Robert Kuzoff – Medical Bioinformatics. Undergraduate researchers in our lab explore medically significant variation in human genomes, transcriptomes, and proteomes. Additionally, we study variation in viral and bacterial pathogens and its consequences for human health.

Below are two examples of projects that have recently been investigated by members of our lab.

Example Project #1 – Mining the Human Genome for Disease Markers

One of the most effective tools to learn about medically significant variation in humans is Genome Wide Association Studies (GWA). Common variants identified through GWAs are analyzed to determine whether they are associated with certain diseases. Progress made through ongoing GWAs has sparked a personalized health care revolution, which entails the customization of medical treatment and evaluation to an individual's genome. A consortium of researchers has begun the 1000 Genomes Project, which samples genetic variation from diverse demographic groups around the world. My research will sample diverse sequences from the 1000 Genomes project for loci known to be correlated with complex diseases to determine the frequency of disease-associated alleles in demographic groups around the world.

Single nucleotide polymorphisms (SNPs) are markers of genetic diversity that are assessed in GWA studies. When SNPs are located in close proximity, they may form linkage disequilibrium blocks (LBDs) that are inherited as a unit. Because sequencing an entire genome is expensive, "proxy SNPs" at variable sites are used by researchers to infer the genotype of a LDB without sequencing it. Statistical analyses then test whether identified proxy-SNPs are associated with complex disease traits. This method has been used to identify alleles associated with diseases such as Early-Onset Alzheimer's, Late-Onset Alzheimer's, Autism, breast cancer, hypertension, obesity, type 2 diabetes, leukemia, melanoma, and Crohn's Disease. My research project will employ a SNP-based survey to assess the frequency of alleles associated with disease in genomes sampled through the 1000 Genomes Project.

Example Project #2 – Comparative Genomics of Influenza

The genome of influenza comprises eight RNA segments, which are periodically subject to horizontal transfer. Accordingly, some RNA segments may share extensive portions of their evolutionary histories, while others share relatively little. RNA-segment re-assortment events complicate efforts to reconstruct phylogenetic relationships among viral strains. In cases where influenza-RNA segments do share a common history, joint analysis of their sequences has the potential to improve the accuracy of phylogeny estimation. However, until recently it has been logistically difficult to identify cases in which influenza-RNA segments do share common histories.

Here we apply a recently developed program, CladeRunner, to this problem and attempt to identify partitions of the influenza B genome that share a common history and are amenable to joint analysis. We used CladeRunner to analyze sequences of all RNA segments from 253 complete genomes of influenza B for evidence of historical

discordance. Our results identified several cases of disparate evolutionary histories and were consistent with those obtained using the well-established incongruence length difference test. Portions of the viral genomes that showed no evidence of historical discordance were analyzed phylogenetically, both in individually and in combination. Phylogenetic trees resulting from analyses of individual and combined partitions were wholly concordant, but the later exhibited substantially greater resolution and lineages within them garnered appreciably higher statistical support. These results suggest that CladeRunner is a useful bioinformatic tool with which to distinguish genome partitions that do share a common evolutionary history. Hence, CladeRunner may be an effective tool in the quest to understand the evolutionary dynamics of influenza B and other dangerous pathogens.