

David Gerard

Assistant Professor

American University

Department of Mathematics and Statistics

3501 Nebraska Ave NW, DMTI 106E

Washington, DC 20016, USA

Phone: +1-202-885-3022

Email: dgerard@american.edu

Homepage: <https://dcgerard.github.io/>

GitHub: <https://github.com/dcgerard/>

Last updated: February 11, 2024

Education

Ph.D. Statistics, University of Washington, June 2015.

Advised by Professor [Peter Hoff](#).

M.S. Statistics, The Ohio State University, June 2012.

B.S. Mathematics, The Ohio State University, June 2010.

B.S. Molecular Genetics, The Ohio State University, June 2010.

Professional Experience

Assistant Professor, American University: 08/2018 – Present.

Postdoctoral Scholar, University of Chicago: 09/2015 – 07/2018.

Advised by Professor [Matthew Stephens](#).

Articles

Peer Reviewed and Published

Gerard, D. (2023). Bayesian tests for random mating in polyploids. *Molecular Ecology Resources* 23(8), p. 1812–1822.

[<doi:10.1111/1755-0998.13856>](https://doi.org/10.1111/1755-0998.13856) [<bioRxiv:2022.08.11.503635>](https://doi.org/10.1101/2022.08.11.503635)

Gerard, D. (2023). Double reduction estimation and equilibrium tests in natural autopolyploid populations. *Biometrics* 79(3), p. 2143–2156.

[<doi:10.1111/biom.13722>](https://doi.org/10.1111/biom.13722) [<bioRxiv:2021.09.24.461731>](https://doi.org/10.1101/2021.09.24.461731)

Gerard, D. (2022). Comment on three papers about Hardy-Weinberg equilibrium tests in autopolyploids. *Frontiers in Genetics* 13, p. 1–6.

[<doi:10.3389/fgene.2022.1027209>](https://doi.org/10.3389/fgene.2022.1027209) [<HAL:hal-03754674>](https://hal.archives-ouvertes.fr/hal-03754674).

Gerard, D. (2021). Scalable bias-corrected linkage disequilibrium estimation under genotype uncertainty. *Heredity* 127(4), p. 357–362.

[<doi:10.1038/s41437-021-00462-5>](https://doi.org/10.1038/s41437-021-00462-5) [<bioRxiv:2021.02.08.430270>](https://doi.org/10.1101/2021.02.08.430270)

Gerard, D. (2021). Pairwise linkage disequilibrium estimation for polyploids. *Molecular Ecology Resources* 21(4), p. 1230–1242.

[<doi:10.1111/1755-0998.13349>](https://doi.org/10.1111/1755-0998.13349) [<bioRxiv:2020.08.03.234476>](https://doi.org/10.1101/2020.08.03.234476)

