Dr Simon Dellicour
Rega Institute, KULeuven
Minderbroedersstaat 10
3000 Leuven, Belgium

17.01.2018

Dear Managing Board of ***PCI Evol Biol***,

We are grateful for the interesting and constructive comments from the Recommender and the two Referees on our preprint manuscript entitled "Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak".

We addressed these comments in a revised version of our preprint manuscript. We include below a point-by-point reply to the issues raised by the Referees. We believe that the revision has significantly improved our manuscript and we hope it can now be found suitable for recommendation.

Sincerely,

Simon Dellicour,
on behalf of all authors

Recommender:

This work offer another very nice illustration of the power of recent advances in phylodynamics when applied to a dataset with dense sampling and rich meta-data (here the location of the infections). It focuses on the recent devastating ebola virus (EBOV) outbreak in West Africa and extends an earlier enormous analysis by Dudas et al. of the by adding a continuous phylogeography approach. It also refines the interpretation of the results by pinpointing the importance of the three capital cities in the magnitude of the outbreak.

Reviewer #1 made some very detailed suggestions and raised a general question about the interpretation of the tree pruning. He/she and Reviewer #2 also made suggestions to broaden the perspective of the article, for instance by discussing epidemiological studies that did not involve phylodynamics to estimate the spread of the epidemics.

In addition to the comments made by the reviewers, I have a couple of my own.

1) Would it be possible to provide confidence intervals for Figure 1 (there are some for panels E, F and G but only for the unprunned tree). The reason why I ask this is because it could help assess the magnitude of the effect. It could also explain why the curves in panels A and B increase at first (I was expecting a steady decrease).
Answer: We understand the request to represent the uncertainty of the estimates. Concerning the density plots in panels A and B, these are based on the posterior distribution of Markov jumps, so already marginalizing over the evolutionary histories. The reason why these densities increase initially is because we consider jumps between administrative areas and the associated distances are determined between the major population centres in these areas. So, this will not include smaller transmission distances within administrative areas.
We attempted to include credible intervals for the different percentages reported on panels A to D, but the different options we explored overload the overall picture that already contains a lot of information. For clarity reasons, we thus finally decided to not include this additional information in the main figures. However, because one of the estimates was sensitive to conditioning on the MCC tree, we now include the alternative versions of Figures 1A-1D as Supplementary Information, and include a note on this sensitivity issue.
Concerning panels 1E-1G, adding credible intervals for all the effective population size dynamic curves (other than the one reported for the un-pruned tree) would lead to too much overlap. In addition, since we used empirical tree distributions for the pruned trees, the uncertainty would also not be adequately accommodated anyway.

2) Figure 1E is really beautiful! I was wondering if there is an explanation to the fact that recent case counts are below the inferred population size.
Answer: Thank you for appreciating our efforts in making these figures. Bearing in mind that different transformations were used for case counts and effective population sizes, there are indeed some differences for the recent estimates. We can only speculate about potential reasons. First, there could be an underreporting of cases in that recent time period, but it is difficult to understand why that would be the case for the period during which the epidemic is largely contained. Alternatively, it could be a period for which the coalescent has difficulties of inferring appropriate population sizes. In this period, prevalence declined such that only a handful of lineages persisted, but these are relatively divergent lineages. So the coalescent could be misled by the maintenance of a relatively large diversity through these remaining lineages, and a smoothed estimate of population size change may simply reflect the loss over time of these lineages. Because we cannot substantiate these explanations and since we only draw on the large-scale correspondence of the coalescent estimates, we did not discuss this further detail in the revised manuscript.

3) Figure 2 is also very nice but I expected to be able to find similarities because the sampled data should be the same. However, even the recent time points (in blue), which should all I guess be sampled, did not seem to be in the same place.
Answer: Because we made forward-in-time simulations and not backward-in-time simulations we cannot expect to reproduce the same coordinates as the sampling distribution. The point of this

posterior predictive procedure is to simulate a diffusion process using the same parameters as those estimated from the data on the same genealogies as estimated from the data. The simulations are then parameterised according to the estimates from the real data, but the outcome of these simulations does not need to match the realised pattern of spread. It serves as a 'landscape-unaware' model of spread against which we can contrast the realized pattern of spread.

