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.

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.

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?

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.

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. 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.

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.

73: I suggest explaining the subscript 1 or 2 which is introduced here.

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.

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

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? 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?

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.

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? 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?

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

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.

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?

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.

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?

197: "*The list of Genbank accession numbers for all sequences is provided in Appendix.*"
Is the appendix missing?

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.

228: "$R_0^{(1)}$ *is assumed to vary over three time intervals*" this should just be two time intervals.

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.

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

Supplementary Fig 3: colour is explained by the colour axis, but there is no explanation for circle size.

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.

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.

Best wishes,
Chris Wymant