

Unraveling the Temporal Origins of the SARS-CoV-2 Omicron Variant: A Phylogenetic Analysis of Sequences from September-October 2021



Zijing (Carol) Zhou¹, Chika Kingsley Onwuamah², Anne Piantadosi³

¹Emory University College of Arts and Sciences, ²Nigerian Institute of Medical Research, ³Emory University School of Medicine



Introduction

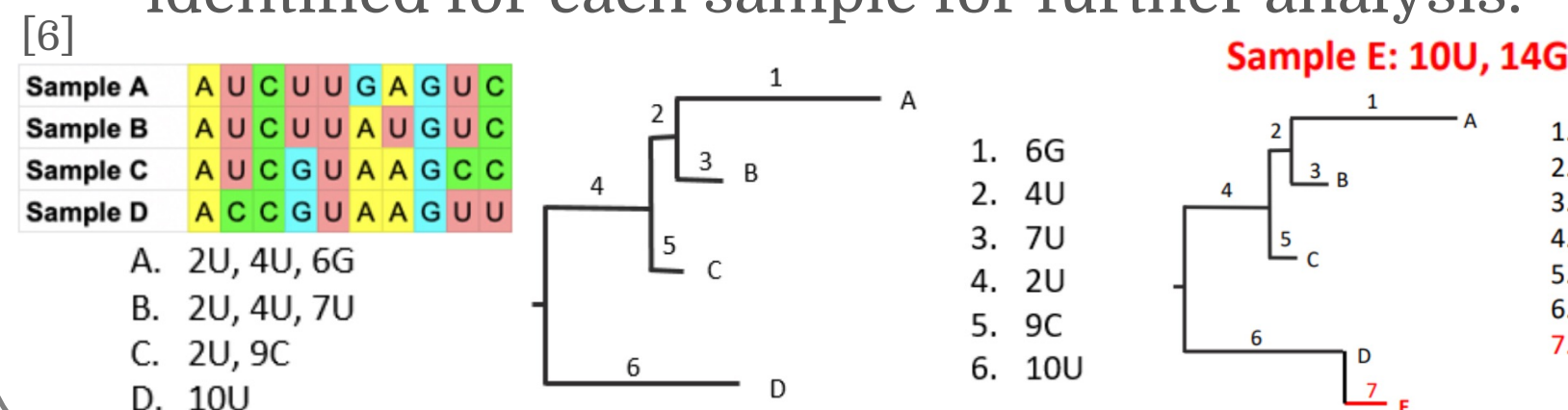
- The Omicron variant (B. 1.1. 529) of SARS-CoV-2 was first identified in South Africa in November 2021^[1].
- However, in a preliminary report, our collaborators indicated that they might have found 12 Omicron samples from Nigeria that dated before November 2021^[2].
- To investigate whether these sequences indicate that Omicron has emerged earlier than previously thought, we performed several phylogenetic analyses.
- We identified the most genetic-related sequences of the 12 Nigerian samples from a global phylogenetic tree using Ultrafast Sample placement on Existing tRees (USHER)^[3] and performed molecular clock inference using TreeTime^[4] and Bayesian Evolutionary Analysis Sampling Trees 2 (BEAST 2)^[5].

Questions

- Are all 12 Nigerian samples Omicron?
- What are the collection dates of the most genetically similar sequences?
- Do the molecular clock phylogenies support the reported collection dates?

Methods

The 12 Nigerian samples were classified by USHER. The most genetically related sequences were identified for each sample for further analysis.



An estimated time-scaled phylogeny was constructed with one subtree (3 Nigerian sequences and 147 other closely related sequences) using TreeTime.

A maximum likelihood tree was constructed (12 Nigerian sequences and 241 other Omicron sequences) using IQTREE^[7] and its temporal signal was checked in TempEst. Two outliers were removed, and a Bayesian phylogenetic analysis was performed using BEAST 2.

