Responses to David Rasmussen

Your preprint has been reviewed by two experts with substantial experience in the field of viral phylodynamics. Both the reviewers and I appreciate that identifying populations and risk factors driving the transmission dynamics of HCV and other chronic infections is an extremely important topic relevant to public health.

Thanks!

However, both reviewers raise substantial concerns about the analysis that I believe need to be addressed before I can offer a recommendation. In particular, one reviewer raises serious concerns about the quality of the phylogenetic reconstruction and therefore the conclusions that can be drawn from a single ML tree.

As we discuss below, we performed sensitivity analyses to address this issue. We also perform ABC analysis on multiple phylogenies inferred using Beast2 to assess the robustness of the estimations.

The other reviewer points out that some of the main conclusions, such as in which risk group the epidemic is growing faster in, are not clearly supported by the data and may in fact be largely influenced by the author's choice of priors (i.e. no prior support is given to the alternative scenario that the epidemic is growing faster in 'classic' hosts).

As discussed in details below, we reran the analysis with larger priors to account for a wider range of scenarios. In short, this does not affect our results.

Both reviewers also point out that the model needs to be more clearly defined, including how individuals are classified into risk groups, and I would suggest maybe briefly describing the model and how different host types are defined before the results section in your revision.

The host classification is performed independently of the sequence data by field epidemiologists based on interviews and risk factors. We clarify this in the article.

We also added a section at the end of the introduction to describe the model.

Many of the reviewers concerns could be addressed by performing a second analysis using the multi-type birth death models implemented in the BDMM package in BEAST 2. The authors make a point in the Discussion that birth-death models might not be applicable here because the two epidemics are linked by transmission rather than mutation, but the 'types' assigned to lineages can either represent the type of the host (as in new or classical) or the type of pathogen (mutant or non mutant) under multi-type birth-death models. Performing this additional analysis with BDMM would allow the authors to compare their ABC methods with more traditional likelihood-based phylodynamic methods, which would lend trust to authors conclusions especially since ABC methods are still in their infancy and many readers might be interested in this comparison. Furthermore, fitting a multi-type birth-death

model in BEAST would allow for joint inference of the phylogeny with the epidemic parameters, addressing the first reviewer's point about tree uncertainty.

As indicated by the reviewer, we mentioned in our manuscript that, to our knowledge, this exact scenario we analyse cannot be captured by Beast (at least using the Beauti interface). First, setting the sampling rate not to start at t=0 has to be done manually, second the BDMM only captures “mutation” models, where a host of type A actually becomes a host of type B, and not transmission models (a host of type A infects a host of type B).

We tried to run an analysis using a similar model, with the important difference that switches in host types occur as mutations instead of being linked to transmission. Unfortunately, perhaps due to our ability to configure the .xml file, the MCMC did not even launch...

In addition to the reviewers many thoughtful comments, I would add the following points as well:

Line 64: "date of the second epidemic" -- what is this second epidemic?
We apologize for the lack of clarity: we meant the epidemic in the new hosts. (To be corrected.)

Line 54: "The width of the posterior distribution indicates our ability to infer a parameter" -- This is not necessarily true... the width of the posterior could be very wide with very long tails, but most of the posterior density could still be centered around a narrow range of values.
Indeed! We now refer to the credibility interval.

Bayesian should be capitalized throughout.
Done

Could the authors comment on why was the infectious period inferred to be so short for "new" hosts?
We believe that the inferred infectious period is short because the time between diagnosis and treatment (and therefore the time between the first detectable viral load and the first undetectable viral load) has decreased significantly over time, and here most of the “new” hosts are successfully treated.

Responses to Louis DuPlessis

Summary

This manuscript investigates the roles of different populations in sustaining ongoing HCV epidemics in France using 213 newly sequenced HCV genomes and an ABC phylodynamics approach to infer epidemiological parameters. In particular the authors examine the difference between so-called classical hosts, who spread the disease through needle-sharing and new hosts that spread HCV through sexual transmission.

