Jared Croyle

jcroyle@berkeley.edu | Berkeley, CA | linkedin.com/in/jared-croyle | jaredcroyle.com

Computational Biologist | Data Engineer + Scientist | Full- Stack ML Software Developer | UC Berkeley

Education

University of California, Berkeley, *B.S. in* Molecular Environmental Biology & Minor in Data Science   May 2025
Relevant Coursework: Machine Learning, Data Analytics, Computational Structures, Practical Genomics, General Genetics

Technical Skills

Languages & Tools: Python, SQL, R, BASH, HTML5/CSS3, Git, GitHub, VSCode, Excel, Linux/MacOS
Data Engineering: Apache Spark, Delta Lake, Databricks, Azure, ETL Pipelines, Jupyter Notebook
ML & Data Analysis: Scikit-learn, Pandas, NumPy, PyTorch, Matplotlib, Seaborn, OpenCV, SHAP, Feature Engineering
Bioinformatics and Infra: AMPtk, DADA2, MEGA11, BEAST2, SnapGene, SRA Toolkit, MLFlow, Docker (basic)

Experience

Genomic Data Engineering Intern, Evolab at UC Berkeley – Berkeley, CA   February 2024 – May 2025
  • Built scalable Python/BASH pipelines to ingest, clean, and normalize 10K+ genomic records on HPC cluster; replicated workflows in Databricks using Spark and Delta Lake.
  • Automated ETL workflows using open-source tools and custom scripts; improved data throughput by 30%.
  • Developed modular validation modules to evaluate predictive models; visualized results using Matplotlib and SHAP for scientific stakeholders.

Genomic Data Intern, Evolutionary Genomics Lab at UC Berkeley – Berkeley, CA   September 2024 – December 2024
  • Designed structured metadata systems using Excel and SQL for 500+ genomic samples to ensure data integrity, reproducibility, and scalability in downstream analytics pipelines.
  • Streamlined QC and cataloging pipelines, supporting long-term genomic research access.

Bioinformatics Fellow (NSF-REU), Stajich Lab at UC Riverside – Riverside, CA   June 2023 – August 2023
  • Engineered reproducible NGS workflows (AMPtk, DADA2) to process 7M+ metabarcode reads, resulting in reduced runtime by 40% via batch processing strategies.
  • Developed Python-based QC and clustering scripts to identify 10K+ ASVs, which contributed to 2 peer-reviewed publications and novel fungal species discovery.

Featured Project

Trunkline: Full-Stack ML Pipeline for Predictive Synthetic Biology – github.com/jaredcroyle/Trunkline-SynBioML
• Architected a full-stack ML pipeline with modular YAML-based config, CLI-based execution, and reproducible runs for synthetic bio data.
• Automated data ingestion, preprocessing, model training, and ensemble prediction using Pandas, Scikit-learn, and SHAP.
• Generated interactive HTML reports with Jinja2 for iteration and explainability (SHAP plots, feature importance, residuals).
• Used for synthetic data benchmarking in biology and gene network prediction; designed for extension by research teams.

Leadership & Service

General Member, Epsilon Eta, University of California, Berkeley   January 2024 – May 2025
  • Mentored undergraduate students in biology and environmental data science via Epsilon Eta.
Science Education Ambassador, The Aquarium of the Pacific – Long Beach, CA   May 2018 – August 2019
  • Supported international biodiversity monitoring and translated research into educational content for 10K+ visitors.

Awards and Publications

  • Kelly, K., Liu, X., **Croyle, J**., and Stajich, J.E. (2025). *Climate impact on the biocrust mycobiome.*
  • **Croyle, J**., Kelly, K., Liu, X., and Stajich, J. E.. *Draft genome sequence of endophytic fungus Coniochaeta sp.* (In Prep)
  • Nona & Innovation Award – Most impactful and innovative bioinformatics software, Nonaworks Hackathon (2024)