

Jared Galloway

Curriculum vitae

jaredgalloway07@gmail.com

+1-406-579-6768



PERSONAL STATEMENT

I am a Data Scientist with a computer and information science background, and a focus on bioinformatics. I have seven years of experience in an academic research environment conducting various flavors of methods development, implementation, and downstream analysis in the biological sciences. My hands-on experience has provided a strong intuition for taking noisy empirical data through the various steps of filtering artifacts, normalization, modeling, and visualization to produce tidy, coherent, and reproducible datasets. My education in product design is often reflected in my work as I find particular interest in creating intuitive and aesthetically pleasing data visualization and software interfaces. I take pride in clearly communicating my work, and enjoy the discussion which follows.

EDUCATION

- **Master of Science, Biology** University of Oregon
Knight Campus - Bioinformatics and Genomics Program *June 2019 – December 2020, Eugene, OR*
- **Bachelor of Science, Computer & Information Science** University of Oregon
College of Arts and Sciences - Departmental Honors - Minor in Product Design *Aug 2013 – July 2018, Eugene, OR*

TECHNICAL SKILLS

- **Languages:** Python, R, C/C++, Bash, Java, Javascript
- **Mathematics:** Advanced Statistics and Bayesian Modeling, Linear Algebra, Discrete Math, Calculus
- **Operating systems:** Linux, OSX, Windows
- **Software:** Git, GitHub Actions, Travis CI, Quay GNU make, cmake, scons, zshell/bash, vim, tmux, VSCode, L^AT_EX, html, markdown, SLURM-based HPC clusters, AWS S3, Conda/Mamba, Docker, Apptainer (formerly Singularity) JAX, Tensorflow/Keras, PyTorch, Pandas, Polars, Numpy, Xarray, Matplotlib, Plotnine (grammar of graphics), Altair (interactive visualization), Scikit-learn, SciPy, BioPython, Nextflow, Snakemake

EXPERIENCE

- **Bioinformatics Analyst II** Fred Hutchinson Cancer Center
Public Health Sciences Division - Matsen Lab *Jan 2021 - Current, Bozeman, MT (Remote)*
 - **Modeling Deep Mutational Scanning datasets:** Developed a novel approach for jointly modeling multiple distinct DMS experimental datasets. The method was validated via simulation and integrated into a JAX-based python package, [multidms](#). Application of this method on empirical data has resulted the identification of specific shifts in mutational effect between three separate homologs of the SARS-CoV-2 spike protein [1].
 - **Algorithmic design for mutational fitness estimation:** Implemented complex algorithm for estimating mutational fitness trajectories (through time) for every possible codon-mutation that may have occurred within a phylogeny (evolutionary history) of > 13 million empirical viral genomes. Using the Snakemake workflow manager, I integrated this together with upstream quality control and downstream analysis to produce the results driving an ongoing study.
 - **Genomic surveillance of viral escape:** Built and deployed custom, public Nextstrain pages for viruses such as [SARS-CoV-2](#), Influenza, Lassa, and others – each of which provide estimates of viral escape to antibodies of interest. This work ultimately has provided analysis tools for various studies, as well as inform the potential effectiveness of these antibodies as therapeutics.
 - **Bioinformatics for studying affinity maturation in germinal centers:** Created an efficient Nextflow pipeline which leveraged high performance computing clusters and software containerization to perform over 50K individual and produce > 140 feature-rich phylogenetic trees of B-cell lineages within germinal centers of engineered mice. This work has provided the foundational results for an ongoing large-scale, multi-lab collaborative study.
 - **Administrative data management:** Contributed to the documentation and execution of best-practice management and storage of sensitive data in various computational environments including HPC file systems and AWS S3 buckets. Additionally was responsible for scrubbing sensitive data from repositories before publication.
 - **Visualization/Communication of results:** Visualized data for researchers and produced finalized figures for publication to clearly communicate patterns in data. Additionally, I regularly presented results and background of my research projects to a range of audiences including lab meetings, collaborative meetings, and team meetings.
 - **Collogue consultation:** Provided various degrees of guidance and education to undergraduate interns, graduate students, postdocs, and various other faculty around the hutch with computational tasks surrounding active research projects as well as employee onboarding. This includes, but is not limited to, algorithmic implementation, command line interface, HPC clusters, batch submission (SLURM managed), workflow management, machine learning model exploration, data formatting, version control, compute environments, and build systems.

