

Project Summary

Overview The proposed project will investigate the genome-wide effects of hybridization and introgression in the genus *Zea*. First, investigators will study incipient speciation between the lowland-adapted *Z. mays* ssp. *parviglumis* (hereafter, *parviglumis*) and the highland-adapted *Z. mays* ssp. *mexicana* (hereafter, *mexicana*). Through field collections, genotyping and common garden studies, the investigators will assess what fraction of the genome is porous to gene flow in hybrid zones, how fitness of these taxa varies across a hybrid zone, and 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 *parviglumis* and *mexicana*; and 2) An ecological timescale in which maize has spread across the Americas and adapted to local conditions. The analysis on an evolutionary timescale will generate basic knowledge on the process of incipient speciation and the porous nature of the genomes of diverging 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 shuffling and 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 represents an unparalleled opportunity for testing these hypotheses and realizing 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 convergent introgression and the extent to which selection determines patterns of gene flow. 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 locally-adapted, parapatric wild teosinte

Zea mays ssp. *parviglumis* (the wild progenitor of maize; hereafter, *parviglumis*) and *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) diverged approximately 60,000 BP (Ross-Ibarra et al., 2009) and have parapatric distributions: while *parviglumis* occurs in the warm lowlands of southwest Mexico,

do objectives go here, or
do we combine rationale with the
intro and put these after?

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 independent hybrid zones show convergent or unique patterns of introgression?*
- B. *Do hybrid populations show higher fitness than parental subspecies in some environments?*
- 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 *parviflumis* ~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. *Does hybridization with locally-adapted relatives facilitate the spread of colonizing species?*
- B. *Is introgression adaptive across multiple geographic scales?*
- C. *Can a widespread species serve as a bridge for gene flow among otherwise allopatric taxa?*

never sure when to use "gene flow", "introgression", or "hybridization". could you make sure we are using appropriately and consistently throughout?

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; 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 (?). 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 maintained by adaptive or purifying selection (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 in independent hybrid populations or are patterns of introgression

no figures in first 4 pages. think this is a problem?

this isn't quite right, but need to explain ecotone as not obvious and this was first use

driven more by local environmental conditions? Has gene flow from native relatives allowed colonizing species to more readily adapt to new environments? 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 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).

The *Zea* study system is uniquely suited to address these questions and advance our understanding of the hybridization biology. The *Zea mays* subspecies *parviglumis* and *mexicana* are distributed across a steep altitudinal 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., 2012c). However, analysis of microsatellite markers genotyped in a range-wide sample has identified elevated admixture between the subspecies in two geographically-distinct, mid-elevation regions of Mexico where the distributions of parental subspecies overlap, suggesting the presence of multiple hybrid zones (Fukunaga et al., 2005). Our recent genome-wide analysis of twelve individuals from a population in one of these hybrid zones 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 or in independent hybrid zones. By expanding our preliminary studies beyond a single population, we can assess how introgression varies across regions and hybrid zones and determine if specific haplotypes are consistently introgressed and widely adaptive or, rather, tied to specific habitats.

In addition to hybridization between *parviglumis* and *mexicana*, gene flow between domesticated maize and various wild *Zea* has been detected based on both hybrid morphologies observed in the field (Wilkes, 1967, 1977) and genetic data (Fukunaga et al., 2005; Ross-Ibarra et al., 2009; ?). Maize domestication from *parviglumis* occurred recently on an evolutionary timescale (\sim 9,000BP; Matsuoka et al., 2002) and was followed by rapid spread of the crop across the Americas over the following millennia (Piperno and Flannery, 2001; Grobman et al., 2012). During this expansion, maize was brought into sympatry with new wild relatives that were likely allopatric to the progenitor of maize (*i.e.*, *parviglumis*) for long periods prior to domestication (Hufford et al., 2012c). Our recent work has provided evidence of introgression from *mexicana* into maize during its colonization of the highlands of the Mexican Central Plateau. We found consistent introgression into several highland maize populations at QTL for phenotypes (*e.g.*, pigment and macrohairs) that distinguish highland *mexicana* from lowland *parviglumis* teosinte, and showed that both *mexicana* phenotypes and higher growth rate were found in maize plants with *mexicana* introgression (Hufford et al., 2013). Our interpretation of these results is that maize received adaptive introgression from *mexicana* that allowed the crop to spread into the highlands of Mexico.

