

## Project Summary

**Overview** The proposed project will investigate the genome-wide effects of hybridization and introgression in the genus *Zea*. First, investigators will study two independent hybrid zones that naturally occur in mid-elevations of Mexico between the lowland-adapted *Z. mays* ssp. *parviflumis* (hereafter, *parviflumis*) and the highland-adapted *Z. mays* ssp. *mexicana* (hereafter, *mexicana*). Through field collections, generation of phenotypic data in common garden studies, and genotyping, the investigators will assess how fitness of these taxa varies across a hybrid zone, the level of convergence in the genetic architecture of introgression across independent hybrid populations and zones, and the evidence for selection on putatively adaptive phenotypic traits across an elevation gradient. Second, investigators will determine the impact of hybridization and introgression between domesticated maize (*Z. mays* ssp. *mays*) and wild *Zea* as maize spread from its center of origin. Population genomic analyses of sympatric collections will be used to test hypotheses about the significance of divergence time in patterning introgression between these taxa, the geographic scale of adaptive introgression, and the potential for maize, a widespread taxon, to serve as a bridge for gene flow between otherwise allopatric and narrowly-distributed *Zea* species.

**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 by 1) Generating novel diversity that has potentially been adaptive in two independent hybrid zones of *parviflumis* and *mexicana*; and 2) Facilitating the spread of maize across the Americas through transference of local adaptation from local wild species to maize. These investigations will generate basic knowledge on the level of convergence in introgression in replicate hybrid zones and the role of introgression in facilitating rapid local adaptation during the spread of a colonizing species.

**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. The investigators have an excellent track record of providing training opportunities for both undergraduate and graduate students in laboratory, computational, and field-based research, 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. In addition, a teaching module on hybridization will be developed for the undergraduate core course "Principles of Biology" at Iowa State University based on the research proposed here. This module will actively engage undergraduates in STEM-based research and will serve as a recruiting tool for independent research supported by this project in the investigators' laboratories. Finally, a proposed exchange program will 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 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 extent of introgression along a chromosome, differentiate neutral from adaptive introgression, and evaluate whether distinct hybrid populations show convergent patterns of introgression. With this expanded perspective, we propose here to 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.

The genus *Zea* includes four species all thought to have diverged within the last few hundred thousand years (Ross-Ibarra et al., 2009). *Zea mays* consists of four annual subspecies: the domesticated ssp. *mays*, the Mexican taxa ssp. *parviglumis* and ssp. *mexicana*, and the Guatemalan ssp. *huehuetenangensis*. *Zea luxurians* (including the populations referred to as *Zea nicaraguensis*) is an annual, flood-tolerant species found in Guatemala, Honduras, and Nicaragua. Finally, the diploid *Zea diploperennis* and its autotetraploid derivative *Zea perennis* are perennial species narrowly distributed along the mountain slopes of western Mexico. 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 investigate Anderson's (1954) hypotheses about the role of hybridization and introgression.

First, analysis of hybridization in the parapatric subspecies *parviglumis* and *mexicana* will allow us to answer questions about the potential for hybridization to generate adaptive variation and how selection acts to pattern gene flow. Second, evaluation of introgression between maize and teosinte will inform our understanding of the role of adaptive introgression during colonization of novel environments and test whether widespread species can act as bridges for gene flow among allopatric congeners.

This proposal builds 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 addressed here. In addition, over the last year we have conducted further 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; Pyhäjärvi 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 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 shaped hybridization in *Zea*

Maize was domesticated in southwest Mexico from *parviflora* ~9,000 BP (Matsuoka et al., 2002) and quickly spread throughout the Americas, bringing it into contact with each of the wild teosintes. Hybridization has been observed everywhere maize grows in sympatry with teosinte, raising questions about the interaction between colonization and gene flow in *Zea*. Our population genetic analyses will assess the following questions:

