Project Summary

Overview The proposed project will investigate the genome-wide effects of hybridization and introgression in the genus *Zea*. First, investigators will study incipient speciation between the lowland-adapted *Z. mays* ssp. *parviglumis* (hereafter, *parviglumis*) and the highland-adapted *Z. mays* ssp. *mexicana* (hereafter, *mexicana*). Through field collections, genotyping and common garden studies, the investigators will assess what fraction of the genome is porous to gene flow in hybrid zones, how fitness of these taxa varies across a hybrid zone, and how allele frequency clines are patterned at adaptive loci. Second, investigators will determine the impact of hybridization and introgression between domesticated maize (*Z. mays* ssp. *mays*) and wild *Zea*. Population genomic analyses of sympatric collections will be used to assess whether maize served as a bridge for gene flow between otherwise allopatric *Zea* species and whether maize received gene flow from wild relatives that facilitated its adaptation to new environments.

Intellectual Merit Much progress has been made in the study of hybridization and introgression through the development of theory, through field-based ecological research, and through genetic analyses based on a limited number of molecular markers. However, much remains to be discovered regarding how these evolutionary processes have shaped genomes. The research proposed here will leverage the genomic resources of the maize model system to investigate how hybridization and introgression have molded the genomes of both wild *Zea* species and domesticated maize on two different timescales: 1) An evolutionary timescale covering 60,000 generations of divergence between *parviglumis* and *mexicana*; and 2) An ecological timescale in which maize has spread across the Americas and adapted to local conditions. The analysis on an evolutionary timescale will generate basic knowledge on the process of incipient speciation and the porous nature of the genomes of diverging species, whereas the analysis on an ecological timescale can inform, for example, the study of biological invasions and the role of introgression in facilitating rapid local adaptation.

Broader Impacts The investigators will achieve societally relevant outcomes in the proposed project by providing STEM training opportunities for undergraduate and graduate students, participating in Iowa State Universitys GK12 Fellowship program, and establishing an exchange program between universities in the United States and Mexico. This project will provide ample training opportunities for both undergraduate and graduate students in laboratory, computational, and field-based research. The investigators have successfully recruited minority students into their research programs in the past and will make every effort to do so as part of this proposed work. The graduate student funded to work at Iowa State University will participate in the University's GK12 program that serves the Des Moines public school system. Through this opportunity, the graduate student will bring their research on maize evolution into the classrooms of the most diverse public school system in the state of Iowa. The *Zea* study system provides an excellent opportunity to deliver evolution training to middle and high school students in a state dominated by maize agriculture. Finally, the proposed exchange program would create an opportunity for students from the United States to conduct research internationally and allow these students to interact with visiting students from Mexico. Through these interactions, students will be better prepared for modern STEM research, which is often highly collaborative and international in nature.

Project Description

Introduction

While the potential roles of hybridization and introgression as agents of evolution have long been appreciated (Anderson, 1948; Anderson and Stebbins, 1954; Stebbins, 1959), only recently have technological innovations allowed for characterization of these processes on a genome-wide scale. Multiple studies have now reported substantial inter-taxon introgression in both plant (Hufford et al., 2013; Renaut et al., 2013) and animal (Consortium, 2012; Staubach et al., 2012; Huerta-Sánchez et al., 2014) species, and introgression has been found across considerable portions of genomes and specifically at loci thought to underlie adaptation.

In the investigation proposed here, we will build upon our previous work in maize (*Zea mays* ssp. *mays*) and its wild relatives the teosintes (*Zea* spp.) to characterize the evolutionary role of hybridization and introgression. As a study system, the genus *Zea* is ideally suited for such research due to its large natural populations and the availability of exceptional genomic resources (Hufford et al., 2012b). Within *Zea* we will be able to assess the genome-wide effects of hybridization and introgression at two different timescales. First, analysis of hybridization and divergence between the subspecies *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *mexicana* will generate basic knowledge of the process of incipient speciation and the porous nature of genomes of diverging species on an evolutionary timescale ($\sim 60,000$ generations). Second, an evaluation of introgression in sympatric maize and teosinte populations on the ecological timescale ($\sim 3,000 - 9,000$ generations) during which domesticated maize colonized the Americas will inform our understanding of the role of introgression in adaptation to new or rapidly changing environments.

Objectives

Objective I: Assess the evolutionary and genomic effects of hybridization in locally-adapted, parapatric wild teosinte

Zea mays ssp. *parviglumis* (the wild progenitor of maize; hereafter, *parviglumis*) and *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) diverged approximately 60,000 BP (Ross-Ibarra et al., 2009) and have parapatric distributions: while *parviglumis* occurs in the warm lowlands of southwest Mexico, *mexicana* is found in the cool highlands of the Central Plateau. Narrow regions of admixture between these wild subspecies have been discovered at middle elevations (Fukunaga et al., 2005; Pyhajarvi et al., 2013). Through targeted collections, generation of high-density genotyping data, population genomic analyses, and common garden experiments, we will address the following research questions:

A. *What fraction of the genome is porous to gene flow in hybrid zones?*

B. *How do the fitness of parental and hybrid populations vary across the hybrid zone?*

C. *Are allele frequency clines at loci associated with fitness traits in* mexicana *and* parviglumis *steeper than those at non-associated loci?*

Objective II: Determine the extent to which hybridization and introgression have altered the *Zea* genus during the post domestication spread of maize

Maize was domesticated in southwest Mexico from *parviglumis* ~9,000 BP (Matsuoka et al., 2002) and quickly spread throughout the Americas, bringing this crop into sympatry with new species of teosinte (Vigouroux et al., 2008). Through a combination of dense genotyping of range-wide samples of maize and teosinte and targeted, full-genome sequencing, we will assess three questions regarding the importance of introgression during the spread of maize:

A. *Was the spread of maize facilitated by gene flow from locally-adapted wild* Zea*?*

B. *What is the geographic scale of adaptive introgression?*

C. *Did maize serve as a bridge for gene flow between previously isolated* Zea *taxa?*

Rationale and Significance

Pioneers in evolutionary biology including G. Ledyard Stebbins and Edgar Anderson recognized the important role hybridization and introgression could play in adaptation and speciation (Anderson, 1948; Anderson and Stebbins, 1954). These evolutionary forces were thought to be particularly influential when environmental conditions encountered by a species were marginal, variable, or new (Stebbins, 1959). More recently, theoretical and empirical investigations of hybridization have focused on hybrid zones, defined as regions where distinct taxa co-occur and mate, resulting in progeny of mixed ancestry (Harrison, 1993).

While hybrid zone theory has progressed and many compelling empirical examples have been identified based on hybrid morphology and genetic marker data (Delmore et al., 2013; Galindo et al., 2013; Parchman et al., 2013; Smith et al., 2013a), several outstanding questions remain. Additional study is needed to determine whether hybrid zones are primarily maintained as tension zones in which hybrids are selected against or as ecotones where hybrids have an advantage under certain environmental conditions (Kruuk et al., 1999; Rasmussen et al., 2012; Smith et al., 2013b). Moreover, genome-wide analysis of the fraction of the genome that is porous to gene flow in hybrid zones is rare and will likely offer considerable insight. For example, recent genomic studies of introgression suggest that rates of gene flow vary substantially across loci, likely as a function of selection for or against introgressed alleles (Hufford et al., 2013; Poelstra et al., 2014). In addition, chromosomal rearrangements (*e.g.*, inversions and translocations) likely play an important role in adaptation and structuring introgression along the genome and may restrict gene flow in hybridizing species (Barb et al., 2014; Guerrero and Kirkpatrick, 2014).

Both wild and domesticated *Zea* offer exciting opportunities to study hybridization. The subspecies *parviglumis* and *mexicana* are distributed across a steep altitudinal gradient and differ for traits that are thought to be adaptive in the highlands such as the presence of macrohairs and stem pigmentation. A recent ecological niche study has found that distributions of these subspecies are quite unique and stable over many thousands of years (Hufford et al., 2012a). However, analysis of microsatellite markers genotyped in a range-wide sample has identified elevated admixture between the subspecies in two geographically distinct mid-elevation regions of Mexico between the distributions of both parental species, suggesting the presence of multiple hybrid zones (Fukunaga et al., 2005). Our recent genome-wide analysis of a population in one of these hybrid zones revealed extensive subspecies admixture across all individuals sampled and relatively short shared haplotypes with individuals from other populations of either *parviglumis* or *mexicana* (Pyhajarvi et al., 2013). This suggests continual gene flow between *parviglumis* and *mexicana* in this hybrid zone over a substantial period of time (Pyhajarvi et al., 2013). Longer shared haplotypes were found in the hybrid population in chromosomal regions identified as potential inversions (Pyhajarvi et al., 2013). These regions may be particularly resistant to gene flow between subspecies, but may also prove adaptive for individuals in some parts of the hybrid zone. Findings from this single hybrid population suggest that gene flow between taxa is ongoing yet varies in extent across the genome. However, very little is known regarding genome-wide patterns of hybridization and subsequent introgression in other populations in this region or in other hybrid zones. Hybrid populations of *mexicana* and *parviglumis* are distributed across elevation gradients in markedly different regions of Mexico and are found at varying distances from each subspecies. By expanding our study of hybridization and introgression across these taxa we can compare evidence supporting ecotone and tension zone hypotheses and, should hybridization prove adaptive, assess how the genomic landscape of hybridization is shaped by natural selection.

