, whatever the dataset used (on NS5A including or excluding the V3 region or on the E1 gene), all reached the same conclusion that the NS5A-dup strains diverged about a century ago from wild-type HCV1-b viruses. Also, the direct sequences of wild-type quasispecies gave greater tree heights than those of NS5A-dup quasispecies, confirming that the latter appeared more recently. Thus, NS5A-dup HCV1-b strains appear to constitute a real emergence of a new HCV1-b family. The clear-cut clustering between wild-type and NS5A-dup strains suggests a reproductive isolation between these two families. Free mixing between wild-type and NS5A-dup strains would have left traces of recombinations in the NS5A gene sequences. The likelihood of such recombination events here suggests multiple



**Fig 6. The origin of the R1 copy of the V3 region (in blue) differs from that of wild-type V3 regions (in black) at the world scale. R2 copies (in red) have a common origin with wild V3. Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with a random relaxed uncorrelated lognormal clock.**

<https://doi.org/10.1371/journal.pone.0174651.g006>

infections from different donors, allowing the mixing of several populations of strains within a patient. Our dataset did not show any evidence of such mixing of populations within any of our patients. Consequently, there is little chance for recombination between the two families of HCV1-b to occur in a short period of several decades. In the absence of recombination events between these strains, the question of reproductive isolation within HCV1-b is still pending. Is recombination lacking because of its rarity in such viruses [18, 48, 49] or is it occurring but deleterious because of a damaging phenotypic effect such as incompatibilities [50, 51]? Toward this, experimental recombination assays *in vitro* would help to shed light on the reality of reproductive isolation.

Also particularly interesting were the phylogenetics of the V3 copies themselves. We found that the first V3 copy did not cluster with the second one. In the case of duplication, a common origin for the two copies would be expected, but our results were at odds with that logic. Indeed, the R1 copy of V3 appeared to come from another population of HCV1-b, as illustrated by Figs 5 and 6, whereas the R2 copy appeared to derive from the V3 region observed in the wild-type NS5A genes. The comparison with NS5A haplotypes from the Los Alamos

database did not permit to clearly identify a potential ancestor for the R1 copy, but comparisons with other genotypes of HCV clearly showed that it did cluster with HCV1-b genotypes.

The mechanism by which the R1 was added to the 5-prime side of the original V3 region remains unclear. We hypothesize here that this extra copy may have been introduced by non-homologous recombination between two divergent HCV1-b strains. Non-homologous recombination in HCV was recently described by Scheel *et al.* [48], who demonstrated the potential for non-homologous recombination in RNA-viruses using cell cultures. To our knowledge, our present work is the first to report a viable non-homologous recombination in a RNA virus that led to the emergence of a new family of HCV1-b about a century ago. Homologous recombination, although rare in certain viruses, has been recognized to involve positive epistasis and produce recombinants with increased fitness [52]. In contrast, non-homologous recombination is often deleterious and likely to produce non-viable genotypes. In their experiment, Scheel *et al.* observed a progressive deletion of the largest extra copies of genomes (1065 nts) introduced by non-homologous recombination, with a rapid evolution toward more fit homologous recombinants with smaller 186 nts duplicated regions following passaging in cell culture [48]. We also observed the occurrence of incomplete V3 regions in three direct sequences of NS5A-dup strains and also in their respective entire quasispecies. Such partial deletions may suggest that the duplicated V3 region is slightly deleterious. Scheel *et al.* [48], in their *in vitro* experiment, showed that the elimination or excision of extra copies occurs in only a few generations. However, recombinants with smaller duplicated sequences appeared remarkably stable, similarly to our observations in the natural HCV recombinants. The same results were also obtained recently by Holmblat *et al.*, who reported that *in vitro* non-homologous recombination between defective poliovirus and coxsackievirus led to durable duplications with no alteration of neurovirulence in mice [53]. They also showed that non-homologous recombinants with large inserts (>70 codons) submitted to successive homologous recombination evolve to a more stable state, and moreover confirmed that recombinants with small duplicated regions were stable. They concluded that these recombinations may be viewed as a model of genetic plasticity for enteroviruses. We did not observe quick elimination of the extra V3 region in our study, quite the contrary they seemed to have persisted for about a century. In contrast to the work of Scheel *et al.* [48], the V3 duplication reported here has experienced “real life” in human organisms, and more recently, faced clinical treatments.

