# SANDEEP VENKATARAM, Ph.D.

sandeepvenkataram@gmail.com　|　LinkedIn　|　Website　|　GitHub　|　240-672-4403　|　Rockville, MD, USA

Computational Biologist / Data Scientist with expertise in deploying analysis pipelines and visualizations to answer genomic and bio-statistical questions.

## EXPERIENCE

**Computational Biologist**　　　　　　　　　　　　　　　　　　　　　　　　　　　　**May '23 - Present**
*Hexagon Bio*　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Menlo Park, CA*

- **Analysis Pipelines for Early Discovery.** Pipelines were developed for 1) analysis of chemical-genetic data generated internally as part of a cross-functional team and 2) generation of a phylogenetic tree across 75k fungal genomes. **Outcomes:** The chemical-genetic pipeline has handled the analysis of over 50,000 screens to date, and the phylogeny was used to identify a fungal strain producing a compound of interest at 100x the titer of the previous production strain.
- **Analysis and Visualizations to support Late-stage Discovery (LSD) efforts. Outcomes:** Deployed QC pipelines and interactive reports that enabled the LSD team and senior leadership to search, sort, filter, visualize and analyze data from assays ranging from human cell viability assays to *in vivo* assays conducted at CROs. These tools were used 3000+ times per month (in a 65-person company) and saved LSD scientists and leadership from 1000s of hours of effort in data analysis and figure generation. Visualizations were presented at board meetings, business meetings, and international conferences.
- **Machine Learning and bio-statistical R&D.** Trained various classifiers on compound SMILES using proprietary data to predict chemical and biological properties. Developed dynamic programming implementations of various published methods to quantify statistical signals within a phylogenetic tree. **Outcomes:** The in-house fungal phylogeny was largely consistent with the published phylogeny, but ML models were trained with limited labeled data resulting in overfitting.
- **Software development and data analysis to support laboratory processes.** Hexagon has developed a custom internal information management system that integrates materials management with assay design, workflow management and data analysis. **Outcomes:** In collaboration with software engineers and laboratory scientists, I contributed new object models for new types of assays (e.g. *in vivo* assays), implemented software tools for laboratory users to upload data into this system, developed QC infrastructure to monitor data quality, and wrote extensive test suites to ensure data integrity and code functionality.
- **Tools:** Python, R, Snakemake, Docker, Google Cloud Services (GCS) / Google Batch

**Independent Research and Education**　　　　　　　　　　　　　　　　　　　　　　**Oct '22 - April '23**
*Independent*　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Rockville, MD*

- Learned how to visualize data from Venkataram *et al.* (2023) as a dynamic dashboard using the Dash framework (GitHub).
- Implemented a Snakemake pipeline for whole-genome DNA analysis (GitHub).

**Postdoctoral scholar**　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Oct '16 - Sept '22**
*University of California, San Diego*　　　　　　　　　　　　　　　　　　　　　　　　*La Jolla, CA*

- Integrated phenotypic measurements from pooled barcoded libraries and individual growth measurements with genotype data from WGS of 400 isolates and simulations of microbial community dynamics to quantify evolution in a community context. With guidance from my advisor (Nature Ecology and Evolution).
- Conducted extensive literature reviews regarding both barcode library construction and analysis (published in the Journal of Evolutionary Biology) and community evolution (published in Philosophical Transactions B).
- Discovered a novel evolutionary phenomenon via metagenomic time-series analysis of microbial populations (PNAS).

**Ph.D. Candidate**　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Sept '10 - Sept '16**
*Stanford University*　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Stanford, CA*

- Founded a novel research direction within the research group using laboratory evolution in microbes to study adaptation in collaboration with mathematicians and yeast geneticists. I was responsible for designing and executing the experiments, conducting genomic and statistical analysis and writing the manuscripts for two first-author publications (Cell, Current Biology).

## SKILLS

**Cross-functional collaboration.**

- With 2 years of industry experience in cross-functional teams and 12 years of academic experience in designing and executing experiments with both wet-lab and computational components, I have extensive experience working in a cross-functional environment developing bioinformatic solutions in a manner that is cognizant of wet-lab limitations.

**NGS analysis.**

- Constructed a data analysis workflow to process > 1000 barcoded chemical genetic screens / week at Hexagon Bio. Developed a QC protocol and associated dashboards to dynamically visualize the data and conduct statistical tests to summarize the results.
- Developed custom python software to process raw Illumina reads from amplicon sequencing of DNA-barcoded mutant libraries, used by researchers at Stanford, UC San Diego and Arizona State (Nature Ecology and Evolution).
- WGS variant calling pipelines identifying SNPs, InDels and structural variants from short-read sequencing data (Cell).
- Time-series metagenomic variant calling (PNAS).