4) About the model choice (HKY+GAMMA and skygrid) the authors refer to Dudas et al. but it seems that the model choice is not really justified over there, e.g. testing for the most appropriate substitution model. If there is actual support, it would be worth mentioning it. Regarding the details about the priors, I guess the xml files will be made available?
Answer: We indeed adopted of nucleotide substitution parameterisation from Dudas ***et al***. that is customised for this Ebola genomic data. The customisation lies in the partitioning of data, with 3 partitions for the three codon positions and an additional partition for the intergenic regions. For each of the 4 partitions, we specify an independent HKY model, an independent gamma distribution for among-site rate heterogeneity (ASRH), and we allow for relative rate differences among the partitions. We did not assess the model fit of this parameterization because:

1) There is a gigantic number of parameterisations to test against in terms of how the data can be partitioned and how these partitions are crossed with specific nucleotide substitution models including or not ASRH. Common substitution model testing procedures do not consider partitioning although adequately modelling rate heterogeneity is far more important than modelling the differences in the type of substitutions.
2) We strongly believe that it is more important to ask whether the estimates of interest are sensitive to the model choice than to ask which model from a pre-specified limited collection of models may fit the data best. Substitution model parameters are nuisance parameters in the inference of the diffusion process that we study. Nevertheless, the HKY model is likely to capture the most pronounced signal about differences in types of substitutions because the kappa estimates all range between 8 and 14 for the partitions.

All the xml files are now available online (https://github.com/ebov/space-time/tree/master/Analyses/continuousDiffusion).

Reviewer 1:

As a preface to my comments, I note that am not familiar with phylogeographic analyses. In the abstract the authors summarise their work as showing that
1. "long-distance dispersal events were not crucial for epidemic expansion",
2. "preventing viral lineage movement to single locations would, in most cases, have had little impact",
3. "urban areas – specifically those encompassing the three capital cities and their suburbs - represented major 'transit centers' for transmission chains, but preventing viral lineage movement to all three simultaneously would have only contained epidemic size to about one third"
4. there was "considerable heterogeneity in dispersal velocity through time", and 5. "announcements of border closures were followed by a significant but transient effect on international virus dispersal".
These conclusions are drawn by performing state reconstruction of the location of ancestral viruses in a large posterior set of timed phylogenies, allowing inferences about the movements of different viral lineages over time, and exploration of counterfactual interventions that would have stopped certain viral movements. I found that authors' work supported all five of these points, and was well described.

However, the major concern I have is with the relevance of these conclusions for public health interventions in practise. The authors' branch cutting model of preventing transmissions means that when one transmission is blocked, all people who would ultimately have their infection as a result of that transmission are assumed to be perfectly immune / not re-exposed. (The authors do "acknowledge that in the interpretation of intervention scenarios, we assume that all other efforts would have remained unchanged when an introduction to a specific location is hypothetically prevented", which is not the same as the no-second-exposure assumption.) This biases all estimates of intervention effect size downwards, by an unknown amount. The effect of this bias is relevant to the authors only considering interventions that block transmission with 100% effectiveness. Modelling by Hollingsworth et al (Nature Medicine 2006) showed that for flu, long distance travel bans had to be

implemented with >99% efficiency in order to slow epidemic growth from a time scale of days to weeks. How much worse would less-than-perfect interventions have been in the current study? I don't know how this question can be addressed with the authors' current method of cutting branches and removing all descendant lineages from the tree. Removing only x% of the descendants, or cutting the branch with only x% probability, would not capture the relevant effect - these would both trivially reduce the effectiveness to x% of its previous value. The relevant effect is that in the counterfactual where most but not all transmissions to a particular area are blocked, those transmissions that did get through can result in the same people as before becoming infected. If this effect cannot be modelled in the current work, it should at least be added as a strong caveat that as only 100% effective interactions have been considered, and impact depends strongly non-linearly on effectiveness, the effect of preventing long-distance transmission in reality is uncertain.

