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.

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?

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.

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).

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?

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.

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).

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.

The authors refer to timed phylogenies (with internal annotation of geographic states) as transmission trees. Transmission trees are not the same as phylogenies.

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.

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.

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.

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.

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"?

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.

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.

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.

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").

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.

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?)

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.

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.

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?

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.

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.

page 6: the authors provide no information about how the wavefront was calculated.

page 7: "in continental African" -> in continental Africa

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.

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?