In addition to the ongoing hybridization between *parviglumis* and *mexicana* occurring since divergence ∼60,000 BP, gene flow between domesticated maize and various taxa of the genus *Zea* has been detected based on both hybrid morphologies observed in the field (Wilkes, 1967, 1977) and genetic data (Fukunaga et al., 2005; Ross-Ibarra et al., 2009). Maize domestication from *parviglumis* occurred recently on an evolutionary timescale (∼9,000BP Matsuoka et al., 2002)) and was followed by rapid spread of the crop across the Americas over the following millennia (Piperno and Flannery, 2001; Grobman et al., 2012). During this diffusion, maize was brought into sympatry with new wild relatives that were likely allopatric to the progenitor of maize (*i.e., parviglumis*) for long periods prior to domestication (Hufford et al., 2012a). Our recent work has provided evidence of introgression from *mexicana* into maize during its earliest colonization of the highlands of the Mexican Central Plateau. We found consistent introgression into several highland maize populations at QTL for phenotypes (*e.g.*, pigment and macrohairs) that distinguish highland *mexicana* from lowland *parviglumis* teosinte, and showed differences in these phenotypes as well as growth rate under cold conditions between maize plants with and without *mexicana* introgression (Hufford et al., 2013). Our interpretation of these results is that maize received adaptive introgression from *mexicana* that allowed the crop to spread into the highlands of Mexico.

Subsequent to its diffusion into the Mexican highlands, maize spread into sympatry with additional teosinte taxa in Guatemala including *Zea luxurians* (hereafter, *luxurians*) and *Zea mays* ssp. *huehuetenangensis* (hereafter, *huehuetenangensis*), each adapted to environmental conditions very different from those of *parviglumis*. Although maize is known to hybridize with both taxa, the extent and adaptive significance of gene flow between these teosintes and maize is unknown. Based on analysis of a small number of resequenced loci, *mexicana* haplotypes appear to be segregating in *luxurians* (Ross-Ibarra et al., 2009). Since *mexicana* and *luxurians* are entirely allopatric in their distributions, this suggests maize may have served as a bridge for gene flow between these two taxa. Further work will be necessary to explore this possibility and to assess if maize has, more generally, altered the genomes of *Zea* species through gene flow during its spread across the Americas (see **Objective II**).

Preliminary Results

Our previous publications suggest *Zea* is a promising model system for exploring the evolutionary role of hybridization and introgression (*e.g.*, Ross-Ibarra et al. 2009; van Heerwaarden et al. 2011; Hufford et al. 2013; Pyhajarvi et al. 2013). To further refine our research questions and provide preliminary results for this proposal we have reanalyzed published data (van Heerwaarden et al., 2011; Fang et al., 2012) of 983 SNPs genotyped across a panel of $> 2,000$ samples including all subspecies and species of teosinte and an Americas-wide sample of maize landraces (*i.e.*, ancient farmer varieties of maize). While the low density of the markers in this data set precludes genome-wide inferences and haplotype-based analyses, the comprehensive taxon sampling makes this an ideal resource for guiding future research.

Evidence for hybrid zones between *parviglumis* and *mexicana*

We have assessed evidence for admixture between *parviglumis* and *mexicana* using the range-wide sampling of hundreds of populations in the 983-SNP data set. The probability of each sample's assignment to *parviglumis* and *mexicana* groups was calculated using the program STRUCTURE (Pritchard et al., 2000). We find that individuals from several mid-elevation populations show appreciable assignment to both *parviglumis* and *mexicana* groups (green underscored individuals in Figure 1) and likely represent hybrid populations.

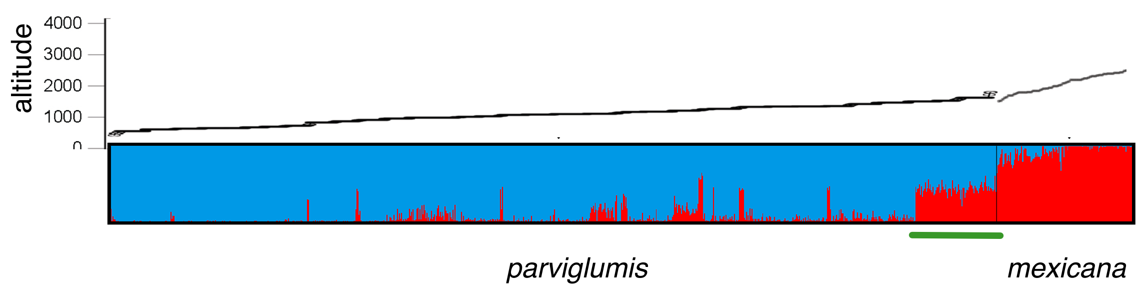


Figure 1: Assignment of *parviglumis* and *mexicana* individuals to K=2 groups using the Bayesian assignment algorithm of STRUCTURE (Pritchard et al., 2000). Individuals are sorted by increasing altitude as indicated by the plot above the bar chart. Individuals from mid-elevation, hybrid zone populations are underscored in green.

Several admixed populations cluster in two geographically distinct regions of Mexico: the eastern Balsas River Basin and eastern Jalisco state. These locations fall at intermediate locations between the main distributions of *parviglumis* and *mexicana* (Panel A, Figure 2). Hybrid populations from eastern Jalisco state are found at higher elevation (average elevation = 1632m) than hybrid populations in the eastern Balsas (average elevation = 1531m) and also show a higher proportion of membership in a *mexicana* (*i.e.*, highland teosinte) group (Panels B and C, Figure 2). These findings suggest that hybrid populations from distinct environments may vary in proportion of ancestry from these two subspecies in a manner that is adaptive. Estimates of pairwise population differentiation also suggest that hybrid populations in the Balsas and Jalisco are distinct in that Jalisco populations are less differentiated from *mexicana* than hybrid populations in the Balsas (Table 1). Not surprisingly, populations in both hybrid zones are less differentiated from *mexicana* and *parviglumis* than these subspecies are from each other.

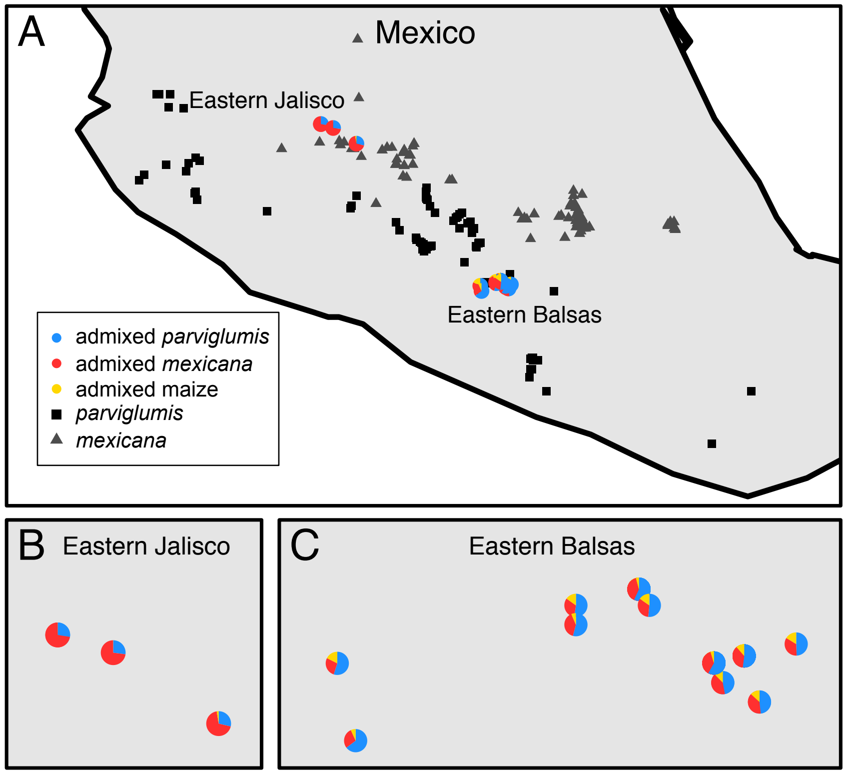


Figure 2: A) Location of two putative hybrid zones of *mexicana* and *parviglumis*. Hybrid populations are represented as pie charts with proportion assigned to *mexicana*, *parviglumis*, and maize groups. Zoomed-in views of the Eastern Jalisco (B) and Eastern Balsas (C) hybrid populations.

Evidence for hybridization across the genus *Zea*

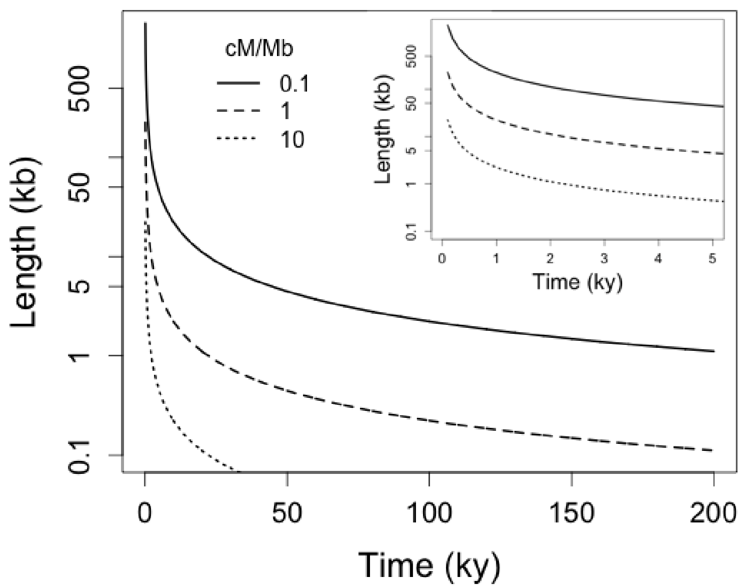
TreeMix Results

Talk about ILS figure here: (Figure 3)

Table 1: Pairwise F_{ST} between teosinte and hybrid populations

Taxon	***parviglumis***	***mexicana***	**Jalisco Hybrids**	**Balsas Hybrids**
parviglumis	—	—	—	—
mexicana	0.107	—	—	—
Jalisco Hybrids	0.059	0.064	—	—
Balsas Hybrids	0.057	0.074	0.034	—

Figure 3: Effect of recombination on the expected length of a shared chromosome segment vs. number of years since divergence or introgression. Shown are three levels of recombination roughly representing high, average, and low recombination regions of the genome.



