Tyler Shimko
Personal Statement

I first came to appreciate the power of biological research in my high school biology class. Through the instruction of my incredibly devoted teacher, I recognized that genomic research had the power to make predictions about and alter the traits of living creatures. Upon graduating from high school I chose to attend the University of Utah, where, with the help of the Office of Undergraduate Research, I would have an immediate opportunity to become deeply involved in the research process and pursue my interests to the fullest. Since beginning my undergraduate studies, I have been fortunate to take part in biological research at three universities across the United States and participate in projects ranging from molecular biology to neurobiology to quantitative genetics. Through these experiences my appetite for discovery has only grown.

My introduction to biological research came during my freshman year at the University of Utah. I joined the laboratory of Dr. Erik Jorgensen to assist in the construction of universal transgene insertion sites within the genome of *C. elegans*. My project eventually resulted in the creation of three distinct transgene landing sites. I presented this research at both the University of Utah's Undergraduate Research Symposium and Bioscience Symposium for Undergraduate Researchers. Additionally, I published an abstract of the work in the university's Undergraduate Research Abstracts Journal. My experience in the Jorgensen lab solidified my interest in biological research. I found it exhilarating to discover, create, and share knowledge with others. However, my exposure to the broader research community alerted me that there were other opportunities to learn new skills and make significant contributions. To expand my research skill set, I sought summer internships in laboratories focused on computational methods in addition to molecular biology. I obtained an offer from Dr. Leonid Kruglyak, of Princeton University at the time.

In Dr. Kruglyak's lab I worked with Dr. Erik Andersen, a post-doctoral fellow. Dr. Andersen had previously completed a genomic mapping experiment wherein he had determined a region of the *C. elegans* genome that conferred resistance to the herbicide paraquat and enlisted my help to construct strains in an attempt to identify the causal genomic variations in that region. I explored different techniques, using both modern molecular biology and more classical genetic crosses, to construct strains with which we could test the hypothesis that we had successfully identified causal genetic variants. I presented my work from Dr. Kruglyak's lab at both the Utah Conference on Undergraduate Research and the National Conference on Undergraduate Research and these results will be published in a manuscript (in preparation). Throughout the course of the summer, I supplemented my hands-on laboratory experience with instruction and practice with the computational methods that Dr. Andersen had employed in mapping experiments. Since both Dr. Andersen and I found our partnership enjoyable and productive, we agreed to continue our collaboration in his new laboratory at Northwestern University in the following summers.

Upon my return to the University of Utah in the fall of 2012, I began a new project in the Jorgensen lab to identify suppressors of the phenotype associated with the null allele of a protein involved in synaptic vesicle endocytosis. I learned new techniques for the design and implementation of genetic screens and applied my new computational skills whenever possible. However, by the end of the academic year, it became apparent that the phenotype of interest was too weak for our suppressor screens to yield any useful information. While I was originally upset that our project and a year's worth of work were for naught, I quickly realized that failure is more of the rule than the exception in biological research. To continue toward a career in research, I would need to learn from failure and to fail gracefully. In this respect, the year had

not been wasted. I looked forward to my summer in the Andersen lab, where I could make use of a developing computational and statistical skill set.

Over the past two summers I have had the opportunity to explore how genetic variation dictates the way in which organisms respond to their environment. In the lab of Dr. Erik Andersen at Northwestern University, I sought to determine the ways in which the genetic variation present in the worldwide population of the model nematode *C. elegans* affects responses to different chemicals including herbicides, pesticides, chemotherapeutic agents, and anthelmintics (compounds used to treat infections of parasitic nematodes). At Northwestern, I helped to construct and optimize a high-throughput screening technique that allows us to measure the effects of the aforementioned compounds on nematodes in a multitude of ways. This pipeline has allowed us to conduct genome-wide association and linkage mapping studies. A description of this process and the associated analyses of the data will be presented in a manuscript that is in preparation. This past summer, I designed and built software to clean and process the data from our screening experiments and run statistical tests to map phenotypic differences to genetic variants. This software, *COPASutils*, has been published in the journal PLoS ONE and made freely available on the Comprehensive R Archive Network. My hope is that this software will gain widespread use in the model organism research community.

During the past two academic years, I have had the pleasure of working in the lab of Dr. Gillian Stanfield at the University of Utah. My project in Dr. Stanfield's lab has focused on identifying mutations that suppress premature sperm activation in male *C. elegans* worms. In my first year in Dr. Stanfield's lab, I carried out a series of crosses and positive phenotype selections as part of a genetic mapping scheme. At the conclusion of my first year in the lab, several strains that displayed the suppressed phenotype were sent for sequencing in the hopes of being able to identify causal mutations. Now, in my second year in the lab, I am beginning to utilize raw genetic sequence and mapping data to identify variants implicated in suppression. Eventually this project may help to identify the causal mutation for each of the individual suppressed strains. My work in Dr. Stanfield's lab will culminate in the publication of my honors thesis.

