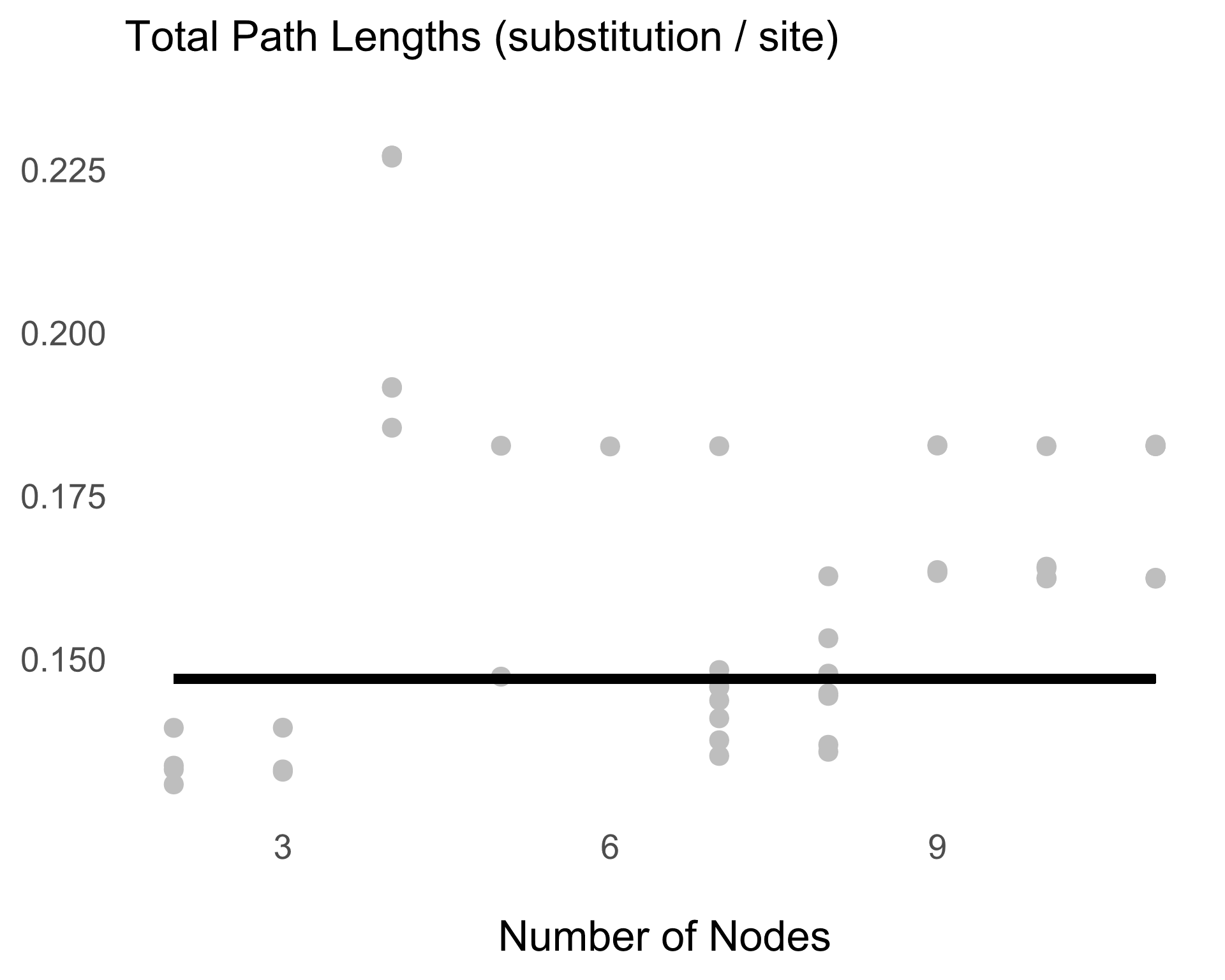
**SARS-like Coronaviruses Have Been Evolving Gradually**

Kevin Surya, Jacob Gardner, and Chris Organ

Having instigated two pandemics so far, how have the broader culprits, SARS-like betacoronaviruses, been evolving? These SARS-like viruses (CoVs) naturally circulate among bats and have periodically jumped to other mammals such as civets, pangolins, and eventually humans [1–3]. Here, we test whether the SARS-like CoV genomes have evolved gradually or in a punctuated manner so that most divergences occurred during transmission events. We further test if the mode of evolution shifted once a pandemic began.

