

## 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. *parviflora* (hereafter, *parviflora*) 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 evidence for selection on putatively adaptive phenotypic traits in hybrid zones. 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 test hypotheses about the role of maize 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 *parviflora* 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 taxa, 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 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. 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 role of hybridization and introgression as agents of evolution has long been postulated (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 evidence of 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 based on full-genome data. Genomic tools provide the capacity to evaluate the genomic extent of introgression, differentiate neutral from adaptive introgression, and evaluate whether distinct hybrid populations show convergent patterns of introgression. With this expanded perspective, we can fully evaluate hypotheses first articulated by Anderson and Stebbins (1954): that 1) Evolution can occur rapidly through selection on novel variation generated by hybridization, and 2) The redistribution of previously isolated species through anthropogenic activity has provided ample opportunity for hybridization to occur (Anderson and Stebbins, 1954). Given its evolutionary history and genomic resources, the *Zea* study system presents an unparalleled opportunity to test these hypotheses and realize the promise of genomic data for hybridization studies.

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 in this system. First, analysis of hybridization in two parapatric and uniquely-adapted teosinte subspecies will allow us to answer questions about the extent to which selection determines patterns of gene flow and whether introgression is convergent across independent hybrid zones. Second, evaluation of introgression between maize and teosinte outside the domestication center will inform our understanding of the importance of adaptive introgression during colonization of novel environments and test the role of widespread species to act as bridges for gene flow among their congeners.

This proposal builds considerably upon a previous full submission to the 2014 NSF-DEB competition that was favorably reviewed but not funded due to concerns over certain aspects of experimental design that we have rectified here. In addition, over the last year we have conducted additional preliminary analyses, solidified collaborations with colleagues in Mexico and Guatemala, and ensured that this project is logically feasible.

## Objectives

### **Objective I: Assess the evolutionary role of hybridization in a naturally occurring teosinte hybrid zone**

*Zea mays* ssp. *parviflora* (the wild progenitor of maize; hereafter, *parviflora*) and *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) diverged approximately 60,000 BP (Ross-Ibarra et al., 2009) and have parapatric distributions: while *parviflora* 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, high-density genotyping, population genomic analyses, and common garden experiments, we will address the following research questions:

- A. *Do hybrid populations show higher fitness than parental subspecies in some environments?*
- B. *Do independent hybrid zones show convergent or unique patterns of introgression?*
- C. *Is there evidence of selection on putatively adaptive phenotypes across hybrid zones?*

### **Objective II: Determine the extent to which anthropogenic movement of maize has enabled hybridization among taxa in *Zea***

Maize was domesticated in southwest Mexico from *parviglumis* ~9,000 BP (Matsuoka et al., 2002) and quickly spread throughout the Americas, bringing it into contact with other teosinte taxa. Hybridization has been observed between maize and each of these teosintes, raising questions about the interaction between colonization and hybridization. Our population genetic analysis will assess the following questions:

- A. *Do the fitness consequences of gene flow from locally adapted populations depend on divergence?*
- B. *Is introgression adaptive across multiple spatial scales?*
- C. *Can a widespread species serve as a bridge for introgression among allopatric species?*

## **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, defined and stable regions of hybridization, referred to as hybrid zones, have been discovered in a number of taxa (reviewed in Harrison, 1993; Shurtliff, 2013; Abbott and Brennan, 2014). It is increasingly clear that the phenomena of hybridization and subsequent introgression shape genomes and influence the trajectory of species as they evolve (Ellstrand, 2014). Hybridization has long been believed to play a role in speciation, and introgression of even a small number of loci can enable a species to adapt and invade novel habitats (Currat et al., 2008; Abbott et al., 2013). For example, we now have strong molecular evidence that hybridization has led to speciation in both plants and animals (reviewed in Mallet, 2007) and that colonization events of non-native species (Lucek et al., 2010) and domesticated crops (He et al., 2011; Hufford et al., 2013) have been facilitated by introgression.

While theory regarding hybridization has progressed and many compelling empirical examples have been identified, several outstanding questions remain. For example, are hybrid zones largely maintained by adaptive or purifying selection or strictly neutral processes (Kruuk et al., 1999; Rasmussen et al., 2012; Smith et al., 2013)? If introgression is adaptive, what is the geographic scale

of this adaptation? Are the same adaptive alleles introgressed across independent hybrid populations or are patterns of introgression driven more by local environmental conditions? Can a widespread congener act as a bridge for gene flow between more narrowly distributed allopatric taxa? Genome-wide analysis of variation in the extent and genetic architecture of introgression — the number and size of introgressed loci — will offer considerable insight regarding each of these questions. For example, the few genomic studies of introgression completed thus far have already suggested that rates of gene flow vary substantially across loci, likely as a function of selection on introgressed alleles (Hufford et al., 2013; Poelstra et al., 2014) and that introgression has occurred at genomic locations underlying adaptive phenotypes that may have facilitated colonization of a non-native taxon (Hufford et al., 2013). As described below, the Zea study system is uniquely suited to address remaining, timely questions in the field of hybridization biology from a genome-wide perspective.

## Research Plan

### **Objective I Assess the evolutionary role of hybridization in naturally occurring teosinte hybrid zones**

The *Zea mays* subspecies *parviglumis* and *mexicana* are distributed across a steep elevational gradient and differ for phenotypes that are thought to be adaptive in the highlands such as the presence of macrohairs, stem pigmentation and shorter flowering time in *mexicana*. A recent ecological niche study has found that the potential distributions of these subspecies are largely allopatric and stable over many thousands of years (Hufford et al., 2012b). However, analysis of microsatellite markers genotyped in a large sample of the two subspecies suggested elevated admixture between them in mid-elevation regions of Mexico where their distributions overlap, potentially indicating the presence of a hybrid zone (Fukunaga et al., 2005). Our recent genome-wide analysis of twelve individuals from one such population revealed abundant small blocks of admixture in all individuals, suggesting long-term gene flow between subspecies (Pyhajarvi et al., 2013). Very little is known, however, about patterns of introgression across populations in this region.

To further refine our research questions and provide preliminary results for this proposal we have reanalyzed published data (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.*, traditional open-pollinated varieties). While the low density of markers in these data precludes genome-wide inferences and haplotype-based analyses, the comprehensive taxon sampling makes this an ideal resource for guiding future research. Using these data we calculated the probability of assignment of samples to *parviglumis* and *mexicana* groups with the program STRUCTURE (Pritchard et al., 2000). We find that individuals from several mid-elevation populations show appreciable assignment to both groups (Figure 1) and likely represent hybrid populations. 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 (mean 1632m) than those in the eastern Balsas (mean 1531m) and also show a higher proportion of membership in the highland teosinte *mexicana* group (Panels B and C, Figure 2). These findings suggest that hybrid populations from distinct environments may vary in the proportion of ancestry from each subspecies in a manner that

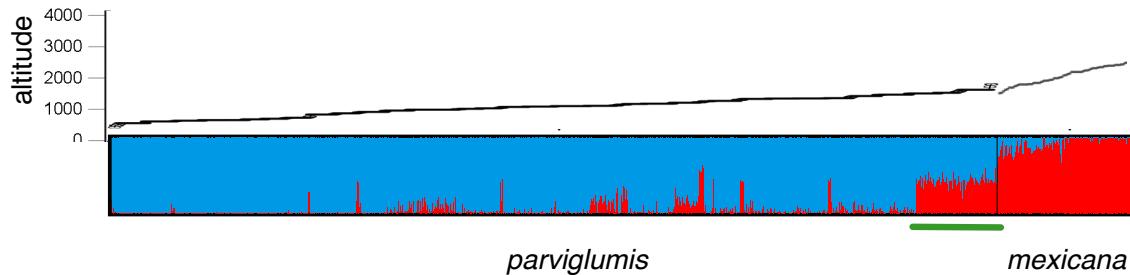


Figure 1: Assignment of *parviflumis* and *mexicana* individuals to K=2 groups using the Bayesian assignment algorithm of STRUCTURE (Pritchard et al., 2000). Individuals are sorted by increasing elevation as indicated by the plot above the bar chart. Individuals from mid-elevation, hybrid zone populations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

is adaptive. Estimates of pairwise population differentiation ( $F_{ST}$ ; data not shown) 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. Not surprisingly, populations in both hybrid zones are less differentiated from *mexicana* and *parviflumis* than these subspecies are from each other. Finally, we have grown plants from populations in both the Jalisco and Balsas hybrid zones in smallscale growth chamber experiments and have found that they have intermediate morphologies between *parviflumis* and *mexicana*.