Answer: We fully agree that our approach of preventing transmission through pruning phylogenetic lineages assumes that relevant people were not exposed to the virus in other ways. Our approach conditions on the reconstruction of a realised pattern, so it is not a model that can take on a different course. We also considered cutting branches with a particular probability, but as the reviewer indicates, this would not adequately capture the effect of less-than-perfect interventions. We follow the reviewer's recommendation to better discuss the limitations of our approach, specifically that we consider no re-exposure and 100% effective interactions:

"***Our phylogenetic approach of assessing hypothetical containment strategies rests on a number of assumptions, with a 100% effectiveness of their implementation being an important one. While it would be straightforward to introduce a probability on the effectiveness of preventing the movement events we target, quantifying the corresponding impact using our phylogenetic measures may not be so relevant. Even if only a fraction of movements is allowed to escape prevention, the resulting transmission chains in the relevant area may have put everyone at risk of infection. In other words, our approach needs to assume that persons that were not infected by a particular lineage, because its transmission was halted, were not exposed to other transmission chains that were not contained. Our phylodynamic approach therefore offers a best-case scenario as starting point, and different degrees of effectiveness and its potential non-linear impact on outcome may be further examined using computational models. Further investigations will be important to assess whether interventions such as travel restrictions can in practice be implemented with reasonable success. In the case of air travel and influenza spread for example, travel restrictions were shown to be practically unfeasible to effectively contain the international spread of a pandemic (Hollingsworth et al. 2006)***".

Modest suggestions & concerns

The authors consider the impact of interventions on reducing epidemic duration. I think that most readers' will be intuitively think that reducing duration is a good thing, because of an assumed link between duration and the final size. However the authors also consider the impact of interventions on the epidemic's final size directly - which is what we really care about - and for a fixed size, reducing epidemic duration is a bad thing. The same sized epidemic concentrated into a shorter time period gives any additional interventions outside of those being modelled less time in which to act. This should be clarified; ideally statements about duration should only be made at the same time as statements about size, less the conclusion be misinterpreted. For example, the authors comment that halting introductions into any one of the three capitals would not have reduced the epidemic duration by much (lower part of Figure S1). The reader will naturally interpret this as "there would have been little point in intervening in only one of the capitals". The authors do not point out that halting introductions into Freetown alone would have reduced the epidemic size by around 40% (upper part of Figure S1). Reducing epidemic size is what we care about - why comment on duration but not size?

Answer: We agree and we now note that epidemic size, and not duration, is the most relevant measure to evaluate the impact of containment strategies before reporting the results and at the end of the results.

Could the authors mention whether a sampling fraction of much less than 100% is expected to affect the inferred dispersal velocity, and if so how? It's not clear to me but I could imagine it results in an overestimate: the serial interval is defined specifically for one infected individual, yet many (95%) of

individuals are missing from the tree, so viral lineages are spending a lot of time outside of sampled hosts.
Answer: With a sampling fraction of about 5% of the known cases, viral lineages are indeed spending time outside of the sampled hosts, but we believe the relevant question is how much time they have spent outside of the reconstructed history. Because our sampling shows a very good correlation with case counts through time in administrative areas, it is likely to adequately capture the large-scale transmission history. So, many additional samples will have come from local transmission chains and would increase the density of tips in our phylogenetic reconstructions, but that would not add considerable much ancestral history to our reconstruction. Because of this and the fact we only took a single representative sequence for a cluster of admin-specific sequences, our dispersal velocity estimate will reflect the rate of lineages that largely spread across different admin regions (and not within admin regions). We further address this in reply to the comment about pruning short-distance transmission.

