MONDAY, November 16, 2015
3:30pm
DBI Room 102

fRNAkenseq: a Web-based RNA Sequencing Analysis Application Developed with the iPlant Collaborative

Allen Hubbard
PhD Student; Bioinformatics and Systems Biology

ABSTRACT: Over the past two years, to deal with the challenges of working with large scale high-throughput sequencing, we have developed computational tools in the form of automated workflows with user interfaces that store and move data within the iPlant computing environment. iPlant is a major cyberinfrastructure project, dedicated to providing computational resources to the research community. One of our applications hosted by iPlant, fRNAkenseq, combines multiple RNA-seq algorithms into a complete fastq to enrichment workflow and allows access to 20,000 genomes compiled by CoGE, another powered-by-iPlant resource focused on comparative genomics. CoGE also allows users to upload personal draft genomes, accessible by fRNAkenseq. Through applications such as fRNAkenseq, we are interested in further advancing the utility of the iPlant computing environment to the animal genomics community. Developing this infrastructure will allow rapid processing of sequencing data for large-scale experiments that will provide us with the power to continue to probe the nature of transcriptome response to stress at the cell and organismal level.

An Automated Thermo-align Approach for Identifying Template Specific Primers and Hybridization Oligos in (repetitive) Genomes

Felix Francis
PhD Student; Bioinformatics and Systems Biology

ABSTRACT: Re-sequencing of target sites in a genome is a widely used technique in modern genomics. Identifying good quality and template specific primers and hybridization oligos, is critical to the success of these applications. A combined thermodynamic and sequence alignment based computational pipeline for automating the identification of template specific primers/hybridization oligos has been developed, which is particularly relevant for studies on organisms with repetitive genomes. During the initial phase of the pipeline, all possible primers of user-defined ranges in size, GC content and melting temperature are extracted for a given genomic locus. An enhanced blast algorithm is then used to evaluate the uniqueness of each candidate sequence with respect to the reference genome. The algorithm uses an end-filling approach to obtain full-length alignments in the presence of mismatches, to allow for exhaustive evaluation of potential mispriming/hybridization sites. A nearest neighbor thermodynamics model is used to compute energy metrics corresponding to DNA duplex hybridization and secondary structure formation under user-defined reaction conditions. In this talk, I will introduce challenges to amplifying/resequencing target regions in repetitive genomes. Results will be presented from testing the pipeline for target-specific amplification in maize, which has a highly repetitive genome. This automated thermo-align tool is expected to facilitate identification of template specific primers and hybridization oligos for a variety of applications in genomics.