

Rebuttal to Forster & Forster Letter to the Editor: Errors in tracing coronavirus SARS-CoV-2 transmission using a maximum likelihood tree.

Carla Mavian, Simone Marini, Mattia Prosperi, Marco Salemi

Submitted to: JMIR Public Health and Surveillance
on: September 29, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
---------------------------------	----------

Preprint
JMIR Publications

Rebuttal to Forster & Forster Letter to the Editor: Errors in tracing coronavirus SARS-CoV-2 transmission using a maximum likelihood tree.

Carla Mavian¹ PhD; Simone Marini² PhD; Mattia Prosperi² PhD; Marco Salemi¹ PhD

¹Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine Emerging Pathogens Institute University of Florida Gainesville US

²Department of Epidemiology, College of Public Health and Health Professions Emerging Pathogens Institute University of Florida Gainesville US

Corresponding Author:

Marco Salemi PhD

Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine

Emerging Pathogens Institute

University of Florida

2055 Mowry Road

Gainesville

US

Abstract

No abstract/not applicable

(JMIR Preprints 29/09/2020:24661)

DOI: <https://doi.org/10.2196/preprints.24661>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible only to logged-in users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a peer-reviewed journal, my article will remain visible to all users.

Original Manuscript

Rebuttal to Forster & Forster Letter to the Editor: Errors in tracing coronavirus SARS-CoV-2 transmission using a maximum likelihood tree.

Carla Mavian^{1,2}, Simone Marini^{1,3}, Mattia Prosperi^{1,3}, Marco Salemi^{1,2#}.

1 Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, FL, USA; 2 Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA; 3 Department of Epidemiology, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA.

Corresponding author: salemi@pathology.ufl.edu

Dear Editor,

Before discussing, in detail, the serious technical issues, and conceptual and theoretical mistakes in the Letter to the Editor by Forster & Forster, we would like to emphasize the following points. First, very recent work by the Stamatakis group (1) has confirmed and further extended our original observation of lack of phylogenetic signal in SARS-CoV-2 sequences from the early phase of the pandemic, which is in line with our main criticism that the Forster *et al.* PNAS paper (2) was based on a superficial analysis of biased and noisy sequence data, resulting at best in misleading conclusions. Second, many of the claims in the paper by Forster *et al.* (2) have been criticized by three independent Letters to the Editor of PNAS (3-5), including a letter from our own group (5), which was signed by over 30 world-renowned experts in phylogenetic analysis, who actually pioneered and contributed to the development of modern phylodynamics. Third, our paper published in JMIR, has also been supported and confirmed by similar findings of other independent investigators (1,6,7) clearly showing that phylogeny-based analyses of SARS-CoV-2 genomic data, available during the early phase of the pandemic, have led to premature conclusions and/or statistically questionable findings, due to lack of phylogenetic signal determined by the sudden

emergence and exponential growth of the virus, as well as a strong sampling bias. Indeed, our paper in JMIR shows that even when new (and more recently sampled) sequences are added to the tree, phylogeographic hypotheses of early SARS-CoV-2 spread in Europe, such as the possible introduction of the virus from Germany to Italy, cannot be proved with sufficient statistical robustness, since the sequence data support several others, equally likely scenarios.

It is true that methods such as contact tracing and mobile phone tracking can be very effective in tracking outbreaks (8). Yet, epidemiological tracing and surveillance by other means was not the focus of our work published on JMIR, which discusses only the unreliability of using SARS-CoV-2 sequence data, without the aid of other contact tracing methods, to infer virus dissemination during the early phase of the pandemic. Therefore, one of the major points raised by Foster & Forster in their rebuttal is irrelevant since it is not pertinent to our manuscript or the interpretation of our findings.

