Project Summary

Maximum of 1 page

Intellectual Merit

Broader Impacts

Project Description

Introduction

Due to their sessile nature, plants are often adapted through natural selection to their local environments (**?**). Understanding the genetic basis of how plants adapt to local conditions – how many loci are involved, the effect sizes of adaptive loci, 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 population growth (**?**). These pressures will require adaptation of crops and wild plants to changing local conditions and cultivation of crops in new locales.

Agricultural species represent promising systems for ongoing research on local adaptation. Most crops were domesticated in narrow geographic centers but encountered and adapted to a wide range of novel environments as agriculture expanded across the globe (**?**). In many instances, traits important for local adaptation have already been identified in crops (**??**). These systems therefore represent compelling opportunities for investigating the genetic architecture of local adaptation. Moreover, insights gained regarding adaptive loci can feed back into modern crop improvement, yielding valuable benefits in the face of rapid environmental change.

Here we propose to use maize adaptation (*Zea mays* ssp. *mays*) to high elevation environments as a model for understanding the genetic basis of local adaptation in plants. Maize was domesticated in the lowlands of southwest Mexico from the narrowly distributed teosinte *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*; **?**). Since domestication, maize has spread worldwide: analysis of cultivation area data indicates maize has the greatest global geographic breadth of 16 staple crops (Figure 1) and is now cultivated on six continents, ranging from southern Chile to Canada and from sea level to well over 3000m in altitude (**?**). Maize has adapted to high elevation in several isolated geographic regions. Moreover, the subspecies of teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) is found only in highland environments, having adapted to highland environments thousands of years prior to maize domestication (**??**). Maize and teosinte thus form an ideal system in which multiple replicated evolutionary experiments will allow us to dissect the genetic architecture of highland adaptation as a model for understanding plant local adaptation.

Aims

We will investigate the genetic basis of highland adaptation in maize by achieving three aims. Table 1 shows the proposed timing of each aim and which CoPIs will be responsible.

1. **Dissect the genetic architecture of highland traits**
2. **Investigate population genetic signatures of highland adaptation**
3. **Characterize functional variation at adaptive quantitative trait loci**

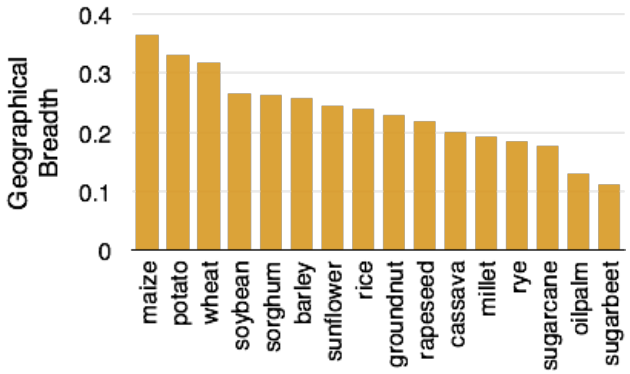


Figure 1: Geographic breadth of the world's 16 staple crops, expressed in percent of land surface area in which each crop is cultivated. Data are from **?**.

h

Year		2015	2016	2017	2018	2019
Aim QTL mapping	1.1	SFG, MBH, RS, JRI	SFG, MBH, RS	SFG, MBH, RS, JRI	–	–
Aim Admix mapping	1.2	MBH	MBH, GC	MBH, GC	–	–
Aim Population genetics	2	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC	JRI, GC	–
Aim Functional analyses	3.2	RS	RS, SFG	RS, SFG	RS, SFG	RS, SFG
Aim RNA-seq	3.3	–	–	JRI, ACJ	JRI, ACJ	JRI, ACJ

Table 1: Proposed timeline of activities, showing which CoPI or Sr. Personnel will be responsible for each objective.

