I intend to pursue a career in academia exploring my research interests at the intersection of mathematical evolution, population genetics, and microbiology. My work reflects this multidisciplinary interest – I have applied mathematical and statistical modelling techniques to a wide variety of biological systems and medical problems including inferring protein evolution in viruses, quantifying evolutionary forces in dogs and wolves, identifying rare disease-causing mutations, and linking human agricultural history with gut microbial demographics.

Intellectual Merit

My undergraduate research at the University of Washington focused on statistical methods development for phylogenetic codon selection models describing protein evolution in RNA viruses. I worked with Dr. Jesse Bloom’s group at the Fred Hutchinson Cancer Research Center and was affiliated with both the Division of Basic Sciences and the Computational Biology program.

In summary, this research consisted of using empirical data from a high-throughput functional assay known as deep mutational scanning to define a null expectation for protein evolution when subject to only purifying selection. We then compared our null expectation to observed evolutionary histories found in natural sequence alignments. Then, evolutionary behaviors which differed from what had been characterized experimentally were interpreted as an indicator for biologically interesting selection.

In the case of viral proteins, this method allows us to detect sites which are subject to strong selection pressure, i.e., selection from the immune system. Understanding the evolutionary forces acting on viral proteins will allow us to better design preventative and therapeutic interventions, e.g., vaccines. We are applying this method broadly to RNA viral proteins, such as those found in influenza, Zika, HIV, and coronaviruses. I shared my work with the broader scientific and public health community via an oral presentation at The Allied Genetics Conference in 2020, which was hosted online, ironically due to a coronavirus. We are preparing a manuscript, on which I am the primary author, for submission to a peer-reviewed journal.

Additionally, as an undergraduate I worked with Professor Kirk Lohmueller’s group at UCLA via the Bruins-In-Genomics Summer Program for Undergraduate Research. We sought to quantify the statistical power of current methods for inferring the distribution of fitness effects (DFE), a fundamental entity in genetics which describes the proportion of all new beneficial, neutral, and deleterious mutations entering a population.

Previous work has shown that between distantly related species, e.g., humans and *Drosophila*, the DFEs are significantly different. However, the timescales and conditions under which the DFE evolves remain unclear. We used empirical genome resequencing data from canids, e.g., breed dogs and arctic wolves, because they provided an excellent model system in which to understand how the DFE evolves over recent timescales. We performed forward population genetics simulations to model evolution in canids and compared the inferred DFE and demographic parameters of simulated and empirical data. After the Summer ended, I continued to work with Kirk’s group remotely on this project, and have since joined UCLA for my Ph.D. We are preparing a manuscript, on which I am a co-author, describing our results and methods for submission to a peer-reviewed journal.

In addition to my ongoing work with Kirk’s group involving canid DFEs, I have also started several other research projects as a student in the Bioinformatics Interdepartmental program. In Autumn 2020, I rotated with the Lohmueller Lab and sought to develop a novel statistical method for detecting recessive deleterious mutations by searching for departures from Hardy-Weinberg Equilibrium.

Recessive deleterious mutations are mutations that are deleterious to health, but which only manifest in diploid organisms when they are two copies of the mutation present. These mutations are known to be difficult to detect because recessive mutations are not guaranteed to manifest in offspring, and deleterious mutations tend to exist in low proportion. Several rare diseases are caused by recessive deleterious mutations, which presents both a compelling medical issue and an interesting statistical question.

We developed a statistical framework where we can compute the null expectation under Hardy-Weinberg equilibrium for the proportion and frequency of how often mutations arise in a given population. We can then evaluate on a mutation-specific basis how the original population differs from the expectation while controlling for confounding factors such as population demographics and genetic drift. Mutations which differ significantly from the null expectation and that are found to be recessive can be identified as recessive deleterious mutations. We performed a simulation study where we applied this framework to simulated data and computed the statistical power of our approach. Next, we will apply this framework to empirical population data with known recessive deleterious mutations to further quantify our statistical power, before finally applying this framework to empirical data without knowledge of recessive deleterious mutations. We hope to provide insight into how and where rare disease-causing mutations arise in the population.

