

## **Statistical Methods for finding conserved patterns in DNA sequences**

Several statistical methods have been developed that use sequence comparison to identify non-coding regions in functional DNA. Hidden Markov Models, originally developed for use in speech recognition, is a statistical method used to analyze DNA and protein sequences. They take into consideration all possible combinations of matches, mismatches and gaps to generate an alignment of a set of sequences. Once an alignment is made, the model can be used to search a sequence database for additional sequences that share the same variation. This method of aligning sequences can be used to find motifs or conserved patterns in DNA and protein sequences. Research this summer focuses on incorporating hidden Markov models into a Motif Discovery Toolkit that is being developed at Oak Ridge National Laboratory. The toolkit contains tools for a large number of computational tasks. The incorporation of hidden Markov models into the toolkit will enable a quick search for similar sequences in long DNA sequences, perhaps an entire genome. Biologists will, then, have a more efficient means of finding transcription factor binding sites and other genomic signals that control gene expression.

Student's Name:

School Student Attends:

Name of Mentor:

Division:

Program:

**Ja'Nera Mitchom**

Fisk University

**Dr. Mike Leuze**

Computer Science and Mathematics

Research Alliance for Minorities