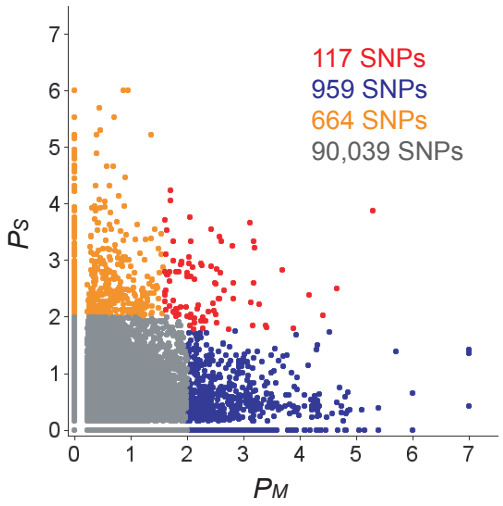
Rationale and Significance

While the genetic basis of local adaptation is generally not well understood, the declining cost of genotyping has enabled a handful of genome-wide studies across populations of model species. For example, **?** 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 across hundreds of accessions of *Medicago truncatula* identified candidate loci for local adaptation and found them to be predictive of growth rate under temperature and soil moisture treatments (**?**). Finally, 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* (**?**) – an enrichment of regulatory variants among loci showing evidence of selection (**?**). While much of local adaptation may involve complex quantitative traits (**?**), the genes important for local adaptation are not necessarily those identified in mapping studies in other populations. In maize, for example, although genome-wide association in the NAM panel suggests that flowering time is largely controlled by many loci of small effect (**?**), adaptive change in flowering time across latitudes has involved loci of large effect on photoperiod (**?**). Traits under selection may also show distinctive patterns of effect size – while several QTL have been shown to be pleiotropic for both ear and tassel traits, the effect sizes of these QTL on ear morphology, which underwent recent strong selection during domestication, are larger than their effect on the tassel (**?**). Though initial genomic studies are beginning to yield valuable insights regarding local adaptation, clearly much remains to be discovered.

Maize and teosinte are an excellent system in which to study local, specifically highland, adaptation. Following domestication, maize spread to the highlands of the Mexican Central Plateau, a migration across more than 1000m of increasing elevation. Colonization of the highlands required adaptation to a number of novel abiotic conditions, including gradients of temperature, precipitation, and elevation. Highland landraces have distinct morphologies (e.g., highly pigmented and hairy leaves and stems) that are believed to confer adaptation to this cooler region (**?**). Our previous genetic analyses (**?**) show that maize has independently adapted to highland environments multiple times, including the southwest US and the Andes of South America, where landraces (i.e., local farmer varieties) are commonly grown above 3000m. Multiple independent instances of highland adaptation in maize and teosinte provide replicated evolutionary experiments, providing power to identify and validate both widespread and population-specific candidate loci for highland adaptation.

Finally, study of the genetic architecture of maize adaptation will provide both basic evolutionary insight and 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 (**?**). **?** further determined that future temperature increases will likely decrease yield across 65% of African maize-growing regions, while all of Africa will see diminished maize yield if increased temperature is accompanied by drought. An understanding of how maize has adapted to challenging environmental conditions in the past will help breeders mitigate yield loss due to future changes.

Figure 2: 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.



Preliminary Results

Preliminary work from project members positions us to make excellent progress on our proposed aims. Co-PIs Ross-Ibarra and Hufford have worked extensively on the population genetics of highland adaptation. **?** explored local adaptation in *parviglumis* and *mexicana* populations, finding a large number of loci showing association with altitude and evidence of selection, as well as highlighting the potential importance of regulatory variants and large inversion polymorphisms. This study identified a putatively adaptive inversion on chromosome four that distinguishes the lowland *parviglumis* from the highland *mexicana* and coincides with a quantitative trait locus associated with traits linked to highland adaptation (**?**). This *Inv4m* inversion is the subject of our functional characterization in Aim 3. **?** also identified populations of *parviglumis* showing extensive admixture with the highland *mexicana* which are the subject of proposed analysis in **??**. **?** identified genomic regions in highland maize that have introgressed from *mexicana*. They showed that plants with *mexicana* alleles showed *mexicana* phenotypes and superior growth under cold conditions, suggesting an adaptive role for introgression and motivating our population genetic analyses in Aim 2. Finally, recent analyses of selection in genotyping data from a wide collection of landraces from the highlands of Mexico and South America finds little overlap in the genes important for adaptation (Takuno *et al.* In Prep; Figure 2), motivating the QTL analysis in **??**.

Co-PI Flint-Garcia and Sr. Personnel Sawers have made important progress on the development of populations for the project. For Aim 1.1 our Mexican cross is already at the F2 generation, and one potential South American cross is now at the F1 generation (Table 2). Back-crosses of the reference genome inbred B73 to a highland Mexican landrace Palomero Toloqueño have been made and selfed to generate a BC1S1 population that will be further developed in Aim 3.2 to dissect the function of the *Inv4m* polymorphism. Both SSRs and SNPs that identify introgressed *mexicana* alleles at *Inv4m* have also been developed.

Co-PI Coop has developed analytical approaches for understanding local adaptation, including methods that allow genome-wide association with environmental variables (**??**), detection of selection in introgressed populations (**?**), and powerful approaches to identify phenotypic selection on quantitative traits **?**. His group is currently working on methods for mapping and studying adaptation in admixed populations.

Table 2: Parental lines for QTL

Population	Parent	Origin (masl)	Status
Mexico	Zapalote Chico	Oaxaca (46)	F2
	Palomero de Jalisco	Jalisco (2520)	
S. America	Araguito	Venezuela (183)	F1
	Sal Prieta	Ecuador (2948)	

Specific Objectives

Aim 1 Genetic architecture of highland traits

One of the primary goals of this proposal is to determine the genetic architecture of highland adaptation. Ultimately, this knowledge will be useful for determining the genes underlying these loci (Aim 3) and the pathways involved in adaptation (Aim 2). These loci can also be used in maize improvement via marker assisted selection. In this aim we wish to determine how many genomic regions control adaptive phenotypes, where these regions are located, and the distribution of allelic effects at these loci. We first perform comparative QTL analysis in two highland x lowland maize crosses (Aim 1.1), then take advantage of historical recombination and greater resolution to map loci in an admixed population of highland and lowland teosinte (Aim 1.2).

