Updated: August 13, 2024

Vivaswat Shastry

Bioinformatics | Computational Biology | Data Science

vivaswats.github.io | vivaswat@uchicago.edu | LinkedIn | Chicago, IL 60615

Experience

- Engineered high-throughput NGS & WGS pipelines for variant calling & deriving insights from population-level sequencing data for 6+ years
    - Experience with diverse data sets from microbes to polyploid plants
- Led statistical modeling & data analysis in 5 peer-reviewed publications (Google Scholar)
    - From phylogenetics of bacterial sequences to metabolomics in mice immunology studies
- Designed & delivered teaching workshops in scientific computing & data visualization for 8+ years
    - Certified Software & Data Carpentry Instructor (from The Carpentries)
    - Optimized workflows daily on high-performance computing (HPC) cluster systems
- Spearheaded advocacy in student-led, university-wide organizations for 10+ years
    - Instituted a Genetics DEI-ship for students as an alternative to a Teaching Assistantship
    - Organized a division-wide ‘Genetics DEI Day’ with >100 participants
- Developed Arduino & Raspberry Pi modules for K-12 math & science aligned with NGSS curriculum
- Amateur stand-up comedian

Education

University of Chicago (Sep 2020 – Jun 2025, *expected*) Chicago, Illinois
Ph.D. in Genetics, Genomics & Systems Biology (Computational Track)

- Conducted a massive simulation study (>100,000 replicates) to assess the impact of extra variables in parameter inference (preprint with results: bioRxiv)
    - Mix of python, R & bash scripts with SLURM array jobs to run in <30 minutes
- Implemented novel method in python to add distant edges in a graph-based framework to derive insights from spatial data (package with user-friendly documentation: FEEMSmix)
    - Decreased computational complexity of algorithm from $\mathcal{O}(nd^2)$ to $\mathcal{O}(n^2d)$ (where $n \ll d$)

University of Wyoming (Jan 2018 – Aug 2020) Laramie, Wyoming
M.S. in Botany

- Developed novel ancestry & genotype estimation software for low-coverage sequencing data in mixed-ploidy individuals using C++ (conda package downloaded >5,000 times: popgen-entropy)
- Improved runtime (10x speedup) and accuracy (30% increase in precision) of a hierarchical Bayesian model with STAN (CRAN package downloaded >12,000 times: CNVRG)

University of Wyoming (Aug 2014 – Dec 2017) Laramie, Wyoming
B.S. in Electrical Engineering, Minor in Computational Science
Honors: *summa cum laude*

- Implemented an MPI-optimized approach that led to strong scaling with ~80% efficiency to extract depth information from a small-motion clip (won ‘Best Poster’ award at RMACC 2017)

Technical Skills

- **Programming**:
    - **python** (NumPy, SciPy, pandas, matplotlib, sklearn, statsmodels, PyTorch, Tensorflow)
    - **R** (dplyr, ggplot2, tidyr, STAN, Rmarkdown, Shiny)
    - **C/C++** (OpenMP, MPI)
    - **MATLAB**
- **Bioinformatics**: samtools, vcftools, bcftools, GATK, plink
- **Version Control & Workflow**: git/GitHub, Jupyter, Bash, SLURM, snakemake

Coursework (selected)

- **Statistics**: Multivariate Data Analysis via Matrix Decompositions, Machine Learning & Applied Statistics, Bayesian Data Analysis, Spatial Statistics, Statistical Theory & Methods
- **Genomics**: Principles of Population Genetics, Human Variation & Disease, Statistical Genetics