Joint inferences of natural selection between sites and populations

Understanding natural selection in populations of humans and other species not only reveals the evolutionary past, but also guides efforts to combat genetic and infectious disease. The input of selected genetic variation is quantified by the distribution of fitness effects (DFE) of new mutations, which plays a key role in determining the genetic architecture of human disease (Di Rienzo 2006) and the tempo at which pathogens adapt (Messer & Petrov 2013b). The DFE can be inferred from contemporary patterns of genetic variation (e.g., Keightley & Eyre-Walker 2007; McVicker et al. 2009), but existing approaches leverage little of the available information and ignore potential correlations between genomic loci and between populations. There is thus a *critical need* for improved methodological and conceptual approaches to inferring the distribution of fitness effects of new mutations.

My *long-term goal* is to develop comprehensive approaches for identifying selection in natural populations and understanding its functional consequences. The *objectives* of this application, which are the next steps toward attainment of that goal, are to develop and apply novel approaches for inferring linked selection within a population and for inferring divergent selection between populations. The *rationale* for the proposed research is that the approaches developed will be broadly applicable, providing a foundation for comprehensive population genomic inferences in humans and other species.

I plan to attain the objectives of this application by pursuing the following two *specific aims:*

Aim 1. Infer linked selection using two-locus statistics.
My group has developed a novel approach for efficiently calculating the statistics of pairs of genetic loci under arbitrary demographic history and selection scenarios (Ragsdale 2016; Ragsdale & Gutenkunst 2017). I propose to apply this approach to quantify several aspects of natural selection in humans and *Drosophila melanogaster*. First, I will use linked neutral variation to infer quantitative models of individual known adaptive variants. Second, I will use linked neutral variation to infer the DFE of new nonsynonymous mutations. Third, I will infer the joint DFE for pairs of new nonsynonymous mutations. Together, these projects will dramatically improve quantitative understanding of natural selection.

Aim 2. Infer joint distributions of fitness effects between populations.
Experiments on modest numbers of mutations have shown that fitness effects often differ between populations (Kondrashov & Houle 1994; Wang et al. 2014), but the broad patterns of such differences are unknown. By introducing the concept of a joint DFE between populations, I propose to quantify divergent selection in humans, *Drosophila melanogaster*, and *Daphnia pulex*. I hypothesize that the degree of divergent selection will vary dramatically with gene function and with population divergence, offering insight into mechanisms of speciation and local adaptation.

The *expected outcomes* of the proposed research are new population genetic inference methods and insights into natural selection in humans and two model organisms. These outcomes are expected to have important *positive impact* on the field of population genetics. The methods will be widely applicable and well-supported, and the results will provide a baseline for future population genomic inferences. More broadly, the methods and resources developed in this research will provide a foundation for quantifying selection in populations of humans and human pathogens, with application to combating genetic and infectious disease.