

JASON A. CORWIN, PH.D., M.S.

jacorwin201@gmail.com | Lyons, CO | (607)-744-2082

linkedin.com/in/jason-a-corwin-196331168/

SENIOR PRINCIPAL DATA SCIENTIST

Full-stack data scientist and computational biologist with demonstrated experience supporting wet-lab molecular biology. Experienced with a wide array of relevant statistical and machine learning models as well as computational platforms (listed below). Adept at designing, building, and administering GPU-enabled local, distributed, and cloud server architectures for research teams. Authored 46 scientific publications and presented at national and international conferences. Core competencies include:

Experimental Design | Frequentist & Bayesian Statistical Analysis |
Supervised & Unsupervised Machine Learning | AWS Solutions Architect | Project & Team Management

EXPERIENCE

THINK BIOSCIENCE, Boulder, CO

September 2023 - August 2024

Sr Pr Data Scientist and Bioinformaticist - Dr. Jerome Fox and Dr. Joel Kralj

Responsible for all advanced analytics, production coding, and DevOps for the Platform Team

- Developed and maintained an ML classifier, pipeline, and PostgreSQL database for automatically identifying and classifying terpenoid compounds from GCMS traces
- Developed a statistical methodology for identifying compound hits in a complex microbial community through error-prone, long-read metagenomic sequencing.
- Developed a pipeline for synthetic peptide development for use in bacterial two-hybrid (B2H) assays
- **AWS Systems Architect:** Handled all coding development, production, and DevOps for informatic pipelines and databases within the AWS Architecture for the Platform Team.
- **LIMS:** developed and managed the server-side the L7 ESP LIMs system

TWIST BIOSCIENCE, South San Francisco, CA

August 2021 - July 2023

Sr Pr R&D Bioinformatics Scientist - Dr. Jean Challacombe

Developed a bioinformatic pipeline for design of RNA-fusion Pan-cancer controls containing >1,000 fusions. Designed library prep optimization.

- Custom design of structural variant (SV) controls cancer detection in cfDNA.
- Automated sequencing off-target and chimera graph visualizations within established pipelines.
- Trained new wet-lab member to run established bioinformatic pipelines.

COLORADO STATE UNIVERSITY, Los Alamos National Laboratory, Fort Collins, CO **October 2018 - August 2021**

CO Post-doctoral Researcher - Dr. Pankaj Trivedi and Dr. Jean Challacombe

Developed and taught statistical pipelines for metagenomic and plant microbiome profiling.

- Modeled epidemiology and geospatial risk of leaf rot disease for sugar beet growers in CO.
- Mentored and managed four graduate students (two MS and two PhDs) through their data collection, bioinformatic and statistical analysis, and thesis writing.
- Designed, built, and administered laboratory server for approximately 10 users.

UNIVERSITY OF COLORADO, Boulder, CO

August 2016 - October 2018

Post-doctoral Researcher - Dr. Noah Fierer

Used multivariate statistics and interaction network analyses on population genomic data to define the genetic architecture shaping plant microbiome interactions in soil.

- Helped develop statistical models and data visualizations to analyze genome data using negative binomial generalized linear mixed models and network-oriented models.
- Managed a team of 5 people to conduct large experiments of 5,000 multi-omic samples in a randomized complete block design.

UNIVERSITY OF NORTH CAROLINA, Chapel Hill, NC**June 2015 - June 2016****Post-doctoral Researcher - Dr. Jeff Dangi (National Academy Member)**

Used unsupervised machine learning and network analysis for microbial probiotics for fertilizer use efficiency.

- Mentored undergraduate research and consulted on statistical and computer vision pipelines.

UNIVERSITY OF CALIFORNIA, Davis, CA, Cold Spring Harbor, NY, Copenhagen**June 2009 - June 2015****Graduate Researcher - Dr. Daniel Kliebenstein**

Taught courses and workshops in statistical analysis and bioinformatics at the Dynamo Center of Excellence at the University of Copenhagen and Cold Spring Harbor Laboratory.

- Designed computational and statistical pipelines for laboratory, including computer vision of laboratory imaging; metabolomic, transcriptomic, and genetic mapping; correlational network analysis and dimension reduction; and analysis of stochastic and bet-hedging processes.
- Managed teams of 2-6 people for tissue sampling, molecular DNA and metabolomic processing, and final data collection, processing, and management.
- Designed, built, and administered laboratory server for approximately 12 users.

EDUCATION

Doctor of Philosophy (PhD) - Quantitative Genetics and Systems Biology of Complex Disease Resistance

University of California, Davis

Master of Science (MS) in Plant Biotechnology**Bachelor of Science (BS) in Plant Science and Molecular Biology**

State University of New York (SUNY), College of Env. Sci. and Forestry, Syracuse, NY

TECHNICAL PROFICIENCIES AND EXPERTISE

R: dplyr, tidyr, MASS, car, lme4, ggplot2**Python:** pandas, matplotlib, seaborn, Numpy, Scikit-Learn, SciPy, Pysam, Pybedtools, SciPy**bash/unix:** grep, awk, sed**Alternate Languages:** (SAS | Perl | SQL (PostgreSQL) | Rust)**Code Version Control:** git | GitHub | BitBucket**Bioinformatics Tools:** FastQC | Bedtools | Samtools | Fastx-tools | Jellyfish | BLAST | GATK | Picard | BWA | Bowtie | Tophat | STAR | STAR-Fusion | Metawrap | IGV**Chemoinformatic Tools:** pyMZML | ProteoWizard**Proteoinformatic Tools:** AlphaFold | SSIPe | EvoEF2**Preferred IDE Platforms:** Jupyter | Rstudio | VSCode**Distributed Computing Platforms:** Amazon Web Service | Databricks | Snowflake**AWS Services:** Cost & Management | RDS | EC2 Management | S3 Storage | CloudFront**LIMS:** L7 ESP Client-side and Server-side Management