- A. *Does the potential for adaptive introgression depend on divergence time?*
- 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). Recent work has focused on defined and stable regions of hybridization, referred to as hybrid zones, have been discovered in a number of taxa (reviewed in Harrison, 1993; Shuriff, 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). 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, we are only beginning to understand the consequences of hybridization in patterning diversity among populations and across genomes. 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). Here, we propose to expand on Anderson's (1954) hypotheses by investigating the role of hybridization in generating novel genomic variation and the extent to which anthropogenic movement has influenced patterns of hybridization and subsequent evolution. The *Zea* study system is uniquely suited to address these questions from a genomics perspective. In addition to its status as a model genetic organism with well-developed genomic resources, maize and its wild relatives have an evolutionary past deeply affected by hybridization and introgression at multiple taxonomic levels.

## 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* (Wilkes, 1967). 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., 2012c). 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 analysis of individuals from one such population revealed abundant small blocks of introgression, suggesting long-term gene flow between subspecies (Pyhäjärvi 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 teosinte taxa and maize landraces (*i.e.*, traditional open-pollinated varieties) from across the Americas. While the low density of markers in these data precludes genome-wide inferences and haplotype-based analyses, the extensive sampling makes this an ideal resource for guiding future research.

We used the software STRUCTURE (Pritchard et al., 2000) to calculate assignment probability of *parviglumis* and *mexicana* samples to two groups. 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* (Figure 2A). Hybrid populations

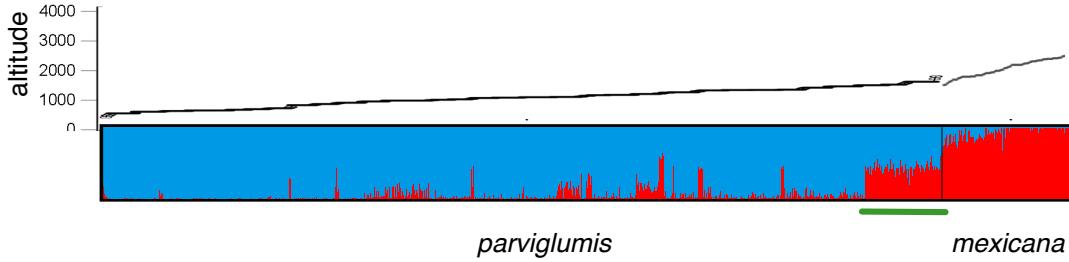


Figure 1: Assignment of *parviglumis* and *mexicana* individuals to K=2 groups using STRUCTURE (Pritchard et al., 2000). Individuals are sorted by increasing elevation, indicated above the bar chart. Individuals from hybrid zone populations are underscored in green.

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 (Figure 2B-C), suggesting that hybrid populations from distinct environments may vary in the proportion of ancestry from each subspecies in a manner that is adaptive. Estimates of genetic differentiation ( $F_{ST}$ ; data not shown) also show that both hybrid zones are less differentiated from *mexicana* and *parviglumis* than these subspecies are from each other. Finally, in preliminary growth chamber experiments we have observed that plants from both the Jalisco and Balsas hybrid zones exhibit stem pigmentation and macrohair abundance intermediate to *parviglumis* and *mexicana*.

These populations provide us with naturally replicated experiments in which to address Anderson's (1954) hypothesis that hybridization will generate novel adaptive variation. The Hufford Laboratory and Senior Personnel Luis Eguiarte will investigate these questions in two hybrid zones of *mexicana* and *parviglumis* through targeted collections, generation of phenotype, genotype and 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?**

There are two dominant evolutionary hypotheses to explain how and why hybrid zones are stable over time. The tension zone hypothesis posits that hybrids have lower fitness relative to their parental taxa across environments due to pre- or post-zygotic hybrid incompatibilities. Tension zones arise at the interface of parental taxa and are typically narrow clines due to the effects of selection against hybrids (Abbott and Brennan, 2014). The bounded hybrid superiority or ecotone hypothesis maintains that hybrids have a selective advantage in intermediate environmental conditions relative to parental species. Under this model, hybrid populations are distributed across an area transitioning between parental habitats (Abbott and Brennan, 2014). Given that hybrid populations in the eastern Balsas and eastern Jalisco span a relatively large area (Figure 2) across environmental conditions intermediate to their parental taxa, we predict that both *parviglumis-mexicana*

