

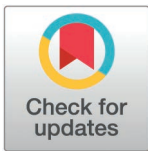
RESEARCH ARTICLE

Genetically distinct within-host subpopulations of hepatitis C virus persist after Direct-Acting Antiviral treatment failure

Lele Zhao^{1,2*}, Matthew Hall^{1,2}, Prahalad Giridhar³, Mahan Ghafari¹, Steven Kemp¹, Haiting Chai⁴, Paul Klenerman⁴, Eleanor Barnes⁴, M. Azim Ansari⁴, Katrina Lythgoe^{1,2,3*}

1 Nuffield Department of Medicine, Pandemic Sciences Institute, University of Oxford, Oxford, United Kingdom, **2** Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom, **3** Department of Biology, University of Oxford, Oxford, United Kingdom, **4** Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom

* lele.zhao@bdi.ox.ac.uk (LZ); katrina.lythgoe@biology.ox.ac.uk (KL)



OPEN ACCESS

Citation: Zhao L, Hall M, Giridhar P, Ghafari M, Kemp S, Chai H, et al. (2025) Genetically distinct within-host subpopulations of hepatitis C virus persist after Direct-Acting Antiviral treatment failure. *PLoS Pathog* 21(4): e1012959. <https://doi.org/10.1371/journal.ppat.1012959>

Editor: Adi Stern, Tel Aviv University, ISRAEL

Received: November 5, 2024

Accepted: February 5, 2025

Published: April 1, 2025

Copyright: © 2025 Zhao 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: The HCV sequences have been deposited in the GenBank database under accession numbers KY620313–KY620880. The haplotype sequences are available on GitHub (https://github.com/lzhao-virevol/hcv_lineage-awareness). The data used in this study is available to the wider scientific community for research purposes since public availability may compromise patient privacy. Access to the data is via application and

Abstract

Analysis of viral genetic data has previously revealed distinct within-host population structures in both untreated and interferon-treated chronic hepatitis C virus (HCV) infections. While multiple subpopulations persisted during the infection, each subpopulation was observed only intermittently. However, it was unknown whether similar patterns were also present after Direct-Acting Antiviral (DAA) treatment, where viral populations were often assumed to go through narrow bottlenecks. Here we tested for the maintenance of population structure after DAA treatment failure, and whether there were different evolutionary rates along distinct lineages where they were observed. We analysed whole-genome next-generation sequencing data generated from a randomised study using DAAs (the BOSON study). We focused on samples collected from patients (N=84) who did not achieve sustained virological response (i.e., treatment failure) and had sequenced virus from multiple timepoints. Given the short-read nature of the data, we used a number of methods to identify distinct within-host lineages including tracking concordance in intra-host nucleotide variant (iSNV) frequencies, applying sequenced-based and tree-based clustering algorithms to sliding windows along the genome, and haplotype reconstruction. Distinct viral subpopulations were maintained among a high proportion of individuals post DAA treatment failure. Using maximum likelihood modelling and model comparison, we found an overdispersion of viral evolutionary rates among individuals, and significant differences in evolutionary rates between lineages within individuals. These results suggest the virus is compartmentalised within individuals, with the varying evolutionary rates due to different viral replication rates and/or different selection pressures. We endorse lineage awareness in future analyses of HCV evolution and infections to avoid conflating patterns from distinct lineages, and to recognise the likely existence of unsampled subpopulations.

requests will be reviewed by the STOP-HCV Data Access Committee. For further details, please see <https://www.expmndm.ox.ac.uk/research/stop-hcv/news/data-access> or contact the Lead PI for the STOP-HCV Consortium, Professor Ellie Barnes (ellie.barnes@ndm.ox.ac.uk).

Funding: LZ and KL are supported by the Wellcome Trust and The Royal Society (107652/Z/15/Z) and the Li Ka Shing Foundation. PK is supported by a Wellcome Senior Fellowship [222426/Z/21/Z]. The research was also supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with funding from the NIHR Oxford BRC. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role in 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.

Author summary

Hepatitis C virus (HCV) remains prevalent globally, particularly straining healthcare systems in low- and middle-income countries. Despite huge efforts to develop an effective HCV vaccine, deployment is at best several years away. In the meantime, Direct-Acting Antivirals (DAAs), which can cure HCV in up to 95% of cases for some but not all HCV genotypes, are the gold-standard treatment. With the widespread use of DAAs in HCV-infected individuals, the emergence and spread of DAA resistance is imminent. HCV has been observed to maintain a structured population within hosts, where distinct subpopulations coexist and fluctuate in frequency. Using novel genomic methods, we demonstrate the continued presence of the unique within-host viral population structure of HCV after DAA treatment failure, meaning resistance-associated mutations may be present but not observed either before or after treatment failure. The distinct lineages exhibited significantly varying evolutionary rates, suggesting either differential selective pressures acting upon the subpopulations, or the infection of different cell types. In light of these findings, we advocate a lineage-aware approach that recognises the complexity of within-host dynamics of HCV since fluctuating subpopulations may harbour distinct genotypic features that could confer different phenotypes, such as drug resistance or high transmissibility. Unless the lineages are teased apart, within-host evolutionary studies risk reporting averaged and inaccurate properties of distinct sub-populations that may be evolutionarily divergent.

Introduction

Infections caused by the hepatitis C virus (HCV) are prevalent globally. While around 30% of infections clear spontaneously, the remaining become chronic if left untreated. Up to 30% of untreated chronic infections result in cirrhosis and liver cancer, leading to an increased risk of death, significantly damaging the health of the individual, and burdening health systems [1]. In recent years, a range of highly effective Direct-Acting Antivirals (DAAs) has proven effective, with efficacy rates up to 95% in some genotypes [2–5], although access to diagnosis and treatment remains low in many low and middle income countries [1]. Without an effective vaccine, reinfections are common in highly affected areas and among at-risk populations [6].

Similar to other RNA viruses, HCV exhibits complex within-host dynamics. It often maintains a unique population structure [7,8], where multiple distinct viral populations coexist but are only observed intermittently across sampling time points [9–13]. Within-host phylogenies have demonstrated the structured viral population of untreated and interferon-treated chronically-infected HCV patients [14–18]. Many biological mechanisms could support such population structure, including physical barriers such as cirrhotic liver tissues [19,20], or the infection of different cell types [21,22]. HCV also has a very low rate of recombination which acts to prevent the genetic mixing of different subpopulations [16], whilst compartmentalisation would prevent cells from being coinfecting with different subpopulations, reducing the effective rate of recombination still further. Segregating populations could lead to differences in replication rates, generation times and/or rates of evolution.

Understanding viral disease dynamics is an indispensable part of the road toward the World Health Organisation's 2030 target of eliminating viral hepatitis. The within-host population structure does not only affect drug resistance mechanisms but could also complicate studies looking at disease transmission, because genetically different populations could be circulating between the time of transmission and the time of sampling [23]. This phenomenon

will inevitably interfere with molecular epidemiology inferences and surveillance efforts, which can rely on previously curated consensus genetic data.

Studying subpopulations and compartmentalisation of HCV is not straightforward because most viral genetic data is from blood samples, so subpopulations can be mixed. With longitudinal sampling, it has been found that sometimes a single population circulates the body, while at other times, multiple distinct populations coexist in the blood [15,16,18]. While compartmentalisation has been proposed as the mechanism underlying these dynamics, it is very difficult to investigate because it requires the isolation of viruses infecting the, yet unknown, compartments. These compartments could be different sections of the same liver or various candidate cell types spread across the body.

Without coordinated proactive measures, the high success rates of DAAs may be heavily challenged by emerging viral drug resistance in the future. Moderate to high prevalence of baseline resistance-associated substitutions (RASs) to DAAs have been found in publicly available sequences and treatment studies, occasionally coupled with lowered sustained virological response (SVR) rates [24–30]. Studies have found that the genetic barriers to DAAs are low across viral genotypes [29,31]. RASs that conferred cross-resistance to multiple DAAs sharing the same target were readily selected *in vitro* [32]. In addition, the detected RASs were naturally inherent to specific viral genotypes [26,29,33]. Different genotypes and demographics vary in their overall SVR rates with DAA treatment [24,25,33,34]. Detailed characterisations of RASs, especially in light of within-host population structure, are essential for treatment guidelines in resource limited regions as it will help to avoid an increase in drug resistance prevalence and to achieve high levels of SVR post treatment.

An important observation has been the maintenance of viral subpopulations after liver transplantation and treatment with ribavirin and interferon, with distinct subpopulations present prior to transplantation or treatment re-emerging months or years later [14,15]. Given the rapid and mass rollout of DAAs, which are highly effective at treating most HCV infections, a key question is whether distinct subpopulations are also maintained in the minority of individuals who fail treatment. To answer this question, we analysed whole-genome deep-sequencing data from samples longitudinally collected from patients that failed DAA treatment. Because of the short-read nature of the data, we used a variety of analytical methods to identify subpopulations, and if they were present to determine their rates of evolution. As a result of the findings, we suggest the within-host evolutionary analyses of HCV viral populations should be lineage-aware as standard.

