University at Buffalo The State University of New York

COMPUTER SCIENCE AND ENGINEERING

Presents

Ben Langmead, Johns Hopkins University

Scalable software for analyzing large collections of DNA sequencing data

Second-generation DNA sequencers are improving rapidly and are now capable of sequencing hundreds of billions of nucleotides of data in about a week for a few thousand dollars. Consequently, archives like the Sequence Read Archive are filling up with petabytes of publicly-available DNA sequencing data, including valuable data from humans with cancer and rare genetic diseases. These datasets are extremely valuable to the life science community, but they are also very hard to use, especially for life scientists with limited access to computational resources.

I will discuss ways in which my laboratory has sought to make second-generation sequencing datasets (and collections thereof) more usable for everyday biological researchers. In particular, I will discuss recent progress toward software tools (called Myrna and Rail) that can analyze many samples worth of mRNA sequencing data at a time, applying a uniform analysis method across all samples. I also discuss how these tools can take advantage of commercial cloud computing services such as Amazon Web Services to perform these computations in a way that (a) doesn't require the user to commit any of their own computing resources, (b) costs just dollars per sample, (c) scales well (i.e. runs quickly) to many hundreds of samples at once.

Bio: Ben Langmead is an Assistant Professor in the Department of Computer Science at Johns Hopkins University. He earned his Ph.D. in Computer Science from the University of Maryland in 2012. His group seeks to make high-throughput biological datasets easy for biomedical researchers to use. The group studies and applies ideas from sequence alignment, text indexing, statistics and parallel programming. His group has produced several high-impact software tools (e.g. Bowtie, Bowtie 2) addressing common genomics research questions. His paper describing Bowtie won the Genome Biology award for outstanding paper in 2009. His group has also released scalable software tools (e.g. Crossbow, Myrna) that use the MapReduce parallel programming model and commercial cloud computing services to analyze large collections of sequencing data. He is the recipient of a Sloan Research Fellowship (2014) and a National Science Foundation CAREER award (2014).

Monday, November 24, 2014

2:00 – 3:00 PM

University at Buffalo – North Campus – Davis 113A

This talk is free and open to the public

For more information, email cse-dept@cse.buffalo or contact (716) 645-3180