

STATEMENT OF PURPOSE: OXFORD DPHIL IN GENOMIC MEDICINE AND STATISTICS

JARED G. GALLOWAY

Genomic and precision medicine research has opened new horizons for human health. With recent advances in technology, quantitative biology research, and access to large datasets, the field has become more equipped to solve complex problems that will play a large role in current and future generations to come. I am applying to graduate school because I believe it is a necessary challenge to nourish my biological skill set and interest in quantitative problem-solving. This training will present an opportunity for me to leverage my computational background and help find creative solutions to problems facing the field at an exciting time for genomics. Moving forward, I would like to help progress the field and participate in the discussions that define how genomics are utilized in medicine. My professional goals are to build tools (software) that biologists can use to help solve a wide variety of problems in genomic medicine research.

Over the course of the past two years I have completed two major genomics research projects and am currently working as a full-time scientific programmer in the Institute of Ecology and Evolution at the University of Oregon. Advised by Drs. Peter Ralph and William Cresko, the first project I took the lead on was studying local polygenic adaptation of stickleback fish populations in Alaska. Using SLiM, I wrote code to run large evolutionary simulations which emulated the geography and evolutionary history of Threespine Alaskan Stickleback. We outlined the rates of gene flow which allowed us to see causal loci from F_{st} scans across the genome as well as the rates at which migration load prevents adaptation. As a first author, the manuscript for this paper, titled *A few stickleback suffice to transport adaptive alleles to new lakes*, is in the final stages of editing and will be posted to the BioArxiv soon, then submitted to Genetics. During my second large research project, I worked with Drs. Ben Haller, Jerome Kelleher, Peter Ralph, and Philipp Messer on the implementation and profiling of *Tree Sequence Recording* in SLiM 3.0. Genealogical tree sequence recording is a strategy for efficiently recording the genealogical history from forward-moving simulations. This history is represented by the forest of trees relating all sampled individuals to each other over every genomic interval. Using TreeSeq, simulations in SLiM can avoid the cost of tracking and propagating neutral mutations as a by-product of obtaining the origins of all sampled genotypes. After successfully implementing *TreeSeq* with rigorous testing, we found simulations where individuals had realistic size genomes experienced a speedup of over 2 orders of magnitude. The paper describing the applications of this strategy, titled *Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes*, was published by Molecular Ecology Resources. My current project working with Dr. Andrew Kern involves using deep learning (neural networks) to infer population genetics parameters from sampled data. For this project, we want to know if the recent advances in deep learning architecture can learn complex patterns in genotype matrices resulting in accurate predictions of recombination and mutation rates. Using my code pipeline, we have found architectures and data prepping heuristics which have resulted in predictions of recombination rates that consistently outperform industry standards such as LD Hat.

The projects described above have exposed me to a wide variety of research environment practices including organization of experiments, scientific writing, and effectively communicating my work. I have gained a familiarity with many concepts in genomics, population genetics, and applied computer science skills such as simulation and machine learning, all of which have laid the foundation for me as a well-versed information scientist. While much of my experience in genomics has been theoretically focused, I aspire to apply my skill set to human health. The applications of genomics in medicine are virtually limitless in areas such as immunotherapy, mental health, gene therapy and more. I believe understanding the polymorphisms in our population and the genetic basis of certain traits is one of the most important aspects of medicine and human health. With a large portion of the problems facing genomics presented in the form of large data, I believe I am in a unique position to approach and help solve these problems. The current advances produced by biology experts along with the flood of data we anticipate has shaped an exciting future for genomic medicine that I am eager to be a part of.