Jiayu SHANG

Artificial Intelligence Thrust, Information Hub
Hong Kong University of Science and Technology (Guangzhou)
jiayushang@hkust-gz.edu.cn
+(86) 18420137826
+(852) 56113626

Education and employment

Jan. 2027 – now	Assistant Professor (Full-time)
	AI Thrust, Information Hub, Hong Kong University of Science and Technology (Guangzhou)
Oct. 2023 – Dec. 2026	Postdoctoral fellow (Full-time)
	Department of Information Engineering, The Chinese University of Hong Kong
Aug. 2019 - Oct. 2023	Doctor of Philosophy (Full-time)
	Department of Electrical Engineering, City University of Hong Kong
Jan. 2019 – Aug. 2019	Research Assistant (Full-time)
	Department of Information Engineering, The Chinese University of Hong Kong
Sep. 2018 – Dec. 2018	Research Assistant (Full-time)
	Department of Information Engineering, The Chinese University of Hong Kong
Jul. 2014 - Aug. 2018	Bachelor of Engineering (Full-time)
	School of Data and Computer Science, Sun Yat-sen University

Research interest

Computational biology and Bioinformatics

Developing deep learning architectures for large-scale genomic data mining, including classification, functional annotation, and pattern discovery in complex microbial communities.

Generative model for therapeutic design

Applying generative models and autonomous AI agents to de novo protein design, bridging computational prediction with real-world biomedical applications.

AI for public and precision Health

Building data-driven frameworks to model pathogen evolution in environmental systems and analyzing longitudinal multi-omics cohort data to uncover microbiome–disease associations in neurodegenerative conditions.

Highlights

High-impact publications

Published **36 peer-reviewed paper**, including **33 JCR Q1** paper, **19 first/corresponding author**, 6 with **IF > 10**, **10 CCF-A** conference/journal papers. **1,521** total citations with **1 ESI highly cited paper**; H-index 16

Well-known computational platforms for viral sequence analysis

Developed **PhaBOX**, a globally recognized web server with **770,000+ views** and **31,000+ submissions** from **1,300+** researchers. Its utilities are featured in elite journals including ***Science***, ***Nature Microbiology***, and ***Cell Host & Microbe***.

Software ecosystem

Established a suite of bioinformatics tools (e.g., PhaGCN, PhaTYP) with **12,000+** download and **500+** stars on GitHub; **9,300+** install via Bioconda.

Translation and entrepreneurship

Founded **Hydro-Brain (**person in charge**)**, a multi-modal AI platform for ecological risk assessment, awarded **200,000 HKD** seed funds from Hong Kong government and City University of Hong Kong.

Grants

As Principal Investigator

1. **Guangdong Provincial Key Laboratory of Wastewater Information Analysis and Early Warning Open Fund Project "**Identification and Risk Prediction of Drug-Resistant Pathogens in Wastewater Based on Deep Learning Technology"; 2025-2027; RMB 80,000

As Participant

1. **RGC Strategic Topics Grants:** Foundation Models-Empowered Ambient Intelligence Systems for Early Diagnosis, Personalized Intervention, and Complex Cross-Disease Interplay Analysis of Aging-Related Degenerative Diseases; HKD 39,170,000 **(Drafting and technical development)**
2. **Hong Kong RGC GRF:** Advancing Plasmid Characterization in Metagenomic Data: Plasmid Protein Annotation and Host Range Identification; HKD 1,038,967 **(Drafting and technical development)**
3. **Hong Kong RGC GRF:** Accurate Characterization of Bacteriophages by Integrating Their Properties with Deep Learning Models; HKD 1,165,790 **(Drafting and technical development)**
4. **Hong Kong RGC GRF:** Characterizing Quasispecies of Known and Novel Viruses from Metagenomic Data ; HKD 731,089 **(Drafting and technical development)**