Results

Subpopulation frequencies fluctuate across timepoints

Approximately 600 chronically infected HCV patients were enrolled in the BOSON study, which assessed the efficacy of different lengths of Sofosbuvir and Ribavirin treatment with or without Peginterferon-Alfa [35]. The three treatment arms were 12 weeks (Sofosbuvir + Ribavirin + Peginterferon-Alfa), 16 weeks (Sofosbuvir + Ribavirin), and 24 weeks (Sofosbuvir + Ribavirin) long. A total of 88 individuals from all three arms failed treatment. Plasma samples were collected before treatment (Baseline), during treatment, 12 weeks and 24 weeks post treatment. For 22 individuals, there were also samples collected before the start of their 12 weeks of retreatment, and occasionally samples were available for 4 weeks, 12 weeks and 24 weeks post retreatment. Of the individuals that failed treatment, two were infected with genotype 2 HCV and the rest were all infected with genotype 3 HCV. Given the specifics of data availability for each of the patients, not all individuals were included in all analyses. For inclusion criteria see [S1 Fig](#).

Since HCV has been reported to have low rates of recombination [16], we hypothesised that if different subpopulations have genetic variants unique to them, then the dynamics of subpopulations could be observed by tracking frequency changes of intra-host Single Nucleotide Variants (iSNVs) through time. We generated frequency trajectories of genomic iSNVs that were below 50% at Baseline, above 50% at Post-Treatment 12 weeks (PT12) and below 50% at Post-Treatment 24 weeks PT24 for 9 individuals with at least 4 such iSNVs (top panel, Fig 1). No frequency threshold was applied to samples after PT24. The majority of these iSNVs were synonymous changes. We then examined the genomic location of these iSNVs (bottom panel, Fig 1). The iSNV positions across the HCV genome did not show a specific genomic region responsible for segregating the subpopulations, and also supported the lack of recombination.

Next, we generated frequency trajectories for all fluctuating iSNVs present in 50 individuals with at least three samples available (S2 Fig). We observed highly concordant iSNV frequencies through time, highly suggestive of genetically distinct subpopulations changing in frequency, with different populations represented by unique haplotypes. To explore this further, we reconstructed viral haplotypes for each sample using CliqueSNV [36], from which we generated haplotype phylogenies for each infection using IQ-TREE2 [37] [see Methods]. The iSNV trajectories were reflected in many individuals' haplotype trees (Fig 2). For example, in Fig 2, patient J, the low frequency lineage was present at around 20% at Baseline, dropped to 0% at 12 weeks post treatment (PT12) and reappeared at around 20% at 24 weeks post treatment (PT24). This low frequency lineage was represented by one baseline haplotype and one PT24 haplotype, with both being on the same branch of the haplotype tree. Similar consistencies of haplotype phylogenies with iSNV trajectories can be seen in other patients (Fig 2 patient A, S3 Fig patient B). However, the correspondence between the allele trajectories and the haplotype trees was not always clear cut, which in some cases may be due to the presence of subpopulations not observed at baseline (S3 Fig patient C).

To visualise if the subpopulations exclusively harboured RASs, we plotted the resistance-associated variant (RAV) frequency trajectories together with the iSNV frequency trajectories (S2 Fig). In most of the individuals, there was either a consistent presence (frequency above 90%) or absence (frequency below 10%) of most RAVs throughout the sampling period. A median of 24 RAVs [range 5-28, mean 23.34] were present at equal to or above 50% frequency throughout the sampling period for the 50 assessed individuals. A median of 3 RAVs [range 0-14, mean 4.22] were present below 50% frequency throughout the sampling period for the assessed individuals. For some positions in 20 individuals, we observed RAV frequency fluctuations (i.e., frequency changing from below 50% to above 50% or vice versa) suggesting complex resistance mechanisms. Specifically, NS5B: 150V, 66T, 120R, 180Q and NS3 67V fluctuated in frequency in 4 individuals, NS5B 90A fluctuated in 5 individuals. In several patients, we observed dynamics suggesting a RAV may have been under selection as a result of the treatment, with a minor subpopulation containing the RAV at baseline becoming the major subpopulation after treatment (e.g., Fig 2 patient A). However, given the highly dynamic nature of within-host HCV subpopulations and the lack of recombination, we cannot say definitively whether the subpopulation became dominant because it harboured the RAV and these observations could also be explained by neutral processes, including drift facilitated by low population size or genetic hitchhiking.

Pretreatment population structure was retained post treatment

Although the iSNV frequency trajectories and haplotype phylogenies were suggestive of structured within-host populations, we are essentially assuming that iSNVs of similar frequencies are linked (on the same genome) and therefore that iSNV frequencies are representative of

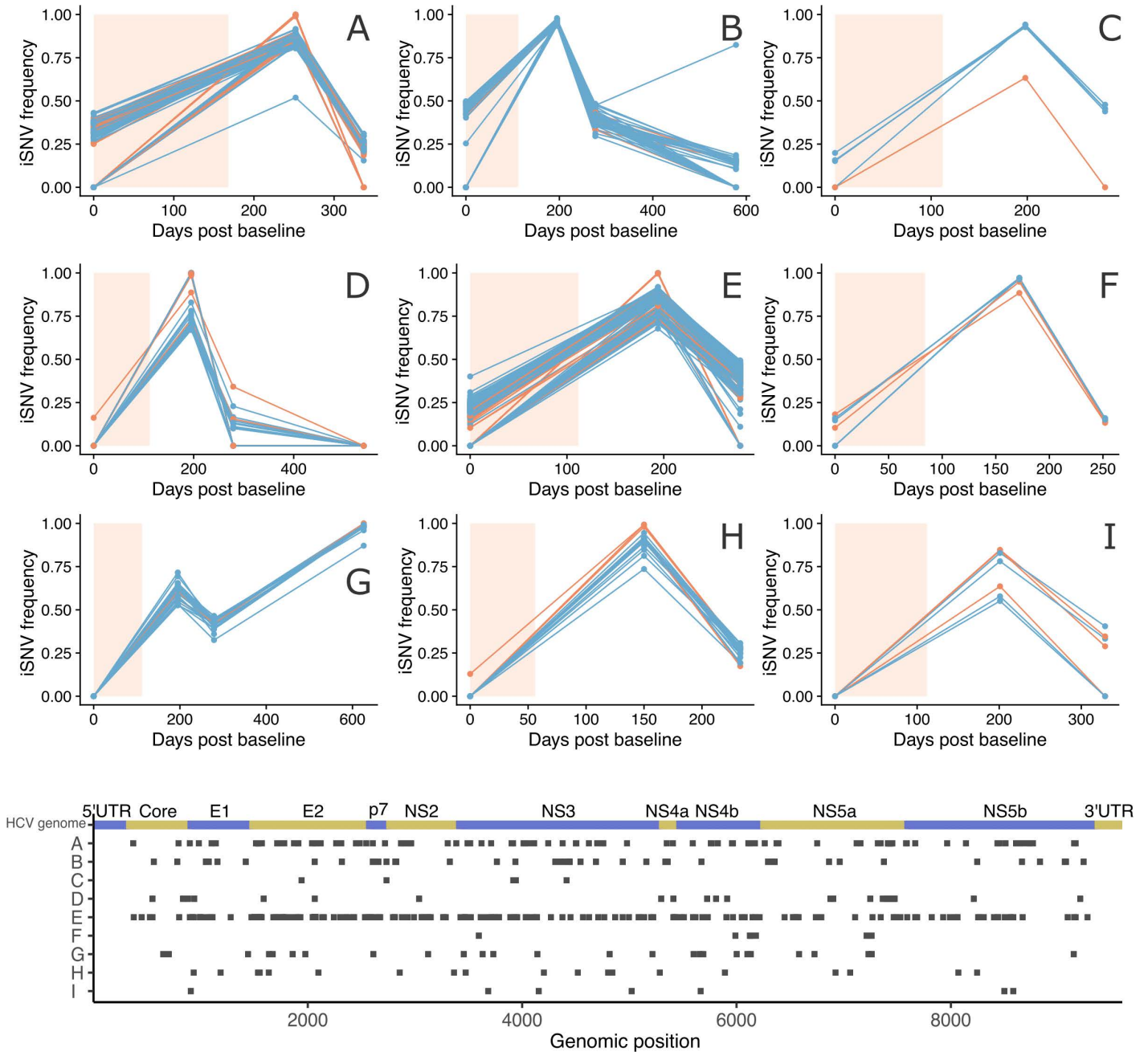


Fig 1. Frequency trajectories of iSNVs that fluctuated (Top) and their positions on the genome (Bottom) (patients A-I). The inclusion criterion was iSNVs that changed from below 50% at Baseline (i.e., 0 days post baseline), to above 50% at PT12 (12 weeks post treatment), and decreased to below 50% at PT24 (24 weeks post treatment). Patient B,D and G have a fourth sample after PT24. Synonymous changes are in blue, nonsynonymous changes are in orange and non-coding sites are in grey. The light peach colour shaded region indicates the duration of DAA treatment (variation in duration was due to the patients being in different treatment groups in the original study, see Foster et al. 2015 for further details [33]).