Questions

- What is the genetic architecture of highland adaptation?
- How much of the genetic architecture is shared between Mexico and South America?
- How much of the genetic architecture is shared between maize and teosinte?

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 2). We make use of specially-inbred landrace lines created by John Doebley (U. Wisconsin) and Seth Murray (Texas A&M), thus simplifying downstream applications and allowing replication of alleles in our functional studies (see Aim 3).

We will self-pollinate F2 plants to create 500 F2:3 families from each population. DNA will be extracted from each of the parents of the F2 plants and sequenced to 20-30X depth on two lanes of Illumina (150bp paired-end reads on a HiSeq 2500 at the UC Davis Genome Center), providing genome-scale SNP data similar to our previous work (HapMap.v2; **?**). F2 plants will be genotyped using genotyping-by-sequencing (GBS; **?**) and run through the standard maize GBS pipeline (**?**) resulting in approximately ∼1M SNPs, allowing straightforward imputation of their full-genome sequence. The genetic map will be created using standard methods (e.g. **?**).

Table 3: Common garden locations

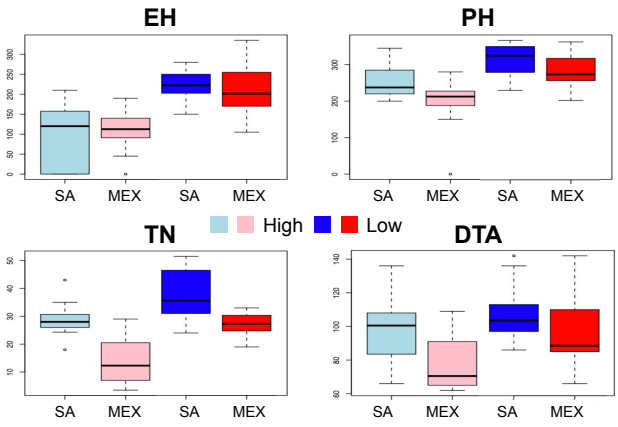
Field Sites	**Lat/Lon**	**Elevation (m)**	**Min/Mean/Max °C**	**Precip (mm)**
Valle de Banderas, Nayarit	20.8, -105.2	54	15.3/25.8/33.7	1184
Irapuato, Guanajuato	20.7, -101.3	1729.0	7.3/20.2/31.7	693
Amealco, Querétaro	19.5, -99.1	2240.0	2.3/15.6/27.0	626
Columbia, Missouri	28.9, -92.2	266.1	-17.8/36.0/40.5	914

Populations will be phenotyped at 3 field locations, including one lowland site (Valle de Banderas in Mexico), one highland site (Irapuato or Queretaro, Mexico), and one temperate site near Columbia, Missouri (Table 3). At each field location, best local practices will be used including fertilizers and pest and weed control.

At each site, the experiment will consist of two replicates in which the 500 entries will be arranged in an augmented alpha lattice design. Parental checks will be included to control for field variation. We will collect a number of phenotypes (Figure 3) using our in-house, barcode-based data collection program. Germination assays in controlled conditions will be conducted 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).

Raw data from each plot will be analyzed using mixed-models incorporating replications and environments. Data will be analyzed across environments to determine whether location (elevation) affects the various phenotypes. 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 **?**). Several iterations of QTL analysis will be conducted: on individual traits, individual traits adjusted for covariates such as flowering time, and multiple traits simultaneously. QTL profiles will be compared across populations (Mexico vs South America) and field sites (elevation) to determine differences in how elevation affects putatively adaptive traits. Comparison of the genetic architecture among traits will inform us of the lability of these traits and their amenability to selection via breeding. Finally, the contrast of each Mexican location to the Missouri location will account for daylength differences and agronomic value in the Midwest.

The expected outcomes of this objective will be 1) A map of QTL underlying phenotypic differences between highland and lowland maize in Mexico and South America, detailing the effect size of each QTL and differences between crosses, and 2) Estimates of fitness differences (PH, BM, SM, and FK (Figure 3)) of highland and lowland plants, as well as F2 with various combinations of QTL, in both environments.



Trait	Phenotype
MH	leaf sheath macrohairs
DTS	days to silking
DTA	days to anthesis
PH	plant height
BM	total plant biomass
EH	ear height
EN	ear number
FK	fifty kernel weight
TN	tassel number
TBN	tassel branch number
TL	tassel length
SM	total kernel mass
RC	root chilling response
GDP	germination depth
SC	stem anthocyanin content
GDP	germination depth
GDT	germination temperature

Figure 3: 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)

Aim 1.2 Admixture mapping in a teosinte hybrid zone

While *mexicana* and *parviglumis* are largely allopatric, the subspecies overlap in two regions of Mexico, eastern Jalisco state and the eastern Balsas River Basin (**?**), and a number of hybrid populations have been documented in these regions (**?**). We have previously documented near equal proportions of ancestry from the two subspecies in one of these populations near the town of Ahuacatitlan in the eastern Balsas (**?**). Growth chamber experiments also suggest plants in this population have higher fitness in cold conditions than other *parviglumis* populations. Moreover, the relatively short length of haplotypes Ahuacatitlan shares with other populations suggests that there has been extensive recombination since the initial admixture event, providing an ideal population for high-resolution admixture mapping of *mexicana* highland adaptation traits.