Subsequent to its diffusion into the Mexican highlands, maize spread into sympatry with additional teosinte taxa in Guatemala including *Zea luxurians* (hereafter, *luxurians*) and *Zea mays* ssp. *huehuetenangensis* (hereafter, *huehuetenangensis*), each adapted to environmental conditions very different from those of the progenitor *parviglumis*. Although maize is known to hybridize with both

do we want these to perfectly mirror objectives?
are they redundant if those comes after objectives?

do we need more broad-scale big picture stuff?
this section very quickly jumps into maize details.

should we move some of these sentences on hufford 2013 down below with objective 2?

I'll leave that up to you...feel free to move them and patch up a transition here it you'd like

taxa, the extent and adaptive significance of gene flow between these teosintes and maize is unknown. Additionally, based on analysis of a small number of resequenced loci, *mexicana* haplotypes appear to be segregating in *luxurians* (Ross-Ibarra et al., 2009). Since *mexicana* and *luxurians* are entirely allopatric in their distributions, this suggests maize may have served as a bridge for gene flow between these two taxa. Further work will be necessary to explore this possibility and to assess if the genomes of *Zea* species have been largely altered through gene flow during the spread of maize across the Americas (see Objective II).

Research Plan

Objective I Evolutionary genomics of teosinte hybrid zones

Objective IA Do independent hybrid zones show convergent or unique patterns of introgression?

Our previous publications suggest *Zea* is a promising model system for exploring the evolutionary role of hybridization and introgression (e.g., Ross-Ibarra et al. 2009; ?; Hufford et al. 2013; Pyhajarvi et al. 2013). To further refine our research questions and provide preliminary results for this proposal we have reanalyzed published data (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.

In order to clearly delineate *parviglumis/mexicana* hybrid zones, we calculated the probability of assignment of samples from these taxa to *parviglumis* and *mexicana* groups using STRUCTURE (?). 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 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 *parviglumis* than these subspecies are from each other.

Currently, few studies have dissected the genome-wide architecture of hybridization in replicate hybrid zones, and little is known therefore about the consistency in patterns of gene flow. The Hufford Laboratory and Senior Personnel Luis Eguiarte will assess the genomic architecture and evidence for selection on introgression in two hybrid zones of *mexicana* and *parviglumis* through population collections, generation of genotype and sequence data, and application of population

an idea on order. currently you go 1) popgen 2) what kind of hybrid zone 3) selection on phenotypes. I think it makes more sense to do: 1) What kind of hybrid zone? (solid, i have come around to liking this one) 2) Are adaptive phenotypes under selection? (what are the alternate hypotheses here?). Maybe if you specify under selection in hybrid pops? Then this builds on ecotone/tension. If hybrids have higher fitness in some pops we should see evidence of selection on some of these traits. if hybrids uniformly suck, putatively adaptive traits should show no evidence of selection in those pops (because they suck due to e.g. BDM incompatibilities or something) 3) This builds

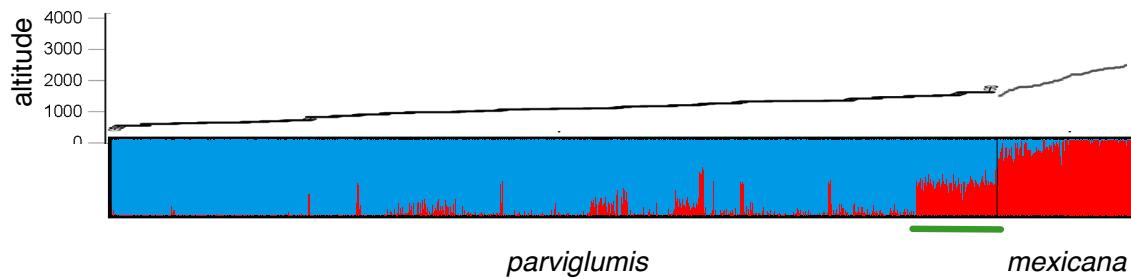


Figure 1: Assignment of *parviglumis* and *mexicana* individuals to $K=2$ groups using the Bayesian assignment algorithm of STRUCTURE (?). Individuals are sorted by increasing altitude as indicated by the plot above the bar chart. Individuals from mid-elevation, hybrid zone populations are underscored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

