Background: Genomic data can be used to explore classic genetic questions in previously impractical ways. For example, we can directly test genetic mechanisms of evolutionary processes, like hybridization. Differential expression between hybrids and their parents has been analyzed across systems, particularly those with hybrid vigor (ex: yeast, poplar, wheat, maize, detailed in [1]). At the extreme is single parent expression (SPE). For a gene to be characterized as demonstrating SPE, it must be expressed within a hybrid and only one parent. SPE has been reported in maize [2,3] and wheat [4] but likely occurs more broadly. Previous analyses have not assessed SPE within a phylogenetic context, revealed its phenotypic impacts, or demonstrated a definitive mechanism. Genomic data can shed light on these questions.

One mechanistic model I will test using genomic data is complementation. In this model, genetic drift fixes distinct slightly deleterious, recessive alleles in parental populations undergoing demographic bottlenecks. When differentiated parents hybridize, non-deleterious alleles mask the effects of deleterious alleles [1]. Deleterious alleles could be in regulatory or genic regions, affecting expression. Single nucleotide polymorphisms (SNPs) or presence/absence variants (PAVs, sequence present in one individual but absent in another) may identify these alleles.

Zea mays ssp. *mays* (maize) is the best study system to explore this question. Specific populations of maize have undergone bottlenecks during post-domestication range expansion and modern breeding. High genetic diversity is retained at the species level (on average, a SNP every 80 base pairs and an indel every 300 base pairs [6]), but this diversity is structured within populations due to genetic drift during bottlenecks. Wide crosses can be made between maize from distinct populations, leading to SPE in hybrids. Maize hybrids have substantially different phenotypes from their inbred parents, often demonstrating increased vigor (*i.e.*, heterosis). Hybrids of different parental combinations are also phenotypically distinct [5]. Heterotic patterns and distinct phenotypes are potentially linked to SPE. As a model system, established genomic resources needed for careful, detailed study of SPE are available, including RNA-seq and whole genome sequences. **Research Question:** Can SPE patterns be explained by the complementation model? Does expression complementation correlate with distinct phenotypes seen in hybrids?

Research Plan: **Aim 1: Assess the extent of SPE in reciprocal crosses spanning a range of divergence times:** Previous work identifying genes displaying SPE in maize has focused on crosses in temperate, modern breeding lines--specifically B73 and Mo17 [3]. While a recent study expanded to crosses with 6 lines [2], no tropical or more divergent lines were included, limiting the phylogenetic range of SPE. I am currently generating diallelic reciprocal crosses of 4 tropical and 4 temperate lines stratified across the broad diversity of maize (Fig. 1). These lines are part of the maize Nested Association Mapping (NAM) population, a highly utilized resource that has revealed the genetic architecture of a number of maize traits. By completing reciprocal crosses, I

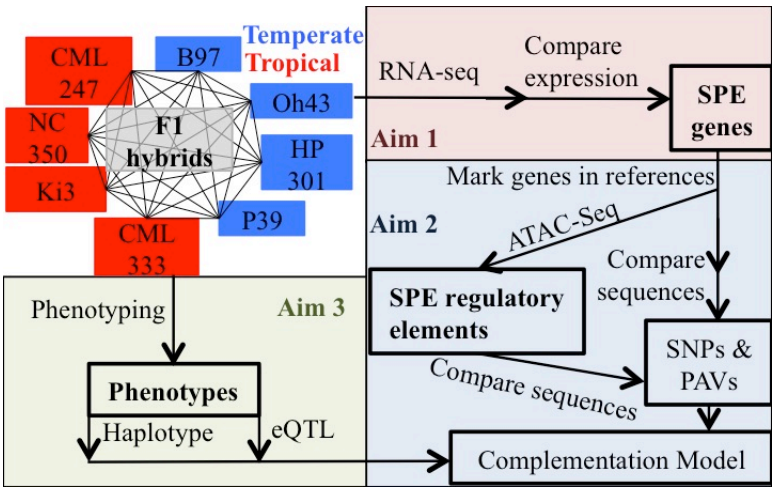

Figure 1: Overall work flow of the study

can separate maternal and paternal effects from true SPE patterns. I will measure gene expression through RNA-sequencing (RNA-seq) of root, shoot, and pollen tissues of F1 hybrids; parental expression has already been measured. I will identify instances of SPE by comparing the expression profiles of F1 hybrids to parents using previously described pipelines [2]. I predict SPE will be detected and will increase with genetic divergence between crossed lines.

Aim 2: Evaluate the complementation model to explain SPE: Previous estimates of the frequency of PAV underlying SPE (~5%) [2] are likely an underestimate since only sequences present in B73 but absent from Mo17 could be identified. The contribution of both coding and regulatory SNPs also remains unclear. Both PAV and SNPs linked to SPE genes can be identified using comparative genomic tools and parental reference genomes (Fig. 1). As a member of the Hufford lab, I will have access to high quality, *de novo* assemblies of parental lines for my crosses, allowing detection of variants underlying SPE. Previously generated open chromatin profiling (ATAC-seq) data sets for the parental lines will delineate regulatory regions associated with these genes. As done in *Arabidopsis* [7], I will search for differences in the accessibility and sequences of regulatory regions upstream of SPE genes. I will test the complementation model by evaluating whether identified SPE genes in the F1 are enriched for PAV and SNPs distinguishing parental lines.

Aim 3: Linking SPE complementation to phenotype: I expect lines with more instances of SPE complementation (Aim 2) to have a greater extent of F1 heterosis. I will generate as well as utilize published phenotypic data sets of F1 hybrids to estimate trait heterosis (yield, plant height, etc.). An expression QTL (eQTL) analysis will indicate if SPE genes are associated with trait heterosis. I will use existing HapMap data from the primary heterotic groups to search for selection signatures in the regions around SPE variants (Aim 2; Fig. 1). Heterotic groups were selected for combining ability and also likely for complementation.

Intellectual Merit: Evaluating genetic models to explain SPE provides an example of how genomic data can be leveraged to address long-standing questions in genetics and explain phenotypic observations. By using crosses spanning a range of divergence times, this study evaluates the evolutionary processes responsible for SPE patterns and directly tests the complementation model. Studying the mechanism of SPE and heterosis has obvious implications for hybrid maize genetics. However, such findings could be applicable across a wide range of systems experiencing hybridization after inbreeding including plants and animals. By building on the computational foundations of maize resources, this study provides a template for similar studies in other systems.

Broader Impacts: This funding will support undergraduate research experiences at Iowa State University through fieldwork, RNA extractions, and computational analyses proposed here. Funding will also support my future collaboration with researchers at CIMMYT in Mexico. While heterosis has been heavily studied in maize, heterosis is exploited in rice, tomatoes, and numerous other crops [1]. Understanding how heterosis is controlled has the potential to impact each of these crop systems and our ability to generate resilient, high-yielding lines.

References: [1] Birchler, J.A., et. al. 2010 *Plant Cell* 22(7):2105-2112. [2] Baldauf, J. A., et. al. 2018 *Current Biology* 28(3):431-437. [3] Stupar, R. M. and Springer, N. M. 2006 *Genetics* 173(4):2199-2210. [4] Sun, Qixin, et. al. 2004 *Plant Science* 166(3):651-657. [5] Birchler, J. A. 2003 *Plant Cell* 15(10):2236-2239. [6] Springer, N., et al. 2009 *PLoS Genetics* 5(11):e10000734. [7] Alexandre, C., et al. 2018 *Molecular Biology and Evolution* 35(4):837-854.