

# BRIAN J ARNOLD, PHD

Princeton NJ

brianjohnarnold@gmail.com | [Web site](#) | [GitHub](#) | [Google Scholar](#)

---

## PERSONAL SUMMARY

Highly adaptive computational biologist with 9+ years of professional data-science experience, specializing in software (Python), machine learning, and detailed analysis of diverse data modalities. Successfully led projects from conception to completion and significantly improved analytical capabilities of research teams by managing large datasets with scalable software. Learns fast, works hard, and navigates uncertainty well.

## SKILLS

**Software:** proficient in Python, R, Bash; familiar with C/C++; comfortable reading code in any language

**Machine learning:** proficient in scipy, scikit-learn, scikit-image, PyTorch, Keras, numpy, Statsmodels

**Bioinformatics:** adept at genomics (short-/long-read), spatial/single-cell transcriptomics, images/videos analyses

**Large datasets:** adept at parallelizing tasks via Snakemake workflows, Python multiprocessing libraries

**Computing:** adept at high-performance computing (SLURM), experience with cloud computing (GCP, AWS)

**Data visualization:** proficient with Matplotlib and Seaborn; experience with Plotly for interactive 3D figures

**Communication:** authored 25 scientific publications, 4 successful grant/fellowship applications, 9 workshops

**Teaching and mentoring:** taught 11 highly acclaimed workshops, designed 7 of these from scratch; patient with beginners

## EXPERIENCE

### Departmental Data Scientist

2023-present

Princeton University, Ecology and Evolutionary Biology

- Segmented 3D 2-photon microscopy videos of brains and directed statistical analyses to decode animal behavior
- Led machine-/deep-learning analyses of 40+ enhancer-readout datasets to characterize cis-regulatory motifs
- Managed 2 trainees on large scale sequencing projects involving 700+ samples and 10+ TB of data
- Gave 2 workshops on machine learning and Snakemake workflows for computational biology

### Biomedical Data Scientist

2020-2023

Princeton University, Computer Science

- Packaged new software to run at scale and optimize neural networks that segment tissues in spatial transcriptomics data
- Authored two modules of new software to detect copy-number variants in bulk tumor sequencing data
- Managed Conda releases and developed software that characterizes evolutionary history of somatic mutations
- Authored Science paper (co-first) on the evolutionary genetics of tusklessness in African elephants
- Taught 4 workshops on Python Packaging, data visualization, machine learning, and R

### Senior Bioinformatics Scientist

2018-2020

Harvard University

- Created and published Snakemake workflow to massively parallelize variant calling
- Designed machine learning classifier to quantify scale of horizontal gene transfer in world's largest flower species
- Authored highly cited review (cover article) in Nature Reviews Microbiology on horizontal gene transfer in bacteria
- Taught 2 workshops on R and genomics

### Consultant

2018

Day Zero Diagnostics

- Disentangled unexpected transmission dynamics within Massachusetts General Hospital using bacterial genomes

### Postdoctoral Fellow

2015-2018

Harvard T.H. Chan School of Public Health

- Inferred recombination parameters from bacterial genomes via Bayesian optimization with Gaussian Process regression
- Designed fast, memory-efficient population genetic simulators in C++ to test novel analyses of bacterial genomes
- Mentored 1 Master's student, 2 undergraduates

### PhD Candidate

2009-2015

Harvard University

- Bomblyes lab: generated a variety of genomic data to study the evolutionary origins of tetraploid *Arabidopsis arenosa*
- Kleckner lab: used immunofluorescence cytology to explore meiotic cell cycle delays in tetraploid yeast

- Wakeley group: developed new population genetic theory for tetraploids using probability theory and Markov chains

## Fulbright Fellow

2009

University of Oulu – Finland

- Sequenced candidate genes to identify mutations associated with hairlessness in the plant *Arabidopsis lyrata*

## EDUCATION

### Harvard University

Cambridge, MA

PhD, Organismic and Evolutionary Biology

2015

*Dissertation: "Evolutionary dynamics of a multiple-ploidy system in Arabidopsis arenosa"*

Primary advisor: Kirsten Bomblies

Secondary advisor: Nancy Kleckner

Committee members: John Wakeley, David Reich, Hopi Hoekstra

### University of Minnesota – Twin Cities

St. Paul, MN

BS (Honors and High Distinction), Plant Biology

2008

## AWARDS AND HONORS

- NIH F32 Postdoctoral Fellowship 2016-2018
- NSF Doctoral Dissertation Improvement Grant 2012
- Two Certificates of Distinction in Teaching (Harvard University) 2010, 2012
- NSF Graduate Research Fellowship 2010-2015
- Herchel Smith Graduate Fellowship (Harvard University) 2009-2011
- James Mills Peirce Fellowship (Harvard University) 2009
- Fulbright Full Grant (Oulu, Finland) 2008-2009
- Merck Index Award: best student in organic chemistry (University of Minnesota – Twin Cities) 2007
- Prentice Hall Book Prize: highest grade in organic chemistry II (University of Minnesota – Twin Cities) 2006

## SELECTED PUBLICATIONS

**\* authors contributed equally to this work**

Chitra U\*, **Arnold B\***, B Raphael (2024). "Quantifying higher-order epistasis: beware the chimera" (*in review*).

Chitra U, **Arnold B**, Sarkar H, Ma C, Lopez-Darwin S, Sanno K, Raphael B (2024). Mapping the topography of spatial gene expression with interpretable deep learning. *RECOMB* 2024.