<https://doi.org/10.1371/journal.ppat.1012959.g001>

sub-population frequencies. However, because we are using short-read sequencing data we cannot *a priori* make this assumption since most mutations will be on different reads. Therefore, we produced the frequency trajectories for pairs of iSNVs that are closer than 150 bp on the genome (S4 Fig), further confirming that frequency concordant iSNVs were on the same genomic background.

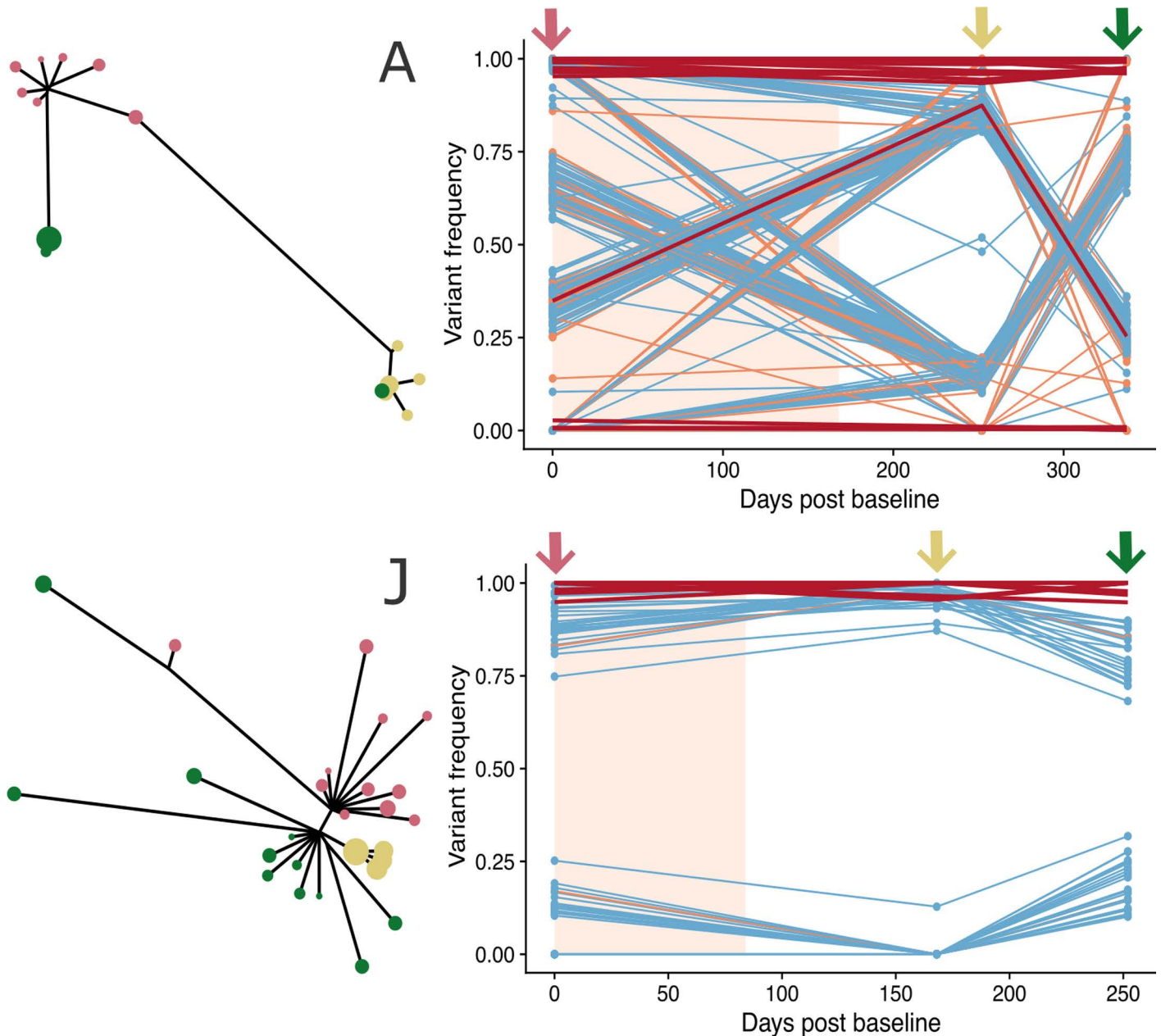


Fig 2. Haplotype phylogenies and iSNV frequency trajectories across sampling time points (patient A and J). All iSNVs with a frequency above 10% in at least one sample were traced across all sampling time points, with synonymous changes in blue, nonsynonymous changes in orange and non-coding in grey. All resistance-associated variants (RAV) trajectories are in red. The light peach colour shade indicates the duration of the DAA treatment. The coloured arrows on top of the trajectories and the coloured tips in the haplotype phylogenies represent the different sampling timepoints (light red: baseline; yellow: PT12 (12 weeks post treatment); green: PT24 (24 weeks post treatment)).

<https://doi.org/10.1371/journal.ppat.1012959.g002>

To further demonstrate the existence of subpopulations, we tested for population structure among sequencing reads spanning sliding windows along the genome, meaning linkage between iSNVs within a given window is known rather than inferred. The short reads meant low resolution in any given window, however, by scanning the entire genome and considering a large number of windows, we were able to build up an aggregate picture for each individual.

We used two cross-validating summary statistics to demonstrate the existence of within-host population structure across the entire genome of HCV.

For each window, we first used a Bayesian sequence-based approach (fastbaps) to determine if sequences showed evidence of population structure persisting through the infection. This population structure was then cross-validated using tree topology (Simmonds Association Index (SAI)). The two summary statistics are very conservative in nature, and therefore a positive result is strongly supportive of population structure, but population structure may be present even if the result is negative. Given that the inclusion of baseline sequences in all clusters was enforced during the analyses, a positive result indicates that population structure was maintained following DAA treatment. Of the 80 individuals with samples that passed the analysis selection criteria, 47 (58.75%) had five or more non-overlapping windows showing evidence of structure, and 23 (28.75%) had five or more non-overlapping windows with the structure validated by both fastbaps and SAI (Fig 3, examples of structured topologies are in S5 Fig).

Lineages that are only intermittently observed across time are less divergent

We have presented multiple lines of evidence supporting the maintenance of subpopulations after treatment in many individuals, and with the relative frequency of observed subpopulations often changing dramatically through time. Although the reconstruction of within-host haplotypes for virus populations can be challenging, the well-segregated subpopulations within HCV infected individuals appear to favour reasonably accurate haplotype reconstruction.

The CliqueSNV [36] reconstructed haplotype trees in our dataset showed different topologies across individuals. Some individuals had only a single lineage throughout the sampling time frame (e.g., Fig 4, patient K), whereas others had more complicated structures. For example, in some cases multiple lineages emerged from different baseline haplotypes (e.g., patient B in Fig 4), and often one or more of the lineages were observed at two sampling time points but missing from an intermediate time point (e.g., for patient B, lower lineage is missing time

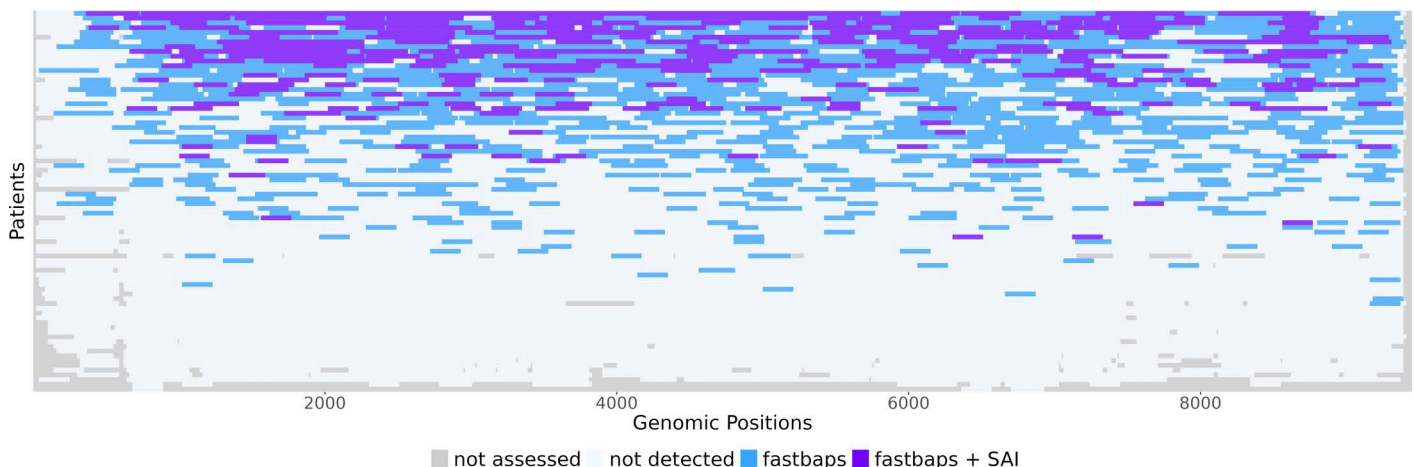


Fig 3. Ocean plot of windows with strong evidence for the maintenance of population structure after treatment. For sliding windows within longitudinally sampled individuals, the windows are coloured purple when evidence that population structure was maintained is strong (supported by fastbaps and Simmonds Association Index), are coloured blue when population structure was supported by fastbaps only, are coloured light blue when the window was assessed and are coloured grey when the window was not assessed.