To test for punctuated evolution [4,5], we regressed total phylogenetic path lengths of SARS-like CoV genomes (root-to-tip distances) on the net number of transmission events (nodes). We acquired a molecular tree (Figure A1) of 52 genomes from [Nextstrain](https://nextstrain.org/groups/blab/sars-like-cov) [6,7]. And we used the maximum likelihood algorithm in BayesTraits 3.0.1 [8], 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 also compared the fit of this model with two others (Figure A2 and A3), where we allowed the regression lines to vary across strains (SARS-CoV, SARS-CoV-2, and SARS-like CoV).

The most likely model, accounting for complexity, is the one with a single regression fit line (BIC = -510.55; lowest ΔBIC = 4.31). Looking at this model (Figure 1), we, however, find little evidence for a punctuated evolution (slope = .0000044 ± .000028, *P =* .438; *R2* = .0049). Diagnostics indicate violations of linear regression assumptions (Figure A4). The node-density artifact [4,9], an underestimation of branch lengths in tree regions with fewer taxa, is present (*δ =* 9.46; Figure A5). But, it does not bias our analysis because we do not detect punctuation. A potentially more serious bias is the undersampling of non-human SARS-like CoVs [10]. Given these caveats, we merely intend to hint that the broader SARS-like betacoronaviruses have been evolving gradually at all times.

**Figure 1.** The number of nodes along the SARS-like betacoronavirus lineages (net transmission events) does not correlate with, nor does it explain the variation (*R2 =* .00049) in total path lengths (substitutions in the genome). PGLS equation: *y =* .15 - .0000044*x*.

**References**

1. Xu, Y. (2020). Unveiling the origin and transmission of 2019-nCoV. Trends Microbiol. *28*, 239–240.

2. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., *et al.* (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269.

3. Lam, T.T.-Y., Shum, M.H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W.Y.-M., Li, W.-J., Li, L.-F., *et al.* (2020). Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature, 1–6.

4. Webster, A.J., Payne, R.J.H., and Pagel, M. (2003). Molecular phylogenies link rates of evolution and speciation. Science *301*, 478–478.

5. Pagel, M., Venditti, C., and Meade, A. (2006). Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science *314*, 119–121.

6. 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.