◦ Research Associate

Computational Biology – Matsen Lab

Fred Hutchinson Cancer Center

March 2020 - Dec. 2020, Seattle, WA

- **PhIP-Seq software and analysis:** Developed a suite of software tools, [phippy](#), for analysis of raw Phage ImmunoPrecipitation Sequencing (PhIP-Seq) data that consisted of; a python package for modeling and computing enrichment statistics, an interactive application for visualizing results of the processed data, and an end-to-end (raw data → results) Nextflow pipeline for encapsulating the common analysis workflows in a portable and reproducible manner. In addition to informing best-practices and future experiments carried out in wet lab, This software was used to process data and produce results for multiple collaborative publications and ongoing studies [3, 4, 5, 11, 13, 15].
- **Neural Networks for global epistasis modeling:** Worked with a graduate student to implement a global epistasis modeling software package, [torchdms](#), using the PyTorch library.
- **Team Organization:** Set up and led team meetings to brainstorm, set, short-term goals, and present results for a range of team members across multiple labs and studies.

◦ Graduate Educator (BIO 410/510)

Biology Department - Graduate Teaching Fellow

University of Oregon

Aug 2019 - March 2020, Eugene, OR

- **Teaching:** Led discussion, taught lessons and proctored exams for four classroom hours each week for a graduate level class focused around Python/Bash scripting, Unix file formatting/parsing, and other algorithmic thinking activities.
- **Assignment Design:** Constructed Docker containers and a Jupyter notebook auto-grade pipeline for students to have a unified environment for Python programming. Hosted office hours, and graded ≈ 100 assignments/exams per week with feedback.

◦ Scientific Programmer

Institute of Ecology and Evolution – Ralph & Kern Co-Lab

University of Oregon

June 2018 - June 2019

- **Software Engineering:** Leveraged Software engineering best-practices such as architecture design, version control, and continuous integration for various open-source software projects including [tskit](#), [stdpopsim](#), and [ReLERNN](#).
- **Experimentation and model development:** Quantified the application of simulated data for training recurrent deep learning models (RNN's) to predict recombination rate estimates across the consensus genome of empirical populations. This work provided the analysis pipeline used in the resulting publication [6].
- **Analysis of population demography inference tools:** Constructed a snakemake pipeline for published analysis outlining a comparison of population demography inference tools [16].

◦ Undergraduate Research Assistant

Institute of Ecology and Evolution – Ralph & Kern Co-Lab

University of Oregon

June 2017 - June 2018, Eugene, OR

- **Algorithmic implementation of tree sequence recording:** Implemented tree sequence recording algorithm into widely used population genetic simulation software, [SLiM](#), and contributed to the writing in the resulting publication [8].
- **Algorithmic benchmarking:** Utilized cluster resources, visualization tools, and more to compare large scale simulations and benchmark performance across a range of metrics including memory allocation and user runtime. This was written up and formally presented in my [undergraduate honors thesis](#).
- **Simulation studies of rapid evolution:** Developed and analyzed simulation models to study migration patterns involved in rapid evolution of stickleback fish on the Alaskan coast. I then lead the publication of this study [7].

◦ Discrete Math Course Grader

Mathematics Department

University of Oregon

June 2016 - June 2017, Eugene, OR

- **Grading:** Graded and constructed feedback for ≈ 100 assignments per week focused on proofs, set theory, and combinatorics.

HIGHLIGHTED SOFTWARE PROJECTS

◦ multidms



Using google's JAX library, this open-source python package provides the utilities for jointly modeling multiple distinct Deep Mutational Scanning datasets. The interface includes a number of tools which help to easily prepare and encode data initialize and fit models, perform out-of-sample prediction, as well as interactively visualizing the resulting parameter values of interest. This software adds to the growing list of tools for modeling DMS data, and has been used to produce results for multiple studies.

◦ phippy



The phippy software suite provides tools for analyzing data from phage display methods that use immunoprecipitation and deep sequencing to capture antibody binding to peptides, often referred to as PhIP-Seq. It has three main components that can be used separately or in conjunction: (i) a Nextflow pipeline, [phip-flow](#), to process raw sequencing data into a compact, multidimensional dataset format and allows for end-to-end automation of reproducible workflows. (ii) a Python API, [phippy](#), which provides interfaces for tasks such as count normalization, enrichment calculation, multidimensional scaling, modeling for significant enrichment of peptides, and finally (iii) a Streamlit application, [phip-viz](#), as an interactive interface for visualizing the data as a heatmap in a flexible manner.

