nature portfolio

Peer Review File

Characterizing Spatial Epidemiology in a Heterogeneous Transmission Landscape Using the Spatial Transmission Count Statistic

Corresponding Author: Dr Justin Bahl

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The work of Lyu et al describes a phylogeographic analysis of SARS-CoV-2 transmission in the state of Texas, considering the Delta variant of the virus. The manuscript shows that the virus was imported into urban centres, where it was maintained, and from which it seeded epidemics in rural areas. Two simple metrics are used to described whether a region acted more as a source or sink for inter-region transmission events, and to describe the important of imports to an area relative to local transmission.

The work is generally of good quality but the novel methods described need further work to establish what is meant by their output. Further, the work needs to be placed more clearly in the context of previous publications describing SARS-CoV-2 outbreaks in rural and urban areas.

1. The use of subsampling is a positive step. Was this repeated (i.e. to look at variance across different subsamples)?

2.. Given that they are highlighted in the abstract, it would be helpful to have more detail on the meaning of the two statistics. With the Source Sink score, it is noted that a score close to 1 indicates that the number of exports is significantly higher than the number of imports. How close to 1 does the score need to be to attain significance, and what does this mean for the results presented? For example, is the score of 0.147 for Houston significantly higher than the score of 0 for Dallas-Fort Worth? The relative values of the numbers provide a heuristic sense of how a region may have behaved, but I am not sure if the results go beyond this: Are the reported differences meaningful?

Similarly for the Local Import score. A ‘value close to 1 indicates that the outbreak is driven by external introductions’. Is this also heuristic, or can significance be attributed to the differences found? Simulated data would provide one way to assess this.

3. I found the mention of population heterogeneity confusing (line 36). Firstly, there are very many papers which consider what I would think of as population heterogeneity. Ram and Schaposnik (Sci Rep 2021) consider heterogeneity by age, which is one way of dividing a population. Secondly, I’m not sure if what is considered here is population heterogeneity, so much as regional heterogeneity: The differences are between environments, rather than between people.

4. I did a very brief literature search for papers on urban-rural differences in SARS-CoV-2 epidemiology. None of the first few hits I came up with were cited, which made me think there is likely greater scope for engagement with the previous literature in this area, even where previous studies have not used sequence data.

Cuadros et al., Ann Epidemiol, 2021 promotes greater interventions in rural, rather than urban areas.
Wells et al, J Roy Soc Interface, 2020 find a link between epidemic size and local connectivity.
Tang et al., NPJ Viruses, 2023 study SARS-CoV-2 spread between rural and urban communities.
Polo et al, Public Health, 2022 simulate outbreaks at the urban-rural interface.

Minor comments:

Line 110: Is unemployment a complication of COVID-19?
Line 347: I think 'Nextstain' is a typo.

Reviewer #2

(Remarks to the Author)
In their work "Characterizing Spatial Epidemiology in a Heterogeneous Transmission Landscape Using a Novel Spatial Transmission Count Statistic" the authors define two spatial transmission count statistic that can be used understand how a virus spreads across within a region and identify which locations (e.g., provinces, counties, etc) act as a source of viral transmission or a sink over time.

I think the work is interesting and could be useful to characterize the spread of a pathogen in a given study area however, I do have some major and minor remarks.

Major comments:

In line, 124 - 125, the authors wrote "we developed a novel phylogeographic pipeline to mitigate genome sampling bias, infer viral phylogenetic relationships, and summarize transmission patterns across multiple scales". This is misleading because of the following:

1. the proportional subsampling scheme reported in the manuscript, 'subsamplerr,' is an R-based script, which is very similar to 'subsampler' a python-based language already published by Brito et al., in 2021.

https://github.com/andersonbrito/subsampler/tree/master
https://doi.org/10.1016/j.cell.2021.03.061

Strikingly, both share similarities in their analytical procedures, package name, and GitHub description: 'subsampling genomic data based on epidemiological time series data'.