In November of year one of the project we will travel to Ahuacatitlan and collect seed from 500 individuals drawn randomly from the population. Seed samples will be transported to Langebio in Irapuato, Mexico for cold storage. A single seed per individual (500 total) will be germinated on filter paper and transplanted into our two Mexican field sites (Table 3). Phenotypes detailed in Figure 3 will be collected for admixture mapping. Many of these traits are known to differ considerably between *parviglumis* and *mexicana* (**?**). Leaflet samples will be collected from plants in the field at the seven-leaf stage, and extracted DNA will be genotyped using GBS. Several computational methods for admixture mapping have already been developed (**?**), but we will augment these with methods currently under development by Co-PI Coop. Current methods are not well suited

to admixture mapping when there are differentially related individuals in the sample, and when natural selection may have systematically distorted admixture at some loci. In natural 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 plant 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 Ahuacatitlan population that incorporate this non-independence into admixture association tests, while accounting for uncertainly in admixture calls along the genome.

Aim 2 Adaptive value of highland alleles

In aim Aim 1 we will map loci corresponding to traits differing between highland and lowland maize and teosinte. In this section we will test the adaptive significance of these QTL in three sets of populations: a worldwide sampling of highland maize, admixed maize populations from Mexico, and admixed populations of *mexicana* and *parvgilumis*.

Questions

- Are highland QTL/loci widespread in highland climes?
- Are loci controlling phenotypic differences between highland and lowland populations adaptive?
- Does natural selection favor introgression from adapted populations?

Aim 2.1 Populations

Global highland maize We will assemble a panel of 500 global highland maize accessions from the public repositories of USDA-ARS and CIMMYT. Our panel will include accessions from highland regions in Guatemala, Central America, the southwestern United States, Ethiopia and other parts of East Africa, and South and East Asia (**?**). This population allows us to test for similarities in the genetic architecture and adaptation to highland environments more broadly across all maize. This population will also highlight the diversity of highland adaptation alleles that can be drawn upon during maize breeding. Data from this worldwide sampling will be augmented by whole-genome-sequencing currently underway in the Hufford and Ross-Ibarra lab of a small set of highland maize lines from Guatemala and the SW U.S.

Highland Mexican landraces **?** documented extensive introgression between *mexicana* teosinte and highland maize landraces, demonstrating an overlap with teosinte QTL for macrohairs and stem pigmentation (**?**). Because of the relatively low-density genotyping used, however, we were limited to identifying large regions of ancient introgression present in most populations. We were also unable to investigate evidence of selection for any of the introgressed regions. Here we propose to resuse 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 colonization of the highlands. Correlations between

genetic differentiation and recombination in these populations will also allow us to investigate selection against introgression (**?**), quantifying the "linkage drag" associated with introgression of potentially beneficial adaptive alleles.

Admixed teosinte populations Here, we will complement the Ahuacatitlan population from Aim 1.2 by sampling four additional admixed populations that have been identified using small-scale SNP data from across the range of each taxon (**?**). We will revisit each of these populations to sample seed, collecting 50 individuals per population. Because these admixture events appear to be ancient (**?**), 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 our this prediction (Figure 4).

Aim 2.2 Analyses

Samples from all populations will be genotyped using GBS. Teosinte populations will be genotyped at higher coverage (48 plex) to decrease genotype error at heterozygous sites. In each population we will apply population genetic approaches which utilize evidence from the site frequency spectrum (**?**) and haplotype structure (**?**) to identify loci under selection. In teosinte populations we will use both haplotype (**?**) and heterozygosity-based (**?**) methods to identify introgressed segments in individual populations. Loci showing evidence of introgression and selection will be compared with those underlying QTL in maize and teosinte populations from Aim 1. Quantitative genetic theory suggests, however, that adaptive phenotypic change can frequently occur without strong selection on individual loci (**?**). To search for evidence of selection on adaptive phenotypes, we will employ recently developed methods from Co-PI Coop (**?**) that provide a powerful statistical framework to identify coordinated shifts in allele frequencies at causative QTL (from **??**) to look for weak selection on alleles underlying highly quantitative traits. These methods will allow us to specifically ask which phenotypes show evidence of selection in individual 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 (**?**).

Expected outcomes: of this objective will be 1) identification of adaptive loci in teosinte and cultivated maize populations, 2) evidence for repeated evolution and comparisons of the genetic architecture of adaptation 3) evidence for selection on individual phenotypic traits, 4) quantification of the potential "linkage drag" or evidence against introgression across other regions of the genome.