<https://doi.org/10.1371/journal.ppat.1012959.g003>

point PT12). We also commonly observed cases where the order of lineage branching was not chronological (e.g., Fig 4, patient L and additional examples in S6 Fig). To further visualise the complex within-host population structure, we generated several haplotype networks using the 1-step and k-step approach from Campo et al. 2014 [7]. The networks showed stably structured connectivity among haplotypes as we increased the maximum genetic distance threshold connecting the haplotypes. These animated networks are available to view online at https://github.com/lzhao-virevol/hcv_lineage-awareness.

We hypothesised that unobserved lineages may be periodically occupying different niches, and this may in turn affect their evolutionary rates. To test if there was evolutionary rate variation across the individuals, we fitted the number of synonymous mutations between haplotypes in the individual trees to two different expected distributions (Fig 5A) using a maximum likelihood framework. We used synonymous mutations as these are more likely to be neutral than nonsynonymous mutations, and therefore less likely to be influenced by selection as a direct result of treatment. A Poisson distributed number of synonymous mutations would be expected, at any given time since baseline, if all viral populations evolve at a single shared rate.

Taking into account the time of sampling post baseline for all reconstructed haplotypes across all individuals in our dataset, we determined the most likely single mutation rate given the observed number of mutations on each haplotype compared to the most similar baseline haplotype. The expected 1000 bootstrapped distributions of mutations using the maximum likelihood rate (Poisson mean: 0.152) was significantly different from the observed distribution of synonymous mutations (Kolmogorov–Smirnov test: $p < 2.2 \times 10^{-16}$). We next introduced additional variation into the expected distribution by fitting a negative binomial distribution. (As well as its common interpretation as the distribution of the number of failures in a given number of equal-probability Bernoulli trials, the negative binomial distribution also arises as a mixture of Poisson distributions with gamma-distributed means.) The 1000 bootstrapped distributions generated with the maximum likelihood parameters (negative binomial shape parameter: 1.296, rate parameter: 0.874) were visually more similar to the observed distribution of synonymous mutations, yet the two overall distributions were still significantly different (Kolmogorov–Smirnov test: $p = 5.386 \times 10^{-6}$). These results support variation in

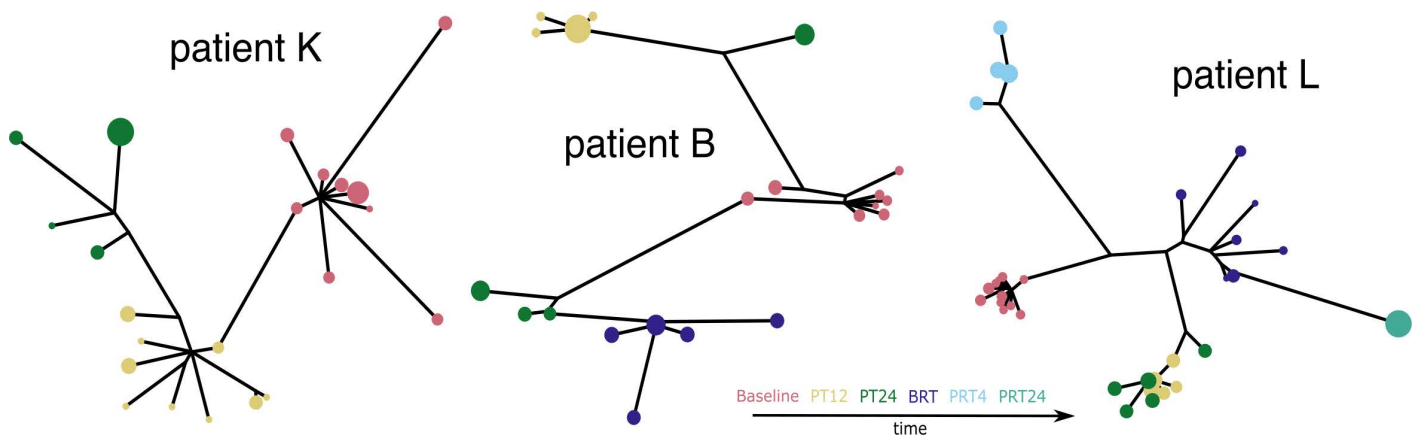


Fig 4. Reconstructed haplotype trees showed different topologies across individuals. Trees showed single subpopulation within-host (patient K), multiple subpopulations within-host (patient B), missing time point in one population (patient B: lower left lineage missing time point PT12), later sequences more similar to baseline (patient L). Baseline sequences are in red, PT12 (post treatment 12 weeks) sequences are in yellow, PT24 (post treatment 24 weeks) sequences are in green, BRT (baseline before re-treatment) sequences are in purple, PRT4 (post re-treatment 4 weeks) sequences are in light blue, PRT24 (post re-treatment 24 weeks) sequences are in light green. The size of the dots are relative haplotype frequencies.

<https://doi.org/10.1371/journal.ppat.1012959.g004>

evolutionary rates among individuals, with the right tail of the observed distribution indicating significant overdispersion among individuals, and we hypothesised this may be a consequence of complex rate dynamics within individuals as well as among.

To investigate the evolutionary rate variation within individuals, we estimated evolutionary rates per within-host lineage using exhaustive model comparison of mixed-effect linear models. For individuals where a single-lineage rate model had the lowest BIC score, one rate was estimated. For individuals where a multiple-lineage rate model was the best model, we separated the lineage rates into ones that were estimated from lineages missing haplotypes

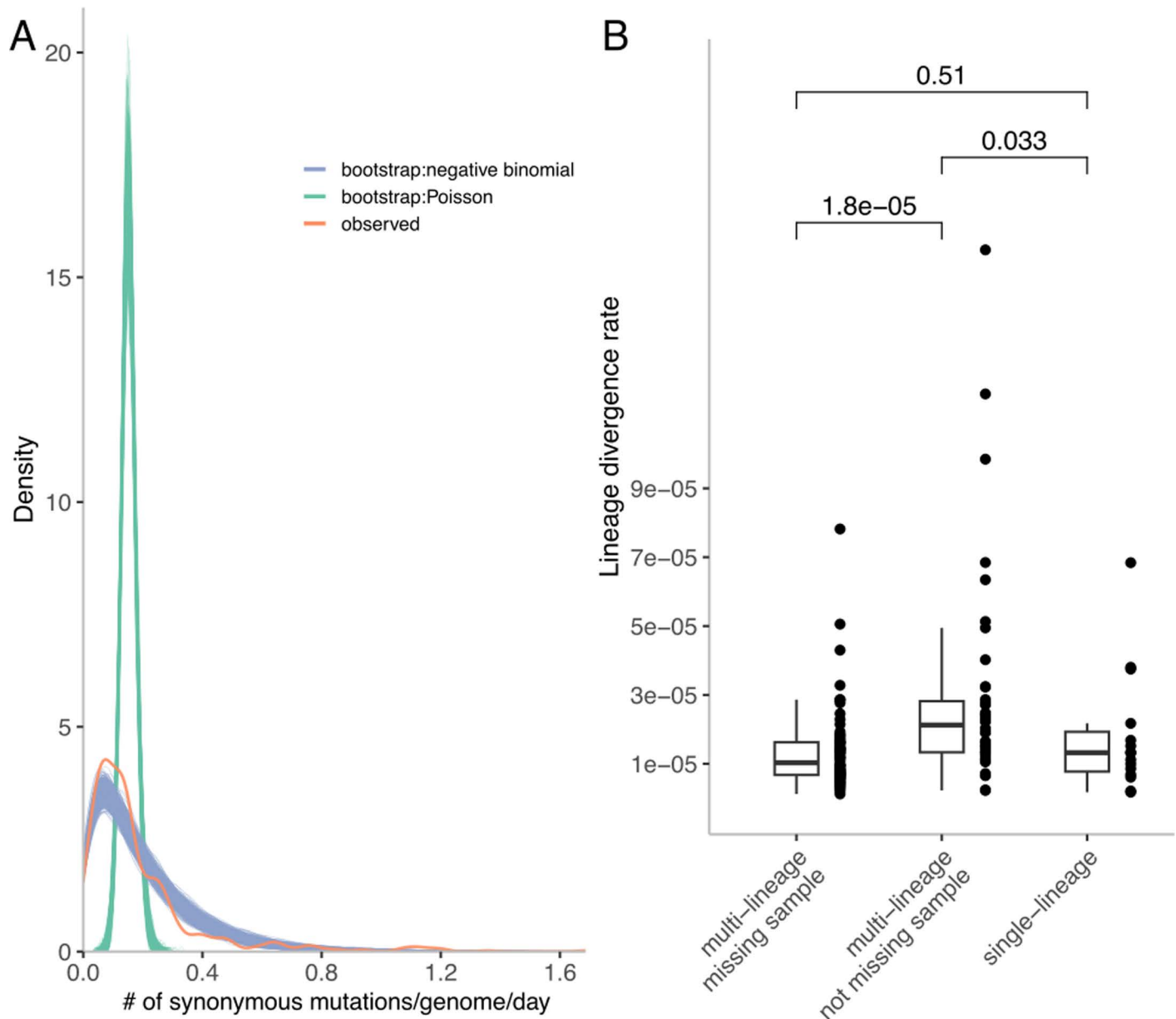


Fig 5. A. Distributions of the observed (red), 1000 Poisson (green) and 1000 negative binomial (blue) model generated expected counts of synonymous mutations per genome per day. B. Estimated rates of lineages from single lineage individuals, and multi-lineage individuals, while distinguishing between lineages where sampling time point(s) was missing or not.