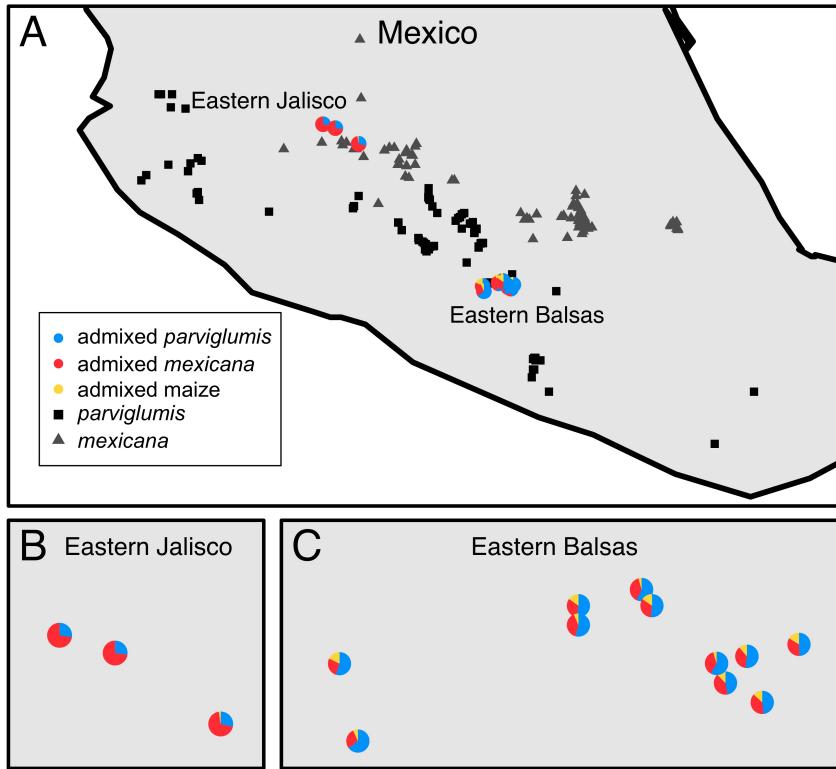


Figure 2: A) Location of two putative hybrid zones of *mexicana* and *parviglumis*. Hybrid populations are represented as pie charts with proportion assigned to *mexicana*, *parviglumis*, and maize groups. Zoomed-in views show the Eastern Jalisco (B) and Eastern Balsas (C) hybrid populations.

hybrid zones will exhibit ecotone dynamics, with selection favoring hybrids across much of the region. We will test this hypothesis by assessing fitness and variation at putatively adaptive phenotypes across both hybrid and parental populations in common garden experiments conducted in Mexico at three elevations: 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*.

We have access to seed from a large number of *mexicana* and *parviglumis* from previous collections, and Senior Personnel Luis Eguiarte has recently collected altitudinal transects of *parviglumis* 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 *parviglumis* 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  $1\ m^2$  quadrats), slope of the terrain, and canopy cover. 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 and Objective IC.

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 is the best option for increasing seed from all populations in a single environment; we have previously had success growing both *parviflumis* and *mexicana* at this site. Forty-five plants per population (three from each sampling site) will be grown and allowed to open-pollinate in isolated plots surrounded by a buffer of emasculated maize plants. 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: Bucerias, Nayarit (40m), Irapuato, Guanajuato (1700m) and Metepec, Mexico (2600m). We have experience working in all three sites and have confirmed that our common gardens will be logistically 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 other traits will be analyzed in Objective IC.