Research Plan

Objective I: Hybrid Zones

A) What fraction of the genome is porous to gene flow in hybrid zones?

Recent empirical investigations have suggested that functional architecture of genomes can lead to islands of speciation (*i.e.*, regions of high differentiation) in hybridizing species (Renaut et al., 2013), but that these regions are not always the same across hybrid populations (Parchman et al., 2013). Currently, very few studies have dissected the genome-wide architecture of hybridization and introgression in replicate hybrid zones and very little is known about the consistency of genome porosity to gene flow. Genome-wide studies in teosinte are feasible at very high marker density (Hufford et al., 2012c, 2013; Pyhajarvi et al., 2013) and are also informed by the genomic resources of maize (Hufford et al., 2012b), often providing detailed functional annotation for loci of interest, a rarity in other natural systems. We will assess the genomic architecture of hybridization in two putative hybrid zones of *mexicana* and *parviglumis* through careful collections and assembly of a study panel, generation of genome-wide marker and sequence data, and application of recently-developed population genomic analyses appropriate to this question.

Panel Construction and Sample Collection: From previous collections, we have access to an extensive sampling of *parviglumis*, *mexicana*, and maize from throughout their respective ranges. Moreover, Senior Personnel Luis Eguiarte and collaborator Salvador Montes-Hernandez (see attached letter of commitment) have recently collected altitudinal transects of *parviglumis* and *mexicana* that extend through both hybrid zones targeted by our project (Díez et al., 2013) and are very familiar with populations in this region. Our current collections will likely not be sufficient for both the genotyping and common garden activities we propose here and we have therefore budgeted for a collection trip during the first year of the project. We will collect from 15 sampling sites in each of 16 populations. Four populations will be sampled from both the eastern Jalisco and eastern Balsas hybrid zones. To the extent possible, we will select hybrid populations in a stratified manner across the elevation gradient found in these regions. Four populations each will also be collected from non-admixed *parviglumis* and *mexicana*, with two populations of each taxon collected from regions proximate to both hybrid zones. We have already obtained the requisite collection permits as well as permits for importing samples into the United States for genetic analysis. Following collection, samples will be divided with half being sent to the lab of Dr. Eguiarte at the Universidad Nacional Autónoma de México (UNAM) for common garden experiments (see Research Question IB) and half being sent to Iowa State University for DNA isolations and subsequent genotyping and full-genome sequencing.

Sample Genotyping and Sequencing: DNA for genotyping will be isolated using a modified CTAB procedure (Saghai-Maroof et al., 1984). For sample genotyping ($n = 240$) we will utilize a reduced representation approach to next-generation sequencing called Genotyping By Sequencing (GBS; Elshire et al. 2011). To date, this method has been implemented to genotype $> 45,000$ maize samples and a bioinformatics pipeline (TASSLE-GBS) has been constructed that allows for genotyping $\sim$1,000,000 SNPs in maize (Glaubitz et al., 2014) using standard GBS data. We have already successfully applied GBS to heterozygous maize and teosinte. We find that, even after filtering for missing data, GBS provides many more markers with minimal ascertainment bias at a fraction of the cost of other available technologies.

In addition to GBS data, we will generate full-genome sequence through the Iowa State University DNA Facility for a single hybrid individual from each hybrid zone. We will generate two lanes of Illumina HiSeq (Rapid Mode, 150 cycles) 150bp, paired-end data per individual. We have developed a bioinformatic pipeline for mapping reads to the existing maize B73 reference genome; this approach has proven successful in identifying variants in teosinte (Chia et al., 2012).

Population Genomic Analyses: We will assess the genome-wide patterns of porosity to gene flow in hybrid zones using both standard and newly developed population genomic analytical methods that are appropriate to this question. Standard measures of differentiation and introgression will include F_{ST}, the proportion of shared and fixed variants in hybrid versus reference allopatric populations, Reich's f statistics (Reich et al., 2009), and genome-wide patterns of linkage disequilibrium. We will also implement recently-developed, haplotype-based methods for detecting hybridization and introgression (*e.g.*, Price et al., 2009; Lawson et al., 2012) that will effectively allow us to model chromosomes from hybrid populations as mosaics of reference allopatric populations of *parviglumis* and *mexicana*. We will assess excess of *parviglumis* or *mexicana* ancestry on a site-by-site basis across hybrid genomes at the population level and will determine whether patterns are conserved across populations within hybrid zones and between hybrid zones. Chromosomal regions showing an excess of ancestry from one taxon in hybrid populations will be inspected for evidence of selection using a combination of site-frequency-, linkage-disequilibrium-, and population-differentiation-based methods (reviewed in Vitti et al., 2013). Chromosomal regions showing strong evidence of selection across individuals within a hybrid zone based on analysis of GBS data will be further dissected using high-density, full-genome data generated for a single individual per hybrid zone. Our full-genome data will facilitate identification of the actual causal polymorphisms (*e.g.*, non-synonymous variants) under selection in hybrid zones.

B) How do the fitness of parental and hybrid populations vary across the hybrid zone?

In order to assess fitness and variation at putatively adaptive traits across both non-admixed and hybrid populations we will conduct common garden experiments in Mexico at three altitudes: 1) Below a hybrid zone in habitat occupied by non-admixed *parviglumis*; 2) Within hybrid zone habitat; and 3) Above a hybrid zone in habitat occupied by non-admixed *mexicana*. Ideally, we would replicate these experiments in transects spanning both hybrid zones. However, we have discussed with our collaborators in Mexico (Ruairidh Sawers and Salvador Montes-Hernandez; see attached letters of commitment) the feasibility of managing six concurrent gardens as well as safety concerns for our students in the state of Guerrero (the location of the eastern Balsas hybrid zone) and have decided we must only propose a single transect of three gardens in the eastern Jalisco hybrid zone. These gardens will be replicated over years two and three of our proposed project. We will continue to work with our collaborators to identify appropriate sites. In our initial discussions with Drs. Sawers and Montes-Hernandez we have identified potential high- and low-elevation sites near Celaya and Bucerias, Mexico respectively. We will explore options for our third garden in the hybrid zone during our collections in the first year of the project. Each garden will consist of three complete blocks including a randomization of three plants from each of 15 sampling sites in the 16 populations described in Objective IA (3 blocks x 3 plants x 15 sites x 16 populations = 2,160 plants per site). Our experiments will gauge relative fitness between *parviglumis*, *mexicana*, and hybrids at each of these sites. We will measure fitness-related phenotypes including percent germination, germination rate, plant height at 15-day intervals, seed set, 100-seed weight, total-above-ground biomass, stomatal conductance and survival as well as putatively adaptive traits across the altitudinal gradient including macrohair density, pigmentation extent, and flowering time. Below we list anticipated results from our common garden experiment supporting either ecotone or tension zone dynamics:

Expected outcomes supporting ecotone dynamics

- Plants from hybrid zones will show higher fitness in hybrid-elevation gardens
- Non-admixed *parviglumis* and *mexicana* will show higher fitness in low- and high-elevation gardens respectively
- Plants from lower elevation hybrid zone populations will have more *parviglumis*-like phenotypes for putatively adaptive traits (*e.g.*, fewer macrohairs, limited pigment, later flowering time), whereas plants from higher elevation hybrid zone populations will resemble *mexicana* (*e.g.*, highly pilose, pigmented, earlier flowering time)

Expected outcomes supporting tension zone dynamics

- Plants from hybrid zones will be less fit in all gardens (*i.e.*, all environments)
- Non-admixed *parviglumis* and *mexicana* will show higher fitness in low- and high-elevation gardens respectively
- Plants from hybrid populations will not show graded phenotypic variation (from *parviglumis*-like to *mexicana*-like) with increasing elevation. Rather, plants will possess random and maladaptive combinations of these traits

C) Are allele frequency clines at loci associated with fitness traits in *mexicana* and *parviglumis* steeper than those at non-associated loci?

We will combine our genome-wide marker data with data collected in our common garden experiments in order to perform genome-wide association analysis of fitness-related phenotypes in *parviglumis*, *mexicana*, and hybrids. We will evaluate both the shape and slope of allele frequency clines at fitness-related loci relative to neutral loci.

Objective II: Genus-wide introgression

Research Question IIA) Was the spread of maize facilitated by gene flow from locally-adapted wild *Zea*?

*background on spread of maize

*one new habitat was highlands

* talk about Hufford 2013: Our recently published study of *mexicana*/maize hybridization in sympatric populations (Hufford et al. 2013) suggests widespread adaptive introgression from *mexicana* into maize.

* but what about spread to (insert something about environment differences) in Guatemala?

* background on luxurians and huehue, information on maize arrival in guatemal from arch. record if available, or on maize in guat from van heerwaarden 2011.

* 4 pairs of huehue/maize and 4 pairs of lux/maize sympatric pops, then at both high and low elevation (or some other gradient?) we pick 1 allopatric teo and 1 allopatric maize, for a total of 6 pops of each taxa = 24 pops total.

* 12 individuals per pop, GBS to 48 plex.

*D stats, g-statistic, pi ¡- no need for phasing

* phase with fastphase, run hapmix or other methods

* look for evidence of selection (Nielsen sweepfinder, etc.)

* can look for overlap with QTL for water-logging etc. traits (Mano papers) or GWAS for related traits in maize

* include formal analysis using Berg's SQuaT approach?

* additional questions: how much diversity remains in lux populations? what is connectivity of these pops? are lux and/or huehue threatened by introgression from maize? important to Guatemala as it is unique diversity to that country

We will also expand the scope of our previous work and assess whether similar patterns of introgression can be detected in maize populations that are sympatric to the Guatemalan teosintes (*huehuetenangensis* and *luxurians*).

