

# Zhenmiao Zhang

zhz142@ucsd.edu, 86-18810979822

## Research Interests

- Telomere-to-telomere genome assembly
- Complete metagenomics, metagenome contig binning
- AI for genomics, AI in drug discovery

## Education Background

03/2024 - Now	<b>Postdoctoral Fellow</b> , Department of Computer Science <b>University of California, San Diego</b> , supervised by <i>Prof. Pavel A. Pevzner</i>
09/2019 - 02/2024	<b>Ph.D.</b> , Department of Computer Science <b>Hong Kong Baptist University</b> , supervised by <i>Dr. Lu Zhang</i>
02/2023 - 11/2023,	<b>Visiting Graduate</b> , Department of Computer Science & Engineering <b>University of California, San Diego</b> , supervised by <i>Prof. Pavel A. Pevzner</i>
09/2015 - 06/2019	<b>Bachelor of Engineering</b> , Computer Science <b>Bachelor of Science</b> , Mathematics and Applied Mathematics <b>Renmin University of China</b>

## Projects

03/2023 - 09/2024	<b>Inferring Synteny Blocks and Segmental Duplications in Complete Genomes</b> We developed GenomeDecoder for inferring synteny blocks and segmental duplications and applied it to analyzing genomic architectures of primate genomes ( <i>Bioinformatics</i> , 2024). This is a companion paper to <i>Complete sequencing of ape genomes</i> ( <i>Nature</i> , 2025).
08/2022 - 04/2025	<b>Contig binning using dynamic contrastive learning with deep language model</b> We developed CompleteBin, a contig binning method that trains a pretrained deep language model using both long and short contigs and clusters the contigs based on their embeddings generated by the language model. ( <i>In submission to Nature Biotech</i> ).
09/2021 - 06/2023	<b>Improved metagenome assembly using short reads with long-range connectivity</b> We developed Pangaea, a metagenome assembler that significantly improved contig contiguity and genome draft quality ( <i>Nature Communications</i> , 2024).
09/2020 - 09/2021	<b>A comprehensive benchmarking of metagenome assemblers</b> We evaluated 19 mainstream metagenome assemblers using 32 metagenomic sequencing datasets ( <i>Briefings in Bioinformatics</i> , 2023).
12/2020 - 02/2021	<b>A unified and versatile linked-read toolkit</b> We developed LRTK to do data simulation, read preprocessing, barcode-aware read alignment, variant calling, and phasing for linked reads ( <i>GigaScience</i> , 2024).
07/2020 - 09/2020	<b>Improved contig binning using assembly and paired-ends graphs</b> We developed METAMVGL, a contig binning tool that integrates assembly and pairedend graphs to improve contig binning ( <i>BMC Bioinformatics</i> , 2021).
09/2019 - 09/2024	<b>Construct reference genome database for human gut microbiomes</b> We created a high-quality reference genome database for human gut microbiomes with 5,785 species clusters and 612 novel species ( <i>Under review at Microbiome</i> ).

## Teaching Experience

09/2022 - 12/2022	<b>COMP1007</b> Introduction to Python and Its Application
01/2022 - 06/2022	<b>COMP7560</b> Information Systems Auditing
01/2022 - 06/2022	<b>COMP4046</b> Information Systems Control and Auditing
01/2021 - 06/2021	<b>COMP1007</b> Introduction to Python and Its Applications
09/2020 - 12/2020	<b>COMP1007</b> Introduction to Python and Its Applications
01/2020 - 06/2020	<b>COMP7250</b> Machine Learning

## Conferences

- Poster, Improve metagenomic binning by recovering short contigs using deep learning, ***RECOMB*** 2023.
- Oral Presentation, Exploring high-quality microbial genomes by assembly of linked-reads with high barcode specificity using deep learning, ***RECOMB-SEQ*** 2023.
- Oral Presentation, METAMVGL: a multi-view graph-based metagenomic contig binning algorithm by integrating assembly and paired-end graphs, ***APBC*** 2021.

## Talks

- Exploring high-quality microbial genomes by metagenome assembly and contig binning using long-range connectivity and deep learning, ***the 26<sup>th</sup> Postgraduate Research Symposium of the Department of Computer Science***, HKBU 2023.
- Nucleotide sequence binning and applications, ***the 25<sup>th</sup> Postgraduate Research Symposium of the Department of Computer Science***, HKBU 2022.
- Metagenome assembly and binning on high-throughput sequencing data, ***the 24<sup>th</sup> Postgraduate Research Symposium of the Department of Computer Science***, HKBU 2021.
- Metagenome assembly: technologies, algorithms and applications, ***the 23<sup>rd</sup> Postgraduate Research Symposium of the Department of Computer Science***, HKBU 2020.

## Professional Activities

- Committee Member, ***BIBM Workshop Computational methods to characterize genomic variants using high-throughput sequencing data***, 2022.

## Grants

- Novel algorithms of metagenome assembly and haplotype phasing using single-tube Long Fragment Read sequencing technology, supported by ***the open project of BGI-Shenzhen***, Shenzhen, 518000, China.

## Awards and Honors

02/2023	Excellent Teaching Assistant Performance Award (COMP1007 Python)
02/2021	Excellent Teaching Assistant Performance Award (COMP1007 Python)
02/2020	Excellent Teaching Assistant Performance Award (COMP1007 Python)
02/2018	Meritorious Winner of International Mathematical Contest in Modelling
12/2014	First Prize in National High School Mathematics Contest, Zhejiang

## Publications

- Yoo et al., Complete sequencing of ape genomes, *Nature*, 2025
- Zhenmiao Zhang\***, Ishaan Gupta, Pavel A Pevzner, GenomeDecoder: Inferring Segmental Duplications in Highly-Repetitive Genomic Regions, *Bioinformatics*, 2025
- Zhenmiao Zhang**, Jin Xiao, Hongbo Wang, Chao Yang, Yufen Huang, Zhen Yue, Yang Chen, Lijuan Han, Kejing Yin, Aiping Lyu, Xiaodong Fang, Lu Zhang\*, Exploring high-quality microbial genomes by assembling short-reads with long-range connectivity, *Nature Communications*, 2024
- Chao Yang, **Zhenmiao Zhang** (co-first), Yufen Huang, Xuefeng Xie, Herui Liao, Jin Xiao, Werner Pieter Veldsman, Kejing Yin, Xiaodong Fang, Lu Zhang\*, LRTK: a platform-agnostic toolkit for linked-read analysis of both human genome and metagenome, *GigaScience*, 2024
- Zhenmiao Zhang**, Chao Yang, Werner Pieter Veldsman, Xiaodong Fang, Lu Zhang\*, Benchmarking genome assembly methods on metagenomic sequencing data, *Briefings in Bioinformatics*, 2023
- Zhenmiao Zhang**, Lu Zhang\*, METAMVGL: a multi-view graph-based metagenomic contig binning algorithm by integrating assembly and paired-end graphs, *BMC Bioinformatics*, 2021