Awards

2024 Research Talent Fellowship in the Hong Kong Research Talent Hub for ITF projects

2023 CHAMPION of Postgraduate Student Research Paper Competition, IEEE (Hong Kong) Computational Intelligence Chapter

2022 Research Tuition Scholarship, City University of Hong Kong

2022 Outstanding Academic Performance Award, City University of Hong Kong

2021 Research student symposium 2021 best paper award **(1st Prize)**, City University of Hong Kong

2021 Research Tuition Scholarship, City University of Hong Kong

2021 Outstanding Academic Performance Award, City University of Hong Kong

2017 The first prize Scholarship, Sun Yat-sen University

Publication (in reverse chronological order)

Co-first authors are underlined; Corresponding author are denoted with *; **IF**: latest impact factor; **Rank**: JCR Impact Factor Rank; IF and Rank are from Clarivate reports of the JCR reports. Conference papers are highlighted with red.

Key Statistics: 15 First-author, 4 Corresponding-author publications; 33 JCR Q1 papers, 6 with IF > 10, 10 CCF-A conference/journal papers. 1,521 Citations; h-index: 16

First or corresponding author

1. Luo, X., Ou-Yang, L., Shen, Y., Guan, J., Cai, D., Zhang, J., Xing, G., Sun, Y., and **Shang, J.,** (2026). "SeekRBP: Leveraging Sequence-Structure Integration with Reinforcement Learning for Receptor-Binding Protein Identification", *Under review*.
2. **Shang, J.,** Peng C., Guan J., Cai, D., Wang, D., and Sun, Y. (2026). "PhaBOX2: an enhanced web server for discovering and analyzing viral contigs in metagenomic data", *Nucleic Acids Research*.
   IF: 13.1; JCR Q1; 13/320 in BIOCHEMISTRY & MOLECULAR BIOLOGY
3. Shen, Y., Shi, K., Yu, C., Zhang, R., Sun, Y. *, and **Shang, J.*** (2026). "PhageMind: Generalized Strain-level Phage Host Range Prediction via Meta-learning", *34th Conference on The Intelligent Systems for Molecular Biology*.
   Top conference in bioinformatic; acceptance rate: 16%; Also appeared in *Bioinformatics* Journal.
4. Luo, X., Ou-Yang, L., Shen, Y., Sun, Y. *, and **Shang, J**. * (2025). "Reinforcement learning-guided sample selection for improved identification of phage receptor binding proteins", *Briefings in Bioinformatics*
   IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY
5. **Shang, J.**, Peng, C., Guan, J., Cai, D., Wang, D., and Sun, Y. (2025). "From genomic signals to prediction tools: a critical feature analysis and rigorous benchmark for phage–host prediction", *Briefings in Bioinformatics*.
   IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY
6. Qu, F., Peng, C., Guan, J., Wang, D., Sun, Y. *, and **Shang, J.***. (2025). "GiantHunter: accurate detection of giant virus in metagenomic data using reinforcement-learning and Monte Carlo tree search", *33rd Conference on The Intelligent Systems for Molecular Biology and the 24th European Conference on Computational Biology*.
   Top conference in bioinformatic; acceptance rate: 17.5%; Also appeared in *Bioinformatics* Journal.
7. **Shang, J.**, Peng, C., Ji, Y., Guan, J., Cai, D., Tang, X., and Sun, Y. (2024). "Accurate and efficient protein embedding using multi-teacher distillation learning", *Bioinformatics*.
   IF: 5.4 JCR Q1; 8/86 in BIOCHEMICAL RESEARCH METHODS
8. Wang, D., **Shang, J.,** Lin, H., Liang, J., Wang, C., Sun, Y., ... and Qu, J. (2024) "Deep learning assisted identification of phage harboring ARGs in the lake replenished by reclaimed water", *Water Research*
   IF: 12.4 JCR Q1; 2/132 in WATER RESOURCES
9. **Shang, J.**, Peng, C., Liao, H., Tang, X., and Sun, Y. (2023) "PhaBOX: A web server for identifying and characterizing phage contigs in metagenomic data." *Bioinformatics Advances*
   IF: 2.8; JCR Q2; 19/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY
10. **Shang, J.**, Peng, C., Tang, X., and Sun, Y. "PhaVIP: Phage VIrion Protein classification based on chaos game representation and Vision Transformer." *31st Conference on The Intelligent Systems for Molecular Biology and the 22nd European Conference on Computational Biology*.
    Top conference in bioinformatic; acceptance rate: 17.9%; Also appeared in *Bioinformatics* Journal.

