

# 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](mailto:dgerard@american.edu)

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

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

Last updated: September 28, 2021

## Education

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

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.

## Articles

### Peer Reviewed and Published

**Gerard, D.** (*in press*). Scalable Bias-corrected Linkage Disequilibrium Estimation Under Genotype Uncertainty. *Heredity*.

<[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://doi.org/10.1101/2020.08.03.234476)>

**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://doi.org/10.1101/2020.08.03.234476)>

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://doi.org/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.** (2021). Double reduction estimation and equilibrium tests in natural autopolyploid populations. *bioRxiv*  
<[doi:10.1101/2021.09.24.461731](https://doi.org/10.1101/2021.09.24.461731)>

## Grants

National Science Foundation, Grant No. [2132247](https://www.nsf.gov/awardsearch/showAward.do?awardNumber=2132247). “A U-statistic approach to population genetics”. \$188,694.00. 09/2021–08/2023. Role: PI.

## Teaching Experience

Regression 08/2021 – 12/2021

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. My teaching website is publicly available at [https://dcgerard.github.io/stat\\_415\\_615/](https://dcgerard.github.io/stat_415_615/).

Data Science 08/2019 – 12/2020

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. My teaching website is publicly available at <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](#). My teaching website for this course is publicly available at [https://dcgerard.github.io/stat\\_412\\_612/](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. My teaching website for this course is publicly available at <https://dcgerard.github.io/stat614/>.

Lecturer of introductory statistics 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. My teaching website for this course is publicly available at <https://dcgerard.github.io/stat234/index.html>.

Lecturer of introductory statistics 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 09/2012 – 06/2013

Led four one-hour recitation sections of introductory statistics every week. These were small group discussions that acted primarily as support for the main content taught during lecture.

Lecturer of introductory statistics 09/2011 - 12/2011

Lectured to large groups of students using a pre-defined course plan.  
Created my own slides to follow along with the lesson plans.

Phylogenetic Workshop Assistant 06/2011 – 08/2011

Led a group of undergraduate researchers studying HIV. We studied the association between evolutionary distance (measured by branch distance along estimated trees) and clinical variables collected by an HIV treatment center in Belgium.

Phylogenetic Workshop Assistant 06/2010 – 08/2010

Led lab sessions on Unix and various phylogenetic program packages.

Led a group of undergraduates researching the evolutionary origins of flightless birds. Individually mentored an undergraduate student on phylogenetic research and statistical analysis with R. We tried to incorporate protein structure into our gene tree estimates while studying the evolutionary relationships of HPV.

Undergraduate Teaching Assistant: 09/2008 – 12/2008

Taught four one-hour recitation sections of intermediate algebra every week.

## Research Project Supervision

### Supervised Master Thesis

Sun, R. (2020). Estimating preferential pairing in polyploids. *Master's thesis*. American University. p. 1–56.

<<http://hdl.handle.net/1961/auislandora:85274>> <ProQuest:28027743>

### 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)

## Research Experience

Postdoctoral Scholar under Professor [Matthew Stephens](#): 09/2015 - 07/2018

Developed empirical Bayes methods to account for hidden confounding in large scale gene expression studies.

Efficiently implemented these methods in R and C++.

Rigorously tested these methods using Monte Carlo methods.

Lead/mentored graduate students on their individual projects.

Authored reports and academic papers.

Research Assistant Under Professor [Peter Hoff](#): 09/2013 - 06/2015

Studied optimal estimation of covariance matrices in Gaussian arrays using the notions of equivariance.

Developed novel tensor decompositions, creating higher-order versions of well-known matrix decompositions.

Developed singular value shrinkage mean estimators for tensor-variate data.

Research Assistant Under Professor [Joe Verducci](#): 09/2011 - 06/2012

Studied monotone association between gene expression and drug resistance within unspecified subtypes of cancer.

Undergraduate Research Fellow Under Professors [Laura Kubatko](#) and [H. Lisle Gibbs](#): 04/2009 - 06/2010

Studied the hybridization of two subspecies of Missouri Rattlesnakes and two populations of Black Ratsnakes.

Duties included sequencing work in the lab, computer programming using Perl, and data analysis with various programs.

## Awards

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. This is the only research award presented by the University of Washington Statistics Department.

Distinguished University Fellowship (September 2010 - June 2012).

Graduated Summa Cum Laude (June 2010).

Medalist Scholarship (September 2006 - June 2010).

## Presentations

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, American University. (September 25, 2018)

## Coding

R Packages: [hwep](#), [ldsep](#), [seqgendiff](#), [updog](#), [tenstr](#), [hose](#), [vicar](#).

Advanced training: R, L<sup>A</sup>T<sub>E</sub>X.

Intermediate training: C++, Shell.

Familiar: Python, Matlab/Octave, Mathematica.

## Reproducible Research

I emphasize extreme reproducibility in my research projects.

Gerard (2021a): [https://github.com/dcgerard/ld\\_simulations](https://github.com/dcgerard/ld_simulations)

Gerard (2021b): [https://github.com/dcgerard/ldfast\\_sims](https://github.com/dcgerard/ldfast_sims)

Gerard (2021c): <https://github.com/dcgerard/hwesims>

Gerard and Stephens (2021): [https://github.com/dcgerard/ruvb\\_sims](https://github.com/dcgerard/ruvb_sims)

Gerard (2020): [https://github.com/dcgerard/reproduce\\_fasims](https://github.com/dcgerard/reproduce_fasims)  
Gerard and Stephens (2020): [https://github.com/dcgerard/mouthwash\\_sims](https://github.com/dcgerard/mouthwash_sims)  
Gerard and Ferrão (2019): [https://github.com/dcgerard/reproduce\\_prior\\_sims](https://github.com/dcgerard/reproduce_prior_sims)  
Gerard et. al. (2018): [https://github.com/dcgerard/reproduce\\_genotyping](https://github.com/dcgerard/reproduce_genotyping)  
Gerard and Hoff (2017): [https://github.com/dcgerard/reproduce\\_sure](https://github.com/dcgerard/reproduce_sure)

## Data Analysis

Data Munging (tidyverse).  
Data Visualization (ggplot2).  
Reproducible Data Analysis (rmarkdown/make).  
Version Control (Git/Github).  
Continuous Integration (Travis CI/AppVeyor).  
R package development.