I found some disconnect between the populations described in the introduction and the ones actually used in the analysis. This may just be a small misunderstanding on my part, but to ensure that everyone is on the same page I wrote down a detailed description of my understanding below.

The weak point of the ABC approach is that all parameter estimates (and some of the model priors) are conditioned on a single independently estimated tree. If this tree is poorly estimated, parameter estimates will likely be biased. From what is presented in the manuscript I am not convinced that the dataset used here has sufficient temporal resolution for the reconstructed tree (in this case the MCC tree from a BEAST analysis) to be correlated to the transmission tree. I also expand on this point below.
We performed a TempEst analysis and the temporal signal is indeed poor.

To address this issue, we re-ran the analysis using a fixed substitution rate as suggested by Reviewer #2.

This was done using 10 different target trees from the Beast2 posterior distribution.

We find that our results are robust to fixing the substitution rate and to varying the phylogeny shape.

I think the application is of high interest and should definitely be investigated. The ABC method is interesting as an alternative to the usual Bayesian phylodynamics approach, since it allows using (almost) arbitrarily complex epidemiological models without the need to calculate the probability of observing a tree under the model. However, it also comes with its own set of limitations, and in this case I think the dataset may not be suitable. In the comments below I offer a few suggestions for addressing this issue. As things stand I trust the estimates of the ABC method, given the tree the authors used, but I don't trust that tree and so I would not extrapolate the results to the real world.

<u>Major comments</u>

1. The different populations of interest and study populations should be presented more clearly.

    - As I understand it, there are 4 populations of interest:
        1. HIV-positive PWID
        2. HIV-positive MSM
        3. HIV-negative MSM using PrEP
        4. HIV-negative MSM not using PrEP
    - (There is some degree of overlap between populations 1 and 2 and population 4 is mostly unsampled).

We apologize for the confusion. In the introduction of the original version, we indeed described the PWID hosts as being HIV-positive when they are in fact HIV-negative. We corrected this in the revised version.

Furthermore, the “classical” population does not only include PWID but potentially any type of transmission route (MSM, blood transfusion, nosocomial transmission...). It is only for the new host that there is a common transmission route (they are all MSM, generally detected during acute HCV infection).

Regarding the MSM hosts (i.e. “new” hosts), we do not have the information about PrEP use, but do know about the HIV state. 78% of the MSM hosts are HIV positive, from which 53% have been detected during or shortly after acute phase. 22% of the MSM hosts are HIV negative and all have been detected during acute phase.

To summarise, the populations are mostly:

1. HIV-negative “classical” hosts
2. HIV-positive MSM
3. HIV-negative MSM

- By the definition in the introduction classical hosts predominantly stem from population 1 and new hosts from the remaining 3 populations, where transmission is sexual.

The host types are obtained from global epidemiological profiles. This means for instance that not all the hosts of the “classical” epidemics have been detected during chronic infection. As indicated above, “classical” hosts type include different profiles: IDU, blood transfusion recipient, MSM detected during chronic phase... Whereas the “new” hosts are all MSM who have been detected during acute phase.

- In the genetic dataset new hosts are represented by all MSM patients (which presumably includes representatives from populations 2,3 and 4). Classical hosts are represented by non-MSM, HIV-negative male patients.
- *Did I understand all of the populations correctly?*

On average the trend is there. But this is only an average on the host profiles established by field epidemiologists.

- **The relationship between classical hosts in the genetic dataset and HIV-positive PWID (population 1) is not clear to me**. Are they also PWID?

Again we apologize for our mistake in the manuscript. The classical hosts include different profiles, and the mode of transmission is mostly unknown and hence qualified as nosocomial. In our dataset, there are no HIV-positive PWID/”classical” host.

Do you have any information about the main mode of transmission in this population or a reason to believe that it is through needle-sharing?

