Metagenomics uses sequencing technologies to study genetic sequences from the whole microbial communities. Binning metagenomic reads is the most fundamental step in metagenomic studies, which is essential for the understanding of the microbial functions, compositions, and interactions in the environment samples. Various taxonomy-dependent and taxonomy-independent approaches have been developed based on the information such as sequence similarity, sequence composition, or k-mer frequency. However, there is still room for improvement, and it is still challenging to bin reads from species with similar or low abundance or reads from unknown species.

In this dissertation, we introduce one taxonomy-independent and three taxonomy-dependent approaches to improve the performance of metagenomic reads binning. The taxonomy-independent method called MBBC, bins reads by considering k-mer frequency in reads without reference genomes. The first two taxonomy-dependent approaches both bin reads by measuring the similarity of reads to the trained Markov Chains from different taxa. The major difference between these two methods is that the first one selects the potential taxa with the taxonomical decision tree, while the second one, called MBMC, selects the potential taxa with the ordinary least squares (OLS) method. The third taxonomy-dependent approach bins reads by combining the methods of MBMC with clustering of Markov Chains trained from assembled reads. By testing on both simulated and real datasets, these tools showed superior or comparable performance with various state of the art methods. We anticipate that our tools can significantly improve the accuracy of metagenomic reads binning and thus be widely applied in the real environmental samples.

Educational Career:
Bachelor's of Electrical Engineering, BS, 2007, HeFei University of Technology
Master's of Electrical Engineering, MS, 2010, University of Science and Technology of China
Master's of Electrical Engineering, MS, 2012, University of Central Florida

Committee in Charge:
Haiyan Hu, Chair, Department of Computer Science
Xiaoman Li, Co-Chair, Department of Computer Science
Shaojie Zhang, Department of Computer Science
Anna Savage, Department of Biology
Annie Wu, Department of Computer Science

Approved for distribution by Haiyan Hu, Committee Chair, on August 23, 2016.

The public is welcome to attend.