Page 5: "The picture that emerges from our phylogeographic analyses is one of multiple moving targets" Evidence for this (i.e. a metapopulation) would be showing that locations of high burden appear and disappear at different times in different places. It may be that the geographically annotated phylogenies show this, but I don't think the authors have shown that here, aside from the example in Fig 2 suggesting a tendency of the epicentre to drift over time. This picture could be shown by plotting the number of reported cases in different locations as a function of time (without need for phylogeographic analysis).
Answer: The demonstration of metapopulation dynamics was indeed more explicit in our previous work, and as part of this, we have already created an animation of both the phylogeographic reconstruction and the case count evolution (the latter is shown by the colour intensity for the administrative regions):
https://github.com/ebov/space-time/blob/master/Visualizations/EBOV_animation.HD.264.mp4
We now explicit cite our previous study when we refer to the metapopulation dynamics.

Page 6: "we remove sequences such that monophyletic clusters of sequences sampled from the same administrative region are only represented by a single sequence. Such clusters would largely represent dispersal within administrative regions, which will be determined by the noise assigned to their location within an administrative region" My concern with this procedure is not the uncertainty in the location of the single sequence representing the cluster (which is what I understand by the 'noise' comment), it is the fact that by preferentially removing sequences that seem to be connected by short-distance transmission before fitting a model of diffusion, the authors will upwardly bias the estimated rate of diffusion. I suggest testing for the presence of this bias. If present, and if including all such sequences renders the problem computationally unfeasible, could the rate of diffusion within such close clusters can be estimated separately, and merged somehow with the estimation where these clusters have been collapsed?
Answer: We believe the Reviewer raises an important point. First, we would like to clarify that for most sequences, there were no specific geographic coordinates available within the administrative area. Therefore, estimating diffusion rates based on clusters of sequences that only constitute samples within an area would not be possible. It is because of the absence of this information as well as the computational challenges of integrating over the administrative areas for unknown locations that we decided to reduce those clusters to one representative sequence. As a result, our procedure ignores a lot of short-distance transmission within administrative areas. To investigate to what extent this would affect our dispersal rate estimate, we now analysed a data set of sequences for which more precise geographic coordinates were available and for which we did not restrict monophyletic clusters of sequences from the same administrative area to a single representative sequence. While this data set therefore accommodates transmission within administrative areas, and also differs in the time interval and total area of sampling (Sierra Leone), the dispersal velocity estimates are remarkable consistent with our previous estimates. This offers reasonable reassurance that our procedure does not result in strong biases. This additional analysis and related results are now described in the Methods, reported in the Results and included as Supplementary Information.

Page 6: "We prune each of the posterior trees by removing the same extant taxa as identified in the MCC tree by the pruning selection process." Why not identify branches with viral lineage movement over distances >d in each posterior tree separately, and then prune the descendant subtree? (Why only

identify these branches in the MCC tree and then prune the same taxa in all posterior trees?) This would seem to make better use of the fact that each posterior tree represents a slightly different evolutionary history, and we want to integrate over all possibilities.
Answer: We indeed had to remove exactly the same taxa across all trees in the posterior so that we could also use the posterior genealogies to infer the demographic trajectories (our proxy of viral effective size through time). We would not be able to do this coalescent inference while averaging over empirical trees that contain different taxa. This is now explicitly mentioned in the Methods section. In addition, we have now examined how sensitive the tree height and tree length summaries are to conditioning on the MCC tree. We report the results in the Supplementary Information in the form of two new figures: one that includes credible intervals for reductions in epidemic size and duration and one that is based on pruning that does not condition on the MCC tree. For the less important epidemic duration measure, we noticed that the reduction associated with preventing spread to administrative locations with >1,000k people is highly uncertain and not represented well by the MCC tree. We also highlight this in the main manuscript and thank the Reviewer for encouraging us to examine this.