**Bioinformatics and Quantitative Analysis**
- Constructed a large-scale phylogenetic tree from sequencing data (50,000+ genomes) using Snakemake and Google Cloud. Developed a python module to access, visualize and conduct statistical tests using this tree.
- Developed C++ software to calculate the iHS statistic and conduct evolutionary simulations
- Used gene ontology analysis to identify overrepresented genes, pathways and functions.
- Conducted principal components analysis and hierarchical clustering to identify phenotypically distinct groups of mutants
- Deployed parametric and non-parametric statistical tests and custom resampling analysis of both categorical and continuous data to test biological hypotheses

**Communication**
- Published original research in high-profile scientific journals including Nature, Nature Ecology and Evolution, Cell, Current Biology, PNAS and Philosophical Transactions B
- Contributed oral presentations at 11 national and international scientific conferences
- Synthesized hundreds of scientific articles while writing two peer-reviewed review papers
- Collaborated extensively with diverse researchers including mathematicians, microbiologists and evolutionary biologists

**Key Words**
- **Computational:** Python (Bokeh, Matplotlib); Docker; R; Google Cloud Services; Snakemake; Dash; Jupyter Notebook; R; Java; C++; Perl; Windows, Unix and HPC environments; Git and GitHub
- **Laboratory:** Barcoded library construction; Microbiology; PCR; DNA isolation; NGS library preparation; Cloning; Flow Cytometry; Chemostat culture.
- **Analytical:** Bioinformatics; Experimental design; Statistics; Whole-genome and PCR amplicon NGS analysis; Software design.
- **Leadership skills:** Critical thinking; Project management; Oral and written communication; Interdisciplinary collaboration.

## EDUCATION

**Ph.D. in Biology**      **Stanford University***(Sept '10 – Sept '16)*

**BA. in Biology and Computer Science**      **Washington University in St. Louis***(Aug '06 – May '10)*

## AWARDS

- Best Talk, *Evolution of Complex Life, Atlanta, GA*  ('19)
- Finalist, NIH Director's Early Independent Award *(DP5)* ('18)
- Runner-up, Crow Award *Genetics Society of America* ('16)

## SELECTED RESEARCH ARTICLES

**Venkataram S**, Kuo HY, Hom EFY and Kryazhimskiy S. "Mutualism-enhancing mutations dominate early adaptation in a two-species microbial community". **Nature Ecology and Evolution**, 2023

**Venkataram S**, Monasky R, Sikaroodi SH, Kryazhimskiy S and Kacar B. "Evolutionary stalling and a limit on the power of natural selection to improve a cellular module". **PNAS**, 2020

Li Y\*, **Venkataram S\***, Agarwala A, Dunn B, Petrov DA, Sherlock G and Fisher DS. "Hidden complexity of yeast adaptation under simple evolutionary conditions". **Current Biology**, 2018

**Venkataram S\***, Dunn B\*, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Herissant L, Blundell JR, Levy SF, Fisher DS, Sherlock G and Petrov DA. "Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast". **Cell**, 2016

Levy SF\*, Blundell JR\*, **Venkataram S**, Petrov DA, Fisher DS and Sherlock G. "Quantitative evolutionary dynamics using high-resolution lineage tracking". **Nature**, 2015

**Venkataram S** and Fay JC. "Is transcription factor binding site turnover a sufficient explanation for cis-regulatory sequence divergence?". **Genome Biology and Evolution**, 2010

\*Equal contribution

## CONFERENCE PRESENTATIONS

**11 Contributed oral presentations (2013-2022) at National and International conferences**

## LEADERSHIP, TEACHING, MENTORING AND SERVICE

- **Conference co-organizer**, Biomedical Computation at Stanford Conference (2012): responsible for fundraising and organizing the venue, advertising, selecting speakers, organizing the program and managing the event.
- **Symposium Organizer and Chair**, "Mapping genotype and phenotype to fitness", SMBE 2018 (Yokohama, Japan)
- **Mentor** for 5 PhD rotation students and 3 undergraduate summer research scholars. Students successfully presented their research at the conclusion of their training.
- **Teaching assistant**: Introductory Computer Science ('07 - '09); Molecular Evolution ('11); Introductory Biology ('12 - '13)
- **Guest lecturer**: Population Genetics ('18); Seminars in Biomedical Research ('21)
- **Volunteer Judge**, Regional High School Science Fairs ('11 - '22)