

# genieR: An R package for inference of demographic history of phylogenies

Fei Xiang<sup>1</sup>, Bethany Dearlove<sup>2</sup>, and Simon Frost<sup>1</sup>

<sup>1</sup> Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom <sup>2</sup> US Military HIV Research Program, Rockville, MD, United States

DOI: [10.21105/joss.00634](https://doi.org/10.21105/joss.00634)

## Software

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

Submitted: 16 January 2018

Published: 05 January 2019

## License

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

## Summary

genieR is an R package (R Core Team (2016)) for the inference of demographic history from reconstructed molecular phylogenies (see Dearlove, Xiang, & Frost (2017) for one application), modeled after the C++ package GENIE (Pybus & Rambaut, 2002). The package processes phylogenetic trees in Newick format, in which taxa may be sampled either at the same time (isochronous) or at different times (heterochronous); the latter are particularly common in the analysis of pathogen sequence data. In addition to extracting basic information from a particular phylogeny such as sampling and coalescent times, genieR also fits a coalescent model (Kingman (1982)) under several demographic scenarios, including constant population size, exponential growth, expansion growth, logistic growth, with both continuous and piecewise variants. The fit of different models can be compared using Akaike's Information Criterion (AIC) (Akaike, 1973) values. genieR uses C++ code as an option for better computational performance, and models can be fit either using maximum likelihood or Markov Chain Monte Carlo (Hastings, 1970). genieR can also simulate phylogenies under different demographic models.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado: Budapest.
- Dearlove, B. L., Xiang, F., & Frost, S. D. W. (2017). Biased phylodynamic inferences from analysing clusters of viral sequences. *Virus Evolution*, *3*(2). doi:[10.1093/ve/vex020](https://doi.org/10.1093/ve/vex020)
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109. doi:[10.2307/2334940](https://doi.org/10.2307/2334940)
- Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability*, *19*, 27–43. doi:[10.2307/3213548](https://doi.org/10.2307/3213548)
- Pybus, O. G., & Rambaut, A. (2002). GENIE: Estimating demographic history from molecular phylogenies. *Bioinformatics*, *18*, 1404–1405. doi:[10.1093/bioinformatics/18.10.1404](https://doi.org/10.1093/bioinformatics/18.10.1404)
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>