We apologize for the lack of clarity. As indicated above, in the classical hosts, the reported transmission route are majoritarily nosocomial but others exist such as injecting drug users or blood transfusion or sexual intercourse.

2. The main informative events for the ABC approach are the branching times in the tree. Since the ABC approach conditions on a single tree, it is extremely important that the timing of the tree is accurate. Judging from the TMRCA estimate, which has an HPD from 1962 to 1997, I think it's likely that many of the other internal nodes have a similar range of uncertainty, which is ignored when conditioning on the MCC tree. The result is that estimates that rely on the timescale (reproduction numbers, infectious periods, origin times) are likely to be biased.

We agree that there is uncertainty in the target tree but, in our opinion, this should make it even more difficult to detect signal and should also weaken the sensitivity analysis. Furthermore, the simulation of the trees is not conditioned on the internal nodes, only on the sampling of the leaves.

To further investigate this, we picked 10 of the trees from the posterior distribution inferred with Beast2 and performed our ABC analysis on each one of them. The ABC results are all similar to our first results and are now shown in Appendix.

> As an example, the authors placed a uniform prior from 1962 to 1997 on t0. However, by conditioning on the MCC tree, the upper bound of this prior implicitly becomes 1981, which explains why no t0 estimates are larger than 1981.

Here we apologize but we do not understand why working with the MCC tree would truncate the prior from [1962,1997] to [1962,1981].

3. In addition, the authors divided the tree into 3 epochs, with different reproduction numbers. However, it is likely that the uncertainty associated with internal nodes stretch across multiple epochs, which call into question the estimates reported here. *(If the procedure was repeated with different trees drawn from the set of posterior trees parameter estimates would be different)*.

Using a different tree would not affect our simulations since the only information we use from the target tree are the leaves (i.e. the sampling dates). This means we can use our simulated data to perform the ABC inference on different target trees with minimal computational effort.

This is now clarified in the manuscript.

- At the least the authors should test for a temporal signal in the data. Since sequences are only sampled over 4 years and HCV has highly variable within-host evolutionary rates, it is possible that there isn't a good clock signal. The authors should also provide more details of the model used for the

BEAST analysis, as using a poorly fitting tree prior can lead to inferring biased branching times.

As indicated above, we performed an analysis using Tempest. This led us to analyse a phylogeny with a fixed molecular clock.

- Even if there is a temporal signal in the data it may still be impossible to obtain accurate estimates of the branching times, since the sampling period is small compared to the tree height. In this case, the analysis could be performed with a fixed or highy constrained clock rate (based on previous analyses of HCV genotype 1a). Alternatively, more genotype 1a sequences from Europe (spanning a bigger sampling period) could be downloaded from Genbank. This will allow internal nodes to be more accurately estimated. If the Lyon sequences form a monophyletic clade, this clade could be pruned from the tree and used for further analyses by the ABC approach.

As suggested, we inferred a phylogeny using Beast2 with a fixed clock rate. The clock value was chosen using existing to be consistent with known HCV evolutionary rates. The results were similar to that obtained when estimating the clock rate.

In the main text, we now report the results with the fixed molecular clock.

4. I believe the authors are estimating the effective reproduction number (Re), not the basic reproduction number (R0).

Indeed: we corrected the text.

5. Figure 1 should be improved. It is impossible to see the structure of the tree close to the present or to read the tip labels.

We improved the tree representation in the revised version.

6. **Line 213:** \nu is not simply a factor of the [mean] number of partners of classical (I1) and new (I2) hosts. The modes of transmission differ between I1 and I2 (blood vs. sexual) and there is also presumably a difference in the per-act likelihood of transmission between the different modes of transmission.
   In addition, \nu is constrained to always be greater than or equal to 1. This appears to be backed up by the data, but is not necessarily always the case. Why is it *a priori* assumed that there is a higher transmission rate between new hosts (MSM) than between classical hosts (PWID)?

