

JOHN STERRETT

+1(731) 819-7433 ◇ Boulder, CO

sterrettJD@protonmail.com ◇ [LinkedIn](#) ◇ [Website](#) ◇ [GitHub](#)

OBJECTIVE

Quantitative Biology PhD with expertise in software development for **bioinformatics** and **statistical learning**. I develop, evaluate, and apply **robust, biologically-inspired tools to improve decision-making**.

EXPERIENCE

Bioinformatics Applications Scientist

Jun 2025 - Present

Integrated DNA Technologies (Gene Writing and Editing)

Boulder, CO

- Built and deployed cloud-ready pipelines for protein engineering, including bioprospecting, de novo design, and protein optimization.
- Led a hackathon team to develop a protein engineering pipeline that outperformed competitors' state-of-the-art.
- Automated reporting for Synthetic Bio Custom Solutions (SCS) quality control and CRISPR Services, saving 1 hour of labor per sample shipped (Quarto, LaTeX, Python).
- Allowed SCS team to offer complex mutagenesis library design, construction, and sequencing-based quality control reports by developing internal Shiny apps integrated with custom Nextflow pipelines and Python packages.

Graduate Researcher

Aug 2021 - May 2025

Lozupone Lab, University of Colorado Anschutz Medical Campus

Aurora, CO

- Developed and validated a scalable bioinformatics [pipeline](#) and [statistical package](#), enabling analysis of a novel sequencing data type characterized by datasets >80,000 features and many terabytes in size.
- Launched data pipelines as a service after analyzing computational requirements to inform pricing decisions.
- Implemented standards for software quality (via continuous integration for testing, linting, documentation), and for data processing on compute clusters (via workflow management), increasing lab-wide research reproducibility.
- Maintained open-source statistical software developed by previous lab members ([AMON](#), [SCNIC](#)) and expanded functionality in community open-source tools including [QIIME2](#) and [Microshades](#).

Bioinformatics Scholar

Nov 2024 - Mar 2025

MantleBio

San Francisco, CA (remote)

- Enhanced usability of bioinformatics-focused cloud computing platform by creating genomic data processing workflows and tutorials in Python, R, and Nextflow and providing clear bug reports to developers.

Graduate Researcher

Jul 2020 - May 2025

Lowry Lab, University of Colorado

Boulder, CO

- Spearheaded statistical analysis (planning, execution, and review) and machine learning model evaluation for microbiome sequencing collaborative projects, identifying key microbial biomarkers for host immune response.

TECHNICAL EXPERTISE

Bioinformatics

Design: bioprospecting, de novo design (auto-regressive, masked language modeling, diffusion models), intelligent mutagenesis, protein folding

Analysis: co-expression network creation and characterization, differential expression, enrichment and pathway analysis, multi-omic risk scoring, phenotype clustering

Processing: fastq quality control, genome assembly, assembly quality control, genome annotation, genome dereplication and strain tracking, phylogenomics, pangenomics

Common data types: short read NGS (transcriptomics, metatranscriptomics, metagenomics, amplicon sequencing), long read sequencing, proteomics, metabolomics

Computing	Workflow management (Snakemake, Nextflow) and integration of workflows with distributed and cloud computing environments (Slurm-managed high performance compute clusters, AWS, and Google Cloud). Git and continuous integration using Github Actions (testing, linting, documentation)
Languages	Python* (advanced), R* (advanced), C++, Julia *Developed/maintained widely-used packages
Statistics	Advanced regression modeling and causal inference, distance-based/multivariate statistics, Bayesian modeling, longitudinal analysis, bootstrapping, permutation tests
Machine Learning	<u>Supervised learning</u> : linear/logistic regression, support vector machines, random forests, gradient boosting, Naive Bayes classifiers <u>Unsupervised learning</u> : dimensionality reduction (PCA, PCoA, t-SNE), clustering <u>Language models</u> : sequence embedding, de novo protein design (custom implementations of autoregressive, masked, and diffusion models) <u>Other</u> : feature selection algorithms, ensemble learning, model evaluation and performance assessment, data simulation for model assessment
Leadership	<u>Mentoring</u> : mentored 20 individuals on statistical analysis in biomedical research <u>Organization</u> : organized community seminar groups: “Genomic and protein language models” and “Workflow management for metagenomics”, organized Interdisciplinary Quantitative Biology Symposium (150+ attendees), recurring guest lecturer in University of Colorado graduate courses on bioinformatics and the gut-brain axis

EDUCATION

PhD in Integrative Physiology, University of Colorado August 2020 – May 2025
 Interdisciplinary Quantitative Biology Program *GPA: 4.0*
Dissertation: Novel multi-omics methods improve understanding of host-microbiome crosstalk in the mammalian gut
Relevant coursework: bioinformatics and genomics, advanced statistics and research methods, causal inference, software engineering, data structures

Bachelor of Science in Nutrition, East Tennessee State University 2016 - 2020
 Coursework- and research-based honors *GPA: 4.0*

PUBLICATIONS

Complete list of 14 peer-reviewed publications (5 first/co-first author) with over 120 citations can be found on [Google Scholar](#). Selected publications for this application include:

- **Poly-omic risk scores predict inflammatory bowel disease diagnosis**
 Sterrett JD*, Arehart CH*, et al. *mSystems*, 2023.
- **SCNIC: Sparse correlation network investigation for compositional data**
 Shaffer M, Thurimella K, Sterrett JD, Lozupone CA. *Molecular Ecology Resources*, 2023.

AWARDS

Graduate Student Award in Research Excellence (Front Range Microbiome Symposium)	2024
Best Talk (BioFrontiers Institute UpGoer 5 Non-Technical Lightning Talks)	2022, 2023
SymbNET International Host-Microbe Symbiosis Scholar	2023
Undergraduate Research Opportunities Program Team Grant (CUB)	2023
William J. Freytag Fellow (CUB)	2021
Integrated Data Science Fellow (National Science Foundation)	2021
College Outstanding Undergraduate Student (ETSU)	2020