11. **Shang, J.**, Tang, X., and Sun, Y. (2023) "PhaTYP: predicting the lifestyle for bacteriophages using BERT." *Briefings in Bioinformatics*.
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

12. Jiang, J. Z., Yuan, W. G., **Shang, J.**, Shi, Y. H., Yang, L. L., Liu, M., ..., and Yuan, L. H. (2023) "Virus classification for viral genomic fragments using PhaGCN2." *Briefings in Bioinformatics*.
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

13. **Shang, J.,** and Sun, Y. (2023) "CHERRY: a Computational method for accurate prediction of virus–pRokarYotic interactions using a graph encoder–decoder model." *Briefings in Bioinformatics*.
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

14. Zhu, Y., **Shang, J.,** Peng, C., and Sun, Y. (2022) "Phage family classification under Caudoviricetes: A review of current tools using the latest ICTV classification framework." *Frontiers in Microbiology*.
IF: 4.5; JCR Q1; 38/163 in MICROBIOLOGY

15. **Shang, J.***, Tang, X., Guo, R., and Sun, Y. (2022) "Accurate identification of bacteriophages from metagenomic data using Transformer." *Briefings in Bioinformatics.*
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

16. **Shang, J.*** and Sun, Y. (2022) "Predicting the hosts of prokaryotic viruses using GCN-based semi-supervised learning." *BMC Biology*.
IF: 4.5; JCR Q1; 15/107 in BIOLOGY

17. **Shang, J.**, Jiang, J., and Sun, Y. "Bacteriophage classification for assembled contigs using graph convolutional network." *29th Conference on The Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology*
Top conference in bioinformatic; acceptance rate: 18.6%; Also appeared in *Bioinformatics* Journal.

18. **Shang, J.** and Sun, Y. (2021) "CHEER: Hierarchical taxonomic classification for viral metagenomic data via deep learning." *Methods*.
IF: 4.3 JCR Q1; 15/86 in BIOCHEMICAL RESEARCH METHODS

19. Du, N., **Shang, J.,** and Sun, Y. (2021) "Improving protein domain classification for third-generation sequencing reads using deep learning." *BMC genomics.*
IF: 3.7 JCR Q2; 52/192 in GENETICS & HEREDITY

20. Li, S., **Shang, J**. *, Duan, Z., and Huang, J. (2018). Fast detection method of quick response code based on run-length coding. *IET Image Processing*.
IF: 2.2; JCR Q3; 130/204 in COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE

Co-author

1. Lu, Y., Shi, Guan, J., Shen, Y., **Shang, J.** , and Sun, Y.* (2026). "TPMM: Three-component Posterior Mixture Model Enables Robust Inverton Detection in Low-Depth Metagenomes and Suggests Potential Viral Invertons", 25th European Conference on Computational Biology.
Top conference in bioinformatic; acceptance rate: 19%; Also appeared in *Bioinformatics* Journal.

2. Peng, C., **Shang, J.** , Guan, J.,and Sun, Y.* (2026). " ViralQC: A Tool for Assessing Completeness and Contamination of Predicted Viral Contigs", 25th European Conference on Computational Biology.
Top conference in bioinformatic; acceptance rate: 19%; Also appeared in *Bioinformatics* Journal.

