The Department of Computer Science and Engineering

University of Nevada, Reno

cordially invites you to a Master's colloquium

A Supervised Strain Classifier

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
with a major in Computer Science.

by

Adrienne E. Breland

Abstract: Several bacterial and viral species are human pathogens and contain strains exhibiting different degrees of virulence. Nucleic acid sequencing enables strain fingerprinting, which is a term used for identifying bacterial and viral strain species and subtypes based on their DNA. Strain fingerprinting methods are becoming increasingly important in the threat of epidemic outbreaks and the possibility of biothreat agents. This thesis examines the use of oligonucleotide word signatures for strain fingerprinting and related classifications. An investigation into word signature differences exhibited by different strains of the same subtype reveals that words not expressed by individual genomes offer the most potential as differentiating features. Thus, a supervised classifier is built with feature sets derived from absent words. Resulting accuracies are high and are listed for five classifications at different levels of phylogenetic resolution: Mixed Pathogens: 100%, Influenza A virus/Influenza B virus:100%, Influenza A virus subtypes (human host):92%, Avian Influenza A virus H5N1 lineages: 94%, Avian to Human Transmission H5N1 lineages:100%. While the data set used does not allow complete confirmation of reported accuracies, it is suggested that this method could be a valuable tool in comparative genomics and enable geographic origination determination of Influenza A virus and other pathogenic isolates.

10:00 am, Friday, March 14, 2008

Scrugham Engineering and Mines (SEM) room 201

For more information contact Dr. Fred Harris @ 784-6571 (Fred.Harris@cse.unr.edu)