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

While we are beginning to understand some of the factors that determine the fate of hybrid individuals and populations, we know little about how the outcome of a hybridization event depends on context and stochastic factors. Under an ecotone model, for example, is the novel adaptive variation hypothesized by Anderson and Stebbins (1954) due to a small number of loci of large effect seen in every hybrid population, or do different hybrid populations exploit unique genetic combinations of parental alleles to adapt to local conditions? Instead, if populations are evolving under tension zone dynamics, do the same loci contribute to hybrid inviability or infertility in every population? The replicated *mexicana*-*parviflumis* hybrid zones to be studied in our proposed project provide an opportunity to test these ideas by assessing evidence of convergence of the genetic architecture – the number and effect size of selected loci – of differences in hybrid fitness.

To accomplish this, we will search for genome-wide patterns of introgression and selection in

the populations sampled in Objective IA. 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. DNA for genotyping will be isolated using the Qiagen DNeasy Plant Mini Kit. Samples will be genotyped using GBS (Genotyping By Sequencing; Elshire et al. 2011) at the Genomic Diversity Facility at Cornell University. A rapid bioinformatics pipeline (TASSLE-GBS; Glaubitz et al., 2014) allows genotyping at  $\sim 1,000,000$  SNPs in *Zea* using standard GBS data; to date this method has been implemented to genotype tens of thousands of maize and teosinte samples. 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), 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 (Chia et al., 2012; Hufford et al., 2012b; 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 we will use a sliding window approach to calculate a number of genotype-based statistics used to quantify differentiation and gene flow, including  $F_{ST}$ , the proportion of shared and fixed variants, relative levels of nucleotide diversity (Geneva et al., 2014) and 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*; these latter approaches will also be applied to our whole-genome sequencing.

We will evaluate *parviflumis* and *mexicana* ancestry on a site-by-site basis across each hybrid genome and determine 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 and potentially allow the identification of candidate causal polymorphisms under purifying (tension zone) or positive directional (ecotone) selection.

Evidence of convergence in the genetic architecture among hybrid populations will support an important role for a limited number of loci in determining the outcomes of hybridization events. If we see little overlap, however, there are at least two possible explanations. First, different phenotypes may be under selection in different hybrid zones (Objective IC). Second, phenotypic evolution in different hybrid populations may make use of different loci segregating in the parental taxa. We have previously observed a similar pattern for high-elevation adaptation in maize landraces (Takuno et al., 2015), and such an outcome seems plausible given the extent of genetic variation in both parental taxa.

## **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 *parviflora* (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. Under ecotone dynamics we would expect selection on these adaptive phenotypes across the entire elevational gradient of teosinte (*i.e.*, *parviflora*, hybrids, *mexicana*), whereas tension zone dynamics would likely result in selection being detected within *parviflora* 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 *parviflora* 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.

**Potential Challenges:** Collection of samples in the eastern Balsas hybrid zone may not be possible due to safety concerns raised by recent violence in the Mexican state of Guerrero. Should our Mexican collaborators advise against sample collection in this region, we will increase seed of existing accessions from these populations for common gardens and genotyping. In addition, some of the population-genetic analyses proposed here 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 many loci, 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 shaped hybridization in *Zea***

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 that had been allopatric to *parviflumis* for long periods prior to domestication (Hufford et al., 2012c). 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 and providing a unique opportunity to test the impact of anthropogenic change on the potential for hybridization among taxa (Anderson and Stebbins, 1954).

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 sample 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).

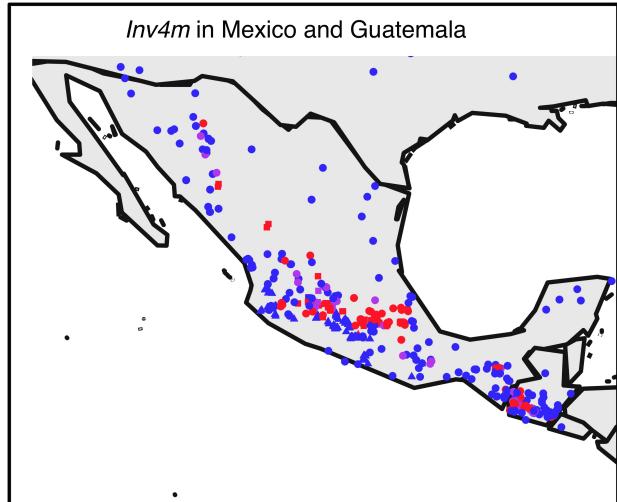
### **Objective IIA Does the potential for adaptive introgression depend on divergence time?**

