

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 machine learning and software (Python and R). Successfully led projects from conception to completion and significantly improved analytical capabilities of research teams by managing large datasets with scalable solutions. Learns fast, works hard, and navigates uncertainty well.

Seeking a dynamic role in a motivated team of biologists and data scientists to use my technical skills for innovative research that uses data of various modalities. Interested in positions that offer both collaborative and leadership opportunities.

SKILLS

Programming: 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 (400+ samples) whole-genome sequencing projects using short- and long-read 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 optimize neural networks that segment tissues in spatial transcriptomics data and run at scale
- 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
- Bomblies 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

SELECTED 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

PEER-REVIEWED 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.
- Kinnaman M, Zaccaria S, Makohon-Moore A, **Arnold B**, Levine M, Gudem G, Arango Ossa J, Glodzik D, Rodríguez-Sánchez M, Bouvier N, Li S, Stockfisch E, Dunigan M, Cobbs C, Bhanot U, You D, Mullen K, Melchor J, Ortiz M, O'Donohue T, Slotkin E, Wexler L, Dela Cruz F, Hameed M, Glade Bender J, Tap W, Meyers P, Papaemmanuil E, Kung A, and C Iacobuzio-Donahue (2024). Subclonal somatic copy-number alterations emerge and dominate in recurrent osteosarcoma. *Cancer Research* 83 (22) 3796-3812.
- Xiao A, Wu F, Bushman M, Zhang J, Imakaev M, Chai PR, Duvallet C, Endo N, Erickson TB, Armas F, **Arnold B**, Chen H, Chandra F, Ghaeli N, Gu X, Hanage WP, Lee WL, Matus M, McElroy KA, Moniz K, Rhode SF, Thompson J, and EJ Alm (2022). Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Research*, 212 (2022) 118070.
- Arnold B**, Huang IT, and WP Hanage (2021). 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.
- Azarian T, Martinez P, **Arnold B**, Qiu, X., Grant, L. R., Corander, J, ... WP Hanage (2020). Frequency-dependent selection can forecast evolution in *Streptococcus pneumoniae*. *PLoS Biology*, 18(10), e3000878.
- 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.
- Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G, Bohutínská M, **Arnold B**, Weisman C, Marhold K, Slotte T, Bomblies K, and L Yant (2019). Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nature Ecology and Evolution* 3(3):457-468
- Shapiro L, Paulson J, **Arnold B**, Scully E, Zhaxybayeva O, Pierce N, Rocha J, Klepac-Ceraj V, Holton K, and R Kolter (2018). An introduced crop plant is driving diversification of the virulent plant pathogen *Erwinia tracheiphila*. *mBio* 9:e01307-18.
- Wadsworth C, **Arnold B**, Sater M, and Y Grad (2018). Azithromycin resistance through interspecific acquisition of an epistasis dependent efflux pump component and transcriptional regulator in *Neisseria gonorrhoeae*. *mBio* 9:e01419-18
- Azarian T, Grant L, **Arnold B**, Hammit L, Reid R, Santosham M, Weatherholtz R, Goklish N, Thompson C, Bentley S, O'Brien K, Hanage WP, and M Lipsitch (2018). The impact of serotype-specific vaccination on phylodynamic parameters of *Streptococcus pneumoniae* and the pneumococcal pan-genome. *PLoS Pathogens*: <https://doi.org/10.1371/journal.ppat.1006966>
- 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.
- Corander J, Fraser C, Gutmann M, **Arnold B**, Hanage WP, Bentley S, Lipsitch M, and N Croucher (2017). Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nature ecology & evolution* 1:1950-1960.
- 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.
- Baduel P, **Arnold B**, Weisman C, Hunter B, and K Bomblies (2016). Habitat-associated life history and stress-tolerance variation in *Arabidopsis arenosa*. *Plant physiology*, 171(1): 437-451.
- 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.
- Wright K, **Arnold B**, Xue K, Šurinová M, O'Connell J, and K Bomblies, (2014). Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. *Molecular Biology and Evolution*, 32(4): 944-955.
- 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.
- Yant L, Hollister JD, Wright K, **Arnold B**, Higgins JD, Franklin F, and K Bomblies, (2013). Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current Biology*. 23, pp. 2151-2156.
- Hollister J, **Arnold B**, Svedin E, Xue K, Dilkes B, and K Bomblies, (2012). Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genetics*, 8(12): e1003093
- Arnold B**, Bomblies K, and J Wakeley (2012). Extending coalescent theory to autotetraploids. *Genetics* 192(1):195-204.

LARGE-TEAM EXPERIENCE

Biomedical Data Scientist

Princeton University, Computer Science

2020-2023

- Contributed copy-number variant calls to National Cancer Institute's Genomic Data Analysis Network
 - Multi-Institutional Italian Lung Diagnostic 2022-2023
 - Diffuse Large B-Cell Lymphoma group 2021-2023
 - Testicular Germ Cell Cancer group 2021-2023

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 2007
 - Supervised lab experiments, graded exams

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

Contributed Talk – “Is pneumococcal evolution sensitive to fitness epistasis?” – 2016 International Symposium on Pneumococci and Pneumococcal Diseases

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

Contributed Talk – “The dynamics of autotetraploid formation in *A. arenosa*” – 2014 Evolution Conference

Poster Presentation – “RADseq reveals biogeography of an autotetraploid” – 2013 Gordon Research Conference on Ecological and Evolutionary Genomics

Poster Presentation – “How do autotetraploids become successful species?” – 2012 Gordon Research Conference on Meiosis

Poster Presentation – “Extending coalescent theory to autotetraploids” – 2011 American Genetics Association Symposium, Guanajuato/Irapuato, Mexico

Contributed Talk – “Evolutionary genetics of autotetraploids” – 2010 Boston Evolutionary Genetics Supergroup Meeting, Cambridge MA