◦ ReLERNN



A python-based CLI for recombination-landscape estimation using recurrent neural networks ([ReLERNN](#)). This software implements a deep learning method for estimating a genome-wide recombination map that is accurate even with small numbers of pooled or individually sequenced genomes. Rather than use summaries of linkage disequilibrium as its input, ReLERNN takes columns from a genotype alignment, which are then modeled as a sequence across the genome using a recurrent neural network.

FIRST/CO-FIRST AUTHOR PUBLICATIONS

- [1] H. K. Haddox, J. G. Galloway, B. Dadonaite, J. D. Bloom, F. A. Matsen, and W. S. DeWitt. Jointly modeling deep mutational scans identifies shifted mutational effects among sars-cov-2 spike homologs. *Under Review at PNAS*, 2023. <http://dx.doi.org/10.1101/2023.07.31.551037>.
- [2] J. G. Galloway, K. Sung, S. S. Minot, M. E. Garrett, C. I. Stoddard, A. C. Willcox, Z. A. Yaffe, R. Yucha, J. Overbaugh, and F. A. Matsen. phipperry: a software suite for phip-seq data analysis. *Bioinformatics*, 39(10), Sept. 2023. <http://dx.doi.org/10.1093/bioinformatics/btad583>.
- [3] M. E. Garrett, J. G. Galloway, C. Wolf, J. K. Logue, N. Franko, H. Y. Chu, F. A. Matsen, and J. M. Overbaugh. Comprehensive characterization of the antibody responses to SARS-CoV-2 spike protein finds additional vaccine-induced epitopes beyond those for mild infection. *eLife*, 11, jan 2022. <https://doi.org/10.7554%2Felife.73490>.
- [4] C. I. Stoddard, J. Galloway, H. Y. Chu, M. M. Shipley, K. Sung, H. L. Itell, C. R. Wolf, J. K. Logue, A. Magedson, M. E. Garrett, K. H. Crawford, U. Laserson, F. A. Matsen, and J. Overbaugh. Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. *Cell Reports*, 35(8):109164, may 2021. <https://doi.org/10.1016%2Fj.celrep.2021.109164>.
- [5] M. E. Garrett, J. Galloway, H. Y. Chu, H. L. Itell, C. I. Stoddard, C. R. Wolf, J. K. Logue, D. McDonald, H. Weight, F. A. Matsen, and J. Overbaugh. High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies. *Cell*, 184(11):2927–2938.e11, may 2021. <https://doi.org/10.1016%2Fj.cell.2021.04.045>.
- [6] J. R. Adrion, J. G. Galloway, and A. D. Kern. Predicting the landscape of recombination using deep learning. *Molecular Biology and Evolution*, 37(6):1790–1808, feb 2020. <https://doi.org/10.1093%2Fmolbev%2Fmsaa038>.
- [7] J. Galloway, W. A. Cresko, and P. Ralph. A few stickleback suffice for the transport of alleles to new lakes. *G3 Genes|Genomes|Genetics*, 10(2):505–514, feb 2020. <https://doi.org/10.1534%2Fg3.119.400564>.
- [8] B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19(2):552–566, feb 2019. <https://doi.org/10.1111%2F1755-0998.12968>.