3. Cai, D., **Shang, J.**, Peng, C., He, L., Shi, M., and Sun, Y. (2026). "Fast and Accurate Identification of Emerging Viral Reassortment from Genome Sequences", *Nucleic Acids Research*.
IF: 13.1; JCR Q1; 13/320 in BIOCHEMISTRY & MOLECULAR BIOLOGY

4. Ji, Y., Guan, J., Liao, H., **Shang, J.,** and Sun, Y. (2026). "PlasRAG: comprehensive plasmid characterization and retrieval through sequence-text alignment", *Genome Biology*.
IF: 9.4; JCR Q1; 10/177 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY

5. Zhao, Q., Wang, D., Lin, H., Zhou, T., Zhang, J., **Shang, J.**, ..., and Zhang, J. (2026). "Unraveling the activity of phage-carrying antibiotic resistance genes in constructed wetlands", *Frontiers in Cellular and Infection Microbiology*.
IF: 4.8; JCR Q1; 32/163 in MICROBIOLOGY

6. Peng, L., Yang, F., Zhang, J., **Shang, J.,** Xu, W., Sheng, S., ..., and Yue, Z. (2025). "Ecological Drivers and Functional Roles of Phage communities in the Yangtze River's Freshwater Ecosystems", *Journal of Hazardous Materials*.
IF: 11.3; JCR Q1; 5/83 in ENGINEERING, ENVIRONMENTAL

7. Wen, C., Guan, J., Uea-Anuwong, T., **Shang, J.**, Peng, C., Tang, X., ..., and Li, F. (2025). "Dissecting the gut microbial communities and resistomes of wild rats from different ecological areas in Hong Kong", *Environmental Research*.
IF: 7.7; JCR Q1; 18/421 in PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH

8. Zhang, J., **Shang, J.,** Liu, B., Zhu, D., Li, Q., Yin, L., ..., and Zou, Y. (2024). "Hot spots of resistance: Transit centers as breeding grounds for airborne ARG-carrying bacteriophages ", *Journal of Hazardous Materials*.
IF: 11.3; JCR Q1; 5/83 in ENGINEERING, ENVIRONMENTAL

9. Peng, C., **Shang, J.,** Guan, J., Wang, D., and Sun, Y. (2024). "ViraLM: empowering virus discovery through the genome foundation model ", *Bioinformatics*.
IF: 5.4 JCR Q1; 8/86 in BIOCHEMICAL RESEARCH METHODS

10. Tang, X., **Shang, J.,** Chen, G., Chan, K. H. K., Shi, M., and Sun, Y. (2024). "SegVir: reconstruction of complete segmented RNA viral genomes from metatranscriptomes". *Molecular Biology and Evolution*.
IF: 5.3 JCR Q1; 24/192 in GENETICS & HEREDITY

11. Ji, Y., **Shang, J.**, Guan, J., Zou, W., Liao, H., Tang, X., and Sun, Y. (2024). "PlasGO: enhancing GO-based function prediction for plasmid-encoded proteins based on genetic structure". *GigaScience*.
IF: 3.9 JCR Q1; 21/136 in MULTIDISCIPLINARY SCIENCES

12. Liao, H., **Shang, J.**, and Sun, Y. (2023) "GDmicro: classifying host disease status with GCN and Deep adaptation network based on the human gut microbiome data", *Bioinformatics.*
IF: 5.4 JCR Q1; 8/86 in BIOCHEMICAL RESEARCH METHODS

13. Guan, J., Peng, C., **Shang, J.**, and Sun, Y. (2023) "Genus-level classification of bacteriophages using a Transformer model", *Briefings in Bioinformatics*
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

14. Tang, X., **Shang, J.,** Ji, Y., and Sun, Y. (2023) "PLASMe: a tool to identify PLASMid contigs from short-read assemblies using Transformer." *Nucleic Acids Research*.
IF: 13.1; JCR Q1; 13/320 in BIOCHEMISTRY & MOLECULAR BIOLOGY

