UC San Diego | Computer Science & Engineering Colloquium and Distinguished Lecture Series | Fall 2015

Mihai Pop

University of Maryland

Monday, September 21
11:00am-12:00pm
CSE 1202, EBU-3B

Metagenomic assembly: the devil is in the details

Metagenomic studies are providing a glimpse at the genomes of the many micro-organisms that inhabit our bodies and our environment. Like in the case of isolate genomes, together the individual sequencing reads can help reduce redundancy and provide downstream analyses will longer segments of DNA. While the assembly of single organisms is well studied and understood theoretically, and a number of highly effective assembly tools exist, metagenomic data pose as of yet not fully addressed computational challenges. Metagenomic data are difficult to assemble due to the wide variance in the abundance of organisms in a sample, as well as due to strain-level differences that distinguish closely related organisms. A number of assembly packages for metagenomic assembly have been developed in recent years, however none has fully solved the metagenomic problem. In my talk I will argue that at least part of the challenge arises from our poor understanding of the needs of biologists end-users. A different assembly strategy is appropriate if one wants to assemble all the genes in a community (but not entire organisms), or if one attempts to extract and completely reconstruct a single organism within a mixture, or if the goal is to track genomic changes of one or more organisms across time. I will also describe our initial progress in developing reference-independent validation approaches for metagenomic assemblies.

Biography

Dr. Pop is an associate professor in the Department of Computer Science and the Center for Bioinformatics and Computational Biology at the University of Maryland, College Park (UMCP). Dr. Pop received his Ph.D. in Computer Science at Johns Hopkins University where he focused on algorithms for computer graphics and Geographic Information Systems (GIS) applications. He then joined The Institute for Genomic Research (TIGR) as a Bioinformatics Scientist, where he was responsible for the development of genome assembly algorithms. During this time, Dr. Pop participated in a number of bacterial and eukaryotic genome projects including important human pathogens such as Bacillus anthracis and Entamoeba hystolitica. Since joining the University of Maryland, Dr. Pop has continued to develop novel approaches for genome assembly and analysis, and has developed extensive expertise in the analysis of metagenomic data. His lab has developed a number of widely used open-source software tools, such as the assembly suite AMOS, the NGS aligner Bowtie, the taxonomic classifier Metaphyler, and the metagenomic assembly package MetAMOS. Most recently he co-led the data analysis working group for the Human Microbiome Project and led the sub-group responsible for the assembly of the data generated in this project.