- Gerard, D.**, & Stephens, M. (2021). Unifying and generalizing methods for removing unwanted variation based on negative controls. *Statistica Sinica* 31(3), p. 1145–1166.
<[doi:10.5705/ss.202018.0345](https://doi.org/10.5705/ss.202018.0345)> <[arXiv:1705.08393](https://arxiv.org/abs/1705.08393)>
- Gerard, D.** (2020). Data-based RNA-seq simulations by binomial thinning. *BMC Bioinformatics* 21(1), 206, p. 1–14.
<[doi:10.1186/s12859-020-3450-9](https://doi.org/10.1186/s12859-020-3450-9)> <[bioRxiv:758524](https://bioRxiv.org/lookup/doi/10.1101/2020.03.11.331111)>
- Gerard, D.**, & Stephens, M. (2020). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics*, 21(1), p. 15–32.
<[doi:10.1093/biostatistics/kxy029](https://doi.org/10.1093/biostatistics/kxy029)> <[arXiv:1709.10066](https://arxiv.org/abs/1709.10066)>
- Gerard, D.** & Ferrão, L. F. V. (2020). Priors for genotyping polyploids. *Bioinformatics* 36(6), p. 1795–1800.
<[doi:10.1093/bioinformatics/btz852](https://doi.org/10.1093/bioinformatics/btz852)> <[bioRxiv:751784](https://bioRxiv.org/lookup/doi/10.1101/2020.03.11.331111)>
- Pollack, J. D., **Gerard, D.**, Makhatadze, G., & Pearl, D. K. (2019). Evolutionary conservation and structural localizations suggest a physical trace of metabolism’s progressive geochronological emergence. *Journal of Biomolecular Structure and Dynamics*, ja, p. 1–20.
<[doi:10.1080/07391102.2019.1679666](https://doi.org/10.1080/07391102.2019.1679666)>
- Gerard, D.**, Ferrão, L. F. V., Garcia, A. A. F., & Stephens, M. (2018). Genotyping Polyploids from Messy Sequencing Data. *Genetics*, 210(3), p. 789–807.
<[doi:10.1534/genetics.118.301468](https://doi.org/10.1534/genetics.118.301468)> <[bioRxiv:281550](https://bioRxiv.org/lookup/doi/10.1101/281550)>
- Gerard, D.**, & Hoff, P. (2017). Adaptive higher-order spectral estimators. *Electronic Journal of Statistics*, 11(2), p. 3703–3737.
<[doi:10.1214/17-EJS1330](https://doi.org/10.1214/17-EJS1330)> <[arXiv:1505.02114](https://arxiv.org/abs/1505.02114)>
- Gerard, D.**, & Hoff, P. (2016). A higher-order LQ decomposition for separable covariance models. *Linear Algebra and its Applications*, 505, p. 57–84.
<[doi:10.1016/j.laa.2016.04.033](https://doi.org/10.1016/j.laa.2016.04.033)> <[arXiv:1410.1094](https://arxiv.org/abs/1410.1094)>
- Gerard, D.**, & Hoff, P. (2015). Equivariant minimax dominators of the MLE in the array normal model. *Journal of Multivariate Analysis*, 137, p. 32–49.
<[doi:10.1016/j.jmva.2015.01.020](https://doi.org/10.1016/j.jmva.2015.01.020)> <[arXiv:1408.0424](https://arxiv.org/abs/1408.0424)>
- Pollack, J. D., **Gerard, D.**, & Pearl, D. K. (2013). Uniquely localized intra-molecular amino acid concentrations at the glycolytic enzyme catalytic/active centers of Archaea, Bacteria and Eukaryota are associated with their proposed temporal appearances on Earth. *Origins of Life and Evolution of Biospheres*, 43(2), p. 161–187.
<[doi:10.1007/s11084-013-9331-8](https://doi.org/10.1007/s11084-013-9331-8)>
- Gerard, D.**, Gibbs, H. L., & Kubatko, L. (2011). Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC evolutionary biology*, 11(1), 291, p. 1–12.
<[doi:10.1186/1471-2148-11-291](https://doi.org/10.1186/1471-2148-11-291)>

Preprints

- Gerard, D.**, Thakkar, M., & Ferrão L.F.V. (2024). Tests for segregation distortion in tetraploid F1 populations. *bioRxiv*.
<[doi:10.1101/2024.02.07.579361](https://doi.org/10.1101/2024.02.07.579361)>

Awarded Grants

External

National Science Foundation, Grant No. [2132247](#). “A U-statistic approach to population genetics”. \$188,694.00. 09/2021–12/2023. Role: PI.

Internal

College of Arts and Sciences Mathias Research Fellowship, the American University Department of Mathematics and Statistics, and the Howard Hughes Medical Institute. “Population Structure Estimation for Polyploids,”. \$5,000. 2022. Role: Faculty Advisor to Toe, S. C.

Department of Mathematics and Statistics Student Research Projects in the Mathematical Sciences. “Genotype Likelihood Approaches to Linkage Disequilibrium Estimation in Polyploids”. \$2000. 2022. Role: Faculty Advisor to Scartz, D. T.

Mellon Grant from the College of Arts and Sciences, American University. “Genome-wide Inference Under Uncertainty,”. \$1,200. 2018–2019. Role: PI.

Teaching Experience

Advanced R 01/2022 – 05/2022

This course prepares students for software development using the R statistical programming language. Topics include: version control via git and its command line interface, package development, unit testing, base R data structures, environments, object oriented programming via S3 and S4, metaprogramming, interfacing with C++ via Rcpp, parallel processing with futures, and batch scheduling jobs via LSF.

<https://dcgerard.github.io/advancedr/>.

