Ever since I learned about DNA, I have been perplexed by it. I beheld the intricacy of life and was bewildered by the fact that it all stemmed from a language of nucleotides. Just a few permutations of *four small letters* created the wild diversity and immense complexity of all life on Earth. If we exactly understood *how* these building-blocks fit together to build an entire organism, we as a species could grow food to feed the world, build organisms to process our waste, construct living buildings to house and care for us, and create medicine for whatever ails us. I wondered: how complex could a system composed of just four small letters be?

The reality is far more overwhelming than I ever envisioned. But as I learned more and came to understand that the complexity of living systems also have complex molecular underpinnings, I became even further invested in determining how the four letter code of DNA created such a multitude of different creatures and phenotypes. As a graduate student, I am now interested in genetic disease and how a particular mutation or other genomic abnormality produces a pathological phenotype. I am also interested in more fundamental questions about the nature of mutational processes and how to efficiently detect mutations.

I realized, however, that to create the knowledge necessary to intimately understand these complex biological systems requires high-quality data—and lots of it. With the flood of big data changing how people collect and analyze information, there is high demand for data scientists to create and improve methods of organizing data. Development of algorithms that quickly process this data and separate signal from noise is more important than ever. Growing up with a love of computers and an engineer father who encouraged mathematical inquiry, I was excited to combine my passions, become a data scientist, and contribute to the disentanglement of the complex network of life. As I felt no single program at my university, Arizona State, could meet my needs, I decided my best course of action would be to dual major in mathematics and molecular biology. Over the course of my undergraduate career, I tackled projects in four labs that have shaped me as a scientist.

Eager to expand my biological and mathematical training, at the end of my freshman year I joined Dr. Sudhir Kumar's computational laboratory where I worked on the FlyExpress database (www.flyexpress.net). The purpose of this database was to collect and store images showing where a gene is expressed throughout development of a fruit fly embryo. The website also offers some analysis tools to help identify genes that may be interacting based on their spatial convergence. I annotated and categorized *Drosophila melanogaster* gene-expression-localization images. I also added metadata and sanitized images that were mined from papers. Some of this work was presented in a poster at a national *Drosophila* conference. I was also able to work with the development team on fixing bugs in the website and analysis tools. In this capacity, I had the opportunity to learn to program, a skill critical to becoming a successful data scientist. Working in a interdisciplinary lab, I fulfilled my desire to combine the biology of gene expression with the mathematics of automated image analysis.

One thing that drew me to working in this lab was the mobile phone apps the lab developed to facilitate database access. I believe that making science easily accessible to the public is important, particularly for publicly funded projects. If someone hears about an interesting gene, he or she can take out his or her phone and quickly view pictures of where that gene is active during *Drosophila* embryo development. As a teaching tool, these colorful images are a fun, visual way to introduce students to a widely-used model organism and concepts like localized gene expression and gene interactions. Using Timetree, another app developed in the Kumar lab, students can discover and visualize times of divergence between any two species. This is an especially important concept for students to understand, particularly due to public resistance to evolutionary biology. I continue developing these types of tools and advocate that labs offer mobile versions of their tools and databases when appropriate as training and educational tools. In an increasingly mobile-focused society, this is a powerful way to easily communicate the importance and impact of research to the public.

In an effort to gain more wet-lab experience and further expand my interdisciplinary training, I joined the materials science lab of Dr. Nicole Herbots. The lab was developing a proprietary technology to prevent fogging and blood accumulation on glass lenses for surgical applications. This position allowed me to learn more about the industrial side of research and explore the interactions between physics and biology. I applied physical theories of light diffraction and fogging and biological theory of protein interaction to develop a product far more effective than any competitor's. This showed me how important interdisciplinary thinking is to solving complex problems. I presented my work in symposia at the physics department and at the honors college. This work was presented by a colleague at several national conferences as well. After some time, I also mentored a high school intern who came to the lab after school. I guided him through the basics of the scientific process and scientific writing as we applied for a grant from the Collegiate Inventor's Competition. During this process, he gained first-hand knowledge about science and science communication that will help him succeed as a burgeoning scientist.