SUPPORTING AUTHOR PUBLICATIONS

- [9] F. C. Welsh, R. T. Eguia, J. M. Lee, H. K. Haddox, J. Galloway, N. V. V. Chau, A. N. Loes, J. Huddleston, T. C. Yu, M. Q. Le, N. T. Nhat, N. T. L. Thanh, A. L. Greninger, H. Y. Chu, J. A. Englund, T. Bedford, F. A. Matsen, M. F. Boni, and J. D. Bloom. Age-dependent heterogeneity in the antigenic effects of mutations to influenza hemagglutinin. Dec. 2023. <http://dx.doi.org/10.1101/2023.12.12.571235>.
- [10] G. E. Boyle, K. Sitko, J. G. Galloway, H. K. Haddox, A. H. Bianchi, A. Dixon, R. E. S. Thomson, R. K. Garge, A. E. Rettie, A. Rubin, R. C. Geck, E. M. J. Gillam, W. S. DeWitt, F. A. Matsen, and D. M. Fowler. Deep mutational scanning of cyp2c19 reveals a substrate specificity-abundance tradeoff. Oct. 2023. <http://dx.doi.org/10.1101/2023.10.06.561250>.
- [11] C. I. Stoddard, K. Sung, Z. A. Yaffe, H. Weight, G. Beaudoin-Bussi eres, J. Galloway, S. Gantt, J. Adhiambo, E. R. Begnel, E. Ojee, J. Slyker, D. Wamalwa, J. Kinuthia, A. Finzi, F. A. Matsen, D. A. Lehman, and J. Overbaugh. Elevated binding and functional antibody responses to sars-cov-2 in infants versus mothers. *Nature Communications*, 14(1), Aug. 2023. <http://dx.doi.org/10.1101/2023.10.06.561250>.
- [12] M. Fourment, C. J. Swanepoel, J. G. Galloway, X. Ji, K. Gangavarapu, M. A. Suchard, and F. A. M. IV. Automatic differentiation is no panacea for phylogenetic gradient computation. *Genome Biology and Evolution*, 15(6), jun 2023. <https://doi.org/10.1093%2Fgbe%2Fevad099>.
- [13] A. C. Willcox, K. Sung, M. E. Garrett, J. G. Galloway, J. H. Erasmus, J. K. Logue, D. W. Hawman, H. Y. Chu, K. J. Hasenkrug, D. H. Fuller, F. A. M. IV, and J. Overbaugh. Detailed analysis of antibody responses to SARS-CoV-2 vaccination and infection in macaques. *PLOS Pathogens*, 18(4):e1010155, apr 2022. <https://doi.org/10.1371%2Fjournal.ppat.1010155>.
- [14] F. Baumdicker, G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, G. Tsambos, S. Zhu, B. Eldon, E. C. Ellerman, J. G. Galloway, A. L. Gladstein, G. Gorjanc, B. Guo, B. Jeffery, W. W. Kretschmar, K. Lohse, M. Matschiner, D. Nelson, N. S. Pope, C. D. Quinto-Cort es, M. F. Rodrigues, K. Saunack, T. Sellinger, K. Thornton, H. van Kemenade, A. W. Wohns, Y. Wong, S. Gravel, A. D. Kern, J. Koskela, P. L. Ralph, and J. Kelleher. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3), dec 2021. <https://doi.org/10.1093%2Fgenetics%2Fiyab229>.
- [15] H. L. Itell, H. Weight, C. S. Fish, J. K. Logue, N. Franko, C. R. Wolf, D. J. McCulloch, J. Galloway, F. A. Matsen, H. Y. Chu, and J. Overbaugh. SARS-CoV-2 antibody binding and neutralization in dried blood spot eluates and paired plasma. *Microbiology Spectrum*, 9(2), oct 2021. <https://doi.org/10.1128%2Fspectrum.01298-21>.
- [16] J. R. Adrion, C. B. Cole, N. Dukler, J. G. Galloway, A. L. Gladstein, G. Gower, C. C. Kyriazis, A. P. Ragsdale, G. Tsambos, F. Baumdicker, J. Carlson, R. A. Cartwright, A. Durvasula, I. Gronau, B. Y. Kim, P. McKenzie, P. W. Messer, E. Noskova, D. O.-D. Vecchyo, F. Racimo, T. J. Struck, S. Gravel, R. N. Gutenkunst, K. E. Lohmueller, P. L. Ralph, D. R. Schrider, A. Siepel, J. Kelleher, and A. D. Kern. A community-maintained standard library of population genetic models. *eLife*, 9, jun 2020. <https://doi.org/10.7554%2Felife.54967>.

PROFESSIONAL REFERENCES

- **Dr. Erick Matsen:** Principal Investigator - matsen@fredhutch.org
- **Dr. Leslie Coonrod:** Associate Director of Master's Program - coonrod@uoregon.edu
- **Dr. Stacey Wagner:** Director of Master's Program - sdwagner@uoregon.edu
- **Dr. Bill Cresko:** Research Advisor - wcresko@uoregon.edu
- **Dr. Peter Ralph:** Principal Investigator - plr@uoregon.edu
- **Dr. Andrew Kern:** Principal Investigator - adkern@uoregon.edu
- **Dr. Ben Haller:** Research Advisor - haller@mac.com

ACHIEVEMENTS

- **Graduate Thesis Project:** Machine learning for identifying synapses.
- **Undergraduate Honors Thesis Project:** Speeding up the Tortoise: a case study in the optimization of forward-moving evolutionary simulations. The manuscript can be found on the [University Arxiv](#)
- **Game Development:** Developed a 3D infinite flyer game, [Squirrel suiter](#) for Windows and Android, available on Google Play.
- **Community Service:** Built trail for [Montana Conservation Corps](#) and acquired 300+ hours community service.