Regression 08/2021 – 12/2021, 08/2022 – 12/2022

This is a standard regression analysis course for students who have a very light statistical background. We cover the simple/multiple linear regression model, interpretation, estimation and testing procedures, diagnostics, model building, and practical considerations. Students are taught how to implement regression using the R statistical programming language.

https://dcgerard.github.io/stat_415_615/.

Data Science 08/2019 – 12/2020, 08/2023 – 12/2023

This course trains students on a suite of tools necessary for a successful career in Data Science, including version control (git), large(ish) data analysis (data.table), and data communication (R Shiny). I also provide a review of basic statistical inference and emphasize how one performs such inference in R. I teach this course almost entirely using GitHub Classroom.

<https://data-science-master.github.io/lectures/>.

Statistical Programming in R 01/2019 – 05/2019, 08/2019 – 12/2019

This course takes students with little or no background in computing and guides them through most of the material covered in R for Data Science.

https://dcgerard.github.io/stat_412_612/.

Biostatistics 08/2018 – 12/2018

Introductory biostatistics at the undergraduate level. This serves as the main statistical methods course for Public Health Science majors. I reviewed basic statistical concepts before introducing the students to ANOVA, linear models, and logistic regression, all within the framework of public health applications.

Statistical Methods 08/2018 – 12/2018

This course was designed to get physical and social science masters students proficient enough to analyze their own data and convey statistical results to their peers. This course focused on actually analyzing data, interpreting results, and effectively communicating statistical conclusions.

<https://dcgerard.github.io/stat614/>.

Lecturer of introductory statistics (University of Chicago) 09/2017 – 12/2017

This course was a standard introduction to Statistics for non-majors plus some light mathematical theory. The content included introductions to probability, inference, and the R statistical computing language.

<https://dcgerard.github.io/stat234/index.html>.

Lecturer of introductory statistics (University of Washington) 06/2013 – 09/2013

This course was designed to introduce Statistics to students who had little mathematical maturity. Course content included comparisons of good sampling designs versus bad sampling designs, conditions under which claims of causality may be made, confounding variables, good and bad measures for answering questions of interest, describing relationships between variables, and good and bad graphics to describe data.

Graduate Teaching Assistant (University of Washington) 09/2012 – 06/2013

Lecturer of introductory statistics (The Ohio State University) 09/2011 – 12/2011

Phylogenetic Workshop Assistant (The Ohio State University) 06/2011 – 08/2011

Phylogenetic Workshop Assistant (The Ohio State University) 06/2010 – 08/2010

Undergraduate Teaching Assistant (The Ohio State University) 09/2008 – 12/2008

Research Project Supervision

Supervised Master's Theses

Sun, R. (2020). Estimating preferential pairing in polyploids. *Master's thesis*. American University. p. 1–47.

[doi:10.57912/23856744](https://doi.org/10.57912/23856744) <[ProQuest:2436381968](https://proquest.com/docview/2436381968)>

Matoka Nana, K. (2023). Exact tests for random mating in autotetraploids. *Master's thesis*. American University. p. 1–38.

[doi:10.57912/23504199](https://doi.org/10.57912/23504199) <[ProQuest:2808426065](https://proquest.com/docview/2808426065)>

Thakkar, M. (2023). Bayesian tests for segregation distortion in experimental tetraploid populations. *Master's thesis*. American University. p. 1–58.

[doi:10.57912/24256936](https://doi.org/10.57912/24256936) <[ProQuest:2847679939](https://proquest.com/docview/2847679939)>

Supervised Research Capstones

Supervised Research, Yuanhan Xu (Graduate), “Improving Fruit Flavor Through Machine Learning and Learning to Rank Algorithms,” Complete. (2021)

Supervised Research, Tianqi Luo (Graduate), “Variational Methods on Genotyping Polyploids,” Complete. (2019)

Supervised Research (joint with Dr. Laurie Bayet), Gemei Li (Graduate), “Multivariate Pattern Analysis on EEG Data,” Complete. (2019)

Other Supervised Research Projects

Tyson, A. (2022–Present), “Tests for unimodal genotype frequencies”.

Hu, H. & Scartz, D. T. (2021–2022), “Genotype likelihood approaches for estimating linkage disequilibrium”.

Habiby, D. (2022), “Genotyping errors in autopolyploid populations of white sturgeon”.

Toe, S. C., Stano, B., & Bandreddi, A. (2022), “Population structure estimation in polyploids”.