Few studies have gauged evidence for selection and dissected the genome-wide architecture of hybridization in replicate hybrid zones. The Hufford Laboratory and Senior Personnel Luis Eguiarte will investigate these questions in two hybrid zones of *mexicana* and *parviflumis* through targeted population collections, generation of phenotype, genotype and full-genome sequence data, and application of quantitative and population genetic analyses as described in the following three sub-Objectives.

#### **Objective IA Do hybrid populations show higher fitness than parental subspecies in some environments?**

Currently, there are two dominant hypotheses amongst evolutionary biologists for the stability of hybrid zones over time: the tension zone and the bounded hybrid superiority zone (i.e., ecotone) hypotheses. Under tension zone dynamics hybrids have lower fitness relative to their parental taxa across environments due to, for example, incompatibility factors that arise when divergent taxa mate. Tension zones arise at the interface of parental taxa and are typically narrow clines due to the effects of purifying selection (Abbott and Brennan, 2014). In contrast, under ecotone dynamics hybrids have a selective advantage in intermediate environmental conditions relative to parental species and their distribution is typically characterized as a smooth transition between parental habitats (Abbott and Brennan, 2014). Hybrid populations of *parviflumis* and *mexicana* span several kilometers (Figure 2) under environmental conditions intermediate to their parental taxa suggesting ecotone dynamics prevail. We will test this hypothesis by assessing fitness and

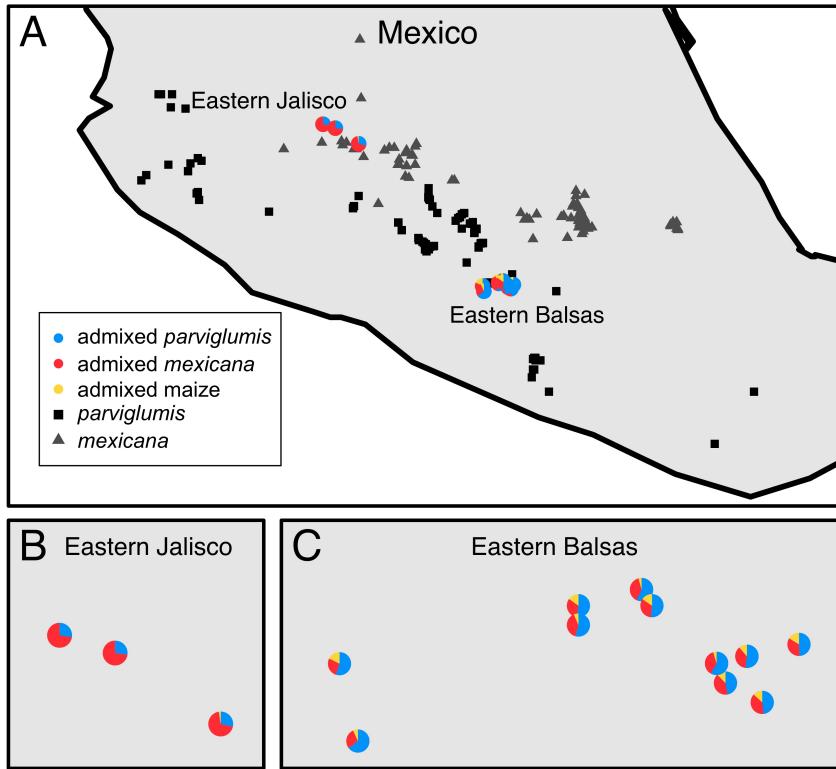


Figure 2: A) Location of two putative hybrid zones of *mexicana* and *parviflumis*. Hybrid populations are represented as pie charts with proportion assigned to *mexicana*, *parviflumis*, and maize groups. Zoomed-in views of the Eastern Jalisco (B) and Eastern Balsas (C) hybrid populations.

variation at putatively adaptive phenotypes across both non-admixed and hybrid populations in common garden experiments conducted in Mexico at three elevations: 1) Below a hybrid zone in habitat occupied by non-admixed *parviflumis*; 2) Within hybrid zone habitat; and 3) Above a hybrid zone in habitat occupied by non-admixed *mexicana*.

From previous collections, we have access to extensive sampling of *mexicana* and *parviflumis*. Moreover, Senior Personnel Luis Eguiarte has recently collected altitudinal transects of *parviflumis* and *mexicana* that extend through both hybrid zones we have identified (Díez et al., 2013) and is familiar with populations in this region. However, our current collections will need to be expanded in order to conduct the activities we propose 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 non-admixed populations each of *parviflumis* and *mexicana* and four populations from both hybrid zones). Sampling sites will be randomly stratified across the elevation gradient of each population. At each sampling site we will collect as much seed as possible from each of five teosinte plants (in the wild, teosinte plants produce an average of ~100 seeds per plant (Wilkes, 1967)). While elevation and correlated trends in temperature and precipitation appear most important in patterning the distribution of hybrids, we will also collect local environmental data at each sampling site including plant density (in 1 m<sup>2</sup> quadrats), slope of the terrain, and canopy cover. A single

soil sample will also be collected from each population. We have already obtained the requisite collection permits as well as permits for importing samples into the United States for the genetic analyses proposed in Objective IB.

Following collection, samples will be sent to Irapuato, Mexico for seed increase (see letter of support from our collaborator Ruairidh Sawers at Langebio in Irapuato). Irapuato is a mid-elevation site and will be best-suited for increasing seed from all populations in a single environment. Forty-five plants per population (three from each sampling site) will be grown in isolated plots surrounded by a detasseled maize buffer and will be allowed to open-pollinate. Seed will be harvested from each plant, keeping each plant's seed separate in order to preserve half-sibling relationships that are necessary for the  $F_{ST}$ - $Q_{ST}$  framework used in Objective IC. This single generation of seed increase in a common environment is necessary to remove maternal effects that may confound phenotypic comparisons of populations in our common garden experiments.

Common garden experiments will be replicated at three sites along an elevation gradient during years two and three of our proposed project. Our lowland site is outside of Bucerias, Nayarit state (40m), our mid-elevation site is in Irapuato, Guanajuato state (1700m), and our highland site is in Metepec, Mexico state (2600m). We have experience working in these sites and have confirmed that our common gardens will be logically feasible. Each garden will consist of three complete blocks including a randomization of a half-sibling family from each sampling site of our 16 sampled populations (3 blocks x 3 half siblings per sampling site x 15 sampling sites x 16 populations = 2,160 plants per garden). We will measure fitness-related phenotypes (percent germination, germination rate, plant height at 15-day intervals, total seed set, 50-seed weight, total-above-ground biomass, stomatal conductance and survival), putatively adaptive phenotypes across the altitudinal gradient (macrohair density, pigmentation extent, and flowering time) and a phenotype for which there is no *a priori* evidence of selection across an elevational gradient (the width of the leaf beneath the first lateral branch at the time of flowering).

Comparison of fitness of hybrid and parental plants across our garden sites will provide evidence regarding ecotone versus tension zone dynamics in teosinte hybrid zones. Ecotone dynamics would be supported by hybrids possessing the highest fitness of all plants in the mid-elevation garden, whereas tension zone dynamics would be supported by hybrids having lower fitness in all gardens. Phenotypic data for putatively adaptive traits and our trait with no evidence of selection will be analyzed in Objective IC.