In working with Dr. Andersen, I have become impressed by his large network of collaborators. Dr. Andersen routinely augments his own expertise by collaborating with individuals across the globe. I have witnessed, firsthand, how a large, diverse, and international group of scientists can approach a problem with different perspectives and collectively arrive at a solution that is superior to that proposed by any individual. In light of this realization, I have begun to make a concerted effort to incorporate an international facet into my education. This year, I have applied to three scholarships, the Marshall, the Churchill, and the Gates Cambridge for the opportunity to complete a year of coursework and research in computational biology at the University of Cambridge in the UK. Should I not be awarded any of these competitive scholarships, I will strive to work internationally for a portion of my graduate career. By working outside of the US, I will bolster my academic network, open new collaborations, and have the opportunity to serve as a science writer, educator, and mentor in the global community.

Throughout my research career, I have attempted to maintain a commitment to openness and transparency in both my academic and professional development. I have made a great effort to ensure that all software that I write is publicly accessible and licensed in such a way that allows for free reuse and modification. Additionally, I have published in open access journals whenever feasible. As I move forward in my academic career I plan to continue this commitment to open, accessible, and reproducible research. I believe that growth within the sciences can be fostered by lowering barriers to research not only in the way in which results are published and

made accessible, but also through outreach and introduction of the research process at an early stage in students' academic careers.

I was fortunate to gain access to academic research early at the University of Utah through the university's Office of Undergraduate Research. To provide my fellow students with similar opportunities, I joined the office as an Undergraduate Research Ambassador and Advisor. In this position, I was able to talk with students and prospective students about the unique opportunities that research had afforded me and direct them to laboratories and research groups pertinent to their interests. I also helped to set up and oversee the university's annual Undergraduate Research Symposium, which allowed me the chance to get to know the projects of my fellow students. My experience with the Undergraduate Research Symposium also taught me a valuable lesson about the importance of communication within the research community. I recognized that the most interesting and effective presenters are those who can most accurately and succinctly describe their research to a non-expert. To improve my own communication skills, I applied for a position on the Public Library of Science's Student Blog. I realized that the Student Blog would be an excellent platform for me to experiment with different communication styles and to have an audience outside of my area of expertise, an opportunity not necessarily available through traditional academic publishing channels.

In my writing for the Student Blog, I have focused on developing, broadly applicable topics within the biological sciences. For instance, my most recent post discussed the implications of increased experimental automation within biology, especially as it relates to molecular biology and the phenotyping of model organisms. I find that keeping the topics interesting, yet not overly technical draws in members of the general public and piques their interest not only about my research, but also that of my peers. I have learned a great deal about effective communication in my experience writing for the blog. I hope to one day apply these communication skills alongside my wide array of research-related skills as a professor and researcher at a major research university.

The ideal professor is one that can skillfully balance the duties of teacher, researcher, and engaged community member. As the son of an elementary and a high school teacher, the values of learning and teaching were instilled in me at an early age. I have felt compelled to share my knowledge and curiosity with my family and friends throughout my life. As a professor, I will get to experience the joy of sharing my passion and life's work with bright, talented individuals; perhaps inspiring some to follow the very path I chose. The prospect of guiding research projects is exciting to me. In the words of Albert Einstein, "As our circle of knowledge expands, so does the circumference of darkness surrounding it." For every question answered, many more wait to take its place. Through research, I will lead a challenging, fulfilling career with endless opportunity for discovery. Critically, my experience as a lifelong learner and educator will be fruitless if not shared with the larger community. As a child, I remember museums as places of wonder, institutions where humanity's collective expertise could coalesce and embolden the curious to ask the important questions. As a professor and expert in my field, I will be honored to contribute back to museums. I hope to share my knowledge through exhibits, trips, and activities to educate people of all ages and backgrounds. I will share with them our current understanding of biology: the field's greatest accomplishments, its most troubling failures, and, above all, its most vexing unanswered questions. I will always strive to be the ideal professor.

Tyler Shimko
Research Proposal

Elucidating how nutrition interacts with genes to influence phenotype
Keywords: quantitative genetics, gene-by-environment interactions, nutrition

In nature, food can be life-sustaining or potentially toxic. This dichotomy is particularly true when bacteria or fungi are the food resources because these organisms can rapidly evolve defense mechanisms that make their consumption perilous. Likewise, some animals might find a particular food source exceptionally nutritious while others may find it completely indigestible. These differences in absorption, nutrition, and toxicity of a food source are controlled by genetic factors. Furthermore, gene-by-environment (GxE) interactions between a consumer and its food often confound our ability to predict phenotype from genotype. It is essential to better understand the molecular interactions between food and consumer that contribute to nutrition and disease.