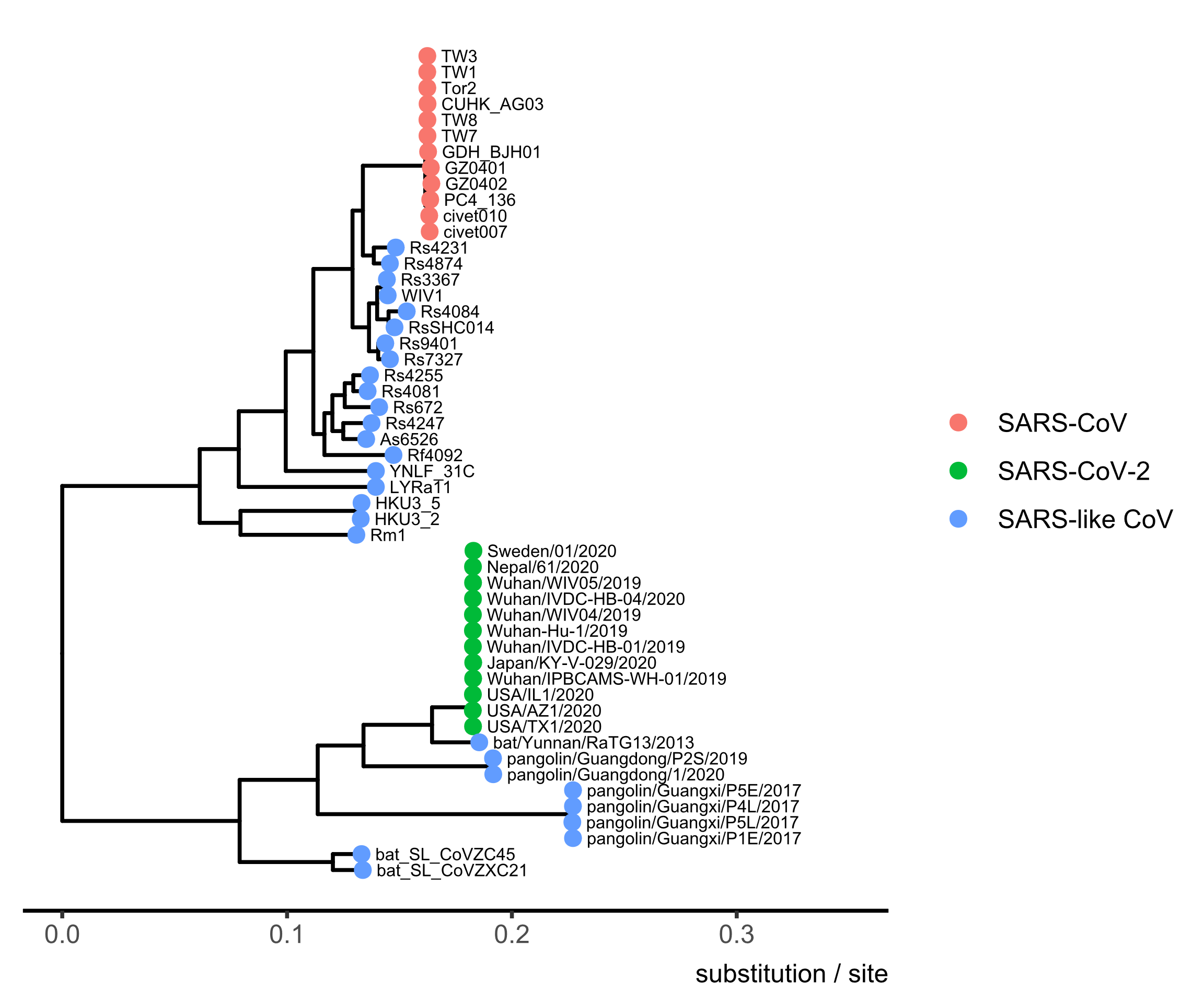
7. Sagulenko, P., Puller, V., and Neher, R.A. (2017). TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. *4*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5758920/ [Accessed March 22, 2020].

8. Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877–884.

