

Project Summary

Overview

The genetic basis of plant adaptation to their local environments remains poorly characterized, despite its relevance to climate change and crop improvement. In this project, the Co-PIs will investigate the genome-wide underpinnings of local adaptation in wild and domesticated populations of maize (*Zea mays*) to high elevation environments. Project collaborators will first identify quantitative trait loci for highland adaptation traits using mapping populations developed from Mexican and South American maize as well as a naturally admixed population of highland and lowland teosinte (*i.e.*, wild maize) and two populations of doubled-haploid introgression lines donated by industry collaborators. These populations will allow for comparison of the genetic architecture and effect sizes of highland traits in distinct geographical regions, across elevations, and in both teosinte and maize. Second, researchers will investigate population genetic evidence of selection through studies of adaptive introgression in maize and teosinte, and adaptive divergence in gene expression between lowland- and highland-adapted maize. Finally, the functional consequences of a putatively adaptive inversion polymorphism identified in highland landraces will be characterized through phenotypic and transcriptomic evaluation of introgression lines.

Project Team: PI Jeffrey Ross-Ibarra (UC Davis); CoPIs Graham Coop (UC Davis), Sherry Flint-Garcia (USDA-ARS), Matthew Hufford (Iowa State), Daniel Runcie (UC Davis), and Ruairidh Sawers (LANGE BIO); and Sr. Personnel Kate Crosby (UC Davis)

Intellectual Merit

Selection shapes the genomes of plants by fine-tuning them to their local biotic and abiotic conditions. Surprisingly little is known about the consistency of this adaptive process across similar environments and the extent to which genomes are altered. Only a handful of investigations characterizing the genetic architecture and effect sizes of locally adaptive loci have been published to date and no such studies have been conducted in an economically important plant. Given the repercussions of local adaptation for conservation and agriculture in the face of climate change and human population pressure, the activities proposed here are both important and potentially transformative. Basic evolutionary insight regarding local adaptation will be provided in the field of population and quantitative genomics and substantial resources will be provided to inform genomic approaches to crop improvement for highland environments.

Broader Impacts

As large datasets becomes increasingly common and important to society -- whether crop phenotyping data from industry field trials or genome-wide-association and ancestry data from human populations -- the ability to analyze and interpret such data becomes ever more valuable. The PIs propose three important steps toward this goal. First, they offer a public workshop to train researchers to collect and track phenotypic data from large field trials. Second, they will continue to develop educational software with the goal of providing students the tools to extract information from genomic data. Third, the PIs will capitalize on their previous experience to organize an international student exchange among members of the team, giving US students experience with large field experiments and Mexican students experience analyzing genomic data.

To enable the translation of their scientific findings to breeders and farmers alike, the PIs will collaborate with the International Maize and Wheat Improvement Center to host a farmer field day in which local farmers and breeders can explore the diversity of maize germplasm and learn about genetic approaches to understanding adaptation.

Finally, the PIs will continue to disseminate knowledge and resources generated in this project via open-source software and publications, deposition of novel germplasm in public repositories, and discussion of results via national and international conferences, as well as informal outreach via social media.

Project Description

Introduction

Due to their sessile nature, plants must adapt to their local environments. Understanding the genetic basis of how plants adapt to local conditions -- the number and effects of adaptive loci, the selected traits and their functional relationships, and the similarity of adaptations among populations and species -- will facilitate improved breeding and conservation strategies. This is particularly pressing given current issues of climate change, habitat loss, and human population growth (Savolainen et al., 2013), which will require adaptation of both crops and wild plants to changing conditions and cultivation of crops in new locales.

Agricultural species represent promising systems for research on local adaptation. While most crops were domesticated in narrow geographic centers, many have spread globally, adapting to a wide range of novel environments (Gepts, 2014). In many instances, traits important for crop adaptation (e.g., flowering time and cold tolerance) have already been identified (Gepts, 2014; Purugganan and Fuller, 2009). Insights gained regarding loci underlying local adaptation can feed back into modern crop improvement, yielding valuable benefits in the face of climate change.

We propose to use the adaptation of maize and its wild relatives (*Zea mays*) to high elevation environments as a model for understanding the genetic basis of local adaptation in plants. Maize (*Zea mays* ssp. *mays*) was domesticated in the lowlands of southwest Mexico from the narrowly distributed teosinte *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*; Matsuoka et al., 2002). Since domestication, maize has spread worldwide, and now exhibits the greatest global geographic breadth of 16 staple crops (Hake and Ross-Ibarra, 2015): maize is cultivated on six continents, ranging from southern Chile to Canada and from sea level to well over 3000m in elevation (Tenaillon and Chancosset, 2011). During its global spread, maize has independently adapted to high elevation environments in multiple geographic regions including Mexico and South America (van Heerwaarden et al., 2011). A related wild relative, the teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*), is endemic to the highlands of central Mexico, having adapted to these environments thousands of years prior to maize domestication (Ross-Ibarra et al., 2009; Hufford et al., 2012). Gene flow from *mexicana* likely played an important role in the highland adaptation of maize in Mexico (Hufford et al., 2013), but *mexicana* is not found in South America and maize adaptation to high elevation in the Andes thus followed an independent evolutionary trajectory (Takuno et al., 2015). Maize and teosinte thus form an ideal system in which multiple replicated evolutionary experiments will allow for dissection of the genetics of highland adaptation and an improved understanding of local adaptation.

This proposal builds considerably upon a previous submission to the 2014 NSF-PGRP competition that was ranked ``Highly Meritorious'' and very favorably reviewed. Since our last submission, we have received a one-year NSF ``Catalyzing New International Collaborations'' grant that has facilitated generation of substantially more preliminary data and further cemented partnerships across our research groups. We have also directly responded to reviewers' concerns (Supplementary Documentation: Response to Prior Reviews).

Aims

We will investigate the genetic basis of highland adaptation in maize and teosinte by achieving three aims. The timeline and contributions of each team member to these aims is described in the management plan (Supplementary Documentation A-2).

1. Compare genetic architectures of convergent highland phenotypes
2. Investigate genomic and functional signatures of highland adaptation
3. Characterize the specific functional role of a putatively adaptive locus

Relevance and Justification

Genome-wide studies across populations of model species are just beginning to unravel the genetic architecture and environmental drivers of local adaptation. For example, Fournier-Level et al. (2011) demonstrated that alleles associated with high fitness in *Arabidopsis thaliana* have a tendency to be both local and linked to climate. Likewise, a recent study of *Medicago truncatula* identified candidate loci for local adaptation and found them to be predictive of growth rate under temperature and soil moisture treatments (Yoder et al., 2014). Our own genome-wide study of teosinte (the wild relatives of maize) revealed an important role for inversion polymorphisms and -- in contrast to results from *Arabidopsis* (Hancock et al., 2011) -- an enrichment of regulatory variants among loci showing evidence of selection (Pyhäjärvi et al., 2013), suggesting an important role for gene expression divergence in local adaptation (Zhao et al., 2015; Fraser, 2013). An important consideration is that, while similar phenotypes may be selected in different regions, the genetic basis of these phenotypes and the specific loci involved may differ among populations or species. In maize, for example, although genome-wide association in the nested association mapping (NAM) panel suggests that flowering time is largely controlled by many loci of small effect (Buckler et al., 2009), adaptive change in flowering time across latitudes has involved loci of large effect on photoperiod (Hung et al., 2012). Therefore, investigating the genetic architecture of convergent traits may uncover novel sources of genetic diversity in key functional traits of use for crop improvement or conservation. Key questions regarding repeated evolution of convergent locally adapted traits will include the level of convergence (same nucleotide, gene, pathway, or tissue), and the source of locally adapted alleles (standing variation, mutation, or introgression).

Maize and teosinte are an excellent system in which to study local adaptation. Following domestication in the lowlands of southwest Mexico, maize spread to the highlands of the Mexican Central Plateau, migrating across more than 2000m of increasing elevation. Colonization of the highlands required adaptation to a number of novel abiotic conditions, including differences as extreme as 25 °C annual mean temperature and 3,000mm annual precipitation (Figure 1A-C). Highland landraces have distinct morphologies (e.g., highly pigmented and hairy leaves and stems shown in Figure 1D) that are believed to confer adaptation to cooler regions (Doebley, 1984) and mimic those of the highland teosinte *mexicana*. Our previous genetic analyses (van Heerwaarden et al., 2011) show that maize has independently adapted to highland environments multiple times, including the southwest US, the Guatemalan highlands, and the Andes of South America. These independent instances of highland adaptation in maize and teosinte provide replicated evolutionary experiments and the power to identify and validate both widespread and population-specific candidate loci

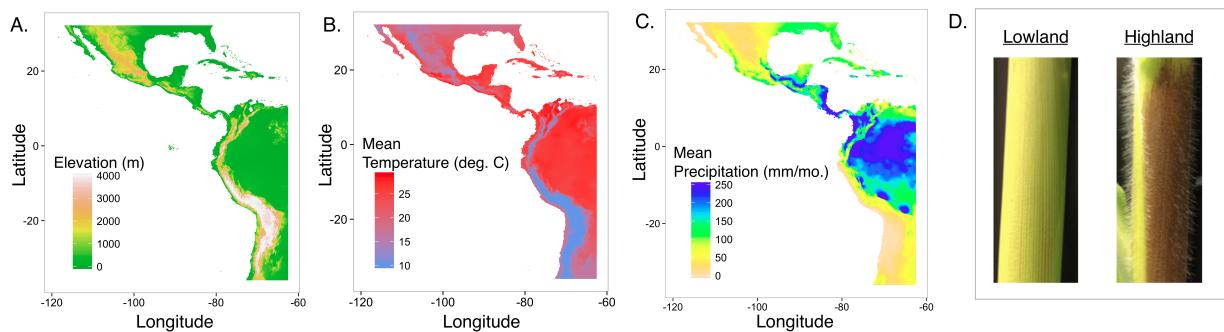


Figure 1: Climate varies considerably across our focal maize habitats of Mexico and western South America. Variables of interest include elevation (A), temperature (B), and precipitation (C). Maize from highland and lowland areas of this region differ considerably in multiple phenotypes such as stem morphology (D).

for highland adaptation.

