# Letter to the Editor: Errors in tracing coronavirus SARS-CoV-2 transmission using a maximum likelihood tree

Peter Forster, Lucy Forster

Submitted to: JMIR Public Health and Surveillance
on: August 15, 2020

# *Table of Contents*

# Letter to the Editor: Errors in tracing coronavirus SARS-CoV-2 transmission using a maximum likelihood tree

Peter Forster[1] PhD; Lucy Forster[2] PhD

[1]McDonald Institute University of Cambridge Cambridge GB
[2]Lakeside Healthcare Group at Cedar House Surgery – NHS St Neots GB

**Corresponding Author:**
Peter Forster PhD
McDonald Institute
University of Cambridge
Downing Site
Cambridge
GB

## Abstract

Tracing and quarantining symptomatic and asymptomatic individuals infected by the novel coronavirus SARS-CoV-2 is an important approach in controlling the current epidemic. Tracing the source of an infection can be achieved by conventional interviews, by mobile telephone tracking, or by phylogenetic tracing of the virus genomes itself, as we have proposed (reference 1)

In a recent critique (reference 2) (see also our reply in reference 3), Mavian and colleagues have disputed our phylogenetic tracing approach and conclude: “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.”

Their underlying reanalysis is however based on a trivial oversight. They analyse genomes collected worldwide in early March 2020 and initially confirm the B-subclade that we had identified, which links a German sequence to an Italian sequence and thence to further Finnish, Mexican, Swiss, and German sequences. However, they then claim “in a new tree inferred just one week later, when more than 135 new full genome sequences were made available on GISAID, the direct link between Germany and Italy […] disappeared due to additional clustering of [five] previously unsampled sequences from Portugal, Brazil, Wales, and [two from] the Netherlands”.

Upon request, Dr Mavian provided us with a file of these five new sequences. Comparing these five in our coronavirus sequence alignment table (freely available at www.fluxus-engineering.com), it transpires that these five sequences are identical to each other and to the pre-existing Italian sequence. Mavian and colleagues appear not to have noticed the identity as they fail to mention it, and instead they present a “maximum likelihood” tree which misleadingly shows these five new sequences and the pre-existing Italian sequence to be separated by apparently deep branches, even though all six are identical. Mavian and colleagues appear to have relied on their computer program without investigating their entered sequences.

Moreover, Mavian and colleagues have not presented the documented patient travel histories of the five new viral sequences. We present these now, using freely available GISAID information and contemporary reports, and find that the Welshman (references 4, 5), both Dutch (reference 6), and the Brazilian (reference 7) all had visited Italy a few days before falling ill. The Portuguese (reference 8) had visited Spain. Thus, in four of the five new cases, the patient’s travel history to Italy confirms the viral sequence matches to the pre-existing Italian sequence. It is therefore unfounded for Mavian and colleagues to claim that the data cannot reveal branching patterns or likely dissemination routes among European countries.”

(JMIR Preprints 15/08/2020:23542)
DOI: https://doi.org/10.2196/preprints.23542

## 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 v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

[editors: Dr Travis Sanchez, Emory University, Travis.Sanchez@emory.edu , Professor Gunther Eysenbach, Uni of Toronto, ed-support@jmir.org]
version 2020Sep19

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

Peter Forster, Lucy Forster

Tracing and quarantining symptomatic and asymptomatic individuals infected by the novel coronavirus SARS-CoV-2 is an important approach in controlling the current epidemic. Tracing the source of an infection can be achieved by conventional interviews, by mobile telephone tracking, or by phylogenetic tracing of the virus genomes themselves, as we have proposed (reference 1).

In a recent critique (reference 2) (see also our reply in reference 3), Mavian and colleagues have disputed our phylogenetic tracing approach and conclude: “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.”

Their underlying reanalysis is however based on a trivial oversight. They analyse genomes collected worldwide in early March 2020 and initially confirm the B-subclade that we had identified, which links a German sequence to an Italian sequence and thence to further Finnish, Mexican, Swiss, and German sequences. However, they then claim “in a new tree inferred just one week later, when more than 135 new full genome sequences were made available on GISAID, the direct link between Germany and Italy […] disappeared due to additional clustering of [five] previously unsampled sequences from Portugal, Brazil, Wales, and [two from] the Netherlands”.

Upon request, Dr Mavian provided us with a file of these five new sequences. Comparing these five in our coronavirus sequence alignment table (freely available at the Fluxus Technology website), it transpires that these five sequences are identical to each other and to the pre-existing Italian sequence. Mavian and colleagues appear not to have noticed the identity as they fail to mention it, and instead they present a “maximum likelihood” tree which misleadingly shows these five new sequences and the pre-existing Italian sequence to be separated by apparently deep branches, even though all six are identical. Mavian and colleagues appear to have relied on their computer program without investigating their entered sequences.

Moreover, Mavian and colleagues have not presented the documented patient travel histories of the five new viral sequences. We present these now, using freely available GISAID information and contemporary reports, and find that the Welshman (references 4, 5), both Dutch (references 6,7), and the Brazilian (reference 8) all had visited Italy a few days before falling ill. The Portuguese (reference 9) had visited Spain. Thus, in four of the five new cases, the patient’s travel history to Italy confirms the viral sequence match to the pre-existing Italian sequence. It is therefore unfounded for Mavian and colleagues to claim that the data cannot reveal branching patterns or likely dissemination routes among European countries.

## References

1. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. PNAS April 28, 2020 117 (17) 9241-9243; doi.org/10.1073/pnas.2004999117, 2020
2. Mavian C, Marini S, Prosperi M, Salemi M. A Snapshot of SARS-CoV-2 Genome Availability up to April 2020 and its Implications: Data Analysis. JMIR Public Health Surveill 2020;6(2):e19170. DOI: 10.2196/19170. PMID: 32412415. PMCID: 7265655, 2020
3. Forster P, Forster L, Renfrew C, Forster M. Reply to Sánchez-Pacheco et al., Chookajorn, and Mavian et al.: Explaining phylogenetic network analysis of SARS-CoV-2 genomes. PNAS 117 (23) 12524-12525; doi.org/10.1073/pnas.2007433117, 2020
4. BBC. Coronavirus: Wales’ first patient urges positivity. BBC News 6 May, 2020
5. Cooper J. Meet Wales’ first COVID patient. Western Mail 8 June, 2020
6. Stoop S. Twee mensen in ‘t Gooi besemt met het coronavirus [Two people in Gooi infected with

coronavirus]. NHNieuws, 4 March, 2020.
7. Saskia de Wijk. Coronabesmetting in Zeewolde leidt tot reacties [Corona infection in Zeewolde leads to reactions]. Zeewolde-Actueel, 5 March 2020
8. GISAID access 412964
9. GISAID access 413638