2. the phylogeographic pipeline used was based on previous tools and workflows including IQ-TREE, Treetime, all of which have been employed in numerous studies and are part of the well-known open-source project, Nextstrain.

The novelty of the study relies on the definition of the two metrics "Source Sink Score", and "Local Import Score" to characterize the imports, exports and local transmission within a study area. Therefore, I strongly recommend that the authors revise the paragraph to emphasize this contribution. Additionally, it is very important that the authors cite 'subsampler' and discuss how 'subsamplerr' advances or differs from 'subsampler'.

In lines 40 – 43, the authors state that the transmissions count statistics i.e., "Source Sink Score" and "Local Import Score" which are estimated from the phylogenetic tree using iq-tree + treetime, help to deal with large dataset with poor genetic diversity. Could the authors please elaborate on this? Specifically, how do these metrics assist in managing issues related to low genetic diversity in large datasets?

In lines 56 – 57, on the significance statement, the authors wrote that they "investigate the role of population density and distribution on phylodynamic inference". In their work, the authors studied the spread of delta variants between 25 metropolitan areas in Texas, which have been categorized in RUCC1 – RUCC3 according to their population density and 1 rural area. While the pattern observed could probably be explained by population density and distribution, such variables were not formally tested. Specifically, such data was not integrated into the phylogeographic analysis, which the authors could have done using, for example, a Bayesian discrete trait analysis with a generalized linear model. This would have allowed the authors to assess the impact of population density and distribution on the spread of the virus, specifically, on the dispersal frequency of the virus, rather than on the phylodynamic inference. Please correct the statement accordingly.

Minor comments

Line 164, no confidence interval for the date of the introduction event?

Regarding the spatial network. How was the weighted network built? Any specific tool?

To build the spatial network, the authors performed ancestral location reconstruction from the global phylogeny from Figure 1 using "26 location traits: 25 metropolitan areas and 1 combined rural area) (Figure S2)". Given that the non-Texas sequences were also included, shouldn't there be 27 location states? A color legend is missing in Figure S2. Please add it for clarity.

In Figure 2, the authors overlay the network analysis on the Texas map. While the analysis was performed at the metropolitan areas level, the map is showing the Texas counties. Could the metropolitan areas instead of the counties be represented on the map? If not, how do the metropolitan areas relate with the Texas counties?

In addition, in Figure 2 different color transparencies are used to show the different weights of the transition events (i.e.,

magnitude of the viral flows). However, as the lines overlap on the map, it is hard to distinguish the colors. I suggest using different line widths instead of varying color transparencies to improve clarity.

Lastly, on lines 297 to 308 the authors discuss about the advantages of the methodology implemented in their study which provides results at minimum cost. What about the limitations or differences with other methods in terms of assessing uncertainty and joint phylogenetic and phylodynamic inference?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I am happy with the response received to my previous comments

Reviewer #3

(Remarks to the Author)
N/A

We thank the reviewers for their careful read and thorough comments. We have addressed all concerns in the revised manuscript and detailed our responses below (in Blue). The line numbers in our response correspond to changes in the revised manuscript. We feel that the revised manuscript is much improved and hope that our responses address all concerns.

Sincerely

Leke Lyu and Justin Bahl

Referee expertise:

Referee #1: viral transmission, epidemiology, genomics

Referee #2: spatial epidemiology, spatial modelling, RNA viruses

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The work of Lyu et al describes a phylogeographic analysis of SARS-CoV-2 transmission in the state of Texas, considering the Delta variant of the virus. The manuscript shows that the virus was imported into urban centers, where it was maintained, and from which it seeded epidemics in rural areas. Two simple metrics are used to described whether a region acted more as a source or sink for inter-region transmission events, and to describe the important of imports to an area relative to local transmission.

The work is generally of good quality but the novel methods described need further work to establish what is meant by their output. Further, the work needs to be placed more clearly in the context of previous publications describing SARS-CoV-2 outbreaks in rural and urban areas.

1. **The use of subsampling is a positive step. Was this repeated** (i.e. to look at variance across different subsamples)?