In addition to providing insight into the genetic mechanisms of local adaptation and recent evolution in maize, the proposed study will provide essential information to help increase or sustain yield in the face of human population growth and climate change. Historical analyses suggest that climate change over the last 30 years has already dramatically impacted maize yields worldwide, slowing gains from breeding and management (Lobell et al., 2011). Recent work has documented farmer-assisted migration of maize to substantially higher elevations in the Andes in response to ongoing climate change (Skarbo and VanderMolen, 2015), and that Mexican farmers in the highlands may be those most vulnerable to changing climates (Bellon et al., 2011). An understanding of how maize has adapted to challenging environmental conditions in the past will help breeders mitigate yield loss due to future changes. **Our proposal directly addresses the PGRP stated goal of ``Development of a genome to systems-level understanding of plant-environmental interactions, especially with respect to adaptation to climate change and response to abiotic and biotic stresses."**

Research Plan

Aim 1 The genetic basis of convergent highland phenotypes

One of the primary goals of this proposal is to determine the genetic architecture of adaptation across multiple, independent colonizations of highland environments in maize and teosinte. Ultimately, these quantitative trait loci (QTL) will be useful for identifying and characterizing the pathways (Aim 2) and specific genes (Aim 2 and Aim 3) involved in adaptation that can then be targeted for maize improvement. In Aim 1, we wish to determine how many genomic regions control adaptive phenotypes, their genomic locations and the distribution of allelic effects. We first perform comparative QTL analysis using populations derived from two highland x lowland maize crosses that will characterize highland adaptation in both Mexico and South America (Aim 1.1). We then take advantage of the historical recombination and greater mapping resolution that can be found in a naturally admixed population of *mexicana* and *parviglumis* to map highland adaptation loci from *mexicana* (Aim 1.2). Finally, we evaluate both *parviglumis* and *mexicana* alleles in a common elite maize background to evaluate their behavior in maize and determine the potential use of *mexicana* alleles in highland maize breeding (Aim 1.3).

Questions

- What is the genetic architecture of highland adaptation?
- How different is the genetic basis of highland adaptation in Mexico and South America?
- Are similar genomic regions responsible for highland adaptation in teosinte?
- How do teosinte alleles affect phenotype in a maize genetic background?

Aim 1.1 QTL mapping of highland adaptation

Our first objective is to identify genomic regions controlling highland adaptation in maize. We will conduct QTL mapping studies of one Mexican and one South American population, each derived by crossing a landrace adapted to lowland conditions with a landrace adapted to highland conditions (Table 1). We make use of landrace inbred lines created by John Doebley (U. Wisconsin) when possible, thus simplifying downstream applications and allowing replication of alleles in current (Aim 2.2) and future functional studies.

In the first year of the project, we will work with Dovetail Genomics (see attached letter of collaboration) to generate *de novo* genome assemblies of the four parents of the mapping populations. These assemblies will provide a much more comprehensive understanding of the genomic basis of highland adaptation (e.g., the role of novel structural rearrangements and transposable element insertions) than would a resequencing approach based on the B73 maize reference. We will self-pollinate F2 plants to create 500 F2:3 families, which allow for replicated measurements in multiple locations. F2:3 plants will be genotyped through Genotyping-By-Sequencing (GBS; Elshire et al., 2011) and run through the standard maize GBS

Table 1: Parental lines for QTL Populations

Population	Parent	Origin (masl)	Inbred	Status
Mexico	Zapalote Chico	Oaxaca (46)	yes	F2:3
	Palomero de Jalisco	Jalisco (2520)	yes	
S. America	Pororo	Bolivia (330)	yes	F2
	Maranon	Peru (2820)	no	

Table 2: Common garden locations

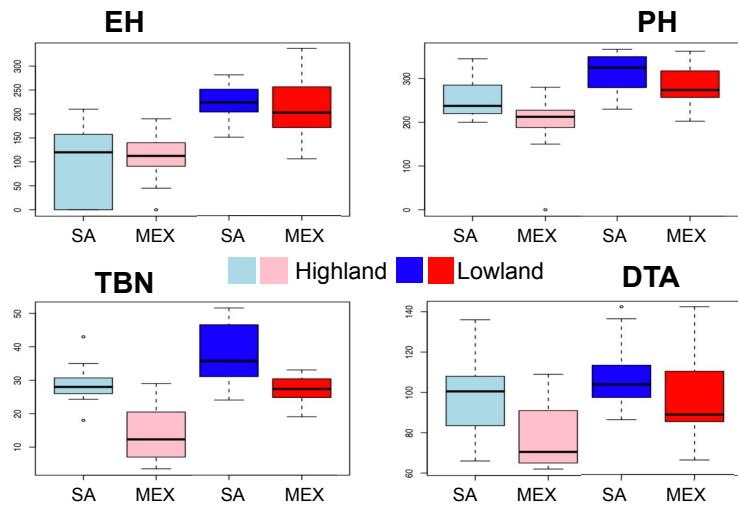
Field Sites	Lat/Lon	Elev (m)	Min/Mean/Max °C	Precip (mm)
V. de Banderas, Nayarit	20.8, -105.2	54	15.3/25.8/33.7	1184
Irapuato, Guanajuato	20.7, -101.3	1729	7.3/20.2/31.7	693
Metepec, Mexico	19.2, -99.5	2582	1.9/13.8/24.2	864

pipeline (Glaubitz et al., 2014) using the parental *de novo* assemblies as references. Based on the current performance of this pipeline, we expect approximately ~1M SNPs, which will allow straightforward imputation of full-genome sequence for all plants. The genetic map will be created using standard methods with a subset of markers (Lander et al., 1987).

Populations will be phenotyped at three field locations in Mexico: lowland, intermediate, and highland (Table 2). Best local practices will be used including irrigation, fertilizer, and pest/weed control across sites. At each location, the experiment will consist of one replicate of each F2:3 population, in which plots of the 500 entries per population will be arranged in an augmented alpha lattice design, with parental checks added to control for field variation. The entire experiment will be repeated a second year. We will collect agronomic, fitness, and elevation-related phenotypes on five plants per plot (Figure 2) using our in-house, barcode-based data collection program following standard protocols for most traits. For macrohair and anthocyanin traits, a fixed-size section of the sheath will be scored using multiple methods (e.g. visual 1-4 scale and image processing for extent, intensity, and spatial patterns) as described in Lauter et al. (2004) during the first season in order to determine the best method for phenotyping in future seasons. Germination success rates under controlled conditions (planting depths of 5 and 20 cm, and temperatures of 7C and 15C) will be evaluated in growth chambers in Ames, Iowa, and root chilling will be evaluated using a custom hydroponic system at the University of California, Davis (see letter of support from Dr. Arnold Bloom) following the protocol of Goodstal et al. (2005).

Raw data from each plot will be analyzed using mixed-models incorporating years, replications, and locations, as well as other design parameters. Data will be analyzed across locations to determine genotype by environment interaction as well as plastic environmental effects on phenotype. Each location will then be analyzed separately to derive least squares means to be used as phenotypic data in QTL analyses. QTL analysis will be conducted using standard software (e.g., SAS; R/qtl, Broman et al., 2003).

Several iterations of QTL analysis will be conducted: on individual traits, individual traits adjusted for covariates such as flowering time, and multiple traits simultaneously. We will test for epistatic (non-additive) interactions among significant QTL (Holland, 1998). QTL profiles will be compared across populations (Mexico vs South America) and among field sites (varying elevation) to determine their effects on adaptive traits. Comparison of the genetic architecture among traits will clarify the lability of these traits and their amenability to selection via breeding.



Trait	Phenotype
MH	leaf sheath macrohairs
ANTH	leaf sheath anthocyanin
DTS	days to silking
DTA	days to anthesis
PH	plant height
EH	ear height
BM	total plant biomass
TIL	tiller number
TBN	tassel branch number
TL	tassel length
EN	ear number
TKM	total kernel mass
50KM	fifty kernel mass
RC	root chilling response
GD	germination depth
GT	germination temperature

Figure 2: Phenotypic differences between a sampling of highland and lowland landraces from Mexico and South America, grown in common garden in Columbia, Missouri (left). List of the phenotypes to be measured in the field (right).

Expected outcomes: 1) A map of QTL underlying phenotypic differences between highland and lowland maize in Mexico and South America and estimates of QTL effect size, and 2) Estimates of fitness differences (PH, BM, TKM, and 50KM (Figure 2)) of highland and lowland plants as a function of their QTL genotype in both environments.

Aim 1.2 Admixture mapping in a teosinte hybrid zone

While *mexicana* and *parviflora* are largely allopatric, they overlap in two areas of Mexico (eastern Jalisco and the eastern Balsas River Basin (Hufford et al., 2012)). A number of hybrid populations of these taxa have been documented in these regions (Fukunaga et al., 2005). We have previously reported near equal proportions of ancestry from the two subspecies in a hybrid population from the eastern Balsas (Pyhäjärvi et al., 2013) and our growth chamber experiments have confirmed that some hybrid plants exhibit highland phenotypes (e.g., pigmented and hairy stems). Higher fitness was also observed in hybrids under cold conditions when compared with non-admixed *parviflora*. In addition, the relatively short lengths of unbroken *mexicana* and *parviflora* haplotypes we have detected in hybrid populations (Pyhäjärvi et al., 2013) suggest there has been extensive recombination since initial admixture, providing an ideal resource for high-resolution admixture mapping of *mexicana* highland adaptation traits. Due to this historical recombination, we anticipate having much higher resolution for mapping highland adaptation traits in naturally admixed teosinte than can be achieved in our synthetic crosses of highland and lowland maize (Aim 1.1). Admixed teosinte will also allow us to assess the genetic architecture of highland adaptation in a third, independent instance (i.e., adaptation that occurred in the wild plant *mexicana* thousands of years prior to domestication).

We have received funding from NSF-CNIC for a field collection during November 2015 of a hybrid population near the town of Santa Rita in the eastern Jalisco *parviflora*-*mexicana* hybrid zone. Seed will be collected from 500 individuals drawn randomly from the population. Seed samples will then be transported to Langebio in Irapuato, Mexico for cold storage. In years 2 and 3, a single seed per individual (500 total) will be germinated and transplanted to the Irapuato field site (Table 2). We will implement agronomic practices in this trial that mirror the more "wild" setting of teosinte (e.g., low planting density, no additional

fertilizer, minimal irrigation) in order to allow for typical trait expression. Phenotypes detailed in Figure 2 will be collected for admixture mapping. Many of these traits are known to differ considerably between *parviflumis* and *mexicana* (Wilkes, 1967). Leaf samples will be collected from plants in the field, and extracted DNA will be genotyped using GBS. While several computational methods for admixture mapping have been developed (Winkler et al., 2010), they are not ideal for use in populations with varying relatedness across individuals or when natural selection has systematically distorted admixture at some loci. In naturally admixed populations these issues can be expected to occur, and will potentially result in false positives due to the non-independence of individuals (a fact accounted for in genome-wide association studies but not in admixture mapping). We will implement novel methods currently under development by Co-PI Coop in our analysis of the Santa Rita population that incorporate non-independence into admixture association tests while accounting for uncertainty in admixture calls along the genome.

Expected outcomes: 1) A map of the location and effect size of QTL underlying phenotypic differences between highland and lowland teosinte, and 2) Empirical testing of novel methods for admixture mapping.