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 begin to investigate the functional genetic basis of adaptive regions. We will first study the phenotypic effects of alleles at *Inv4m*, an inversion corresponding to a chromosome 4 QTL that has introgressed into highland maize (Aim 3.2). Then we will use RNA sequencing

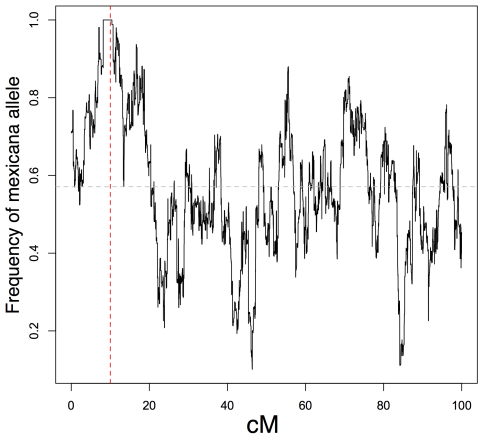


Figure 4: Analysis of 100 generations of simulated admixture between *mexicana* and *parviglumis* across a 100cM chromosome. A beneficial *mexicana* allele with selection strength $s = 0.1$ is introgressed at position 10cM (red vertical line), showing that deviation from background variation in ancestry (horizontal gray line) can be used to detect selection in admixed populations.

data to find differences in expression, plasticity, and identify potential candidate loci within QTL (Aim 3.3).

Aim 3.1 Questions

- What are the phenotypic consequences of introgressing a single adaptive QTL?
- What are the functional differences of different alleles of an adaptive QTL?
- How do maize and teosinte differ in expression response to highland and lowland environments?
- Can RNA-seq help refine QTL to identify candidate genes?

Aim 3.2 Functional evaluation of *Inv4m*

In this objective we propose to functionally characterize the genetics of *Inv4m*, an inversion polymorphism located on chromosome 4 (169-180Mb). Our previous work (**??**) indicates that this region is supported by a robust signature of introgression from *mexicana*, shows broad distribution among highland races, and overlaps with a QTL identified in a *parviglumis* x *mexicana* cross (**?**) associated with leaf pigmentation and pubescence.

We will generate heterogeneous inbred families (HIFs; **?**) from a cross of the highland landrace Palomero Toluqueño (PT) to the reference genome inbred B73. PT is a popcorn originating from the highland valleys of central Mexico that is considered basal to the Mexican highland landrace radiation (**?**); it also exhibits the highest level of *mexicana* introgression among characterized material (**?**). Furthermore, inspection of the PT genome sequence (**?**) shows that PT carries the *mexicana* allele at *Inv4m* (**?**). We will screen an existing collection of ∼150 B73 x PT BC1S3 families (three generations of selfing after 1 generation of back-cross) to identify HIFs segregating for B73 and PT haplotypes using microsatellite markers that distinguish B73 and PT alleles in this region. HIFs will be self-pollinated to generate pairs of near-isogenic lines (NILs) homozygous for the B73 or PT haplotype. While different in the candidate region, NIL pairs will share a common genetic background outside this region, including a sizable (25%) contribution of PT, capturing potential epistatic effects important to expression of the candidate phenotype. NILs will be genotyped by GBS both to confirm the extent of introgression around *Inv4m* and to characterize this shared background. A total of 6 HIF derived NIL pairs (i.e. 12 lines), will be characterized

in our three field sites (Table 3) and evaluated for phenotypes described in Figure 3. In each site, we will plant 3 replicate rows of our NILs and the B73 and PT parents. Data will be analyzed broadly as described in **??**, both treating the introgression region as a single block, or considering individual markers.

While the probable lack of recombination between haplotypes of *Inv4m* (**?**) facilitates generation of test materials and assessment of the region as a block, it nonetheless hampers downstream efforts to dissect phenotypic effects and fine map the loci involved. To address this issue, we will also generate a series of NILs by marker-assisted recurrent backcross to B73 using a collection of seven diverse donor varieties: 3 lowland haplotypes represented by the 2 lowland parents of our mapping populations (Table 2) and an inbred *parviglumis*; 4 highland haplotypes represented by the 2 highland parents of our mapping populations (Table 2), an inbred *mexicana* and the Palomero Toluqueño haplotype segregating in our HIFs. All 3 highland maize varieties are predicted to carry the *mexicana* inverted haplotype at the *Inv4m* region. Each of these parents either have resequenced genomes (**??**) or will be sequenced as part of this project in Aim 1.1. It is anticipated that this material will be phenotyped selectively in light of initial results generated by analysis of HIFs in the early part of the project.

Expected outcomes: 1) Estimation of phenotypic effects of the *Inv4m* candidate region among lowland and highland teosinte and landrace maize; 2) identification of differences in phenotypic effect among NIL pairs, indicative of background dependent epistatic interaction among genes 3) Dissection of the highland haplotype on the basis of phenotypic variation among NILs carrying the inverted form; 4) Generation and identification of material suitable for future fine mapping through crossing of genetically/functionally divergent inversions from NILs.

