Abstract:
A crucial problem in genome assembly is the discovery and correction of mis-assembly errors in draft genomes. We developed a method that will enhance the quality of draft genomes by identifying mis-assembly errors using paired short read sequence data and optical mapping data. An important aspect of this work is that it highlights the need to use another source of information in order to identify mis-assembly errors, and that optical mapping data can be used as this source of information when it is combined with proper analysis methods. Hence, in the first half of this talk we will present a method for aligning assembled contigs to a genome wide optical map. This is the only non-proprietary method capable of completing the alignment for the budgerigar genome in a reasonable amount of time at the current moment.

In the second half of this talk, we present the details of our method that detects and corrects misassembly errors. Our results show we are able to correct 60% of extensively mis-assembled contigs in a draft of the Francisella tularensis genome, and between 31% and 100% of extensively mis-assembled contigs in various draft genomes of loblolly pine. The tools developed as this work will be applicable in any scientific study that uses the comparison of draft genomes to draw conclusions as to the causal relationships of genetic variation and biological response could benefit from our research. For example, researchers studying cancer, autism, schizophrenia, Parkinson’s disease, and many other complex, multifactorial traits and diseases that require the use and comparison of high-quality draft genomes.

Bio:
Christina Boucher is an Assistant Professor in the Department of Computer Science at Colorado State University (CSU). She has co-authored more than 30 publications in peer-reviewed journals and conferences, including high-profile venues such as ESA 2015, ISMB 2015, and RECOMB 2013. She is also the recipient of a Best Paper Award at ESA 2015, and a co-PI on a $2,200,000 USDA NIFA grant.

Prior to joining CSU in 2012, Dr. Boucher was a post-doctorate researcher under the supervision of Pavel Pevzner at the University of California, San Diego. During this time, she was one of the primary developers of SEQuel, a method for correcting micro-errors, which are defined as small insertions and deletions, and substitution errors (ISMB 2012). She received her doctorate in computer science from the University of Waterloo in Canada in 2010 under the supervision of Dr. Ming Li. She received several research awards as a doctorate candidate, including a Google Anita Borg Scholarship, a NSERC Post-Graduate Scholarship, the President's Graduate Scholarship, and a NSERC Postdoctoral Fellowship.