Awards

Editor’s choice article for *Heredity* Volume 127 Issue 4, September 2021.

Highlighted article for *Genetics* Volume 210 Issue 3, November 2018.

Z.W. Birnbaum Award for the best general exam of the 2013-2014 academic year.

Distinguished University Fellowship (September 2010 - June 2012).

Graduated Summa Cum Laude (June 2010).

Medalist Scholarship (September 2006 - June 2010).

Presentations

Gerard, David, Mira Thakkar, & Luis Felipe Ventrone Ferrão. “Tests for Segregation Distortion in F1 Populations of Tetraploids.” Tools for Polyploids Workshop 2024, San Diego, California, 10–11 January 2024.

<https://www.polyploids.org/sites/default/files/2024-01/f1pres.pdf>

https://youtu.be/tBkB_su3r74

Gerard, D. C., “Bayesian Tests for Random Mating in Autopolyploids,” Poster Presented at: Plant & Animal Genome Conference 30. <https://doi.org/10.5281/zenodo.7557197> (January 16, 2023)

Scartz, D. Thomas, Hu, Hanwei, & Gerard, David. “Comparison of LD Estimation Methods Under Genotype Uncertainty”. Poster Presented at: Joint Statistical Meetings (JSM), Washington, DC. <https://doi.org/10.5281/zenodo.6987180> (August 8, 2022)

Gerard, D. C., “Double Reduction Estimation and Equilibrium Tests in Natural Autopolyploid Populations,” Poster Presented at: Plant & Animal Genome Conference XXIX. <https://pag.confex.com/pag/xxix/meetingapp.cgi/Paper/43285> (January 10, 2022)

Gerard, D. C., “Double Reduction Estimation and Equilibrium Tests in Natural Autopolyploid Populations,” Mathematics and Statistics Colloquium, American University. <https://youtu.be/mxTvokBhA8U> (October 19, 2021). Slides available at <https://doi.org/10.5281/zenodo.5589844>.

Gerard, D. C., “The updog R package for genotyping polyploids,” Polyploid Webinar, University of Arizona, Online. Scope: International. <https://youtu.be/9NSOQ3tLwxw> (June 15, 2020)

Gerard, D. C., “Priors for Genotyping Polyploids,” R. Clifton Bailey Statistics Seminar Series, George Mason University, Department of Statistics, HUB, Room 3 4400 University Drive, Fairfax, VA 22030. Scope: Regional. (October 18, 2019)

Gerard, D. C., “Priors for Genotyping Polyploids,” 2019 WNAR/IMS/JR Annual Meeting, Portland, Oregon. (June 24, 2019)

Gerard, D. C., “Priors for Genotyping Polyploids,” Mathematics and Statistics Colloquium, American University. (September 25, 2018)

Coding

Most of my research involves a large computational component. I make sure to implement and release all methods I develop.

Authored R Packages:

CRAN: [hwep](#), [ldsep](#), [seqgendiff](#), [updog](#), [tenstr](#),

GitHub: [hose](#), [vicar](#), [rmexact](#), [phwelike](#), [hexocto](#).

Advanced training: R, \LaTeX .

Intermediate training: C++, Shell.

Familiar: Python, Matlab/Octave, Mathematica.

Reproducible Research

I emphasize extreme reproducibility in my research projects. All analysis scripts are publicly posted, with code linked to data. Most of the repositories below require only a single command line to reproduce all of the results of the paper.

Gerard (2023a): <https://github.com/dcgerard/hwesims>

Gerard (2023b): https://github.com/dcgerard/rmbayes_sims

Gerard (2022): <https://github.com/dcgerard/hwesupp>

Gerard (2021a): https://github.com/dcgerard/ld_simulations

Gerard (2021b): https://github.com/dcgerard/ldfast_sims

Gerard and Stephens (2021): https://github.com/dcgerard/ruvb_sims

Gerard (2020): https://github.com/dcgerard/reproduce_fasims

Gerard and Stephens (2020): https://github.com/dcgerard/mouthwash_sims

Gerard and Ferrão (2019): https://github.com/dcgerard/reproduce_prior_sims

Gerard et. al. (2018): https://github.com/dcgerard/reproduce_genotyping

Gerard and Hoff (2017): https://github.com/dcgerard/reproduce_sure