### **Objective IB Do independent hybrid zones show convergent or unique patterns of introgression?**

While we cannot predict whether tension zone or ecotone dynamics define teosinte hybrid zones prior to completing our common garden experiments, assessment of genome-wide patterns of introgression will be an essential next step regardless of this outcome. If we find that hybrids are selected against and exist within two discrete tension zones, then our primary interest will be in investigating the level of convergence in hybrid inviability loci (*i.e.*, Dobzhansky-Muller incompatibilities) across populations and independent hybrid zones. Should hybrid populations show superior fitness in our intermediate garden alone and follow an ecotone model, we will instead investigate convergence in the genetic architecture of adaptive introgression.

During the initial seed increase at our Irapuato field site in year one, leaf tissue will be collected from a single plant per sampling site ( $n = 240$ ; see Objective IA) at the 5-7 leaf stage, stored in silica,

and shipped to Iowa State University for DNA isolations, subsequent genotyping and full-genome sequencing. DNA for genotyping will be isolated using the Qiagen DNeasy Plant Mini Kit. For sample genotyping ( $n = 240$ ) we will utilize the services of the Genomic Diversity Facility at Cornell University to implement 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 tens of thousands of 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 will multiplex only 48 individuals per lane (instead of the standard 384 used for inbred lines) to minimize missing data and errors in identifying heterozygous genotypes. Based on our experience with diverse maize and teosinte (e.g., Takuno et al., 2015; Mezmouk and Ross-Ibarra, 2014), 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 150bp, paired-end data per individual. We have previous experience dealing with whole genome shotgun data (Gore et al., 2009; Chia et al., 2012; Hufford et al., 2012a; da Fonseca et al., 2015), and have recently developed and implemented an open-source pipeline for read-mapping and SNP calling (<https://github.com/RILAB/paap>) using the existing maize B73 reference genome.

We will assess genome-wide patterns of ancestry (*parviflumis* versus *mexicana*) in hybrid individuals using several approaches. First, standard measures of differentiation including  $F_{ST}$ , the proportion of shared and fixed variants, and relative levels of nucleotide diversity (Geneva et al., 2014) will be calculated in sliding windows along the genome. These will be complemented by window-based methods using counts of derived alleles (Martin et al., 2015). Second, we will implement haplotype-based methods for detecting 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 our allopatric reference populations of *parviflumis* and *mexicana*.

We will evaluate *parviflumis* and *mexicana* ancestry on a site-by-site basis across each hybrid genome and determine level of convergence at three levels: individual, population, and independent hybrid zone. 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 inspected in our whole-genome resequencing data to refine haplotype boundaries. Whole genome sequence data will also potentially allow the identification of candidate causal polymorphisms under purifying (tension zone) or positive directional (ecotone) selection.

#### **Objective IC Is there evidence for selection on putatively adaptive phenotypes across hybrid zones?**

As mentioned above, stem pigmentation, macrohair density, and flowering time are phenotypes thought to be under selection in teosinte across an elevational gradient. Pigmented and pilose plants have an advantage in retaining heat at high elevation (for a discussion of highland adaptation in the context of maize see Eagles and Lothrop 1994). Additionally, *mexicana* flowers much earlier

than *parviflumis* (Rodriguez et al., 2006), which may represent an adaptation to shorter growing seasons at high elevation. We will combine the phenotypic data collected in our half-sibling progeny arrays in the common garden experiments proposed in Objective IA with our genome-wide marker data obtained during Objective IB from the mother of these arrays in order to evaluate evidence for selection on these potentially adaptive phenotypes. A method recently developed by Ovaskainen et al. (2011) and implemented in the software DRIFTSEL (Karhunen et al., 2013) is particularly suited to this purpose. The method builds upon the  $F_{ST}$ - $Q_{ST}$  framework (Leinonen et al., 2013) for comparison of population differentiation and quantitative phenotype divergence and allows the signature of selection on a given phenotypic trait to be distinguished from genetic drift. The strength of evidence for selection based on DRIFTSEL for putatively adaptive phenotypes (pigment, macrohairs, flowering time) will be compared to that of phenotypes with no *a priori* evidence of selection across an elevational gradient (culm diameter and leaf width). Under ecotone dynamics we would expect selection on these adaptive phenotypes across the entire elevational gradient of teosinte (*i.e.*, *parviflumis*, hybrids, *mexicana*), whereas tension zone dynamics would likely result in selection being detected within *parviflumis* and *mexicana* but not in hybrid zones. The evidence for selection on these putatively adaptive phenotypes will be considered in comparison to our measured neutral trait, the width of the leaf beneath the first lateral branch at the time of flowering.

In addition, we will conduct association analyses to connect genotype to phenotype using GBS data described in Objective IB and phenotypic data for potentially adaptive traits and traits gauging fitness. Association analysis will be conducted using TASSEL5.0 (Bradbury et al., 2007). Significant associations will then be cross-referenced with regions of excess ancestry from *parviflumis* or *mexicana* in hybrid populations and zones identified in Objective IB, particularly those that show evidence of selection based on additional population genetic summary statistics. This final combination of data and analyses could reveal the phenotypes, loci, and ancestry source under selection within hybrid zones.

#### **Preliminary Results:**

#### **Potential Challenges:**

Some of the population-genetic analyses proposed may have difficulty with the high rate of missingness and heterozygous error in GBS data. If these approaches do not work or force us to remove too much of our data, we will instead take advantage of approaches designed to estimate admixture from genotype-likelihoods calculated on low-coverage sequence data (Skotte et al., 2013). We have already designed pipelines to work with genotype-likelihoods (*e.g.* <https://github.com/arundurvasula/angsd-wrapper>) and can utilize these methods to calculate standard diversity statistics as well.

## **Objective II Determine the extent to which anthropogenic movement of maize has enabled hybridization among *Zea* taxa**

Following its domestication from *parviflumis*, maize spread rapidly across the Americas (Piperno and Flannery, 2001; Grobman et al., 2012), colonizing novel environments distinct from that inhabited by its wild ancestor. As its range expanded, maize came into contact with other wild teosinte

add here  
prelim growth  
chamber ex-  
pers. that  
suggest hy-  
brids better  
than par-  
ents in some  
pops.

add here  
that if field  
collects not  
work, we al-  
ready have  
smaller sets  
of seed from  
hybrid pops  
collected by  
collabs

that had been allopatric to *parviflumis* for long periods prior to domestication (Hufford et al., 2012b). Hybridization between maize and each of these taxa has been documented based on morphological (Wilkes, 1967, 1977) and genetic (Doebley, 1990; Fukunaga et al., 2005; Ross-Ibarra et al., 2009; van Heerwaarden et al., 2011) data, raising a number of questions about the role of gene flow in the recent evolution of both maize and its wild relatives.

In this objective, the Ross-Ibarra Laboratory will use population genetic approaches to address three questions that arise from this natural experiment. First, we will use maize and teosinte populations in Guatemala to investigate whether divergence time between taxa affects the possibility of adaptive introgression. Second, we will take advantage of multiple pairs of sympatric maize and teosinte populations from both Mexico and Guatemala to test hypotheses about the geographic scale of adaptive introgression. Finally, we will use population-level comparisons of maize and all the diploid taxa in *Zea* to test the hypothesis that maize has served as a bridge for gene flow between otherwise allopatric teosinte (Ross-Ibarra et al., 2009).