Most models of speciation predict that introgression between taxa should decrease with increasing divergence between them (Harrison and Larson, 2014), as the establishment of mechanisms contributing to isolation (e.g. postzygotic isolation; Orr and Turelli, 2001) is expected to accelerate over time. 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),

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: *parviflumis*, circles: domesticated maize). Data from Fang et al. (2012)



has a much larger genome (Tenaillon et al., 2011), and divergence time between the two taxa has been estimated at  $\sim$ 150,000 generations (Ross-Ibarra et al., 2009). Although 150,000 generations is not extremely long, in less than half this time alleles at multiple loci conferring post-mating prezygotic isolation from maize have established in *mexicana* (Evans and Kermicle, 2001; Kermicle and Evans, 2010; Kermicle et al., 2006). Prezygotic isolation mechanisms in plants are not thought to evolve more rapidly than postzygotic factors (Widmer et al., 2009), and we thus predict that the increased genetic divergence from *luxurians* is sufficient to lead to decreased gene flow with maize. Although we will not be able to characterize the geographic extent of hybridization with *luxurians*, evidence of selection against introgression would suggest tension zone dynamics as described in Objective IA. If we do observe adaptive introgression, these ideas predict it will nonetheless 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 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 in year one of the grant. 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). Genotyping and analysis of introgression will closely follow methods described in Objective IB. We will genotype 12 individuals from each population using GBS (but multiplexing

96 per lane). 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 ~ 30X using paired-end 150bp reads on an Illumina Hi-Seq 3000; no deep whole-genome sequence currently exists for either taxon. We will complement these with existing data from five Guatemalan landrace maize individuals that were sequenced to high depth as part of a USDA-funded project. 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 *parviflora*), and the humid lowlands of Guatemala (with *luxurians* and *huehuetenangensis*). In addition to the populations sampled in Objective IIA, in year two of the

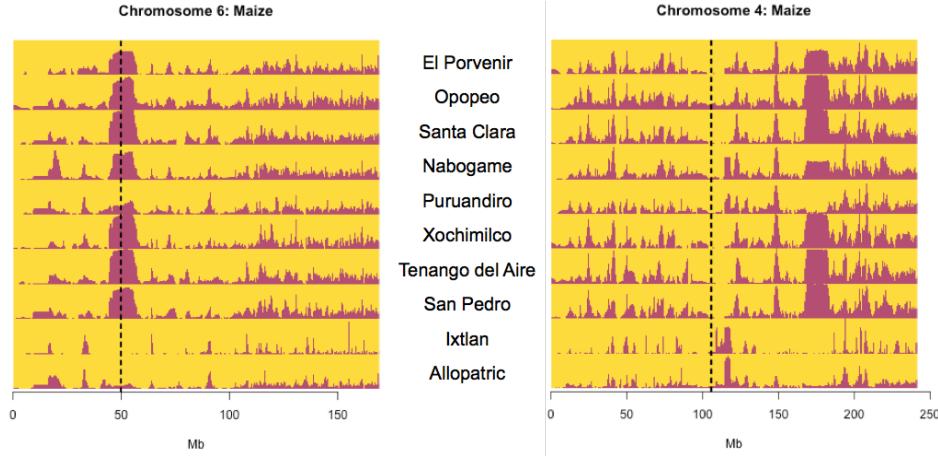


Figure 4: Introgression from *mexicana* to maize along chromosomes 4 and 6 (Hufford et al., 2013). Each row represents a population of maize, and the frequency of the *mexicana* allele at a given position is shown in dark red. Consistent signals of introgression are evident in several chromosomal regions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