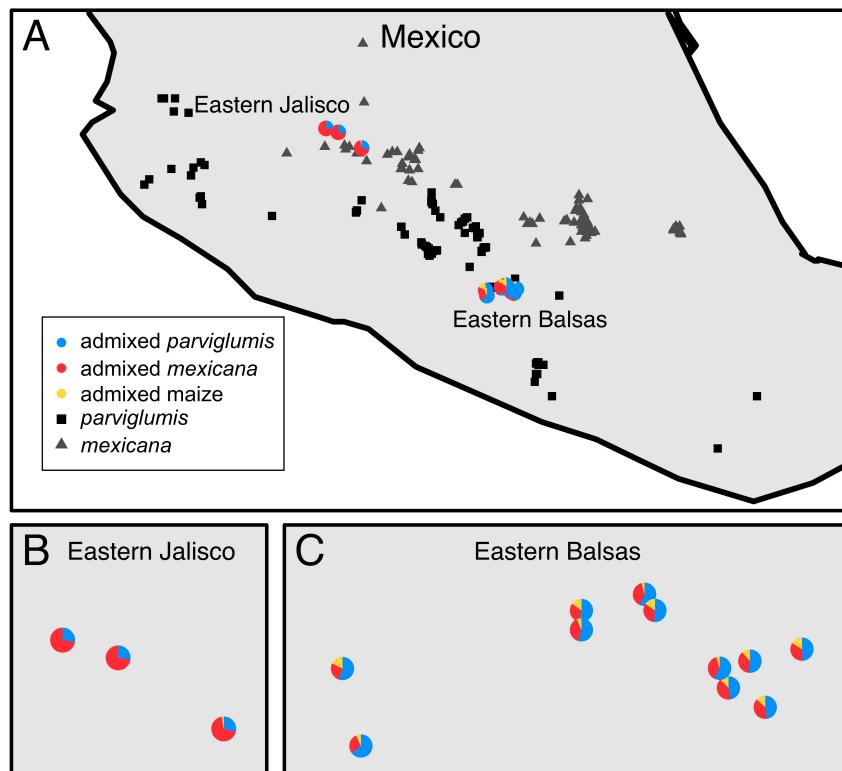


Figure 2: A) Location of two putative hybrid zones of *mexicana* and *parviglumis*. Hybrid populations are represented as pie charts with proportion assigned to *mexicana*, *parviglumis*, and maize groups. Zoomed-in views of the Eastern Jalisco (B) and Eastern Balsas (C) hybrid populations.

genomic analyses appropriate to this question.

Panel Construction and Sample Collection: From previous collections, we have access to extensive sampling of *mexicana* and *parviglumis*. Moreover, Senior Personnel Luis Eguiarte and collaborator Salvador Montes-Hernandez (see attached letter of commitment) have recently collected altitudinal transects of *parviglumis* and *mexicana* that extend through both hybrid zones targeted by our project (Díez et al., 2013) and are familiar with populations in this region. Our current collections will likely be insufficient for both the genotyping and common garden 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. 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)). We will also measure plant density, slope of the terrain, and canopy cover at each sampling site. These data will be useful covariates for estimating any potential maternal effects in our common garden experiments. Four populations will be sampled from both the eastern Jalisco and eastern Balsas hybrid zones. To the extent possible, we will select hybrid populations in a stratified manner across the elevation gradient found in these regions. Four populations each will also be collected from non-admixed *parviglumis* and *mexicana*, with two populations of each taxon collected from regions proximate to both hybrid zones. We have already obtained the requisite collection permits as well as permits for importing samples into the United States for genetic analysis. Following collection, samples will be sent to the lab of Dr. Eguiarte at the Universidad Nacional Autónoma de México (UNAM) for cold storage until common garden experiments are planted (see Objective IB). During year two of our proposed project, leaf tissue will be collected from a single plant per sampling site ($n = 240$) in our mid-elevation common garden (see Objective IB) at the 5-7 leaf stage, stored in silica, and shipped to Iowa State University for DNA isolations and subsequent genotyping and full-genome sequencing.

Sample Genotyping and Sequencing: DNA for genotyping will be isolated using a modified CTAB procedure (Saghai-Marof et al., 1984). 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; see attached letter of commitment from Dr. Sharon Mitchell). 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 ~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 thousands of diverse maize and teosinte, 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., 2012b), 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.

i struggle with "pop genomic".
do we use sometimes, always?
seems like we should be consistent about why/when we use pop genetic vs. pop genomic

FYI silica isn't best for GBS. if possible to ship fresh, better. and remind me to tell you the sampling protocol. if you're doing plates, it will save you billions of years down the road to sample directly into plates correctly in the field

another FYI:
CTAB sucks
for WGS.
fine for GBS
though

bad news:
nobody
in my lab
can get
that damn
pipeline
working. i
haven't tried
myself. leave
it in for now,
just another
FYI