Myers M, **Arnold B**, Bansal V, Balaban M, Mullen K, Zaccaria S, B Raphael (2024). HATCHet2: clone-and-haplotype-specific copy number inference from bulk tumor sequencing data. *Genome Biology* 25 (1):1-28.

Mirchandani C, Shultz A, Thomas G, Smith S, Baylis M, **Arnold B**, Corbett-Detig R, Enbody E, and T Sackton (2024). A fast, reproducible, high-throughput variant calling workflow for population genomics. *Molecular Biology and Evolution* 41 (1) msad270.

**Arnold B**, Huang IT, and WP Hanage (2022). Evolutionary genomics and horizontal gene transfer in bacteria. *Nature Reviews Microbiology* <https://doi.org/10.1038/s41579-021-00650-4>.

**† highlighted as the cover article for April issue of Nature Reviews Microbiology**

Campbell-Staton S\*, **Arnold B\***, Golcalves D, Poole J, Long R, and R Pringle (2021). Ivory poaching and the rapid evolution of tusklessness in African elephants. *Science* 374, 483-487.

**† featured in over 40 media interviews including New York Times, Science magazine, and Nature magazine**

Cai L, **Arnold B**, Xi Z, Khost D, Patel N, Hartmann C, ... C Davis (2021). Deeply Altered Genome Architecture in the Endoparasitic Flowering Plant *Sapria himalayana* Griff. (Rafflesiaceae). *Current Biology*, 31(5), 1002-1011.e9.

**Arnold B**, Sohail M, Wadsworth C, Corander J, Hanage WP, Sunyaev S, and Y Grad (2019). Fine-scale haplotype structure reveals strong signatures of positive selection in a recombining bacterial pathogen. *Molecular Biology and Evolution* 37(2):417-428.

Pensar J, Puranen S, **Arnold B**, MacAlasdair N, Kuronen J, Tonkin-Hill G, Pesonen M, Xu Y, Sipola A, Sánchez-Busó L, Lees J, Chewapreecha C, Bentley S, Harris S, Parkhill J, Croucher N, and J Corander (2019). Genome-wide epistasis and co-selection study using mutual information. *Nucleic Acids Research* 47(18):e112-e112.

**Arnold B**, Gutmann M, Grad Y, Sheppard S, Corander J, Lipsitch M, and WP Hanage (2018). Weak epistasis may drive adaptation in recombining bacteria. *Genetics* 208(3):1247-1260.

- Arnold B**, Lahner B, DaCosta J, Weisman C, Hollister J, Salt D, Bomblies K, and L Yant (2016). Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences of the USA*, 113(29): 8320-8325.
- Arnold B**, Kim S, and K Bomblies. (2015). Single origin of autotetraploid *Arabidopsis arenosa* followed by interploidy admixture. *Molecular Biology and Evolution*, 32(6):1382-1395.
- Arnold B**, Corbett-Detig R, Hartl D, and K Bomblies (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, Vol 22: 3179-3190.
- Arnold B**, Bomblies K, and J Wakeley (2012). Extending coalescent theory to autotetraploids. *Genetics* 192(1):195-204.

## TEACHING EXPERIENCE

### Princeton University workshops

- Introduction to computational biology workflows ([GitHub](#)) 2024
  - Designed and taught three 2-hour sessions
  - 26 attendees
- Introduction to machine learning for Ecology and Evolutionary Biology ([GitHub](#)) 2023
  - Designed and taught two 1-hour sessions
  - 45 attendees
- Introduction to machine learning ([GitHub](#)) 2023
  - Designed and taught one of five 1-hour sessions
  - 90+ attendees
- Data visualization python ([GitHub](#)) 2022
  - Designed and taught one 2-hour session
  - 20+ attendees
- Level up your python 2022
  - Assisted instructor for one 2-hour session
  - 20+ attendees
- Best practices in python packaging 2021
  - Designed and co-taught one 3-hour session
  - 10+ attendees
- Introduction to data analysis with R 2021
  - Designed and taught one 2-hour session
  - 10+ attendees

### Harvard University workshops

- Introduction to R (with tidyverse) 2019-2020
  - Taught three 2-hour sessions
  - 15+ attendees
- Read mapping and variant calling 2019-2020
  - Taught one 3-hour session
  - 5+ attendees

### Harvard University classes

- Genetics and Genomics 2012
  - Taught weekly recitations, graded homework and exams; received teaching award
- Coalescent Theory 2010
  - Taught weekly recitations, graded homework and exams; received teaching award

### University of Minnesota classes

- Biochemistry 2008
  - Taught weekly recitations, graded exams
- Organic Chemistry 2007
  - Supervised lab experiments, graded exams

- General Botany
  - Supervised lab experiments, graded exams

2007

### **SELECTED PRESENTATIONS**

Contributed Talk – “New analyses for copy number, tumor evolution, and spatial transcriptomics” – 2022 National Cancer Institute’s Genomic Data Analysis Network

Invited Talk – “Characterizing copy number aberrations and intratumor heterogeneity with machine learning” – 2022 University of Edinburgh

Invited Talk – “Fine-scale haplotype structure reveals strong signatures of positive selection in a recombining bacterial pathogen” – 2019 University of Nottingham Departmental Seminar

Invited Talk – “Genomic landscape of linked selection in *N. gonorrhoeae*” – 2018 Society for Molecular Biology and Evolution Satellite Workshop on “Genome Evolution in Pathogen Transmission and Disease”

Contributed Talk – “Weak epistasis may drive adaptation in recombining bacteria” – 2017 Society for Molecular Biology and Evolution

Invited Talk – “Evolutionary dynamics of a multiple ploidy system in *A. arenosa*” – 2015 University of Oslo Departmental Seminar Series