Aim 1.3 Teosinte alleles in a maize background

Aim 1.1 will identify genomic regions associated with highland maize phenotypes. Aim 1.2 extends this approach with higher resolution admixture mapping in a teosinte hybrid zone. To understand the phenotypic consequence of teosinte alleles in a maize background and assess their utility for maize breeding, we will bridge these two approaches using two doubled haploid (DH; completely homozygous line) populations containing 12.5% teosinte developed by DuPont Pioneer (see attached letter of support). For both populations, a donor teosinte parent (*parviflumis* or *mexicana*) has been crossed and back-crossed twice to the same elite DuPont Pioneer inbred prior to DH production. Each population consists of 200 lines and includes introgressions that together span the entire maize genome. Both populations have already been genotyped for more than 50,000 SNPs. Using an experimental design similar to Aim 1.1, we will evaluate a single replicate of both populations at all three sites (Table 2) in each of years 2 and 3 of the grant, phenotyping these for traits listed in Figure 2. QTL and epistasis analyses will be done in parallel to Aim 1.1 using the same procedures and software. Although mapping here will be relatively low resolution given population size and the size of introgressed regions, it will allow explicit evaluation of teosinte alleles in a maize background, and in conjunction with Aim 1.1 will allow comparison of *mexicana* and highland maize alleles at each QTL.

Expected outcomes: 1) Comparison of teosinte allele effects in a maize background, and 2) Evaluation of the utility of teosinte alleles for maize improvement and highland adaptation.

Aim 1 Preliminary Results:

We have made important progress to set the stage for activities in Aim 1. The South American and Mexican maize populations to be utilized in Aim 1.1 are at the F2 and F2:3 generation respectively (Table 1). We have established all necessary field sites at three elevations (Table 2) and have conducted preliminary trials to ensure each site is suitable for project goals. In January of 2015, we held a workshop at our lowland field site that was attended by seven project members, and successfully transferred high-throughput phenotyping methods developed by Co-PI Flint-Garcia across research groups. We have analyzed published genotype data (Fang et al., 2012) to confirm that the Santa Rita teosinte population in Aim 1.2 is admixed and verified in a small growth chamber experiment that plants from this population are variable for highland traits.

Aim 1 Potential Challenges:

Dovetail Genomics has been quite successful in *de novo* assembly of several animal genomes but are only beginning to apply their method to plants. We are currently working with them on an unrelated teosinte assembly; if this does not meet our quality needs we will search for another provider (e.g., NRGENE). However, we note that even a partial assembly will be very useful in the low-copy, genic fraction of the genome, for alignment to divergent haplotypes, for genotyping, and for identification of large-scale structural rearrangements. For admixture mapping in Aim 1.2, we have targeted the Santa Rita population due to its higher

proportion of *mexicana* ancestry and polymorphism for highland traits. Our current marker density for individuals in this population does not allow for accurate estimation of *mexicana* and *parviglumis* haplotype lengths. Prior to large-scale phenotyping of individuals from this population, we will generate high-density marker data using GBS for 12 Santa Rita individuals and reference, non-admixed *mexicana* and *parviglumis* individuals to accurately infer haplotype lengths. Should admixture be quite recent and haplotypes longer than needed for high-resolution mapping, we will instead use samples from the Ahuacatitlan population which has a lower proportion of *mexicana* ancestry, but has already been demonstrated to have small haplotype blocks (Pyhäjärvi et al., 2013).

Aim 2 Population Genetics of Highland Adaptation

In Aim 1 we employ a top-down QTL approach to map loci corresponding to traits differing between highland and lowland maize and teosinte. In Aim 2, we will use a complementary, bottom-up population genetic approach (Ross-Ibarra et al., 2007) to characterize genomic signatures of adaptation associated with elevation and to identify which traits and genomic regions are adaptive.

Questions

- Does natural selection favor introgression from adapted populations?
- Are loci controlling highland-lowland phenotypic differences adaptive?
- Is there evidence for convergent evolution in independently adapted highland populations?
- Did natural selection on gene expression contribute to high elevation adaptation?

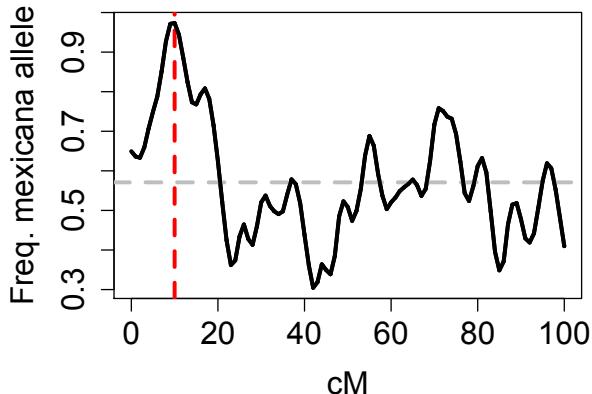
Aim 2.1 Population genetics of adaptive introgression

We have documented extensive introgression between *mexicana* teosinte and highland maize landraces Hufford et al. (2013), demonstrating an overlap of introgression with QTL for macrohairs and stem pigmentation in teosinte (Lauter et al., 2004). However, due to the relatively low-density genotyping used, we were limited to identifying large regions of ancient introgression and unable to investigate evidence of selection. Here we propose to reuse the same nine sympatric and two allopatric populations, sampling 18 individuals from each. These populations provide an opportunity to compare selection on maize alleles (QTL from Aim 1.1) to those from *mexicana* and ask whether adaptive introgression is local and ongoing or largely a single event that occurred during maize colonization of the highlands. Correlations between genetic differentiation and recombination in these populations will also allow us to investigate selection against introgression (Brandvain et al., 2014), quantifying the linkage drag associated with introgression of teosinte alleles into a maize background.

In addition to analyzing introgression into landraces, we will investigate evidence for adaptive introgression in hybrid populations of teosinte. We will complement the Santa Rita population from Aim 1.2 with samples (already collected by collaborators) of 50 individuals from each of four additional admixed populations identified using data from Fang et al. (2012). Because these admixture events appear to be ancient (Pyhäjärvi et al., 2013), replicate populations should provide high resolution to assess parallel evolution and phenotypic selection. As these populations are at the extreme high elevational range of *parviglumis*, we predict we will see evidence of adaptive introgression from *mexicana*. Population genetic theory predicts that adaptive loci which have introgressed due to natural selection should show distinct signals of elevated admixture, and our preliminary simulation results bear out this prediction (Figure 3).

Samples from all populations will be genotyped using GBS. Teosinte populations will be genotyped at higher coverage (48 plex) to decrease errors in calling heterozygotes. In each population we will apply population genetic approaches utilizing evidence from both the site frequency spectrum (Nielsen et al., 2005) and haplotype structure (Voight et al., 2006) to identify loci under selection. In teosinte populations we will use both haplotype (Price et al., 2009) and heterozygosity-based (Geneva et al., 2015) methods to identify introgressed segments in individual populations. Loci showing evidence of introgression and selection will be compared to those underlying QTL in maize and teosinte populations from Aim 1 and those

Figure 3: Analysis of 100 generations of simulated admixture between *mexicana* and *parviglumis* across a 100cM chromosome. A beneficial *mexicana* allele with selection strength $s=0.1$ is introgressed at position 10cM (red vertical line), showing that deviation from background variation in ancestry (horizontal gray line) can be used to detect selection in admixed populations.



showing evidence of selection based on expression data from Aim 2.2. Quantitative genetic theory suggests, however, that adaptive phenotypic change can occur without strong selection on individual loci (Le Corre and Kremer, 2012). To search for evidence of selection on highland phenotypes, we will employ recently developed methods from Co-PI Coop (Berg and Coop, 2014) that provide a powerful statistical framework to identify coordinated shifts in allele frequencies at causative QTL (from Aim 1). These methods will allow us to identify which phenotypes mapped in Aim 1 or other populations (e.g. Wallace et al., 2014) show evidence of selection and in which populations. Comparison among populations of maize and teosinte will highlight patterns of repeated evolution, indicative of the possibility that standing genetic variation or multiple pathways (a larger mutational target) can be utilized by plants to achieve similar phenotypic outcomes (Ralph and Coop, 2010).

Expected outcomes: 1) Identification of adaptive loci in teosinte and cultivated maize populations, 2) Evidence for or against convergent evolution among populations and subspecies 3) Identification of selection on individual phenotypic traits, 4) Quantification of selection against introgression across other regions of the genome.

Aim 2.2 Population genetics of gene expression adaptation

In Aim 2.1, we will study the population genetics of introgression to infer loci under selection. Here we will use population genetic variation in gene expression traits to pinpoint selected genes, and to link those genes to functional traits. We and others have documented considerable gene expression divergence associated with maize domestication (Swanson-Wagner et al., 2012; Lemmon et al., 2014), and an enrichment of regulatory variants among loci showing evidence of selection in teosinte (Pyhäjärvi et al., 2013). Among genetic loci that affect gene expression, *cis*-regulatory variants often account for the largest differences in expression (Song et al., 2013; Buil et al., 2014), drive additive gene expression variation that may be efficiently targeted by selection (Ronald and Akey, 2007; Lemmon et al., 2014), and are relatively stable across environments, tissues, and genetic backgrounds (Springer and Stupar, 2007; Buil et al., 2014).

We will use allele-specific expression (ASE) to scan the genome for genes that have undergone adaptive divergence (Q_{ST} , Leinonen et al., 2013) in the *cis*-control of gene expression between high and low elevation maize landraces in Mexico and South America. ASE assays measure expression differences between the two alleles of a gene in the same tissue, controlling for environmental and technical variation among individuals and samples. These expression differences are caused by genetic differences in linked *cis*-regulatory elements. ASE therefore can directly isolate functional genetic variation *in-situ* without large mapping populations. By identifying genes that show *cis*-regulatory divergence between low and high eleva-

tion populations, we will learn both the genomic loci involved in local adaptation, and the molecular pathways they control.

We will select 20 outbred landraces each from high ($> 2000\text{m}$) and low ($< 1600\text{m}$) elevation sites in Mexico and South America from a panel of individuals we have previously analyzed (Takuno et al., 2015), in addition to the four parents of our QTL populations in Aim 1.1. We will create F1 hybrid families of all 84 accessions by crossing each landrace to B73. We will grow two plants from each F1 family at the high and low elevation Mexican field locations in parallel with the QTL populations. From each plant, we will sample leaf and seedling stem tissue as rapidly as possible centered around mid-day on the day when the majority of plants reach the v4 leaf stage. Tissue will be flash-frozen in liquid nitrogen and transferred to a dry ice bath in the field for transport to Langebio. We will make strand-specific RNAseq libraries (Zhong et al., 2011) and pool for multiplexed sequencing in 32 lanes on the HiSeq3000 at the UC Davis Genome Center (100bp paired-end reads, 10M reads/sample). To prevent mapping biases that could cause erroneous ASE calls, particularly against more divergent highland alleles, we will re-sequence the exomes of all 80 outbred parents using the Nimblegen Maize SeqCap EZ, followed by multiplexed sequencing (12/lane for 50X coverage of all transcribed regions), and map reads to parent-specific pseudo-transcriptomes (Lemmon et al., 2014).