<https://doi.org/10.1371/journal.ppat.1012959.g005>

from intermediate sampling timepoints and lineages that did not miss haplotypes from intermediate sampling time points. By doing so, we assumed that if the intermediate haplotype was not observed, then the subpopulation it belonged to was at low or undetectable frequency. The estimated rates for the lineages fell within the expected range for HCV neutral evolution reported by previous studies ($2.05\text{--}8.21 \times 10^{-5}$ s/s/d [16] and $1.11\text{--}1.13 \times 10^{-3}$ s/s/yr [38]), with lineages that had missing observations having rates significantly lower than the lineages not missing observations (Mann-Whitney U test: $p = 1.868 \times 10^{-5}$) (Fig 5B). These lowered divergence rates suggest that when not present in circulating blood, subpopulations may be subject to lower selection pressures, and therefore there is less opportunity for synonymous mutations to hitchhike with selected-for mutations to high frequency. Alternatively, viruses in these subpopulations might have a longer viral generation time, perhaps due to infecting different cell types. This would reduce the rate at which new mutations are introduced into the population per unit time, resulting in a lower rate of neutral evolution [39]. The rate estimates from single-lineage individuals were intermediate between the multi-lineage individual rates for lineages with no missing observations, and those with missing observations. This may be because the haplotypes were not well-resolved within these individuals and divergence signals from lineages of differing rates were merged into the single lineage.

Discussion

Here, we have systematically demonstrated for the first time that HCV pre-existing within-host population structure is often maintained after DAA treatment failure. Moreover, we found considerable variation in evolutionary rates among individuals, and between different lineages within individuals, with lineages that are absent at intermediate time points tending to have lower rates of evolution. To do this we analysed whole-genome short-read virus sequences generated as part of the BOSON study, a randomised control trial to test the effectiveness of different treatment regimens of a second-generation DAA. We used a number of corroborating analyses, including tracking iSNV frequencies through time, scanning sliding window phylogenies across the genome, and haplotype reconstruction. This work has implications not only for our understanding of HCV biology, but also for how in the future to screen for DAA tolerance or resistance mutations before treatment and after treatment failure, which we argue should be “lineage-aware”.

Previous work has also demonstrated the presence of distinct viral lineages during HCV infections, but the question remains whether their existence poses a biological challenge for treatment efforts. Even in patients who ultimately experience viral rebound, DAA treatment induces a large reduction in the HCV viral population size, but not so large that viral diversity is not maintained; for sofosbuvir treatments at least [35]. Because the resistance landscape is affected by many environmental factors such as host HLA genotype, INFL4 reactivity, DAA resistance is almost never the result of a single specific drug resistant mutation in the virus that initiated the infection [34]. The genetic resistance mechanism has proven a challenge to tease apart [40]. While the cure rate of HCV infections by DAA is high, the undisrupted population structure in patients experiencing rebound after DAA treatment is concerning.

Others have characterised the mutational spectrum of HCV post DAA treatment and the development of (multi-)drug resistance from pre- to post-DAA treatment [30,41,42]. However, the role that structured populations have in the response to DAA treatment is unknown. It is possible that the presence of population structure makes the viral population more resilient to environmental disturbances caused by immunity or treatment, meaning immune- or drug-sensitive populations could be preserved within the human body, enabling them to re-emerge when immune or treatment pressures ease. One recent study showed differential

persistence of RASs after treatment failure among different genotype infected individuals [43], where high-level resistant RASs are less likely to persist compared to low- and medium-level resistant RASs. They also observed that RASs could disappear as early as 3 months after treatment failure. If it proves to be the case that population structure facilitates DAA persistence, longer treatment duration and/or treatments that target both observed and unobserved subpopulations may be advisable. If the population structure arises due to compartmentalisation, whether it be physical barriers caused by cirrhotic liver tissues or non-hepatic cells across the body, treatment strategies may need to consider the coverage of all compartments, as another study showed different RASs are present in different compartments in the liver [19].

Consideration of maintenance of HCV within-host population structure after treatment failure may also be critical in efforts to control the virus. In particular, we have seen time and time again that we should never be complacent over the emergence and spread of drug resistance, as witnessed by recent increases of rates of transmitted resistance to antiretrovirals in HIV infections [44], after decades where these rates were consistently low and considered relatively unproblematic. For HCV, there is still a large amount that is unknown, but the answers may be critical in our fight for its global eradication, which with the advent of DAAs should be our aspiration. For example, dynamic viral population structures were not seen within all the individuals we analysed, which could be because they did not exist, or because they existed but we did not observe them [45]. Important questions remain: how widespread is within-host HCV population structure, what is the mechanism underlying the structure, why do unobserved populations have a lower rate of evolution, and what role does the viral population structure have in treatment failure and ultimately the spread of drug resistant and tolerant mutations?

It would be sensible to hypothesise that the establishment and maintenance of structured populations is complex and dependent on the interaction between host, viral, and environmental factors. Understanding the roles of these different factors could then feed into our decisions on how best to screen individuals for pre-existing virus resistance mutations, treat them, and to follow them up, so as to prevent onward transmission of resistant mutations. As a specific example, relying on the viral genome consensus at a single sampling time point for drug resistance testing risks not capturing resistant genotypes that are present but do not happen to be circulating in the blood or at sub-consensus frequency at the time of sampling. We therefore advocate for lineage awareness in evolutionary analysis of HCV infections.

Here, we analysed whole-genome short-read sequencing data. This has the advantage over previous studies looking at the within-host evolution of HCV since it covers the whole genome, rather than just the E1/E2 region that is more typically studied. However, the reliance on short Illumina reads, rather than on longer reads that are often used in this type of analysis, meant we could not directly test for population structure at each time point. We therefore used a sliding window approach that scanned across the whole genome, and tested for population structure within each window using two cross-validating statistics. This methodology ensured high specificity in the detection of population structure, but does mean there are likely many cases where population structure existed but we did not detect it with high confidence. Moreover, for part of the analysis we relied on a haplotype reconstruction tool, with the resemblance of the reconstructed haplotypes to the true haplotypes dependent on the power of the reconstruction tool and the diversity and composition of viruses within the samples. Using genuine haplotypes directly from long-read deep-sequencing data would be preferable, but while Illumina short-read sequencing is still the norm, our methods provide a roadmap for how to analyse this type of data.

Another drawback of our study is that we did not have samples pre-baseline. We therefore assumed that all subpopulations were observed at the first sampling time point. This

assumption will inevitably inflate evolutionary rate estimates for lineages that were present but not observed at baseline, but which were then observed later on, since they will have appeared to have diverged a large distance from the observed baseline population (S3 Fig, patient C). Future work may be able to resolve this limitation by incorporating the possibility of absent subpopulations at the first sampling time point.

In summary, we have demonstrated that within-host population structure is an important component of HCV within-host evolution in response to treatment, and we therefore advocate for lineage awareness in future evolutionary analysis of HCV. A better understanding of the within-host dynamics of HCV infections would benefit future treatment strategies, drug resistance outlook and transmission mitigation.

Methods

Samples, sequencing and bioinformatics

Blood samples were collected from participating patients of the completed and published BOSON clinical trial [35] and permission was granted for the use of samples for further studies. The BOSON clinical trial (ClinicalTrials.gov ID: NCT01962441) took place at 78 locations in United States, Australia, Canada, New Zealand, and United Kingdom. For the patients that did not achieve sustained viral response (SVR) during the study, samples were collected before, during and post treatment. For several individuals, baseline and post treatment samples were also available as they underwent re-treatment. The mean HCV RNA measured for the enrolled individuals were around 6.2-6.3 log₁₀ IU/mL, with a minimum titre of 4 log₁₀ IU/mL at enrolment.

RNA was extracted from 500 µl of plasma using the NucliSENS easyMAG system (bioMérieux) into 30 µl of water, of which 5 µl was processed with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) with published modifications to the manufacturer's protocol [46]. A 500-ng aliquot of the pooled library was enriched using the xGen Lockdown protocol (Rapid Protocol for DNA Probe Hybridization and Target Capture Using an Illumina TruSeq Library [v1.0]; Integrated DNA Technologies) with a comprehensive panel of HCV-specific, 120-nucleotide DNA oligonucleotide probes (IDT), designed using a published algorithm [47]. The enriched library was sequenced on the Illumina MiSeq v2 platform to produce paired 150-bp reads.

