

Intellectual merit: My passion for genetics stems from the attractive idea that we can explain the vast diversity of life through simple models of genetic inheritance. I was introduced to the depth of genetics through research and immediately found myself engrossed in it. I began by working with Dr. Jeffrey Ross-Ibarra at the University of California, Davis, who recruited me after we met at a conference. With his close mentorship, I began a project on the genetics of maize domestication. At this point, I knew nothing about maize, domestication, or genetics, but I slowly learned through many meetings and readings. This mentorship would prove instrumental in my decision to become a researcher and drive my desire to mentor others in science.

The key question I tackled was this: what genetic changes happened during maize domestication and how has natural selection been affected by domestication? My task was to model the demographic history, that is, the changes in population size, using a method called approximate Bayesian computation. We found that the domesticated maize population was reduced to 5% of the ancestral size and that this reduced the ability of natural selection to remove harmful variants in this population. This has important implications for crop scientists, who use artificial selection to weed out harmful variants. My contribution was many simulations of the demographic model, sampling over a range of parameters, and retaining those that matched the observed data. These results were published in a paper I co-authored (Beissinger et al 2016, *Nature Plants*). Perhaps more importantly, this paper fueled my passion for genetics. With a simple set of models and inference frameworks, we were able to make powerful statements about evolution. I dove into the literature of population genetics, reading as many papers as I could.

Separately, I undertook another project in Dr. Ross-Ibarra's lab. Another love of mine is programming, and I saw an immediate need in the population genetics community for easy to use software. Many geneticists lack a programming background so I sought to create software that addresses this usability gap. I was not the only who saw this need, and quickly found myself in charge of a collaboration between our lab in Davis and Dr. Peter Morrell's lab in the University of Minnesota. I led a project to create a wrapper and visualization software, called ANGSD-wrapper, for population genetics, which tries to address the growing cross-disciplinary nature of the research. This was a successful project that required writing software that could be used by many researchers, coordinating several people across institutions, and extensively testing software. In the end, it resulted in a first author publication (Durvasula et al 2016, *Mol Eco Res*). I found it very rewarding to be able to create a tool that allows biologists to conduct research that would otherwise not be possible, and have since looked for ways to incorporate the idea of broad usability into my research.

By this point, I had immersed myself into the world of genetics research. At the end of my undergraduate education, I applied and was accepted to an internship abroad in Vienna, Austria with Dr. Angela Hancock. The project was to study the demographic history of *Arabidopsis thaliana*, the principle laboratory plant species.

In the months prior to my arrival, a large consortium called the 1001 Genomes Project undertook the sequencing of over 1000 *A. thaliana* accessions. They collected plants across Europe and Asia (the presumed native range) and cataloged the genetic diversity. Dr. Hancock's group was interested in plants on the Canary Islands, off the coast of Africa, in order to study local adaptation to climate. In order to understand where these plants came from, she collected and sequenced the genomes of plants from North Africa that were not included in the 1001 Genomes project. My project was to compare the genomes of European and North African plants to these Canary island plants and attempt to discover where they came from. During the course of this analysis, I discovered that the African plants were very different from the European plants

and harbored a large amount of older genetic variation, suggesting the plant had deep roots in Africa. This was in stark contrast to the prevailing understanding of the species history, which thought the populations in Africa were peripheral to those in Europe.

Our focus changed from the Canary Islands to Africa and we dug deeper. We contacted collaborators with expertise in sequencing ancient DNA and had them sequence specimens saved in an herbarium in South Africa. We found that these plants also harbored deep connections to Africa, rather than being recently introduced as previously thought. Our study brought about a new understanding of the history of this well-studied plant – suggesting a very ancient history on the African continent. Follow up work is underway to understand the kinds of adaptations that may exist in climates very different to that in Europe and Asia.

During the course of this project, I led an international team with members working in Austria, Germany, Japan, and Spain in a large sequencing effort that resulted in a new understanding of an important species for understanding biology. This project resulted in a first author publication which was covered by *Science* and *Nature Plants* and highlighted as ‘exceptional’ by the Faculty of 1000 (Durvasula et al 2017, *Proc. Natl. Acad. Sci. U.S.A.*).

These research experiences solidified my resolve to become a researcher and compelled me to apply to graduate school. I began my graduate career at the University of California, Los Angeles, an institution that attracted me due to its fiercely interdisciplinary genetics department, which spans the range of research from very applied clinical genetics to basic research questions about the nature of evolution.