For each gene, we will measure the \log_2 expression ratios of each landrace allele relative to B73. Due to independent assortment of alleles in the outbred parents of our two sequenced F1 individuals, we will assay on average 30 of the 40 sampled alleles per elevation per population, providing high power to estimate expression divergence associated with elevation at each field location and in each tissue. With two individuals of each family at each of the two field locations, on average 22.5 of the 40 alleles will be assayed in both environments allowing for tests of field location \times elevation effects on ASE. Replicate individuals of F1 families from the inbred QTL population parents will provide estimates of among-individual variation in ASE.

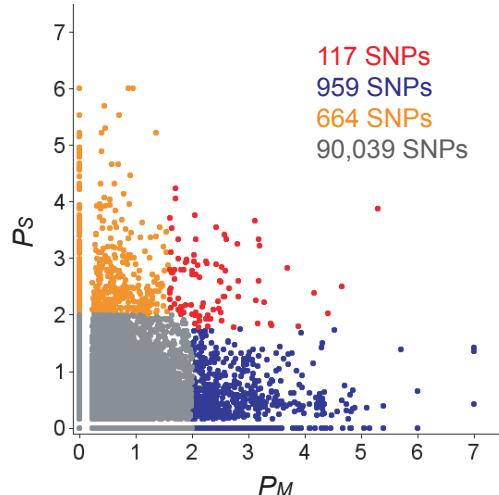
Individually, high gene expression divergence (high Q_{ST}) associated with elevation for a particular gene is suggestive of selection on that gene. Coordinated divergence across groups of related genes provides much stronger evidence of the action of natural selection. First, we will test if genes under QTL identified in Aim 1 or adaptively introgressed regions identified in Aim 2.1 show enrichments for high Q_{ST} between high and low elevation populations, providing additional evidence that these loci were important for local adaptation. Second, we will identify molecular pathways and gene function groups where a large proportion of genes diverged in the same direction (higher or lower expression) between low and high elevation populations using the sign test (Orr, 1998; Bullard et al., 2010). Such coordinated changes are unlikely unless the activities of the pathways themselves have been shaped by selection.

Expected outcomes: 1) A map of genes within and outside Aim 1 QTL that show evidence for adaptive divergence in gene expression between high and low elevation landraces in Mexican and South American populations. 2) Candidate gene pathways and functional groups that underwent directional selection for gene expression activity during adaptation to high or low elevation environments.

Aim 2 Preliminary Results:

We have worked extensively on the population genetics of highland adaptation in maize and teosinte. Pyhäjärvi et al. (2013) explored local adaptation in *parviflumis* and *mexicana* populations, finding loci showing evidence of selection and association with elevation and highlighting the importance of regulatory variants and large inversions. Hufford et al. (2013) identified genomic regions in highland maize that have introgressed from *mexicana* and demonstrated that maize with *mexicana* alleles showed highland phenotypes and superior growth under cold conditions, suggesting an adaptive role for introgression and motivating our population genetic analyses in Aim 2.1. Finally, Takuno et al. (2015) explored selection from a collection of maize from the highlands of Mexico and South America and found little overlap in the genes important for adaptation (Figure 4), consistent with an important role for selection on standing genetic variation in lowland maize and *parviflumis*. All germplasm necessary for Aim 2 has already been collected, and many of the B73 F1s needed in Aim 2.2 have already been made; the remainder will be finished by year 1 of the grant.

Figure 4: Little overlap of adaptive loci between continents. Shown is a scatter plot of $-\log_{10}$ empirical p-values of genetic differentiation (F_{ST}) in Mexico (P_M on x-axis) and S. America (P_S on y-axis). SNPs showing evidence of selection are highlighted in blue (Mexico), orange (S. America), or red (both Mexico and S. America), along with the number of SNPs in each category.



Aim 2 Potential Challenges:

GBS data is known to have a high heterozygote error rate, potentially complicating the identification of haplotypes. We have experience working with haplotype analyses using GBS data (Takuno et al., 2015), but can also take advantage of methods to detect introgression (e.g. Geneva et al., 2015) and selection (e.g. Nielsen et al., 2005) that do not require haplotype information.

Field-collections of tissue for RNAseq may be challenging; if we are unable to collect sufficiently homogeneous and well-staged tissue in either field location, we will repeat the experiment in a greenhouse at UC Davis.

Exome sequence provides our best reference for mapping RNA-seq reads, but will miss novel genes or transcripts not included in the array. To test for this we will also map reads to the *de novo* genome assemblies of the parents in Aim 1.1.

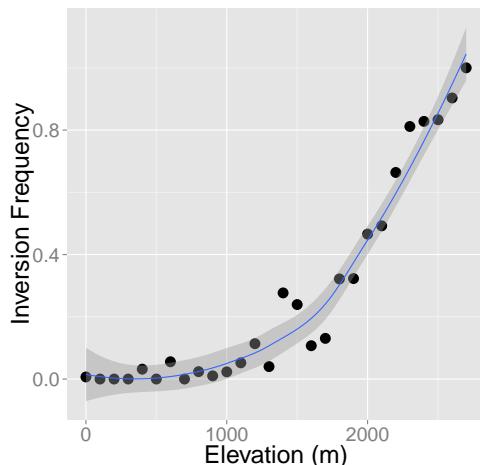
Aim 3 Functional characterization of adaptive QTL

After mapping QTL for highland adaptation (Aim 1) and studying their adaptive significance (Aim 2), in this aim we will investigate the functional genetic basis of a putatively adaptive region, *Inv4m*, an inversion polymorphism located on the long arm of chromosome 4 (169-180Mb). Our previous work (Hufford et al., 2013; Pyhäjärvi et al., 2013) identified a robust signature of introgression of the inverted haplotype from *mexicana* into maize in the Mexican highlands. This region overlaps with a QTL identified in a *parviflumis* x *mexicana* cross (Lauter et al., 2004) associated with leaf pigmentation and pubescence and shows a dramatic cline in frequency from complete absence in maize from the Mexican lowlands to fixation at the highest elevations of the Mexican Central Plateau (Figure 5). We will first study the phenotypic effects of alleles at *Inv4m* introgressed into a common B73 background (Aim 3.1). We will then use RNA sequencing to characterize the molecular effect of these alleles on genome-wide gene expression traits and to identify potential candidate genes within this QTL (Aim 3.2).

Questions

- What are the phenotypic consequences of introgressing a single adaptive QTL?
- Do highland alleles confer greater or lesser plasticity to highland vs lowland environments?
- Can RNA-seq help refine QTL to identify candidate genes?

Figure 5: Frequency cline of the *Inv4m* inversion in Mexico. *Inv4m* is nearly absent in low elevation (<1500m) populations, but rises to frequencies of > 90% in the highlands. Data from (Hearne et al., 2014)



Aim 3.1 Functional evaluation of *Inv4m* in the field

To evaluate the *Inv4m* polymorphism, we have generated BC4 NILs from crosses of highland haplotypes into the reference genome inbred B73. We selected as donors the Mexican highland landraces Palomero Toluqueño (PT) and Conico (CO), and one *mexicana* teosinte. PT is a popcorn originating from the highland valleys of central Mexico, is considered basal to the Mexican highland landrace radiation (Reif et al., 2006), and exhibits the highest level of *mexicana* introgression among characterized material (Matsuoka et al., 2002). CO is an economically important race in the central Mexican highlands and is proposed to be a derivative of PT. We introgressed the highland allele at *Inv4m* using simple sequence repeat markers and a PCR assay designed to detect a SNP diagnostic of the inverted *mexicana* haplotype (Hufford et al., 2013). This strategy has been validated by GBS analysis of PTxB73 families, confirming introgression of *Inv4m* in its entirety, with no recombination detected across a region of ~20Mb. In addition, we will include a B73 BC4S4 NIL carrying a lowland teosinte (*parviflumis*) haplotype selected from an existing introgression library made by CoPI Flint-Garcia.

We will evaluate NILs at three elevations during Years 2 and 3 (Table 2) for a range of fitness and agronomic traits (Figure 2). We will characterize the NILs *per se* and as crosses to a highland hybrid tester (CML457xCML459). We have grown B73 and NIL materials successfully in all of our proposed sites, and can be confident of obtaining meaningful trait measurements. In each trial, we will plant and collect phenotypic data from the 10 genotypes (4 NILs -- PT, CO, *mexicana* and *parviflumis* -- and B73 individually and each crossed to the tester) in a randomized complete block design trial with ten replicates of plots of ten plants (total of 100 individuals per genotype).

Expected outcomes: 1) Estimation of phenotypic effects of lowland and highland haplotypes of the *Inv4m* candidate region, including genotype x environment effects on the basis of replicated trials at three elevations.

Aim 3.2 Transcriptional reprogramming driven by *Inv4m* in response to cold

To further characterize the effects of the *Inv4m* polymorphism and identify functional traits underlying its phenotypic effects, we will measure genome-wide gene expression differences between B73 and the four NILs characterized in Aim 3.1. Genes inside the inversion with divergent expression between highland and lowland alleles are good candidates for loci underlying the phenotypic effects of this inversion, potentially dissecting the large linkage block. Also, altered co-expression profiles of genes in known pathways will provide insight into molecular mechanisms underlying these phenotypic effects (Swanson-Wagner et al., 2012). However, the environmental, developmental, and tissue contexts where such functional differences will be most important are unknown, necessitating a systematic approach. We will focus on the response to

Table 3: RNAseq tissues, as described in the B73 gene atlas (Sekhon et al., 2011)

Growth Stage	Tissues
V1	Pooled leaves, Primary root
V3	Stem and SAM, First leaf and sheath, Topmost leaf
V5	Shoot tip, Tip of stage-2 leaf, Base of stage-2 leaf

cold during early development when low temperatures are likely to be a strong selective force in highland environments. Our previous results show that highland genotypes are able to maintain a higher growth rate than lowland genotypes in cool temperatures (Hufford et al., 2013). We are particularly interested in genes that respond to temperature in B73, but are constitutively activated or repressed in the highland NILs (but not the *parviflumis* NIL), paralleling the growth rate results.

We will grow plants of the five genotypes in growth chambers set to warm (32C/25C day/night) or cold (23C/11C) temperatures. We will sample eight tissues from each genotype, corresponding to tissues of the B73 gene atlas (Sekhon et al., 2011, Table 3) 3hrs after lights-on based on developmental stage, pooling tissue from three plants per genotype per tissue. These tissues were chosen to maximize the diversity of gene expression profiles identified in the atlas during early development. The whole experiment will be replicated four times for a total of 320 samples. These will be barcoded for multiplex sequencing in 32 lanes on the HiSeq3000 at the UC Davis Genome Center, aiming for 20M 50bp paired-end reads per sample. Reads will be mapped to the B73 genome, or the de novo assembly of the Palomero de Jalisco line from Aim 1.1 inside introgressed regions to prevent mapping biases, and differential expression of genes, pathways, and gene sets will be tested with the *R* packages *voom* and *limma* (Ritchie et al., 2015).