15. Ji, Y., **Shang, J.**, Tang, X., and Sun, Y. (2023) "HOTSPOT: Hierarchical hOst predicTion for aSsembled Plasmid cOntigs with Transformer." *Bioinformatics.*
IF: 5.4 JCR Q1; 8/86 in BIOCHEMICAL RESEARCH METHODS

16. Tang, X., **Shang, J.**, and Sun, Y. (2022) "RdRp-based sensitive taxonomic classification of RNA viruses for metagenomic data." *Briefings in Bioinformatics*.
IF: 7.7; JCR Q1; 2/67 in MATHEMATICAL & COMPUTATIONAL BIOLOGY

17. Cai, D., **Shang, J.**, and Sun, Y. (2022) "HaploDMF: viral haplotype reconstruction from long reads via deep matrix factorization." *Bioinformatics*.
IF: 5.4 JCR Q1; 8/86 in BIOCHEMICAL RESEARCH METHODS

Selected Software

Accumulating **770,000+** pageviews, **31,000+** submissions from **1,300+** researchers worldwide on webserver; **12,000+** download and **500+** stars on GitHub; **9,300+** install via Bioconda.

1. **PhaBOX**: a webserver for conducting comprehensive virus analysis.
Availability: https://github.com/KennthShang/PhaBOX/
Web server: https://phage.ee.cityu.edu.hk/

2. **PhaTYP**: predicting the lifestyle for bacteriophages using BERT.
Availability: https://github.com/KennthShang/PhaTYP

3. **PhaGCN**: phage contig classification using Graph neural network.
Availability: https://github.com/KennthShang/PhaGCN

4. **CHERRY:** phage host prediction using graph convolutional encoder and a link prediction decoder.
Availability:https://github.com/KennthShang/CHERRY

Conference and Talks

Conference proceedings:

2026 "PhageMind: Generalized Strain-level Phage Host Range Prediction via Meta-learning ", *34th Conference on The Intelligent Systems for Molecular Biology*, Washington, D.C., United States of America

2025 "GiantHunter: accurate detection of giant virus in metagenomic data using reinforcement-learning and Monte Carlo tree search", *33rd Conference on The Intelligent Systems for Molecular Biology and the 24th European Conference on Computational Biology*, Liverpool, United Kingdom

2023 "PhaVIP: Phage VIrion Protein classification based on chaos game representation and Vision Transformer", *31st Conference on The Intelligent Systems for Molecular Biology and the 22nd European Conference on Computational Biology*, Lyon, France

2021 “Bacteriophage classification for assembled contigs using graph convolutional network”, *29th Conference on The Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology*, VIRTUAL due to the SARS-COV-2

Invited talks:

2026 "PhaBOX2: an enhanced web server for discovering and analyzing viral contigs in metagenomic data", *The 20th Annual Conference on Persistent Organic Pollutants,* Zhuhai, China

2025 "Illuminate dark matters in microbial communities via deep learning", *School of Mathematics*, *South China University of Technology*, Guangzhou, China

2025 "Identification and risk prediction of antibiotic-resistant pathogens in wastewater based on deep learning", *Beijing Normal University*, Zhuhai, China

2022 "Hierarchical taxonomic classification for viral metagenomic data via deep learning", *CIMDA/Oxford Seminar Series*, VIRTUAL due to the SARS-COV-2

2021 **"**Understanding the language of life with AI"**,** *RECOMB satellite conference on Comparative Genomics,* VIRTUAL due to the SARS-COV-2

Service

Guest Editor

Frontiers in Cellular and Infection Microbiology, *Microorganisms*

Editorial Board Member

BMC Bioinformatics 2023-present

Program committee

IEEE International Conference on Bioinformatics and Biomedicine 2026

ISCB member

International Society for Computational Biology, 2021-present

Reviewer

Nature Communications, Microbiome, Genome Biology, RNA, Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Briefings in Bioinformatics, BMC Bioinformatics, BMC Genomics, Cell Report, Scientific Data.

