Research Statement

Yue Hao

My research focuses on modeling the pattern of unbalanced gene loss after whole-genome duplication events, finding the causes of biased fractionation and the evolutionary transitions influenced by gene and genome duplications. Polyploidy is increasingly seen as an important route to evolutionary innovation, and I am interested in better understanding post-polyploidy functional evolution, and of the nature of the functional links and the degree of co-evolution inherent in the interacting macromolecules that make up the cell. By incorporating evolutionary models and statistical inferences, my goal is to synthesize biological insights and develop computational tools for hypothesis generation and predictive biology.

One of my dissertation chapters describes modeling of the ancient whole genome triplication shared by the tribe Brassiceae, which contains many economically valuable crops such as cabbage, collard greens, kale, broccoli, cauliflower, Chinese cabbage, turnip and mustard. Plants in this tribe demonstrate high morphological diversity within each species. Thus, it is a fascinating system to study the effect of polyploidy during evolution. I used a phylogenomic pipeline called POInT (Polyploidy Orthology Inference Tool) to trace the history of gene loss across a set of modern genomes that all descend from the same polyploid ancestor and phylogenetically model the resolution of the Brassica triplication, confirming that the triplication occurred in two distinct steps and resulted in modern species with three different subgenomes. With the high resolution subgenome assignment, I then looked at the function of genes from different subgenomes. Better understanding of how genes interact and which gene copies are lost after hybridization could help us improve breeding programs for crops.

I am also interested in using comparative genomics to unravel the genetic changes underlying major evolutionary transitions, such as the origin of mammalian pregnancy. Using a comparative genomics pipeline ORIS (ORthology Inference using Synteny), I inferred orthology relations between human genes and genes from each of 43 other vertebrate genomes and reconstruct the ancestral states of these orthologs. orthologs appeared along with the eutherian mammal origin are associated with functions such as transcription regulation, keratinization and immune response. However, most of the genetic toolkit of mammalian pregnancy involved the repurposing of many genes that had pre-existing biological functions. Genes that play a role in the placental mammal origin are not only an evolutionary point of interest but also could serve as novel drug targets for reproductive and/or immune diseases.

During my PhD, I also worked on multiple collaborative projects and gained experience in fields other than comparative genomics. For example, in a metagenomics project, we examined the effect of diet shift on sheep gut microbiome at the metabolic level. In another project, we analyzed the RNA-seq data from mice brain tissues and found differentially expressed genes that could explain the behavioral changes between wild island mice and laboratory mice.

In the future, I plan to explore the population-level variations for organisms that experienced ancient polyploidy and to further understand the effect of ancient whole genome duplication/triplication on species diversity, adaptation and domestication.