The reads were demultiplexed. Low-quality reads were trimmed using QUASR (v7.0120), and adapters were removed using Cutadapt (v1.7.1). Human reads were removed using Bowtie (v2.2.4). HCV reads were extracted by mapping to 162 ICTV reference sequences (tblastn) for *de novo* assembly with Vicuna (v1.3) and read mapping with *shiver* [48]. HCV genotypes were assigned by similarity to ICTV reference sequences. Consensus sequences were generated using *shiver* (v1.7.3) and used to construct a phylogenetic tree. Samples were removed from further analyses if they did not cluster with the other consensus sequences from the same individual. All phylogenetic tree reconstructions in this study were done using IQ-TREE2 with the GTR+F+R6 model [37]. A flowchart illustrates the sample processing and filtering for the study (S1 Fig).

Baseline iSNV and resistance-associated variant (RAV) frequency trajectories

A total of 266 samples were mapped by *shiver*, which generated base frequency files (i.e., files containing base counts per genomic position) for tracking baseline iSNV frequencies for 84 longitudinally sampled individuals. A iSNV position required a minimum depth of 100 reads and an iSNVs allele frequency of at least 10% for at least one time point, to be included in the

analysis. After quality filtering and sample availability selection, the iSNV frequency trajectories were produced for 50 individuals with three or more samples. iSNVs that were fixed (increased to and remained at above 90%) or purged (decreased to and remained at below 10%) after the first time point were not assessed. For 9 out of the 50 individuals, frequencies of pairs of iSNVs that are closer than 150 bp were also extracted from the same sequencing reads to confirm linkage between these iSNVs.

In addition, a list of 65 candidate resistance-associated variants (RAVs) was compiled from previous publications [26,49–51] including DAA resistant genotypes, DAA resistant genotypes associated with HLA and INFL4 genotypes (S1 Table). Amino acid frequencies of RAVs in individuals of this study were directly extracted from mapped reads. No minimum frequency threshold was applied for the RAVs, but frequency was only assessed when the RAVs were present on two or more mapped reads. All genomic positions are in reference to the H77 genome (GenBank accession: NC_038882).

Haplotype reconstruction

We used CliquesSNV [36] to reconstruct haplotypes for the samples in our study (command in S1 Text). One phylogenetic tree per individual was constructed using the haplotypes (aligned with MAFFT (default settings)).

Population structure in sliding windows

The bam files (N=266) of individuals with two or more longitudinal samples with HCV reads mapped were processed using phyloscanner (phyloscanner_make_trees.py) [52]. Phyloscanner generated alignments and trees for 210 bp sliding windows with 50 bp increments across the entire HCV genome. The window width was chosen through visual inspection of analysis results from a random subset of the samples to strike a balance between maximum window width (or read length) and read abundance spanning the window. Next, the k parameter in phyloscanner, determining the routines for identifying contaminant reads and multiple infections was optimised for HCV as described in S1 Text using sequencing reads from 570 patients from the BOSON study. The phyloscanner_analyse_trees.R script was used to remove possible contaminants within each sample.

After decontamination, fastbaps [53], a hierarchical bayesian clustering tool, was applied to divide the sequences into 2 clusters within each phyloscanner window (Dirichlet prior type = 'optimise.baps', number of initial clusters = 2). For each individual patient, genomic windows containing a minimum of 2 sequences from at least 2 different samples were assessed. If both clusters contained sequences from the baseline (pre-treatment) sample and sequences from one or more later (post-treatment) samples, the window was deemed to have sequence-based support for a baseline structured population being maintained throughout treatment. To validate the sequence-based support for the maintenance of population structure, Simmonds' Association Index (SAI) [54], a tree-topology-based statistic for compartmentalisation, was also calculated for the phylogeny constructed for each window. This step took the clusters assigned by fastbaps and tested how stable the clusters were by bootstrapping the phylogenetic trees. A SAI < 0.1 [55] indicates strong support for stable clustering.

Subpopulation divergence rate comparison

To estimate the neutral rate of evolution, the number of synonymous substitutions was counted for each non-baseline haplotype to its closest baseline haplotype. A maximum likelihood model with a Poisson distribution with one overall rate for all individuals in the study was fitted to the observed synonymous substitutions. To accommodate different rates of

evolution among individuals, a second maximum likelihood model with a negative binomial distribution was also fitted. Fitting to a negative binomial distribution relaxed the assumption that rates are identical amongst individuals, instead taking each to be drawn from a gamma distribution.

To determine if evolutionary rates vary among subpopulations within the same individuals, we estimated an evolutionary rate per subpopulation. For each individual's haplotype tree, an exhaustive model comparison analysis was carried out to find the best model of evolutionary rates that explained the distribution of synonymous substitutions across time. Specifically, the non-baseline haplotypes were divided into different hypothetical subpopulations if their closest baseline haplotype and sampling time were different. Linear mixed-effect models were compared for all possible combinations of the hypothetical subpopulations and the best model was determined by the lowest BIC score (see [S7 Fig](#) for example). The sets of subpopulations and their rates were extracted and categorised into individuals with only one subpopulation, individuals with multiple subpopulations where some subpopulations are missing haplotypes from intermediate time points between the first and last time point of that lineage, and subpopulations that were not missing haplotypes from intermediate time points between the first and last time point of that lineage.

Supporting information

S1 Fig. Flowchart for sample processing and quality filtering.

(DOCX)

S2 Fig. Variant frequency trajectories across sampling time points for all patients with three or more samples available. All nucleotide variants with a frequency above 10% in at least one sample were traced across all sampling time points, with synonymous changes in blue, nonsynonymous changes in orange and non-coding in grey. All relevant resistance-associated variants (RAV) trajectories are in red. The peach-coloured shade indicates the duration of the DAA treatment. All variants that became fixed (frequency increased to and remained at above 90%) or were purged (frequency decreased to and remained at below 10%) after the first sampling time point were not included. Patient IDs were assigned randomly, for example, there is no relevance between patient A and patient A2.

(DOCX)

S3 Fig. Haplotype phylogenies and iSNV frequency trajectories (patient B and C). All iSNVs with a frequency above 10% in at least one sample were traced across all sampling time points, with synonymous changes in blue, nonsynonymous changes in orange and non-coding in grey. All resistance-associated variants (RAV) trajectories are in red. The peach-coloured shade indicates the duration of the DAA treatment. The coloured arrows on top of the trajectories and the coloured tips in the haplotype phylogenies represent the different sampling time points (light red: baseline; yellow: PT12 (12 weeks post treatment); green: PT24 (24 weeks post treatment); purple: BRT (baseline before re-treatment)).

(DOCX)

S4 Fig. Frequency trajectories of pairs of variants (patient A to I). Panel G - 6584:6728 shows the trajectories of paired variants (day 0 frequency > 5%) between genomic position 6584 and 6728 for Patient G. Panel B - 1050:1076 shows the trajectories of all paired variants (day 0 frequency > 5%) between genomic position 1050 and 1076 for Patient B. Panel E - 2579:2675 shows the trajectories of paired variants (day 0 frequency > 5%) between genomic position 2579 and 2675 for Patient E.

(DOCX)

S5 Fig. Phylogenies constructed from 210 bp windows of the HCV genome sequences from four different individuals. The top two sets of phylogenies show structure maintenance while the bottom two sets do not show clear structure maintenance of the within-host viral population. The left side of each set is the molecular phylogeny, where branch length corresponds to the number of substitutions. The right side of each set is the time-calibrated phylogeny, where branch length corresponds to sampling time. The tip labels indicate the sampling time point and the number of occurrences the sequence appeared within the window. Red: Baseline (B), yellow: 12 weeks post treatment (PT12), green: 24 weeks post treatment (PT24), and blue: baseline before retreatment (BRT). (DOCX)

S6 Fig. Six example patient haplotype trees where the order of lineage branching was not chronological. Baseline sequences are in red, PT12 (post treatment 12 weeks) sequences are in yellow, PT24 (post treatment 24 weeks) sequences are in green, BRT (baseline before re-treatment) sequences are in purple, PRT4 (post re-treatment 4 weeks) sequences are in light blue, PRT24 (post re-treatment 24 weeks) sequences are in light green. The tip labels consisted of the sampling time point and the CliqueSNV-estimated frequency of the haplotype within the sample. Note the haplotype frequencies from the same sample might not sum up to 1 as the minimum frequency to call a haplotype was set to 0.05. (DOCX)

S7 Fig. An example of the exhaustive model comparison using linear mixed-effect models for one patient's haplotype tree. The first column lists all combinations of post Baseline haplotype grouping, as indicated by coloured tip labels. The second column shows the fitting of the linear mixed-effect models with the corresponding groupings and the third column shows the BIC score of the model fit. The bolded BIC score of -108.78 was the lowest among all models tested and suggested that for this individual, time point 1 and 3 haplotypes belong to the same lineage, while time point 2 haplotypes belong to a different lineage. (DOCX)

S1 Text. Additional method specifications. (DOCX)

S1 Table. List of 65 resistance-associated variants (RAVs). (DOCX)

Author contributions

Conceptualization: Lele Zhao, Katrina Lythgoe.

Data curation: Haiting Chai, M Azim Ansari.

Formal analysis: Lele Zhao, Matthew Hall, Prahalad Giridhar, Mahan Ghafari, M Azim Ansari.

Funding acquisition: Katrina Lythgoe.

Methodology: Lele Zhao.

Software: Steven Kemp.