We suggest that far from being deleterious, the supplementary V3 region may provide an epidemiological advantage for HCV1-b. This hypothesis is supported by the observation of better demographic dynamics (Fig 1) and intra-host mutation rates (Fig 2) for NS5A-dup. Other examples of advantageous duplications in viruses can be found in the literature. For example, a duplication in the P1 gene of *Potyviridae*, a family of plant viruses, was reported to be the consequence of a former genomic recombination and to largely contribute to their adaptation to a wide range of host species [54]. In another report on the bovine viral diarrhea virus, another *Flaviviridae* belonging to the Pestivirus genus, a duplicated NS3 gene was associated with increased disease severity [55]. In the same way, our results showed that the introduction of an extra copy of V3 did not alter the fitness of HCV strains, and the results of our recent clinical work suggest that such strains may have more negative effects on the patients than their wild counterparts [27, 47].

## Implications for viral emergence

The evolutionary rate of the NS5A-dup gene was greater than that of the wild-type locus. HPM allowed us to obtain a good estimate of the mutation rate for each quasispecies. The capacity of

HPM to compare intra-host dynamics has been demonstrated in HIV [45]. Its results are more accurate than those obtained from working on each quasispecies independently.

The difference in evolutionary rates between NS5A-dup and wild loci had no consequence on the demographic dynamics of the different families as illustrated by the demographic curves estimated under the Bayesian Skyline Model in Fig 1. Also, our clinical study did not observe differences in viral load levels between patients infected by the NS5A-dup strain and those infected by wild strains [27].

Thus, although carrying two V3 regions may not confer a net advantage in the intra-host population growth rate, it also does not appear to have any deleterious effects. Moreover, we have detected the strain in a number of patients. Its persistence in patients infected at different times again argues in favor of the absence of a deleterious effect of the extra V3 copy. In other words, the acquisition of a supplementary V3 region should be considered as selectively neutral at least, if not positive.

In the present work, we showed for the first time the impact of non-homologous recombination on the emergence of a new family of HCV1-b strains. Recombination, although rare in viruses, could nonetheless play an important role in the potential of emergence.

However, the recombination rate in HCV may be underestimated, probably due to clinical diagnostic methods, which usually focus on a single small region rather than large genomic portions. Here we report a non-homologous recombination, an event thought to be rarer than homologous recombination. Our finding may have important consequences on HCV epidemiology. Indeed, as we demonstrated, the NS5a-dup strains may form a distinct clade, supposing a recent reproductive isolation from wild strains. This new population should have its own dynamics, and thus considering HCV1-b as a single population entity for epidemiology could be a source of bias. Toward the question as to whether NS5A-dup represents a new emergent HCV-1b cluster, it would be of use to question the occurrence of such non-homologous recombination events in other viruses for which recombination is assumed to be more likely.

## Supporting information

**S1 Fig. Alignments of two batches of direct sequences taken from wild type NS5A strains and NS5A-dup mutants strains respectively.** The first V3 domain, labelled V3-R1 (light grey square), is located between positions 2353 and 2383, the second V3 domain, labelled V3-R2 (dark grey square), is tandemly located thereafter out to position 2414. This second domain is aligned with the unique V3 domain from the wild-type NS5A strains (light brown square). Dashes (-) in the NS5A-dup sequences indicate shorter deletions in the last 6 sequences. (TIF)

**S2 Fig. Phylogeny of the E1 locus. NS5A-dup strains (in blue) are divergent from wild ones (in red).** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock. (TIF)

**S1 Table. Description of the strains of HCV1-b.** For each strain the locality and date of sampling are given. Sequences of NS5A harboring two V3 regions are indicated in bold and italics, the others are wild type. Direct sequences are given in S1A Table and strains studied for quasispecies analysis are given in S1B Table. (DOCX)

## Acknowledgments

The authors thank K. Erwin for proofreading the English manuscript.

The authors thank the AC11-HCV Study and AC33-Virology groups from the ANRS, and the following clinicians: Angers—I Fouchard-Hubert, P Calès; Avicenne—D Roulot; Besançon—V Di Martino; Bordeaux—V de Lédighen; Brest—JB Nousbaum, F Tanné; Clermont-Ferrand—A Abergel; Dijon—A Minello, P Hillon; Grenoble—JP Zarski, V Leroy, MN Hilleret; Limoges—V Loustaud-Ratti; Lyon—V Zoulim; Nantes—J Gournay; Paris-Cochin—P Podévin; Paris-Pitié Salpêtrière—P Lebray; Rennes—D Guyader; Saint-Etienne—M. Ecochard, X. Roblin; Strasbourg—M. Doffoël; Toulouse—S Métivier, K Barange; Tours—L D'Alteroche, Y Bacq.

