**Bayesian Statistics Research During COVID Times**

**Max Thrush Hukill, Class of 2021**

The novel coronavirus pandemic has caused all of us to adapt, and our work this summer has been no exception. The main thrust of our work this summer has included three projects: the analysis of a behavioral assay describing cricket responses to auditory stimulus; a mixture model describing the rapidly changing genome of *Plasmodium falciparum* (the malaria parasite); and the analysis of bumble bee pollination experiments using both classical and Bayesian statistical methods.

For the crickets project, our group collaborated with Professor Hadley Horch’s lab to help interpret the data of a new class of experiment. Seeking to understand how crickets respond to external auditory stimuli (as bats hunt crickets via echolocation), the Horch Lab trained an artificial intelligence engine known as DeepLabCut to monitor cricket movement in real time. To decipher the convoluted signal that DLC outputs, several programs were developed to track salient features (for example, how a cricket’s body angle changes depending on from which side the sound emanates, or how the magnitude of the sound affects the twitch of the insect). Together, we’ve successfully produced a proof-of-concept work for these methods, demonstrating that this class of experiment does indeed produce analyzable data. Once the strain of the pandemic subsides, we are all very eager to gather enough data to draw meaningful statistical inferences on cricket behavior.

During our investigations of the *P. falciparum* genome, we focused on the genetic code of both the apicoplast and mitochondrion of the malaria parasite. While the genome of the eukaryotic parasite itself is intractably tangled, their simpler organelles divide independently and in a far simpler manner (as the apicoplast and mitochondria are really embedded prokaryotes). However, this relative simplicity is still a staggering mixture model. Learning about the recent developments in the field required integrating cellular biology and genetics with probability theory and algorithm design. Additionally, we developed our bioinformatics toolkit, designing a pipeline to clean and process vast quantities of genomic data. By comparing the malaria genomic features in the data with their geographic origin, researchers have uncovered mysteries of human migration. Exploring the intricacies of this process with such an interdisciplinary lens was a particularly rewarding experience.

Lastly, bee pollination is a well-studied and fascinating field. Working with the experiments done by Prof. Patty Jones in Ecology, we’re in the process of linking classical statistics methods with a Bayesian approach. The experiments are concerned with pollination patterns and training in response to various nectar compositions. The typical approach modelling approach involves a generalized linear mixed model of regression, whereby each variable affects the duration of a pollination visit in a direct, conspicuous manner. However, by scaling up to a Bayesian hierarchical regression, we can model the experiment more holistically, representing the entire journey a bee takes. This requires a continuous-time Markov chain, which effectively treats any given pollination journey as a series of states (drawn from a discrete state space) with corresponding amounts of time spent in each state (exponentially-distributed dwell times). The potential for this method of analysis to streamline the current statistical protocol is exciting: an entire class of behavioral experiments could benefit from this more nuanced approach, likely helping to uncover discoveries currently hidden from view.

**Faculty Mentor: Prof. Jack O’Brien**

**Funded by: Surdna Foundation Undergraduate Research Fellowship Program**