Results

Sample ID	Collection Date	Nextstrain clade	Pango lineage	Genetically Similar Sequence
hCoV-19/Nigeria/NM-22-96030/2021	2021-08-13	21K (BA.1)	BA.1.1	USA/WY-WYPHL-22026856/2022 EPI_ISL_12573597 2022-01-21
hCoV-19/Nigeria/NM-22-90209/2021	2021-08-13	21K (BA.1)	BA.1.15	Nigeria/LSB352/2021 EPI_ISL_14056900 2021-12-26
hCoV-19/Nigeria/GEL/09-21/0176/2021	2021-09-03	21K (BA.1)	BA.1.1	England/ALDP-3723FA0/2022 OW019735.1 2022-02-14
hCoV-19/Nigeria/GEL/09-21/0513/2021	2021-09-07	21K (BA.1)	BA.1.1	Scotland/EDB42217/2022 OY393969.1 2022-01-25
hCoV-19/Nigeria/GEL/09-21/0548/2021	2021-09-07	21K (BA.1)	BA.1	Nigeria/NM-22-100087/2022 EPI_ISL_11297617 2022-01-06
hCoV-19/Nigeria/GEL/09-21/0668/2021	2021-09-09	21K (BA.1)	BA.1	England/PHEC-3X047X9B/2021 OX852117.1 2021-12-21
hCoV-19/Nigeria/GEL/09-21/0723/2021	2021-09-09	21K (BA.1)	BA.1	England/PHEC-3X047X9B/2021 OX852117.1 2021-12-21
hCoV-19/Nigeria/GEL/09-21/1095/2021	2021-09-14	21K (BA.1)	BA.1	England/PHEC-3X047X9B/2021 OX852117.1 2021-12-21
hCoV-19/Nigeria/GEL/09-21/1848/2021	2021-09-26	21K (BA.1)	BA.1.1	France/BRE-LBZCentre-60C620030031/2022 EPI_ISL_14864918 2022-01-03
hCoV-19/Nigeria/NM-22-98355/2021	2021-09-27	21K (BA.1)	BA.1	USA/TG1266422/2022 ON518978.1 2022-03-05
hCoV-19/Nigeria/GEL/09-21/1975/2021	2021-09-28	21K (BA.1)	BA.1.17	Australia/VIC44980/2022 EPI_ISL_11176640 2022-02-28
hCoV-19/Nigeria/GEL/10-21/0493/2021	2021-10-08	21K (BA.1)	BA.1	USA/WY-WYPHL-22023370/2022 EPI_ISL_11348986 2022-01-24

Sequences are closely related on the subtree USHER generated

Table 1. The Nextstrain clade and Pango lineage of the 12 Nigerian samples classified by USHER are confirmed as Omicron BA.1. For each sample, USHER generated a tree identifying the most genetic-related sequences. The collection dates of these closely related sequences are from late December 2021 to March 2022. Highlighted sequences indicate that they are closely related to each other.