## Author Contributions

**Conceptualization:** HLGG FLF CP CL.

**Data curation:** HLGG AP PV CL.

**Formal analysis:** HLGG CL.

**Funding acquisition:** HLGG JI CP FLF.

**Investigation:** HLGG AP EB CH OP AD PV SV SA VT FA AAR EAG JBB YB PT PA CGG DB SL ASS HS BP GL SMC FSK JMP JI CL.

**Methodology:** HLGG AP CH SV CL.

**Project administration:** HLGG FLF.

**Resources:** HLGG AP EB CH CL.

**Supervision:** HLGG JI FLF CL.

**Validation:** HLGG AP EB CL.

**Visualization:** HLGG CL.

**Writing – original draft:** HLGG CL.

**Writing – review & editing:** HLGG AD FLF CL.

## References

1. Kirkegaard K, Baltimore D. The mechanism of RNA recombination in poliovirus. *Cell*. 1986; 47(3):433–43. Epub 1986/11/07. PMID: [3021340](https://pubmed.ncbi.nlm.nih.gov/3021340/)
2. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol*. 2011; 9(8):617–26. Epub 2011/07/05. <https://doi.org/10.1038/nrmicro2614> PMID: [21725337](https://pubmed.ncbi.nlm.nih.gov/21725337/)
3. Gmyl AP, Korshenko SA, Belousov EV, Khitrina EV, Agol VI. Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? *RNA*. 2003; 9(10):1221–31. Epub 2003/09/18. <https://doi.org/10.1261/rna.5111803> PMID: [13130136](https://pubmed.ncbi.nlm.nih.gov/13130136/)
4. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol*. 2007; 81(14):7620–8. Epub 2007/05/12. <https://doi.org/10.1128/JVI.00083-07> PMID: [17494080](https://pubmed.ncbi.nlm.nih.gov/17494080/)
5. Weaver SC, Kang W, Shirako Y, Rumenapf T, Strauss EG, Strauss JH. Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol*. 1997; 71(1):613–23. Epub 1997/01/01. PMID: [8985391](https://pubmed.ncbi.nlm.nih.gov/8985391/)
6. Kew O, Morris-Glasgow V, Landaverde M, Burns C, Shaw J, Garib Z, et al. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science*. 2002; 296(5566):356–9. Epub 2002/03/16. <https://doi.org/10.1126/science.1068284> PMID: [11896235](https://pubmed.ncbi.nlm.nih.gov/11896235/)
7. Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology*. 2013; 57(4):1333–42. Epub 2012/11/23. <https://doi.org/10.1002/hep.26141> PMID: [23172780](https://pubmed.ncbi.nlm.nih.gov/23172780/)

