Schraiber, Joshua G.
Research Statement

I am a computational biologist dedicated to understanding the evolutionary forces that shape variation from the molecular to the macroscopic level. I utilize high-powered computing to build theoretical models and develop novel statistical methods that are then tested by rigorous empirical analysis. My work is highly interdisciplinary and I work in a range of model systems, from yeast to pigs. Empirically and theoretically, I am primarily interested in the intersection of population genetics with the evolution of quantitative and complex traits.

Population genetics

Population genetics is the lens through which we look at patterns of genetic variation to make inferences about evolutionary history. Both demography and natural selection leave signatures in the genome, and I work on inferring and disentangling these patterns in high-throughput sequencing data.

Allele frequencies in ancient DNA and experimental evolution. Modern advancements in ancient DNA (aDNA) and experimental evolution have opened a long-shut window into the temporal dynamics of evolution. Because all the information about natural selection acting to change allele frequency is contained in the temporal dynamics of the allele, we can now gain unprecedented insight into the history of alleles subject to natural selection. I am driven to use this data to make inferences about the strength of positive selection in nature, and to ask about the frequency of natural selection acting on standing genetic variation.

Past work: I attacked the mathematical difficulty of analyzing selection by building novel inference strategies based on the probabilities of individual allele frequency paths. In doing so, I quantified the impact of weak natural selection on allele frequency changes and found that weak selection is probably impossible to detect, even with extremely densely sampled time series.

Current work: Building on my theoretical results, I developed a Bayesian method to reconstruct allele age and selection intensity from allele frequency trajectories. I applied this framework to aDNA from humans and horses. Notably, while I found the expected signal of strong positive selection in favor of lactase persistence in Europeans, I also inferred that the lactase allele substantially predates the onset of agriculture, suggesting that selection acted on standing genetic variation.

Future work: Utilizing diversity at linked, neutral sites provides the most powerful tool for detecting recent selection without using aDNA. Inspired by this fact, I will develop a Bayesian method to incorporate linked neutral diversity into inferences from allele frequency time series. This approach fits naturally within the path probability framework that I developed. By conditioning on the path of the selected allele and the genealogy at the linked neutral sites, I will develop a Markov chain Monte Carlo algorithm to improve inferences about natural selection. Using whole genome aDNA data from ancient humans, horses, dogs and pigs, I will apply this method genome-wide to gain insight into the role of natural selection in shaping diversity across a wide variety of species. Because this methodology will make use of all the available information contained in the data, I will be able to definitively answer questions about the frequency of selection on standing variation.

Admixture and natural selection in natural populations. High-throughput sequencing lets us see into the evolutionary history of both model and non-model organisms. Interbreeding among local populations (or even different species) of organisms has important implications for understanding both the evolutionary past, as well as the future in the face of human interference with natural habitats. Similarly, we can gain insight into both past and future by understanding how natural selection has shaped organisms.

Past work: With Martien Groenen's group, I sought to understand the impact of geological and human activity on diversity within and between pig species. Using high-coverage, full-genome data, we found that pigs in island Southeast Asia had experienced multiple periods of secondary contact due to rising and falling sea levels. We also found a substantial role of human translocation in shaping the genomes of wild suids.

Current work: Because of the importance of domestication in shaping human history, I am examining the genomic signatures of domestication in pigs. We find that the process of domestication was diffuse, and that interbreeding between domestic pigs and wild boars was common in both Asia and Europe. Despite the rampant signatures of admixture, we found islands of domestication, including indications of parallel targets of selection in European and Asian domestic pigs.

Future work: Continuing my collaboration with Martien Groenen and Greger Larson, I will utilize aDNA to further refine our understanding of pig domestication. In particular, we will analyze the timescales of gene flow between wild and domestic pigs. Doing so will require novel methodological advances to take full advantage of the available aDNA. Moreover, in collaboration with my postdoctoral co-adviser, Joshua Akey, I will examine demography and admixture in budding yeast. The demographic and selective history of wild yeast is understudied compared to its importance as a model organism in molecular biology. Using a new, full-genome dataset from worldwide samples of yeast, we will analyze population structure and look for signatures of local adaptation, which has only rarely been identified in wild yeasts. We will also look for the genomic signal of domestication in yeast, and compare it to the signal found in multicellular organisms.