We conducted a sensitivity analysis (Figure 4) by generating nine additional genome datasets for Texas. These datasets were created using the same proportional sampling scheme as the original dataset. We then ran the phylogeographic workflows on each dataset (Methods: Lines 218-224). We found All 10 replicates supported RUCC-1 regions as the predominant viral sources, as these regions consistently showed the highest Source Sink Scores (Results: Lines 318-327). The robustness of both the Source Sink Score and the Local Import Score varied across regions. We believe that data availability

and volume significantly impact the robustness of these metrics (Discussion: Lines 389-397).

2. Given that they are highlighted in the abstract, it would be helpful to have more detail on the meaning of the two statistics. With the Source Sink score, it is noted that a score close to 1 indicates that the number of exports is significantly higher than the number of imports. How close to 1 does the score need to be to attain significance, and what does this mean for the results presented? For example, **is the score of 0.147 for Houston significantly higher than the score of 0 for Dallas-Fort Worth?** The relative values of the numbers provide a heuristic sense of how a region may have behaved, but I am not sure if the results go beyond this: **Are the reported differences meaningful?**

Similarly for the Local Import score. A 'value close to 1 indicates that the outbreak is driven by external introductions'. Is this also heuristic, or can significance be attributed to the differences found? Simulated data would provide one way to assess this.

Thank you for this comment. We appreciate the opportunity to clarify the interpretation of the Source Sink Score and Local Import Score. These metrics are heuristic and designed to provide a relative understanding of how regions behave within the context of local-scale outbreak. The role of a population as a source or sink evolves dynamically as the outbreak progresses and host immunity develops. Therefore, the Source Sink Score should be interpreted as a comparative measure, emphasizing relative differences between regions rather than absolute values (Discussion: Lines 362–364).

The sensitivity analysis tests the robustness of these metrics, offering insight into when a Source Sink Score is similar or significantly different between areas. For instance, in 6 out of 10 replicates, the Source Sink Score for Houston was higher than for Dallas-Fort Worth, indicating that the values are similar and may not represent a significant difference between the two regions. In contrast, all 10 replicates indicate that the Source Sink Score for Houston is higher than all rural regions, suggesting a more substantial difference (Table S6). While we might be tempted to interpret the difference between Houston and San Antonio (Houston > San Antonio for all 10 replicates) as Houston is a more important source than San Antonio. However, the Local Import Score suggests that the outbreak in San Antonio is not dependent on introductions. We, therefore, feel that comparing the magnitude of the metrics would be inappropriate. Both regions are sources to rural areas. The difference between Houston and San Antonio is harder to interpret.

Additionally, as estimated statistics, the Source Sink Score and Local Import Score have the potential to be integrated into a Bayesian phylodynamic framework. This integration would enable the calculation of Bayesian Credible Intervals for these scores, providing a robust measure of their uncertainty. This approach may assist in interpreting the difference among Source Sink Scores between regions, thereby allowing for more precise regional comparisons (Discussion: Lines 399–407). We have further explored this integration in follow-up work currently available on medRxiv (https://www.medrxiv.org/content/10.1101/2024.09.18.24313896v1).

3. I found the mention of population heterogeneity confusing (line 36). Firstly, there are very many papers which consider what I would think of as population heterogeneity. Ram and Schaposnik (Sci Rep 2021) consider heterogeneity by age, which is one way of dividing a population. Secondly, I'm not sure if what is considered here is population heterogeneity, so much as regional heterogeneity: The differences are between environments, rather than between people.

We adopted the term "regional heterogeneity" as it more accurately captures the diversity observed across different regions in our analysis (Abstract: Line 37).

4. I did a very brief literature search for papers on urban-rural differences in SARS-CoV-2 epidemiology. None of the first few hits I came up with were cited, which made me think there is likely greater scope **for engagement with the previous literature in this area**, even where previous studies have not used sequence data.

Cuadros et al., Ann Epidemiol, 2021 promotes greater interventions in rural, rather than urban areas.