Research Question IIB) What is the geographic scale of adaptive introgression?

* in many plants local adaptation exists on a very fine geographical scale (examples, citations)

* previous work suggested introgression from mexicana allowed maize to adapt to highlands, and most of the introgression was ancient, suggesting adaptation to a broad set of challenges associated with higher elevations

* but we did see differences among populations (cite examples)

* now with increase resolution we can look at individual sympatric population pairs

* we also know parv hybridizes with maize, but unclear if any of the sympatric introgression is adaptive * 3 pairs of mex/maize sympatric pops * 3 pairs of parv/maize sympatric pops * introgression, search for evidence of selection w/in individual pops

Research Question IIC) Did maize serve as a bridge for gene flow between previously isolated Zea taxa?

Our initial survey of divergence and gene flow in *Zea*, based on a set of 26 Sanger sequence loci, found evidence for admixture between allopatric populations of *mexicana* and *luxurians* at multiple loci (Ross-Ibarra et al., 2009). As there is no evidence to suggest that these populations overlapped in their recent history, we took these results to suggest that maize, which is known to hybridize with both taxa, may have served as a bridge for gene flow between the two teosintes. Further support for this idea comes from genotyping of a SNP diagnostic for the two haplotypes at the inversion locus *Inv4m* (Fang et al., 2012; Pyhajarvi et al., 2013; Hufford et al., 2013). While the inverted haplotype at this locus appears to be derived in *mexicana* (Pyhajarvi et al., 2013), the SNP allele diagnostic of the inverted haplotype is nonetheless found in both maize and *luxurians* samples from Guatemala (Fig. **??**), suggesting the possibility that the inversion has moved from *mexicana* into *luxurians* via a maize intermediate. An alternative explanation, however, is that the haplotype was polymorphic in the common ancestor of both taxa, and remains segregating in each due to incomplete lineage sorting. Simple estimates of the length of shared haplotypes expected to be unbroken by recombination suggest that, over the $sim140,000$ divergence time between *mexicana* and *luxurians* (Ross-Ibarra et al., 2009), we might well expect to see shared haplotypes of even several kb in length in low recombination regions of the genome (Fig. **??**). The high-density, genome-wide data generated here will provide an opportunity to test whether observed patterns of haplotype sharing between previously allopatric *Zea* are due to recent introgression from maize. If shared haplotypes have come from introgression from maize over the last few thousand years, the genome-wide distribution of shared haplotype lengths should reveal longer shared segments (Fig. **??**) than if haplotype sharing is due to incomplete lineage sorting alone. Because maize has been found to hybridize with all species of *Zea* (Wilkes, 1977), we will extend this objective to the perennial taxa *Zea diploperennis* (hereafter, *diploperennis*) and *Zea perennis* (herafter *perennis*) as well.

We will sample 12 individuals from of 2 populations of *diploperennis* and *perennis* and genotype these using GBS. These populations, combined with samples from other teosinte populations in research questions **??** and , will provide us a representative sample of wild teosinte populations from across the Americas. We will use a number of methods, including HapMix (**?**)XXX), relative nucleotide differences (**?**)XXX), and STRUCTURE (**?**) to compare across these sets of populations to identify shared haplotype segments. *add FINE structure? something else? more detail here?* Although we do not have a genetic map for most teosinte taxa, evidence from comparison among maize populations finds remarkable stability of the genetic map at a relatively coarse scale (**?**)XXX). We will thus use the NAM genetic map to estimate recombination rates genome-wide, and using estimates of divergence time between taxa, generate an expected distribution of shared haplotype lengths. This expected distribution will be compared to that observed in pairwise comparisons between taxa to look for evidence of recent introgression. We will also scan a large sample of $> 16,000$ maize samples that have been genotyped with GBS (www.panzea.org) to look for the presence of such haplotypes in domesticated maize.

Broader Impacts

Our efforts to broaden the impact of the research proposed here will begin within our groups through our commitment to effectively mentor volunteer undergraduate interns as well as graduate students and/or postdoctoral scholars funded by the project. Students and postdocs will receive one-on-one training from the investigators and senior personnel on laboratory, computational, and field research methods. Mentees will also be encouraged and funded to present their work at scientific conferences. Our groups have an excellent mentoring track record with four undergraduate students in the last five years publishing their work in scholarly journals and multiple underrepresented minorities participating in our research.

ISU GK12 Fellowship Program

In addition to the student and postdoc mentoring that will occur within our groups, as part of our broader impact activities each year one of our graduate students will participate in Iowa State University's GK12 Fellowship program: *Symbi*, `http://www.gk12.iastate.edu/default.asp`. The selected graduate student will spend one full day each week in a middle or high school science classroom for the entire academic year of the Des Moines Public School District. This is the largest and most diverse school district in Iowa with over 50% underrepresented minority student enrollment and over 70% of students receiving free or reduced-cost lunch. The graduate student will introduce the K12 students to the scientific process through inquiry-based activities, relate the students science curriculum to real world examples, work with students on their science fair projects, and serve as a role model in a STEM profession. Furthermore, the graduate student will introduce students to his/her research project on hybridization and introgression in *Zea*, a topic that is particularly well suited for teaching evolution in Iowa given the important role that maize plays in the Iowan economy. In introducing his/her dissertation research, the graduate student will engage Des Moines students in how research is conducted and provide STEM content professional training to his/her partner teacher. The GK12 Fellow will work with approximately 150 students on a regular weekly basis. Student assessments from this program have shown that a significant number of students like science more after having a GK Fellow in the classroom. Teachers report that having a GK12 Fellow in their classroom is excellent professional development. The PI will also visit the classroom and will support the selected graduate student in their development of appropriate material for the K12 audience.

US-Mexico Exchange Program

Finally, we will establish a student exchange program between the Eguiarte Laboratory at UNAM in Mexico and the Hufford and Ross-Ibarra Laboratories in the United States. The Ross-Ibarra Laboratory has run an NSF-supported, US-Mexico exchange program for the last three years. All of the exchange students involved in the program have continued on to additional graduate work, and two have earned authorship on forthcoming papers from their internship. We will build upon the success of this program. A student from the Eguiarte group will spend 2-3 months in either the Hufford or Ross-Ibarra Laboratory learning the GBS methodology and/or honing his/her skills in population genomic analysis, whereas a student from the Hufford and/or Ross-Ibarra Laboratories will travel to Mexico to participate in sample collection trips and to obtain expertise in common garden field experiments. This exchange will build capacity in all groups involved and will provide a valuable international research experience for a graduate student supported by the grant.

Senior Personnel Claudia Calderon has previously led international student research trips and will assist in preparing students from both the United States and Mexico for the exchange program. A survey will be given to both exchange students and faculty in order to gauge expectations prior to the trip and facilitate collaborations amongst the labs. The survey will also assess students' knowledge and preconceived ideas regarding their travel destinations. A meeting (online or face-to-face) with the cohort of students traveling will help address these pre-conceptions and reduce cultural misunderstandings. Suggestions will be given to students of how to prepare before the trip (visa, immigration requirements) and how to communicate with their peers and others during their exchange. Students will be given information regarding the facilities where they will be staying, transportation to be used, food and water safety, the availability of telecommunications and general safety guidelines.

Results From Prior NSF Support

Ross-Ibarra: #1238014: Biology of Rare Alleles in Maize and Its Wild Relatives

$13,311,185 ($2,368,767 to Ross-Ibarra and $1,206,211 to Flint-Garcia), 05/15/13-04/30/18. PI Edward Buckler, co-PIs J. Doebley, J. Holland, S. Flint-Garcia, Q. Sun, P. Bradbury, S. Mitchell, J. Ross-Ibarra

Intellectual merit In the first year we have developed accurate imputation approaches, found evidence for the importance of deleterious variants and non-genic polymorphisms in heterosis and GWAS, documented differences in recombination among the parents of the NAM population, and found population genetic evidence suggesting the importance of demography and purifying selection across the genome. The grant has produced 18 total publications in its first year (only publications involving PIs Flint-Garcia and Ross-Ibarra are shown below).

Broader impacts In the first year this project has included 10 postdoctoral and 12 graduate trainees. The GBS workshop and traveling maize exhibit continue to be popular and successful. A new version of the teacher-friendly guide to maize evolution has been revised and published online.

Publications Peiffer et al. (2013); Romay et al. (2013); Wills et al. (2013); Mezmouk and Ross-Ibarra (2014); Peiffer et al. (2014); Sood et al. (2014)

Ross-Ibarra: #0922703: Functional Genomics of Maize Centromeres

$5,008,031 ($754,409 to Ross-Ibarra). 09/01/09-08/31/14. PI Kelly Dawe, co-PIs J. Birchler, J. Jiang, G. Presting, J. Birchler, J. Ross-Ibarra

Intellectual merit Centromeres are regions of the genome that organize and regulate chromosome movement, yet the biology of centromeres remains poorly understood. Co-PI Ross-Ibarra's group has focused in particular on the evolutionary genetics of centromeres. This work has demonstrated the remarkable evolutionary lability of centromere tandem repeats, but has shown that there is little evidence in maize for coevolution between centromere sequence and kinetochore proteins. Ongoing work from the Ross-Ibarra lab seeks to characterize kinetochore proteins, assess the phylogenetic evidence for longer-term coevolution, and understand patterns of centromere and genome size variation in natural populations.

Broader impacts Co-PI Ross-Ibarra has established an international student exchange program as part of this grant. Data and result of this project have been disseminated via publications and presentations as well as deposited in the maize genetics community database `www.maizegdb.org`. Former trainees on the grant include Dr. Matthew Hufford (Co-PI on the current grant).

