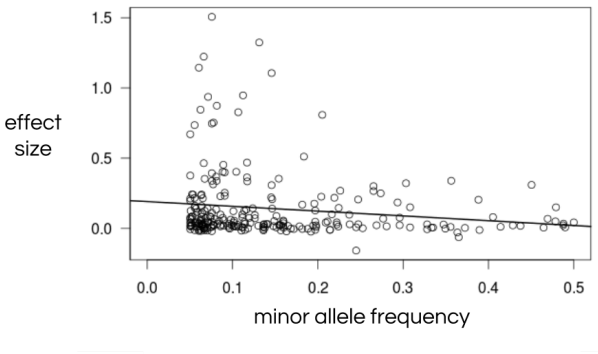
I study the **population genomics of quantitative traits** and seek to understand the evolutionary forces shaping quantitative trait variation. In particular, I incorporate population genomics into studies of quantitative traits by investigating how selection has acted on the loci controlling trait variation, an approach that is now feasible on a genome-wide scale. Systematically describing the evolutionary forces that maintain variation for quantitative traits is important not only for our understanding of evolutionary biology, but has important consequences for ecology, medicine, and agriculture.

My lab will investigate the population genomics of quantitative traits at two levels: variation within populations and adaptive divergence between populations by generating and interpreting genomic data from wild plant systems and developing new methods to analyze publicly-available crop plant datasets. Below I discuss three main research areas and describe past accomplishments, current work, and future directions within these areas.

1. What evolutionary forces maintain genetic variation for complex traits within populations?
Genetic variation for traits persists within populations despite the expectation that widespread stabilizing selection should remove variation. This variation could either be neutral, under purifying selection and maintained by mutation-selection balance, or under balancing selection. While understanding the maintenance of genetic variation within populations has been a long-term goal of evolutionary biology, there has been little systematic investigation of these questions on a genome-wide scale. Below, I describe my work developing *Capsella grandiflora*, an obligately-outcrossing Brassicaceae species, into a model for understanding within-population variation.

In my Ph.D., I investigated the forces maintaining genetic variation for traits by conducting a genome-wide association mapping project for gene expression in a single population of *C. grandiflora*. I mapped *cis*-regulatory expression quantitative loci (eQTLs) for ~6,000 genes, allowing me to investigate selection across thousands of traits. The allele frequencies of eQTLs were significantly rarer than those expected based on carefully matched permuted data, suggesting that these variants are under purifying selection. As would be expected under purifying selection, the effect size and minor allele frequency of eQTLs were negatively correlated, even when controlling for ascertainment biases. In total, my results demonstrate that purifying selection generally acts on the genetic variants affecting gene expression, consistent with the explanation that gene expression variation within this population is maintained under mutation-selection balance. My study was among the first to systematically demonstrate that selection acts on a large number of loci that control gene expression variation within a natural population. (Josephs et al. 2015)



eQTL minor allele frequency is negatively correlated with effect size. Each circle is an eQTL.

I am expanding my efforts to understand the selective forces acting on genetic variants for trait variation by incorporating new mapping approaches. For example, I have mapped whole coexpression networks, allowing me to find loci that affect *trans*-regulation of expression in hundreds to thousands of other genes. I am using this data to address the following questions: What selective pressures act on genetic variants that have large and/or pleiotropic effects? Do these differ from the purifying selection I observed on *cis*

eQTLs? I am testing these large effect *trans*-eQTLs for enrichment in population genomic measures of balancing selection to see if there is a significant role of balancing selection in maintaining variation at large-effect *trans*-regulatory elements.

Balancing selection may not always leave clear signatures in sequence data, so understanding the role that balancing selection plays in maintaining variation will require investigating specific mechanisms of balancing selection, such as temporally-fluctuating selection. **In my lab, I will continue my work investigating the role of balancing selection in maintaining genetic and phenotypic variation within a focal population of *C. grandiflora* by generating a time series of genomic sequence data from this population**, complementing the dataset that I generated in my Ph.D. Directly investigating temporal changes in allele frequency, including at previously identified eQTL, has the potential to reveal the effects of both directional and balancing selection at the sequence and trait level.