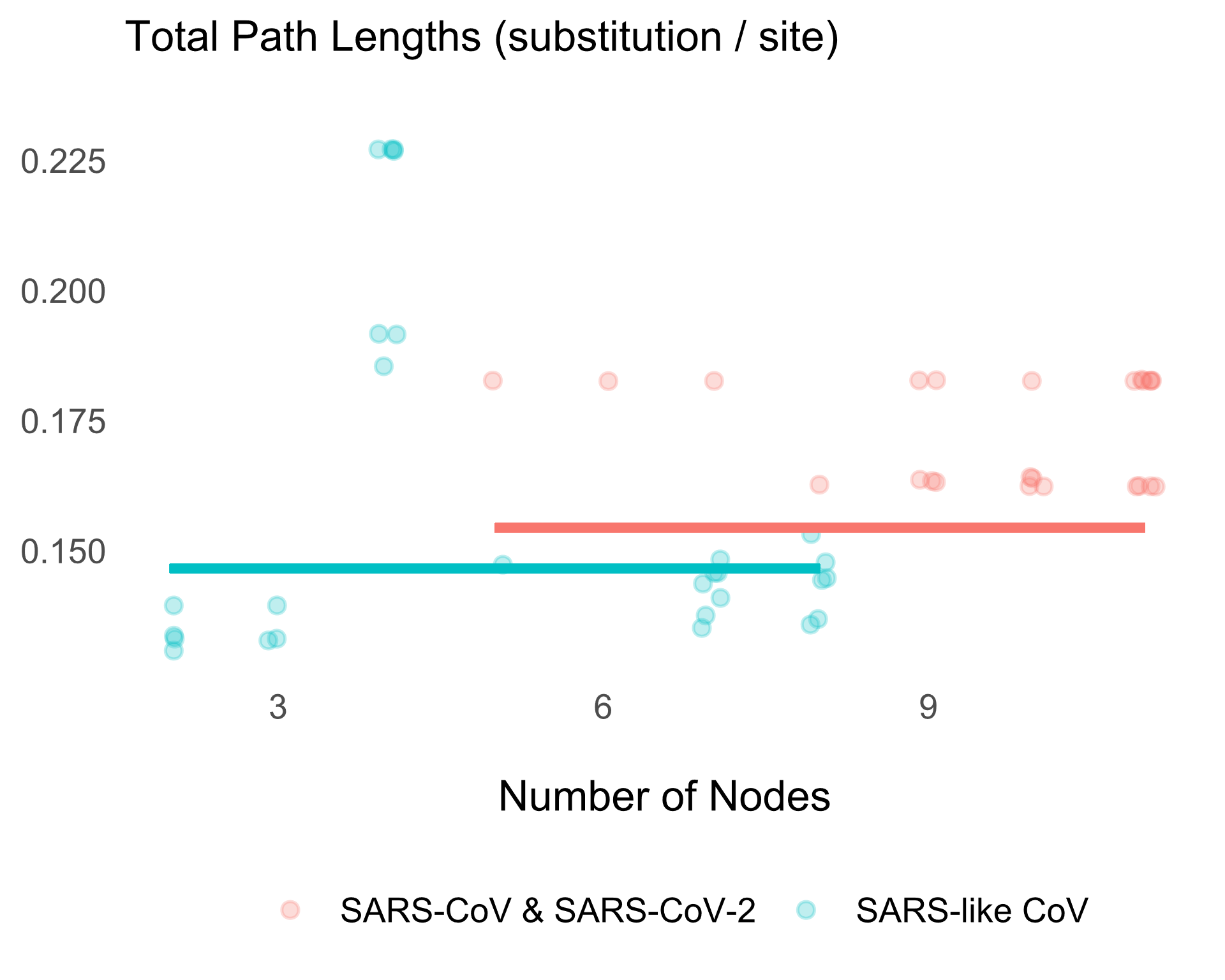
9. Venditti, C., Meade, A., and Pagel, M. (2006). Detecting the node-density artifact in phylogeny reconstruction. Syst. Biol. *55*, 637–643.

10. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nat. Med. *26*, 450–452.