Publications Shi et al. (2010); Chia et al. (2012); Fang et al. (2012); Hufford et al. (2012b,c, 2013); Melters et al. (2013); Kanizay et al. (2013); Pyhajarvi et al. (2013)

References Cited

E. Anderson and Jr. Stebbins, G. L. Hybridization as an evolutionary stimulus. *Evolution*, 8(4): 378–388, 1954.

Edgar Anderson. Hybridization of the habitat. *Evolution*, 2(1):1–9, 1948.

Jessica G. Barb, John E. Bowers, Sebastien Renaut, Juan I. Rey, Steven J. Knapp, Loren H. Rieseberg, and John M. Burke. Chromosomal evolution and patterns of introgression in helianthus. *Genetics*, 2014. doi: 10.1534/genetics.114.165548. URL `http://www.genetics.org/content/early/2014/04/24/genetics.114.165548.abstract`.

JM Chia, C Song, PJ Bradbury, D Costich, N de Leon, J Doebley, RJ Elshire, B Gaut, L Geller, JC Glaubitz, M Gore, KE Guill, J Holland, MB Hufford, J Lai, M Li, X Liu, Y Lu, R McCombie, R Nelson, J Poland, BM Prasanna, T Pyhajarvi, T Rong, RS Sekhon, Q Sun, MI Tenaillon, F Tian, J Wang, X Xu, Z Zhang, SM Kaeppler, J Ross-Ibarra, MD McMullen, ES Buckler, G Zhang, Y Xu, and D Ware. Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet*, 44(7):803–807, 2012.

The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405):94–98, 2012. doi: 10.1038/nature11041. URL `http://dx.doi.org/10.1038/nature11041`.

K. E. Delmore, R. A. Brenneman, R. Lei, C. A. Bailey, A. Brelsford, E. E. Louis, and S. E. Johnson. Clinal variation in a brown lemur (eulemur spp.) hybrid zone: Combining morphological, genetic and climatic data to examine stability. *Journal of Evolutionary Biology*, 26(8):1677–1690, 2013. doi: 10.1111/jeb.12178.

Concepcin M. Díez, Brandon S. Gaut, Esteban Meca, Enrique Scheinvar, Salvador Montes-Hernandez, Luis E. Eguiarte, and Maud I. Tenaillon. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*, 199(1):264–276, 2013. ISSN 1469-8137. doi: 10.1111/nph.12247. URL `http://dx.doi.org/10.1111/nph.12247`.

RJ Elshire, JC Glaubitz, Q Sun, JA Poland, K Kawamoto, ES Buckler, and SE Mitchell. A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PLoS One*, 6(5): e19379, 2011.

Z Fang, T Pyhajarvi, AL Weber, RK Dawe, JC Glaubitz, J Gonzalez Jde, C Ross-Ibarra, J Doebley, PL Morrell, and J Ross-Ibarra. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, 191(3):883–894, 2012.

K Fukunaga, J Hill, Y Vigouroux, Y Matsuoka, J Sanchez, KJ Liu, ES Buckler, and J Doebley. Genetic diversity and population structure of teosinte. *Genetics*, 169(4):2241–2254, 2005.

J. Galindo, M. Martnez-Fernndez, S. T. Rodrguez-Ramilo, and E. Roln-Alvarez. The role of local ecology during hybridization at the initial stages of ecological speciation in a marine snail. *Journal of Evolutionary Biology*, 26(7):1472–1487, 2013. doi: 10.1111/jeb.12152.

Jeffrey C. Glaubitz, Terry M. Casstevens, Fei Lu, James Harriman, Robert J. Elshire, Qi Sun, and Edward S. Buckler. Tassel-gbs: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, 9(2):e90346, 02 2014. doi: 10.1371/journal.pone.0090346. URL `http://dx.doi.org/10.1371%2Fjournal.pone.0090346`.

Alexander Grobman, Duccio Bonavia, Tom D. Dillehay, Dolores R. Piperno, Jos Iriarte, and Irene Holst. Preceramic maize from paredones and huaca prieta, peru. *Proceedings of the National Academy of Sciences*, 109(5):1755–1759, 2012. doi: 10.1073/pnas.1120270109. URL `http://www.pnas.org/content/109/5/1755.abstract`.

Rafael F Guerrero and Mark Kirkpatrick. Local adaptation and the evolution of chromosome fusions. *Evolution*, 2014.

R. G. Harrison. Hybrids and hybrid zones: Historical perspective. In R. G. Harrison, editor, *Hybrid Zones and the Evolutionary Process*, pages 3–12. Oxford University Press, New York, 1993.

Emilia Huerta-Sánchez, Xin Jin, Zhuoma Bianba, Benjamin M Peter, Nicolas Vinckenbosch, Yu Liang, Xin Yi, Mingze He, Mehmet Somel, Peixiang Ni, et al. Altitude adaptation in tibetans caused by introgression of denisovan-like dna. *Nature*, 2014.

Matthew B Hufford, Enrique Martínez-Meyer, Brandon S Gaut, Luis E Eguiarte, and Maud I Tenaillon. Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PloS one*, 7(11):e47659, 2012a.

MB Hufford, P Bilinski, T Pyhajarvi, and J Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends Genet*, 28(12):606–615, 2012b.

MB Hufford, X Xu, J van Heerwaarden, T Pyhajarvi, JM Chia, RA Cartwright, RJ Elshire, JC Glaubitz, KE Guill, SM Kaeppler, J Lai, PL Morrell, LM Shannon, C Song, NM Springer, RA Swanson-Wagner, P Tiffin, J Wang, G Zhang, J Doebley, MD McMullen, D Ware, ES Buckler, S Yang, and J Ross-Ibarra. Comparative population genomics of maize domestication and improvement. *Nat Genet*, 44(7):808–811, 2012c.

MB Hufford, P Lubinksy, T Pyhajarvi, MT Devengenzo, NC Ellstrand, and J Ross-Ibarra. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477, 2013.

LB Kanizay, T Pyhajarvi, EG Lowry, MB Hufford, DG Peterson, J Ross-Ibarra, and RK Dawe. Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in zea mays. *Heredity (Edinb)*, 110(6):570–577, 2013.

L. E. B. Kruuk, S. J. E. Baird, K. S. Gale, and N. H. Barton. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, 153(4):1959–1971, 1999.

Daniel John Lawson, Garrett Hellenthal, Simon Myers, and Daniel Falush. Inference of population structure using dense haplotype data. *PLoS Genet*, 8(1):e1002453, 01 2012. doi: 10.1371/journal.pgen.1002453. URL `http://dx.doi.org/10.1371%2Fjournal.pgen.1002453`.

Y Matsuoka, Y Vigouroux, MM Goodman, G J Sanchez, E Buckler, and J Doebley. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci U S A*, 99(9):6080–6084, 2002.

DP Melters, KR Bradnam, HA Young, N Telis, MR May, JG Ruby, R Sebra, P Peluso, J Eid, D Rank, JF Garcia, JL Derisi, T Smith, C Tobias, J Ross-Ibarra, I Korf, and SW Chan. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*, 14(1):R10, 2013.

S Mezmouk and J Ross-Ibarra. The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda)*, 4(1):163–171, 2014.

T. L. Parchman, Z. Gompert, M. J. Braun, R. T. Brumfield, D. B. McDonald, J. A. C. Uy, G. Zhang, E. D. Jarvis, B. A. Schlinger, and C. A. Buerkle. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, 22(12):3304–3317, 2013. doi: 10.1111/mec.12201.

JA Peiffer, SA Flint-Garcia, N De Leon, MD McMullen, SM Kaeppler, and ES Buckler. The genetic architecture of maize stalk strength. *PloS one*, 8(6):e67066, 2013.

JA Peiffer, MC Romay, MA Gore, SA Flint-Garcia, Z Zhang, MJ Millard, CA Gardner, MD McMullen, JB Holland, PJ Bradbury, and ES Buckler. The genetic architecture of maize height. *Genetics*, 2014.

D. R. Piperno and K. V. Flannery. The earliest archaeological maize (zea mays l.) from highland mexico: New accelerator mass spectrometry dates and their implications. *Proceedings of the National Academy of Sciences*, 98(4):2101–2103, 2001. doi: 10.1073/pnas.98.4.2101. URL `http://www.pnas.org/content/98/4/2101.abstract`.

J. W. Poelstra, N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Mller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344(6190):1410–1414, 2014. doi: 10.1126/science.1253226. URL `http://www.sciencemag.org/content/344/6190/1410.abstract`.

AL Price, A Tandon, N Patterson, KC Barnes, N Rafaels, I Ruczinski, TH Beaty, R Mathias, D Reich, and S Myers. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, 5(6):e1000519, 2009.

Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

T Pyhajarvi, MB Hufford, S Mezmouk, and J Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome Biol Evol*, 5(9):1594–1609, 2013.

Josepth B. Rasmussen, Michael D. Robinson, Alice Hontela, and Daniel D. Heath. Metabolic traits of westslope cutthroat trout, introduced rainbow trout and their hybrids in an ecotonal hybrid zone along an elevation gradient. *Biological Journal of the Linnean Society*, 105(1):56–72, 2012. doi: 10.1111/j.1095-8312.2011.01768.x.

David Reich, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh. Reconstructing indian population history. *Nature*, 461(7263):489–494, 2009. doi: 10.1038/nature08365. URL `http://dx.doi.org/10.1038/nature08365`.

S. Renaut, C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J. M. Burke, and L. H. Rieseberg. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4:1827–, 2013. doi: 10.1038/ncomms2833. URL `http://dx.doi.org/10.1038/ncomms2833`.

MC Romay, MJ Millard, JC Glaubitz, JA Peiffer, KL Swarts, TM Casstevens, RJ Elshire, CB Acharya, SE Mitchell, SA Flint-Garcia, MD McMullen, JB Holland, ES Buckler, and CA Gardner. Comprehensive genotyping of the usa national maize inbred seed bank. *Genome Biol*, 14(6):R55, 2013.