Page 6/7: the authors describe a procedure for comparing the number of border crossings inferred with the number expected by chance (in a border-unaware simulation), and plot the level of evidence for a discrepancy between the two in Figure 4. Some kind of direct plot of the number of border crossings over time (or the fraction of movements that cross a border, to normalise to the growing epidemic size) would provide a more intuitive visualisation of the effect the authors are trying to test for - a transient decrease after border closures - than Figure 4. As there, the rate of crossing within-country borders could be plotted for comparison. Quantifying the level of evidence is clearly important, but the plot I suggest could be included in addition. This would also show the magnitude of the effect, about which no information is currently given (only the frequency with which N_inferred < N_simulated is presented).
Answer: We agree and plots with the frequency of border crossing for both the estimated and simulated distributions are included as Figure S2.

Minor points

The authors could consider citing Ratmann et al. (Science Translational Medicine 2016), who also identified transmission patterns using viral phylogenies and quantified the impact of removing transmissions between certain groups of individuals.
Answer: Thank you for the suggestion, we now refer to this study when starting our exposé on the phylogenetic pruning procedure.

The authors refer to timed phylogenies (with internal annotation of geographic states) as transmission trees. Transmission trees are not the same as phylogenies.
Answer: We have now replaced the expression “transmission tree” by “phylogenetic tree”.

I was confused in a few places whether the velocity being discussed was the velocity of the epidemic wavefront or the velocity of an individual viral lineage.
Answer: We have edited the text to avoid the confusion between these two concepts (“velocity of the epidemic wavefront” is now clearly distinguished from “the weighted/mean dispersal velocity”).

I read the first two sentences of the abstract as stating that viral genomic data is critical for viral molecular epidemiology. This is essentially tautological, analogous to saying that measurements of human height are critical for studies of human height. A statement of interest would be about how important the study conclusions are. If the intended point is the importance of being able to do viral molecular analyses rapidly, the statement should be reworded to clarify.
Answer: The second sentence has now been modified.

Page 1: "but preventing viral lineage movement to all three simultaneously would have only contained epidemic size to about one third". A three-fold reduction in the total size of a large epidemic is substantial; the "but... only" sounds odd.
Answer: We corrected the text.

Page 1: "the impact that specific intervention strategies made, had or could have made". The first two items in this three-item list are identical.
Answer: Corrected.

Figure 1 C&D: "% of locations in each range" - does this mean that all samples from the same location only contribute 1 to each bin? If so, this doesn't seem very informative. Or should this be "% of samples in each range"?
Answer: Indeed, it was actually the percentage of samples. This mistake has been corrected.

Page 2: "this GLM approach identified a gravity model of transmission". A modelling analysis does not 'identify' the correct model, it tests multiple models and may find that one is a better description of reality than the others. All models are wrong, but some are useful.
Answer: This has been reworded.

Figure 2: the legend states that "nodes are coloured according to a colour scale" for which the minimum and maximum are stated, but what the colour itself actually represents is left unsaid. It's fairly obvious but this could easily be clarified.
Answer: This has been clarified.

Page 3: "Although about 27% of the genome samples were from these administrative areas" And what fraction of the total reported cases? It is mentioned in the introduction that sampling intensity correlates well with the infection burden, but it would be helpful to clarify for this example.
Answer: This specific information has been added.

Figure 3: the x axis and the legend both state that panel C shows dispersal velocity; I think it shows distance. The legend also contains "1,000 trees sampled sampled" (duplicated "sampled").
Answer: Panel C shows the density for velocity (in km per infection). The duplicate "sampled" has been removed.

Page 4: "This result shows that preventing viral lineage movement to these locations halts the dynamic spread of lineages, which in turn continue to generate numerous clusters of cases in other locations, even in different capitals. This stands in contrast to a model of separate, independent and local chains of transmission in each capital city." A simpler summary would be "This result shows that there was transmission between the capitals." Transmission between capitals (including via external locations as intermediates) is necessary and sufficient for the observation, given the branch-cutting approach for modelling blocked transmissions.
Answer: We agree, and have modified the text accordingly.

