BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Yue Wu

eRA COMMONS USER NAME (c`redential, e.g., agency login): WU.YUE

POSITION TITLE: Postdoctoral Fellow

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Start Date MM/YYYY	Completion Date MM/YYYY	FIELD OF STUDY
Nanjing University	BS	09/2012	07/2016	Biology
University of Georgia	PHD	09/2017	07/2022	Bioinformatics
Stanford University	Postdoctoral Fellow	08/2022	Present	Multi-omics/Time series

A. Personal Statement

My research and training experience prepare me for interdisciplinary research regarding biological time-series data in biochemistry, metabolomics, and cohort health data. My long-term goal is to become an independent investigator, studying the time-series profile of human health, particularly in method development and stratification of metabolic diseases.

Before graduate school, I carried out research on diverse topics, ranging from plant evolution and epigenomics to data science in cancer. As an undergraduate, my research focused on computational methods for phylogenetics and omics profiles, which leads to multiple publications on systems of humans, plants, fungi, and archaea (Dr. Zhu-Qing Shao and Dr. Jian-Qun Chen) (4). I then worked on The Cancer Genome Atlas (TCGA) data, integrated omics, and modeled cancer metabolism (Dr. Ying Xu, Dr. Chi Zhang, and Dr. Sha Cao). During this period, I found metabolomics and modeling metabolic time series an interesting direction, which is both technically approachable and full of unanswered questions.

The University of Georgia provided the natural next step in my career, where flexible collaboration and interdisciplinary interactions happened. During graduate school time, I was co-supervised by Dr. Arthur Edison, a world leader in metabolomics and nuclear magnetic resonance (NMR), and Dr. Jonathan Arnold, a real all-arounder leader in computational biology. With their mentoring, knowledge, and support, I built an interdisciplinary system biology project, measuring, analyzing, and modeling the metabolic time series of *Neurospora crassa*. I contributed to a novel experimental NMR approach for time-series metabolic profiling (3), produced a computational method for feature extraction (2), and built multiple machine learning systems for automatic microscopic image annotation. We perturbed the metabolic network of *Neurospora* with different carbon sources, and I uncovered regulation related to carbon usage through network modeling and dimensionality reduction (1). I also built a new package to automatically deconvolute NMR data and partly solved a longstanding problem in metabolomics through innovation in optimization approaches (Dr. Frank Delaglio and Dr. Arthur Edison). During my graduate career, I organized a local symposium, volunteered with science organizations, mentored a few graduate students, and taught an omics class.

For my postdoctoral training, I will continue my topics on biological time-series data, transit into chronic diseases, and model the precise profile of human health through time-series multi-omics. My sponsor Dr. Michael P. Snyder is a world leader in multi-omics, cohort study, precision medicine, and system biology, and has an extensive record of training successful postdoctoral fellows. His lab has strong support in study design, nutrient analysis, secure data pipeline, and experimental and computational processes for omics data.

The proposed research will combine my background in time-series modeling and the resources of multi-omics and cohort study in Snyder Lab, which prepares me for the transition to applying dynamic profiling to human health problems. In addition, the proposed training plan will boost my career development,

particularly in grant writing, large collaboration project management, and student mentoring. My choice of the sponsor, research project, and training will give me a solid foundation for my goal of studying dynamics in health data as an independent investigator.

1. **Wu Y**, Judge MT, Edison AS, Arnold J. Uncovering in vivo biochemical patterns from time-series metabolic dynamics. PLoS One. 2022;17(5):e0268394.
2. **Wu Y**, Judge MT, Arnold J, Bhandarkar SM, Edison AS. RTExtract: time-series NMR spectra quantification based on 3D surface ridge tracking. Bioinformatics. 2020;36(20):5068-75.
3. Judge MT, **Wu Y**, Tayyari F, Hattori A, Glushka J, Ito T, et al. Continuous in vivo Metabolism by NMR. Front Mol Biosci. 2019;6:26.
4. **Wu Y**, Wu P, Wang B, Shao ZQ. Genome-Wide Analysis Reveals Ancestral Lack of Seventeen Different tRNAs and Clade-Specific Loss of tRNA-CNNs in Archaea. Front Microbiol. 2018;9:1245.