J Ross-Ibarra, M Tenaillon, and BS Gaut. Historical divergence and gene flow in the genus zea. *Genetics*, 181(4):1399–1413, 2009.

M A Saghai-Maroof, K M Soliman, R A Jorgensen, and R W Allard. Ribosomal dna spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences*, 81(24):8014–8018, 1984. URL `http://www.pnas.org/content/81/24/8014.abstract`.

J Shi, SE Wolf, JM Burke, GG Presting, J Ross-Ibarra, and RK Dawe. Widespread gene conversion in centromere cores. *PLoS Biol*, 8(3):e1000327, 2010.

Katie L. Smith, Joshua M. Hale, Laurne Gay, Michael Kearney, Jeremy J. Austin, Kirsten M. Parris, and Jane Melville. Spatio-temporal changes in the structure of an australian frog hybrid zone: A 40-year perspective. *Evolution*, 67(12):3442–3454, 2013a. doi: 10.1111/evo.12140.

Katie L. Smith, Joshua M. Hale, Michael R. Kearney, Jeremy J. Austin, and Jane Melville. Molecular patterns of introgression in a classic hybrid zone between the australian tree frogs, litoria ewingii and l.paraewingi: evidence of a tension zone. *Molecular Ecology*, 22(7):1869–1883, 2013b. doi: 10.1111/mec.12176.

S Sood, S Flint-Garcia, MC Willcox, and JB Holland. Mining natural variation for maize improvement: Selection on phenotypes and genes. In *Genomics of Plant Genetic Resources*, pages 615–649. Springer, 2014.

Fabian Staubach, Anna Lorenc, Philipp W. Messer, Kun Tang, Dmitri A. Petrov, and Diethard Tautz. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (¡italic¿mus musculus¡/italic¿). *PLoS Genet*, 8(8):e1002891, 08 2012. doi: 10.1371/journal.pgen.1002891. URL `http://dx.doi.org/10.1371%2Fjournal.pgen.1002891`.

G. Ledyard Stebbins. The role of hybridization in evolution. *Proceedings of the American Philosophical Society*, 103(2):231–251, 1959.

J van Heerwaarden, J Doebley, WH Briggs, JC Glaubitz, MM Goodman, J de Jesus Sanchez Gonzalez, and J Ross-Ibarra. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci U S A*, 108(3):1088–1092, 2011.

Y Vigouroux, JC Glaubitz, Y Matsuoka, MM Goodman, G J Sanchez, and J Doebley. Population structure and genetic diversity of new world maize races assessed by dna microsatellites. *Am J Bot*, 95(10):1240–1253, 2008.

Joseph J. Vitti, Sharon R. Grossman, and Pardis C. Sabeti. Detecting natural selection in genomic data. *Annual Review of Genetics*, 47(1):97–120, 2013. doi: 10.1146/annurev-genet-111212-133526. URL `http://dx.doi.org/10.1146/annurev-genet-111212-133526`. PMID: 24274750.

H. G. Wilkes. Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize. *Economic Botany*, 31(3):254–293, 1977.

HG Wilkes. Teosinte: the closest relative of maize. *Teosinte: the closest relative of maize.*, 1967.

DM Wills, CJ Whipple, S Takuno, LE Kursel, LM Shannon, J Ross-Ibarra, and JF Doebley. From many, one: Genetic control of prolificacy during maize domestication. *PLoS Genetics*, 9(6): e1003604, 2013.

Budget Justification

Personnel

No funding is requested for the PI, Co-PIs, or any Senior Personnel.

Other Personnel

Graduate students Funds are requested to support two graduate students each for 6 months during the academic year for each year of the project. At UC Davis, the current pay rate for doctoral students at 50% FTE is $27,319 during the academic year. Included is the estimated annual salary increase of 3%. The two students will be working on analysis of GBS data in the introgression and admix population genetic sections of **??**, and will likely help with QTL analysis and sequencing in **??**, and potentially RNA-seq analysis in **??**.

Technician Funds are requested for the first three years of the grant for a 50%-time technician (Laboratory Assistant III) to extract DNA and RNA, prepare genomic and transcriptomic sequencing libraries, and perform root chilling experiments. The salary for this positions is set at $36,000 ($18,000 for 50% time), with an annual increase of 5%.

Fringe Benefits

Fringe benefits are applied to personnel salaries using the university approved rates:

- Graduate students - 1.3% for all years.
- Technician - 50.4%(1/1/2015-6/31/2015), 53.4%(6/31/2015-6/31/2016), 55.7%(6/31/2016-6/31/2017), 57.3%(6/31/2017-12/31/2017)

Equipment

No equipment funds are requested.

Travel

Travel for the PI and Co-PI Coop and one student to 1 domestic conference each year is budgeted at $3,000. Travel for one of the Senior Personnel or Co-PIs to participate in the field workshop is budgeted at $1,000 each year.

Travel for Senior Personnel and members of their group to manage field experiments and phenotype is budgeted at $12,000 each of the first 3 years. Travel for both Senior Personnel to 1 international conference each year is budgeted at $3,000 per year.

Participant Support

Our exchange program proposes to exchange two students per year between the US and Mexico. We are requesting funds to pay for 2 exchange students per year of the grant. These funds will cover student subsistence ($1,800 a month to include housing and subsistence) for 3 months, visa costs ($500), and round-trip travel to Mexico ($2,000).

Other Direct Costs

Materials and Supplies In each of the first three years of the grant, $15,000 is requested in materials and supplies. $10,000 of this is for laboratory supplies for PI Ross-Ibarra for library prep for whole genome sequencing, RNA sequencing, and DNA extraction and preparation for GBS. This also includes funds for supplies for root chilling experiments to be done at UC Davis. In each of the five years, $2,500 is budgeted for standard office supplies, computer supplies (extra storage for our cluster, backup drives for lab members), and other miscellaneous expenses for Co-PI Coop and PI Ross-Ibarra.

Whole genome sequencing The genomes of each of the four parental lines of our QTL mapping populations will be resequenced to a depth of 20-30X using 2 lanes of paired end 150bp reads on an Illumina HiSeq 2500. Current lane costs are approximately $2,200 per lane, and library preparations costs are approximately $100, for a total cost of $18,000.

GBS Genotyping-by-sequencing will be performed for our introgression and admixture population genetic analyses. GBS will be performed at the Institute for Genomic Diversity at Cornell. Current prices are $60 per sample to run samples at 48-plex. We will genotype 360 individuals for our introgression analysis in year 1 for a cost of $21,600, and 144 individuals in year 2 for a cost of $8,640.

RNA sequencing In total, RNA sequencing will be performed on 384 individuals (8 inbreds x 2 stages x 2 tissues x 2 environments x 3 replicates + 8 NILs x 2 genotypes x 2 stages x 2 environs x 3 replicates). Cost to prepare RNA libraries in our lab are approximately $100 per library, and sequencing costs for single-end 50bp reads at the UCD Genome Center are approximately $1,000 per lane. Multiplexing 12 barcodes per lane, this comes out to 32 lanes of sequence and a total cost of $70,400.

Field fees Fees for the field experiments in our highland and lowland field sites (Table **??**) are approximately $60,000 the first three years of the experiment to allow development of the mapping populations and two replicates of the phenotyping. These fees include land rental and basic management (planting, watering, weeding, fertilizing), as well as station fees to hire manual labor for phenotyping. These fees decrease to $10,000 in the last two years of the proposal as subsequent field experiments including evaluation of NILs and RNA-seq lines, will be considerably smaller. Field fees total $200,000 across the five years of the grant.

Graduate Student Tuition Tuition for graduate students is charged to the project in proportion to the amount of effort the graduate student will work on the project. For a graduate student employed on the project for 9 academic months at 50% FTE, the tuition charge is $31,546 in FY 2015 to account for out-of-state tuition, $17,266 in FY 2016 and increasing 5% each subsequent year.

Publication Costs In year two $1,500 is requested for publication fees to an open access journal. In subsequent years $3,000 is requested annually.

Total Direct Costs

Total direct costs for UCD come to $874,643. Subawards to USDA-ARS and Iowa State total $1,218,560.

Indirect Costs

Indirect costs are calculated on Modified Total Direct Costs (Total Direct costs less graduate student fees and participant support and subaward funding beyond the first $25,000) using F&A rates approved by US Department of Health and Human Services. For this project, F&A rates of 55.5% were used from Jan. 1, 2015 through June 30, 2015, 56.5% from July 1, 2015 through June 30, 2016, and 57% from July 1, 2016 until the end of the project.

Facilities, Equipment, and Other Resources

Facilities, Equipment & Other Resources

UC Davis

Dr. Ross-Ibarra has four standard laboratory benches as part of a shared lab space at UCD. The shared space is the single largest lab space on campus, and provides for seamless interaction between the labs housed there. The space currently houses three other PIs, all working on the genetics and genomics of economically important plant taxa (Dubcovsky, Neale, Dandekar). The lab is equipped with standard equipment and tools for molecular biology, including freezers and refrigeration, a shared liquid handling robot, thermal cyclers, centrifuges, gel rigs, balances, and standard molecular biology supplies. A dedicated low-humidity refrigerator for seed storage is available through the university, and low-humidity storage cabinets for tissues and temporary seed storage are in the laboratory. Dr. Ross-Ibarra occupies half of a large office suite that includes a conference room and cubicle space for 25 people. Both Macintosh and PC workstations are available for student and postdoc employees. The PI is a contributing partner in a large computer cluster, giving the lab dedicated access to 192 processors, with the opportunity for use of nearly 800 additional CPU as resources allow. Recent (2013) additions to the cluster have provided it with additional CPU as well as six new shared high-memory (512Gb RAM) nodes, one of which is dedicated to the Ross-Ibarra lab. Dr. Ross-Ibarra is a faculty member of the UC Davis Genome Center, a large facility that includes bioinformatics, genotyping, metabolomics, proteomics, and expression analysis cores able to perform a variety of genomics analyses at cost for UC Davis faculty. The Genome Center also rents time on its equipment, including a bioanlyzer and library preparation robots. As a member of the Genome Center, Dr. Ross-Ibarra also has access to their additional computational facilities. UC Davis has also entered into a recent partnership with BGI (formerly the Beijing Genomics Institute) to provide additional high-throughput sequencing services via a new Sacramento-based sequencing facility.