We agree that the transmission route is different but, since the transmission route remains uncertain, we did not model it mechanistically (which would anyway be unlikely to bring additional insights given the limited size of the dataset). More generally, adding more parameters in the model would require a larger dataset and additional summary statistics.

In the version of the analysis shown in the main text, we have modified the prior for \nu to account for the limitation above.

Minor comments

1. What is meant by the phylogeny was estimated in PhyML and then rooted using BEAST? Does this mean that you performed a BEAST analysis on the sequence data using the maximum-likelihood tree as a starting tree?

We apologize for this error: the phylogeny was also estimated using Beast2 (PhyML was used in earlier analyses to combine it with LSD).

This is corrected in the new version of the the manuscript.

2. **Line 57:** I don't know the shape of the posterior distribution for the TMRCA, but regardless of how peaked it is, I wouldn't place much confidence on the TMRCA being in the early 1980s when the HPD interval stretches from 1962 to 1997.

We agree. This is why we use the HPD interval minimum and maximum values as our prior to estimate the tMRCA using ABC.

3. Do the authors use the epidemiological data from the Dat'AIDS cohort anywhere? Is there a relation between the cohort and the sequences sampled in Lyon?

No, we did not use the data from the Dat'AIDS cohort. We mentioned it because the HCV cohort from Lyon is part of the Dat'AIDS cohort. We clarified the text on this issue.

4. The AIC is not a good method for Bayesian model comparison. To choose between a strict and relaxed clock it makes more sense to check if the HPD for the coefficient of variation (for the relaxed clock model) excludes 0.

We now fix the clock rate given the lack of temporal signal.

5. **Line 242:** Please provide more details about how coalescent trees are drawn from the simulated trajectories and sampling dates (are the sampling dates fixed to the truth or are they also from the model?).

The phylogenies are simulated in a two-step procedure.

First, we simulate epidemiological trajectories using our compartmental model and Gillespie's stochastic event-driven simulation algorithm. For these trajectories, we record the number of individuals in each compartment through time and the reactions that occur at each step (recovery of a host, transmission between hosts, etc.). The sampling dates are

fixed and originate from the target tree. For each simulation, a sampling date is randomly associated to a host compartment given a prior on the proportion of the dates for each host.

Second, we add the sampling reactions and their dates, associated to a host compartment, to the trajectories. The simulation of a phylogeny starts from the last sampling date and follows the epidemiological trajectory backward-in-time. Each backward step in the trajectory can induce a tree modification: a sampling reaction will lead to a labelled leaf in the phylogeny, a transmission reaction can lead to the coalescence of two sampled lineages.

So to answer the question itself, in the simulation, the dates of the leaves are that of the target phylogeny. However, the nature of the leaf label (new or classical host) is drawn at random.

This has been clarified in the manuscript.

6. Using the word "cluster" for heterogeneous and homogeneous clades is misleading as cluster has a specific interpretation in an epidemiological setting. Clade is a better term to use here.

This is true indeed. We made the change.

7. Heterogeneous clusters (should be clade) of type Y are defined as clades where more than 70% of leaves are of type Y (and includes clades where all but one leaf are of type Y). What about very heterogeneous clades, where type Y makes up between 50 and 70% of leaves? Are these clades completely ignored? In addition, did you use a greedy algorithm to find the biggest possible clades or do you have nested homogeneous and heterogeneous clades? (the same branch can appear in multiple clades).

Clades where Y represents 50 to 70% of the leaves are indeed ignored.

Given the structure of phylogeny, where only a few new hosts tips are included in the new hosts clusters, we do not expect the cut-off value (here set to 70%) to have a large effect. But, theoretically, we could compute the summary statistics with a cut-off of 50%.

The algorithm will not compute summary statistics on a branch that appears in different homogeneous clades (or heterogeneous clades) because we always select the largest clade. However, it does allow to compute summary statistics of a branch that appears in both a homogeneous clade and in a heterogeneous clade, because for each clade we capture two different phylogeny properties: homogeneity and heterogeneity.