Page 4: "The same extent of spatial spread is not achieved for the data set restricted to dispersal events <250 km, indicating that relatively long-distance dispersal events contributed to the maximum epidemic wavefront distance." How much smaller is the spatial spread? (How big is the contribution of these long-distance events?)
Answer: This information has been added in the text: "***Our phylogeographic estimates of the epidemic wavefront through time indicate that EBOV spread up to ~500 km from its location of origin in about 8 to 9 months (Fig. 3A). With a maximum wavefront distance of ~400 km, the same extent of spatial spread is not achieved for the data set restricted to dispersal events <250 km, indicating that relatively long-distance dispersal events contributed to the maximum epidemic wavefront distance***".

Page 5: "short-distance dispersal realised by human mobility" All mobility here is human. Do the authors mean mobility on foot? If so this seems unwarranted - there is no examination of the method of transport. Cars, bikes etc. could have been used to cover short distances. Best just to leave this statement at "short-distance dispersal" I think.
Answer: Corrected.

Page 5: "If viral lineage movement to each of the capitals was prevented, beginning from the onset of the epidemic, then epidemic size could have been reduced by 15% to 37%. This emphasises the importance of these cities". The size of the resulting reduction needs to be balanced against the total number of cases that occurred in the cities. If the latter is 15% to 37%, the observation would be compatible with the cities exporting no cases at all.
Answer: For clarity, we now contrast this result with the percentages of reduction obtained when preventing lineage movement to all the capitals together: "***If viral lineage movement to a single capital could have been prevented, beginning from the onset of the epidemic, then epidemic size could have been reduced by 15% to 37%. In contrast, preventing lineage movement to all the capitals reduced epidemic size to about one-third, while their sample size percentage and case count percentage are 28% and 39%, respectively. This theoretical result emphasises the importance of urban transmission, but at the same time, it indicates that no single capital was critical for the maintenance of all co-circulating lineages***".

Page 5: "justify the use of a continuous diffusion process... at least for relatively restricted geographic scales" Preferable to complete the latter sentence with "such as...". A few hundred kilometres?
Answer: We prefer not to associate a particular distance to this, as it remains to be determined how large the scale can be. Moreover, the appropriate scale will also depend on the mobility dynamics in the area of interest. We refer to a 'restricted scale' to remain cautious.

Page 6: "When unique sampling coordinates are not available for every sequence... we associate a random coordinate within the administrative area of sampling to each sequence." Was this random coordinate drawn once only for each such sequence, fixing one realisation of a stochastic effect across many replicates, or re-drawn many times (for each realisation of the diffusion model)? That latter would seem to be better, as the advantage of the kind of Bayesian analysis used by the authors is to integrate over uncertainties in the model.
Answer: In a first attempt, we indeed tried to integrate out every location over the relevant administrative area. While this approach works well for a restricted number of tips, we were confronted with significant mixing issues when trying to do this for all tips. We are currently working on new integration techniques for unobserved locations that we will hopefully be able to use for such purposes in the future.

Page 6: the authors describe the form of the distributions used for priors, but not the parameters of the priors. Could all relevant information be captured by providing the BEAUti file as supplementary information? If so this would greatly facilitate replication of the analysis.
Answer: Yes, the xml files will be made available.

Page 6: the authors provide no information about how the wavefront was calculated.
Answer: We have now added a sentence to describe this as well as a reference to the Methods section.

Page 7: "in continental African" -> in continental Africa.
Answer: Corrected.

Page 7: "As illustrated in Figure 3, posterior and posterior predictive diffusion histories are roughly similar except for the position and orientation of branches in the West African study area." Consider clarifying what it is about these histories that is similar, given that the position and orientation of branches seems to be the dominant feature. Consider removing the final "in the West African study area", which slightly confuses the sentence, hinting that branches are different inside West Africa but similar outside of it.
Answer: We have now clarified this sentence (and this should actually refer to Figure 2): "***Figure 2 illustrates the difference in position and orientation of branches between a diffusion history reconstructed from the data and a diffusion history simulated using our posterior predictive simulation procedure***".