Population Genomic Analyses: We will assess the genome-wide patterns of ancestry 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. Second, we will attempt to 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 allopatric reference populations of *parviflumis* and *mexicana*. Finally, if phasing genotypes (Scheet and Stephens, 2006) proves difficult given high levels of missing data or heterozygous error, we will make use of software 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. We will evaluate excess of *parviflumis* or *mexicana* ancestry on a site-by-site basis across hybrid genomes at the population level and will determine whether patterns are conserved across populations within hybrid zones and between hybrid zones. Chromosomal regions showing an excess of ancestry from one taxon in hybrid populations will be inspected for evidence of selection using a combination of site-frequency-, linkage-disequilibrium-, and population-differentiation-based methods (reviewed in Vitti et al., 2013). Chromosomal regions showing strong evidence of selection across individuals within a hybrid zone based on analysis of GBS data will be inspected in our whole-genome resequencing data to refine haplotype boundaries. Whole genome sequence data will also allow estimation of the age of the introgression and potentially identification of candidate causal polymorphisms.

Objective IB Do hybrid populations show higher fitness than parental subspecies in some environments?

In order to assess fitness and variation at putatively adaptive phenotypes across both non-admixed and hybrid populations we will conduct common garden experiments in Mexico at three altitudes: 1) Below a hybrid zone in habitat occupied by non-admixed *parviflumis*; 2) Within hybrid zone habitat; and 3) Above a hybrid zone in habitat occupied by non-admixed *mexicana*. Common garden experiments will be replicated over years two and three of our proposed project.

Discussions with collaborators in Mexico (Ruairidh Sawers and Salvador Montes-Hernandez; see attached letters of commitment) raised concerns about the safety of students at field sites in the state of Guerrero (the location of the eastern Balsas hybrid zone; see US State Department Travel Warning at <http://travel.state.gov/content/passports/english/alertswarnings/mexico-travel-warning.html>) and the feasibility of managing six concurrent gardens. We thus propose a single transect of three replicate gardens in the eastern Jalisco hybrid zone. In our initial discussions with Drs. Sawers and Montes-Hernandez we have identified potential high- and low-elevation sites near Celaya and Bucerias, Mexico respectively. We will explore options for our third hybrid zone garden during our collections in the first year of the project. The hybrid zone is less than 50 kilometers from the host institution of Dr. Sawers (Langebio in Irapuato, Mexico) and identification of an appropriate site should be straight-forward. Each garden will consist of three complete blocks including a randomization of three plants from each of 15 sampling sites in the 16 populations described in Objective IA (3 blocks x 3 plants x 15 sites x 16 populations = 2,160 plants per garden). We will measure fitness-related phenotypes (percent germination, germination rate, plant height at 15-day intervals, 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 phenotypes for which there is no *a priori* evidence of selection across an elevational gradient (culm diameter and the width of the leaf beneath the first lateral branch at the time of flowering). Analysis of relative fitness of hybrid and parental plants across our garden sites will provide evidence regarding ecotone vs. 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 gardens, whereas tension zone dynamics would be supported by hybrids having lower fitness in all gardens. Phenotypic data for putatively adaptive traits and traits with no evidence of selection will be analyzed in Objective IC.

culm diameter is associated with Inv1n which shows strong altitudinal clines – see Zhou's paper. i've never looked to see if the phenotype shows a cline but based on that i suspect it might!

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

Stem pigmentation, macrohair density, and flowering time in particular are 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 *parviglumis* (Rodriguez et al., 2006), which may represent an adaptation to shorter growing seasons at high elevation. We will combine our genome-wide marker data obtained in Objective IA with phenotypic data collected in our common garden experiments in Objective IB 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 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).

other thoughts: 1) somewhere you need to say these are half-sib families so you can do qst-fst. do you want to explain qst-fst? 2) if these are half sib families and you have a reasonable % selfs, we have code that can impute and phase mom and kids. up to you if worth mentioning; I can provide github link

In addition, we will conduct association analyses to connect genotype to phenotype using GBS data described in Objective IA 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 *parviglumis* or *mexicana* in hybrid populations and zones identified in Objective IA, 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.

culm diameter is associated with Inv1n and which does show an elevation cline. mexicana has thicker culms than parv i believe

Preliminary Results:

Potential Challenges:

Objective II Crop-wild introgression during maize diffusion

Following domestication, maize spread rapidly across the Americas, colonizing novel environments different from that inhabited by its wild ancestor *parviglumis*. During this diffusion, maize came

in contact with each of the other wild teosinte taxa in the genus *Zea*. Hybridization between maize and each of these taxa has been previously documented (see Doebley, 1990, for a review), 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 address three questions that arise from this natural experiment. First, given previous evidence of the potentially adaptive significance of introgression from the teosinte *mexicana*, we ask whether maize colonization of the tropical lowlands of Guatemala was also facilitated by adaptive introgression from the teosintes *luxurians* and *huehuetenangensis*. Second, building on our previous observations in the highlands of the Central Plateau, we seek to define the geographic scale over which introgression may be adaptive. Finally, we return to the observation of haplotype sharing between allopatric teosinte taxa (Ross-Ibarra et al., 2009) and propose to test whether these results are best explained by incomplete lineage sorting or the possibility that domesticated maize may have served as a bridge for indirect gene flow among teosinte populations.