#### **Objective IIA Do the fitness consequences of gene flow from locally adapted populations depend on divergence?**

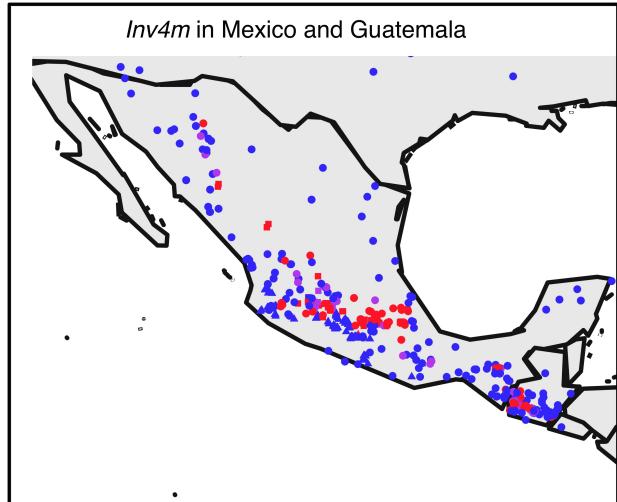
Most models of speciation predict that introgression between taxa should decrease with increasing divergence between them (Harrison and Larson, 2014), as differences contributing to postzygotic isolation are expected to increase as the square of divergence time (Orr and Turelli, 2001). Maize colonization of novel environments in Guatemala that are inhabited by distinct teosinte species offers an opportunity to test the importance of divergence in restricting the potential for adaptive introgression.

Following domestication, maize spread southward into Guatemala into conditions distinctly more tropical than those found in its center of origin in southwest Mexico. In comparison to southwest Mexico, Guatemalan winters are warmer, annual fluctuation in temperature is lower, and there is nearly double the annual precipitation. Upon arrival in Guatemala, maize came into contact with two new teosintes *Zea mays* ssp. *huehuetenangensis* (hereafter, *huehuetenangensis*) and *Zea luxurians* (hereafter, *luxurians*). Both teosintes exhibit a number of adaptations to tropical environments including differences in root architecture, flooding tolerance, and delayed flowering (Wilkes, 1967; Mano et al., 2006). Hybrids between maize and both taxa (Wilkes, 1967) have been observed, providing the opportunity for adaptive introgression.

If divergence impacts the potential for adaptive introgression, we predict different patterns of maize-teosinte hybridization for each of these taxa. Divergence time between maize and *huehuetenangensis* is not known, but it has been classified as a subspecies of *Zea mays* (Doebley et al., 1990) and we predict that adaptive introgression from *huehuetenangensis* is likely, as we have previously observed from subspecies *mexicana* (Hufford et al., 2013). *Zea luxurians*, however, exhibits a number of morphological differences from *Zea mays* (Doebley and Iltis, 1980), has a much larger genome (Tenaillon et al., 2011), and divergence time between the two taxa has been estimated at ~150,000 generations (Ross-Ibarra et al., 2009). We predict decreased introgression from *luxurians* due to the increased likelihood of loci causing postzygotic isolation. If adaptive introgression is observed, we predict it will be at fewer loci due to potential linkage with incompatibility factors. Because gene flow from colonizing, domesticated maize is likely to be maladaptive (Hufford et al., 2013), we also predict decreased introgression from maize into both teosinte taxa. Though incompatibilities are expected to be stronger in *luxurians*, this species also has a relatively

mention tcb1

Figure 3: Geographic distribution of alleles at a SNP diagnostic of the highland maize inversion *Inv4m*. Genotypes are shown in color (blue: homozygous standard, purple: heterozygous, red: homozygous inverted) and shapes represent taxa (squares: *mexicana*, triangles: *parviflora*, circles: domesticated maize). Data from Fang et al. (2012)



low effective population size (Ross-Ibarra et al., 2009) and selection against maladaptive introgression may be less efficient. Evidence supporting this possibility comes from the observation of alleles diagnostic of highland Mexican maize both in Guatemalan maize (Figure 3) and *luxurians* samples from Guatemala (Fang et al., 2012).

To test our predictions about the role of divergence time, the Ross-Ibarra Laboratory, in collaboration with Senior Personnel Claudia Calderon, will sample six populations each of *luxurians* and *huehuetenangensis* stratified across their elevational range in Guatemala. We will sample four sympatric sites for each teosinte, collecting samples from both the maize and teosinte populations at each site. Two populations as distant as possible from domesticated maize will be used as controls, taken to represent ancestral populations absent of admixture. Finally, four Guatemalan maize populations allopatric to teosinte will also be sampled, for a total of 24 maize and teosinte populations. We will be assisted in our collection by Hector Tuy of the Institute of Agriculture, Natural Resources and Environment (IARNA) in Guatemala (see attached letter of commitment). We will genotype 12 individuals from each population using GBS (as described in Objective IB but multiplexing 96 per lane), and analysis of introgression and selection will follow methods described in Objective IB. Evaluation of the adaptive nature of observed introgressions will follow Hufford et al. (2013), using data on QTL for putatively adaptive phenotypes (Omori and Mano, 2007; Mano et al., 2008) and small-scale growth-chamber experiments to compare phenotype and growth rate under different conditions.

In addition to our GBS data, we will sequence one allopatric and one sympatric individual each of *luxurians* and *huehuetenangensis* to  $\sim 30X$  using paired-end 150bp reads on an Illumina Hi-Seq 3000; no deep whole-genome sequence currently exists for either taxon. Combined with our existing, high-depth landrace maize sequences (see Preliminary Results), these sequences will allow additional tests of introgression, refine introgressed or selected regions and potentially identify candidate adaptive polymorphisms within the regions of interest.

These analyses will test the role of divergence time in limiting the potential for adaptive introgression and help establish whether observations of adaptive introgression from crop wild relatives may be a common occurrence that has facilitated the spread of domesticated taxa beyond their original habitat. The results will also provide useful baseline information on patterns of genetic

diversity in two teosinte taxa of conservation concern within Guatemala and of interest for novel root phenotypes for breeding including root angle, adventitious root formation, and the formation of aerenchyma (Omori and Mano, 2007; Mano and Omori, 2007).

### **Objective IIB Is introgression adaptive across multiple spatial scales?**

Due to their sessile nature, plants must adapt to their local environments. The geographical scale of adaptation varies widely, however, from large portions of a species range (Lowry and Willis, 2010; Fang et al., 2014) to extremely local adaptation on the scale of a few meters (Hamrick and Holden, 1979). While there are now several examples of introgression facilitating colonization and adaptation (reviewed in Bock et al. 2015) we know little about the geographic scale at which this occurs. Alleles that have spread throughout a wide geographic range in local populations have been tested in multiple genetic backgrounds and multiple microenvironments, and as such may have a higher likelihood of being adaptive in new populations via introgression. In contrast, the selective benefit of alleles that are adaptive on a very local scale in a single population may depend more strongly on genetic background or particular aspects of the microenvironment and may fare worse when introgressed into a new population. These arguments lead us to predict that adaptive introgression across maize populations found in sympatry with teosinte will be dominated by alleles beneficial over a larger geographic area. We should thus see parallel patterns across most populations, with relatively few loci showing evidence of adaptive introgression only in one population. Consistent with this prediction, our previous analysis of adaptive introgression from *mexicana* found highly similar patterns of introgression across all high-elevation sympatric populations, with few examples of introgressed regions at high frequency in a single population (Figure 4). However, the relatively low density SNP data in our prior work provide poor resolution to identify all but the strongest selection signatures (Tiffin and Ross-Ibarra, 2014) and suffer from ascertainment bias that may limit detection of diverged haplotypes.

To investigate the scale of adaptive introgression in maize, we will analyze a set of sympatric pairs of maize and teosinte populations from the highlands of Central Mexico (with *mexicana*), the Pacific coast of Mexico (with *parviflumis*), and the humid lowlands of Guatemala (with *luxurians* and *huehuetenangensis*). In addition to the populations sampled in Objective IIA, we will sample 12 maize and 12 teosinte from each of three *mexicana* and *parviflumis* sites previously identified as sympatric with domesticated maize (Hufford, 2010; Hufford et al., 2013). We will endeavor to sample different environments within the range of each teosinte to maximize our opportunity to identify differential local adaptation. Maize and teosinte individuals from each population will be genotyped using GBS, and we will test for introgression and selection following methods proposed in Objective IB. We will also make use of whole-genome resequencing data to better characterize introgressed haplotypes and identify potential candidate loci. Resequencing data will already be available for maize landraces (see Preliminary Results) and most teosinte (Objective IIA and Preliminary Results), but only two low-coverage inbred *mexicana* genomes are currently available (Chia et al., 2012). We will thus resequence two *mexicana* genomes to complement these data.

