SDSU
presents
a thesis defense for
Master of Science
degree in
Computer Science

Tuesday,
December 8, 2015

3:00pm
GMCS 418

SAN DIEGO STATE
UNIVERSITY
Leadership Starts Here

Julian Raheema

Contig Clustering of Metagenomics (CCOM):
A Tool That Generates Population Genomes (Bins) To Analyze
And Capture Uncultured Genomes

Abstract

Recently, a great quantity of academic research and energy has been invested in the study of metagenomics. It helps us to read the DNA code extracted from microorganisms that can't be cultivated in the laboratory. It helps us to explore those microbial communities that play a big role for our ecological system. These DNA samples are collected primarily in two different ways: either spatially or temporality separated. The DNA is sent to a laboratory for DNA extraction and sequencing. The typical metagenomics workflow starts with collecting samples from the environment and filtering them for purity. The samples are sequenced assemble them using assembler software, and finally input these samples into binning tools to extract the population genome from the collected samples. Contig Clustering of Metagenomics (CCOM) is a web-based tool that analyzes population genomes using three binning tools. It takes one contigs (assembled from reads files) and at least three reads files to generate the genome population. MetaBAT and GroopM generate the population genomes, and the third tool crAss returns the destination matrix and plots for the inputs. Using my new web-based tool show that MetaBAT is faster than GroopM and crAss, but GroopM generates more bins than MetaBAT. After the Contig Clustering of Metagenomics finish running, it returns to the user the results as a downloadable link with execution time for each tool.

Thesis Committee

Robert Edwards, Thesis Chair, Department of Computer Science
Roger Whitney, Department of Computer Science
Elizabeth Dinsdale, Department of Biology