Expected outcomes: 1) Lists of genes and *a priori* gene sets differentially expressed according to *Inv4m* genotype. 2) Candidate genes inside the *Inv4m* inversion and other QTL identified in Aim 1 that may control highland phenotypes, particularly the maintenance of photosynthesis and growth under cold conditions.

Aim 3 Preliminary Results:

All NIL stocks in Aim 3.1 have been advanced to BC4 and have been confirmed to carry *Inv4m*. We have begun to self-pollinate the NIL stocks to obtain families homozygous for the highland haplotypes, and all stocks will be ready by year 2. The B73 x *parviflumis* BC4S4 families that carry a lowland haplotype in the *Inv4m* region are already available from an existing collection generated by Co-PI Flint-Garcia.

Aim 3 Potential Challenges:

Poor performance of B73-based material in the field may be problematic, especially in the highland environment: our use of additional test-cross stocks is to address this potential difficulty. Additionally, we recognize that the use of BC4 material limits our power to detect epistatic interactions. In the case of pigmentation, where well characterized loci are known to lie outside our candidate region, we will also move *Inv4m* haplotypes into appropriate tester backgrounds to allow pigmentation expression.

The causal polymorphisms in *Inv4m* may not act through gene expression perturbations of linked genes, or in the specific tissues or environments that we test in Aim 3.2. We have attempted to design a thorough sampling strategy to maximize the chance of observing gene expression differences if they exist. But regardless of whether we can identify candidates for causal genes, we will learn about molecular mechanisms underlying the phenotypic effects of the inversion through inspecting genome wide gene expression differences.

Broader Impacts of the Proposed Work

Our proposal seeks to broaden educational opportunities and scientific outreach through an exchange program, a phenotyping workshop, and a set of farmer field days in collaboration with International Maize and Wheat Improvement Center (CIMMYT) in Mexico. 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. Finally, the germplasm created as part of this proposal will likely be of use to other researchers and in breeding programs.

Exchange Program

We propose an international student exchange program between our teams in the US and Mexico. Our goal is to involve students directly in research while fostering intercultural exchange and promoting future international research opportunities. Participating Mexican students will learn computational management of large datasets that can be introduced to their respective laboratories and peers. American exchange students will benefit from experience in highland and lowland environments as well as opportunities to work with landraces and teosinte in the field. Over the course of the grant, we will fund 10 graduate or undergraduate students for 3-month research internships in one of the collaborating laboratories. Students will participate in research projects directly relating to the research focus of the grant, including developing mapping populations, mapping traits, or analysis of population genetic or expression data, with the expectation that such research will often contribute to publications. Students will give a presentation to both their home and host lab detailing their work over the 3-month period. Each of the PIs will participate, sending students to Mexico and/or accepting students from Mexico for internships. PI Ross-Ibarra will manage the program, as he is fluent in Spanish and has past experience with a very similar program (NSF 0922703). Over the last four years his lab has hosted eight Mexican students, two of whom coauthored a publication resulting from their work, and a third who has continued on to a PhD program in the U.S.

Phenotyping Workshop

The USDA-ARS group in Columbia, MO has developed a streamlined phenotypic data collection system utilizing a handheld barcode device, barcoded plant tags, and barcoded phenotyping tools in order to maximize efficiency. We will host a phenotyping workshop in Columbia, MO during each year of the grant in order to transfer this system to other research institutions. The phenotyping workshop will include topics on experimental design, setting up the FieldBook database (creating locations, traits, and projects, assigning plots and measurements to projects, generating plant tags, loading the program and trait groups to the Palm for data collection), and data collection (specific traits related to local adaptation of interest to our group, synchronizing data with the desktop/laptop database, managing data conflicts, running reports). This proposal will provide travel support for instructors. The workshop will be free but participants will be expected to pay for their own travel and purchase their own Palm handheld (a few devices will be available for participants not wishing to purchase a device ahead of time). The workshop will be held in late summer so participants can gain hands-on experience in data collection in the corn field. We have already held one successful workshop in 2014 for lab members of each of the participating labs as part of our CNIC funding. Workshop announcements will be posted to multiple email lists such as the Corn Breeding Research, Maize bionet, and evoldir list-servs, the National Association of Plant Breeders Newsletter, etc. in order to attract breeding and genetics researchers from as many plant communities as possible. Co-PI Flint-Garcia already has experience organizing such events, having been involved in the recruitment of participants for the 2015 Panzea GBS workshop to be held in Columbia, MO. Surveys will be administered after each workshop in order to gauge the value of the workshop and make improvements for future years.

Farmer Field Days

Working with the CIMMYT seedbank (see letter of support from the head of the maize germplasm at CIMMYT, Dr. Denise Costich), we will co-host annual farmer field days in the highland field site during years

2-5. Field days serve as a way to regenerate valuable highland maize germplasm while demonstrating and sharing such diversity with the agricultural community. Following the format of the successful trial field day in November 2014, demonstration plots of diverse highland landrace and improved material will be planted for presentation to participating farmers, including a subset of the material to be evaluated in the experimental portion of this project. Project members will attend field days and engage in dialogue with the agricultural community, promoting the diversity of highland maize and explaining the scientific basis of the project. Concomitantly, project members will gain invaluable insight into the nature of highland maize cultivation, traits important to farmers, and how farmer selection impacts maize evolution in the highlands.

Educational Software

Population and quantitative genetics are key to understanding genetics and evolution, and basic understanding of genetic variation is important for all people due to the rise of personal genomics and genomic medicine (e.g. Redfield, 2012). We will develop undergraduate teaching modules in population and quantitative genetics using data from this project. These will be tested and integrated into large undergraduate teaching courses (evolutionary biology and genetics) at UC Davis and graduate courses at UC Davis and Iowa State (ecological genomics). We have already begun to develop and distribute some of these resources, such as genome-scale demonstrations of Hardy Weinberg Equilibrium. These underscore the usefulness of population genetics in describing real world patterns and expose students to the wealth of genomic data being collected. Other examples will include using association mapping data to demonstrate quantitative genetics models and explaining concepts of genetic and genealogical ancestry using genomic identity by descent. These modules will be prepared in the open source language R to ensure that they are easily used, modified, and distributed. They will be designed to be tailored for use at multiple levels: from basic concepts in introductory classes to programming exercises for upper division courses. Modules will be publicly distributed via Github (see Data Management Plan).

Germplasm Resources

This project will generate multiple germplasm resources that can be used for mapping additional phenotypes (our F2:3 populations) or investigate introgressions from exotic lines (our NIL populations). Such material could be of interest to the Germplasm Enhancement of Maize project as well as to public and private breeders in the US, Mexico, and abroad. Seed generated in this project will be deposited in the USDA-ARS Maize Stock Center with backups kept at Iowa State and USDA-ARS Missouri. In addition, seed will be made available in Mexico through the Mexican national agronomic agency INIFAP. Finally, seed from our collections of teosinte will enhance sampling and provide diversity not currently present in germplasm banks and will be deposited for curation at CIMMYT.

Results From Prior 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 postdoctoral 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, Hufford: USDA #2009-65300-05668: Scanning for Climate Change: High-throughput Discovery of Loci for Advanced Breeding in Maize

\$300,000. 09/01/12-08/31/14. PI Jeffrey Ross-Ibarra, co-PI M.B. Hufford

Intellectual merit This proposal set out to use population genetic methods to identify loci showing large allele frequency differences between highland and lowland maize, and assess whether population genetics could provide evidence of parallel adaptation. We found that maize had adapted to the highlands of central Mexico via introgression from teosinte, and are currently writing up our results showing little overlap (and theory showing why there should be little overlap) between selected loci in S. America and Mexico.

Broader impacts Ross-Ibarra has released code for data analysis and trained a number of undergraduate students on this project.

Publications Hufford et al. (2013); Pyhäjärvi et al. (2013)

Coop: #1262327: Collaborative Research: ABI Innovation: Visualization And Statistics For Spatial Population Genomic Analysis.

\$327,156. 05/01/13-04/30/16. PI Graham Coop

Intellectual merit We are developing a set of spatial statistics methods based on Gaussian random fields for the analysis of geographic population genomics data, we have also developed theory both for linkage disequilibrium in contact zones and a set of theoretic results about adaptation in a geographic setting.

Broader impacts The R package of the BEDAZZLE software has been released online, and has already been used by many molecular ecologists. We have developed a freely available, online population genetics textbook as well as a series of R exercises aimed at graduate students.

Publications Bradburd et al. (2013); Ralph and Coop (2014a,b); Sedghifar et al. (2015); Bradburd et al. (2015)

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

\$13,311,185 (\$3.2M to Ross-Ibarra and \$1.2M to Flint-Garcia), 05/15/13-04/30/18. PI Edward Buckler, co-PIs J. Doebley, J. Holland, S. Flint-Garcia, Q. Sun, S. Mitchell, J. Ross-Ibarra

Intellectual merit In the first two years we have developed accurate imputation approaches, found evidence for the importance of deleterious variants and non-genic polymorphisms in heterosis and GWAS, documented differences in recombination among the parents of the NAM population, and found population genetic evidence suggesting the importance of demography and purifying selection across the genome. The grant has produced > 20 total publications in its first two years (only publications involving PIs Flint-Garcia and Ross-Ibarra are shown below).

Broader impacts This project has included 12 postdoctoral and 12 graduate trainees, a GBS workshop and traveling maize exhibit, and an on-line maize evolution resource for teachers.