In Winter 2020, I worked with Professor Nandita Garud’s group on developing a demographic model for common commensal gut bacteria. Many population genetics tools and techniques were designed with eukaryotic samples in mind, so performing demographic inference on a set of prokaryotic samples like the human gut microbiome required a new set of tools. I implemented a bioinformatics pipeline which takes in microbiome data and outputs the genetic and evolutionary information of the most prominent singular strain from each individual host, resulting in “pseudo-isolate” data, on which we can perform standard population genetics inference.

We found that, across different demographic model constraints and across different species, there is a consistent population contraction in bacterial demographic history which overlaps with the onset of human agriculture. We posit that the shift from hunter-gatherer societies to farming communities greatly impacted the human diet, resulting in a loss of microbial diversity. Moving forward, we plan to further explore this hypothesis using non-population-genetics methods, e.g., phylogenetic or tree-based approaches, to verify our findings.

In Spring 2020, I worked with Professor Sriram Sankararaman’s group on statistical methods development for admixture graph inference. Admixture graphs consist of phylogenetic trees augmented with additional nodes representing admixture events, which enables a more faithful representation of evolutionary history than simpler graphs such as phylogenetic trees. I tested several novel search heuristics to exhaustively evaluate all topologies while efficiently searching parameter space. We found that current methods for admixture graph inference are often guaranteed to be confounded by local optimum and return the incorrect network topology, suggesting that different approaches are necessary for accurately inferring evolutionary histories.

Broader Impacts

My experiences as an early-career researcher have greatly benefited from working with mentors who are excellent scientists and, more importantly, excellent role models. As an NSF funded graduate student, I would aim to similarly dedicate my time and support towards fostering future scientists. From personal experience, programs like the Bruins-In-Genomics Summer Program for Undergraduate Research at UCLA offer a great opportunity for training young scientists.

This program consists of an intensive two months of research in which undergraduate researchers work directly with a graduate student or postdoctoral mentor and a PI. Mentorship often persists following the conclusion of the program as students receive additional funding for GRE prep courses, application fees for graduate programs, and travel grants to present their research. Through this program I will recruit undergraduate students from a wide range of backgrounds, academic and otherwise, to help answer the similarly broad interdisciplinary questions I am pursuing through my PhD studies.

In addition to research mentorship, I have been and will be involved in teaching and developing coursework. Previously as an undergraduate, I had the opportunity to serve as an instructor for an upper-division laboratory course offered by the University of Washington Department of Microbiology. My responsibilities for this position included developing a comprehensive 10-week lesson plan regarding a wide variety of cell-and-tissue culturing techniques for environmental bacteria, facilitating a two-hour laboratory section twice a week, holding office hours, and writing, grading, and proctoring written examinations. As a I graduate student, I will continue my work in education by serving as a teaching assistant in the fourth year of my PhD.

Additionally, I am passionate about improving working conditions for graduate students and other early-career researchers. I am pursuing this passion in two ways: first, I represent graduate students and early-career research fellows in my PhD program as part of Student Researchers United – a multi-campus wide effort to collectively bargain for improved working conditions for graduate student researchers across the University of California, and second, I serve as one of the elected representatives for my graduate program in my institution’s Graduate Student Association.

Student Researchers United is a collective bargaining unit of graduate student researchers across the University of California. The goal of this organization is to secure legal rights and protections for trainees, improve working conditions, and advocate for increased research funding. My role in this organization has been focused on drafting communications with local politicians to advocate for increased funding for research.

I serve as a representative for my graduate program on the Biological Sciences Council, which is one of thirteen academic groups which comprise the Graduate Student Association at UCLA. We meet biweekly to discuss ongoing topics and issues which graduate students face at UCLA as well as track and manage a budget which funds cross-disciplinary events across campus to facilitate meaningful relationships and collaborations with other graduate students. I prioritize advocating for workshops to provide assistance and guidance to UCLA undergraduates who are interested in entering graduate training in the Biological sciences.