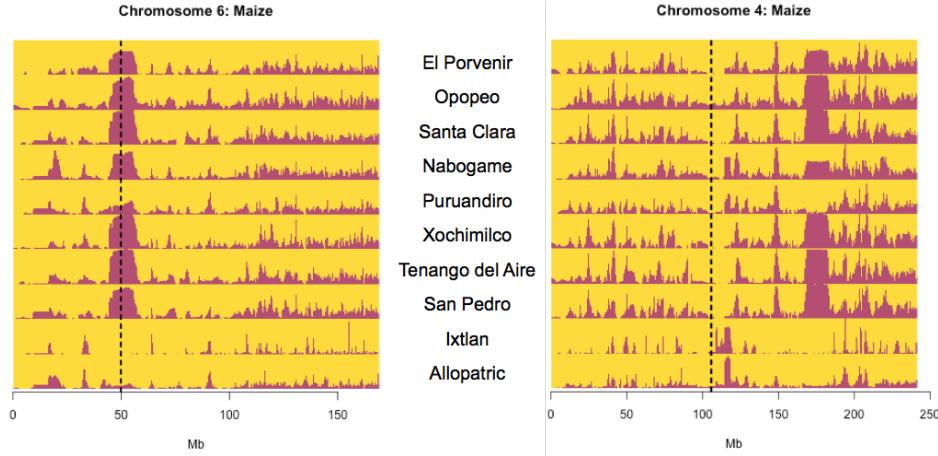


Figure 4: Introgression from *mexicana* to maize along chromosomes 4 and 6 (Hufford et al., 2013). Each row represents a population of maize, and the frequency of the *mexicana* allele at a given position is shown in dark red. Consistent signals of introgression are evidence in several chromosomal regions across all populations except the two lowest elevation sympatric populations (Ixtlan and Puruandiro) and the allopatric population.

#### Objective IIC Can a widespread species serve as a bridge for introgression among allopatric species?

While the role of endemic taxa in facilitating colonization or invasion has been investigated in some taxa (Bock et al., 2015), whether colonizing taxa have the potential to serve as a bridge for gene flow among allopatric endemic taxa has not been tested. Population genetic analysis of divergence in *Zea* using 26 loci identified evidence of recent admixture between *luxurians* and both *parviflumis* and *mexicana* (Ross-Ibarra et al., 2009). These teosinte are currently allopatric; while there is little fossil evidence with which to estimate ranges, ecological niche modeling suggests that the ranges of both *parviflumis* and *mexicana* have not changed appreciably for tens of thousands of years (Hufford et al., 2012b). Domesticated maize is currently found in sympatry with each of these taxa and is known to hybridize with each, suggesting the possibility that maize may have served as a bridge for gene flow between otherwise allopatric teosinte (Ross-Ibarra et al., 2009).

Our preliminary analyses of additional data are consistent with these conclusions. We have already shown that the inverted allele of the *Inv4m* inversion polymorphism has likely introgressed from *mexicana* into maize in the highlands of central Mexico. The *mexicana* haplotype at this locus is also found in highland maize in Guatemala (Figure 3) and in all of the samples of *mexicana* genotyped by Fang et al. (2012), suggesting maize may have facilitated the movement of the inverted allele from *mexicana* into *luxurians*. Analysis of individual genome sequences from *Zea diploperennis* (hereafter, *diploperennis*) and *luxurians* (Tenaillon et al., 2011) further support this idea, revealing evidence of admixture between *parviflumis* and *luxurians* as well as between maize and *diploperennis* (Figure 5).

Speciation and divergence in *Zea* is relatively recent, however, and an alternative explanation for at least some of these results is that polymorphisms found in ancestral populations continue to segregate in multiple taxa. Simple estimates of the length of shared haplotypes expected to

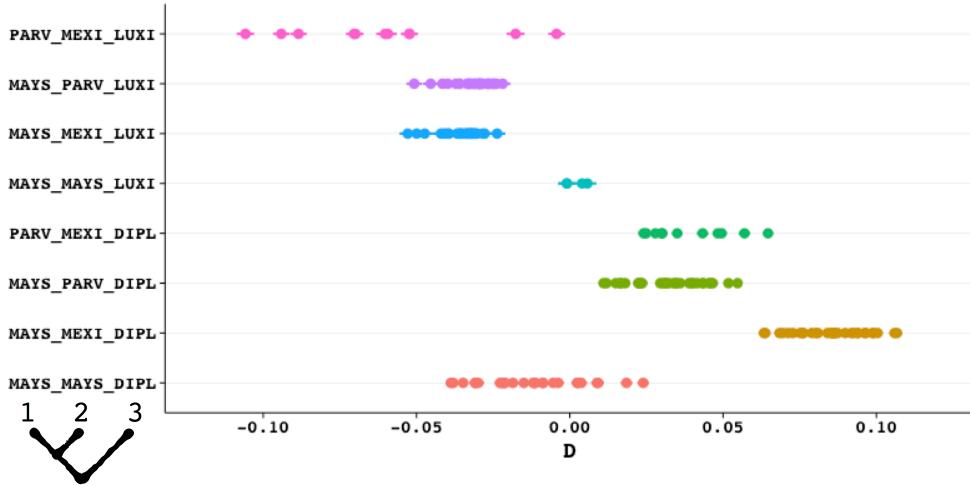
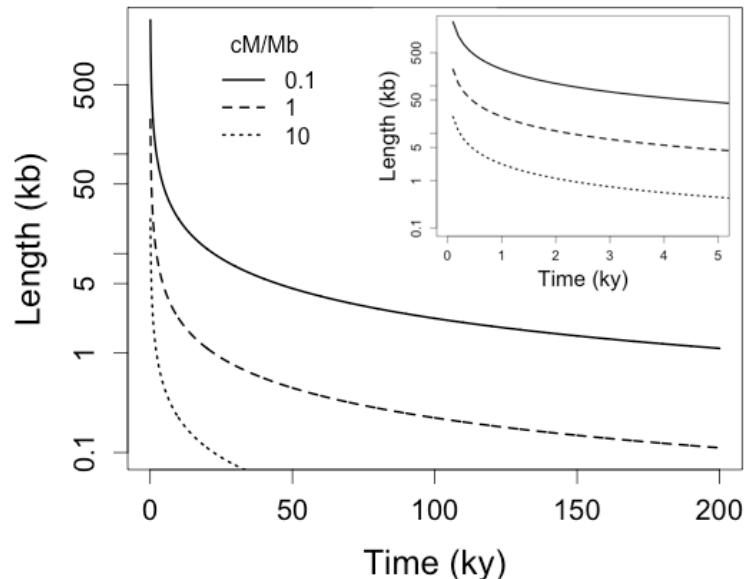


Figure 5: Genome-wide evidence for admixture among *Zea*. In the absence of admixture, taxa 1 and 2 on the tree should share similar numbers of derived alleles with taxa 3 due to the stochastic nature of incomplete lineage sorting. Admixture leads to deviations from this expectation as measured by the  $D$  statistic (Green et al., 2010). Negative values of  $D$  indicate admixture between taxa 1 and 3, while positive values indicate admixture between taxa 2 and 3.

be unbroken by recombination over the  $\sim 150,000$ -generation divergence time between *mexicana* and *luxurians* (Ross-Ibarra et al., 2009) suggest we might expect to see shared haplotypes of even several kb in length in low recombination regions of the genome (Figure 6).