Dr. Coop's dry space is located on the 3rd floor of the Storer building, which houses the Department of Evolution and Ecology. The space is newly renovated space and consists of 3 offices that can seat a 8 total of people, and a conference room. In addition members of the lab have access to an additional conference room and other offices shared with the Begun, Langley, Lott, Kopp and Turelli groups. This group is part of the larger Center and Graduate Group for Population Biology, one of the leading graduate training programs in ecology and evolution in the world. Each current member of Dr. Coops group has a quad-core Mac pro. The computers are loaded with all the necessary software (Word, R, Mathematica etc.) and are connected to the university network as well as to color and black and white printers. The Coop lab has access to the genome center computational facilities: http://www.genomecenter.ucdavis.edu/core-facilities/.

Iowa State

Project components completed in the Hufford Laboratory will include mapping population development, DNA isolation and PCR, and population genetic analysis of genotyping data. Population development will be carried out in field space available at the Curtiss Farm of Iowa State University (ISU). This facility is equipped with irrigation, tractors, tillage equipment, planters, and combines. Seed processing and cold storage facilities are also available on the ISU campus. The Hufford Laboratory has all equipment necessary for DNA isolation and PCR including centrifuges, thermal cyclers, an ultra-low freezer, water baths, a pH meter, balances, and an electrophoresis system. A gel imaging system and a NanoDrop spectrophotometer for DNA quantification are accessible through the Center for Plant Responses to Environmental Stresses at ISU. The DNA Facility at ISU provides access to cutting-edge genomic technology including HiSeq and MiSeq Illumina sequencing and library preparation for both paired-end and mate-pair approaches. Data analyses will be carried out using the High Performance Computing clusters available at ISU. Dr. Hufford currently has access to the Lightning3 cluster which has a mix of Opteron based servers, consisting of 18 SuperMicro servers with core counts ranging from 32 to 64 and 256 to 512 GB of memory.

USDA-ARS, Missouri

Dr. Flint-Garcia has 600 sq. ft of laboratory space in Curtis Hall, on the University of Missouri campus. The laboratory is fully equipped for molecular genetics, including a chemical hood, a Beckman table top centrifuge with multiple tube buckets, a Tetrad four plate thermalcycler, several freezers, ultra-low freezers and refrigerators, water baths, a pH meter, and balances. In the building, laboratory personnel have ready access to ultracentrifuges and rotors, growth chambers, an autoclave, lyophilizers, a Sorvall high speed preparative centrifuge with four rotors, a shaker-incubator for bacterial cultures, a chromatography cabinet, electrophoresis equipment for DNA, RNA protein and DNA sequence analysis, a plate reading spectrophotometer/flourometer, a pulse-field electrophoresis system, six Thermolyne thermalcyclers, and four Tetrad four plate thermalcyclers. Dr. Flint-Garcia has multiple personal computers, and computing resources including weekly data backups, direct access to a Sun Ultra10 Unix Workstation and NT server for data sharing, and IT support from USDA-ARS. In addition, the co-PI has access to the Lewis bioinformatics cluster (over 180 compute nodes with more than 1200 processor cores and 5400 GB of memory) via the University of Missouri Bioinformatics Core Facility. Dr. Flint-Garcia has 120 sq. ft of office space and ample office and desk space for postdocs, technicians and graduate students. Dr. Flint-Garcia shares two ABI 3100 DNA sequencers, an ABI 7900HT RTPCR machine, and a Beckman NxP robot used primarily for DNA extractions with Mel Oliver and Mike McMullen, and other USDA scientists in the unit. Dr. Flint-Garcia has access to greenhouse and field space (with irrigation capability; University of Missouri South Farm and Bradford Research Center), seed processing and cold storage space, and use of winter nursery facilities in Puerto Rico. The co-PI has access to a complete set of field equipment including multiple tractors, tillage equipment, a 4-row plot planter, and a 2-row plot combine.

LANGEBIO

Langebio's mandate is to conduct top-ranked research while promoting genomic knowledge for the protection and sustainable use of Mexican biodiversity. Its unique location in the agricultural center of Mexico facilitates field sampling and field experimentation. We have ample experience growing maize in nurseries located on the West Coast (Valle de Banderas, Nayarit), in Central Mexico (Irapuato; Celaya, Guanajuato), and have begun to establish additional sites in the high valleys of Central Mexico (Queretero; Estado de Mexico). We regularly conduct field expeditions to collect plants in both the dry regions of Northern Mexico (maize collections in Chihuahua, Lamiaceae throughout the Northeast) and the lower valleys of the Eje Volcanico and Costa del Pacifico (Teocintle and maize, Solanaceae, and Cucurbitaceae). Research at Langebio is supported by greenhouse facilities and two service units: Genomics and Mass Spectrometry, both of them equipped with state-of-the-art instrumentation, including several next-generation sequencing machines and diverse mass spectrometry equipments. Other facilities include a computation cluster and a specialized clean room for ancient DNA analysis.

SEE APPENDIX A-1 UPLOADED AS A SUPPLEMENTARY DOCUMENT

Supplementary Documentation

Postdoctoral Researcher Mentoring Plan

The current proposal requests funding for two postdoctoral researchers, one each at Iowa State and USDA-ARS in Columbia. Nonetheless, we expect additional postdocs to join the group via alternative funding opportunities (fellowships, etc.) and anticipate that postdocs in the labs of all the PIs may collaborate to a greater or lesser degree on this project. Much of our thinking on postdoctoral mentoring comes directly from our own mentorship experience – PIs Flint-Garcia, Hufford, and Ross-Ibarra were all postdoctoral scholars on NSF-funded programs. For this project, the PI at each institution will act as mentor and supervisor for each postdoc, holding regular weekly meetings to assess progress and set goals. One clear goal will be first authorship on submitted papers, with the expectation of approximately one first author paper per year of duration of the postdoc.

Interaction and experience presenting and discussing science will be highly encouraged. All groups will have internal lab meetings (the Coop and Ross-Ibarra labs at UC Davis hold joint lab meetings) at which postdocs and graduate students will be given numerous opportunities to hone their presentation skills. The Coop, Ross-Ibarra and Hufford labs currently host weekly journal clubs in which postdocs gain additional training in reading, presenting, and dissecting scientific literature. Members of the Ross-Ibarra and Flint-Garcia labs also attend a weekly journal club as part of another collaborative project (NSF #1238014). In addition, we will organize a monthly group meeting via web-conference in which one lab member presents on their research progress. UC Davis has a ReadyTalk license allowing inexpensive web-conference hosting. All of our institutions have seminar series specifically for postdoctoral and graduate students to practice presentation skills; members of our labs will be encouraged to attend these.

Another important aspect of training will be experience mentoring graduate students and undergraduates. Postdocs will be given the opportunity to supervise undergraduate and/or graduate students on projects related to the grant. Previous efforts to encourage such supervision in our labs have been very successful, with postdoc-mentored students presenting conference posters on their research or earning authorship on papers. Supervisory experience has proven helpful for postdocs applying for jobs, especially in industry.

Postdocs will be encouraged to write and apply for external funding, including fellowships and grant proposals. Both the Ross-Ibarra and Coop labs have a documented history of successful funding with postdoctoral scholars as Co-PIs, providing valuable training (and even initial funding) for the scholars' future academic careers.

Postdocs in the Hufford and Flint-Garcia labs will take part as trainers in the annual phenotyping workshop under supervision of Co-PI Flint-Garcia. This will provide additional training in high-throughput phenotyping as well as valuable teaching experience.

Finally, postdocs will be encouraged to take advantage of professional development programs offered by their local institutions. All of our institutions have infrastructure in place for professional development of postdocs and offer training in responsible conduct of research, grantsmanship, mentoring, career development, authorship of journal papers, and teaching.

Supplementary Documentation

Sharing of Results and Management of Intellectual Property

Data Types

This proposal will generate sequence data, genotype, phenotype data, analytical software, teaching resources, germplasm, and publications.

Data Access, Sharing

All sequence data (RNA-seq, whole genome sequencing, and fastq files from genotyping by sequencing) will be submitted immediately upon completion of data quality control to the NCBI sequence read archive (SRA), along with passport information on each parent. A "hold until publication" embargo will be requested at the SRA. Before publication, data will also be made publicly available via the Figshare website (`www.figshare.com`), a free public website allowing dissemination and archiving of large datasets. Data will be released in accordance with the Toronto agreement (2009. Nature 461:168-170. `www.nature.com/nature/journal/v461/n7261/full/461168a.html`) under the stipulation that no whole-genome analyses be performed until we have published our initial analyses. RNA-seq data will include metadata as stipulated by MIAME (`http://www.ncbi.nlm.nih.gov/geo/info/MIAME.html`) and will also be deposited in the NCBI GEO database.