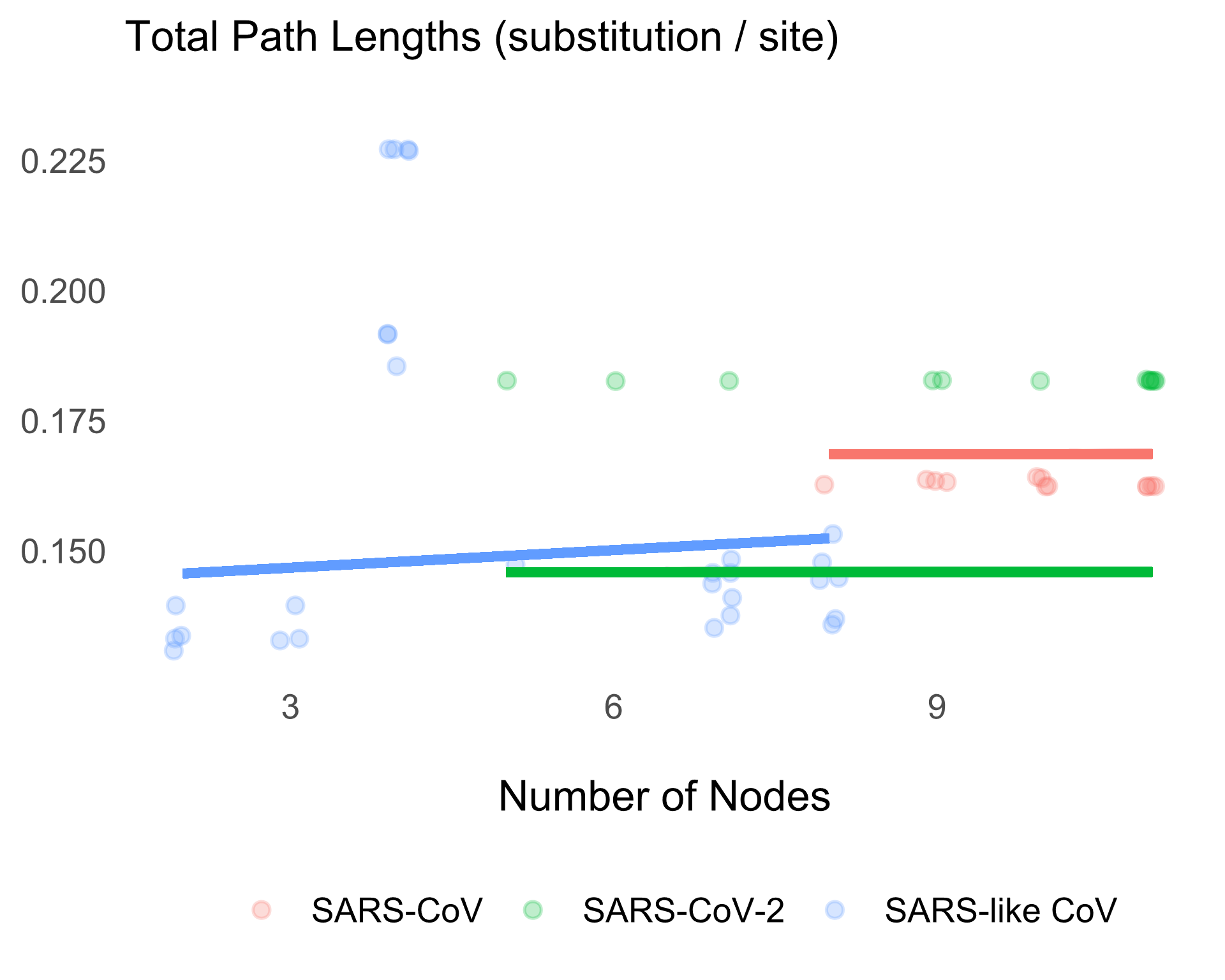
**Appendix**



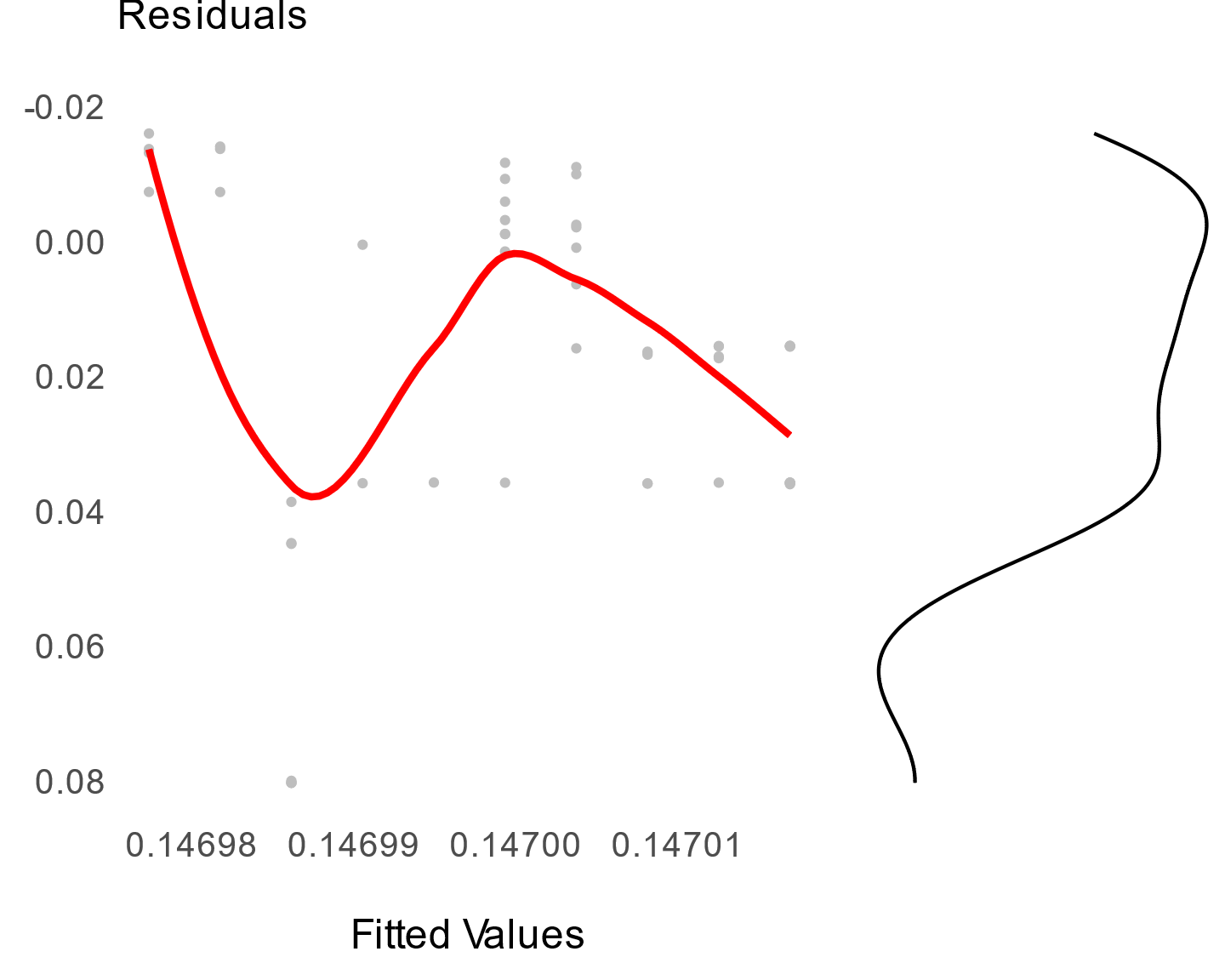
**Figure A1.** Molecular phylogeny of SARS-like betacoronaviruses downloaded from [Nextstrain](https://nextstrain.org/groups/blab/sars-like-cov). This tree contains 52 genomes (12 SARS-CoV, 12 SARS-CoV-2, and 28 SARS-like CoV).



