



Utah Chapter North Regional Meeting

**Thursday September 8, 2011
Utah State University – Logan Campus
ENGR 203, 5:30 P.M.**

[This meeting also serves as the colloquium for the Department of Mathematics and Statistics, with **refreshments and mingling** beginning at 4:30 in Lund Hall.]

Speaker: Hongmei Jiang
Department of Statistics
Northwestern University

Title: Statistical and computational challenges for metagenomics analysis based on next-generation sequencing data



Abstract: Next-generation sequencing technologies greatly promote the field of metagenomics which studies multiple genomes recovered directly from an environment, without the need of culturing them. Based on the short reads sequenced from a metagenomic sample, we would like to identify the multiple species or genomes contained in the sample and to estimate their relative abundance. One widely used approach is to employ sequence homology by aligning sequence reads to known reference sequence databases using a comparison program such as BLAST and assigning the reads to the taxa of the taxonomy tree based on the best match or multiple high-scoring hits. Due to homogeneity of DNA sequences, low coverage sequencing, and large volume of experimental data, estimating the relative abundance of multiple genomes accurately becomes very challenging. Here, we propose a mixture model to estimate the relative abundance of the species and to assign the reads in a global framework. The method is comprehensively tested on simulated metagenomic data and is able to accurately estimate the relative abundance of the genomes. We also apply the proposed method on several metagenomic real datasets. The current statistical and computational methods that are being developed to analyze the metagenomics data and the challenges will be highlighted in the talk.