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

Maize was domesticated from *parviflora* in the Balsas River valley on the Pacific Coasts of Mexico. As maize spread from its center of origin, it encountered and adapted to a wide range of different environments.

We have previously documented the importance of introgression in facilitating maize adaptation to the highlands of the Mexican Central Plateau (Hufford et al., 2013; Takuno et al., 2015). As maize spread from its origins in the Pacific Coast of Mexico, it encountered a number of different environments in addition to the highlands. By only a few thousand years after domestication, for example, maize had arrived in the humid lowlands of Guatemala (Neff et al., 2006). Conditions in Guatemala are substantially more tropical than the southwest coast of Mexico inhabited by *parviflora*, with warmer winters, lower annual fluctuation in temperature, and more than double the annual precipitation. Upon arrival in Guatemala, maize came into contact with the wild teosintes *huehuetenangensis* and *luxurians*. These taxa exhibit a number of adaptations to their environment including differences in root architecture, flooding tolerance, and delayed flowering (Wilkes, 1967; Mano et al., 2006). Maize is known to hybridize with both taxa (Wilkes, 1967), raising the possibility that the process of adaptive introgression we observed in the highlands of Central Mexico is mirrored in the humid lowlands of Guatemala.

To address this question we will sample six populations of both *luxurians* and *huehuetenangensis*, stratified across their elevational range in Guatemala. Two of these populations will be chosen to be as distant as possible from domesticated maize, and used as a control representative of ancestral haplotypes or allele frequencies. For the remaining four populations of each taxon, we will sample populations of both the teosinte and a sympatric or nearby maize landrace population. Additionally, two maize populations from similar environment, but allopatric to teosinte, will also be chosen, for a total of 24 maize and teosinte populations. We will be assisted in our collection efforts by Mario Fuentes López of the Fundit Organization in Guatemala (see attached letter of commitment). We will genotype 12 individuals from each population using GBS. Two individuals (either two *luxurians* or one *luxurians* and one maize) will be fully sequenced. There is currently no full genome sequence of *luxurians*, and this will allow us to delineate introgressed or selected regions, identify copy number variants, and potentially identify candidate adaptive polymorphisms

within the regions of interest. Analysis of introgression and selection will follow methods described in Objective IA. If introgression from either of these teosintes has been adaptive to colonization of Guatemala, we predict we will find evidence of introgression in the majority of maize populations, that these regions will show evidence of recent selection in maize, and that they will overlap with QTL for phenotypes likely to be adaptive in these environments (e.g. Omori and Mano, 2007; Mano et al., 2008). We also predict that the same regions should show evidence of selection against introgression from maize into *luxurians*. Results from these analyses will help establish whether observations of adaptive introgression from *mexicana* are an anomaly unique to the highlands of Mexico, or whether adaptive introgression from crop wild relatives may be a common occurrence that has facilitated the spread of domesticated taxa beyond their original habitat.

Finally, this aim will also provide useful baseline information on patterns of genetic diversity in *luxurians* and *huehuetenangensis*, both 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). Almost nothing is known about diversity in these taxa, and questions regarding their evolutionary history, long-term survival, the risk of diversity loss or extinction due to excessive hybridization with maize, and the relationship and connectivity among populations will all be furthered by the results obtained here.