grant we will sample 12 maize and 12 teosinte from each of three *mexicana* and *parviglumis* sites previously identified as sympatric with domesticated maize (Hufford, 2010; Hufford et al., 2013). Seed from numerous populations of each taxa are already available in our labs. We will sample populations from 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. We will take advantage of resequencing data from Objective IIA as well as more than 30 maize landrace genomes, including samples from high- and low- elevations in both Mexico and Guatemala, that we have sequenced as part of a USDA-funded project. Only two low-coverage inbred *mexicana* genomes are currently available (Chia et al., 2012), however, so we will resequence two *mexicana* genomes to complement these data.

In addition to resolving the scale of adaptive introgression and local adaptation, results from these analyses will inform questions about convergence in the genetic architecture of selection in hybrid populations (Objective IB).

#### **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 *parviglumis* and *mexicana* (Ross-Ibarra et al., 2009). These teosinte are currently allopatric; while there is little

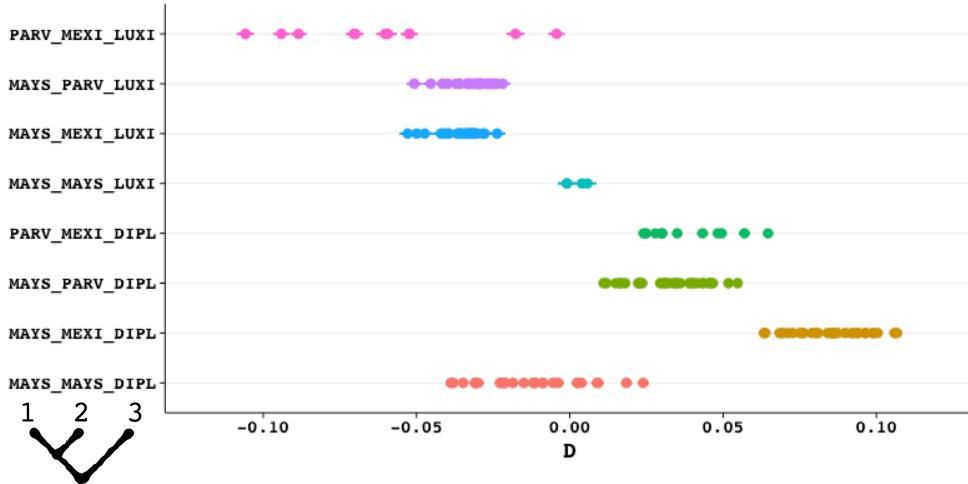


Figure 5: Genome-wide evidence for admixture among *Zea*. In the absence of admixture, taxa 1 and 2 on the tree should share similar numbers of derived alleles with taxa 3 due to the stochastic nature of incomplete lineage sorting. Admixture leads to deviations from this expectation as measured by the  $D$  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.

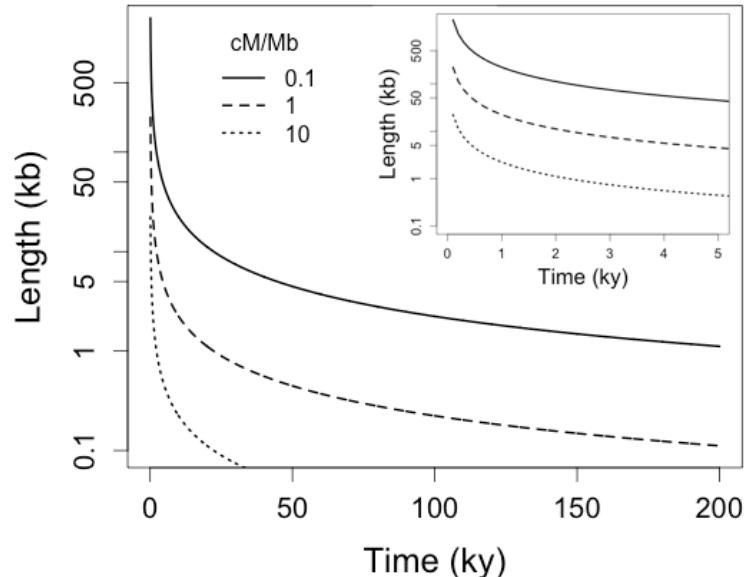
fossil evidence with which to estimate ranges, ecological niche modeling suggests that the ranges of both *parviglumis* and *mexicana* have not changed appreciably for tens of thousands of years (Hufford et al., 2012c). Domesticated maize is currently found in sympatry with each of these taxa and is 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 *parviglumis* 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 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

