ARIELLA GLADSTEIN, PhD

algladstein@gmail.com

https://github.com/agladstein ⋄ www.linkedin.com/in/ariella-gladstein

Accomplished computational biologist with deep expertise in population genetics, machine learning, and large-scale genomic data analysis. Proven ability to design, develop, and implement machine learning algorithms to analyze genomic datasets, collaborate with multidisciplinary teams, and communicate findings effectively.

Expertise: *population genetics, computational genomics, machine learning, cloud computing, statistics*

EDUCATION

University of Arizona, Tucson, AZ　　　　　　　　　　　　　　　　　*August 2018*
PhD in Ecology and Evolutionary Biology, Minor in Mathematics

Beloit College, Beloit, WI　　　　　　　　　　　　　　　　　　　　*May 2011*
B.S. in Mathematical Biology & Russian, Cum Laude
Departmental Honors: Mathematical Biology

EXPERIENCE

Research Scientist, Computational Biology　　　　　　　　　　　　　2021 - 2024
Ancestry team　　　　　　　　　　　　　　　　　　　*Embark Veterinary, Remote*
Developed dog breed ancestry and relative-matching products, enhancing computational efficiency while maintaining rigorous quality standards. Played a key role in improving pipeline automation from 40% to over 95% and scaling processes from hundreds of thousands to millions of samples. Collaborated with team members and engineers to improve local ancestry inference, implementing algorithm improvements for uncertainty estimates. Developed a machine learning approach for village dog classification, enhancing accuracy and scalability. Conducted literature reviews, tested, and reported recommendations on published methods for population clustering and relative inference, including advanced machine learning methods. Collaborated with a cross-functional team on a science initiative projected to save $7.5 million annually. Served on the Diversity, Equity, Inclusion, and Belonging (DEIB) committee, co-authoring an Inclusive Norms document, designing and analyzing company-wide DEIB surveys, and contributing to various DEIB initiatives.

Postdoctoral Fellowship　　　　　　　　　　　　　　　　　　　　　2018 - 2020
Schrider Lab　　　　　　　*Department of Genetics, University of North Carolina, Chapel Hill*
Developed deep learning models for population genetics inference. Contributed to writing documentation and tutorials for the population genomics tskit software ecosystem. Actively participated in the PopSim Consortium, contributing to the development of the stdpopsim library.

Dissertation Research　　　　　　　　　　　　　　　　　　　　　　2011 - 2018
Hammer Lab　　　　　　　　　*Ecology and Evolutionary Biology, University of Arizona*
Developed and applied computational methods for population genetics analysis, addressing ascertainment bias in identifying runs of homozygosity and demographic inference. Created and released software for high-throughput genomic simulations. Applied Approximate Bayesian Computation (ABC) to model the demographic history of Ashkenazi Jews.

SKILLS

Programming	Python, R, Bash, unit testing, profiling, multiprocessing, Git
Machine Learning	Scikit-learn, Keras, TensorFlow
Data Science	Numpy, Pandas, SciPy, Matplotlib, Seaborn, Plotly
Cloud Computing	Linux/Unix, AWS, SageMaker, JupyterHub, Kubernetes
Genomics	PLINK, Bedtools, VCFtools, phasing, imputation, haplotype detection, population structure, genome simulation
Reproducibility	Workflow development (Pegasus, Makeflow, Snakemake), Docker, Jupyter, Knitr
Documentation	LaTeX, Markdown, reStructuredText, JupyterBook, Confluence

TEACHING EXPERIENCE

Introduction to using stdpopsim Workshop   2020, *Virtual, PopSim Consortium*
Designed and presented interactive workshop on using stdpopsim, the standard library for reproducible, bug-free simulations of genetic datasets from published demographic histories.

NSF Cyber Carpentry Workshop   2019, *University of North Carolina, Chapel Hill*
Assisted with 2-week workshop, where participants learned best practices for data-intensive computing, cloud infrastructure, and deep learning through hands-on projects.

CyVerse Container Camp   2018, *University of Arizona*
Assisted with 3-day hands-on workshop on using containerized workflows for more reproducible science.

Software Carpentry Workshop   2017, *University of Arizona*
Assisted with Software Carpentry hands-on workshop on basic Unix/Bash, Python, and Git

ECOL 320 Genetics   2014 - 2016, *University of Arizona*
Graduate Teaching Assistant

SELECT PUBLICATIONS, POSTERS, & TALKS

Gladstein A.L.[*], et al. A Novel Machine Learning Method for Classification of Village Dogs. Probabilistic Modeling in Genomics. 2023. Cold Spring Harbor, NY. (Poster)

Baumdicker, F.[*], Bisschop, G.[*], Goldstein, D.[*], Gower, G.[*], Ragsdale, A. P.[*], Tsambos, G.[*], Zhu, S.[*], …, **Gladstein, A. L.**, …, Kelleher, J. 2022. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics.* Volume 220, Issue 3. `https://doi.org/10.1093/genetics/iyab229`

Adrion, J. R.[*], Cole, C. B.[*], Dukler, N.[*], Galloway, J. G.[*], **Gladstein, A. L.**[†], Gower, G.[*], Kyriazis, C.C.[*], Ragsdale, A.P.[*], Tsambos, G.[*], …, Gravel, S., Gutenkunst, R.N., Lohmeuller, K.E., Ralph, P.L., Schrider, D.R., Siepel, A., Kelleher, J., Kern, A.D. 2020. A community-maintained standard library of population genetic models. *eLife.* 9:e54967. doi: `https://doi.org/10.1101/2019.12.20.885129`

Gladstein A.L.[†], Schrider R.D. Demographic model selection with deep learning. Probabilistic Modeling in Genomics. 2019. Aussois, France. (Poster)

Gladstein A.L.[*] and Hammer M.F. 2019. Substructured population growth in the Ashkenazi Jews inferred with Approximate Bayesian Computation. *Molecular Biology and Evolution.* 36(6): 1162-1171. doi: `https://dx.doi.org/10.1093/molbev/msz047`

Gladstein A.L.[*] et al. 2018. SimPrily: A Python framework to simplify high-throughput genomic simulations. *SoftwareX*, 7, 335-340. `https://doi.org/10.1016/j.softx.2018.09.003`

Gladstein A.L.[†], et al. The effect of SNP array ascertainment bias on the distribution of runs of homozygosity lengths. Annual Meeting of the American Society for Human Genetics. 2015. Baltimore, MD. (Poster)

AWARDS

NSF XSEDE compute allocations (1e6 CPU hrs, 6,500 GPU hrs, 5.5 Tb storage)	*2017 - 2020*
NIH Computational and Mathematical Modeling of Biological Systems Traineeship ($71,064)	*2013-14*
NSF Integrative Graduate Education and Research Traineeship in Genomics ($97,083)	*2011-13*

PROJECT MANAGEMENT

Managed team of 5, including computer scientists, software engineer, and mathematician, working on code development and high throughput computing for bioinformatics	*2017*

OTHER ACTIVITIES

Circus Arts   2012 - present
Aerial silks, aerial rope, lyra, flying trapeze, tightwire, handbalancing, contortion

Figure Skating   1998 - 2013, 2021 - present
Singles freestyle

---

[*]Presenter
[†]First author