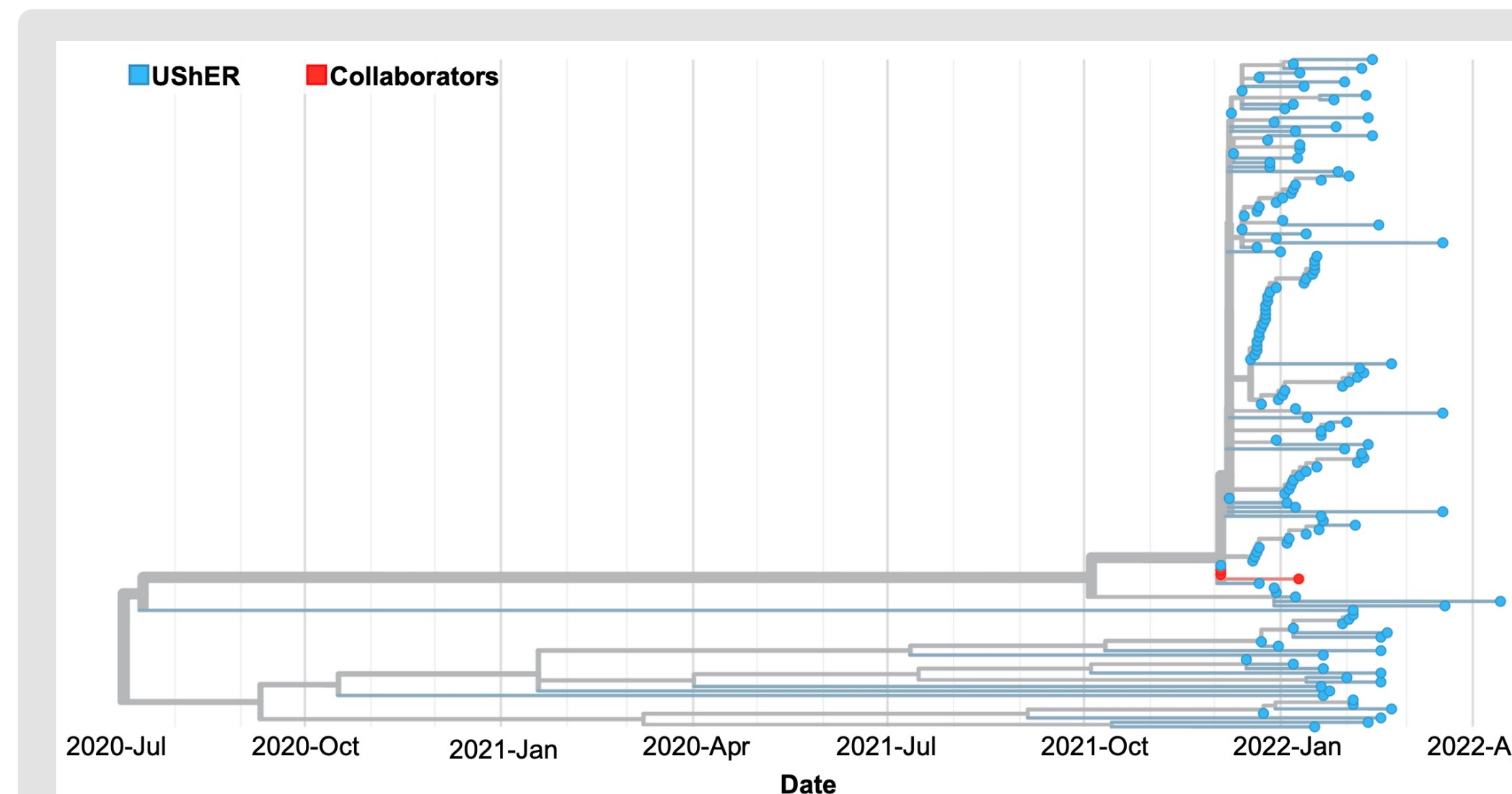


Figure 1. An estimated time-scaled phylogenetic tree was constructed with the subtree containing 3 Nigerian samples using TreeTime and visualized in Auspice^[8]. The red dots are the 3 Nigerian samples. The estimated collection dates for the 3 samples are in December 2021 and January 2022, later than the reported collection dates (September 2021).