Phenotypic data and genotypes from sequencing and GBS will be uploaded to Figshare, along with appropriate metadata associated with other publications, links to germplasm, SRA experiments, Github code, etc. Phenotypic data will be recorded digitally in the field using the high-throughput techniques developed by Dr. Flint-Garcia. Data will be uploaded at the end of each day into the FieldBook database developed by Dr. Flint-Garcias USDA-ARS group and immediately backed up at a remote location. Data will be grouped into projects, and each project is associated with a unique digital object identifier (DOI). Drs. Ross-Ibarra and Coop have already used Figshare extensively to share and archive data, preprints, and code (see `http://figshare.com/authors/Jeffrey_Ross-Ibarra/98899` and `http://figshare.com/authors/Graham_Coop/101524`). Data on Figshare is publicly available and searchable. We will submit data as soon as we complete quality control, but again with explicit stipulations as to the analyses that the data can be used for prior to our initial publication. All appropriate metadata including plant ID, data collector, sequence run, field location, etc. will be associated with genotype and phenotype data deposited to Figshare.

Analytical software and code from this project will be hosted on Github, a version-controlled public git repository. Upon submission of papers all code will be made publicly available. Drs. Ross-Ibarra and Coop have already done this extensively (see `https://github.com/rossibarra`, `https://github.com/rilab`, and `https://github.com/cooplab`). Publication of all code will ensure reproducibility of all analyses conducted.

Presentations and teaching resources from our field workshop will be made publicly available via Figshare as well.

All data, code, and presentations will be made publicly available via a creative commons CC by 2.0 license (http://creativecommons.org/licenses/by/2.0/) allowing free access to reuse, redistribute, and modify, requiring only citation of the license and the original source.

All publications resulting from this project will be submitted to one or more preprint servers (e.g. arXiv, bioRxiv, PeerJ) such that they will be publicly available immediately upon submission of the paper for publication.

Data Archiving

All data, code, presentations, and publications will be made publicly available online (see above). Prior to public release, all data will be hosted locally. Dr. Ross-Ibarra will maintain a backup of all raw genotyping, sequence, and phenotyping data. His lab maintains a DROBO distributed backup server (currently $>$ 8Tb of free space) which is robust to single disk failure. All analytical code will be hosted on Github, which maintains version-controlled backups, as private repositories until release.

Both our F2:3 families and our near isogenic lines will require multiple generations of development until they are mature resources for mapping traits related to highland adaptation. We will archive a sample from each generation of population development in temperature- and humidity-controlled facilities at Iowa State University and Langebio. Sample accession data will be securely stored in a MySQL server hosted at the University of California, Davis and backed up on a weekly basis offsite. International agreements prohibit some of the maize and teosinte germplasm collected in Mexico from being stored and distributed by USDA. We will, however, deposit small quantities of seed from all our collections with the CIMMYT germplasm bank in Mexico, and deposit samples of our mapping populations (F2:3 seed) in the USDA-ARS Maize Stock Center at the University of Illinois. Both centers provide public access to seed.

Supplementary Documentation

Management Plan

Communication

All team members will communicate on a monthly basis via a scheduled conference call. UC Davis has a ReadyTalk license allowing inexpensive web-conference hosting; other institutions will call in and can share slides, video, their desktop, and audio. During these calls we will discuss progress, problems and solutions, as well as ways to more efficiently collaborate and coordinate among laboratories. One member from each of two labs will present an update of thier work. Postdocs and all students will be expected to participate.

Team members will hold an annual meeting in person each year as a satellite meeting to a conference (either Plant and Animal Genome or the annual Maize Genetics Conference). PIs not able to make the meeting will join via teleconference. Annual meetings will consist of PIs reporting progress during the past year and goals for the upcoming year.

Outreach

The exchange program will be coordinated between team members. Management of visa and travel costs will be done through UC Davis, as DR. Ross-Ibarra's program has experience with international exchange with Mexico.

Dr. Flint-Garcia will coordinate the annual phenotyping workshop, held each year in Columbia. The workshop will be timed to coincide with data collection at the end of the field season each year. The workshop will be advertised broadly (evoldir list-serv, maizegdb, etc.). Attendees will be expected to pay their own travel and purchase a handheld device

Research

Total research commitment to this grant for each PI will be:

- Angelica Cibrian Jaramillo: 5%
- Graham Coop: 5%
- Sherry Flint-Garcia: 10%
- Matthew Hufford: 15%
- Jeffrey Ross-Ibarra: 10%
- Ruairidh Sawers: 15%

Below are details of the responsibilities of each team member during each year of the grant, with initials as shown in summary Table 2. Although one group will take lead for writing publications, it is anticipated that several team members and members of their groups will be coauthors on many of these publications.

Year	1	2	3	4	5
?? QTL mapping	SFG, RS, JRI	SFG, MBH, RS, JRI	SFG, MBH, JRI	SFG, JRI	SFG
?? Admix mapping	MBH	MBH, GC	MBH, RS, GC	MBH, GC	–
?? Population genetics	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC
?? Functional analyses	RS	RS, SFG	RS, SFG, MBH	RS, SFG	RS, SFG
?? RNA-seq	–	ACJ	JRI, ACJ, RS	JRI, ACJ	JRI, ACJ

Table 2: Proposed timeline of activities showing which team members will be responsible for each objective. Team member names are abbreviated: MBH, Matthew Hufford; JRI, Jeffrey Ross-Ibarra; SFG, Sherry Flint-Garcia; GC, Graham Coop; RS, Ruairidh Sawers; ACJ, Angelica Cibrian Jaramillo

Year 1

?? SFG will generate seed of F2:3 for Mexico and F2 for S. American cross. JRI will sequence parents of both crosses. RS will choose highland site.

?? MBH will collect seed from Ahuacatitlan. GC will focus on developing methods for admix mapping.

?? MBH will collect seed from additional admixed populations. JRI will genotype samples from highland Mexico maize. AJC and MBH will put together dataset of global highland maize. GC will work on methods for selection in admix populations.

?? RS will screen HIFs and advance the population. Initial crosses for allelic series will be performed.

Year 2

?? SFG and RS will grow mapping population at each of 3 locales. SFG, RS, and MBH will phenotype populations in field. SFG will genotype F2 plants. JRI will phenotype root chilling.

?? MBH will genotype samples. RS and MBH will grow samples at two locations. GC will begin data analysis.

?? JRI will genotype seed from additional admix populations. MBH will genotype global highland maize collection. JRI and GC will begin data analysis of introgressed highland maize.

?? RS will advance the HIF and NIL populations.

?? ACJ will choose NILs for RNAseq analysis

Year 3

?? SFG and RS will grow second replicate of mapping population at each of 3 locales. SFG, RS, and MBH will phenotype populations in field. SFG and JRI will build map and begin QTL analysis.

?? GC and MBH will complete data analysis and begin writing.

?? JRI and GC will work on data analysis of admixed teosinte and highland Mexico maize. MBH will begin data analysis of global highland maize.

?? RS, SFG, and MBH will grow and phenotype HIFs and NILs. RS will advance both populations.

?? ACJ and RS will grow NILs and donors in highland and lowland environment. ACJ will extract RNA. JRI will perform RNAseq library prep and sequencing.

Year 4

?? SFG and JRI will perform QTL analysis

?? GC and MBH will write paper.

?? JRI and GC will finish data analysis and begin papers for admixed teosinte and highland Mexico maize. MBH will finish analysis of global highland maize.

?? RS will advance both populations. SFG and RS will analyze data.

?? JRI and ACJ will analyze RNAseq data.

Year 5

?? SFG will write paper.

?? MBH, JRI, and GC will write papers.

?? RS will advance both populations. SFG and RS will write paper.

?? JRI and ACJ will write paper.

Supplementary Documentation

Plans for Undergraduate and Graduate Student Mentoring

Undergraduate Students

Only Iowa State has requested funding for undergraduate students, but it is anticipated that undergraduate students will participate in unfunded internship roles at UC Davis and possibly USDA-ARS through the University of Missouri. Undergraduates will be partnered directly with a graduate student or postdoc. Unpaid undergraduate interns will be expected to develop specific research projects, and are expected to present on the progress of their work during regular group meetings. In addition to research experience in the lab or in the field, undergraduates will be encourage to attend regular lab meetings, and lab journal clubs; this is already regularly the case for students working with Drs. Ross-Ibarra and Coop. UC Davis undergraduates have also presented their work at university-sponsored research conferences and numerous students have earned authorship on peer-reviewed publications. Students will be given opportunities to develop data analysis and management skills, both through the field management system of Dr. Flint-Garcia, and through learning basic statistical and bioinformatics tools such as R and Unix at UC Davis or Iowa State. Undergraduate students will be provided guidance about potential careers in biology and plant science (see, for example, `http://www.slideshare.net/jrossibarra/forgradschool`).

Graduate Students

The current proposal requests funding for graduate students only at UC Davis, although it is hoped that additional students will participate in this grant through other funding mechanisms (institutional support, competitive fellowships, etc.). Students will be trained in order to prepare them for research careers (academic or otherwise). All students will be expected to take part in internal lab meetings (the Coop and Ross-Ibarra labs at UC Davis hold joint lab meetings) at which pthey will be given numerous opportunities to hone their presentation skills. The Coop, Ross-Ibarra and Hufford labs currently host weekly journal clubs in which students gain additional training in reading, presenting, and dissecting scientific literature. All members of the Ross-Ibarra and Flint-Garcia labs also attend a weekly journal club as part of another collaborative project (NSF #1238014). In addition, we will organize a monthly group meeting via web-conference in which one lab member presents on their research progress. UC Davis has a ReadyTalk license allowing inexpensive web-conference hosting. All of our institutions have seminar series specifically for postdoctoral and graduate students to practice presentation skills; members of our labs will be encouraged to attend these. Graduate students on the grant will be expected to produce first-author papers for peer-review as part of their project, and encouraged to contribute to additional papers as middle author. Students will be expected to attend and present a poster or talk at a scientific conference each year; UC Davis provides several opportunities for travel funds to support students in this manner. Finally, issues of ethics and organization will be included in training. These will include authorship, reproducibility, and basic scientific ethics. For example students will be encourage to pursue open science, including the submission of preprints and pre-publication data release and students will be required to maintain Github repositories of their computational work to ensure reproducibility and transparency.