Publications Peiffer et al. (2013); Romay et al. (2013); Wills et al. (2013); Mezmouk and Ross-Ibarra (2014); Peiffer et al. (2014); Hirsch et al. (2014); Sood et al. (2014); Tiffin and Ross-Ibarra (2014); Makarevitch et al. (2015); da Fonseca et al. (2015)

References Cited

- Mauricio R Bellon, David Hodson, and Jon Hellin. Assessing the vulnerability of traditional maize seed systems in mexico to climate change. *Proceedings of the National Academy of Sciences*, 108(33):13432–13437, 2011.
- JJ Berg and G Coop. The population genetic signature of polygenic local adaptation. *PLoS Genetics*, In press, 2014.
- *Gideon Bradburd, Peter L Ralph, and Graham Coop. A spatial framework for understanding population structure and admixture. *bioRxiv*, page 013474, 2015.
- *GS Bradburd, PL Ralph, and GM Coop. Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, 67(11):3258–3273, 2013.
- Y Brandvain, AM Kenney, L Flagel, G Coop, and A Sweigart. Speciation and introgression between *mimulus nasutus* and *mimulus guttatus*. *PLoS Genetics*, In press, 2014.
- KW Broman, H Wu, S Sen, and GA Churchill. R/qtI: Qtl mapping in experimental crosses. *Bioinformatics*, 19(7):889–890, 2003.
- *ES Buckler, JB Holland, PJ Bradbury, CB Acharya, PJ Brown, C Browne, E Ersoz, S Flint-Garcia, A Garcia, JC Glaubitz, MM Goodman, C Harjes, K Guill, DE Kroon, S Larsson, NK Lepak, H Li, SE Mitchell, G Pressoir, JA Peiffer, MO Rosas, TR Rocheford, MC Romay, S Romero, S Salvo, H Sanchez Villeda, HS da Silva, Q Sun, F Tian, N Upadyayula, D Ware, H Yates, J Yu, Z Zhang, S Kresovich, and MD McMullen. The genetic architecture of maize flowering time. *Science*, 325(5941):714–718, 2009.
- Alfonso Buil, Andrew Anand Brown, Tuuli Lappalainen, Ana Viñuela, Matthew N Davies, Hou-Feng Zheng, J Brent Richards, Daniel Glass, Kerrin S Small, Richard Durbin, Timothy D Spector, and Emmanouil T Dermitzakis. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet*, 47(1):88–91, December 2014.
- J H Bullard, Y Mostovoy, S Dudoit, and R B Brem. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci. U.S.A.*, 107(11):5058–5063, March 2010.
- *Rute R da Fonseca, Bruce D Smith, Nathan Wales, Enrico Cappellini, Pontus Skoglund, Matteo Fumagalli, José Alfredo Samaniego, Christian Carøe, María C Ávila-Arcos, David E Huffnagel, et al. The origin and evolution of maize in the southwestern united states. *Nature Plants*, 1(1), 2015.
- JF Doebley. Maize introgression into teosinte-a reappraisal. *Annals of the Missouri Botanical Garden*, pages 1100–1113, 1984.
- RJ Elshire, JC Glaubitz, Q Sun, JA Poland, K Kawamoto, ES Buckler, and SE Mitchell. A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PLoS One*, 6(5):e19379, 2011.
- *Z Fang, T Pyhäjärvi, AL Weber, RK Dawe, JC Glaubitz, J Gonzalez Jde, C Ross-Ibarra, J Doebley, PL Morrell, and J Ross-Ibarra. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, 191(3):883–894, 2012.
- A Fournier-Level, A Korte, MD Cooper, M Nordborg, J Schmitt, and AM Wilczek. A map of local adaptation in *arabidopsis thaliana*. *Science*, 334(6052):86–89, 2011.
- H B Fraser. Gene expression drives local adaptation in humans. *Genome Research*, 23(7):1089–1096, July 2013.
- K Fukunaga, J Hill, Y Vigouroux, Y Matsuoka, J Sanchez, KJ Liu, ES Buckler, and J Doebley. Genetic diversity and population structure of teosinte. *Genetics*, 169(4):2241–2254, 2005.

Anthony J. Geneva, Christina A. Muirhead, Sarah B. Kingan, and Daniel Garrigan. A new method to scan genomes for introgression in a secondary contact model. *PLoS ONE*, 10(4):e0118621, 04 2015.

Paul Gepts. The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology*, 18(0):51 -- 59, 2014.

JC Glaubitz, TM Casstevens, F Lu, J Harriman, RJ Elshire, Q Sun, and ES Buckler. Tassel-gbs: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*, 9(2):e90346, 2014.

F John Goodstal, Glenn R Kohler, Leslie B Randall, Arnold J Bloom, and Dina A St Clair. A major qtl introgressed from wild *lycopersicon hirsutum* confers chilling tolerance to cultivated tomato (*lycopersicon esculentum*). *Theoretical and Applied Genetics*, 111(5):898--905, 2005.

Sarah Hake and Jeffrey Ross-Ibarra. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*, 4, 2015.

AM Hancock, B Brachi, N Faure, MW Horton, LB Jarymowycz, FG Sperone, C Toomajian, F Roux, and J Bergelson. Adaptation to climate across the *arabidopsis thaliana* genome. *Science*, 334(6052):83--86, 2011.

Sarah Hearne, Charles Chen, Ed Buckler, and Sharon Mitchell. Unimputed gbs derived snps for maize landrace accessions represented in the seed-maize gwas panel. <http://hdl.handle.net/11529/10034> International Maize and Wheat Improvement Center [Distributor] V2 [Version], 2014.

*Candice N Hirsch, Sherry A Flint-Garcia, Timothy M Beissinger, Steven R Eichten, Shweta Deshpande, Kerrie Barry, Michael D McMullen, James B Holland, Edward S Buckler, Nathan Springer, et al. Insights into the effects of long-term artificial selection on seed size in maize. *Genetics*, 198(1):409--421, 2014.

JB Holland. Computer note. epistacy: A sas program for detecting two-locus epistatic interactions using genetic marker information. *Journal of Heredity*, 89(4):374--375, 1998.

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

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

*HY Hung, LM Shannon, F Tian, PJ Bradbury, C Chen, SA Flint-Garcia, MD McMullen, D Ware, ES Buckler, JF Doebley, and JB Holland. Zmcct and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc Natl Acad Sci U S A*, 109(28):E1913--E1921, 2012.

ES Lander, P Green, J Abrahamson, A Barlow, MJ Daly, SE Lincoln, and L Newburg. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1(2):174--181, 1987.

N Lauter, C Gustus, A Westerbergh, and J Doebley. The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics*, 167(4):1949--1959, 2004.

V Le Corre and A Kremer. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, 21(7):1548--1566, 2012.

Tuomas Leinonen, R J Scott McCairns, Robert B O'Hara, and Juha Merilä. QST--FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat Rev Genet*, 14(3):179--190, February 2013.

Zachary H Lemmon, Robert Bukowski, Qi Sun, and John F Doebley. The Role of *cis* Regulatory Evolution in Maize Domestication. *PLoS Genet*, 10(11):e1004745, November 2014.

DB Lobell, W Schlenker, and J Costa-Roberts. Climate trends and global crop production since 1980. *Science*, 333(6042):616--620, 2011.

*Irina Makarevitch, Amanda J Waters, Patrick T West, Michelle Stitzer, Candice N Hirsch, Jeffrey Ross-Ibarra, and Nathan M Springer. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS genetics*, 11(1):e1004915, 2015.

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

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

R Nielsen, S Williamson, Y Kim, MJ Hubisz, AG Clark, and C Bustamante. Genomic scans for selective sweeps using snp data. *Genome research*, 15(11):1566--1575, 2005.

H Allen Orr. Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data. *Genetics*, 149(4):2099--2104, August 1998.

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

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

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

Michael D Purugganan and Dorian Q Fuller. The nature of selection during plant domestication. *Nature*, 457(7231):843--848, 2009.

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

P Ralph and G Coop. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647--668, 2010.

*Peter L Ralph and Graham Coop. Convergent evolution during local adaptation to patchy landscapes. *bioRxiv*, page 006940, 2014a.

*Peter L Ralph and Graham Coop. The role of standing variation in geographic convergent adaptation. *bioRxiv*, page 009803, 2014b.

RJ Redfield. Why do we have to learn this stuff? -- a new genetics for 21st century students. *PLoS Biology*, 10(7):e1001356, 07 2012.

JC Reif, ML Warburton, XC Xia, DA Hoisington, J Crossa, S Taba, J Muminović, M Bohn, M Frisch, and AE Melchinger. Grouping of accessions of mexican races of maize revisited with ssr markers. *Theoretical and Applied Genetics*, 113(2):177--185, 2006.

M E Ritchie, B Phipson, D Wu, Y Hu, C W Law, W Shi, and G K Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl. Acids Res.*, 43(7):e47--e47, April 2015.

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

James Ronald and Joshua M Akey. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. *PLoS ONE*, 2(7):e678, 2007.

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

Jeffrey Ross-Ibarra, Peter L Morrell, and Brandon S Gaut. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences*, 104(suppl 1): 8641–8648, 2007.

O Savolainen, M Lascoux, and J Merilä. Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11):807–820, 2013.

*Alisa Sedghifar, Yaniv Brandvain, Peter L Ralph, and Graham Coop. The spatial mixing of genomes in secondary contact zones. *bioRxiv*, page 016337, 2015.

Rajandeept Sekhon, Haining Lin, Kevin L Childs, Candice N Hansey, C Robin Buell, Natalia de Leon, and Shawn M Kaeplker. Genome-wide atlas of transcription during maize development. *Plant J.*, 66(4): 553–563, March 2011.

Kristine Skarbø and Kristin VanderMolen. Maize migration: key crop expands to higher altitudes under climate change in the andes. *Climate and Development*, Online Ahead of Print:1–11, 2015.

Gaoyuan Song, Zhibin Guo, Zhenwei Liu, Qin Cheng, Xuefeng Qu, Rong Chen, Daiming Jiang, Chuan Liu, Wei Wang, Yunfang Sun, Liping Zhang, Yingguo Zhu, and Daichang Yang. Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F1 hybrids. *BMC Plant Biol.*, 13(1):221, 2013.

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

Nathan M Springer and Robert M Stupar. Allele-Specific Expression Patterns Reveal Biases and Embryo-Specific Parent-of-Origin Effects in Hybrid Maize. *The Plant Cell* ..., 19(8):2391–2402, August 2007.

R Swanson-Wagner, R Briskine, R Schaefer, MB Hufford, J Ross-Ibarra, CL Myers, P Tiffin, and NM Springer. Reshaping of the maize transcriptome by domestication. *PNAS*, 2012.

Shohei Takuno, Peter Ralph, Kelly Swarts, Rob J Elshire, Jeffrey C Glaubitz, Edward S Buckler, Matthew B Hufford, and Jeff Ross-Ibarra. Independent molecular basis of convergent highland adaptation in maize. *bioRxiv*, page 013607, 2015.

MI Tenaillon and A Charcosset. A European perspective on maize history. *Comptes rendus biologies*, 334 (3):221–228, 2011.

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

BF Voight, S Kudaravalli, X Wen, and JK Pritchard. A map of recent positive selection in the human genome. *PLoS Biology*, 4(3):e72, 2006.

Jason G Wallace, Peter J Bradbury, Nengyi Zhang, Yves Gibon, Mark Stitt, and Edward S Buckler. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS genetics*, 10(12):e1004845, 2014.

HG Wilkes. *Teosinte: the closest relative of maize*. PhD thesis, Harvard University, 1967.

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

CA Winkler, GW Nelson, and MW Smith. Admixture mapping comes of age. *Annual review of genomics and human genetics*, 11:65–89, 2010.

JB Yoder, J Stanton-Geddes, P Zhou, R Briskine, ND Young, and P Tiffin. Genomic signature of adaptation to climate in medicago truncatula. *Genetics*, 2014.