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2022 – Present	Postdoctoral Researcher, Stanford University
2017 – 2022	Graduate Research Assistant, UGA
2016 – 2017	Intern, Computational Systems Biology Lab, UGA
2017 – Present	Member, Metabolomics Association of North America
2022 – Present	Member, American Heart Association
2021	Mentor, Anti-racism in the Life Sciences, UGA
2021	Mentor, Research Experiences for Undergraduates (REU), UGA
2019	Organizing committee, Machine learning Mini-symposium, UGA
2019	Teaching Assistant, Bioinformatics and Omics, UGA

Honors

2020	XSEDE Startup grant (Manager)
2019	Graduate school travel grant, UGA
2019	PEARC19 student program travel grant
2016	Outstanding graduate with honor, Nanjing University
2014 – 2015	First-class People's Scholarship
2013 – 2014	Silver Medal, iGEM (International Genetically Engineered Machine) competition

C. Contributions to Science

1. **Knowledge extraction from time-series metabolic systems.** Recent developments in omics approaches provide a comprehensive view of the biological system at one time point. However, the understanding of the dynamic response to environmental perturbation is still limited in both data collection and computational analysis. I contributed to an NMR approach to collecting time-series metabolic data (a). I then designed the computational method to efficiently extract chemical information from the high-dimensional heavy dataset (b). This provides rich information regarding dynamic metabolic processes under different environments. I uncovered biological regulation in carbon metabolism and glycogen utilization from this high-dimensional time series, through modeling and time-series analysis (c). I built a new efficient workflow to understand metabolic dynamics and regulation, which can be expanded to other fermentation systems and the study of metabolic disease in humans.
    a. **Wu Y**, Judge MT, Edison AS, Arnold J. Uncovering in vivo biochemical patterns from time-series metabolic dynamics. PLoS One. 2022;17(5):e0268394.
    b. **Wu Y**, Judge MT, Arnold J, Bhandarkar SM, Edison AS. RTExtract: time-series NMR spectra quantification based on 3D surface ridge tracking. Bioinformatics. 2020;36(20):5068-75.
    c. Judge MT, **Wu Y**, Tayyari F, Hattori A, Glushka J, Ito T, et al. Continuous in vivo Metabolism by NMR. Front Mol Biosci. 2019;6:26.
    d. Cao S, Zhou Y, **Wu Y**, Song T, Alsaihati B, Xu Y. Transcription regulation by DNA methylation under stressful conditions in human cancer. Quantitative Biology. 2017 November; 5(4):328-337. doi: 10.1007/s40484-017-0129-y.
2. **Automation in phenotyping biological systems.** New experimental approaches (e.g., microscopic devices) enable the recording of thousands of samples in a short time, which greatly promotes the phenotyping of plants, fungi, and human tissues. However, image annotation and information extraction

are still manual intensive. I built multiple frameworks to classify phenotypes (e and f) through ResNet in PyTorch, associate with genomic information, and uncover important structures through feature importance evaluation. I also built image segmentation programs through Detectron2 to annotate different symbiosis structures of Arbuscular mycorrhiza and worm population. Automation in phenotyping greatly expands the sample size in association studies and can be used for medical diagnosis.

   e. Krach EK, **Wu Y**, Skaro M, Mao L, Arnold J. Wild Isolates of Neurospora crassa Reveal Three Conidiophore Architectural Phenotypes. Microorganisms. 2020;8(11).
   f. Krach EK, Skaro M, **Wu Y**, Arnold J. Characterizing the gene–environment interaction underlying natural morphological variation in Neurospora crassa conidiophores using high-throughput phenomics and transcriptomics. G3 Genes|Genomes|Genetics. 2022;12(4).

3. **Phylogenetics analysis of gene family and disease resistance.** Through the years, I contributed to computational methods in phylogenetics, evolution, and omics data. I analyzed genomics (i) and epigenomics (h) in plants and discovered molecular mechanisms related to symbiosis and disease resistance. I then used a similar phylogenetic approach to study tRNA evolution and metabolism in Archaea (g). I obtained a broad understanding and technical capabilities across multiple fields and omics. The evolution patterns of gene families can reveal mechanisms in crop disease resistance.

   g. **Wu Y**, Wu P, Wang B, Shao ZQ. Genome-Wide Analysis Reveals Ancestral Lack of Seventeen Different tRNAs and Clade-Specific Loss of tRNA-CNNs in Archaea. Front Microbiol. 2018;9:1245.
   h. Wu P, **Wu Y**, Liu CC, Liu LW, Ma FF, Wu XY, et al. Identification of Arbuscular Mycorrhiza (AM)-Responsive microRNAs in Tomato. Front Plant Sci. 2016;7:429.
   i. Shao ZQ, Xue JY, Wu P, Zhang YM, **Wu Y**, Hang YY, et al. Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. Plant Physiol. 2016;170(4):2095-109.

Complete List of Published Work in My Bibliography:
https://www.ncbi.nlm.nih.gov/myncbi/1zo_dHxCPX0kb/bibliography/public/

D. Scholastic Performance

YEAR	COURSE TITLE	GRADE
	UNIVERSITY OF GEORGIA	
2017	Responsible Conduct Research	A
2018	Stat Inference for Life Sci	A
2018	Systems Biology	A
2018	Mathematical Biology	A
2018	Bioinformatics Algorithms	A
2019	Applied Time Series	A
2018 - 2022	Binf Sem	S
2018 - 2022	Curr Topics Resch	S

In the University of Georgia, the grade S (U) indicates satisfactory (unsatisfactory) participation in certain required courses.