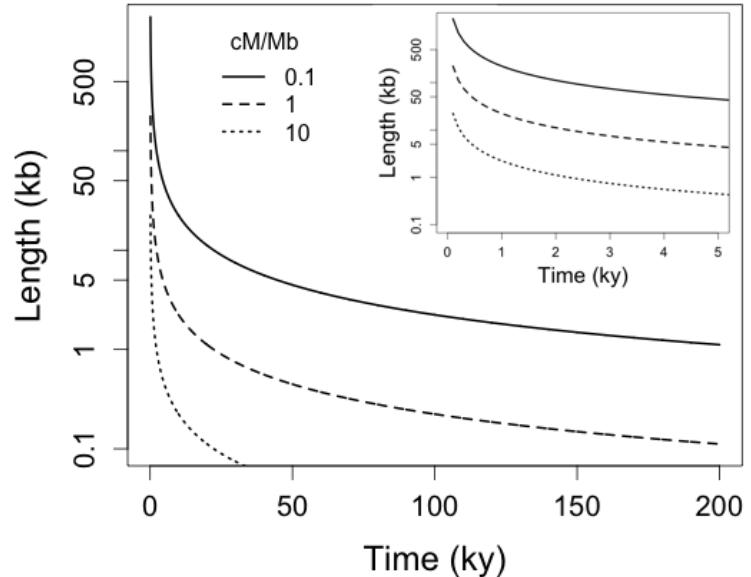
Objective IIB Is introgression adaptive across multiple spatial scales?

Due to their sessile nature, plants must adapt to their local environments. The scale of local adaptation varies widely, from large geographic regions (Lowry and Willis, 2010; Fang et al., 2014) to fine scale adaptation within a population (Hamrick and Holden, 1979). While there are now several examples of introgression facilitating adaptation in plant populations, 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 have adapted on a 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.

We observed widespread introgression of *mexicana* alleles in highland maize across the Central Plateau (Hufford et al., 2013), suggesting that some *mexicana* alleles increased the fitness of maize across a wide geographic area. We also observed a number of introgressed regions at high frequency in only one or a few of the sympatric maize populations, suggestive of introgression. Unfortunately our previous SNP data provides poor resolution to identify selection (?) and suffers from ascertainment bias that limits detection of *mexicana* haplotypes (Hufford et al., 2013).

To understand the scale of adaptive introgression and local adaptation in maize, we propose here to 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 *parviglumis*), and the humid lowlands of Guatemala (with *luxurians* and *huehuetenangensis*). In addition to re-analysis of the populations sampled in Objective IIA, we will sample three pairs of *mexicana* and *parviglumis* previously identified as sympatric with domesticated maize (Hufford, 2010; Hufford et al., 2013). For each of *luxurians*, *mexicana*, *parviglumis*, and *huehuetenangensis* we will endeavor to sample populations from different environments within the range of the taxon. Maize and teosinte individuals from

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



each population will be genotyped using GBS, and we will test for introgression and selection following methods proposed in Objective IA. Based on these results we will select two individuals (of maize or *mexicana*) for whole-genome shotgun sequencing. The improved resolution offered by whole genome sequence will allow us to better characterize introgressed haplotypes and identify potential candidate loci. We will quantify the overlap between regions of the genome showing evidence for selection and those showing evidence of introgression in all or a subset of sympatric populations. While we have prior evidence for adaptive introgression from *mexicana* into maize, we do not know whether introgression from *parviglumis*, *luxurians*, or *huehuetenangensis* may have proven adaptive in any population. If we find evidence of both introgression and selection, one hypothesis is that we may expect to find adaptive introgression limited to only those regions which are broadly adaptive across a wide geographic area. In this model, introgression may have been important for initial colonization of a new geographical area, such as the high altitudes of the Central Plateau, but continual gene flow from sympatric populations is selected against by farmers as it includes numerous deleterious alleles at domestication-related loci (c.f. Hufford et al., 2013). Alternatively, because there is considerable environmental variation even within a geographic area such as the Central Plateau (Hufford et al., 2012c; Pyhajarvi et al., 2013), we may observe evidence for substantial adaptive introgression in localized pairs of populations, with little overlap among populations. Under this model, adaptive introgression not only allows colonization of new regions, but also better adapts maize to local conditions.

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

Our initial survey of divergence and gene flow in *Zea*, based on a set of 26 Sanger-sequenced loci, found evidence for admixture between allopatric populations of *mexicana* and *luxurians* at multiple loci (Ross-Ibarra et al., 2009). As there is no evidence to suggest that these populations overlapped in their recent history, we took these results to suggest that maize, which is known to hybridize