Reponses to Chris Wymant

I preface my remarks by noting that I am unfamiliar with approximate Bayesian computation, and cannot comment on how well it was executed here.

The authors perform phylodynamic analysis of viral genetic sequence data from two epidemiologically distinct populations of individuals infected with HCV, to test the hypothesis that the populations have different transmission dynamics. The question is both relevant for public health and interesting for epidemiology, and the method used is appropriate. I have some substantive concerns about the conclusions, and some minor concerns and suggestions.

Substantive concerns

The main conclusion of the paper is that the epidemic in new hosts is growing faster than that in classical hosts, however the confidence with which this conclusion can be made is not stated. The doubling time in classical hosts since 1997 is estimated to be 0.58 - 10.13 years, and the doubling time in new hosts 0 - 3.51 years. The relevant quantity for the conclusion is the posterior for the ratio of these two parameters. The authors do present a ratio comparing the two hosts with regards to the reproduction number, and find that the confidence intervals do not exclude 1; if the same is true of the doubling time, which seems plausible given the similar parameter dependencies of R_0 and doubling time, the main conclusion is not supported. The same point applies to the other host parameters inferred to be different: assortativity and recovery/removal rate. The parameters themselves need not be redefined, but the posteriors of their ratios should be examined to support claims of differences.

91: "*We also estimate that 'new' hosts transmit HCV 6.50[2.56; 9.81] times more than 'classical' hosts (parameter v)*"
Unless I have misunderstood, the authors' choice of the prior Unif(1, 10) for the parameter nu means it was impossible to come to any conclusion other than a greater transmission rate for new hosts. The prior for nu should have equal weight above and below 1 if we are to learn how the data inform the ratio of transmission rates; here it has zero weight below 1. I think that treating the two hosts equally a priori (i.e. imposing a host-type interchangeability symmetry) implies that the prior for log(nu) should be symmetrical about 0.

Thank you for the suggestion. We now run the same ABC analyses with a prior of \nu in Unif(0,10) to include both hypotheses.

The only description I can find for the method of going from the birth-death model equations to a phylogeny is that it "*resembles that developed by [Saulnier et al.] and uses Gillespie's stochastic simulation algorithm...* [and] *generates phylogenies of infections using the coalescent approach based on simulated trajectories and sampling dates. Importantly, we assume that the virus can still be transmitted after sampling.*" I think more detail would be appropriate here as it is an important part of the method. For example, it is unclear to me whether/how the model handles the fact that the new host samples represent high sampling of a small underlying population, whereas the classical host samples represent lower sampling of a larger underlying population. This may be clear for readers familiar with these

inference methods, but reassurance for other readers would be appreciated. The Saulnier et al. paper states that "In the SI-DR model [similar to the that of the current work, I think]... the number of new infections also depends on the susceptible population size, but there is no sampling because the model assumes that the sampling dates are known." This seems relevant but I'm unsure whether it addresses my particular concern about the conclusion, below.

As explained in our response to Reviewer #1, phylogenies are simulated in a two-step procedure where the trajectories are first simulated in a forward-in-time process using Gillespie's SSA. Then, we merge the sampling dates into the trajectories and simulate phylogenies using a coalescent process from the last sampling date. A sampling event will lead to a leaf in the phylogeny and a transmission event can lead to the coalescence of two sampled lineages. Note that each sampling date is randomly associated to a host compartment. By using the coalescent process, we assume the sampling rate to be low.

This is clarified in the revised version of the manuscript.