At UCLA, I am currently co-advised by two professors, Dr. Kirk Lohmueller and Dr. Sriram Sankararaman. With these professors, I am studying the forces that shape and maintain genetic variation. With Dr. Lohmueller, I have already co-led and completed a project that is available as a preprint and currently submitted to a journal (Huber and Durvasula et al 2017 *bioRxiv*). In this work, we explored the distribution of dominance coefficients for new mutations. In genetics, dominance is the relationship between the fitness of individuals with 1 copy of a mutant allele and individuals with 2 copies of a mutant allele. It is fundamental to genetics, yet there has been little work in quantifying it using genomic data from natural populations. In this project, we used data from my previous paper (Durvasula et al 2017, *PNAS*) as well as from another related plant species to estimate the distribution of dominance. Our work shows that the amount gene expression needed, per gene, is the determining factor in the degree of dominance of new mutations. This work answers a classic question in genetics using simple models and is exactly the type of research I am excited to continue in Dr. Lohmueller’s lab. The next project I am working on in this lab involves understanding how stabilizing selection is affected by population history. Models of stabilizing selection, that is selection on a complex trait at an optimum, suggest that the effective population size will affect the variance of a phenotype, but not the mean. Preliminary results suggest that we can see these differences when measuring gene expression in both humans and *A. thaliana* in different populations. This will have important implications for understanding how complex traits evolve.

In parallel, in Dr. Sankararaman’s lab, I am working on a method to detect introgression from archaic hominin species in modern humans (see Research Statement). Here, we use data from simulations to train a machine learning model to distinguish between regions of archaic hominin DNA and regions of modern human DNA. In this project, I am particularly excited about applying machine learning to population genetics as it is underutilized thus far. My future work at UCLA will involve merging statistical machine learning with genomics to answer the new questions that large data sets will allow us to ask.

Broader Impacts: As an undergraduate, I benefitted greatly from close mentorship from faculty, graduate students, and postdocs. Using my position as an NSF funded graduate student, I would like to dedicate my time to developing curious students into future scientists. I have already begun mentoring a computer science student in genomics (see Research Statement) and I am looking forward to mentoring more as I progress through graduate school.

At UCLA, we have a program specific to genomics, Bruins in Genomics (BIG), that focuses on recruiting undergraduate level students. This intensive summer program introduces students to the world of research, pairing them with a mentor and a PI. Two months of research culminates in a presentation, which is important for students applying to graduate schools. I will recruit students through this program from a wide range of backgrounds to help answer the interdisciplinary questions being pursued during my PhD. For example, I can mentor students with biology backgrounds in computer science as well as students with computer science backgrounds in biology.

An extension I would like to incorporate involves getting local high school students involved as well, which would allow us to draw on the diverse community in Los Angeles. In particular, I would like to reach out to students who never saw themselves as researchers or scientists. While I always thought science was very interesting, I never pictured myself doing it because it is often portrayed as the realm of straight A students and excellent test takers. I struggled with grades in high school and college, and it wasn't until Dr. Ross-Ibarra let me research in his lab that I realized I could do it too. I hope to extend this to other students and shape the next generation of researchers.

Another way I will reach out to students is through teaching. In addition to research, I also had the opportunity to assist in teaching a graduate level course at UC Davis. While I was developing ANGSD-wrapper, Dr. Ross-Ibarra and another professor, Dr. Andrew Whitehead, began to develop a course in ecological genomics for graduate students. This class was designed for students with less of a computational background and was designed around ANGSD-wrapper as a tool for ecological genomic analysis. I taught other students for the first time and found that I truly enjoyed sharing my knowledge with other students.

Since attending UCLA, I have explored this passion in two ways. First, I started a graduate student-only journal club called "Frydomics" where I and other students interested in genomics meet to discuss papers. Given the cross-disciplinary nature of genomics, we are able to teach each other about various topics. This is especially useful to new graduate students and undergraduates, who have less experience and are often eager to learn more. We have students from both biology and computer science backgrounds, which allows us to learn and teach across disciplinary boundaries. Second, I will be a teaching assistant for a class in population genetics taught by my advisor, Dr. Lohmueller in Spring 2018. I will work closely with Dr. Lohmueller in developing course material and in particular, I will focus on assignments to help biologists learn how to code, which will be important for students that choose to pursue research. In the future, I intend to work with others at UCLA to bring Software Carpentry, a nonprofit dedicated to teaching programming to students, to campus and lead courses as an instructor.

At UCLA, I am able to explore my diverse interests in problems in genetics. I intend to continue to bridge the interdisciplinary gap between genetics and computer science, teaching many students in both disciplines along the way. With NSF funding, I will look forward to advancing scientific knowledge, expanding my research interests, and helping to develop the next generation of scientists.