Aim 3.3 Gene expression

In this objective we will use RNA sequencing to evaluate expression differences among lines, across environments, and among loci. We will first grow the 8 inbred lines which serve as parents of our allelic series analysis in Aim 3.2 in our highland and lowland field site (Table 3) to identify genes responsive to these environments. From each inbred we will sample leaf and root tissue from three plants at each of two time stages (seedling and flowering adult). Tissue will be flash frozen and sent to UC Davis for extraction and sequencing (multiplexed 12 individuals per lane of an Illumina HiSeq 2500) at the UC Davis Genome Center. Each individual will be barcoded, providing 3 biological replicates for each tissue/time/environment combination. We will use edgeR (**?**) to assess differences in expression across environments. We will then compare differentially expressed (DE) genes to QTL from Aim 1, loci showing selection identified in Aim 2, and introgressed regions showing phenotypic differences in Aim 3.2. These results will help narrow down potential candidate genes in QTL and serve as functional validation of loci showing population genetic evidence of selection. The data will also allow investigation of the relationship between phenotypic plasticity and adaptive change (c.f. **?**) via comparison of DE genes among environments for a single inbred to differences in DE genes among inbreds.

Our second approach will be a targeted analysis of transcriptomic changes in the *Inv4m* NIL lines from Aim 3.2. Using NILs generated from each of the same 7 inbred donors (alongside an additional replicate of B73), we will evaluate shoot tissues of three plants sampled at seedling and flowering stage for each of the two genotypes (homozygous B73, homozygous donor). Samples will be extracted and sequenced as described above. These analyses will allow us to refine potential

candidate loci within introgressed segments of our NILs, moving us closer to a functional characterization of observed phenotypic differences. Whole-transcriptome comparison of the NILs to the donor lines will also permit differentiation between cis and trans regulation of expression within the *Inv4m* region, and analysis of co-expression networks (c.f. **?**) will highlight the effects of introgressed genes on expression patterns in the rest of the genome, enabling us to begin to dissect the genetic pathways involved in adaptive highland traits.

Expected outcomes: 1) Identification of candidate genes showing plastic differential expression within lines across environments 2) Identification of candidate genes showing differential expression among lines from different environments 3) Detailed information on the effects of introgressed segments on genome-wide expression.

Broader Impacts

Exchange Program

We propose an international student exchange program between the PIs in the U.S. and Senior Personnel at LANGEBIO in Mexico. Over the course of the grant, we propose to fund 10 graduate or undergraduate students for 3-month research internships in one of the collaborating laboratories. Students involved will participate in research projects directly relating to the research focus of the grant, including developing mapping populations, mapping traits, population genetic analysis, or analysis of next-generation data. The expectation is that such research will often lead to co-authorship on publications. Students will be asked to give two presentations, one to the host lab upon arrival, talking about the lab/university they came from and research there, and another to their 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 similar exchange program (NSF 0922703). Over the last four years, his lab has hosted 6 Mexican students who have worked on various computational aspects of centromere evolution. Two of those students have earned authorship on a paper to be submitted later this year and one has gone on to a PhD program in the U.S.

Our goal is to involve students directly in research while at the same time fostering intercultural exchange and promoting future international research opportunities. It is particularly appropriate for the study of maize, a crop with significant cultural and economic impact in both Mexico and the U.S. Participating Mexican students will learn new analytical methods – especially computational management of large datasets – that can be introduced to their respective laboratories and peers. American exchange students will similarly benefit from experience with large field experiments and efforts to functionally characterize individual loci. The hope is that Mexican undergraduate students involved may be recruited to graduate programs in the U.S., ideally to work in the lab of one of the PIs, and that American undergraduate students will be exposed to international opportunities for research, graduate education, and collaboration.

Phenotyping workshop

The USDA-ARS group in Columbia 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 during each year of the grant. Through this workshop, Dr. Flint-Garcia's state-of-the-art system will be transferred to other research institutions to aid in large-scale data collection. The phenotyping workshop will include topics on Experimental Design, setting up the FieldBook database, and Data Collection. Experimental design topics include understanding where variation comes from, how to control for environmental/field variability and experimental error; heritability and repeatability. The need for consistent data collection and high-throughput will be emphasized. FieldBook database setup topics include setting up Palm handheld users, locations, traits, projects, assigning plots to projects, assigning traits and measurements to projects, generating barcoded plant tags, and loading the program and trait groups to the Palm to prepare for data collection. Topics to be covered in Data Collection include data collection for specific traits related to local adaptation of interest to our group, synchronizing data from the palm with the desktop/laptop database, managing data conflicts between the palm and the database, running reports, and exporting data. This proposal will provide travel support for instructors. The workshop will be free but participants will be expected to purchase their own Palm handheld and pay for their own travel. The workshop will be held each year in late summer so that the participants can gain hands-on experience in data collection in the corn field.

