Personal Statement

Broader Impact Upon entering university, I knew right away that I wanted to seek out opportunities for scientific research. Part of this desire came simply from a personal curiosity that manifest as a desire for understanding. Throughout my life, I have always been driven by learning, and seeking to understand how and why anything is the way it is. Scientific research offers an avenue to seek that understanding. But further than simple curiosity, I was drawn to the potential to contribute something entirely new. Everything that I had learned in school had been new to me, but already well established to the general wealth of knowledge. In scientific research, I saw an opportunity for creativity that I had previously only experienced through music. There exist unsolved problems, and the job of a scientst is to seek out these problems and find a solution, adding something completely new to the sum of what we as a society know – this is what drove me to seek out a professor to work with and learn from in my first semester. I've been fascinated for a long time with genetics, so I started as a molecular and cell biology major, and joined the Nelson Lab at UConn under Dr. Craig Nelson. In this lab, I assisted graduate students, participated in lab meetings, and learned about the techniques and methods necessary to conduct successful biological research. However, the most important thing that I learned during my time there was not related to biology at all. For one of the projects, a graduate student in the lab was working on clustering of single-cell genomic data, and at the advice of Dr. Nelson I enrolled in a programming class to gain the necessary skills to assist the graduate student. It was through this opportunity that I discovered my passion for computer science, and learned about the field of bioinformatics. The problems in molecular biology inspired me – from using genetic information to learn more about the history of life and the machinery that drives it to developing strategies to combat and cure disease. But it was computer science – from developing and analyzing algorithms to writing software – that I knew I wanted to spend my time doing each day. Bioinformatics and Computational biology is a uniquely interdisciplinary field which allows me to combine the two, and I knew that I had found something worth pursuing.

Intellectual Merit The second semester of my sophomore year, I switched my major to computer science, and joined the Computational Biology Lab with Dr. Mukul Bansal. Since then, I have worked on multiple complex biological problems. I am consistently drawn to these problems because it is easy to see the downstream effects that solutions to these problems would have. When I work to improve viral transmission network inference, I know that an improved understanding of disease transmission means that epidemiologists and doctors will have an improved understanding and ability to combat that disease. The problems which I encounter in bioinformatics allow scientists to use the massive amount of genetic data and computational power available to us to better understand life, all the way from differentiation of the earliest stem cells to old age. And even when you strip that layer away, it remains that the problems are really interesting in their own right from a computer science perspective. Computer science is very much a problem solving field – you have an input, and a desired output, and the first task is to find or develop a series of steps to get from A to B. The second step is actually implementing that series of steps as a program, a process which is never as trivial as it first sounds. Genetic data exists as long streams of

one of four characters, from which we can extract all manner of interesting conclusions, but only if the right algorithm is used. It is a creative process as much as a technical one, and in order to contribute something new. I have been a musician my entire life, and computer science research allows me to apply that creativity towards solving biologically meaningful problems.

Relevant Background

Intellectual Merit At the start of my junior year, I took on my own project within the lab. I applied for UConn's university scholar program, which focused on a large-scale research project. Although I was not selected for that program, I learned how to conduct a literature review, design a research project, and write a proposal. The goal of the project was to use information about the host of each viral strain to reconstruct a highly accurate viral phylogeny, with the ultimate goal of improved inference of viral transmission. Throughout the year, I built TreeFix-VP (Viral Phylogeny). I worked with both high and low level programming languages to write software which would evaluate the cost of a given phylogeny using transmission data in a highly scalable manner, and to conduct a heuristic search around an input phylogeny to find and return the lowest cost tree. I did preliminary testing using HCV outbreak data obtained through a CDC collaboration, and then began on a robust testing using simulated data to evaluate the improvement in accuracy of phylogenetic inference in a variety of settings. In addition, I compared TreeFix-VP with other methods commonly used for phylogenetic inference. Throughout the year, I continued to present on my progress both within the lab and in the department Bioinformatics Seminar.

Testing proved to be more difficult than I had anticipated, as it required the integration of multiple different software tools, use of a sophisticated simluation framework, and further research into the set of experiments which would best evaluate the performance of TreeFix-VP. Through this experience, I was able to reach out to and work with other researchers in the field, and I designed a set of experiments to test the improvement in phylogenetic accuracy across a wide range of parameters. I spent the summer following my junior year working full time on this project, and was able to overcome this challenge. By the start of my senior year, I had results which showed that TreeFix-VP was able to improve the accuracy of viral sequence phylogenies, and also had a positive downstream effect on the the reconstruction of transmission networks. The next step became to share and communicate this work. Throughout the semester, I have had the opportunity to present my work at the UConn Bioinformatics Seminar, at the CANGS Workshop at IEEE ICABBS 2018, and at the UConn Fall Frontiers poster exhibition. I am currently working on writing a manuscript for publication in an ICABBS conference publication. Over the course of this project, I have learned how to design and perform indpendent interdisciplinary research, and how to clearly and effectively disseminate the results of my research. Between my classes, internships, and lab work, I have become a more skilled biologist and computer science. Moreover, I have contributed something new to the understanding of viral evolution and transmission.

In addition to my research work, a number of other experiences have prepared me to be a successful graduate student and researcher. I have participated and won awards in multiple hackathons, where I have had the opportunity to work with a team, identify a problem,

design and implement a solution to that problem, and present the solution - all within 24 hours. Through these experiences, I have learned as much about problem solving and communication as I have about computer programming. The summer following my sophomore year, I interned at Optum Technology. I was a Scrum master for my team, a role in which I helped guide the progress of our project and communicate with management. Although this was an industry position, the project to which we were assigned was much more of an exploratory research project. We were tasked with evaluating the viability of machine learning for automatic insurance claim adjudication. In addition to this, we developed a machine learning pipeline and guide manual which was distributed to developers within Optum to facilitate the use of machine learning for general problem solving. I am also currently working as a teaching assitant for Theory of Computation. In both of these positions, I have had to break down complex topics and present information in a way that is easily understandable by a diverse audience. Each of these skills directly translate to being a successful graduate student and researcher.

Broader Impacts My work on TreeFix-VP was driven by the desire to impart real change, by improving the accuracy of viral transmission networks. The best way to combat the outbreak of a disease is to understand what factors contribute to its transmission, and that can only be done if we have an accurate picture of how the disease spread. Although only a smaller part of that much larger picture, it drives me to know that my work with TreeFix-VP to reconstruct accurate viral phylogenies will lead to a reduction in disease transmission. Throughout my computer science classes, I have been particularly interested in graph and network algorithms, so the study of disease networks is a natural intersection between what I find interesting and something that is socially impactful.

Future Goals

My goal is to pursue graduate study and research in the field of bioinformatics and eventually a career in bioinformatics research. The most interesting problems to me are those tangential to health, because those are the problems that I see as being concretely meaningful. I want to continue research that deals with networks, where I have interests in two areas. The first is a continuation of research with disease transmission networks, where my goal is to contibute to the understanding of what factors impact the spread of disease and how this spread can best be combatted. My second area of interest is in genetic regulatory networks, where I aim to develop further understanding of how gene expression is regulated and the way that information flows throughout cells. These areas of work interest me both as complex computational problems and meaningful biological problems. I will not only develop theoretical solutions to these problems, but create implementations which can by used by other researchers and by health professionsals. By completing a PhD in bioinformatics, I will be able to achieve my goals of engaging in creative activity, tackling interesting, challenging, and socially impactful problems, and furthering the knowledge for future researchers by contributing something entirely new to the field of bioinformatics