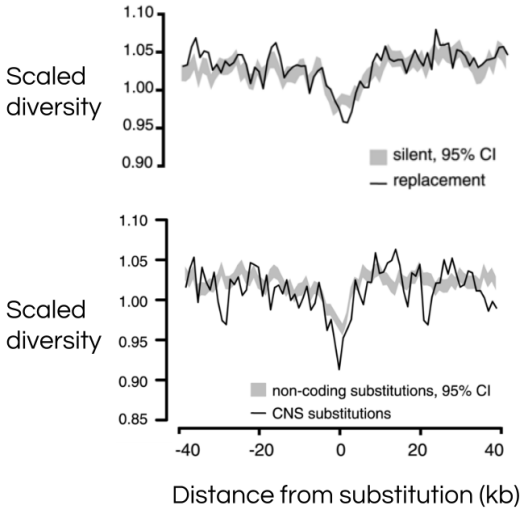
2. What traits are involved in polygenic adaptation across space and time?
Adaptation in quantitative traits often occurs through polygenic adaptation: small shifts in the allele frequencies of many loci that are not detectable through standard selection scans. Polygenic adaptation may be especially common in domestication, where a number of quantitative traits have been under selection, both consciously and unconsciously. I am developing and applying methods for identifying polygenic adaptation in publicly available genomic datasets in maize, allowing me to detect specific traits that show evidence of polygenic adaptation across the landscape of domesticated maize and in time, throughout maize breeding.

While polygenic adaptation is difficult to detect using standard population genetic methods like Fst, it leaves valuable information in the covariance between allele frequency and the direction of effect on the trait. This covariance has been used to successfully develop methods that leverage genome-wide association study (GWAS) results to detect polygenic adaptation. However, current methods for detecting polygenic adaptation may be led astray when GWAS are conducted in populations with significant population structure, as is the case in maize GWAS and for many other plant systems. Not controlling sufficiently for population structure in the GWAS will lead to severely skewed signatures of selection but the loci discovered in a GWAS that does control for population structure will have different allele frequency distributions than the underlying neutral background, potentially creating false signals of selection. I am currently working to develop methods that are robust to structure-related complications.

While these methods will be applicable to a number of traits and systems, I am particularly interested in investigating polygenic adaptation in plastic traits. Many organisms respond dynamically to their environments through plasticity. Evolutionary biologists have long been interested in understanding if, and how, selection shapes plastic responses, but few have sought to systematically detect adaptative plasticity. The many GWAS conducted on crop lines grown in multiple years, locations, and environments provide an ideal data source for investigating these questions. **My research group will apply the approaches I am developing to detect polygenic adaptation to detect adaptive divergence in plasticity in a number of publicly available datasets from domesticated crop species.**

3. How do positive and negative selection shape genomic sequence variation and divergence?
Genetic variation and divergence in traits within and between species is ultimately based in genetic sequence variation and divergence, so understanding the forces that shape sequence variation is crucial. I have made two significant contributions in this area:

Little is known, especially in plant genomes, about the relative strength of selection in noncoding regions compared to coding regions or the effects of linked selection on genetic diversity. To address these gaps, I used population genomic data from *C. grandiflora* to test for the signature of selective sweeps in the *Capsella* lineage in coding and conserved non-coding sequence. Neutral diversity was significantly reduced around fixed replacement substitutions and fixed substitutions in conserved noncoding sequence, consistent with recurrent selective sweeps. These results imply that plant populations can experience high rates of positive selection in both coding and non-coding regions. (Williamson*, Josephs* et al. 2014)



Expression level is often negatively correlated with sequence divergence, but the selective forces responsible this correlation were unknown. I measured expression in *C. grandiflora* leaves using RNAseq and compared the strength of selection acting on genes with different expression levels using population genomic methods. While positive selection did not vary between expression classes, negative selection was significantly stronger in high expression genes. In addition, I collaborated with Ramesh Arunkumar to show that positive and negative selection are stronger in pollen-expressed genes compared to other genes (Arunkumar, Josephs et al. 2013). These results demonstrate that variation in rates of selection across the genome can be predicted to some extent by where and how much genes are expressed and that the negative correlation between expression and sequence divergence is driven by negative selection. I have further shown that negative selection is stronger in genes that show higher coexpression network connectivity and genes that lack local regulatory variation (Josephs et al. *in prep*, Williamson*, Josephs* et al. 2014; Arunkumar, Josephs et al. 2013)

There is still much to be done to fully understand the evolutionary forces that shape genome-wide patterns of sequence variation. In particular, linked selection has dramatic consequences for sequence variation (Josephs & Wright 2016) but we still lack a complete view of its importance in a wide range of species or clear methods for disentangling the action of linked selection from demography. **My lab will continue to investigate the evolutionary forces shaping genomic sequence variation in both wild and domesticated plant species.**