The authors note that a legitimate 'interrogation' about the study (would 'limitation' be better?) is that for new hosts sampling is expected to be 'high', whereas for classical hosts it is 'representative' (I assume this implies lower), but do not comment on the relevance of this for the results. I have a hazily formed concern that this could bias the inference in the direction of the conclusion drawn, even in the null hypothesis of no difference in transmission dynamics. Specifically, if mixing is mostly assortative, new hosts are sampled densely from a small population, and classical hosts are sampled sparsely from a large population, might that give a tree with the observed structure - few recent coalescents for classical hosts but many for new hosts - even for the null hypothesis? I think this could be easily tested with the existing simulation. If this nuisance effect does exist, does it manifest itself in the tree in a manner that (a) is different from the effect of interest - transmission dynamics - and (b) can be modelled, thus permitting the two effects to be quantitatively disentangled?

Sampling bias is a research topic in itself (and we know colleagues who are working on this exact same question). We now discuss how the (likely) higher sampling rate in the new epidemic could affect the results. Regarding the assortativity, we cannot rule it out but cannot readily include it in the model because it would require to explicitly model the population sizes of each host type, which we tried but without success given the limited size of the dataset and the absence of prevalence data. Another bias could be with regard to the speed of spread of the epidemics because the coalescent classically assumes a negligible sampling rate. However, we think this bias should lead to us underestimating the speed of spread of the new epidemics (we simulate forward in time first without removing sampled hosts so a high sampling rate should make the realised R0 smaller than the simulated one; neglecting sampling should therefore allow us to simulate the epidemics with a lower R0 than the "true" one), which we already find to be larger than the classical one.
In summary, we acknowledge the limitation about the population structure and think that the R0 of the new epidemics is potentially underestimated.

Regardless of whether the previous concern is grounded, running the inference on the simulated null hypothesis would be reassuring. For example, this would have detected the bias introduced by the prior on nu. Allowing a piecewise-through-time R_0 for one host type but not the other might also lead to subtle differences in the inference of growth even for the null hypothesis.
We could indeed have also estimated a value of the R0 of the new epidemics in the time points where it is not reported but decided to keep our prior there because a large part of the study relies on the epidemiological data independent from the sequence.

However, one way to answer this question is that when we tried a similar null hypothesis, with only acute infections (which could be seen as “new” hosts) and chronic infections (“classical” hosts) we were unable to perform simulations that matched the phylogeny (there were not enough host of one type or the other when performing our backward-in-time phylogeny inference).

Minor concerns and suggestions

Given the paper's main aim, it would benefit from a more precise definition of what the difference between the two host categories is, as the distinction seems to me to be slightly vague. The most explicit statement is lines 42-25:
"*‘classical’ hosts (typically HIV-infected patients with a history of opioid intravenous drug use) and ‘new’ hosts (HIV-infected and HIV-negative MSM, detected during or shortly after acute HCV phase, potentially using recreational drugs such as cocaine or cathinones)*"
By only describing a number of features of the typical classical host and a number of features of the typical new host, it's unclear how to categorise any individual who doesn't fit either description (are they excluded here?), or who fits both. What is the deciding factor? Is it whether the individual is an MSM or not, or whether the most likely exposure route is thought to be needle exchange or not, or something else? I assume that the timing of the diagnosis is correlated with category but does not affect the categorisation decision (as I think it is at best a proxy for patient characteristics of real interest), in which case distinguishing deciding factors and correlating factors for the categorisation would be helpful.
The host classification is performed independently of the sequencing data by field epidemiologists based on interviews and risk factors. The “new” hosts are MSM detected during acute phase whereas the “classical” hosts type includes different profiles such as IDU, men who received blood transfusions, MSM detected during chronic phase, etc... The deciding factor is indeed the most likely transmission route.

I suggest making the categorisation precise in the abstract as well as the main text. A related and very minor point: perhaps there are equally concise but more informative names for the host types (MSM / non-MSM, IDU / non-IDU), which would remove the slightly distracting need to keep the names in quotation marks at every occurence, while avoiding any confusion about 'new hosts' being a fixed category over time, not newly infected individuals.
Thank you for the suggestion: we tried to clarify the host type names in the revised version. However, given the answer above, the “classical” hosts including different epidemiological profiles, we decided to keep our notation.