Caenorhabditis elegans and its various strains of bacterial food **[1]** serve as an ideal model to study the genetic interactions between consumer and food resource. Genetic variation between strains of bacterial food impacts the animal's ability to locate prey and the overall health of the animal **[2, 3]**. Additionally, genetic variation in the *C. elegans* population modulates phenotypic traits in response to nutrition of the bacterial food **[4]**. Here, I propose to utilize computational and molecular tools to dissect the genetic interplay between *C. elegans* and its bacterial food.

Quantitative trait locus (QTL) mapping has been used in a wide range of organisms including the yeast *S. cerevisiae* **[5]**, the roundworm *C. elegans* **[6]**, and the plant *A. thaliana* **[7]** to connect genetic variations to the molecular mechanisms controlling phenotypic variation. This technique leverages the natural genetic variation within a species to identify specific genes implicated in phenotype determination. Moreover, genetic tools available for bacterial gene mapping, such as transposon sequencing, can elucidate the bacterial genes responsible for triggering *C. elegans* phenotypic variation. The powerful genetic and computational tools available in *C. elegans* and bacteria will be used to identify genes in both species that contribute variation to the animal's phenotype.

Hypothesis: *Genetic variation in both bacterial food and the consumer interact to drive differences in the physiology of the animal.*

Aim 1: Determine the *C. elegans* genes that interact with molecular components of the bacteria. I will first measure body length, population size, and optical density, which represent proxies for the overall health of an organism, of each strain in an existing population of *C. elegans* recombinant inbred lines (RILs) after they are grown in liquid culture containing a single strain of bacterial food. Each RIL strain contains a mixture of two divergent genetic backgrounds **[8]**. Data collection will be repeated multiple times with different bacterial strains to identify species to which the animal's phenotypic response is variable. These phenotype data will be collected in an automated fashion using a COPAS BIOSORT large-particle flow cytometer. These initial experiments will facilitate the identification of bacterial strains that differentially affect *C. elegans* across the RIL population.

After collecting preliminary data, I will develop a suite of computational tools in the R statistical computing environment to augment the existing packages *COPASutils* **[9]** and *R/qtl* **[10]** for reading, processing, and mapping with the phenotype data. This software will extend the functionality of the above packages by utilizing differences in phenotype between a control and treatment condition to identify genomic regions that interact with molecules present in the bacterial food **[11]**. Beyond identifying QTL, the software I develop will define a confidence interval around each QTL peak, which will limit the number of genes possibly implicated in the animal's response to its food source. To identify the responsible gene, I will construct near-

isogenic lines that contain a region of one of the parental strains in the background of the other. Because QTL mapping relies on genetic differences, the isolation of specific genetic variants in a "control" genetic background employed by this technique will help to attribute specific phenotypic characteristics to a well-defined genetic interval. Through repeated backcrossing and phenotyping, the region will be narrowed iteratively until only the gene necessary to replicate the original effect will remain in the region. The genes identified through this process will represent the contribution to variable phenotype from the genome of the animal.

Aim 2: Identify bacterial genes associated with differences in *C. elegans* physiological phenotypes. *C. elegans* differs from many other organisms, such as yeast **[12]**, in that they consume live bacteria instead of simple sugars as a food source. Consequently, the animals are exposed to many different compounds when they consume their food, any one of which could cause a variable response between individual animals. To identify the bacterial metabolites that cause the variability in *C. elegans* phenotypes, I will create a gene disruption library from a strain of bacteria to which the animals vary in their response using a suitable bacterial transposon, similar to what has been done to determine that *C. elegans* requires the vitamin B12 **[13]**. These mutagenized bacterial strains will then be sequenced using transposon sequencing (Tn-seq) to identify the genes affected by the mutagenesis **[14]**.

Following sequencing, I will feed a representative sample of RILs mutant bacteria and measure their phenotypes as described in Aim 1. I will then construct a computational pipeline to identify genomic regions of the bacteria that, when mutagenized, cause changes in the genetic architecture, the underlying genetic basis, of the trait, as evidenced by gain or elimination of QTL. This finding would indicate that an animal's phenotype is dependent both on its genotype at a particular locus as well as the nutritional content of its food. Furthermore, changes in genetic architecture would imply that the mutagenized bacterial gene is necessary for the GxE effect to be realized within the animal. The genes identified through this pipeline will uncover the contribution to variable phenotype from the genome of the bacteria.