We will make use of both genotyping and whole-genome sequence data to test whether observed patterns of haplotype sharing among allopatric teosinte are consistent with recent introgression via maize or can be better explained by segregating ancestral polymorphism. Although the limited sampling in Ross-Ibarra et al. (2009) only identified shared haplotypes between *mexicana* and *luxurians*, maize is known to hybridize with all taxa in the genus. Genome-wide genetic marker data do not currently exist for either *diploperennis* or its autotetraploid derivative *Zea perennis* (hereafter, *perennis*); we will sample 12 individuals from each of two accessions already available to us of both *diploperennis* and *perennis* for genotyping using GBS. We will also generate full-genome sequence for one *diploperennis* (in addition to one current sequence – see Preliminary Results) and an additional *luxurians*. These samples, combined with our other teosinte populations (Objective IIA and Objective IIB), will provide a representative sample of all wild teosinte populations. We will use window-based approaches (e.g., Martin et al., 2015) to estimate the size of putatively introgressed chromosomal segments from GBS data; these will be combined with more powerful haplotype-based analysis of full genome sequence data (Price et al., 2009). Haplotypes shared between two teosinte as a result of gene flow via a maize intermediate will be of recent origin given the timing of maize domestication and subsequent range expansion. Recent introgression events will result in longer shared haplotypes (Figure 6) and should be distinguishable from the genome-wide distribution of shared haplotype lengths as well as lengths predicted based on divergence time and local recombination rates. Estimates of local recombination rates will utilize either a 1cM-resolution map from maize (Wallace et al., 2014) or a teosinte map under development in

Figure 6: 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 maize genome.



the Ross-Ibarra lab (see Potential Challenges).

We will complement the comparative analysis of teosinte with a survey of maize diversity data. Although we cannot exhaustively test for the presence of putatively shared haplotypes in all of maize, we can survey publicly available data — including GBS genotypes for more than  $> 4,000$  maize landraces (Hearne et al., 2015) and whole genome sequence of more than 1000 diverse lines (<http://www.panzea.org/#!genotypes/cct1>) — to search for such haplotypes in domesticated maize.

**Preliminary Results** We have made arrangements with collaborators to assist in collection of maize, *luxurians* and *huehuetenangensis* from Guatemala. The Ross-Ibarra and Hufford labs already have sufficient seed for all other samples for Objective IIC and Objective IIB.

As part of a USDA grant on climate change to Drs. Ross-Ibarra and Hufford, we have sequenced more than 30 maize landrace accessions from across the range, including regions of sympatry with teosinte in Mexico and Guatemala. We have also recently sequenced multiple wild-collected *parviflumis* and a single *diploperennis* genome as part of another NSF-funded project. These samples will prove invaluable for analysis of introgression and selection; comparison of these samples with published data corroborates earlier Sanger data suggesting admixture between allopatric teosinte (Figure 5).

**Potential Challenges** No high-resolution genetic map currently exists for any teosinte, but we are currently working on producing such a map for *parviflumis* as part of a different NSF project. Evidence from comparisons among maize populations finds remarkable stability of the genetic map at a relatively coarse scale (Rodgers-Melnick et al., 2015) suggesting differences in the genetic map are unlikely to dramatically affect our estimates.

need suggestions as to challenges here. one possibility: if we don't get lux/huehue seeds, what do we do? other potential pitfalls?

## 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.

do we have a management plan? I think might be worth going to smaller font and adding a timeline. thoughts?

## 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 four years. All of the exchange students involved in the program have continued on to additional graduate work, and two have earned authorship on published or forthcoming papers from their internship. We will build upon the success of this program. Each year a graduate or undergraduate student from the Eguiarte group spending 2-3 months in either the Hufford or Ross-Ibarra Laboratory learning GBS methodology and/or population genomic analysis, and a student from the Hufford 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 students with valuable international research experience.

Senior Personnel Claudia Calderón 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.

add something about undergraduate researchers in my lab?

is this correct? every year?

## Results From Prior NSF Support

i edited both rare alleles and the centromere grant. include one or both as you see fit

## **NSF #1404974: US-Mexico Planning Visit and Workshop to Assess the Genomic Basis of Local Adaptation in Maize**

\$34,650. 09/01/14-08/31/15. PI Matthew Hufford, co-PI J. Ross-Ibarra, G. Coop, Senior Personnel S. Flint-Garcia, Collaborators R. Sawers and A. Cibrian-Jaramillo

**Intellectual merit** Through planning meetings and a phenotyping workshop in Mexico, this project has established a new international collaboration and laid the foundation for work proposed in a pending Plant Genome Research Program proposal. Planning meetings helped coordinate generation of preliminary data described in this proposal and the phenotyping workshop transferred high-throughput methods across our research groups.

**Broader impacts** Participants in the workshop included graduate students and postdoctoral scholars from the US and Mexico, providing STEM training and international scientific experience.

**Publications** Funding is for organizational purposes; no publications are expected from this award.

## **NSF #1238014: Biology of Rare Alleles in Maize and Its Wild Relatives**

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

**Intellectual merit** In the first two years of this proposal we have developed imputation approaches, found evidence for the importance of deleterious variants in determining heterosis, documented copy number variation in natural teosinte populations, and found population genetic evidence suggesting the importance of demography in patterning purifying selection across the genome.

**Broader impacts** The project so far has included 4 postdoctoral and 1 graduate trainees in the Ross-Ibarra lab. and the project GBS workshop and museum exhibit continue to be popular.

**Publications** Tiffin and Ross-Ibarra (2014); Takuno et al. (2015); da Fonseca et al. (2015); Hake and Ross-Ibarra (2015); Makarevitch et al. (2015)

## References Cited

- R. Abbott, D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. M?st, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. V?in?I?, J. B. W. Wolf, and D. Zinner. Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2):229–246, 2013.
- R. J. Abbott and A. C. Brennan. Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1648), 2014.
- E. Anderson. Hybridization of the habitat. *Evolution*, 2(1):1–9, 1948.
- E. Anderson and J. Stebbins, G. L. Hybridization as an evolutionary stimulus. *Evolution*, 8(4):378–388, 1954.
- D. G. Bock, C. Caseys, R. D. Cousens, M. A. Hahn, S. M. Heredia, S. HÄEbner, K. G. Turner, K. D. Whitney, and L. H. Rieseberg. What we still don't know about invasion genetics. *Molecular Ecology*, 24(9):2277–2297, 2015.
- P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635, 2007.
- J. Chia, C. Song, P. Bradbury, D. Costich, N. de Leon, J. Doebley, R. Elshire, B. Gaut, L. Geller, J. Glaubitz, M. Gore, K. Guill, J. Holland, M. Hufford, J. Lai, M. Li, X. Liu, Y. Lu, R. McCombie, R. Nelson, J. Poland, B. Prasanna, T. Pyhajarvi, T. Rong, R. Sekhon, Q. Sun, M. Tenailion, F. Tian, J. Wang, X. Xu, Z. Zhang, S. Kaepller, J. Ross-Ibarra, M. McMullen, E. 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.
- T. H. G. Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405):94–98, 2012.
- M. Currat, M. Ruedi, R. J. Petit, and L. Excoffier. The hidden side of invasions: Massive introgression by local genes. *Evolution*, 62(8):1908–1920, 2008.
- R. R. da Fonseca, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. A. Samaniego, C. Carøe, M. C. Ávila-Arcos, D. E. Hufnagel, et al. The origin and evolution of maize in the southwestern united states. *Nature Plants*, 1(1), 2015.
- C. M. Díez, B. S. Gaut, E. Meca, E. Scheinvar, S. Montes-Hernandez, L. E. Eguiarte, and M. I. Tenaillon. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*, 199(1):264–276, 2013.
- J. Doebley. Molecular evidence and the evolution of maize. *Economic Botany*, 44(3):6–27, 1990.
- J. Doebley et al. Molecular systematics of zea (gramineae). *Maydica*, 35(2):143–50, 1990.

