

The Novel Coronavirus is Mutating Gradually

Kevin Surya, Jacob Gardner, and Chris Organ

Here, we describe and compare the mode of SARS-CoV-2 genomic evolution to its tempo, which so far has been much slower than that of SARS-CoV [1] and the seasonal flu [2]. Once SARS-CoV-2 jumped and spread among humans, has it been mutating in a punctuated manner, so that a large proportion of their genomic divergences occurred during transmission events? Or has SARS-CoV-2 evolution been gradual?

To test for punctuated evolution [3,4], we regressed total phylogenetic path lengths of SARS-CoV-2 genomes (root-to-tip distances) on the net number of transmission events (nodes). We acquired a molecular tree of 3,958 genomes from [Nextstrain](#) [5,6]. And we used the maximum likelihood algorithm in BAYESTRAITS 3.0.1 [7], under a phylogenetic generalized least squares (PGLS), to estimate the parameters of the regression above. Punctuation would be consistent with a strong positive correlation.

We, however, find little evidence for a punctuated genome evolution (slope = $-0.0063 \pm .0038$; $R^2 = .064$; LR [vs. intercept-only model] = 263.79; Figure 1). Diagnostics indicate no severe violations of linear regression assumptions (Figure A2). Additionally, the node-density artifact [3,8], an underestimation of branch lengths in tree regions with fewer taxa, does not seem to be present ($\delta = -.062$; Figure A3). Altogether, the evidence suggests that SARS-CoV-2 genomes are evolving gradually, with much of the mutations occurring in between net transmission events. The tempo and mode of SARS-CoV-2 evolution are likely linked. We expect the mutation rate of the novel coronavirus to be steady, and that drugs and vaccines (e.g. [9,10]) under development will still be effective in the future.

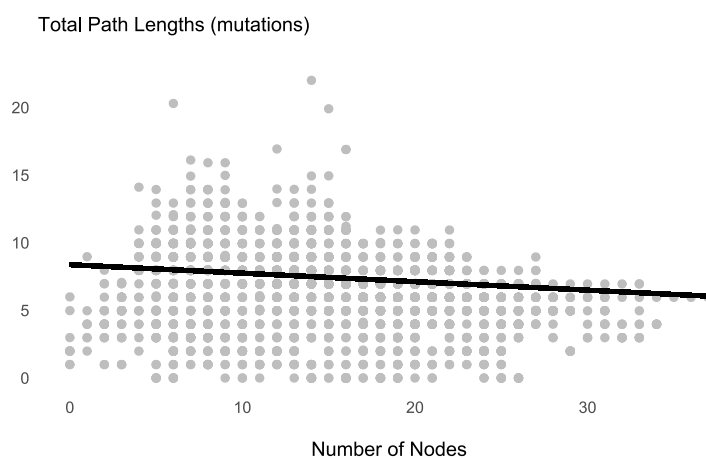


Figure 1. The number of nodes along the SARS-CoV-2 lineages (net transmission events) does not correlate with, nor does it explain the variation ($R^2 = .064$) in total path lengths (accumulated mutations in the genome). PGLS equation: $y = 8.38 - .063x$.

References

1. Jia, Y., Shen, G., Zhang, Y., Huang, K.-S., Ho, H.-Y., Hor, W.-S., Yang, C.-H., Li, C., and Wang, W.-L. (2020). Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. *bioRxiv*, 2020.04.09.034942.
2. Kupferschmidt, K. (2020). Mutations can reveal how the coronavirus moves—but they're easy to overinterpret. *Science*. Available at: <https://www.sciencemag.org/news/2020/03/mutations-can-reveal-how-coronavirus-moves-they-re-easy-overinterpret> [Accessed April 16, 2020].
3. Webster, A.J., Payne, R.J.H., and Pagel, M. (2003). Molecular phylogenies link rates of evolution and speciation. *Science* *301*, 478–478.
4. Pagel, M., Venditti, C., and Meade, A. (2006). Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* *314*, 119–121.
5. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* *34*, 4121–4123.
6. Sagulenko, P., Puller, V., and Neher, R.A. (2017). TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* *4*.
7. Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* *401*, 877–884.
8. Venditti, C., Meade, A., and Pagel, M. (2006). Detecting the node-density artifact in phylogeny reconstruction. *Syst. Biol.* *55*, 637–643.
9. Le, T.T., Andreadakis, Z., Kumar, A., Román, R.G., Tollefsen, S., Saville, M., and Mayhew, S. (2020). The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.*
10. Sheahan, T.P., Sims, A.C., Zhou, S., Graham, R.L., Pruijssers, A.J., Agostini, M.L., Leist, S.R., Schäfer, A., Dinnon, K.H., Stevens, L.J., *et al.* (2020). An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.*