Supervision: Paul Klenerman, Eleanor Barnes, M Azim Ansari, Katrina Lythgoe.

Visualization: Lele Zhao.

Writing – original draft: Lele Zhao, Katrina Lythgoe.

Writing – review & editing: Lele Zhao, Matthew Hall, Mahan Ghafari, Katrina Lythgoe.

References

1. WHO. "Hepatitis C"; 2024. Available from: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>.
2. Sulkowski MS, Gardiner DF, Rodriguez-Torres M, Reddy KR, Hassanein T, Jacobson I, et al. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. *N Engl J Med*. 2014;370(3):211–21. <https://doi.org/10.1056/NEJMoa1306218> PMID: [24428467](https://pubmed.ncbi.nlm.nih.gov/24428467/)
3. Feld J, Jacobson I, Hézode C, Asselah T, Ruane P, Gruener N. Sofosbuvir and Velpatasvir for HCV Genotype 1, 2, 4, 5, and 6 Infection. *N Engl J Med*. 2015;373:2599–607.
4. Wyles D, Ruane P, Sulkowski M, Dieterich D, Luetkemeyer A, Morgan T. Daclatasvir plus Sofosbuvir for HCV in Patients Coinfected with HIV-1. *N Engl J Med*. 2015;373:714–25.
5. Kowdley KV, Gordon SC, Reddy KR, Rossaro L, Bernstein DE, Lawitz E, et al. Ledipasvir and sofosbuvir for 8 or 12 weeks for chronic HCV without cirrhosis. *N Engl J Med*. 2014;370(20):1879–88. <https://doi.org/10.1056/NEJMoa1402355> PMID: [24720702](https://pubmed.ncbi.nlm.nih.gov/24720702/)
6. Dietz J, Lohmann V. Therapeutic preparedness: DAA-resistant HCV variants in vitro and in vivo. *Hepatology*. 2023;78(2):385–7. <https://doi.org/10.1097/HEP.000000000000405> PMID: [37055017](https://pubmed.ncbi.nlm.nih.gov/37055017/)
7. Campo DS, Dimitrova Z, Yamasaki L, Skums P, Lau DT, Vaughan G, et al. Next-generation sequencing reveals large connected networks of intra-host HCV variants. *BMC Genomics*. 2014;15 Suppl 5(Suppl 5):S4. <https://doi.org/10.1186/1471-2164-15-S5-S4> PMID: [25081811](https://pubmed.ncbi.nlm.nih.gov/25081811/)
8. Palmer BA, Schmidt-Martin D, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, et al. Network Analysis of the Chronic Hepatitis C Virome Defines Hypervariable Region 1 Evolutionary Phenotypes in the Context of Humoral Immune Responses. *J Virol*. 2015;90(7):3318–29. <https://doi.org/10.1128/JVI.02995-15> PMID: [26719263](https://pubmed.ncbi.nlm.nih.gov/26719263/)
9. Alfonso V, Mbayed VA, Sookoian S, Campos RH. Intra-host evolutionary dynamics of hepatitis C virus E2 in treated patients. *J Gen Virol*. 2005;86(Pt 10):2781–6. <https://doi.org/10.1099/vir.0.81084-0> PMID: [16186232](https://pubmed.ncbi.nlm.nih.gov/16186232/)
10. Culasso ACA, Baré P, Aloisi N, Monzani MC, Corti M, Campos RH. Intra-host evolution of multiple genotypes of hepatitis C virus in a chronically infected patient with HIV along a 13-year follow-up period. *Virology*. 2014;449:317–27. <https://doi.org/10.1016/j.virol.2013.11.034> PMID: [24418566](https://pubmed.ncbi.nlm.nih.gov/24418566/)
11. Palmer B, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, Fanning L. Analysis of the evolution and structure of a complex intrahost viral population in chronic hepatitis C virus mapped by ultradeep pyrosequencing. *J Virol*. 2014;88:13709–21.
12. Li H, McMahon BJ, McArdle S, Bruden D, Sullivan DG, Shelton D, et al. Hepatitis C virus envelope glycoprotein co-evolutionary dynamics during chronic hepatitis C. *Virology*. 2008;375(2):580–91. <https://doi.org/10.1016/j.virol.2008.02.012> PMID: [18343477](https://pubmed.ncbi.nlm.nih.gov/18343477/)
13. Wang X-H, Netski DM, Astemborski J, Mehta SH, Torbenson MS, Thomas DL, et al. Progression of fibrosis during chronic hepatitis C is associated with rapid virus evolution. *J Virol*. 2007;81(12):6513–22. <https://doi.org/10.1128/JVI.02276-06> PMID: [17329332](https://pubmed.ncbi.nlm.nih.gov/17329332/)
14. Gray RR, Strickland SL, Veras NM, Goodenow MM, Pybus OG, Lemon SM, et al. Unexpected maintenance of hepatitis C viral diversity following liver transplantation. *J Virol*. 2012;86(16):8432–9. <https://doi.org/10.1128/JVI.00749-12> PMID: [22623804](https://pubmed.ncbi.nlm.nih.gov/22623804/)
15. Raghwan J, Rose R, Sheridan I, Lemey P, Suchard MA, Santantonio T, et al. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection. *PLoS Pathog*. 2016;12(9):e1005894. <https://doi.org/10.1371/journal.ppat.1005894> PMID: [27631086](https://pubmed.ncbi.nlm.nih.gov/27631086/)
16. Raghwan J, Wu C-H, Ho C, De Jong M, Molenkamp R, Schinkel J. High-resolution evolutionary analysis of within-host hepatitis C virus infection. *Journal of Infectious Diseases*. 2019;219:1722–9.
17. Riaz N, Leung P, Bull RA, Lloyd AR, Rodrigo C. Evolution of within-host variants of the hepatitis C virus. *Infect Genet Evol*. 2022;99:105242. <https://doi.org/10.1016/j.meegid.2022.105242> PMID: [35150893](https://pubmed.ncbi.nlm.nih.gov/35150893/)
18. Ramachandran S, Campo DS, Dimitrova ZE, Xia G, Purdy MA, Khudyakov YE. Temporal variations in the hepatitis C virus intrahost population during chronic infection. *J Virol*. 2011;85(13):6369–80. <https://doi.org/10.1128/JVI.02204-10> PMID: [21525348](https://pubmed.ncbi.nlm.nih.gov/21525348/)
19. Sorbo MC, Carioti L, Bellocchi MC, Antonucci F, Sforza D, Lenci I, et al. HCV resistance compartmentalization within tumoral and non-tumoral liver in transplanted patients with hepatocellular carcinoma. *Liver Int*. 2019;39(10):1986–98. <https://doi.org/10.1111/liv.14168> PMID: [31172639](https://pubmed.ncbi.nlm.nih.gov/31172639/)
20. Pérez PS, Di Lello FA, Mullen EG, Galdame OA, Livellara BI, Gadano AC, et al. Compartmentalization of hepatitis C virus variants in patients with hepatocellular carcinoma. *Mol Carcinog*. 2017;56(2):371–80. <https://doi.org/10.1002/mc.22500> PMID: [27163636](https://pubmed.ncbi.nlm.nih.gov/27163636/)