In supplementary Figure S1, some of the filled bars reach greater values that the corresponding open bars, e.g. for epidemic size for Freetown. How is possible that halting only those introductions into

Freetown that occur after July 2014 has greater reduction on the epidemic size than halting all introductions into Freetown at any time?
Answer: We apologize for the lack of clarity about this. In the case of the delayed intervention strategies, percentages of reduction were computed relative to period impacted by the intervention strategy (after June 2014). This explains why we can observe higher reductions in this case. This has now been clarified (in the Results section as well as in the legend of Figure 1).

Review 2:

The study by Dellicour and colleagues makes use of phylodynamic analyses for studying the spatial spread of Ebola during the 2013-2016 outbreak in West Africa. The authors extended their previously published phylogeographic framework to examine 1) the potential effect of intervention strategies - such as border closures - and 2) the process of spatial spread by introducing a continuous diffusion process (as opposed to the discrete approach in their earlier analysis). The methods are state-of-the-art and described in sufficient detail. The main findings of the study suggest that the Ebola epidemic was mainly driven by short- rather than long-distance dispersal. Furthermore, the study corroborates the notion that urban transmission was a major contributor to the characteristic spatial transmission dynamics that was observed in West Africa. I found the study rather technical and applying its findings in public health practice is maybe somewhat limited. However, the study is certainly a valuable contribution to the field of phylodynamics and provides an excellent example how genomic analyses can be used to infer the spatial spread of epidemics.

What I was missing a bit was a deeper discussion and comparison of the results to other studies outside the field of phylodynamics that investigated the spatial spread of Ebola and the impact of control interventions (e.g., border closures). The authors briefly mention two key papers using gravity-type models by Backer & Wallinga (ref. 27) and Kramer et al. (ref. 28). Others have estimated the velocity of Ebola spread at 1004 km per year (Zinszer et al., 2015, Lancet Infect Dis, PMID: 26333328) which seems to be in rough agreement with Fig. 3D.
Answer: We agree and added a comparison to specific modelling results to the discussion. In addition, we now also better integrate modelling studies in the first paragraph of the introduction. Specifically, we refer to containment efforts, how they have been investigated using mathematical modelling, and how sequence data and phylodynamic approaches may complement these to evaluate long-range interventions.

I also have a question related to how the authors call this velocity (mean dispersal distance per infection). Shouldn't it be per generation? In my view, it is not a single infection that spreads, but an epidemic that expands over subsequent generations.
Answer: We are afraid that the word 'generation' may be too vague, as there is host generation, viral generation, transmission generation, etc. If we say 'per infection', we do not imply that it is a single infection that spreads just as we would not imply a single generation if we would use 'per generation'.

Minor comments:

- Methods: The authors associate a random coordinate within the entire administrative area for sampled sequences that have the same geographic coordinates. I was wondering whether this assumption could introduce any sort of bias. For example, if all sequences came from exactly the same place in an otherwise large area, wouldn't associating random coordinates suggest wider spatial spread than what effectively happened?
Answer: For a detailed answer, we refer to our reply to the related comment by Reviewer 1. In short, we have now performed an additional analysis to show that our rate estimate is robust to this (as reported in the new Appendix in Supplementary Information).

- Fig. 1A: What is the dashed line on the peak of the distribution of lineage dispersal distances supposed to show?

Answer: It corresponds to the distribution of distances associated with lineage dispersal occurring after June 2014. We apologize for the fact that this information was lacking in the previous version of the legend.

- Fig. 1C: I could not find the dashed line that is described in the figure caption.
Answer: We have now solved this issue.

- Fig. 3: The word "sampled" appears twice in the second sentence.
Answer: Corrected.

- Fig. 4: I could not find a reference to this figure in the main text of the paper.
Answer: Corrected.