



Shaopeng Liu

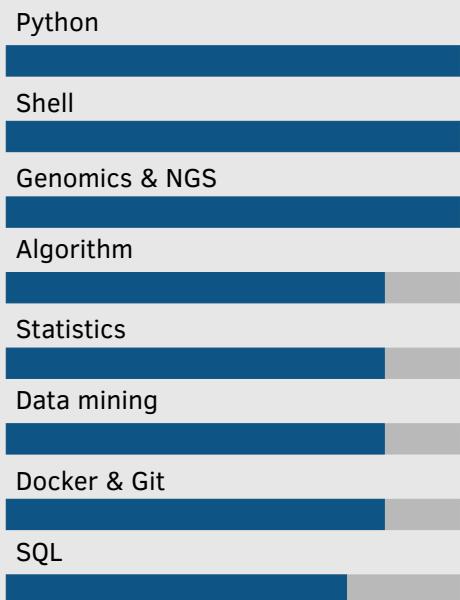
Ph.D. candidate
Bioinformatics

sml6467@psu.edu
 +1 3145852220
 State College, PA, 16803
 PDF: Curriculum vitae

Keywords

- 1. Bioinformatics
- 2. Metagenomics
- 3. K-mer-based algorithms
- 4. Genomic data science
- 5. Biomedical knowledge graphs

Skills



Summary

Dedicated bioinformatician with over 6 years of expertise in processing genomic datasets and executing complex computational tasks. Proficient in crafting and executing extensive pipelines and innovative algorithms. Presently engaged in the development of sketching-based algorithms and graph learning techniques for metagenomic samples. Eager to harness efficient algorithms, multi-omic data, artificial intelligence, and biomedical knowledge graphs to tackle health-related challenges.

Education

- 2019-2024 Ph.D. in Bioinformatics & Genomics
2016-2017 M.Sc. in Biostatistics
2010-2014 B.Sc. in Biology

Pennsylvania State University
Washington University in St. Louis
Wuhan University

Experience

Computational Biology intern (2023.6-8, 23andMe)

- created pipelines for the analysis and benchmarking of multiome single-cell datasets.

Bioinformatic intern (2022.5-8, Gilead Sciences)

- developed innovative k-mer-based algorithms for HBV genotyping in clinical samples.

Doctoral researcher in computational biology (2020-2024, PSU)

- proved, implemented, and benchmarked a truncation-based containment Min-Hash algorithm CMash for multi-resolution estimation of Jaccard and containment indices in metagenomic analysis;
- utilized k-mer-sketching-based methods to develop comprehensive pipelines to explore the "microbial dark matter";
- worked in the Biomedical Data Translator Consortium that aims at providing a graph-based reasoning tool for biomedical and translational studies;
- constructed a metagenomic-specific knowledge graph by integrating public resources and adopted graph learning methods to for data mining purposes in metagenomics.

Bioinformatician (2017-2019, WUSTL)

- engaged in collaborative efforts with diverse teams and conducted bioinformatic analyses on a spectrum of genomic and epigenomic datasets, including ChIP-seq, RNA-seq, WGBS, ATAC-seq, and single-cell RNA data.
- worked in the Data Coordination Center (DCC) of TaRGET II Consortium, streamlined and set the routine pipelines for RNA-seq, ChIP-seq, WGBS, and ATAC-seq data; and processed part of data with preliminary quality analysis;
- developed AIAP, an advanced ATAC-seq analysis pipeline for comprehensive analysis of ATAC-seq dataset; and also implemented the pipeline into Docker and Singularity

Publication

- 2023 *Primer: Analysis of human metagenomic data (*In-progress*)
2023 *MKG: a microbial knowledge graph for metagenomic data mining (*In-progress*)
2023 *Fast, lightweight, and accurate metagenomic functional profiling using FracMinHash sketches (*Submitted*)

*: first or co-first author



Shaopeng Liu

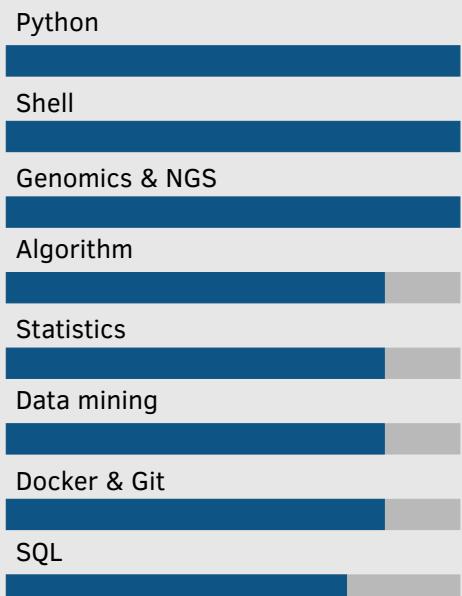
Ph.D. candidate
Bioinformatics

sml6467@psu.edu
 +1 3145852220
 State College, PA, 16803
 PDF: Curriculum vitae

Keywords

1. Bioinformatics
2. Metagenomics
3. K-mer-based algorithms
4. Genomic data science
5. Biomedical knowledge graphs

Skills



2023	*Connecting Syncmers to FracMinHash: similarities and advantages (<i>Submitted</i>)
2023	ARAX: a graph-based modular reasoning tool for translational biomedicine. <i>Bioinformatics</i>
2022	Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science[J]. <i>Clinical and Translational Science</i>
2022	Progress toward a universal biomedical data translator[J]. <i>Clinical and Translational Science</i>
2022	*CMash: fast, multi-resolution estimation of k-mer-based Jaccard and containment indices. <i>Bioinformatics</i>
2021	*AIAP: A Quality Control and Integrative Analysis Package to Improve ATAC-seq Data Analysis. <i>Genomics, Proteomics & Bioinformatics</i>
2021	Deconstructing Stepwise Fate Conversion of Human Fibroblasts to Neurons by MicroRNAs. <i>CELL STEM CELL</i>
2020	Comparison of differential accessibility analysis strategies for ATAC-seq data. <i>Scientific reports</i>
2019	Conditional Activation of NF- κ B Inducing Kinase (NIK) in the Osteolineage Enhances Both Basal and Loading-Induced Bone Formation. <i>Journal of Bone and Mineral Research</i>
2018	The role of Twist1 in mutant huntingtin-induced transcriptional alterations and neurotoxicity. <i>Journal of Biological Chemistry</i>
2017	Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development. <i>Proceedings of the National Academy of Sciences</i>
2014	Multistage Regulator Based on Tandem Promoters and CRISPR/Cas. <i>ACS Synthetic Biology</i>

Awards

2019-2021	The Graham Endowed Fellowship	Pennsylvania State University
2019	The J. Lloyd Huck Graduate Fellowship	Pennsylvania State University
2019	The Braddock Scholarship	Pennsylvania State University
2016	WU Scholarship and Grants	Washington University in St. Louis
2013	Golden medal	International Genetically Engineered Machine (iGEM)
2010-2012	College Scholarship	Wuhan University