Another hazily formed concern: is there a significant difference in the distribution of sampling dates for the two host categories? If so, might this have an effect on the inference of growth due to some right-censoring effect? This could also perhaps be tested using the existing simulation.
There are 145 sampling dates associated with the IDU label (“classical” hosts) and 68 with the MSM label. For the “classical” hosts, the mean value of sampling dates is 2015.86 [2014.85;2018.33] (numbers in brackets indicate the 95% confidence interval). For the “new” hosts, the mean sampling date is 2016.20 [2014.85;2018.33]. So we do not expect a bias for one or the other.

73: I suggest explaining the subscript 1 or 2 which is introduced here.
This was added in the new version.

82: "*that ‘new’ hosts transmit HCV 6.50[2.56; 9.81] times more than ‘classical’ hosts (parameter v).*"
I suggest rephrasing for clarity, something like
"the rate at which new hosts transmit is 6.50[2.56; 9.81] times greater than for classical hosts."
Following this with mention of the different recovery/removal rates makes the similar values of R_0, despite the very different transmission rates, more intuitive.
Done! Thank you.

94-95: the symbols nu and gamma are introduced here - it would be good to define them in words here.
Done!

95-96: "*a rapid growth of the epidemic in ‘classical’ hosts imposes a lower growth in ‘new’ hosts.*" Can the reason for this be clarified?
What we meant here is that the rooting in time of the phylogeny generates some constraints. One of these has to do with the speed of spread. If the doubling time of the first epidemics is on an extreme side of the distribution, then we are not able to simulate epidemics in which the doubling time of the new epidemics is on the same extreme side of the distribution.

The prior for R_0 for the classical hosts in recent years constrains it to be less than 0.13 with 100% certainty. I would not describe this as “as little informative as possible” as the authors do - it seems very restrictive. And given the above point about negative correlation between the two growth rates, might this restrictive choice of prior push inference of nu to higher values?
We apologize for our mistake here: a comma was forgotten. The prior for the R_0 from the second time interval (after 1997) is from 0.1 to 3 following a uniform distribution.

Figure 3: I suggest using a smaller point size to plot simulated points, such that there is negligible overlap between different points even at the region of highest density. If many simulated points are on top of each other away from where the data point lies - which we can't tell currently - the data point could actually be in a low-density tail away from where the simulated distribution is concentrated.

Thank you for the suggestion.

97-98: "*epidemics with the same R_0 but a longer infection duration have a lower doubling time and therefore a weaker epidemiological impact.*" What is meant by "weaker epidemiological impact", a smaller outbreak size perhaps?
Yes, a lower doubling time means a smaller outbreak in the same amount of time.

More explicitly, is the logic here that if we fix R_0 and increase the infection duration and thus the doubling time, the same sized outbreak could only be explained by an earlier onset, but that is not permitted by the prior on the onset, based on the tree?
This is correct as well.

104: "*Even if*" -> "However, if"
Done.

112-116: I do not understand the point made here. It sounds like it could be paraphrased as "cross-validation is inappropriate for testing estimation of those parameters for which the tree is highly informative", which does not make sense, so I assume I am misunderstanding.
We apologize for the lack of clarity. Here, we used a threshold of 70% for the summary statistics on heterogeneous clusters given the structure of our target phylogeny. But, for the cross-validation we might have used a simulated phylogeny as a target that has a different structure with more heterogeneous clusters that our summary statistics may have not captured.

The main point we wanted to make here is that the cross-validation tests scenarios that have little to do with the observed phylogeny (where clearly there is a high degree of clustering). The summary statistics we designed have been chosen based on this tree shape. A reformulation of our statement would be that our ability to infer parameters really depends on the summary statistics chosen. Therefore, to infer assortativity values lower than 0.5 (which biologically would have little sense in this context), other summary statistics should be developed.

