Unlike message RNAs (mRNAs) whose information is encoded in the primary sequences, the cellular roles of non-coding RNAs (ncRNAs) originate from the structures. Therefore studying the structural conservation in ncRNAs is important to yield in-depth understanding of their functionalities. In the past years, many computational methods have been proposed to analyze the common structural patterns in ncRNAs using comparative methods. However, the RNA structural comparison is not a trivial task, and the existing approaches still have numerous issues in efficiency and accuracy. In this dissertation, we will introduce a suite of novel computational tools that extend the classic models for ncRNA secondary and tertiary structure comparisons.

For RNA secondary structure analysis, we first developed a computational tool, named PhyloRNAalifold, to integrate the phylogenetic information into the consensus structural folding. The underlying idea of this algorithm is that the importance of a co-varying mutation should be determined by its position on the phylogenetic tree. By assigning high scores to the critical covariances, the prediction of RNA secondary structure can be more accurate. Besides structure prediction, we also developed a computational tool, named ProbeAlign, to improve the efficiency of genome-wide ncRNA screening by using high-throughput RNA structural probing data. It treats the chemical reactivities embedded in the probing information as pairing attributes of the searching targets. This approach can avoid the time-consuming base pair matching in the secondary structure alignment. The application of ProbeAlign to the FragSeq datasets shows its capability of genome-wide ncRNAs analysis.

For RNA tertiary structure analysis, we first developed a computational tool, named STAR3D, to find the global conservation in RNA 3D structures. STAR3D aims at finding the consensus of stacks by using 2D topology and 3D geometry together. Then, the loop regions can be ordered and aligned according to their relative positions in the consensus. This stack-guided alignment method adopts the divide-and-conquer strategy into RNA 3D structural alignment, which have improved its efficiency dramatically. Furthermore, we also have clustered all loop regions in non-redundant RNA 3D structures to de novo detect plausible RNA structural motifs. The computational pipeline, named RNAMSC, was extended to handle large-scale PDB datasets, and solid downstream analysis was performed to ensure the clustering results are valid and easily to be applied to further research. The final results contain many interesting variations of known motifs, such as GNAA tetraloop, kink-turn, sarcin-ricin and t-loop. We also discovered novel functional motifs that conserved in a wide range of ncRNAs, including ribosomal RNA, sgRNA, SRP RNA, GlnS riboswitch and twister ribozyme.

Major: Computer Engineering

Educational Career:
Bachelor's of Civil Engineering, BS, 2002, Wuhan University
Master's of Computer Science, MS, 2005, Huazhong University of Science and Technology
Master's of Computer Engineering, MS, 2010, University of Central Florida

Committee in Charge:
Shaogiezhang, Chair, Computer Science
Ratan Guha, Department of Computer Science
Sumit Jha, Department of Computer Science
Kenneth Stanley, Department of Computer Science
Hojun Song, Texas A&M University