Broader impacts: The protocols developed in this project can be adapted to other model organisms or agriculturally important livestock. Studies using adaptations of the tools developed in this project on model organisms will help to further define genetic interactions between metazoans and their plant, animal, or bacterial food sources. Additionally, this research will be important for identifying and manipulating GxE effects in livestock animals to increase yield or decrease sickness and malnourishment in the food resources depended upon by humans. For this reason, it will be critical that this research is as accessible as possible. I plan to take full advantage of the resources available, including open-access funding, to release all data, code, and communications in a manner consistent with the "gold" standard of open-access publication. Ultimately, I hope this strategy facilitates work by other scientists in developing countries where definition and manipulation of GxE interactions involved in livestock production is critical to maintaining an adequate food supply.

References: 1. Félix, M.-A. *et al.* (2012). BMC Biology **2.** Werner, KM, *et al.* (2014). The Journal of Biological Chemistry. **3.** MacNeil, LT, *et al.* (2014). Worm. **4.** Walhout, AJ. (2014). Cell Metabolism. **5.** Ehrenreich, IM, *et al*. (2010). Nature. **6.** Gaertner, BE, *et al.* (2010). Genetics Research. **7.** Keurentjes, JJB, *et al*. (2007). Proceedings of the National Academy of Sciences. **8.** Rockman, MV, *et al.* (2008). Genetics. **9.** Shimko, TC, *et al*. (2014). PLOS ONE. **10.** Broman, KW, *et al*. (2003). Bioinformatics. **11.** Gutteling, EW, *et al*. (2007). Heredity. **12.** Bhatia, A, *et al*. (2014). G3. **13.** Watson, E., *et al*. (2013). Cell. **14.** van Opijnen, T, *et al*. (2014). Current Protocols in Molecular Biology.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit
Excellent

Explanation to Applicant
Excellent intellectual merits! The PI has a strong academic records with multiple awards and recognition. As an undergraduate researcher PI has first-authored and published R package COPASutil tool in an Open Access Journal.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Excellent

Explanation to Applicant
The PI has demonstrated track record activities in his campus increasing research awareness to other undergraduates through his role as Undergraduate Research Ambassador for the University's Office of Undergraduate Research. The PI has authored the Biology Student Blog and aims to become a professor in future. The PI has a plan to make the research outcomes/ the tools he developed available to others to use. COPASutil tool already published in PlOsOne is an example. Explaining more outreach activities would make the proposal even stronger.

Summary Comments
This application package is very strong. Both the statements are well written and the provided letters support the intellectual merits. Research plan to integrate research with education (or other broader impact activities) could have made the propsal even stronger.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit
Excellent

Explanation to Applicant
The candidate has a very strong resume and strong letters of recommendations. He is independent and creative, and the proposal reflect an original research area for the candidate's laboratory. I have no reservations that this candidate will be a successful scientist. However, the proposed work appears to be overambitious, and consists of a general hypothesis and too much emphasis on experimental techniques, with little information provided on the scientific discoveries. There is a lack of focus on the actual set of genes that will need to be studied. The use of QTL without narrowing down the gene sets may result in a study that becomes much longer than anticipated. The proposed work could be greatly improved by focusing first on aim 2 and narrowing down on the type of category of compound (lipids, proteins, etc.) to be studied. Then, the genes in C. elegans involved in breaking down these compounds could be specifically studied.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Very Good

Explanation to Applicant
The broader impact relates to the impact of the techniques proposed more than the actual scientific discovery. It is not clear how

some of the results from the proposed work can be translated to yield improvements in livestock animals.

Summary Comments

A very creative proposal that lacks specifics on the scientific discoveries and relies primarily on the description of the technical methodolgoies that will be utilized

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit

Excellent

Explanation to Applicant

The applicant has excellent academic record, and has carefully selected his elective courses that would provide him with a good background for quantitative genetic research. Further the applicant has been proactive and secured several undergraduate research fellowships, and has gained good experience in C. elegans genetics as well as computational tools, the core qualifications for this research topic. The letters of recommendation strongly corroborate the applicant's academic qualifications and research experience.

Broader Impacts Criterion

Overall Assessment of Broader Impacts

Excellent

Explanation to Applicant

The applicant had well thought about the broader impacts of this research. He envisions the application of his proposed methodology to metazoans, especially to livestock production in developing countries, to achieve adequate food supplies. He is also committed to making his work freely available for the benefit of other scientists.

Summary Comments

The applicant aims to study changes in host physiology due to the interaction between the genetic variation in the host and the genetic variation in the organisms it consumes as food. The well proposed research strategy and scope of broader impacts gives it a strong score on both Intellectual merit and Broader impacts. The strong recommendation letters reflect the applicants educational qualification and excellent research skills.