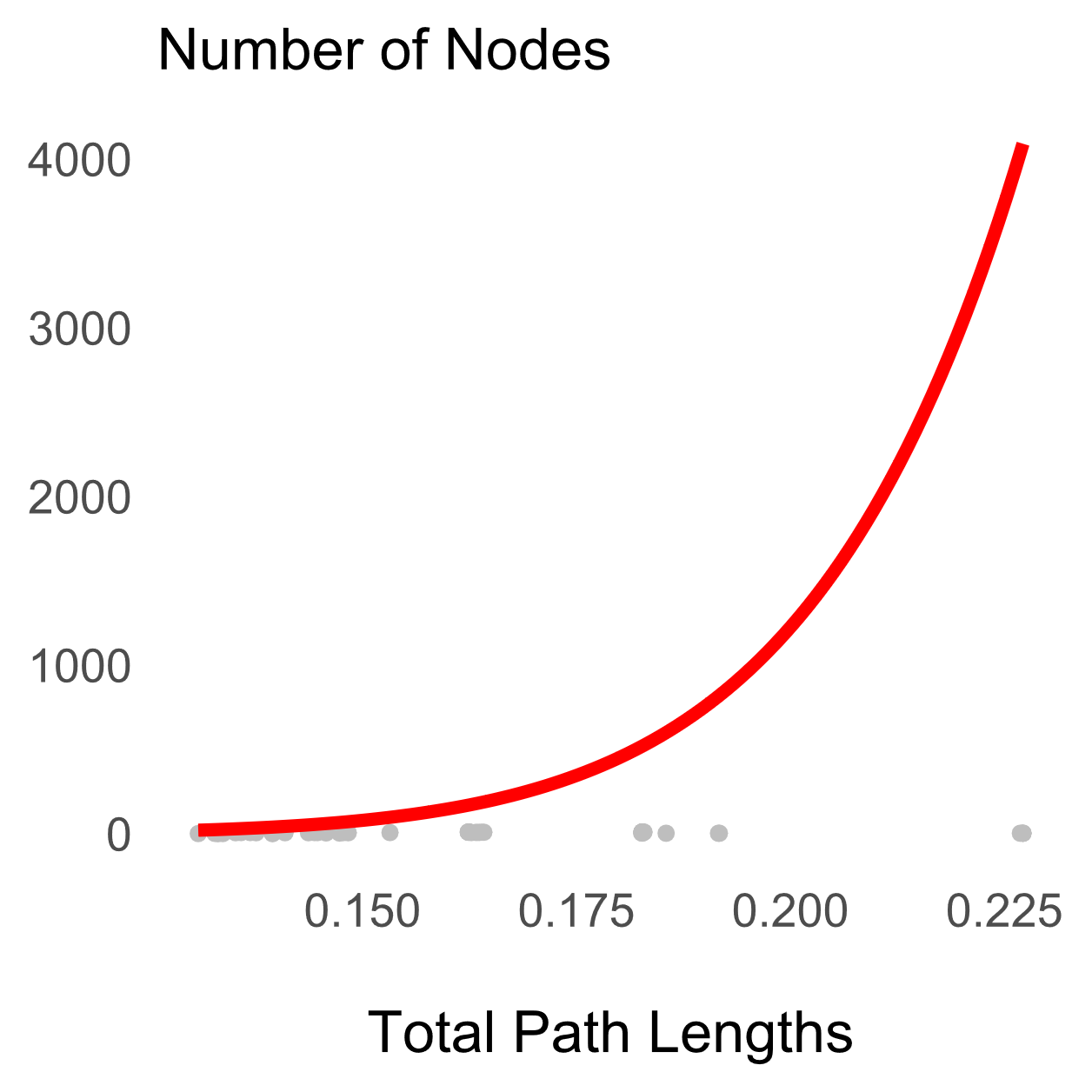
**Figure A2.** SARS-like betacoronaviruses have likely been evolving gradually before and during the two pandemics (slope [SARS-CoV & SARS-CoV-2] = .0000038 ± .000027, *P* = .445; slope [SARS-like CoV] = -.0000010 ± .0012, *P* = .500; *R2 =* .049). Having two regression fit lines, however, do not increase the likelihood enough relative to the model with only one line (ΔBIC = -5.32). PGLS Equation (SARS-CoV & SARS-CoV-2): *y =* .154 + .0000038*x*. PGLS Equation (SARS-like CoV): *y =* .147 - .0000010*x*. The intercepts seem off because they represent phylogenetic means.