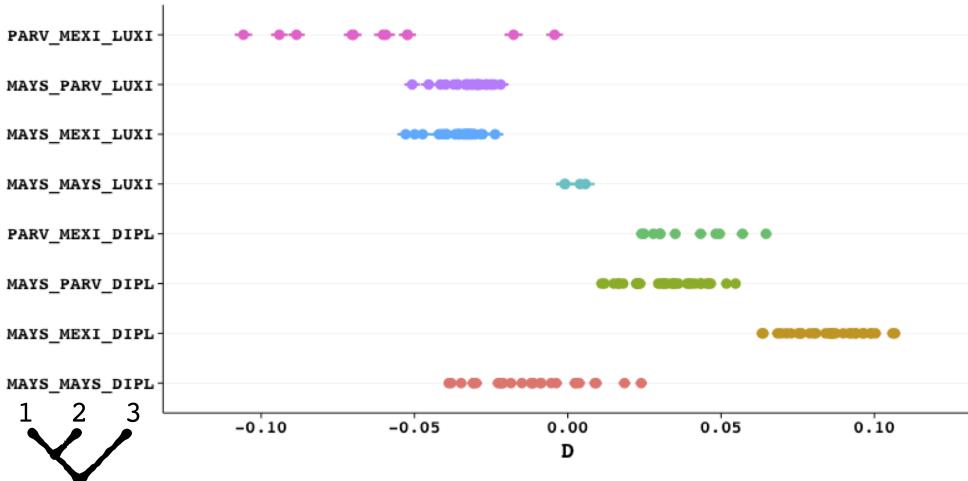


Figure 4: Genome-wide evidence for admixture among *Zea*. In the absence of admixture, taxa 1 and 2 on the tree shown should share similar numbers of derived alleles with taxa 3 due to incomplete lineage sorting. Admixture leads to deviations from this expectation as measured by the D statistic (?). Negative values of D indicate admixture between taxa 1 and 3, while positive values indicate admixture between taxa 2 and 3.

with both taxa, may have served as a bridge for gene flow between them. Further support for this idea comes from patterns of haplotype segregation at an inversion locus on chromosome 4 (*Inv4m*; Fang et al., 2012; Pyhajarvi et al., 2013; Hufford et al., 2013). The inverted haplotype at this locus appears to be derived in *mexicana* (Pyhajarvi et al., 2013), but has introgressed into maize in the highlands of Mexico, an apparent example of adaptive introgression (Panel A, Figure ??; Hufford et al., 2013). We have screened the > 2000 samples in this data set for a SNP diagnostic for the inverted *mexicana* haplotype and have found that the *mexicana* haplotype is segregating in maize in the highlands of both Mexico and Guatemala (Panel B, Figure ??). The haplotype is also present in all samples of *luxurians* genotyped in Fang et al. (2012). These preliminary results suggest that the inversion has moved from *mexicana* into *luxurians*, perhaps via a maize intermediate.

Our preliminary analyses identified shared haplotypes in allopatric teosinte taxa, suggesting that maize may have served as a bridge for gene flow between *mexicana* and *luxurians* and potentially other *Zea* taxa (see above). However, an alternative explanation is that these loci were polymorphic in a common ancestor and continue to segregate due to incomplete lineage sorting. Simple estimates of the length of shared haplotypes expected to be unbroken by recombination suggest that, over the ~ 140,000-generation divergence time between *mexicana* and *luxurians* (Ross-Ibarra et al., 2009), we might well expect to see shared haplotypes of even several kb in length in low recombination regions of the genome (Figure 3). The high-density, genome-wide data generated here will provide an opportunity to test whether observed patterns of haplotype sharing between previously allopatric *Zea* are due to recent introgression from maize. If shared haplotypes have come from introgression from maize over the last few thousand years, the genome-wide distribution of shared haplotype lengths should reveal longer shared segments (Figure 3) than if haplotype sharing is due to incomplete lineage sorting alone.

No high-resolution genetic map currently exists for any teosinte, but the Ross-Ibarra group is currently working on producing such a map for *parviglumis* as part of a different NSF project (# 1238014), and evidence from comparisons among maize populations finds remarkable stability of the genetic map at a relatively coarse scale (Bauer et al., 2013) suggesting differences in the genetic map are unlikely to dramatically affect our estimates. Will will use our teosinte genetic map or the published NAM map (McMullen et al., 2009) to generate an expected distribution of shared haplotype lengths along the genome based on expected divergence times between taxa.

Although the limited sampling in Ross-Ibarra et al. (2009) only identified shared haplotypes between *mexicana* and *luxurians*, maize has been found to hybridize with all species in *Zea* (Wilkes, 1977). We will thus include in our analysis here the perennial taxa *Zea diploperennis* (hereafter, *diploperennis*) and *Zea perennis* (hereafter, *perennis*). We will sample 12 individuals from each of two populations of *diploperennis* and *perennis* and genotype these using GBS. These populations, combined with samples from other teosinte populations included in Objective IIA and Objective IIB, will provide us with a representative sample of wild teosinte populations from across the Americas. Shared haplotypes will be identified (see methods in Objective IA) and compared to the expected distribution of haplotype lengths to look for evidence of recent introgression consistent with the hypothesis that maize has served as a bridge for gene flow among teosinte taxa. Although we cannot exhaustively test for the presence of such haplotypes in all of maize, we can survey published GBS data for more than > 16,000 maize samples (www.panzea.org) to assess the frequency of such haplotypes in domesticated maize.