- J. F. Doebley and H. H. Iltis. Taxonomy of zea (gramineae). i. a subgeneric classification with key to taxa. *American Journal of Botany*, pages 982–993, 1980.
- H. A. Eagles and J. E. Lothrop. Highland maize from central mexico—its origin, characteristics, and use in breeding programs. *Crop Science*, 34:11–19, 1994.
- N. C. Ellstrand. Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, 101(5):737–753, 2014. doi: 10.3732/ajb.1400024. URL <http://www.amjbot.org/content/101/5/737.abstract>.
- R. Elshire, J. Glaubitz, Q. Sun, J. Poland, K. Kawamoto, E. Buckler, and S. Mitchell. A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PLoS One*, 6(5):e19379, 2011.
- Z. Fang, T. Pyhajarvi, A. Weber, R. Dawe, J. Glaubitz, J. Gonzalez Jde, C. Ross-Ibarra, J. Doebley, P. Morrell, and J. Ross-Ibarra. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, 191(3):883–894, 2012.
- Z. Fang, A. M. Gonzales, M. T. Clegg, K. P. Smith, G. J. Muehlbauer, B. J. Steffenson, and P. L. Morrell. Two genomic regions contribute disproportionately to geographic differentiation in wild barley. *G3: Genes|Genomes|Genetics*, 4(7):1193–1203, 2014.
- K. Fukunaga, J. Hill, Y. Vigouroux, Y. Matsuoka, J. Sanchez, K. Liu, E. Buckler, and J. Doebley. Genetic diversity and population structure of teosinte. *Genetics*, 169(4):2241–2254, 2005.
- A. Geneva, C. Muirhead, and L. Lovato. An improved sequence measure used to scan genomes for regions of recent gene flow. *arXiv preprint: 1403.1552*, 2014.
- J. C. Glaubitz, T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler. Tassel-gbs: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, 9(2):e90346, 02 2014.
- M. Gore, J. Chia, R. Elshire, Q. Sun, E. Ersoz, B. Hurwitz, J. Peiffer, M. McMullen, G. Grills, J. Ross-Ibarra, D. Ware, and E. Buckler. A first-generation haplotype map of maize. *Science*, 326(5956):1115–1117, 2009.
- R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, et al. A draft sequence of the neandertal genome. *science*, 328(5979):710–722, 2010.
- A. Grobman, D. Bonavia, T. D. Dillehay, D. R. Piperno, J. Iriarte, and I. Holst. Preceramic maize from paredones and huaca prieta, peru. *Proceedings of the National Academy of Sciences*, 109 (5):1755–1759, 2012.
- S. Hake and J. Ross-Ibarra. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*, 4:e05861, 2015.
- J. Hamrick and L. Holden. Influence of microhabitat heterogeneity on gene frequency distribution and gametic phase disequilibrium in avena barbata. *Evolution*, pages 521–533, 1979.

- 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.
- R. G. Harrison and E. L. Larson. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105(S1):795–809, 2014.
- Z. He, W. Zhai, H. Wen, T. Tang, Y. Wang, X. Lu, A. J. Greenberg, R. R. Hudson, C.-I. Wu, and S. Shi. Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet*, 7(6):e1002100, 06 2011.
- S. Hearne, C. Chen, E. Buckler, and S. Mitchell. Unimputed gbs derived snps for maize landrace accessions represented in the seed-maize gwas panel. <http://hdl.handle.net/11529/10034>, 2015. Accessed: 2015-02-16.
- E. Huerta-Sánchez, X. Jin, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, et al. Altitude adaptation in tibetans caused by introgression of denisovan-like dna. *Nature*, 2014.
- M. Hufford, X. Xu, J. van Heerwaarden, T. Pyhajarvi, J. Chia, R. Cartwright, R. Elshire, J. Glaubitz, K. Guill, S. Kaepller, J. Lai, P. Morrell, L. Shannon, C. Song, N. Springer, R. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M. McMullen, D. Ware, E. Buckler, S. Yang, and J. Ross-Ibarra. Comparative population genomics of maize domestication and improvement. *Nat Genet*, 44(7):808–811, 2012a.
- M. Hufford, P. Lubinksy, T. Pyhajarvi, M. Devengenzo, N. Ellstrand, and J. Ross-Ibarra. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477, 2013.
- M. B. Hufford. *Genetic and ecological approaches to guide conservation of teosinte (*Zea mays ssp. parviglumis*), the wild progenitor of maize*. PhD thesis, University of California, Davis, 2010.
- M. B. Hufford, E. Martínez-Meyer, B. S. Gaut, L. E. Eguiarte, and M. I. Tenaillon. Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PloS one*, 7(11):e47659, 2012b.
- M. Karhunen, J. Meril??, T. Leinonen, J. M. Cano, and O. Ovaskainen. driftsel: an r package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources*, 13(4):746–754, 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.
- D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush. Inference of population structure using dense haplotype data. *PLoS Genet*, 8(1):e1002453, 01 2012.
- T. Leinonen, R. S. McCairns, R. B. O'Hara, and J. Merilä. Qst-fst comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, 14(3):179–190, 2013.

- D. B. Lowry and J. H. Willis. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol*, 8(9):e1000500, 09 2010.
- K. Lucek, D. Roy, E. Bezault, A. Sivasundar, and O. Seehausen. Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in switzerland. *Molecular Ecology*, 19(18):3995–4011, 2010.
- I. Makarevitch, A. J. Waters, P. T. West, M. Stitzer, C. N. Hirsch, J. Ross-Ibarra, and N. M. Springer. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genetics*, 11(1):e1004915, 2015.
- J. Mallet. Hybrid speciation. *Nature*, 446(7133):279–283, 2007.
- Y. Mano and F. Omori. Breeding for flooding tolerant maize using teosinte as a germplasm resource. *Plant Root*, 1:17–21, 2007.
- Y. Mano, F. Omori, T. Takamizo, B. Kindiger, R. Bird, and C. Loaisiga. Variation for root aerenchyma formation in flooded and non-flooded maize and teosinte seedlings. *Plant and Soil*, 281(1-2):269–279, 2006.
- Y. Mano, F. Omori, B. Kindiger, and H. Takahashi. A linkage map of maize  $\times$  teosinte zea luxurians and identification of qtls controlling root aerenchyma formation. *Molecular breeding*, 21(3):327–337, 2008.
- S. H. Martin, J. W. Davey, and C. D. Jiggins. Evaluating the use of abba–baba statistics to locate introgressed loci. *Molecular biology and evolution*, 32(1):244–257, 2015.
- Y. Matsuoka, Y. Vigouroux, M. 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.
- S. Mezmouk and J. Ross-Ibarra. The pattern and distribution of deleterious mutations in maize. *G3: Genes| Genomes| Genetics*, 4(1):163–171, 2014.
- F. Omori and Y. Mano. Qtl mapping of root angle in f2 populations from maize ???b73???\* teosinte ???zea luxurians???. *Plant Root*, 1:57–65, 2007.
- H. A. Orr and M. Turelli. The evolution of postzygotic isolation: accumulating dobzhansky-muller incompatibilities. *Evolution*, 55(6):1085–1094, 2001.
- O. Ovaskainen, M. Karhunen, C. Zheng, J. M. C. Arias, and J. Meril. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics*, 189(2):621–632, 2011.
- 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.