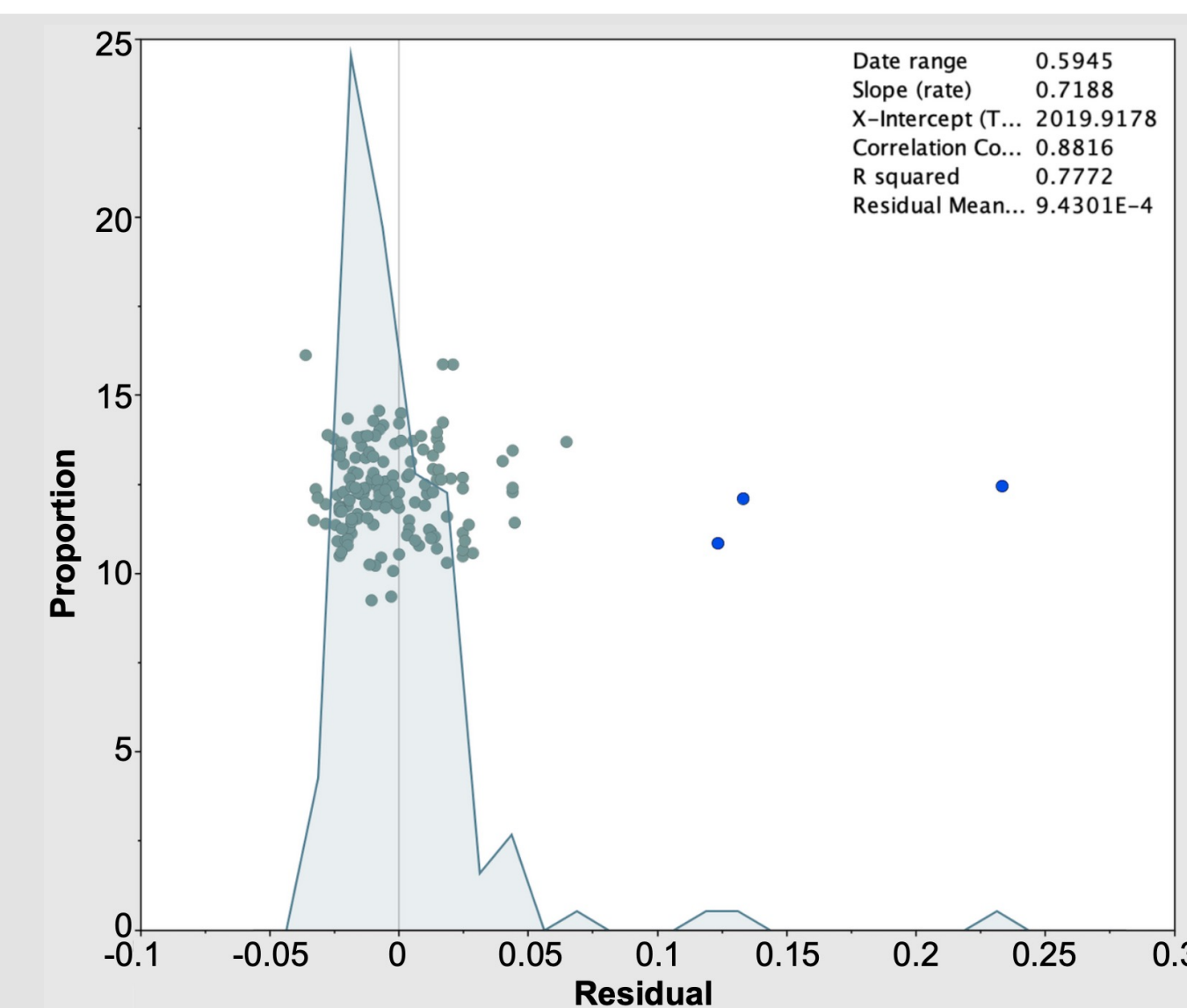


Figure 2. A residual (the tangential deviation from the regression line) plot was generated by TempEst using the time tree in Figure 1. The dark blue dots are the 3 Nigerian sequences. While TreeTime reconstructed a time-scaled tree, the collection dates of the 3 Nigerian sequences are not explained by time, and the sequences are still clear outliers, indicating that their reported dates may not be correct.

