

ARUDHIR (ARU) SINGH



arudhir@gmail.com

224.619.0499

Microbial bioinformatics *in nebulo*. Force multiplier for experimental biology through sequencing-based analyses; strong proponent of the democratization of and minimal barriers to bioinformatics science through intuitive interfaces and human-centric outputs. Lead developer for several high-impact production NGS pipelines.

SKILLS

Bioinformatics

Genomics: Alignment, SNP and SNV analysis, assembly, annotation, comparative genomics
Transcriptomics: Mapping-based quantification, differential expression, pathway enrichment
Epigenomics: ATAC-seq
Organisms: Gram positive and negative bacteria, yeasts (*S. cerevisiae*, *P. pastoris*, *Y. lipolytica*)

Molecular Biology

Cloning: Restriction, Type IIS, Gibson
Automation: Hamilton, Echo
Purification: Plasmid, genomic, protein
Organisms: *E. coli*, *S. cerevisiae*

Programming

Languages: Python (advanced), Bash (intermediate), Java (beginner), C++ (beginner)
Deployment: Docker, AWS, GitLab
Query: GraphQL, SQL

EXPERIENCE

Computational NGS 3, Ginkgo Bioworks

Jan 2022 – Aug 2022

- Lead developer and project manager for hybrid genome assembly pipeline on AWS Batch
 - Performs genome assembly using seven different short and long read assemblers;** returns short-read polished assemblies, processed reads, and quality metrics to guide further curation as required
 - Reduced time required to create quality assemblies by a factor of ten** between 2021 and 2022
 - Employed by several bioinformaticians to generate **over fifty reference-quality bacterial assemblies**
 - Coordinated the generation of gold-standard datasets and curated assemblies for twelve high-value chassis strains and **decreased error-rates by up to several orders of magnitude** across standard contiguity, consistency, and completeness metrics
 - Brought pipeline to production-quality by including thorough logging, stable UX, and CI/CD coordinated through GitLab
- Developed AWS Batch pipeline to contextualize short read scaffolds from several thousand diverse environmental isolates
 - Collaborated with experimentalists to ensure the resultant data is relevant, actionable, and accessible
 - Focused on taxonomic identification, functional annotations, and virulence factors
- Standardized computational on-boarding across several computational and experimental teams, **reducing time to on-board computational biologists across the company from months to minutes**
 - Collaborated with DevOps and IT engineers to create a "batteries-included" bioinformatics Ansible playbook
 - Managed and prioritized feedback from a diverse group of users and communicated priorities to DevOps
 - Decreased overhead involved in adhering to best practices and subsequent troubleshooting

Systems Biologist 2, Ginkgo Bioworks

Jul 2020 – Jan 2022

- Lead developer for transcript quantification pipeline used on over one hundred bulk RNAseq datasets, several organisms, and utilized by dozens of biologists and bioinformaticians
 - Containerized Snakemake workflow performing rRNA filtering, read QC, quasi-mapping based transcript quantification, and collected summary statistics
- Coordinated the genome assembly, annotation, and analysis of 28 strains relevant to the human skin microbiome
- Created several reports relating observed fermentation profiles in *S. cerevisiae* strains to underlying transcriptome
 - Transcript quantification, differential expression, and pathway enrichment
 - Confirmed weak acid stress driving decreased productivity in engineered strain
 - Identified diverse transcriptional profiles in wild isolates despite similar production profiles
 - Provided evidence that lack of observed production was due to mixed culture rather than promoter exhaustion
- Created seven-lecture NGS analysis workshop designed for wetlab scientists
 - Lectured for two hours weekly with discussion sections and homework assignments
 - Tailored course content to common themes present in observed analysis requests; theory and practice involving FASTQ manipulation, alignment, *de novo* assembly, and variant calling
- Helped rapidly iterate on prototypes of a server-less AWS Batch architecture from DevOps
 - Identified bugs, quirks, and areas of improvement over several iterations
 - Demonstrated and promoted usage through developing several analysis pipelines and on-boarding other informaticians

EXPERIENCE (CONT.)

Strain Engineer 1, Ginkgo Bioworks

Jul 2018 - Jul 2020

- Developed and ran several sessions of an introduction to Python course designed to teach fundamental computer science concepts and practical programmatic techniques for experimentalists (over fifty students, over a dozen informatics converts)
- Purified and delivered proteins of interest to external collaborator
- Wrote report advocating the decision to license BioCyc and coordinated the subsequent server setup
- Developed Python package for rapid microbial pathway enrichment and analysis from RNA-seq data
- Assembled probiotic strains and validated genome stability for external customer prior to FDA submissions
- Lead developer and maintainer of several Python command-line interfaces to interact with internal LIMS data
- Functional multi-omics to elucidate a cryptic enzyme participating in core metabolic pathway using transcriptomics and proteomics; comparative genomics in PATRIC showed similar gene architecture in a number of gram negative strains

Strain Engineer Intern, Ginkgo Bioworks

Jul 2017 - Jul 2018

- Developed prototype large scale one-pot combinatorial library assembly workflow
 - Optimized wetlab protocol by experimenting with methods and transferring the Echo and Hamilton
 - Developed WGS-based NGS pipeline to identify and quantify constructs from transformed isolates
 - Developed Nanopore-based NGS pipeline to identify and quantify constructs from mixed strain liquid cultures
 - Delivered hundreds of sequence-verified constructs to external client
- Variant analysis of long-term evolution experiment (1 year+) using *breseq*

Undergraduate Intern, Energy Biosciences Institute

Oct 2015 - May 2016

- Screened candidate proteins for sequence motifs associated with high-affinity xylose transportation from *Yarrowia lipolytica* and introduced them into *Saccharomyces cerevisiae* to aid co-fermentation of glucose and xylose.

Team Lead, iGEM

Jan 2015 - Oct 2016

- Led a team of ten undergraduates to submit and present an entry in the Information Processing track at iGEM
- Spearheaded the standardization of a pre-existing synthetic cellular analog memory construct as well as integrated enhanced usability and modularity for the system in *Escherichia coli*
- Trained new members of the team in lab protocols and proper experimental procedures
- Planned and delegated daily wet-lab and dry-lab tasks and managed monthly expenses

Undergraduate Researcher, iGEM

Jan 2014 - Oct 2015

- Engineered *Lactobacillus plantarum* to degrade theobromine through a heterologous N-demethylase pathway to counteract chocolate toxicity in small animals
- Developed and utilized a Python2 script to expedite diagnostic restriction digests

EDUCATION

University of Illinois, Champaign IL (2012-2016)

Bachelor of Science (B.S.)

Major: Engineering Physics

Concentration: Computational Physics

Minor: Bioengineering, Mathematics