mitolina

MITOchondrial LINeage Analysis
Andersen, Mikkel Meyer

Published in:
The Journal of Open Source Software

DOI (link to publication from Publisher):
10.21105/joss.01266

Creative Commons License
CC BY 4.0

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Andersen, M. M. (2019). mitolina: MITOchondrial LINeage Analysis. The Journal of Open Source Software, 4(34), [1266]. https://doi.org/10.21105/joss.01266

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: May 07, 2020
mitolina: MITOchondrial LINeage Analysis

Mikkel Meyer Andersen

1 Department of Mathematical Sciences, Aalborg University, Denmark

Summary

This R (R Core Team, 2018) package, mitolina (MITOchondrial LINeage Analysis), contains functionality to simulate and analyse populations of mitochondrial genomes (mitogenomes). This is achieved using both R and C++ via Rcpp (Eddelbuettel & Balamuta, 2017) for efficient computations.

Mitochondria are of interest for example in forensic genetics as they are very resistant to degradation and it is sometimes possible to construct a mitochondrial DNA profile when it is not possible to construct a traditional DNA profile (Andersen & Balding, 2018; Butler, 2009). This happens for example when the biological sample does not contain cell nuclei or when the sample is degraded (for example by time or environment).

Just as DNA profiles based on the Y chromosome are paternal lineage markers (inherited from fathers to boys) (Andersen & Balding, 2017; Butler, 2009), then DNA profiles based on the mitogenome are maternal lineage markers (inherited from mothers to children) (Andersen & Balding, 2018; Butler, 2009). This software operates under the maternal inheritance only model, i.e. that mtDNA is only passed on by mothers to children. It is often helpful to perform simulations of populations in lineage marker research as recent research on using Y chromosomal DNA profiles in forensic genetics demonstrate (Andersen, 2018; Andersen & Balding, 2017; Andersen, Caliebe, Jochens, Willuweit, & Krawczak, 2013; Andersen, Eriksen, & Morling, 2013).

This R package, mitolina, is based on the R package malan (Andersen, 2018) that simulates populations of Y chromosomes. The packages are fundamentally different in two aspects caused by the way that paternal and maternal lineage markers behave genetically. For example, with the mitogenome it is necessary to simulate both females and males (as males have their mother’s mitogenome), at least in the generations where the profiles must be used. Also, the genetic DNA profiles used in forensic genetics are different for the two types of lineage markers. A mitogenomic DNA profile can be seen as a profile of many thousands binary markers whereas a Y-profile consists of 10-20 integer valued markers.

The simulation model allows for flexible simulations by first simulating a genealogy (with population sizes at each generation specified by a vector for number of females and a vector for number of males) with various parameters such as variance in reproductive success (Andersen & Balding, 2017). It is possible to impose mitogenomes in various ways and several sets of mutation rates are included (Översti et al., 2017; Rieux et al., 2014; Soares et al., 2009). There are also 588 forensic-quality haplotypes representing three U.S. populations from (Just et al., 2015) included; they can for example be used to distribute founder haplotypes.

The mitolina software makes it possible to query the population and its individuals in multiple ways. For example to count the number of females or males in the population.
with a certain mitogenome. Or obtain the distribution of number of meioses between a queried contributor and the individuals in the population with a matching mitogenome.

The documentation of mitolina consists of manual pages for the various available functions, an article describing how to use the package (vignette), and unit tests.

Research using this software in interpretation of DNA profiles based on the mitogenome in forensic genetics is already published (Andersen & Balding, 2018) and the aim is that this software can help further research in this important topic.

I would like to thank David J Balding for helpful discussions.

References

Andersen, M. M. (2018). malan: MAle Lineage ANalysis. The Journal of Open Source Software, 3(25). doi:10.21105/joss.00684

Andersen, M. M., & Balding, D. J. (2017). How convincing is a matching Y-chromosome profile? PLOS Genetics, 13(11), e1007028. doi:10.1371/journal.pgen.1007028

Andersen, M. M., & Balding, D. J. (2018). How many individuals share a mitochondrial genome? PLOS Genetics, 14(11), e1007774. doi:10.1371/journal.pgen.1007774

Andersen, M. M., Caliebe, A., Jochens, A., Willuweit, S., & Krawczak, M. (2013). Estimating trace-suspect match probabilities for singleton Y-STR haplotypes using coalescent theory. Forensic Science International: Genetics, 7(2), 264–271. doi:10.1016/j.fsigen.2012.11.004

Andersen, M. M., Eriksen, P. S., & Morling, N. (2013). The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. Journal of Theoretical Biology, 329, 39–51. doi:10.1016/j.jtbi.2013.03.009

Butler, J. M. (2009). Fundamentals of Forensic DNA Typing. Academic Press.
Eddelbuettel, D., & Balamuta, J. J. (2017). Extending R with C++: A Brief Introduction to Rcpp. PeerJ Preprints, 5, e3188v1. doi:10.7287/peerj.preprints.3188v1

Just, R. S., Scheible, M. K., Fast, S. A., Sturk-Andreaggi, K., Röck, A. W., Bush, J. M., Higinbotham, J. L., et al. (2015). Full mtGenome reference data: Development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. Forensic Science International: Genetics, 14, 141–155. doi:10.1016/j.fsigen.2014.09.021

Översti, S., Onkamo, P., Stoljarova, M., Budowle, B., Sajantila, A., & Palo, J. U. (2017). Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity. Scientific Reports, 7. doi:10.1038/s41598-017-05673-7

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., Ruiz-Linares, A., et al. (2014). Improved Calibration of the Human Mitochondrial Clock Using Ancient Genomes. Molecular Biology and Evolution, 31(10), 2780–2792. doi:10.1093/molbev/msu222

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., et al. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. American Journal of Human Genetics, 84, 740–759. doi:10.1016/j.ajhg.2009.05.001