* I have been working on analyzing RNA-seq data of soybean developing embryos for the past 3 years. I have developed a computational pipeline for analysis of this RNA-seq data that has a potential to be applied to any species with known reference genome. I have developed and used several algorithms and tools for mining of this time-course RNA-seq data. In addition to finding differentially expressed genes, I have analyzed differentially expressed splice variants of known genes leading to investigation of alternative splicing events such as exon skipping, intron retention, and antisense transcripts. We have further analyzed these splice variants regarding to their domain composition, coding potential, signaling and metabolic pathways, and co-expression network.
* I am interested in the function of long noncoding RNAs (lncRNAs. lncRNAs’ functions have been implicated in many diseases such as cancer. For this purpose, I have developed an SVM classifier with over 98% accuracy that can classify transcripts into coding and noncoding classes with integration of several features related to sequence and structure of a transcript. This SVM classifier is currently trained, tested, and applied on Arabidopsis and soybean transcripts. I am going to extend this classifier to human, mouse, zebrafish, and Drosophila.