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.



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. These will be genotyped in year two. In addition to sequence data generated in Objective IIA and Objective IIB, we have in hand a single genome sequence of *diploperennis* from another NSF-funded project. We will additionally sequence a single additional individual of each *diploperennis* and *luxurians* to ensure multiple high-quality haplotypes for each species. 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 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 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.

**Potential Challenges** If we have difficulty collecting sufficient populations of *luxurians* and *huehuetenangensis*, we will instead make use of the more limited germplasm accessions for these two taxa available at the USDA and CIMMYT germplasm repositories.

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.

## 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 post-doctoral 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.

Normal avenues of research dissemination (publications, conferences) will be enhanced by public hosting and distribution of code and teaching resources, as well as public release of presentations and article preprints. Both the Ross-Ibarra and Hufford labs have an excellent track record of open science, including code (via <http://github.com>), presentations (via <http://slideshare.com>), data (via NSF <http://iplantcollaborative.org>, <http://figshare.com>, and <http://datadryad.org>), and posting of publications as freely available preprints before publication (via <http://arxiv.org>, <http://biorxiv.org>, <http://figshare.com>, and <http://peerj.com>). Both PIs actively promote their science via social media as well, including for example a combined twitter following of ~ 2,000.

In addition to these broader impacts, we propose both an undergraduate STEM module for classroom instruction at ISU and an international student exchange program with collaborators in Mexico.

## STEM Training for Undergraduates

Senior Personnel Dr. Calderon, a former Howard Hughes Medical Institute Teaching Fellow at the University of Wisconsin, will develop a teaching module on the topic of hybridization to be implemented in the course “Principles of Biology” that is taught each Fall by Dr. Hufford at Iowa State University. This course is a requirement for many ISU majors including Biology, Chemistry, and Genetics, with enrollment in Dr. Hufford’s section averaging ~300. The inquiry-driven module will incorporate the data we propose to collect here in order to provide students with an opportunity to experience the biology that we practice in our research programs. A written evaluation will be used to measure student performance and to identify strategies for improving the module.

Both Dr. Hufford and Dr. Ross-Ibarra have requested funds to support undergraduate students to work in their research laboratories. Dr. Hufford will recruit undergraduates who show

particular interest in the hybridization teaching module to spend a year working on the topic in his lab and will help them develop an independent project related to this subject matter. Dr. Ross-Ibarra will recruit students from his undergraduate Genetics course. These students will work with the postdoctoral scholar and Dr. Ross-Ibarra to design and carry-out research projects related to the grant; these could include projects such as growth-chamber experiments assessing fitness of populations showing adaptive introgression, greenhouse crossing experiments to evaluate the potential for hybridization, or analysis of resequencing data to look at patterns of copy number variants or transposable element transposition in hybrid populations.

### **US-Mexico Exchange Program**

We will build on the success of an NSF-funded US-Mexico exchange program that the Ross-Ibarra Laboratory has run 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. Here, 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. During years two and three of the project, a graduate or undergraduate student from the Eguiarte group will spend 2-3 months in either the Hufford or Ross-Ibarra Laboratory learning GBS methodology and/or population genomic analysis techniques, 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. Funds have been requested to support travel and subsistence for participating students (two students during both years two and three). Senior Personnel Claudia Calderon will lead the exchange program, as she has considerable past experience organizing international student research trips and working with international students in different countries. She will consult with students and faculty several times before each visit to work through logistical (visa, housing) and cultural concerns.