Software

A good understanding of population and quantitative genetics is key to a students understanding of genetics and evolution, but these subjects are often conceptually quite difficult. An understanding of genetic variation and its phenotypic effects is also an increasingly important part of being an informed citizen, due to the rise of personal genomics and genomic medicine (e.g. **?**). The large amount of population genetic and association data being generated offers a superb chance to motivate these subjects using real data. We will develop undergraduate teaching modules in population and quantitative genetics using data from this project. These modules will be tested and integrated into large undergraduate teaching courses (introductory 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, e.g. genome-scale demonstrations of Hardy Weinberg Equilibrium (HWE) using human HapMap data. Such demonstrations underscore the usefulness of basic population genetics in describing real world patterns, and begin to expose students to the wealth of genomics data being collected. Other examples will include: using association data from our admixed populations 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 statistical program R, to ensure that they are easily used, modified, and distributed, and to expose students to programming in biology. The modules will be designed so that they can be tailored for use at a variety of levels from teaching basic concepts to large undergraduate classes to providing the raw data for programming exercises for upper division courses.

The modules will be publicly distributed via Github (see Data Management Plan) in a fully open manner. The use of github will allow others to modify and extend the modules and to share and track these modifications.

Germplasm resources

This project will generate multiple germplasm resources. Seed from the F2 parents will allow use of this mapping population to study additional phenotypes of interest (e.g. root morphology and growth). Seed from our NIL populations will allow investigation of genome-wide introgressions from a variety of exotic lines. Such material could be of interest to the Germplasm Enhancement of Maize (http://www.public.iastate.edu/ usda-gem/) project as well as to public and private breeders both in the US and abroad. In Mexico, for example, the highland niche represents a key target market for an emerging private sector of small breeding companies established following deregulation in 1990s, and is one of the areas most vulnerable to climate change (**?**). While highland adapted hybrids are available, these are largely derived from lowland sub-tropical material with little or no contribution of the highland landraces and the germplasm developed here could be an important contribution to furthering such programs. Finally, seed from our collections of teosinte will enhance the sampling of these subspecies and provide additional diversity not currently present in germplasm banks. Seed from our mapping populations will be deposited in the USDA-ARS Maize Stock Center at the University of Illinois, and backups will be kept at Iowa State and Missouri.

Results From Prior NSF Support

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

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

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

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

Publications ??????

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

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

Intellectual merit Centromeres are regions of the genome that organize and regulate chromosome movement, yet the biology of centromeres remains poorly understood. Co-PI Ross-Ibarra's group has focused in particular on the evolutionary genetics of centromeres. This work has demonstrated the remarkable evolutionary lability of centromere tandem repeats, but has shown that there is little evidence in maize for coevolution between centromere sequence and kinetochore proteins. Ongoing

work from the Ross-Ibarra lab seeks to characterize kinetochore proteins, assess the phylogenetic evidence for longer-term coevolution, and understand patterns of centromere and genome size variation in natural populations.
Broader impacts Co-PI Ross-Ibarra has established and currently runs an international student exchange program as part of this grant. Data and result of this project have been disseminated via publications and presentations as well as deposited in the maize genetics community database `www.maizegdb.org`. Former trainees on the grant include Dr. Matthew Hufford (Co-PI on the current grant).
Publications ?????????

Coop: #1262645: Collaborative Research: ABI Innovation: Visualization And Statistics For Spatial Population Genomic Analysis.

$314,260, with an effective date of 05/01/13. Award Duration: 36 months.
Intellectual merit We are developing a set of spatial statistics methods based on Gaussian random fields for the analysis of geographic population genomics data. The first method based on this approach has just been published, allowing a sound statistical framework to distinguish the effects of geographic and ecological distance on genetic isolation.
Broader impacts The R package of the software has been released online, and has already been used by many molecular ecologists.
Publications ?

Flint-Garcia: #0820619: Genetic Architecture of Maize and Teosinte

$ 9,823,000. 3/1/2009-2/28/2013. PI Edward Buckler, co-PIs J. Doebley, T. Fulton, S. Flint-Garcia, J. Holland, S. Kresovich, M. McMullen, Qi Sun.
Intellectual merit This project extends over more than a decade, and has pioneered the characterization of population genetic and evolutionary parameters of maize diversity, developed resources to connect this genetic diversity to phenotype through both association and joint linkage-association mapping, conducted fine scale analysis of domestication and agronomic QTL, and recently expanded to whole-genome analysis of diversity, evolution, and phenotype. Overall, the maize diversity project has developed a wide range of approaches and broadened understanding of the maize genome, evolution and adaptation, genetic mapping, and the agricultural improvement of maize. The project successfully released and analyzed the maize Nested Association Mapping (NAM) population, collaborated on making first and second generation haplotype maps for maize, resolved domestication traits, developed a range of novel statistical approaches for association mapping, and dissected complex traits such as flowering time, kernel composition, disease resistance, height, and inflorescence and leaf morphology.
Broader impacts The outreach program included a traveling science museum exhibit on maize diversity, evolution and genetics (seen by at least 300,000 people at five venues to date, including the famous Corn Palace in South Dakota), online Teacher Friendly Guide to the Evolution of Maize, seven Genotyping-By-Sequencing (GBS) workshops (held at primarily at Cornell but has also been held in Kenya), and training of postdocs, graduate students and undergraduates, the vast majority of which have continued in scientific careers. Former trainees on this grant include Dr. Flint-Garcia and Dr. Ross-Ibarra (PIs of the current grant), only their publications are shown below.
Publications ??????????????????????