Teaching

Teacher assistant at City University of Hong Kong (2020-2023):

Multi-Dimensional Data Modeling and its Applications (Postgraduate)
Machine Learning for Signal Processing Application (Postgraduate)
Data Structures and Algorithms (Undergraduate)

Teacher assistant at Sun Yat-sen University (2016-2018):

Artificial Intelligence (Undergraduate)
Digital Image Processing (Undergraduate)
Radio Frequency Identification (Undergraduate)

Teacher assistant at Udacity Company (2017-2018):

Deep learning (nanodegree program)
Data scientist (nanodegree program)

Student Advising

Co-advised four PhD students and one visiting student in City University of Hong Kong, resulting in corresponding-author publications on high impact journal and top conference in bioinformatics.

Advising undergraduate students on their Final Year Project at City University of Hong Kong. Topic including:

- ✧ Apply convolutional neural network to find sequence features in genomic data.
- ✧ Graph convolutional neural network for host prediction.
- ✧ Combining RAG and LLM for a reliable Q/A system.

Advising MSc students on their Dissertation at City University of Hong Kong. Topic including:

- ✧ Comparing Large Language Models with Traditional Biology: How do LLMs Understand DNA and Proteins.
- ✧ Benchmarking Protein Language Models for "Dark Matter": Function Prediction in Viral Metagenomics.
- ✧ Unveiling the Hidden Kingdom of Metagenomic Data: Fungi Identification and Characterization

Collaborations (in alphabetical order)

Associate Prof. CHEN Liangjun, *Zhongnan Hospital of Wuhan, Wuhan University*

- ➢ **Multi-omics & Neurodegeneration:** Leading a longitudinal cohort study to investigate the role of bacteriophage-driven gut and oral microbiome remodeling in the onset and progression of Alzheimer's disease **(Ongoing).**
- ➢ **Generative AI & Antibody Design:** Leveraging generative models and AI agents to design novel therapeutic antibodies for hemorrhagic fevers (e.g., SFTSV and Hantaan virus), bridging the gap between AI prediction and clinical application **(Ongoing).**

Associate Prof. WANG Donglin, *School of Environmental Science and Engineering, Shandong University*

- ➢ **AI-Driven Viral Discovery:** Developed deep learning and reinforcement learning frameworks to identify ARG-carrying bacteriophages and giant viruses in the Yangtze River and reclaimed water systems **(3 papers published)**.
- ➢ **Eco-Risk Modeling:** Building data-driven aquatic ecological health assessment framework. Leveraging multimodal models to integrate multi-source environmental data (high-throughput metagenomics and physicochemical sensors) from extensive multi-matrix sampling to forecast pathogen risks in the South-to-North Water Diversion Project **(One manuscript in preparation / Ongoing).**

Prof. SHI Weifeng, *School of Life Sciences and Biotechnology , Shanghai Jiao Tong University*

- ➢ **Deep-Sea Metagenomics:** Conducting a large-scale viral survey of 1,300+ Mariana Trench samples to uncover novel bacteriophage diversity and unique genomic adaptations to extreme hadal environments **(One manuscript in preparation).**

Associate Prof. YUE Zhengfu, *Chinese Academy of Tropical Agricultural Sciences*

- ➢ **Environmental Viromics & ARG Tracking:** Characterized the ecological drivers of phage communities and the dissemination of airborne ARG-carrying bacteriophages across diverse ecosystems, including the Yangtze River and transit hubs **(2 publications)**.

Prof. ZHANG Rui, *Institute for Advanced Study, Shenzhen University*

- ➢ **Precision Phage Therapy:** Developed strain-level host range prediction models using deep learning to accelerate the selection of therapeutic phages against multidimensional bacterial pathogens **(one publication and one manuscript in preparation)**.