

# MATTHEW GIBSON

genetics <> bioinformatics <> cloud engineering

@ matthew.j.s.gibson@gmail.com    📍 USA    🔗 <https://home.gibsonmatthew.com>  
📄 <https://github.com/gibsonMatt>

## SUMMARY

---

I am a geneticist by training, a data scientist, and an engineer. I hold a PhD in evolutionary genetics and bioinformatics, and have worked as both a senior data scientist developing cloud-native genetics pipelines for early-stage startups as well as a bioinformatics engineer in the pharmaceutical industry building infrastructure for drug discovery. With a publication portfolio of 12 peer-reviewed articles, I have also contributed to advancing knowledge in population genomics, phylogenomics, evolutionary genetics, and low-pass sequencing.

## EXPERIENCE

---

### Bioinformatics Engineer

📅 June 2024 – Current

Eli Lilly (Lilly Research Labs)

- built and maintained novel AWS infrastructure (terraform) for reproducible research
- implemented distributed usage tracking for a multi-tenant cloud platform, reducing cloud spend through custom observability
- wrote and maintained R software packages to support streamlined access to research data from R clients

### Data Scientist/Senior Data Scientist

📅 May 2022 – April 2024

Gencove, Inc.

- wrote and maintained production code for public-facing statistical genetic pipelines (low-pass imputation, PRS, WGS)
- drove adoption of low-pass sequencing at consumer genetic testing company through simulation of biobank-scale GWAS
- served as technical lead for new WGS analysis product, lowering execution costs through optimized workflows with Sentieon

## RELEVANT TECHNICAL SKILLS AND TOOLS

---

- **Languages:** Python, R (Shiny + tidyverse), bash, Terraform, Javascript/Typescript, Nextflow, SQL
- **Bioinformatics:** samtools, bcftools, GATK/Sentieon, ANGSD, snakemake, PLINK, STRUCTURE, treemix, salmon, Seqera platform
- **Cloud:** AWS (IAM, S3, Glue, Athena, Lambda, WAF, EC2, Step Functions, Parallel Cluster, Datasync, EKS), k8s
- **Computing:** GitHub CI/CD, docker, singularity, SLURM, git, Unix, Flask, Django, Apache Spark

## SELECTED PUBLICATIONS

---

- Stewart, C.M., **Gibson, M.J.S.**, Parsa, J., & Li, J.H. (2024). A comparison between low-cost library preparation kits for low coverage sequencing. *BioRxiv*. <https://doi.org/10.1101/2024.01.30.578044>
- **Gibson, M.J.S.**, Torres, M.L., Brandvain, Y., & Moyle, L.C. (2021). Reconstructing the history and biological consequences of a plant invasion on the Galapagos islands. *eLife*. <https://doi.org/10.7554/eLife.64165>
- **Gibson, M.J.S.** & Moyle, L.C. (2020). Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Molecular Ecology*. <https://doi.org/10.1111/mec.15477>

*Full list of publications available on my Google Scholar*

## EDUCATION

---

### Ph.D. Evolution and bioinformatics Indiana University

📅 Aug 2016 – May 2022

-included formal coursework in machine learning  
-obtained \$120k in USDA funding (NIFA)

### B.S. Genetics

#### University of Kansas

📅 2011 – 2016

## PROJECTS

---

### Omics Hub Databricks Platform

- led strategic platform evaluation and deployment of enterprise data warehouse for omics analytics, accelerating adoption through cross-functional collaboration. Developed framework enabling self-service usage while mentoring junior engineers

### Statistical tool for detecting hemiplasy

- wrote statistical software (HelST) for predicting the risk of false inference in phylogenetic data sets. This tool is hosted on GitHub and published in eLife.

### Bovine InfiniSEEK

- developed novel algorithms for "InfiniSEEK" product which combines low-pass sequencing and targeted high coverage genotyping

### Statistical methods for detecting introgression

- led an international project using genome sequencing, HMM local ancestry inference, and admixture mapping to dissect the basis of fruit color variation in Galápagos tomato. This work is published in eLife.