21. Gismondi MI, Díaz Carrasco JM, Valva P, Becker PD, Guzmán CA, Campos RH, et al. Dynamic changes in viral population structure and compartmentalization during chronic hepatitis C virus infection in children. *Virology*. 2013;447(1–2):187–96. <https://doi.org/10.1016/j.virol.2013.09.002> PMID: [24210114](https://pubmed.ncbi.nlm.nih.gov/24210114/)
22. Ducoulombier D, Roque-Afonso A-M, Di Liberto G, Penin F, Kara R, Richard Y, et al. Frequent compartmentalization of hepatitis C virus variants in circulating B cells and monocytes. *Hepatology*. 2004;39(3):817–25. <https://doi.org/10.1002/hep.20087> PMID: [14999702](https://pubmed.ncbi.nlm.nih.gov/14999702/)
23. Rose R, Rodriguez C, Dollar J, Lamers S, Massaccesi G, Osburn W. Inconsistent temporal patterns of genetic variation of HCV among high-risk subjects may impact inference of transmission networks. *Infect Genet Evol*. 2019;71(1):1–6.
24. Childs K, Davis C, Cannon M, Montague S, Filipe A, Tong L, et al. Suboptimal SVR rates in African patients with atypical genotype 1 subtypes: Implications for global elimination of hepatitis C. *J Hepatol*. 2019;71(6):1099–105. <https://doi.org/10.1016/j.jhep.2019.07.025> PMID: [31400349](https://pubmed.ncbi.nlm.nih.gov/31400349/)
25. Gupta N, Mbituyumuremyi A, Kabahizi J, Ntaganda F, Muvunyi CM, Shumbusho F, et al. Treatment of chronic hepatitis C virus infection in Rwanda with ledipasvir-sofosbuvir (SHARED): a single-arm trial. *Lancet Gastroenterol Hepatol*. 2019;4(2):119–26. [https://doi.org/10.1016/S2468-1253\(18\)30382-0](https://doi.org/10.1016/S2468-1253(18)30382-0) PMID: [30552056](https://pubmed.ncbi.nlm.nih.gov/30552056/)
26. Smith D, Magri A, Bonsall D, Ip CLC, Trebes A, Brown A, et al. Resistance analysis of genotype 3 hepatitis C virus indicates subtypes inherently resistant to nonstructural protein 5A inhibitors. *Hepatology*. 2019;69(5):1861–72. <https://doi.org/10.1002/hep.29837> PMID: [29425396](https://pubmed.ncbi.nlm.nih.gov/29425396/)
27. Chen Z-W, Li H, Ren H, Hu P. Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data. *Sci Rep*. 2016;6:20310. <https://doi.org/10.1038/srep20310> PMID: [26842909](https://pubmed.ncbi.nlm.nih.gov/26842909/)
28. Palladino C, Ezeonwumelu I, Mate-Cano I, Borrego P, Martínez-Román P, Arca-Lafuente S. Epidemic history and baseline resistance to NS5A-specific direct acting drugs of hepatitis C virus in Spain. *Scientific Reports*. 2020;10:13024.
29. Patiño-Galindo JÁ, Salvatierra K, González-Candelas F, López-Labrador FX. Comprehensive Screening for Naturally Occurring Hepatitis C Virus Resistance to Direct-Acting Antivirals in the NS3, NS5A, and NS5B Genes in Worldwide Isolates of Viral Genotypes 1 to 6. *Antimicrob Agents Chemother*. 2016;60(4):2402–16. <https://doi.org/10.1128/AAC.02776-15> PMID: [26856832](https://pubmed.ncbi.nlm.nih.gov/26856832/)
30. Takeda H, Ueda Y, Inuzuka T, Yamashita Y, Osaki Y, Nasu A, et al. Evolution of multi-drug resistant HCV clones from pre-existing resistant-associated variants during direct-acting antiviral therapy determined by third-generation sequencing. *Sci Rep*. 2017;7:45605. <https://doi.org/10.1038/srep45605> PMID: [28361915](https://pubmed.ncbi.nlm.nih.gov/28361915/)
31. Kliemann DA, Tovo CV, Gorini da Veiga AB, Machado AL, West J. Genetic Barrier to Direct Acting Antivirals in HCV Sequences Deposited in the European Databank. *PLoS One*. 2016;11(8):e0159924. <https://doi.org/10.1371/journal.pone.0159924> PMID: [27504952](https://pubmed.ncbi.nlm.nih.gov/27504952/)
32. Fernandez-Antunez C, Wang K, Fahnøe U, Mikkelsen LS, Gottwein JM, Bukh J, et al. Characterization of multi-DAA resistance using a novel hepatitis C virus genotype 3a infectious culture system. *Hepatology*. 2023;78(2):621–36. <https://doi.org/10.1097/HEP.000000000000353> PMID: [36999539](https://pubmed.ncbi.nlm.nih.gov/36999539/)
33. Vo-Quang E, Soulier A, Ndebi M, Rodriguez C, Chevaliez S, Leroy V, et al. Virological characterization of treatment failures and retreatment outcomes in patients infected with “unusual” HCV genotype 1 subtypes. *Hepatology*. 2023;78(2):607–20. <https://doi.org/10.1097/HEP.000000000000379> PMID: [36999537](https://pubmed.ncbi.nlm.nih.gov/36999537/)
34. Dietz J, Di Maio VC, de Salazar A, Merino D, Vermehren J, Paolucci S, et al. Failure on voxilaprevir, velpatasvir, sofosbuvir and efficacy of rescue therapy. *J Hepatol*. 2021;74(4):801–10. <https://doi.org/10.1016/j.jhep.2020.11.017> PMID: [33220331](https://pubmed.ncbi.nlm.nih.gov/33220331/)
35. Foster GR, Pianko S, Brown A, Forton D, Nahass RG, George J, et al. Efficacy of sofosbuvir plus ribavirin with or without peginterferon-alfa in patients with hepatitis C virus genotype 3 infection and treatment-experienced patients with cirrhosis and hepatitis C virus genotype 2 infection. *Gastroenterology*. 2015;149(6):1462–70. <https://doi.org/10.1053/j.gastro.2015.07.043> PMID: [26248087](https://pubmed.ncbi.nlm.nih.gov/26248087/)
36. Knyazev S, Tsyvina V, Shankar A, Melnyk A, Artyomenko A, Malygina T, et al. Accurate assembly of minority viral haplotypes from next-generation sequencing through efficient noise reduction. *Nucleic Acids Res*. 2021;49(17):e102. <https://doi.org/10.1093/nar/gkab576> PMID: [34214168](https://pubmed.ncbi.nlm.nih.gov/34214168/)
37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*. 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015> PMID: [32011700](https://pubmed.ncbi.nlm.nih.gov/32011700/)

38. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol.* 2011;11:131. <https://doi.org/10.1186/1471-2148-11-131> PMID: 21595904
39. Lythgoe KA, Lumley SF, Pellis L, McKeating JA, Matthews PC. Estimating hepatitis B virus cccDNA persistence in chronic infection. *Virus Evolution.* 2021;7:veaa063. <https://doi.org/10.1093/ve/veaa063>
40. Smith DA, Fernandez-Antunez C, Magri A, Bowden R, Chaturvedi N, Fellay J, et al. Viral genome wide association study identifies novel hepatitis C virus polymorphisms associated with sofosbuvir treatment failure. *Nat Commun.* 2021;12(1):6105. <https://doi.org/10.1038/s41467-021-25649-6> PMID: 34671027
41. Yamashita T, Takeda H, Takai A, Arasawa S, Nakamura F, Mashimo Y. Single-molecular real-time deep sequencing reveals the dynamics of multi-drug resistant haplotypes and structural variations in the hepatitis C virus genome. *Scientific Reports.* 2020;10:2651. <https://doi.org/10.1038/s41598-020-59663-5>
42. Nakamura F, Takeda H, Ueda Y, Takai A, Takahashi K, Eso Y. Mutational spectrum of hepatitis C virus in patients with chronic hepatitis C determined by single molecule real-time sequencing. *Scientific Reports.* 2022;12:7083.
43. Dietz J, Müllhaupt B, Buggisch P, Graf C, Peiffer K-H, Matschenz K, et al. Long-term persistence of HCV resistance-associated substitutions after DAA treatment failure. *J Hepatol.* 2023;78(1):57–66. <https://doi.org/10.1016/j.jhep.2022.08.016> PMID: 36031158
44. WHO. “HIV Drug Resistance Report 2024”; 2024. Available from: <https://iris.who.int/bitstream/handle/10665/376039/9789240086319-eng.pdf>
45. Gray RR, Salemi M, Klenerman P, Pybus OG. A new evolutionary model for hepatitis C virus chronic infection. *PLoS Pathog.* 2012;8(5):e1002656. <https://doi.org/10.1371/journal.ppat.1002656> PMID: 22570609
46. Batty EM, Wong THN, Trebes A, Argoud K, Attar M, Buck D, et al. A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. *PLoS One.* 2013;8(6):e66129. <https://doi.org/10.1371/journal.pone.0066129> PMID: 23762474
47. Bonsall D, Ansari MA, Ip C, Trebes A, Brown A, Klenerman P, et al. ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Res.* 2015;4:1062. <https://doi.org/10.12688/f1000research.7111.1> PMID: 27092241
48. Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D, et al. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus Evol.* 2018;4(1):vey007. <https://doi.org/10.1093/ve/vey007> PMID: 29876136
49. Ansari MA, Aranday-Cortes E, Ip CL, da Silva Filipe A, Lau SH, Bamford C, et al. Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *Elife.* 2019;8:e42463. <https://doi.org/10.7554/eLife.42463> PMID: 31478835
50. Ansari MA, Pedergnana V, L C Ip C, Magri A, Von Delft A, Bonsall D, et al. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet.* 2017;49(5):666–73. <https://doi.org/10.1038/ng.3835> PMID: 28394351
51. Wing PAC, Jones M, Cheung M, DaSilva S, Bamford C, Jason Lee W-Y, et al. Amino Acid Substitutions in Genotype 3a Hepatitis C Virus Polymerase Protein Affect Responses to Sofosbuvir. *Gastroenterology.* 2019;157(3):692–704.e9. <https://doi.org/10.1053/j.gastro.2019.05.007> PMID: 31078622
52. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, et al. PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Mol Biol Evol.* 2018;35(3):719–33. <https://doi.org/10.1093/molbev/msx304> PMID: 29186559
53. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* 2019;47(11):5539–49. <https://doi.org/10.1093/nar/gkz361> PMID: 31076776
54. Wang TH, Donaldson YK, Brettell RP, Bell JE, Simmonds P. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J Virol.* 2001;75(23):11686–99. <https://doi.org/10.1128/JVI.75.23.11686-11699.2001> PMID: 11689650
55. Lewis MJ, Frohnen P, Ibarondo FJ, Reed D, Iyer V, Ng HL, et al. HIV-1 Nef sequence and functional compartmentalization in the gut is not due to differential cytotoxic T lymphocyte selective pressure. *PLoS One.* 2013;8(9):e75620. <https://doi.org/10.1371/journal.pone.0075620> PMID: 24058696