- J. W. Poelstra, N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. M??ller, 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.
- A. Price, A. Tandon, N. Patterson, K. Barnes, N. Rafaels, I. Ruczinski, T. 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.
- J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
- T. Pyhajarvi, M. Hufford, S. Mezmouk, and J. Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome Biol Evol*, 5(9):1594–1609, 2013.
- J. B. Rasmussen, M. D. Robinson, A. Hontela, and D. 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.
- 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.
- E. Rodgers-Melnick, P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, 112(12):3823–3828, 2015.
- J. G. Rodriguez, J. J. Sanchez, B. Baltazar, L. De la Cruz, F. Santacruz-Ruvalcaba, J. Ron, and J. B. Schoper. Characterization of floral morphology and synchrony among Zea species in Mexico. *MAYDICA*, 51(2):383–398, 2006.
- J. Ross-Ibarra, M. Tenaillon, and B. Gaut. Historical divergence and gene flow in the genus zea. *Genetics*, 181(4):1399–1413, 2009.
- Q. R. Shurtliff. Mammalian hybrid zones: a review. *Mammal Review*, 43(1):1–21, 2013.
- L. Skotte, T. S. Korneliussen, and A. Albrechtsen. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3):693–702, 2013.
- K. L. Smith, J. M. Hale, M. R. Kearney, J. J. Austin, and J. Melville. Molecular patterns of introgression in a classic hybrid zone between the australian tree frogs, *litoria ewingii* and *I.?????paraewingi*: evidence of a tension zone. *Molecular Ecology*, 22(7):1869–1883, 2013.
- F. Staubach, A. Lorenc, P. W. Messer, K. Tang, D. A. Petrov, and D. Tautz. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet*, 8(8):e1002891, 08 2012.
- G. L. Stebbins. The role of hybridization in evolution. *Proceedings of the American Philosophical Society*, 103(2):231–251, 1959.

- S. Takuno, P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. Independent molecular basis of convergent highland adaptation in maize. *Genetics*, 2015. doi: 10.1534/genetics.115.178327. URL <http://www.genetics.org/content/early/2015/06/15/genetics.115.178327.abstract>.
- M. I. Tenaillon, M. B. Hufford, B. S. Gaut, and J. Ross-Ibarra. Genome size and transposable element content as determined by high-throughput sequencing in maize and zea luxurians. *Genome biology and evolution*, 3:219–229, 2011.
- P. Tiffin and J. Ross-Ibarra. Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12):673–680, 2014.
- J. van Heerwaarden, J. Doebley, W. Briggs, J. Glaubitz, M. 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.
- J. J. Vitti, S. R. Grossman, and P. C. Sabeti. Detecting natural selection in genomic data. *Annual Review of Genetics*, 47(1):97–120, 2013.
- J. G. Wallace, P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E. S. Buckler. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet*, 10(12): e1004845, 2014.
- H. Wilkes. *Teosinte: the closest relative of maize*. PhD thesis, Harvard University, 1967.
- H. G. Wilkes. Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize. *Economic Botany*, 31(3):254–293, 1977.

## **Facilities, Equipment, and Other Resources**

### **Iowa State University**

Project components completed in the Hufford Laboratory at Iowa State University (ISU) will include DNA isolation and preparation for genotyping and population genetic analysis of genotyping and full-genome sequence data. The Hufford Laboratory has all equipment necessary for DNA isolations, quality control and preparation for genotyping 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. Genotyping will be carried out using a reduced representation approach to next-generation sequencing known as Genotyping by Sequencing (GBS) at the Genomic Diversity Facility at Cornell University (see letter of commitment from Sharon Mitchell). Full-genome sequencing will be carried out at the DNA Facility at ISU, which provides access to cutting-edge genomic technologies 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. Broader impacts at ISU will be facilitated by the *Symbi* program. *Symbi* was Iowa's first GK12 program and represents a partnership between the Des Moines Public School System and Iowa State University. Staff members from *Symbi* have experience facilitating over 30 previous graduate student fellows in communicating their science to grade school students and will assist the graduate student funded by this project to do the same (see letter of commitment from *Symbi*).

### **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. Dr. Ross-Ibarra 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 bioanalyzer 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.

### **Partners in Mexico and Guatemala**

Senior Personnel on this project include Luis Eguiarte of UNAM in Mexico City and Claudia Calderón, a Guatemalan national. This project will benefit greatly from their many years of experience working in the field in Mexico and Guatemala respectively. In addition we have confirmed commitments from Ruairidh Sawers of Langebio in Irapuato, Mexico and Salvador Montes-Hernandez of Inifap in Celaya, Mexico (see attached commitment letters) to assist with common garden experiments. Between Dr. Sawers and Dr. Montes-Hernandez, our collaborators have ample experience growing maize and teosinte in nurseries located on the West Coast (Valle de Banderas, Nayarit), in Central Mexico (Irapuato and Celaya, Guanajuato), and in the high valleys of Central Mexico (Queretaro, Estado de Mexico). They also 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 (teosinte and maize, Solanaceae, and Cucurbitaceae). A commitment has also been confirmed (see attached letter) from Mario Fuentes López to assist with teosinte collection in Guatemala.

# Data Management Plan

## Data Types

This proposal will generate genotype and full-genome sequence data, phenotype data, analytical code, germplasm, and publications.

## Data Archiving, Plan for Sharing, Public Access Policy

**Genotype and Sequence Data** All data will be made publicly available and stored online. However, prior to public release, all data will be hosted locally. Drs. Hufford and Ross-Ibarra will maintain a backup of all raw genotyping and sequencing data. Dr. Hufford has access to 144Tb of free data storage through the College of Liberal Arts and Sciences at Iowa State and Dr. Ross-Ibarra maintains a DROBO distributed backup server (currently > 8Tb of free space) which is robust to single disk failure. All sequence data (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. Just prior to publication, genotypes will be made publicly available via the Figshare website ([www.figshare.com](http://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](http://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.

**Phenotype Data** Phenotypic data will be recorded digitally in the field using a high-throughput protocol developed by Dr. Sherry Flint-Garcia at USDA and the University of Missouri. Data will be uploaded at the end of each day into the FieldBook database developed by Dr. Flint-Garcia 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). Phenotypic data will then be uploaded to Figshare, along with appropriate metadata including plant ID, data collector, field location. Data on Figshare are 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.

as in a web DOI? or is this just a unique ID? i thought DOI was a specific thing for websites?

**Analytical Software and Code** 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. Hufford and Ross-Ibarra have already done this extensively (see <https://github.com/mbhufford> and <https://github.com/rossibarra> and <https://github.com/rilab>). Publication of all code will ensure reproducibility of all analyses conducted. All data and code 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.

**Germplasm** Sample accession data will be securely stored in MySQL servers hosted at Iowa State University and 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 and Guatemala 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, which provides public access to seed.

**Publications** 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.

## **Supplementary Documentation**

### **Postdoctoral Researcher Mentoring Plan**

The current proposal requests funding for one postdoctoral researchers at UC Davis and one at Iowa State. We also hope additional postdocs may join the group via alternative funding opportunities (fellowships, etc.) and anticipate that postdocs funded on other grants 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 Hufford and Ross-Ibarra were both postdoctoral scholars on NSF-funded programs. For this project, each PI will act as mentor and supervisor for postdocs in their lab, holding 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. Both labs will have internal lab meetings at which postdocs and graduate students will be given numerous opportunities to hone their presentation skills. Both labs currently host weekly journal clubs in which postdocs gain additional training in reading, presenting, and dissecting scientific literature. Members of the Ross-Ibarra lab are also encouraged to write critiques of the papers read in journal club, which are occasionally posted as blog posts and have engendered dialog with the article's authors. This provides excellent training in reviewing and scientific communication. 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. Both 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. The postdoc at UC Davis will be explicitly tasked with helping recruit undergraduate students and mentoring and supervising their research. We have included \$1,000 per year in the budget for the postdoc to use to support undergraduate student research. Such efforts in our labs have been very successful, with postdoc-mentored students presenting conference posters on their research or earning authorship on papers. Lab alumni have confirmed the utility of supervisory experience in applying for jobs, especially in industry.

Postdocs will be encouraged to write and apply for external funding, including fellowships and grant proposals. The Ross-Ibarra lab has 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.

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. The Ross-Ibarra lab has regular contact and collaboration with both USDA researchers and the seed industry, and postdocs interested in careers outside of universities should have ample opportunity to explore these options. To date, the Ross-Ibarra lab has successfully placed postdocs in government, industry, and tenure-track faculty positions and continues to work to support postdocs pursuing a wide range of careers.