

MATTHEW GIBSON

data science <> software engineering

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

SUMMARY

I am a data scientist with software development know-how. I hold a PhD in genetics and bioinformatics, and currently work as a Senior Data Scientist developing production-quality code for biotechnology applications. With a publication portfolio of 10 research articles in journals like eLife and Molecular Ecology, I have contributed to advancing knowledge in population genetics, bioinformatics, and statistics. My technical skills extend across Python, R, and Bash, supplemented by experience in languages like JavaScript (Next.js). As I envision my next professional chapter, I am excited to develop a more profound engineering skill set through enriching experiences.

EXPERIENCE

Senior Data Scientist

📅 Jan 2023 – Current

Gencove, Inc.

- wrote production code for new public-facing computational tooling using common CI/CD infrastructures
- retained customers by rapidly resolving their issues through a combination of deep genetic knowledge and an ability to quickly deploy hotfix solutions
- designed and led R&D projects related to applications of low pass sequencing at large scales
- contributed core features leveraging REST and GraphQL APIs in Django as well as front end web development in Javascript

Data Scientist

📅 May 2022 – Jan 2023

Gencove, Inc.

- deployed informatic and statistical solutions to high-throughput genomics problems in agriculture and human health
- developed and supported a deep sequencing (WGS) variant calling pipeline

TECHNICAL SKILLS

- **Languages:** Python (advanced), R (advanced), bash (advanced), C# (beginner), Javascript (beginner)
- **Bioinformatics:** RADseq assembly, samtools, bcftools, variant calling with GATK, low-coverage genotyping in ANGSD, short read mapping/alignment, RNAseq analysis, Sentieon
- **Computing:** AWS infrastructure, GitLab CI/CD, docker, docker compose, SLURM, git, Unix tooling, Flask, Django
- **Statistics:** linear modeling/regression, mixed-effects models, GWAS, hidden Markov modeling, demographic modeling with maximum likelihood.

SELECTED HOBBY PROJECTS

- **bioneer:** a Python LLM tool for bioinformaticians. I developed a command line tool that implements dynamic few-shot prompt injection using langchain and ChromaDB to fine tune a large language model at run-time and provide high quality responses. The tool is made with best-practices and production scaling in mind, and is hosted publicly on my GitHub
- **Ludum Dare 47 Game Jam:** planned and designed a video game from scratch as part of a game jam. The theme of the jam was "stuck in a loop". A colleague and I built the game "Escape Velocity" using Unity C#, which you can play here!
- **R shiny apps.** I've made R shiny interfaces for visualizing a multivariate model of population evolution as well as for creating publication-ready maps.

EDUCATION

Ph.D. Evolution and bioinformatics

Indiana University-Bloomington

📅 Aug 2016 – May 2022

B.S. Genetics, with honors

University of Kansas

📅 2011 – 2016

PROJECTS

Computational 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

- co-created the innovative product "Infiniseek", which combines low-pass sequencing and target capture into a single assay. Infiniseek provides high quality variant calls in regions critical to breeders.

Statistical methods for detecting introgression in Galápagos tomato

- led a multi-year, international project using reduced representation genome sequencing, HMM local ancestry inference, and admixture mapping to dissect the genetic basis of fruit color variation in Galápagos tomato. This was the flagship project of my PhD.

Landscape genomics in wild tomato

- used high-fidelity genomic and climate data from >200 wild plant samples to understand the climatic drivers of adaptation and identify novel functional genes.