Appendix

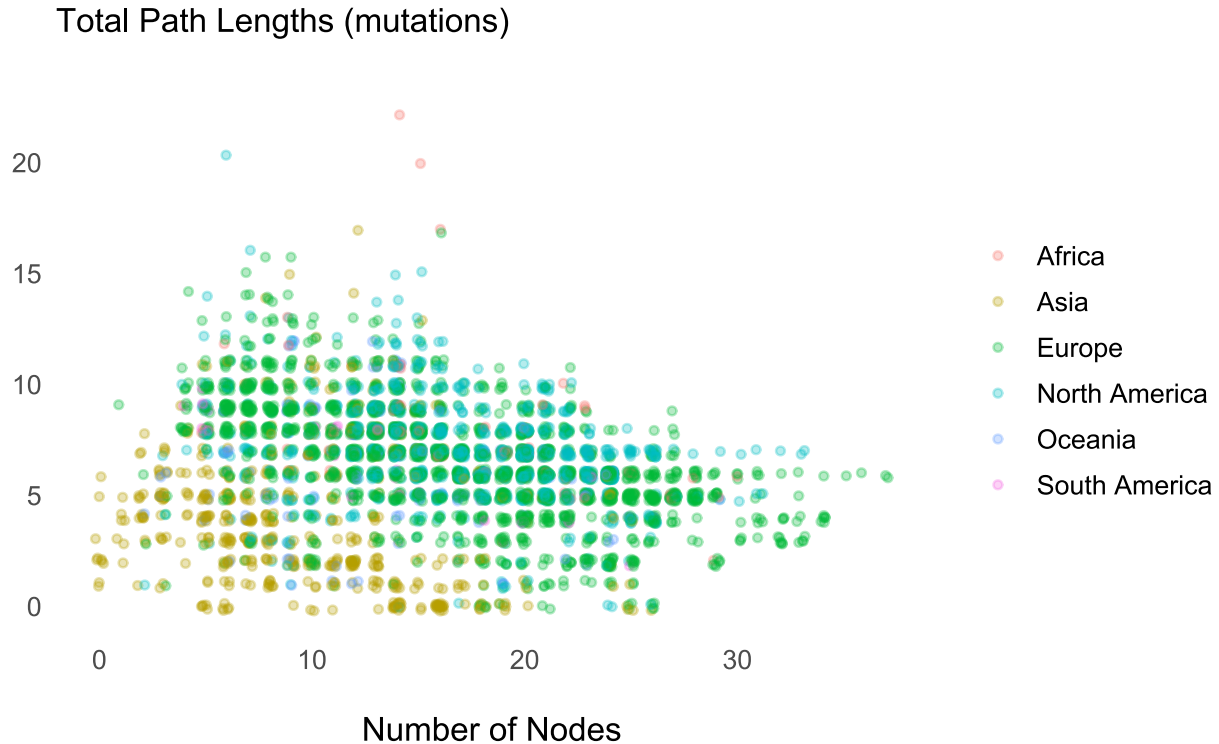


Figure A1. There may also not be punctuations of the SARS-CoV-2 genomic evolution at the continental level. See the interactive plot in the supplementary file folder to explore this data further. But, note that every continent has multiple introductions (see [Nextstrain](#)).

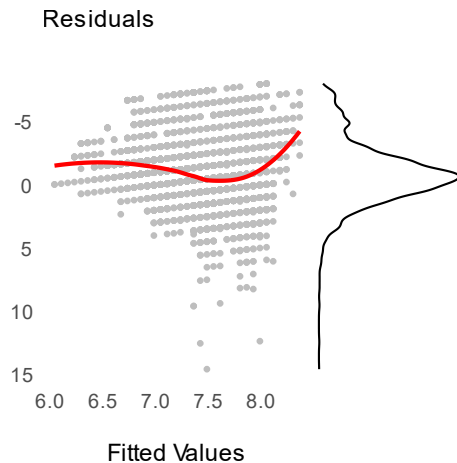


Figure A2. Regression diagnostics. The residuals vs. fitted values plot does not indicate a severe violation of residual homogeneity. The histogram shows that the residuals are normally distributed with a slight right skew.

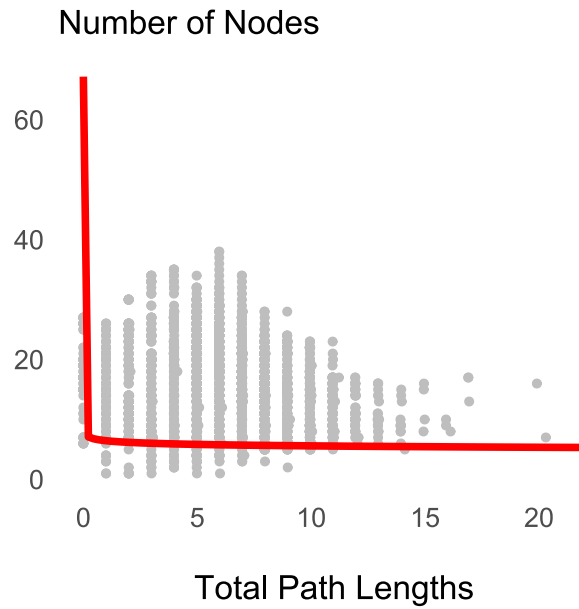


Figure A3. The node-density effect, which can bias our punctuation test, does not seem to be present ($\delta = -.062$). (A symptom of the artifact is a curvilinear relationship with $\delta > 1$.) Equation: $y = 6.46x^{-.062}$.