Forster & Forster misinterpret the message of our paper, based on targeted sentences/paragraphs taken out of context. There was no trivial or otherwise oversight in our analysis. If anything, the subsequent isolation of sequences from Portugal, Brazil, Wales, and Netherlands patients, which were identical to the pre-existing Italian sequence makes precisely our point: “it is not possible with the present data to decide which branching pattern (and, therefore, which phylogeographic reconstruction) most likely represents actual dissemination routes among European countries”. Forster & Forster go on discussing how the Welshman, both Dutch, and the Brazilian patients all had visited Italy a few days before falling ill. This is an interesting information that may suggest such patients were infected in Italy, after all contact tracing is presently considered the golden standard for tracking SARS-CoV-2 dissemination, but it has very little to do with the central problem raised in our manuscript: branching patterns in the phylogeny alone, especially when based on several identical sequences from different geographic areas (one of the very definitions of lack of

phylogenetic signal in any basic textbook (9), which Forster & Forster seem to ignore) cannot distinguish among different, and equally likely dissemination scenarios. In fact, Table 1 in our manuscript shows, as expected given the presence of several identical sequences sampled over a short time interval in different countries, that alternative topologies underlying alternative dissemination scenarios are equally likely. Besides, SARS-CoV-2 incubation period and lack of symptoms during early infection, should caution against firm conclusions on the directionality of infection even when further details on travel history are available.

Forster & Forster critique of Figure 2 exemplifies the extent of their misreading of our manuscript. As clearly stated in the legend, the maximum likelihood tree in Figure 2 is displayed as a “cladogram”, which means that branch lengths are not drawn proportional to genetic distance or time-scaled. Cladograms are branching diagrams only showing cladistic relationship among taxa, where branches have arbitrary length chosen for best display purposes (9). It seems Forster & Forster misread the figure legend, as the confusion between cladogram and phylogram implied in their rebuttal would be quite egregious for any scientist with a basic background in phylogenetic analysis. Foster & Forster have not discovered any flaw. The fact that previously unsampled sequences from Portugal, Brazil, Wales, and the Netherlands are identical to the Italian sequence, is *exactly* the point we are making: such sequences altogether have no phylogenetic signal (defined as the minimum amount of genetic diversity required to generate resolved phylogenies (9)). In a phylogeny with branch lengths drawn proportional to genetic distances, such sequences would appear to cluster tightly along very short branches of actual zero length, simultaneously arising from a common ancestor. This is what we call in the paper star-like signal, which is obviously associated with phylogenetic noise, i.e. the inability to discern the exact evolutionary relationship among sequences (other than trivially say that they are all identical and, thus, related through a most recent common ancestor). In summary, while it is true that identical sequences are likely linked by close transmissions, it is also important to remember that, in the absence of phylogenetic information, it

would be impossible to establish the correct sequence of events through phylogeny reconstruction alone, which is the whole point of our paper.

References

1. Morel, B., Barbera, P., Czech, L., Bettisworth, B., Hübner, L., Lutteropp, S., Serdari, D., Kostaki, E.-G., Mamais, I., Kozlov, A.M. *et al.* (2020) Phylogenetic analysis of SARS-CoV-2 data is difficult. *bioRxiv*, 2020.2008.2005.239046.
2. Forster, P., Forster, L., Renfrew, C. and Forster, M. (2020) Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences*, 202004999.
3. Chookajorn, T. (2020) Evolving COVID-19 conundrum and its impact. *Proceedings of the National Academy of Sciences*, 202007076.
4. Sánchez-Pacheco, S.J., Kong, S., Pulido-Santacruz, P., Murphy, R.W. and Kubatko, L. (2020) Median-joining network analysis of SARS-CoV-2 genomes is neither phylogenetic nor evolutionary. *Proceedings of the National Academy of Sciences*, 202007062.
5. Mavian, C., Pond, S.K., Marini, S., Magalis, B.R., Vandamme, A.-M., Dellicour, S., Scarpino, S.V., Houldcroft, C., Villabona-Arenas, J., Paisie, T.K. *et al.* (2020) Sampling bias and incorrect rooting make phylogenetic network tracing of SARS-COV-2 infections unreliable. *Proceedings of the National Academy of Sciences*, 202007295.
6. MacLean, O.A., Orton, R., Singer, J.B. and Robertson, D.L. (2020).
7. Worobey, M., Pekar, J., Larsen, B.B., Nelson, M.I., Hill, V., Joy, J.B., Rambaut, A., Suchard, M.A., Wertheim, J.O. and Lemey, P. (2020) The emergence of SARS-CoV-2 in Europe and North America. *Science*, eabc8169.
8. Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D. and Fraser, C. (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, eabb6936.
9. Salemi, M., Vandamme, A.-M., & Lemey, P. (2009) *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press, Cambridge, UK.