Ultimately, I decided industry work was not for me, and I desired a larger emphasis on computation and analysis of large datasets. I found this in Dr. Paul Davies' lab. Dr. Davies is part of the Physical Sciences Oncology Center and is interested in examining cancer from an evolutionary perspective. I worked on my Honors thesis under his direction. For the project, I mined gene families from every human gene and compared the evolutionary history of the species in these families to estimate an evolutionary age of these genes. I also used the annotations of these genes to find patterns in the ages and functions of genes mutated in cancer. Through this analysis, I identified a set of well-conserved genes that may be responsible for increased genome instability in cancer. A paper describing this work has been submitted for publication.

The entirety of the data used and analyses performed is open source and freely available to download and use on Github (www.github.com/adamjorr/thesis-2015). I strongly believe in open source scientific software and welcome contributions to any software I produce. Transparency is vital to the scientific process and making my software open source allows others to review every analysis I perform. This makes my work have higher impact, as anyone can slightly modify my analyses and easily incorporate them into their own projects. I also provide scripts to download and analyze the same data I use in my original experiment. This makes any analyses I do highly reproducible.

In an effort to organize more outreach events and increase knowledge about open-source software and computational methods, I became an instructor for Software Carpentry, an organization devoted to increasing scientists' computational skills. After obtaining certification, I organized and taught a two-day workshop at my university. We were able to instruct 24 students in basic Unix, git, and R skills that will help improve the quality and efficiency of their work. The workshop was in high demand, and all 24 spots in the workshop had been reserved less than a day after registration opened. I am currently in the process of organizing another workshop with a higher capacity to help improve scientists' computational skills in my research community. As I continue my career, I will continue to advocate for open-source software as a way to improve the quality and reproducibility of research.

As I completed my thesis work, I thought about the causes of cancer and how the genetic instability of the entire diverse tumor population contributed to development of the disease. I became interested in learning more about mutational processes. To this end, I joined Dr. Reed Cartwright's lab in my last year of my undergraduate career and have continued working in his lab for graduate school. His work focuses on computational mutation research and statistical models of evolution, a perfect fit for my interests. With Dr. Cartwright's experience with software development, I have since significantly improved my programming and data management skills and will continue to do so. I have most recently started a project identifying and mapping mutations in an individual of the species *Eucalyptus melliodora.* Using the mutations and mutation rates identified by my method, I have been able to recover the physical topology of the tree from the phylogeny estimated by somatic mutations. I was able to present this result in a poster at the Society for Molecular Biology and Evolution's 2016 conference in Gold Coast, Australia. Work on this project in particular has prepared me to carry out the project I propose in the Research Statement.

I saw first-hand the usefulness of the educational tools I helped create at Arizona State University's (ASU) outreach event Night of the Open Door. During this event, ASU invites people from the surrounding communities to explore the campus, research labs, and various departments of the school. Each department and lab puts together an activity targeted at children, parents, and others to generate interest in science and educate the community on the research that's being done at the department. I helped the Kumar lab devise and operate an activity for the event. We introduced guests to the TimeTree app and gave children plastic animals, challenging them to guess when their ancestors diverged and imagine what that ancestor might have looked like. We then showed them how to use the app to see how close they came to literature estimates.

In the second event I helped the Cartwirght lab display the evolution of a basic spatial model that colorfully displayed the different alleles present in a population. This way, people could visualize how a population evolves in time and spreads out spatially. It also served as a colorful, eye-catching backdrop for our other activities. We also taught guests about the genetics behind certain cat and dog traits, then allowing them to "create" and color an animal using an assortment of genes. These guests took home not just plastic figures and sketches of pets, but an appreciation for science and how science impacts their daily lives. The smiles on even the smallest childrens' faces as they had fun doing science and learning about the world around them still serves as an inspiration to me during long nights at the lab, and I will certainly continue doing all I can to promote curiousity and the joy of science.

Small grants for interdisciplinary graduate work are hard to find. As an interdisciplinary scientist, this fellowship would give me the freedom to further educate students on the merits of computational and data management skills without fear of losing funding. My interdisciplinary background gives me a unique perspective on biology, math, and computer science that allows me to engage with and prepare a diverse population of students for tackling complex problems in biology. At the same time, it will allow me to further improve my own skills so that I can continue to develop and share tools and algorithms to disentangle the sophisticated web that is biology.