

Letter to the Editor

# Errors in Tracing Coronavirus SARS-CoV-2 Transmission Using a Maximum Likelihood Tree. Comment on “A Snapshot of SARS-CoV-2 Genome Availability up to April 2020 and its Implications: Data Analysis”

---

Peter Forster<sup>1</sup>, PhD; Lucy Forster<sup>2</sup>, PhD

<sup>1</sup>McDonald Institute, University of Cambridge, Cambridge, United Kingdom

<sup>2</sup>Lakeside Healthcare Group at Cedar House Surgery – NHS, St Neots, United Kingdom

**Corresponding Author:**

Peter Forster, PhD  
McDonald Institute  
University of Cambridge  
Downing Site  
Cambridge, CB2 3ER  
United Kingdom  
Email: [pf223@cam.ac.uk](mailto:pf223@cam.ac.uk)

**Related Articles:**

Comment on: <https://publichealth.jmir.org/2020/2/e19170/>

Comment in: <https://publichealth.jmir.org/2020/4/e24661/>

(*JMIR Public Health Surveill* 2020;6(4):e23542) doi: [10.2196/23542](https://doi.org/10.2196/23542)

**KEYWORDS**

coronavirus; SARS-CoV-2; genome; phylogenetic; network; tracing; tracking; transmission; dissemination; infection; geography; route

Tracing and quarantining symptomatic and asymptomatic individuals infected by the novel coronavirus SARS-CoV-2 is an important approach to 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 in our work [1].

In a recent critique in *JMIR Public Health and Surveillance* [2] (also see our reply [3]), Mavian and colleagues have disputed our phylogenetic tracing approach and concluded: “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 analyzed genomes collected worldwide in early March 2020 and initially confirmed 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 on the Fluxus Technology website [4]), it transpires that these five sequences are identical to each other and to the pre-existing Italian sequence. Mavian and colleagues [2] appear not to have noticed the identity as they fail to mention it; 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 they are identical. Mavian and colleagues [2] appear to have relied on their computer program without investigating their entered sequences.

Moreover, Mavian and colleagues [2] have not presented the documented patient travel histories of the five new viral sequences. We present these now, using freely available GISAID (Global Initiative on Sharing All Influenza Data) information [5] and contemporary reports, and find that all patients (ie, the Welshman [6,7], both Dutch [8,9], and the Brazilian [GISAID access 412964]) had visited Italy a few days before falling ill. The Portuguese (GISAID access 413648) 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 [2] to claim that the data cannot reveal branching patterns or likely dissemination routes among European countries.

### Conflicts of Interest

None declared.

### References

1. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci U S A* 2020 Apr 28;117(17):9241-9243 [FREE Full text] [doi: [10.1073/pnas.2004999117](https://doi.org/10.1073/pnas.2004999117)] [Medline: [32269081](https://pubmed.ncbi.nlm.nih.gov/32269081/)]
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 Jun 1;6(2):e19170. [doi: [10.2196/19170](https://doi.org/10.2196/19170)] [Medline: [32412415](https://pubmed.ncbi.nlm.nih.gov/32412415/)]
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. *Proc Natl Acad Sci U S A* 2020 Jun 09;117(23):12524-12525. [doi: [10.1073/pnas.2007433117](https://doi.org/10.1073/pnas.2007433117)] [Medline: [32439706](https://pubmed.ncbi.nlm.nih.gov/32439706/)]
4. Fluxus Technology. URL: <https://www.fluxus-technology.com/> [accessed 2020-10-13]
5. GISAID. 2020. URL: <https://www.gisaid.org/> [accessed 2020-10-13]
6. Coronavirus: Wales' first patient urges positivity. *BBC News*. 2020 May 6. URL: <https://www.bbc.com/news/uk-wales-52561210> [accessed 2020-10-13]
7. Cooper J. Meet Wales' first COVID patient. *Western Mail*. 2020 Jun 8. URL: <https://www.pressreader.com/uk/western-mail/20200608/282076279099626> [accessed 2020-10-13]
8. Stoop S. Twee mensen in 't Gooi besmet met het coronavirus (Two people in Gooi infected with coronavirus). *NHNieuws*. 2020 Mar 4. URL: <https://www.nhnieuws.nl/nieuws/263083/twee-mensen-in-t-gooi-besmet-met-het-coronavirus> [accessed 2020-10-13]
9. de Wijk S. Coronabesmetting in Zeewolde leidt tot reacties (Corona infection in Zeewolde leads to reactions). *Zeewolde-Actueel*. 2020 Mar 5. URL: <https://www.zeewolde-actueel.nl/nieuws/algemeen/207051/coronabesmetting-in-zeewolde-leidt-tot-reacties> [accessed 2020-10-13]

### Abbreviations

**GISAID:** Global Initiative on Sharing All Influenza Data

*Edited by T Sanchez, T Derrick; submitted 15.08.20; peer-reviewed by C Mavian; comments to author 18.09.20; revised version received 19.09.20; accepted 23.09.20; published 11.11.20*

*Please cite as:*

*Forster P, Forster L*

*Errors in Tracing Coronavirus SARS-CoV-2 Transmission Using a Maximum Likelihood Tree. Comment on "A Snapshot of SARS-CoV-2 Genome Availability up to April 2020 and its Implications: Data Analysis"*

*JMIR Public Health Surveill* 2020;6(4):e23542

URL: <http://publichealth.jmir.org/2020/4/e23542/>

doi: [10.2196/23542](https://doi.org/10.2196/23542)

PMID: [33174859](https://pubmed.ncbi.nlm.nih.gov/33174859/)

©Peter Forster, Lucy Forster. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 11.11.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.