Quantitative genetics and complex traits

Many traits closely related to organismal fitness have a complex genetic basis. In the century since Fisher reconciled Mendelian genetics with the observations of biometricians, we have made significant strides in understanding the genetic basis of complex traits. However, we still struggle to understand how evolution at the molecular level maps onto evolution at the phenotypic level. To bridge this gap, I work both within traditional quantitative genetic modeling as well as forging novel methods to jointly model sequence and phenotypic evolution.

Comparative phenomics. Evolution gives us the opportunity to see how closely related species respond to different environmental conditions. By comparing these species phenotypically, we gain insight into the forces that shape evolution at the phenotypic level. I develop novel phylogenetic comparative methods, both to analyze more traditional phenotypes such as body size and to analyze functional genomic phenotypes such as RNA expression.

Past work: One of the great debates in evolutionary biology is the relative importance of punctuated vs. gradual evolution. I co-developed an approach to address this question with comparative data using models known as Lévy processes. With this method, we found evidence that the ancestor of great apes had experienced a burst of brain size evolution. Furthermore, working with Rachel Brem, I developed a method to detect signatures of natural selection in RNA-sequencing datasets. Because these datasets typically contain small numbers of species, I leveraged the power of agglomerating the signal across genes in predefined functional categories. I applied this method and found a strong signature of stabilizing selection acting on gene expression across the *Saccharomyces sensu stricto* species, and several examples of lineage-specific natural selection.

Current work: Working with Joshua Akey, I am generating RNA sequencing data from *Saccharomyces* species and strains subjected to different environmental stresses. I found that a core set of genes act in a conserved stress response pathway; however, each species examined also has a lineage-specific stress response. Interestingly, we find that the genes involved in lineage specific responses show evidence of accelerated evolution at both the level of RNA expression and DNA sequence.

Future work: I will continue to address questions about punctuated evolution. Specifically, I will apply my method to several large comparative datasets, and ask about the frequency of punctuated evolution across a variety of clades. Because the debate about punctuated vs. gradual evolution has largely been one of paleontologists vs. neonatologists, I will augment the method to analyze phenotypic data from fossil taxa. By rigorously analyzing comparative data from both extant and extinct species, I will shed light on long-standing questions about the tempo and mode of evolution at the phenotypic level.

Evolution of DNA sequence and phenotype. I want to explain how natural selection acting at the level of phenotype impacts the variation we see at the molecular level. To approach this question, I have begun to develop joint models of sequence and phenotypic evolution.

Past work: I proposed a model of phenotypic evolution that is explicitly based on mapping mutations onto genealogies at loci that control the trait in question. We found that the details of the mutational effect distribution leave characteristic signatures in the distribution of a quantitative trait, suggesting that is possible to make inferences about mutational effects from quantitative trait variation in natural populations.

Current work: I am conducting research funded by the NSF Postdoctoral Fellowship in Biology with Joshua Akey and Jonathan Wakefield to assay within and between species variation in DNA sequence, RNA expression, and chromatin accessibility in the *Saccharomyces sensu stricto* species. We will be assaying genome sequence, RNA expression and chromatin accessibility in population samples from 5 different *Saccharomyces* species. Ultimately, we will use patterns of polymorphism and divergence to assess the signatures of selection acting at these three distinct levels of phenotype. I am also developing a Bayesian approach to infer the parameters of mutational effects on RNA expression levels. This will allow us to understand the raw mutational input upon which natural selection acts. We will apply this method to analyze several population-level expression datasets to ask if the parameters of mutational input are similar in different taxa.

Future work: Like much work in theoretical population genetics, I began by assuming that quantitative traits evolve neutrally. An important refinement of this model will be to incorporate the effects of weak selection on the trait in question. I will use a perturbative approach based on the ancestral selection graph to understand how weak selection changes the distribution of a quantitative trait in a population. Once this is computed, I will extend the approach I am currently developing with Jonathan Wakefield to infer the parameters of natural selection acting on RNA expression. A full, mechanistic understanding of how selection operates from molecules to phenotypes will require the development of population genetic models that relate sequence evolution to phenotypic divergence. By combining my experience developing models of sequence and phenotypic evolution, I will develop first-principles models that jointly describe the evolution of genotype and phenotype. I will then test these models by applying them to publically available datasets of molecular phenotypes, including assays of chromatin accessibility and RNA expression