8. Omland LH, Farkas DK, Jepsen P, Obel N, Pedersen L. Hepatitis C virus infection and risk of cancer: a population-based cohort study. *Clin Epidemiol*. 2010; 2:179–86. Epub 2010/09/25. PMID: [20865115](#)
9. Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol*. 2006; 45(4):529–38. Epub 2006/08/02. <https://doi.org/10.1016/j.jhep.2006.05.013> PMID: [16879891](#)
10. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology*. 2014; 59(1):318–27. Epub 2013/10/12. <https://doi.org/10.1002/hep.26744> PMID: [24115039](#)
11. Galli A, Bukh J. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol*. 2014; 22(6):354–64. Epub 2014/03/19. <https://doi.org/10.1016/j.tim.2014.02.005> PMID: [24636243](#)
12. Gonzalez-Candelas F, Lopez-Labrador FX, Bracho MA. Recombination in hepatitis C virus. *Viruses*. 2011; 3(10):2006–24. Epub 2011/11/10. <https://doi.org/10.3390/v3102006> PMID: [22069526](#)
13. Kalinina O, Norder H, Mukomolov S, Magnusius LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol*. 2002; 76(8):4034–43. <https://doi.org/10.1128/JVI.76.8.4034-4043.2002> PMID: [11907242](#)
14. Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, Sandres-Saune K, et al. New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol*. 2007; 81(8):4357–62. Epub 2007/02/03. <https://doi.org/10.1128/JVI.02639-06> PMID: [17267503](#)
15. Shi W, Freitas IT, Zhu C, Zheng W, Hall WW, Higgins DG. Recombination in hepatitis C virus: identification of four novel naturally occurring inter-subtype recombinants. *PLoS One*. 2012; 7(7):e41997. Epub 2012/08/23. <https://doi.org/10.1371/journal.pone.0041997> PMID: [22911872](#)
16. Mes TH, van Doornum GJ. Recombination in hepatitis C virus genotype 1 evaluated by phylogenetic and population-genetic methods. *J Gen Virol*. 2011; 92(Pt 2):279–86. Epub 2010/11/12. <https://doi.org/10.1099/vir.0.024414-0> PMID: [21068218](#)
17. Morel V, Fournier C, Francois C, Brochet E, Helle F, Duverlie G, et al. Genetic recombination of the hepatitis C virus: clinical implications. *J Viral Hepat*. 2011; 18(2):77–83. Epub 2011/01/18. <https://doi.org/10.1111/j.1365-2893.2010.01367.x> PMID: [21235686](#)
18. Palmer BA, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, Fanning LJ. Analysis of the evolution and structure of a complex intrahost viral population in chronic hepatitis C virus mapped by ultradeep pyrosequencing. *J Virol*. 2014; 88(23):13709–21. Epub 2014/09/19. <https://doi.org/10.1128/JVI.01732-14> PMID: [25231312](#)
19. Gerotto M, Dal Pero F, Loffreda S, Bianchi FB, Alberti A, Lenzi M. A 385 insertion in the hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed cryoglobulinemia type 2. *Blood*. 2001; 98(9):2657–63. Epub 2001/10/25. PMID: [11675335](#)
20. Walker PR, Ketunuti M, Choge IA, Meyers T, Gray G, Holmes EC, et al. Polymorphisms in Nef associated with different clinical outcomes in HIV type 1 subtype C-infected children. *AIDS Res Hum Retroviruses*. 2007; 23(2):204–15. Epub 2007/03/03. <https://doi.org/10.1089/aid.2006.0080> PMID: [17331028](#)
21. Wen HL, Si LY, Yuan XJ, Hao SB, Gao F, Chu FL, et al. Complete genome sequencing and analysis of six enterovirus 71 strains with different clinical phenotypes. *Virology*. 2013; 454(1):115. Epub 2013/04/13. <https://doi.org/10.1186/1743-422X-10-115> PMID: [23577963](#)
22. Shackelton LA, Holmes EC. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol*. 2004; 12(10):458–65. Epub 2004/09/24. <https://doi.org/10.1016/j.tim.2004.08.005> PMID: [15381195](#)
23. Holmes EC. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol*. 2003; 11(12):543–6. Epub 2003/12/09. PMID: [14659685](#)
24. Simon-Loriere E, Holmes EC. Gene Duplication Is Infrequent in the Recent Evolutionary History of RNA Viruses. *Mol Biol Evol*. 2013. Epub 2013/03/15.
25. Gritsun TS, Gould EA. The 3' untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. *Virology*. 2006; 354(1):217–23. Epub 2006/10/26. PMID: [17063566](#)
26. Gritsun TS, Gould EA. Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. *J Gen Virol*. 2006; 87(Pt 11):3297–305. Epub 2006/10/13. <https://doi.org/10.1099/vir.0.82235-0> PMID: [17030864](#)
27. Le Guillou-Guillemette H, Ducancelle A, Bertrais S, Lemaire C, Pivert A, Veillon P, et al. Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients. *J Clin Virol*. 2015; 69:203–9. Epub 2015/07/26. <https://doi.org/10.1016/j.jcv.2015.06.096> PMID: [26209408](#)

28. Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol.* 2004; 85 (Pt 9):2485–502. Epub 2004/08/11. <https://doi.org/10.1099/vir.0.80204-0> PMID: 15302943
29. Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, Friebe P, et al. Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog.* 2008; 4(3): e1000035. Epub 2008/03/29. <https://doi.org/10.1371/journal.ppat.1000035> PMID: 18369481
30. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell RA, et al. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature.* 2010; 465(7294):96–100. Epub 2010/04/23. <https://doi.org/10.1038/nature08960> PMID: 20410884
31. Pawlotsky JM, Aghemo A, Dusheiko G, Forns X, Puoti M, Sarrazin C, editors. *EASL Recommendations on treatment of hepatitis C, 2014.* EASL Annual Meeting; 2014; London.
32. Inchauspe G, Zebedee S, Lee DH, Sugitani M, Nasoff M, Prince AM. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proc Natl Acad Sci U S A.* 1991; 88(22):10292–6. PMID: 1658800
33. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, Lunel F, et al. Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol.* 1998; 79 (Pt 6):1373–81.
34. Veillon P, Payan C, Gaudy C, Goudeau A, Lunel F. [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris).* 2004; 52 (9):505–10.
35. Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C, Lunel F. Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. *World J Gastroenterol.* 2007; 13(8):1195–203. Epub 2007/04/25. PMID: 17451199
36. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. *Bioinformatics.* 2005; 21(3):379–84. Epub 2004/09/21. <https://doi.org/10.1093/bioinformatics/bth485> PMID: 15377502
37. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004; 5:113. Epub 2004/08/21. <https://doi.org/10.1186/1471-2105-5-113> PMID: 15318951
38. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013; 30(12):2725–9. Epub 2013/10/18. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
39. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9(8):772. Epub 2012/08/01.
40. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985; 22(2):160–74. Epub 1985/01/01. PMID: 3934395
41. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012; 29(8):1969–73. Epub 2012/03/01. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748
42. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 2006; 4(5):e88. Epub 2006/05/11. <https://doi.org/10.1371/journal.pbio.0040088> PMID: 16683862
43. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 2005; 22(5):1185–92. Epub 2005/02/11. <https://doi.org/10.1093/molbev/msi103> PMID: 15703244
44. Suchard MA, Kitchen CM, Sinsheimer JS, Weiss RE. Hierarchical phylogenetic models for analyzing multipartite sequence data. *Syst Biol.* 2003; 52(5):649–64. Epub 2003/10/08. PMID: 14530132
45. Edo-Matas D, Lemey P, Tom JA, Serna-Bolea C, van den Blink AE, van 't Wout AB, et al. Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intra-host evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. *Mol Biol Evol.* 2011; 28 (5):1605–16. Epub 2010/12/08. <https://doi.org/10.1093/molbev/msq326> PMID: 21135151
46. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. *TRENDS in Ecology and Evolution.* 2003; 18(9):481–8.
47. Petsaris O, Vallet S, Le Guillou-Guillemette H, Veillon P, Gouriou S, Barbier G, et al. Duplication of the V3 domain in hepatitis C virus (1b) NS5A protein: Clonal analysis and physicochemical properties related to hepatocellular carcinoma occurrence. *J Clin Virol.* 2016; 74:19–25. Epub 2015/12/01. <https://doi.org/10.1016/j.jcv.2015.11.011> PMID: 26655076
48. Scheel TK, Galli A, Li YP, Mikkelsen LS, Gottwein JM, Bukh J. Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. *PLoS Pathog.* 2013; 9(3):e1003228. Epub 2013/04/05. <https://doi.org/10.1371/journal.ppat.1003228> PMID: 23555245

49. Simmonds P. Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. *J Virol.* 2006; 80(22):11124–40. Epub 2006/09/08. <https://doi.org/10.1128/JVI.01076-06> PMID: 16956935
50. Carrasco P, de la Iglesia F, Elena SF. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. *J Virol.* 2007; 81(23):12979–84. Epub 2007/09/28. <https://doi.org/10.1128/JVI.00524-07> PMID: 17898073
51. Sanjuan R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A.* 2004; 101(22):8396–401. Epub 2004/05/26. <https://doi.org/10.1073/pnas.0400146101> PMID: 15159545
52. Vuillaume F, Thebaud G, Urbino C, Forfert N, Granier M, Froissart R, et al. Distribution of the phenotypic effects of random homologous recombination between two virus species. *PLoS Pathog.* 2011; 7(5):e1002028. Epub 2011/05/17. <https://doi.org/10.1371/journal.ppat.1002028> PMID: 21573141
53. Holmblat B, Jegouic S, Muslin C, Blondel B, Joffret ML, Delpyroux F. Nonhomologous recombination between defective poliovirus and coxsackievirus genomes suggests a new model of genetic plasticity for picornaviruses. *MBio.* 2014; 5(4):e01119–14. Epub 2014/08/07. <https://doi.org/10.1128/mBio.01119-14> PMID: 25096874
54. Valli A, Lopez-Moya JJ, Garcia JA. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J Gen Virol.* 2007; 88(Pt 3):1016–28. Epub 2007/02/28. <https://doi.org/10.1099/vir.0.82402-0> PMID: 17325376
55. Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, Akashi H. Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus. *J Gen Virol.* 2003; 84(Pt 2):447–52. Epub 2003/02/01. <https://doi.org/10.1099/vir.0.18773-0> PMID: 12560578