Li Zhao, Janneke Wit, Nicolas Svetec, and David J Begun. Parallel Gene Expression Differences between Low and High Latitude Populations of *Drosophila melanogaster* and *D. simulans*. *PLoS Genet*, 11(5):e1005184 EP --, May 2015.

Silin Zhong, Je-Gun Joung, Yi Zheng, Yun-ru Chen, Bao Liu, Ying Shao, Jenny Z Xiang, Zhangjun Fei, and James J Giovannoni. High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation. *Cold Spring Harb Protoc*, 2011(8):pdb.prot5652--pdb.prot5652, August 2011.

Budget Justification

A Personnel

Funds are requested for the summer salary of Co-PI Runcie for years 2-5 at the rate of 2/12th a base pay of \$93,333. Salary is also requested for Sr. Personnel Kate Crosby for 12 months per year, starting at \$48,000. In year four of the grant Dr. Crosby would transition to an Assistant Project Scientist position, with a base salary of \$56,385. Dr. Crosby will lead the population genetic analysis of introgression and admixture, and consult on QTL and admix mapping.

B Other Personnel

1 Postdoctoral Scholars

Funds are requested to support a postdoc for 12 months per year, for all five years of the proposal with a base salary of \$44,566. The postdoc would lead the two RNA expression projects, the growth-chamber project of Aim 3.2 in year 1-2, and the field-based project of Aim 2.2 in years 3-5. The postdoc will be responsible for tissue collection, genotyping and data analysis.

2 Technician

Funds are requested for the first three years of the grant for a 75%-time technician (Assistant Specialist I) to extract DNA and RNA, prepare GBS, genomic and transcriptomic sequencing libraries, perform root chilling experiments, facilitate genotyping/sample prep for collaborating labs, and coordinate the summer exchange program. In the latter two years of the grant this is reduced to 25% time to continue facilitating any genotyping and administration of the summer exchange program. The base salary for this positions is \$42,144.

3 Graduate students

Funds are requested to support one graduate student at 50% FTE at \$28,000 during the academic year for each year of the project. The student will work with Dr. Coop on the population genetic analysis of the admixed populations in Aim 2.

C Fringe Benefits

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

- Faculty: 18%-18.5% in FY 2017-2018, 18.5%-19.1% in FY 2018-2019, 19.1%-19.7% in FY 2019-2020, 19.7%-20.03% in FY 2020-2021
- Postdocs: 17% - 18% in FY 2016-2017, 18%-18.5% in FY 2017-2018, 18.5%-19.1% in FY 2018-2019, 19.1%-19.7% in FY 2019-2020, 19.7%-20.03% in FY 2020-2021
- Graduate students: 1.3% for all years
- Assistant Project Scientist: 40.9%-42.1% in FY 2019-2020, 42.1%-43.4% in FY 2020-2021
- Assistant Specialist 38.4%-38.5% in FY 2016-2017, 38.5%-39.7% in FY 2017-2018, 39.7%-40.9% in FY 2018-2019, 40.9%-42.1% in FY 2019-2020, 42.1%-43.4% in FY 2020-2021

D Equipment

No equipment funds are requested.

E Travel

Domestic travel for the PIs Ross-Ibarra, Coop and Runcie, as well as one graduate student and two postdocs is budgeted at \$7,000 per year. This covers travel to the PI meeting or a domestic conference each year. Travel to Mexico is budgeted at \$4,000 per year and includes travel for fieldwork in the common garden sites, travel to the phenotyping workshop, and travel to the farmer field day.

F Participant Support

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

Starting in year 2, our highland maize farmer field day program at CIMMYT is budgeted at \$45,000 per year (see attached letter of support from Dr. Denise Costich, CIMMYT).

G Other Direct Costs

1 Materials and Supplies

In each of the first three years of the grant, \$15,000 is requested in materials and supplies for PI Ross-Ibarra for library prep for whole genome sequencing, RNA sequencing, and DNA extraction and preparation for GBS. This also includes funds computer supplies (storage for computer cluster, backup hard drives, etc.) and supplies for root chilling experiments to be done at UC Davis. In each of the five years, \$1,500 is budgeted for standard office supplies, computer supplies (desktop computer, backup hard drives), and other miscellaneous expenses for Co-PI Coop.

In years 1-3 of the grant, \$10,000 is requested for Co-PI Runcie for materials and supplies including standard office supplies, computer supplies including extra storage for our cluster and backup drives for lab members, and standard lab supplies including glass and plasticware, gels and chemicals. In years 4-5, this total decreases to \$5,000 to support primarily computational and office supplies.

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

Exome sequencing In year 2, the exomes of the 80 landraces used in Aim 2.2 will be sequenced using the Nimblegen Maize SeqCap EZ kit. Total costs for exome capture and library prep total \$13,000 for 80 samples. The samples will be multiplexed and sequenced in 4 lanes on the HiSeq3000 at the UC Davis Genome Center using the PE125bp kit to achieve 25X coverage of each sample for \$10,000.

RNA sequencing In Aim 3.2, 320 RNAseq libraries will be generated. In Aim 2.2, 640 RNAseq libraries will be generated. Cost to prepare RNAseq libraries in our lab are approximately \$50 per library, totaling \$48,000 between years 1 and 3. In [subsec:rnaeq] libraries will be multiplexed and sequenced in 32 lanes on the HiSeq3000 using the PE50bp kit, totaling \$48,000 in year 1. In Aim 2.2 libraries will be multiplexed and sequenced in 32 lanes on the HiSeq3000 using the PE125bp kit, totaling \$75,000 in year 3.

2 Publication Costs

In year two \$3,000 is requested for publication fees for two papers in an open access journal. In subsequent years \$4,500 is requested annually.

3 Subawards

The budget includes subwards to Iowa State (total \$882,867) and USDA-ARS (total \$476,171). We are also requesting a subaward to LANGE BIO (total \$500,137), as three of the project field sites are located in Mexico and will require substantial coordination and frequent visits for phenotyping, sampling, and other research activities. A Mexican co-PI also helps ensure the success of our exchange program and provides a host institution for U.S. students.

4 Graduate Student Tuition

Tuition for graduate students is charged to the project in proportion to the amount of effort the graduate student will work on the project. For a graduate student employed on the project for 9 academic months at 50% FTE, the tuition charge is \$32,786 in FY 2015-2016 to account for out-of-state tuition, \$19,451 in FY 2017 and increasing 10% each subsequent year for in-state tuition.

H Total Direct Costs

Total direct costs for UCD, including \$1,859,175 of subawards, comes to \$3,673,143.

I Indirect Costs

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

Facilities, Equipment, and Other Resources

Facilities, Equipment & Other Resources

UC Davis: Dr. Ross-Ibarra and Dr. Runcie together have seven 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. Runcie has access to considerable desk space and a conference room in a large shared office across from the lab. The PIs are both contributing partners in a large computer cluster, giving the labs dedicated access to 192 and 96 processors, respectively, with the opportunity for use of nearly 2,000 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. Both labs have in-building access to two Conviron PRG15 growth chambers, and nearby access to others through the UC Davis Controlled Environmental Facility: <http://greenhouse.ucdavis.edu/cef/description.html>.

Dr. Coop's dry space is located on the 3rd floor of the Storer building, which houses the Department of Evolution and Ecology. The space is newly renovated and consists of 3 offices that can seat a total of eight people, and a conference room. In addition, members of the lab have access to a separate conference room and other offices shared with the Begun, Langley, Lott, Kopp and Turelli groups. This group is part of the larger Center and Graduate Group for Population Biology, one of the leading graduate training programs in ecology and evolution in the world. Each current member of Dr. Coop's group has a quad-core Mac pro. The Coop lab also has access to the genome center computational facilities: <http://www.genomecenter.ucdavis.edu/core-facilities/>.

Iowa State: Project components completed in the Hufford Laboratory will include mapping population development, DNA isolation and PCR, and population genetic analysis of genotyping data. Population development will be carried out in field space available at the Curtiss Farm of Iowa State University (ISU). This facility is equipped with irrigation, tractors, tillage equipment, planters, and combines. Seed processing and cold storage facilities are also available on the ISU campus. The Hufford Laboratory has all equipment necessary for DNA isolation and PCR including centrifuges, thermal cyclers, an ultra-low freezer, water baths, a pH meter, balances, and an electrophoresis system. A gel imaging system and a NanoDrop spectrophotometer for DNA quantification are accessible through the Center for Plant Responses to Environmental Stresses at ISU. The DNA Facility at ISU provides access to cutting-edge genomic technology including Pacific Biosciences and HiSeq/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. Dr. Hufford has also recently collaborated with Research IT in the College of Liberal Arts

and Sciences at ISU to build the largest memory (1.5TB of RAM) computer on campus and has utilized this to successfully complete multiple genome assembly projects. Graduate students and postdocs in the Hufford group have access to desk space in multiple offices in Bessey Hall.

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

LANGEBIO: LANGEBIO is a recently founded unit of the Mexican graduate education and research institution CINVESTAV. LANGEBIO currently hosts 16 diverse research groups within the broader campus of CINVESTAV Irapuato. The institute's mandate is to conduct top-ranked research while promoting genomic knowledge for the protection and sustainable use of Mexican biodiversity. Its unique location in the agricultural center of Mexico facilitates field sampling and field experimentation. Dr. Sawers' group occupies half a laboratory bay in the main Langebio building, fully equipped for molecular biology, and including bench space for 12 people. In addition, Dr. Sawers has dedicated on site access to 65 sq. m of greenhouse space and humidity/temperature controlled seed storage. Further institutional facilities include a genomics/sequencing centre, computer cluster (66 nodes, 535 cores and 2000GB of memory) and microscopy facility (standard and fluorescent stereomicroscopes, compound microscopes with fluorescence and DIC capabilities, confocal microscope, laser capture dissection microscope, transmission and scanning electron microscopes), supported by a full administrative department. Off campus, Dr. Sawers and his group have successfully used the field sites detailed in the proposal over a number of seasons.

SEE APPENDIX A-1 UPLOADED AS A SUPPLEMENTARY DOCUMENT

Supplementary Documentation: Postdoctoral Researcher Mentoring Plan

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

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

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

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

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

The postdoc in the Ross-Ibarra lab will gain outreach experience by co-organizing farmer field days in Mexico with Dr. Denise Costich of CIMMYT and will also have the opportunity to work with and supervise exchange students. Additionally, the postdoc will host several informal workshops on computational tools for population/quantitative genetics and ecological niche modelling.

Finally, postdocs will be encouraged to take advantage of professional development programs offered by their local institutions and to attend conferences each year to present results and build relationships with other leaders in the field. 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. As a group, our labs have already had success placing postdoctoral scholars in careers industry, government, and academic positions, and we will continue to encourage postdocs to explore a range of career opportunities.

