

relped: Build Relatedness Pedigrees

Ryan A. Hagenson¹ and Caitlin J. Curry¹

¹ Omaha's Henry Doorly Zoo and Aquarium

DOI: [10.21105/joss.02124](https://doi.org/10.21105/joss.02124)

Software

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

Editor: [Marcos Vital](#) ↗

Reviewers:

- [@djmitche](#)
- [@rabdill](#)

Submitted: 19 December 2019

Published: 12 March 2020

License

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

Summary

Relatedness is a measure of shared genes and ranges from 0.0 (no relation) to 1.0 (exact clones) (Queller & Goodnight, 1989). By numerically tracing genetics through a family tree, or pedigree chart, we can establish: 0.5 relatedness between parent-offspring (50% of the offspring's genes are from one parent), ~0.5 relatedness of full-siblings (50% of genes from each parent individually, but not the same 50% to each offspring), ~0.25 relatedness of half-siblings (50% of genes from one parent individually, but not the same 50%) – this trend toward 0.0 continues the more distantly related two individuals are genetically.

Given the relatedness of a set of individuals as calculated by tools such as ML-Relate (Kalinowski, Wagner, & Taper, 2006), in the past the compiling biologist had to manually draw the resulting pedigree chart via manual reasoning on inferred connections – `relped` automates this process. By using a combination of relatedness values (with optional normalization) and codified entries, `relped` allows incremental refinement of the resulting pedigree network by replacing values with codes as familial determinations are definitively called. As well, there are options to include parentage and/or demographics information to better match the typical appearance of a standard pedigree chart. With parentage information provided, `relped` can differentiate between parent-offspring and full-sibling (both ≈ 0.5 related) relationships. Pedigree network connections made between individuals are consistent between runs of `relped`; therefore, the pedigree network can be redrawn by repeatedly running `relped` until the pedigree network is visually satisfactory. Given the findings of Staples et al. (2016), by default relationships beyond the 9th-level (e.g., 4th cousins) are not mapped. Any individuals related beyond this cutoff are left unmapped, but a list of these IDs can be saved.

Statement of Need

While relatedness can be calculated from a pedigree (Lynch & Walsh, 2018), there are not tools to build a pedigree chart from relatedness alone. This is primarily due to the amount of information required to produce the hierarchical tree structure within a pedigree chart. The valuable information from a pedigree chart is how individuals within the sampled population connect.

The primary use of `relped` is for visualization of genetic relationships when in-depth parentage information is not readily available. This tool creates a pedigree network to attain a better understanding of the connections formed from genetic pairwise comparisons within a sampled population.

A pedigree chart is a visual depiction of relationships across generations. A pedigree chart relies heavily on complete family histories and parentage information, which may be unknown or unattainable in some populations. Relationships between individuals are then reliant on calculated genetic relatedness (r_{xy} ; i.e., (Kalinowski et al., 2006; Peakall & Smouse, 2012; Queller & Goodnight, 1989)).

In contrast to a pedigree chart, `relped` creates a pedigree network to illustrate genetic connections to visualize relationships in a sampled population. The network can be generated with or without known parentage or demographic information. A dataset with minimal information (e.g., r_{xy} only of a subset of the population) will produce a network with genetic connections based on the information provided. As more information is provided, connections are better resolved. The resulting network built from relationship calls, parentage, and demographic information for all individuals in a population should result in a network that most closely resembles a pedigree chart.

References

- Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-relate: A computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6(2), 576–579. doi:[10.1111/j.1471-8286.2006.01256.x](https://doi.org/10.1111/j.1471-8286.2006.01256.x)
- Lynch, M., & Walsh, B. (2018). *Genetics and analysis of quantitative traits*. Sinauer Associates.
- Peakall, R., & Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28(19), 2537–2539. doi:[10.1093/bioinformatics/bts460](https://doi.org/10.1093/bioinformatics/bts460)
- Queller, D. C., & Goodnight, K. F. (1989). Estimating Relatedness Using Genetic Markers. *Evolution*, 43(2), 258. doi:[10.2307/2409206](https://doi.org/10.2307/2409206)
- Staples, J., Witherspoon, D. J., Jorde, L. B., Nickerson, D. A., Below, J. E., & Huff, C. D. (2016). PADRE: Pedigree-aware distant-relationship estimation. *The American Journal of Human Genetics*, 99(1), 154–162. doi:[10.1016/j.ajhg.2016.05.020](https://doi.org/10.1016/j.ajhg.2016.05.020)