

ShUSHER: private browser-based placement of sensitive genome samples on phylogenetic trees

Alexander Kramer^{1, 2}, Yatish Turakhia³, and Russell Corbett-Detig^{1, 2}

1 Department of Biomolecular Engineering, University of California Santa Cruz. Santa Cruz, CA 95064, USA **2** Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA **3** Department of Electrical and Computer Engineering, University of California, San Diego; San Diego, CA 92093, USA

DOI: [10.21105/joss.03677](https://doi.org/10.21105/joss.03677)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Mikkel Meyer Andersen](#) ↗

Reviewers:

- [@kblin](#)
- [@bede](#)

Submitted: 27 August 2021

Published: 14 October 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

ShUSHER (Shh: private Ultrafast Sample placement on Existing tRees) is a browser-based application for local analysis of sensitive biological sequence data. It uses USHER, a previously developed algorithm ([Turakhia et al., 2021](#)), to place user-provided genome samples on an existing phylogenetic tree and return subtrees surrounding those samples. It displays visualizations of the results using Auspice, part of the Nextstrain project ([Hadfield et al., 2018](#)). UCSC hosts a [web tool](#) with similar functionality, but it performs server-side computation on user samples, making it unsuitable for samples containing Protected Health Information (PHI) or other sensitive data. There are two main components of ShUSHER: a port of the existing C++ USHER code to WebAssembly and a user interface built with React, which together perform computation on user-provided samples entirely client-side in a web browser. The web application is accessible at <https://shusher.gi.ucsc.edu/>.

Statement of need

Phylogenetic trees are often used to help trace the origin, spread, and evolution of viruses. The continuously growing number of sequenced SARS-CoV-2 genomes has quickly overwhelmed the capabilities of many existing tree construction methods. USHER is a method that can efficiently place newly sequenced genomes on existing, large phylogenetic trees. Many researchers have been using the USHER web tool hosted at UCSC to place their samples on existing global trees comprising millions of SARS-CoV-2 sequences. However, some jurisdictions consider viral genomes PHI, in which case the viral sequence data cannot be transmitted over the Internet. Presently, researchers cannot use the UCSC web tool for such data and must instead install and run USHER locally with a command-line application. ShUSHER is an alternative to this, providing a user-friendly platform for researchers to run USHER in their web browser using an existing tree of sequences while keeping their data private. Currently, ShUSHER supports placement of user samples on a [global tree](#) of publicly available SARS-CoV-2 samples ([McBroome et al., 2021](#)), and it may be extended in the future to support private analysis of other pathogen sequences.

Acknowledgements

ShUSHER uses or adapts code from several open source projects. The authors thank the many developers of [Nextclade](#), [Auspice](#), [auspice.us](#), and [USHER](#). We also thank Angie Hinrichs for

developing and maintaining essential web infrastructure upon which ShUSHER relies. This work was supported by CDC award BAA 200-2021-11554 and by a gift from the Eric and Wendy Schmidt Foundation.

References

- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34. <https://doi.org/10.1093/bioinformatics/bty407>
- McBroome, J., Thornlow, B., Hinrichs, A. S., Maio, N. D., Goldman, N., Haussler, D., Corbett-Detig, R., & Y. Turakhia. (2021). A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. In *bioRxiv*. Cold Spring Harbor Laboratory. <https://doi.org/10.1101/2021.04.03.438321>
- Turakhia, Y., Thornlow, B., Hinrichs, A. S., Maio, N. D., Gozashti, L., Lanfear, R., Haussler, D., & Corbett-Detig, R. (2021). Ultrafast Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nature Genetics*, 53. <https://doi.org/10.1038/s41588-021-00862-7>