Supplementary Documentation A-1: Sharing of Results and Management of Intellectual Property

Data Types

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

Data Access, Sharing

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

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

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

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

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

All manuscripts 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. Manuscript preprints of published work will be updated with the final (unformatted) version of the document in accordance with publisher guidelines.

Finally, the group will host a group webpage on Github (the Ross-Ibarra lab website is already hosted there). The webpage will provide a description of and links to all of the products described above.

Data Archiving

All data, code, presentations, and publications will be made publicly available online (see above). Prior to public release, all data will be hosted locally. Dr. Ross-Ibarra will maintain a backup of all raw genotyping, sequence, and phenotyping data. His lab maintains a DROBO distributed backup server (robust to single disk failure) which will be expanded to include disks designated for this project. Analytical code will be backed up at Github, which maintains version-controlled backups.

Sample seed of each mapping population (and the generations of development) will be archived in temperature- and humidity-controlled facilities at Iowa State University and USDA-ARS Missouri. Sample accession data will be securely stored in a MySQL server hosted at the University of California, Davis and backed up on a weekly basis offsite. International agreements prohibit some of the maize and teosinte germplasm collected in Mexico from being stored and distributed by USDA. We will, however, deposit small quantities of seed from all our collections with the CIMMYT germplasm bank in Mexico, and deposit samples of our mapping populations (F2:3 seed) in the USDA-ARS Maize Stock Center at the University of Illinois. Both centers provide public access to seed.

Supplementary Documentation A-2: Management Plan

Communication

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

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

All team members (students, postdocs, PIs) are fluent in English. PIs Ross-Ibarra and Sawers are fluent in Spanish, and PIs Hufford, Flint-Garcia, Runcie, and Sr. Personnel Crosby all have a working proficiency in Spanish. We thus do not expect any language complications for communication within the group.

Data and Code

As described in more detail in Supplementary Document A-1, code will be hosted on Github, and data will be deposited in public repositories. Links to and descriptions of both data and code will be hosted on a central project website on Github.

Outreach

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

Dr. Flint-Garcia will coordinate the annual phenotyping workshop, held each year in Columbia, MO. The workshop will be timed to coincide with data collection at the end of the field season each year. The workshop will be advertised broadly (Corn Breeding Research (managed by Dr. Flint-Garcia), Maize bionet, and evoldir list-servs, the National Association of Plant Breeders Newsletter, etc.). Attendees will be expected to pay their own travel and purchase a handheld device. Surveys will be administered after each workshop in order to gauge the value of the workshop and make improvements for future years.

Dr. Ross-Ibarra will work with Dr. Denise Costich at CIMMYT to coordinate the annual farmer field days in the highland field site in late fall (the 2014 field day was November 21) in years 2-5. We will rely on the CIMMYT infrastructure to advertise the event and invite participants. Multiple project members will attend field days each year to engage in dialogues with the agricultural community, promote the diversity of highland maize and explain the scientific basis of this project.

Research

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

- Graham Coop: 2%
- Sherry Flint-Garcia: 5%
- Matthew Hufford: 10%
- Jeffrey Ross-Ibarra: 17%
- Daniel Runcie: 17%
- Ruairidh Sawers: 15%

Table 1: Summary of proposed timeline of activities showing which team members will be responsible for each objective. Details in text. Team member names are abbreviated: MBH, Matthew Hufford; JRI, Jeffrey Ross-Ibarra; SFG, Sherry Flint-Garcia; GC, Graham Coop; RS, Ruairidh Sawers; DR, Daniel Runcie; KC Kate Crosby

	Year: 1	2	3	4	5
Aim 1.1 QTL mapping	SFG, MBH	SFG, MBH, RS, JRI	SFG, MBH, RS, JRI, KC	SFG, JRI, KC	SFG
Aim 1.2 Admix mapping	GC	MBH, GC, RS	MBH, GC, RS	MBH, GC	MBH, GC
Aim 1.3 Teosinte DH mapping	--	MBH, RS	MBH, RS, GC	MBH, GC, RS	--
Aim 2.1 Population genetics	JRI, GC	JRI, GC, KC	JRI, GC, KC	JRI, GC	JRI, GC
Aim 2.2 Allele-specific expression	MBH, RS, SFG	RS, DR	DR	DR, JRI, RS	DR, JRI, RS
Aim 3.1 Inversion NILs	RS	RS	RS	RS, SFG	RS, SFG
Aim 3.2 Inversion RNA-seq	DR	DR	RS, DR	--	--

Below are details of the responsibilities of each team member during each year of the grant, with initials as shown in Table 1. Although one group will take the lead for writing publications, it is anticipated that several team members and members of their groups will be coauthors on many of these publications.

Year 1

Aim 1.1 SFG will generate seed of F2:3 for S. American cross. MBH will sequence *de novo* parents of both crosses.

Aim 1.2 GC will develop methods for admix mapping.

Aim 2.1 JRI will genotype samples from highland Mexico maize. GC will work on methods for selection in admix populations.

Aim 2.2 SFG, MBH and RS will generate F1 stocks for RNAseq field experiments.

Aim 3.1 RS increase seed for NILs and generate test-cross stocks.

Aim 3.2 DR will grow the NILs in growth chambers for RNAseq.

Year 2

Aim 1.1 RS will grow the mapping populations at each of 3 locations. SFG, RS, and MBH will phenotype populations in field and growth chambers. SFG will genotype F2 plants. JRI will phenotype root chilling. MBH will begin comparative genomic analysis of *de novo* assemblies.

Aim 1.2 MBH will genotype samples. RS and MBH will grow samples at the mid-altitude location. GC will begin data analysis.

Aim 1.3 RS will grow Pioneer teosinte BC2DH lines at each of 3 locations. RS and MBH will phenotype populations.

Aim 2.1 JRI will genotype seed from additional admix populations. JRI, KC and GC will begin data analysis of introgressed highland maize.

Aim 2.2 RS and DR will grow the F1 populations at two Mexican field sites and collect tissue for exon capture and RNAseq. DR will sequence the exomes of the F1 parents.

Aim 3.1 RS will grow and phenotype NIL stocks at three locations.

Aim 3.2 DR will do RNAseq for the growth chamber experiment.

Year 3

Aim 1.1 RS will grow a second replicate of the mapping populations at each of 3 locations. SFG, RS, and MBH will phenotype populations in field. SFG, JRI, and KC will build map and begin QTL analysis. MBH will complete comparative genomic analysis of *de novo* assemblies and write paper.

Aim 1.2 MBH will genotype samples. RS and MBH will grow samples at the mid-altitude location. GC will continue data analysis.

Aim 1.3 RS will grow a second replicate of Pioneer teosinte BC2DH lines at each of 3 locations. RS and MBH will phenotype populations. MBH, GC and RS will begin analysis.

Aim 2.1 JRI, KC and GC will work on data analysis of admixed teosinte and highland Mexico maize.

Aim 2.2 DR will do RNAseq on the collected tissue.

Aim 3.1 RS will grow and phenotype NIL stocks at three locations.

Aim 3.2 DR and RS will write the paper.

Year 4

Aim 1.1 SFG, KC, and JRI will perform QTL analysis

Aim 1.2 GC and MBH will continue analysis.

Aim 1.3 MBH, GC and RS will complete analysis and write paper.

Aim 2.1 JRI and GC will finish data analysis and begin papers for admixed teosinte and highland Mexico maize.

Aim 2.2 DR, JRI and RS will analyze RNAseq data and redo analysis in growth chambers if necessary.

Aim 3.1 SFG and RS will analyze data.

Year 5

Aim 1.1 SFG will finish analysis and write paper.

Aim 1.2 GC and MBH will write paper.

Aim 2.1 JRI, and GC will write papers.

Aim 2.2 DR, JRI and RS will finish analysis and write paper.

Aim 3.1 SFG and RS will write paper.

Supplementary Documentation A-4: Response to Prior Reviews

While the panel rated our previous submission as "highly meritorious", both the panel and individual reviewers raised specific concerns. Below we paraphrase those concerns and our response.

Panel Summary

Intellectual Merit: While the panel lauded the overall project, there was concern that the project was too ambitious given the modest research budget, raising concerns about our ability to carry out all aspects of the research.

Response: While we have expanded the RNA-seq section to address other reviewer concerns (see below), we have added a new CoPI with expertise in transcriptome analysis, 3 new postdoctoral scholars (PIs Runcie, Ross-Ibarra, and Sawers) in UCD and Langebio, increased student support for PI Coop, and modestly increased funds for travel and research supplies to account for the additional personnel. We feel the expanded team is well equipped to tackle the research proposed.

Broader Impacts: Both the panel and reviewers were concerned that approaches to recruit trainees for the phenotyping workshop were not well described.

Response: Workshop announcements will be posted to multiple email lists such as the Corn Breeding Research (managed by Dr. Flint-Garcia), Maize bionet, and evoldir list-servs, the National Association of Plant Breeders Newsletter, etc. in order to attract breeding and genetics researchers from as many plant communities as possible. Dr. Flint-Garcia has already been involved in the recruitment of participants for the 2015 Panzea GBS workshop to be held in Columbia, MO.

Reviewers

Population development Creating admixture populations and NILs will take many years.

Response: We have made substantial progress in population development: all test materials will be available by the end of Year 1 and ready for evaluation from 2017 onwards. Admixture seed is already collected directly from wild populations, allowing us to take advantage of the mapping potential of this material without further development.

Phenotyping: Additional detail is needed on how phenotypic traits would be measured.

Response: We have added additional detail to the phenotyping, especially on anthocyanin and macrohairs.

Teosinte admixture: Will results from teosinte be relevant to securing crop yields? This aim is not clearly linked to overall project goals.

Response: Selection in both maize and teosinte has resulted in adaptation to similar highland conditions. Because of the long history of recombination, admixture mapping in teosinte will provide much greater resolution for mapping loci than is possible in our synthetic populations. We have more clearly explained these advantages and better integrated this aim within the larger scope of the project. We have also added Aim 1.3 in which we explicitly test the effects of teosinte alleles in a maize background to evaluate the relevance of teosinte alleles for improving maize.

RNA-seq analyses: Additional detail is needed on how tissues and developmental stages will be chosen for RNAseq analysis, and the criteria used to identify adaptation-specific genes.

Response: We have substantially revised this section. In Aim 3.2, we will match tissues to those assayed in the Maize gene atlas (Sekhon et al., 2011) for comparison to other studies in the common B73 background. To identify gene expression traits that are important for adaptation, we have added Aim 2.2 where we will use population genetic techniques (Qst and the sign test) to identify genes and pathways that show statistically robust signatures of adaptation.

Potential Pitfalls: Possible pitfalls should be discussed given the ambitious scope of the proposal.

Response: We have included brief paragraphs at the end of each aim discussing potential challenges that could be encountered and strategies for ensuring success of the project.

Supplementary Documentation A-5: Plans for Undergraduate and Graduate Student Mentoring

Undergraduate Students

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

Graduate Students

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