133-135: "*That the duration of the infectious period in new hosts is in the same order of magnitude as the time until treatment suggests that the majority of the infection may be occurring during the acute phase.*"
Would this statement be more accurate if "during the acute phase" were replaced by "before treatment"? If so, isn't this observation - that treatment mostly stops infectiousness - just a sanity check on the results, rather than evidence for conclusions about the acute phase?
Classically, treatment is given at the end of the acute infection. This is because patients can recover naturally. This is still the case with the new treatments (protease inhibitors) but the difference is that they are more successful at treating the infection than the old interferon-based therapy. So it makes sense that with the new treatments, host infection duration would be truncated at the end of the acute infection..

146 onwards: I don't understand the hypothesis being tested here. I think it is that previously the epidemic consisted only of people who were diagnosed a long time after infection, but

more recently there is a separate epidemic of people who were diagnosed shortly after infection. This only makes sense to me if diagnosis time is merely a proxy for another characteristic that is causal for difference in dynamics.
Here, we wanted to validate the fact that the differences in the structure and the labels in the phylogeny are not due to the infection phase (acute vs. chronic), but rather to the epidemiological profile (classical hosts vs new hosts). This is because new/MSM hosts are all detected during acute infections.

166: "*Another promising perspective would be to combine sequence and incidence data. Although this could not be done here due to the limited sampling*"
Why does limited sampling prevent incidence data from being included? Presumably the fewer sequences that are available, the more informative the background incidence is? Or is the point that the sequences were all obtained in a short time window during which incidence was fairly constant?
Unfortunately it was impossible to use incidence data here because it is too parcellar. In general, the sampling rate is poorly known for HCV.

197: "*The list of Genbank accession numbers for all sequences is provided in Appendix.*"
Is the appendix missing?
Yes, we apologize for that. We added the accession numbers in the revised version.

Table 1: two columns are used to show the priors for gamma_1 and gamma_2, though they are the same; however only one column is used for a_1 and a_2. I think this is only a presentational difference rather than communicating some difference between the treatment of gamma_i and a_i, in which case using the same approach for both would be clearer.
This was modified in the new version.

228: "*R_0^(1) is assumed to vary over three time intervals*" this should just be two time intervals.
Thank you for the suggestion.

231: "*which we assume to have decreased R_0.*" It would be helpful to clarify here that this is through decreasing beta, keeping gamma_1 fixed.
This was modified in the text.

Supplementary Fig 2: I suggest having just three lines in the legend which match exactly those used, for clarity.
Yes, thank you for the suggestion.

Supplementary Fig 3: colour is explained by the colour axis, but there is no explanation for circle size.
We added the explanation in the revised version. The intensity of the color and the size of the circles are both proportional to the correlation coefficients.

Supplementary Fig 5: I do not understand the plot. It's unclear to me whether I'm looking at three different distributions of the parameter itself, or of the error in the parameter resulting

from three different processes (in which case I don't understand all three processes), or a mixture of the two.
We performed a leave-one-out cross-validation where we used 100 times a simulated phylogeny as a target phylogeny. We measured the median of each parameter distribution (the prior distribution, the prior distribution reduced after the simulations and the ABC posterior distribution) noted θ. For each type of distribution and each simulation scenario (100 target phylogenies) we computed the mean relative error such as :

$$MRE = \frac{1}{100}\sum_{i=1}^{100}\left|\frac{\theta_i}{\Theta} - 1\right| \text{ where } \Theta \text{ is the true value.}$$

In this figure, squares represent mean relative errors and their standard errors. Each row indeed represents a process: using only the prior distribution (in theory the worse we can do), using the prior distribution reduced after the simulations (i.e. removing parameter combinations that could not be simulated) and finally the posterior distribution resulting from the ABC analysis.

It would be helpful to include supplementary plots of prior and posterior together, for all parameters. For example if posteriors do not tend to zero well before the edge of the uniform priors chosen, this would suggest that the latter were overly restrictive.
We do show the prior and the posterior distributions for the parameters that we infer using our ABC method in Figure 2.

This is now clarified in the text.