### **Results From Prior NSF Support**

#### **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 #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 focused in particular on the evolutionary genetics of centromeres, demonstrating the lability of centromere tandem repeats but also showing little evidence in maize for coevolution between centromere sequence and kinetochore proteins.

**Broader impacts** Co-PI Ross-Ibarra established a successful international student exchange program. Former trainees on the grant include Dr. Matthew Hufford (PI on the current grant).

**Publications** Shi et al. (2010); Chia et al. (2012); Fang et al. (2012); Hufford et al. (2012a,b, 2013); Melters et al. (2013); Kanizay et al. (2013); Pyhäjärvi et al. (2013)

## References Cited

- 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. HÃŒEbner, 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. Kaepler, 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.
- 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.
- M. Evans and J. Kermicle. Teosinte crossing barrier1, a locus governing hybridization of teosinte with maize. *Theoretical and Applied Genetics*, 103(2-3):259–265, 2001.
- 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.
- 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.
- 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, P. Bilinski, T. Pyhajarvi, and J. Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends Genet*, 28(12):606–615, 2012a.
- M. Hufford, X. Xu, J. van Heerwaarden, T. Pyhajarvi, J. Chia, R. Cartwright, R. Elshire, J. Glaubitz, K. Guill, S. Kaeplier, 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, 2012b.
- M. Hufford, P. Lubinsky, 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. parviflora*), 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, 2012c.
- L. Kanizay, T. Pyhajarvi, E. Lowry, M. Hufford, D. Peterson, J. Ross-Ibarra, and R. Dawe. Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in zea mays. *Heredity (Edinb)*, 110(6):570–577, 2013.
- 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.
- J. Kermicle, S. Taba, and M. Evans. The gametophyte-1 locus and reproductive isolation among zea mays subspecies. *Maydica*, 51(2):219, 2006.
- J. L. Kermicle and M. M. Evans. The zea mays sexual compatibility gene ga2: naturally occurring alleles, their distribution, and role in reproductive isolation. *Journal of Heredity*, 101(6):737–749, 2010.
- 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.
- 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.
- D. Melters, K. Bradnam, H. Young, N. Telis, M. May, J. Ruby, R. Sebra, P. Peluso, J. Eid, D. Rank, J. Garcia, J. Derisi, T. Smith, C. Tobias, J. Ross-Ibarra, I. Korf, and S. Chan. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*, 14(1):R10, 2013.
- 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. Ryall, 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. Pyhäjärvi, M. Hufford, S. Mezmouk, and J. Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome Biol Evol*, 5(9):1594–1609, 2013.
- 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.
- J. Shi, S. Wolf, J. Burke, G. Presting, J. Ross-Ibarra, and R. Dawe. Widespread gene conversion in centromere cores. *PLoS Biol*, 8(3):e1000327, 2010.
- 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.
- 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.
- A. Widmer, C. Lexer, and S. Cozzolino. Evolution of reproductive isolation in plants. *Heredity*, 102(1):31–38, 2009.
- 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. 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.

### **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 commitment from Ruairidh Sawers of Langebio in Irapuato (see attached commitment letter) to assist with common garden experiments. 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 from Hector Tuy (see attached letter) from IARNA 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 what analyses the data can be used for prior to our initial publication.

**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 researcher 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 of the Ross-Ibarra and Hufford Labs and Senior Personnel Eguiarte and Calderon via web-conference in which one lab member will present 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 postdocs at both UC Davis and Iowa State will be explicitly tasked with helping recruit undergraduate students, mentoring them and supervising their research. Both institutions have included funds to support undergraduate student research. Such efforts in our labs have previously 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. UC Davis and Iowa State 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.