Budget Justification

Personnel

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

Other Personnel

Graduate students Funds are requested to support two graduate students each for 6 months during the academic year for each year of the project. At UC Davis, the current pay rate for doctoral students at 50% FTE is $27,319 during the academic year. Included is the estimated annual salary increase of 3%. The two students will be working on analysis of GBS data in the introgression and admix population genetic sections of AIM2, and will likely help with QTL analysis and sequencing in Aim1, and potentially RNA-seq analysis in Aim 3.

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

Fringe Benefits

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

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

Equipment

No equipment funds are requested.

Travel

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

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

Participant Support

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

Other Direct Costs

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

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

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

RNA sequencing : In total, RNA sequencing will be performed on 192 individuals (8 inbreds x 2 stages x 2 tissues x 2 environments x 3 replicates + 8 NILs x 2 genotypes x 2 stages x 2 environs x 3 replicates). Cost to prepare RNA libraries in our lab are approximately $100 per library, and sequencing costs for single-end 50bp reads at the UCD Genome Center are approximately $1,000 per lane. Multiplexing 12 barcodes per lane, this comes out to 32 lanes of sequence and a total cost of $70,400.

Field fees: Fees for the field experiments in our highland and lowland field sites 3 are approximately $60,000 the first three years of the experiment to allow development of the mapping populations and two replicates of the phenotyping. These fees include land rental and basic management (planting, watering, weeding, fertilizing), as well as station fees to hire manual labor for phenotyping. These fees decrease to $10,000 in the last two years of the proposal as subsequent field experiments including evaluation of NILs and RNA-seq lines, will be considerably smaller. Field fees total $200,000 across the five years of the grant.

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

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

Total Direct Costs

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

Indirect Costs

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

Facilities, Equipment, and Other Resources

Facilities, Equipment & Other Resources

UC Davis

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

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

Iowa State

Project components completed in the Hufford Laboratory will include mapping population development, DNA isolation and PCR, and population genetic analysis of genotyping data. Population development will be carried out in field space available at the Curtiss Farm of Iowa State University (ISU). This facility is equipped with irrigation, tractors, tillage equipment, planters, and combines. Seed processing and cold storage facilities are also available on the ISU campus. The Hufford Laboratory has all equipment necessary for DNA isolation and PCR including centrifuges, thermal cyclers, an ultra-low freezer, water baths, a pH meter, balances, and an electrophoresis system. A gel imaging system and a NanoDrop spectrophotometer for DNA quantification are

accessible through the Center for Plant Responses to Environmental Stresses at ISU. The DNA Facility at ISU provides access to cutting-edge genomic technology including HiSeq and MiSeq Illumina sequencing and library preparation for both paired-end and mate-pair approaches. Data analyses will be carried out using the High Performance Computing clusters available at ISU. Dr. Hufford currently has access to the Lightning3 cluster which has a mix of Opteron based servers, consisting of 18 SuperMicro servers with core counts ranging from 32 to 64 and 256 to 512 GB of memory.

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

LANGEBIO

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

Supplementary Documentation

Data Management Plan

Data Types

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

Data Access, Sharing

Sequence data of the parental lines will be deposited to NCBI sequence read archive (SRA) along with passport information on each parent.

Phenotypic data and genotypes from sequencing and GBS will be uploaded to Figshare, where it can be associated with other data (publications, links to germplasm, SRA, code). Data will be grouped into projects, and each project is associated with a unique digital object identifier (DOI). PIs 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.

Analytical software and code from this project will be hosted on github, a version-controlled public git repository. Upon submission of papers all code will be made publicly available. PIs 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.

All appropriate metadata including plant ID, data collector, sequence run, field location, etc. will be associated with genotype and phenotype data deposited to figshare.

Presentations and teaching resources from our field workshop will also be made publicly available via the Slideshare website.

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

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.

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. PI Ross-Ibarra will maintain a backup of all raw genotyping, sequence, and phenotyping data. His lab maintains a DROBO distributed backup server (currently ¿8Tb of free space) which is robust to single disk failure. All analytical code will be hosted on github, which maintains version-controlled backups, as private repositories until release.

Seed will be maintained in climate-controlled conditions at Iowa State. International agreements prohibit some of the maize and teosinte germplasm collected from being stored by USDA. We will

deposit small quantities of seed from all our collections with the CIMMYT germplasm bank in Mexico, and deposit samples of our mapping populations in the USDA-ARS Maize Stock Center at the University of Illinois. Both centers provide public access to seed.

Postdoctoral Researcher Mentoring Plan:

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

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

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

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

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