

YANGYANG LI

yangyangli.top | github.com/cauliyang | linkedin.com/in/yangyang-liz5

SELF-INTRODUCTION

I am a fourth-year Ph.D. candidate at the intersection of deep learning, algorithmic efficiency, and computational biology. My research harnesses cutting-edge machine learning techniques and high performance computing to unravel complex biological enigmas. With a proven track record in developing innovative research software and data-driven solutions, I excel at translating intricate scientific challenges into actionable insights. Driven by a passion for pushing the boundaries of bioinformatics, I am poised to make significant contributions in a dynamic, research-intensive environment where I can continue to evolve as a scientist and innovator.

RESEARCH EXPERIENCE AND PROJECT

Northwestern University <i>Ph.D. in Bioinformatics</i>	Chicago, US <i>June 2022 – June 2025</i>
<ul style="list-style-type: none">Developed a genomic language model that improved chimera artifact detection accuracy by 80% in Nanopore Direct RNA Sequencing, enhancing data reliability for genomic researchArchitected PxBLAT, a high performance Python library interfacing with BLAT (C-based command-line tool), achieving a 20% speed improvement in genomic sequence alignmentsDeveloped a novel graph algorithm that increased non-linear transcript identification accuracy in long-read sequencing data, facilitating improved understanding of complex RNA structuresCreated Aurora, a web application for intuitive graph algorithm visualization, reducing algorithm comprehension time and enhancing collaboration among bioinformatics researchersFormulated a deep generative model tailored for sequencing data simulation	
University of Minnesota <i>Ph.D. in Bioinformatics and Computational Biology</i>	Minneapolis, US <i>Sep. 2020 – June 2022</i>
<ul style="list-style-type: none">Innovated a novel algorithm that improved detection of non-linear structural variations in transcriptomes, enabling more comprehensive analysis of complex RNA structures and alternative splicing eventsDeveloped a transformer-based deep learning model to predict causality between gene fusion and structural variationLed a rigorous evaluation of 10+ leading tools for alternative splicing variant detection, resulting in optimized tool selection guidelines that increased detection accuracy across diverse genomic datasetsCourses (Grade): Advanced Machine Learning (A), Introduction to Data Mining (A), Adv. Algs. & Data (B)	
China Agricultural University <i>Master in Crop Bioinformatics</i>	Beijing, CN <i>Sep. 2018 – June 2020</i>
<ul style="list-style-type: none">Leveraged machine learning techniques to identify key genetic markers across 1,400 maize genomics datasets, leading to improvement in predicting high-yield agronomic traits for crop enhancementSpearheaded a comprehensive genome-wide association study (GWAS) mapping genetic variations to maize ear attributes across 450 diverse natural populations	

WORK EXPERIENCE

Northwestern University IT Organization <i>Student Consultant in Data Science, Statistics, and Visualization</i>	Chicago, US <i>June 2024 – now</i>
<ul style="list-style-type: none">Analyzed complex collaboration data and created compelling social network visualizations, contributing to a funding applicationDeveloped a Rust-based data summarization tool that processed millions of records, reducing analysis time by 75% and improving decision-making efficiencyDesigned and led a comprehensive workshop on data visualization with Seaborn	

EDUCATION

Northwestern University <i>Ph.D in Bioinformatics. GPA: 3.7</i>	Chicago, US <i>June 2022 – June 2025</i>
University of Minnesota <i>Ph.D. in Bioinformatics and Computational Biology. GPA: 3.68</i>	Minneapolis, US <i>Sep. 2020 – June 2022</i>
China Agricultural University <i>Master in Crop Bioinformatics. GPA 3.14</i>	Beijing, CN <i>Sep. 2018 – June 2020</i>
Northeast Agricultural University <i>Bachelor of Arts in Agricultural Engineering. GPA 3.04</i>	Harbin, CN <i>Sep. 2014 – June 2018</i>

TECHNICAL SKILLS

Languages and Frameworks: C++, Python, Rust, Pytorch, Jax, Candle, GGML
Development Stack: Neovim, GDB, Git, Numpy, Pandas, Matplotlib, Docker, GitHub Action, CMake, HTML, GCC, Clang, Linux, L^AT_EX
Specializations: Algorithm Development, Concurrency Programming, Data Analysis and Visualization, Natural Language Processing

GRANTS AND HONORS

- First place of Computation and Data Exchange (CoDEx) Interactive Visualization Challenge (2024)

CONFERENCE TALK

- Computation and Data Exchange (CoDEx) Visualization Challenge
- Workshop: Data Visualization with Seaborn

PUBLICATIONS

Li, Yangyang, & Yang, R. (2024, 12). PxBLAT: an efficient python binding library for BLAT. *BMC Bioinf.*, 25(1), 1–8. doi: 10.1186/s12859-024-05844-0

Fry, J., **Li, Yangyang**, & Yang, R. (2022, 09). ScanExitronLR: characterization and quantification of exon splicing events in long-read RNA-seq data. *Bioinformatics*. doi: 10.1093/bioinformatics/btac626