Wells et al, J Roy Soc Interface, 2020 find a link between epidemic size and local connectivity.

Tang et al., NPJ Viruses, 2023 study SARS-CoV-2 spread between rural and urban communities.

Polo et al, Public Health, 2022 simulate outbreaks at the urban-rural interface.

Thank you for the suggestion. These papers are very helpful in explaining the socioeconomic and demographic differences between rural and urban areas. We included them in both the introduction and discussion sections to better engage with the existing literature.

Minor comments:

Line 110: Is unemployment a complication of COVID-19?

We removed this statement.

Line 347: I think 'Nextstain' is a typo.

Corrected.

Reviewer #2 (Remarks to the Author):

In their work "Characterizing Spatial Epidemiology in a Heterogeneous Transmission Landscape Using a Novel Spatial Transmission Count Statistic" the authors define two spatial transmission count statistic that can be used understand how a virus spreads across within a region and identify which locations (e.g., provinces, counties, etc) act as a source of viral transmission or a sink over time.

I think the work is interesting and could be useful to characterize the spread of a pathogen in a given study area however, I do have some major and minor remarks.

Major comments:

In line, 124 - 125, the authors wrote “we developed a novel phylogeographic pipeline to mitigate genome sampling bias, infer viral phylogenetic relationships, and summarize transmission patterns across multiple scales”. This is misleading because of the following:

1. the proportional subsampling scheme reported in the manuscript, ‘subsamplerr,’ is an R-based script, which is very similar to ‘subsampler’ a python-based language already published by Brito et al., in 2021.

https://github.com/andersonbrito/subsampler/tree/master
https://doi.org/10.1016/j.cell.2021.03.061

Strikingly, both share similarities in their analytical procedures, package name, and GitHub description: ‘subsampling genomic data based on epidemiological time series data’.

2. the phylogeographic pipeline used was based on previous tools and workflows including IQ-TREE, Treetime, all of which have been employed in numerous studies and are part of the well-known open-source project, Nextstrain.

The novelty of the study relies on the definition of the two metrics “Source Sink Score”, and “Local Import Score” to characterize the imports, exports and local transmission within a study area. Therefore, I strongly recommend that the authors revise the paragraph to emphasize this contribution. Additionally, it is very important that the authors cite ‘subsampler’ and discuss **how ‘subsamplerr’ advances or differs from ‘subsampler’.**

We have now revised the paragraph to properly cite the work by Brito et al. and the Nextstrain platform (Introduction: Lines 131–133). Additionally, we have highlighted Brito’s work when introducing the subsampling scheme in the methods section (Methods: Lines 148–152). We apologize for the oversight in citation and the misleading statement. *Subsamplerr* builds upon the algorithm developed by Anderson F. Brito. We adapted their work into an R package to facilitate the visual exploration of sampling heterogeneity and the implementation of proportional sampling schemes.

In lines 40 – 43, the authors state that the transmissions count statistics i.e., “Source Sink Score” and “Local Import Score” which are estimated from the phylogenetic tree using iq-tree + treetime, help to deal with large dataset with poor genetic diversity. Could the authors please elaborate on this? Specifically, **how do these metrics assist in managing issues related to low genetic diversity in large datasets?**

We agree that the statement was not appropriate, and we have removed it from the draft to avoid any confusion. Densely sampled SARS-CoV-2 datasets can complicate

phylogenetic analysis due to the combination of large dataset size and few mutations. This creates a complex tree search, with many local optima on the likelihood surface. While our method is useful for translating phylogenetic results into epidemiological inference, it does not address the challenges of phylogenetic inference itself.

In lines 56 – 57, on the significance statement, the authors wrote that they **"investigate the role of population density and distribution on phylodynamic inference"**. In their work, the authors studied the spread of delta variants between 25 metropolitan areas in Texas, which have been categorized in RUCC1 – RUCC3 according to their population density and 1 rural area. While the pattern observed could probably be explained by population density and distribution, such variables were not formally tested. Specifically, such data was not integrated into the phylogeographic analysis, which the authors could have done using, for example, a Bayesian discrete trait analysis with a generalized linear model. This would have allowed the authors to assess the impact of population density and distribution on the spread of the virus, specifically, on the dispersal frequency of the virus, rather than on the phylodynamic inference. Please correct the statement accordingly.