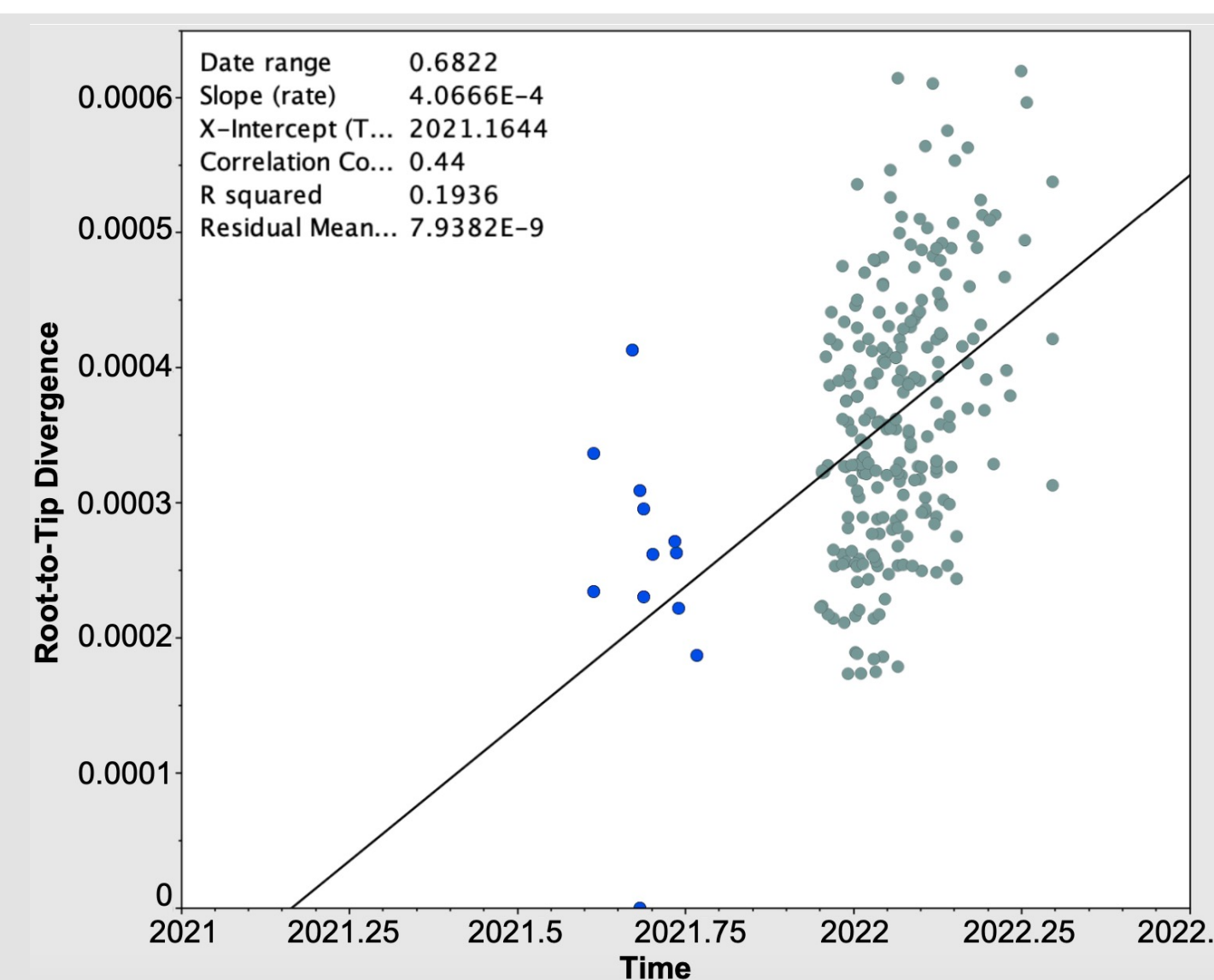


Figure 3. A plot of the divergence from the root of the tree against time of collection was generated by TempEst. There is a weak correlation (the correlation coefficient is 0.44) between the dates of the tips and the divergence for the global tree, and the slope is 4×10^{-4} which is not far from our expected slope of 4×10^{-3} . This indicates a BEAST run could be done. The dark blue dots are the 12 Nigerian sequences.

Conclusion

- Since the 12 Nigerian sequences are most closely related to sequences with later collection dates (Table 1), the collection dates of the Nigerian sequences could be inaccurate.
- The molecular clock phylogeny (Figure 1) and the residual plot (Figure 2) also infer that the 3 Nigerian sequences diverge later than November 2021.
- There is a relationship between time and the divergence of the roots for the global tree, meaning the sequences are suitable for a BEAST run (Figure 3).

Future Directions

- We will complete the Bayesian phylogenetic analysis to have a more accurate time tree and evolutionary rate which will help answer the question of whether the dates for these sequences are accurate.

References

- Omicron and its Subvariants: A Guide to What We Know (n.d.). Yale Medicine. Available at: <https://www.yalemedicine.org/news/5-things-to-know-omicron> (Accessed April 11, 2024).
- Onwuamah, C., Iwalokun, B., Ahmed, R., Toyé, E., Amoo, O., Ndodo, N., et al. (2022). The trend of SARS-CoV-2 variants from July 2021 to Jan 2022 in metropolitan Lagos, the epicentre of the COVID-19 pandemic in Nigeria. doi: [10.21203/rs.3.rs-1602118/v1](https://doi.org/10.21203/rs.3.rs-1602118/v1)
- Turakhia, Y., Thornlow, B., Hinrichs, A. S., De Maio, N., Gozashti, L., Lanfear, R., et al. (2021). Ultrafast Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat Genet 53, 809–816. doi: [10.1038/s41588-021-00862-7](https://doi.org/10.1038/s41588-021-00862-7)
- Sagunenko, P., Puller, V., and Neher, R. A. (2018). TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol 4, vex042. doi: [10.1093/ve/vex042](https://doi.org/10.1093/ve/vex042)
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Computational Biology 10, e1003537. doi: [10.1371/journal.pcbi.1003537](https://doi.org/10.1371/journal.pcbi.1003537)
- Carreon-Valencia, T. (CDC/NIOSH/WTCHP) (n.d.). Real-time phylogenetics with USHER.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32, 268–274. doi: [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300)
- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et al. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123. doi: [10.1093/bioinformatics/bty407](https://doi.org/10.1093/bioinformatics/bty407)

Author Contribution

Chika Kingsley Onwuamah and Anne Piantadosi conceived of the research idea and provided the data and guidance. Zijing Zhou conducted the data analysis and made the poster. All authors edited the poster.

Acknowledgements

I thank Tyshawn Ferrell for his help with TreeTime and my friend's cat for mental support.