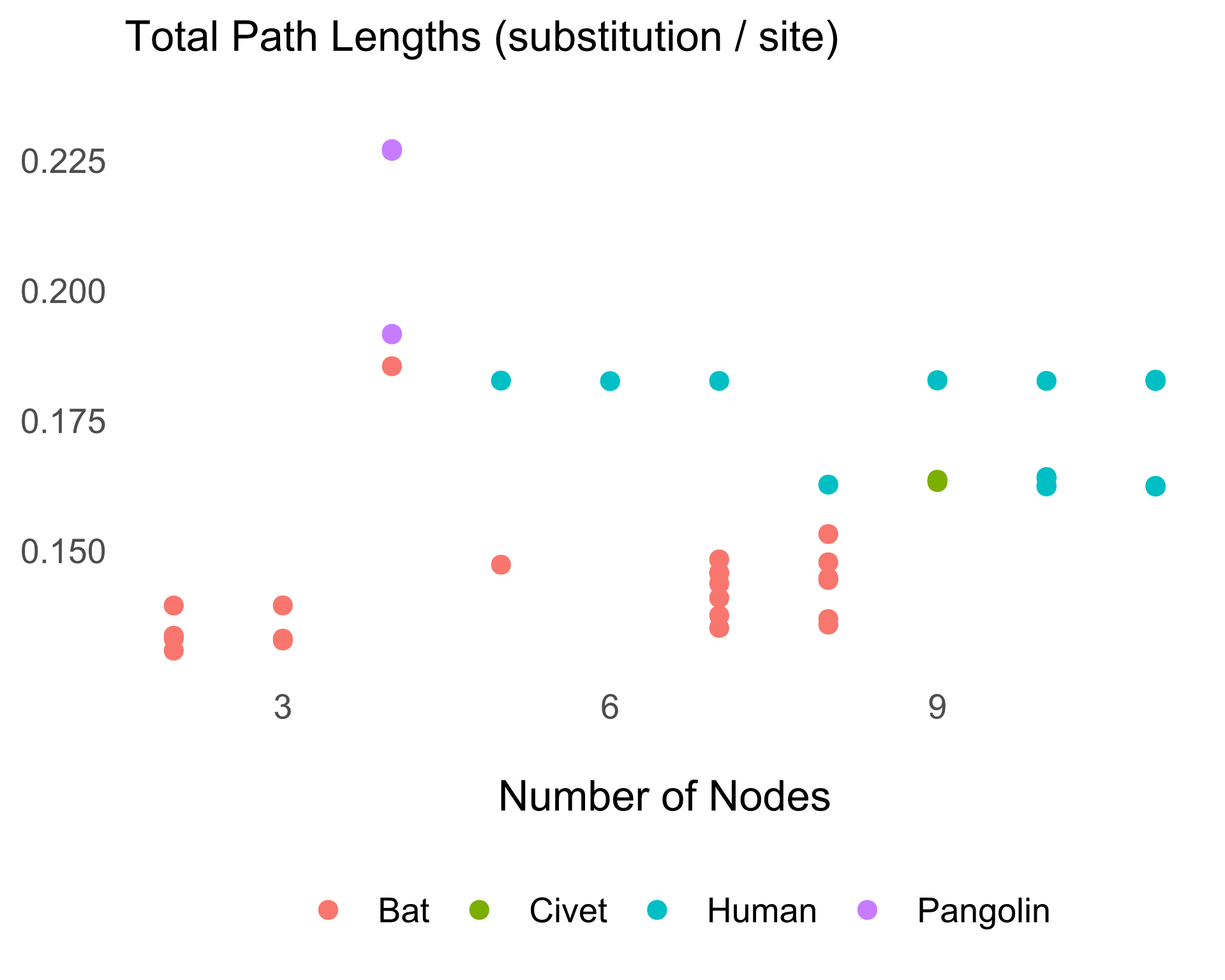
**Figure A3.** SARS-like betacoronaviruses have likely been evolving gradually before and during the two pandemics (slope [SARS-CoV and SARS-CoV-2; fixed to be equal] = .0000032 ± .000026, *P* = .451; slope [SARS-like CoV] = 0.0011 ± .0012, *P* = .187; *R2 =* .14). Having three regression fit lines, however, do not increase the likelihood enough relative to the model with only one line (ΔBIC = -4.31). PGLS Equation (SARS-CoV): *y =* .169 + .0000032*x*. PGLS Equation (SARS-CoV-2): *y =* .146 + .0000032*x*. PGLS Equation (SARS-like CoV): *y =* .143 + 0.0011*x*. The intercepts seem off because they represent phylogenetic means.



**Figure A4.** Regression diagnostics indicate assumption violations. The residuals vs. fitted values plot shows that residuals are not constant across the fitted values. Also, the distribution of the residuals is right-skewed.



**Figure A5.** The node-density artifact is obvious (*δ =* 9.46). But, it does not bias our analysis because we do not find evidence for punctuation. (A symptom of the artifact is a curvilinear relationship with *δ >* 1.) Equation: *y =* 5,039,309,975*x*9.46.



**Figure A6.** SARS-like betacoronaviruses seem to have been evolving gradually, at least in bats and humans, where we have a decent sample size.