Objective IID 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.

need more here

Objective IIE Potential Challenges

need more here

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.

add something about undergraduate researchers in my lab?

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 forthcoming papers from their internship. We will build upon the success of this program. A student from the Eguiarte group will spend 2-3 months in either the Hufford or Ross-Ibarra Laboratory learning the GBS methodology and/or honing his/her skills in population genomic analysis, whereas a student from the Hufford and/or Ross-Ibarra Laboratories will travel to Mexico to participate in sample collection trips and to obtain expertise in common garden field experiments. This exchange will build capacity in all groups involved and will provide a valuable international research experience for a graduate student supported by the grant.

Senior Personnel Claudia 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.

Results From Prior NSF Support

Hufford, Ross-Ibarra, Coop, Flint-Garcia, Sawers: #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-PIs 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 amongst principal investigators and laid the foundation for the work proposed in the current 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 phenotyping workshop included graduate students and post-doctoral scholars from the United States and Mexico, providing STEM training and an international scientific experience.

Publications Funding is for organizational purposes and generation of preliminary data; no publications have been produced under this award.

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

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

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

Broader impacts Co-PI Ross-Ibarra has established an international student exchange program as part of this grant. Data and results of this project have been disseminated via publications and presentations as well as deposited in the maize genetics community database www.maizegdb.org. Former trainees on the grant include Dr. Matthew Hufford (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); Pyhajarvi et al. (2013)

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.
- J. G. Barb, J. E. Bowers, S. Renaut, J. I. Rey, S. J. Knapp, L. H. Rieseberg, and J. M. Burke. Chromosomal evolution and patterns of introgression in helianthus. *Genetics*, 2014.
- E. Bauer, M. Falque, H. Walter, C. Bauland, C. Camisan, L. Campo, N. Meyer, N. Ranc, R. Rinc?nt, W. Schipprack, et al. Intraspecific variation of recombination rate in maize. *Genome biology*, 14(9):R103, 2013.
- 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. Tenaillon, 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.
- 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.
- 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.

- 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.
- 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.
- R. F. Guerrero and M. Kirkpatrick. Local adaptation and the evolution of chromosome fusions. *Evolution*, 2014.
- R. F. Guerrero, F. Rousset, and M. Kirkpatrick. Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587):430–438, 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.
- 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.
- 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. 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, 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. 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, 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.
- 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.
- 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.
- 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.
- M. McMullen, S. Kresovich, H. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S. Mitchell, B. Peterson, G. Pressoir, S. Romero, M. Oropeza Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J. Glaubitz, M. Goodman, D. Ware, J. Holland, and E. Buckler. Genetic properties of the maize nested association mapping population. *Science*, 325(5941):737–740, 2009.
- 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.
- H. Neff, D. M. Pearsall, J. G. Jones, B. Arroyo, S. K. Collins, and D. E. Freidel. Early maya adaptive patterns: mid-late holocene paleoenvironmental evidence from pacific guatemala. *Latin American Antiquity*, pages 287–315, 2006.
- 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.
- 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.
- T. L. Parchman, Z. Gompert, M. J. Braun, R. T. Brumfield, D. B. McDonald, J. A. C. Uy, G. Zhang, E. D. Jarvis, B. A. Schlinger, and C. A. Buerkle. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, 22(12):3304–3317, 2013.
- 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.
- 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.
- 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.
- M. A. Saghai-Maroof, K. M. Soliman, R. A. Jorgensen, and R. W. Allard. Ribosomal dna spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences*, 81(24):8014–8018, 1984.
- P. Scheet and M. Stephens. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*, 78(4):629–644, 2006.
- 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.
- 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 l.?????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>.
- Y. Vigouroux, J. Glaubitz, Y. Matsuoka, M. Goodman, G. J. Sanchez, and J. Doebley. Population structure and genetic diversity of new world maize races assessed by dna microsatellites. *Am J Bot*, 95(10):1240–1253, 2008.

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.

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), a free public website allowing dissemination and archiving of large datasets. Data will be released in accordance with the Toronto agreement (2009. Nature 461:168-170. www.nature.com/nature/journal/v461/n7261/full/461168a.html) under the stipulation that no whole-genome analyses be performed until we have published our initial analyses.

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.