

C. J. Battey, PhD

Computational Biologist
NIH F32 Postdoctoral Fellow

email: cjbattey@gmail.com
twitter: [@cj_battey](https://twitter.com/cj_battey)
website: cjbattey.com
github: github.com/cjbattey

Summary

I'm a computational biologist specializing in spatial population genetics and applications of machine learning to genetic data. I have a Ph.D. in biology from the University of Washington and have published research in evolutionary genetics and bioinformatics in top journals including *eLife*, *GENETICS* and *Evolution*. My recent projects have focused on applications of deep learning to modeling spatial genetic variation in humans and improving controls for population structure in genome-wide association studies.

Education

University of Washington, Seattle 2013-2018 Ph.D., Biology
University of California, Berkeley 2006-2010 B.A., Integrative Biology & English

Experience

- 2020-** NIH F32 Postdoctoral Fellow, Kern Lab, Dept of Biology, University of Oregon, Eugene, OR
- Developing simulation and machine learning tools for predicting traits from genomes.
- 2018-2020** Postdoctoral Researcher, Kern Lab, Dept. of Biology, University of Oregon, Eugene, OR
- Conducted research on genetic variation in spatially varying populations
- Developed a supervised machine learning method for predicting spatial locations from genotypes.
- Tested applications of variational autoencoders to visualizing genetic variation
- 2013-2018** PhD Student. Dept. of Biology, University Washington, Seattle, WA
- Research on demographic modeling and population structure in migratory birds
- Taught labs in population genetics, comparative anatomy, physiology, and conservation biology
- Developed interactive teaching apps for population genetics: <https://cjbattey.shinyapps.io/driftR/>
- 2010-2012** Staff Scientist. Cardno ENTRIX, Concord, CA
- Survey and monitoring of protected species in northern California.
- Managed team conducting rapid environmental assessments for small utility projects.

Selected Publications

full list at https://scholar.google.com/citations?hl=en&user=BxCj0TUA AAAAJ&view_op=list_works&sortby=pubdate

"Predicting Geographic Location from Genetic Variation with Deep Neural Networks"

C. J. Battey, Peter L. Ralph, Andrew D. Kern. 2019. in press, *eLife*.
<https://elifesciences.org/articles/54507>

"Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data"

C. J. Battey, Peter L. Ralph, Andrew D. Kern. 2020. *GENETICS*. <https://doi.org/10.1534/genetics.120.303143>

"Minor Allele Frequency Thresholds Strongly Affect Population Structure Inference with Genomic Data Sets"

Ethan Linck and C. J. Battey. *Mol Ecol Resour.* 2019. <https://doi.org/10.1111/1755-0998.12995>

"Evidence of Linked Selection on the Z Chromosome of Hybridizing Hummingbirds."

C. J. Battey. 2020. *Evolution*. <https://doi.org/10.1111/evo.13888>

Funding & Awards

NIH NRSA F32 Postdoctoral Fellowship, 2020-2022
"Computational Methods for Next-Generation GWAS"

NSF Doctoral Dissertation Improvement Grant, 2016-2018
"Investigating Genomic Signatures of Range Shifts and Demographic Change in Migratory Hummingbirds"

WRF Hall Fellowship, UW Dep Biology, 2013

Travel awards: Society for Molecular Biology and Evolution, American Society of Naturalists, American Ornithologists' Union, UW Dept Biology, UW Graduate School

Skills

Code: R, python, javascript+html+css, awk, bash

Bioinformatics: illumina whole genome assembly, alignment, phasing, and SNP calling with GATK, samtools, and ANGSD.

Computing: SLURM, Amazon EC2, Unix tools, git.

Statistics: Linear models, mixed models, classical stats (t-tests, anovas, etc), GWAS, genetic demographic modeling.

Machine Learning: Decision trees, SVMs, neural networks (tensorflow), variational autoencoders

Geospatial analysis & visualization: time series and geospatial modeling in R (raster, sp, ggplot2). Experience in geospatial modeling with tensorflow neural networks.