ANNI ZHANG, PhD • caozhichongchong@gmail.com
78 High Tower Rd, South Windsor, CT 06074 • Lab website Google Scholar

Objective

I'm exploring opportunities for work that is useful and interesting, most likely using my skills in AI/ML, bioinformatics, algorithms, microbiology, genomics, and/or statistics

About Me

I value fast feedback and checking risky things first. I identify interesting questions before looking at data, and I often build a small proof of concept to test whether an idea has practical or business value. I prefer to study a problem end-to-end on a small dataset before optimizing or scaling a pipeline. I also value efficient communication: I ask and welcome clarification questions early, and I avoid large-group discussions to keep meetings more efficient.

Skills

- Experienced in generating project ideas, building, and leading teams
- Comfortable with Python, Bash, Git, HPC workflows, and algorithm design
- Comfortable with sequence alignment, variant calling, multi-omics integration, microbiology
- Familiar with AI/ML model design, sequence-based LLMs, graph neural network applications

Work Experience

Assistant Professor at Nanyang Technological University, 2024 – present

Establish collaboration

- In 1.5 years, I have established projects with experts across three continents in cancer therapy, virology, gut microbiology, neuroscience, social sciences, computational biology, structural biology, bacterial engineering, and phage therapy. I am the lead PI on a $1.0 M Singapore MOE Tier 2 grant (awarded), and a co-PI on a $40 M Harvard–NTU NRF CREATE application, two $10 M MOE Tier 3 applications (shortlisted by the university), and a $1.0 M MOE Tier 2 application (shortlisted by the university).

Establish product vision and lead the science team

I designed research directions and project ideas for PhD students who subsequently received full scholarships from the university. I lead a team of a former Google software engineer, a bioinformatician, and two microbiologists. My long-term aim is to design bacteria-based therapies via precise engineering. My strategy is to generate new training data by applying biological principles to data that we already have, without relying on new experiments. Projects include:

- Develop a deep neural network to predict bacterial protein-DNA binding. The idea is to use biological principles such as autoregulation and local regulation to extract information from all (~0.5 million) bacterial genomes.
- Identify druggable microbiome targets to treat diseases using a graph neural network trained on LLM-powered knowledge graph. The idea is to predict new links between microbial metabolites/genes and diseases that would show up in the next 5 years, generating new hypotheses for wet-lab validation.

- Develop a deep neural network to predict protein-ligand binding. The idea is to apply gene annotation tools to all (~0.5 million) bacterial genomes to pair each protein sequence with the molecules that it interacts with.
- Identify positively selected (beneficial) genes in the gut microbiome using population genetics and multi-omics analyses. The idea is to compare survival rates between strains that acquired genes via horizontal gene transfer and strains that did not.

Mentoring

- Primary principle investigator (PI) for 3 PhD students, 2 undergraduate students, and 2 master's students
- Co-PI for 3 PhD students
- Thesis advisory committee member for 9 PhD students
- Thesis examiner for 4 PhD students and 2 master's students

Teaching

- Coordinate and teach a 315-student undergraduate course, BS1009 Introduction to Computational Thinking. Teach Python programming and logic; lead hands-on computational labs.
- Guest lecture for BIO-0196 Exploratory Topics in Biology – Microbial Genome Evolution, invited by Dr. Fatima Aysha Hussain at Tufts University (April 2026 - scheduled). Introduce sequence alignment, limitations of existing tools, the X-Mapper algorithm, and a practical demo.

Science communications

- Invited speaker at 12 seminars and conferences since mid-2024

Community service

- Working Group Member, National Infectious Disease Research Strategy. Serve as one of the six experts in the national working group developing Singapore's infectious disease research strategy for the next 5–10 years.
- Chair of Applied and Environmental Microbiology Conference Gordon Research Seminar (GRS). For this two-day seminar, I designed sessions, invited speakers, selected short talks and session chairs from submitted abstracts, and organized a career panel.
- Host visits and seminars of Dr. Xiaoqian Yu's (University of Montreal) and Dr. Klas Udekwu's (University of Idaho & SLU).
- Nominated for 2026 Asian Young Scientist Fellowship
- Organizing committee member for the NTU-Sorbonne Université Joint Winter School "AI for Sciences". I designed the AI for Biology session and selected the speakers.

Senior Postdoc / Research Associate at Massachusetts Institute of Technology, 2018 – 2024

Establish product vision and lead the science team

- X-Mapper: fast and accurate sequence aligner for microbial sequencing data. The idea is to use gapped x-mers to address microbial genomic diversity. Jeffry Gaston, Eric Alm, and **An-Ni Zhang**. "X-Mapper: fast and accurate sequence alignment via gapped x-mers". *Genome Biology* 26, 15 (2025). NTU news
- Quick-variants - fast and accurate microbial genetic variant identifier. Jeffry Gaston, Eric Alm, and **An-Ni Zhang**. "Fast and Accurate Variant Identification Tool for Sequencing-Based Studies." *BMC Biology* 22, no. 1 (2024): 90.

Independent research project

- Combine population genetics and multi-omics analyses to quantify gut microbiome CRISPR immunity update rate in healthy individuals to guide phage therapy. **An-Ni Zhang**, Jeffry

Gaston, …, and Eric Alm. "CRISPR spacer acquisition is a rare event in human gut microbiome." *Cell Genomics* 5, no. 1 (2025). MIT news Cell Genomics Jan 2025 Cover Image

- Use multi-omics analysis and theoretical frameworks to evaluate AMR risk to human health, prioritizing 3.6% of AMRs as the highest risk to enhance surveillance efforts. **An-Ni Zhang**, Jeffry Gaston, …, and Eric Alm. "An omics-based framework for assessing the health risk of antimicrobial resistance genes." *Nature communications* 12, no. 1 (2021): 1-11.

Mentor juniors

- I mentored 4 PhD students and 1 postdoc, organized group meetings and lab activities, and supported research design, career development, and communication among postdocs, PhD students, and our PI.

Science communications

- Invited speaker at 23 seminars and conferences since 2018

Teaching

- MIT Kaufman Teaching Certificate. I learned principles of backward course design and active learning, and designed the syllabus for "Bioinformatics for Microbial Sequencing Data."
- Guest lecturer for Course 20.265 "Genetics for Biological Engineering" at MIT (two lectures)

Education

University of Hong Kong 2014-2018

PhD in Microbiology, Bioinformatics, and Environmental Engineering

Thesis: "PAOs in phosphorus removal reactors and ARGs in environmental microbiomes: new insights from bioinformatics." Trained in Python and bioinformatics; applied to environmental microbiology, including identifying mobile genetic elements associated with antimicrobial resistance spread.

Main publications

An-Ni Zhang, …, and Tong Zhang. "Online searching platform for the antibiotic resistome in bacterial tree of life and global habitats." *FEMS Microbiology Ecology* 96, no. 7 (2020): fiaa107. (**Editor's Choice article**)

An-Ni Zhang, …, and Tong Zhang. "Mining traits for the enrichment and isolation of not-yet-cultured populations." *Microbiome* 7, no. 1 (2019): 1-13.

An-Ni Zhang, …, and Tong Zhang. "Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection." *Microbiome* 6, no. 1 (2018): 1-14.

Nanjing University 2010-2014

BS in Environmental Biology

- I studied antimicrobial resistance mechanisms in yeast by competitively growing a yeast knockout collection in the presence of triclosan, sequencing, and quantifying each knockout strain using barcodes.