Due to the format requirements of *Communications Medicine*, we removed the significance statement, and portions of its content have now been incorporated into the abstract (Abstract: Lines 58–63). We agree with the reviewer that the original statement regarding phylodynamic inference was not appropriate, and it has been removed from the draft.

The primary objective of this study is to develop a novel method for translating phylogeographic signals into epidemiological insights. We applied these new statistics to compare and better understand transmission heterogeneity between urban and rural areas in Texas. However, we did not aim to explore the underlying causes of this heterogeneity, such as the impact of population density and distribution on virus spread. While we recognize this is an important question, it falls outside the scope of our current study. Our focus is on developing and applying novel metrics to assess transmission differences across regions.

Still, we appreciate the reviewer's suggestion regarding Bayesian phylogeographic analysis, which is indeed inspiring. As estimated statistics, the Source Sink Score and Local Import Score have the potential to be incorporated into a Bayesian phylodynamic framework. We have expanded on this idea as a future direction in the discussion section (Discussion: Lines 399–407). Additionally, we have further explored this integration in a follow-up study, which is currently available on medRxiv, and we would be happy to share it with you (https://www.medrxiv.org/content/10.1101/2024.09.18.24313896v1).

Minor comments

Line 164, no confidence interval for the date of the introduction event?

We added the estimation on confident interval (Results: Lines 262-264).

Regarding the spatial network. How was the weighted network built? Any specific tool?

We constructed a weighted, undirected network to capture the viral flow between metropolitan areas in Texas. Each metropolitan area is represented as a node, and the edge carries weight corresponding to the spatial transmission counts. We processed the various network data objects using the 'igraph' package in R. We revised the method section to make it more readable (Methods: Lines 209-216).

To build the spatial network, the authors performed ancestral location reconstruction from the global phylogeny from Figure 1 using "26 location traits: 25 metropolitan areas and 1 combined rural area) (Figure S2)". **Given that the non-Texas sequences were also included, shouldn't there be 27 location states?** A color legend is missing in Figure S2. Please add it for clarity.

To understand the spatial transmission of SARS-CoV-2 in Texas, we estimated ancestral location states on the phylogeny using 27 location traits: one contextual trait and 26 subregions of Texas (comprising 25 metropolitan areas and one combined rural area). We have corrected the description in the results section (Results: Lines 270–272), and we added the color legend in Figure S2.

In Figure 2, the authors overlay the network analysis on the Texas map. While the analysis was performed at the metropolitan areas level, the map is showing the Texas counties. **Could the metropolitan areas instead of the counties be represented on the map?** If not, how do the metropolitan areas relate with the Texas counties?

In addition, in Figure 2 different color transparencies are used to show the different weights of the transition events (i.e., magnitude of the viral flows). However, as the lines overlap on the map, it is hard to distinguish the colors. **I suggest using different line widths instead of varying color transparencies to improve clarity.**

The original Figure 2 has been updated and is now presented as Figure 2A. In the revised figure, county boundaries have been removed, and line widths are now proportional to transmission counts, improving clarity. Overall, the figure is more readable and visually communicates the data more effectively.

Lastly, on lines 297 to 308 the authors discuss about the advantages of the methodology implemented in their study which provides results at minimum cost. What about the limitations or differences with other methods in terms of assessing uncertainty and joint phylogenetic and phylodynamic inference?

Thank you for pointing this out. We have expanded the discussion to address the limitations of our study. Specifically, we have noted potential biases in the estimation of local transmission counts, and our sensitivity analysis indicates that data availability and

volume significantly impact the robustness of these metrics. These additions are reflected in Lines 382–387 and Lines 395–397.