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 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 (Savolainen et al., 2013) which 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 have encountered and adapted to a wide range of novel environments as agriculture expanded across the globe (Gepts, 2014). In many instances, traits important for local adaptation have already been identified in crops (Gepts, 2014; Purugganan and Fuller, 2009). 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 the adaptation of maize (*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*; Matsuoka et al., 2002). 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 (Tenaillon and Charcosset, 2011). As it spread across the globe, maize has independently adapted to high elevation in several isolated geographic regions (van Heerwaarden et al., 2011). Moreover, a second teosinte subspecies *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) is found only at high elevation, having adapted to these environments thousands of years prior to maize domestication (Ross-Ibarra et al., 2009; Hufford et al., 2012a). 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.

Aims

We will investigate the genetic basis of highland adaptation in maize by achieving three aims. The proposed timing of each aim and which team members will be responsible can be found in the management plan.

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 Monfreda et al. (2008).
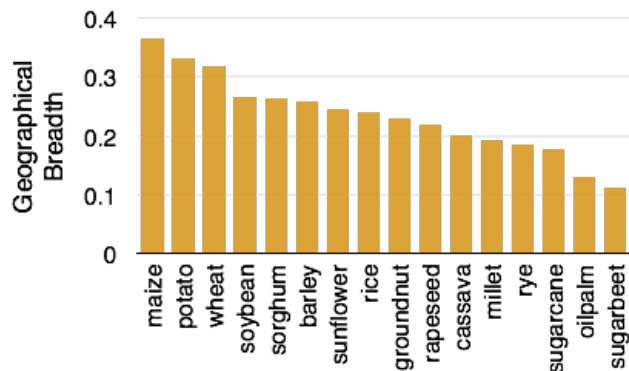
Relevance and Justification

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, 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). 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* (Hancock et al., 2011) – an enrichment of regulatory variants among loci showing evidence of selection (Pyhajarvi et al., 2013). While much of local adaptation may involve complex quantitative traits (Le Corre and Kremer, 2012), 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 (Buckler et al., 2009a), adaptive change in flowering time across latitudes has involved loci of large effect on photoperiod (Hung et al., 2012c). Traits under selection may also show distinctive patterns of effect size – while several QTL are 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 on the tassel (Brown et al., 2011). 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 in the lowlands of southwest Mexico, 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 cooler regions (Doebley, 1984). Our previous genetic analyses (van Heerwaarden et al., 2011) 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 and the power to identify and validate both widespread and population-specific candidate loci for highland adaptation.

Finally, in addition to providing insight into recent evolution in the maize genome, the study of the genetic architecture of maize adaptation 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., 2011b). Lobell et al. (2011a) 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. **Our proposal thus 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."**

Preliminary Results

Preliminary work from project members positions us to make excellent progress on our proposed aims. Drs. Ross-Ibarra and Hufford have worked extensively on the population genetics of high-land adaptation. Pyhajarvi et al. (2013) 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 (Lauter et al., 2004). This *Inv4m* inversion is the subject of our functional characterization in Aim 3. Pyhajarvi et al. (2013) also identified populations of *parviglumis* showing extensive admixture with the highland *mexicana* which are the subject of proposed analysis in Aim 1.2 and Aim 2. Hufford et al. (2013) 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, Takuno et al. (In Prep) explore selection in genotyping data from a wide collection of landraces from the highlands of Mexico and South America and find little overlap in the genes important for adaptation (Figure 2), motivating the QTL analysis in Aim 1.1.

Drs. Flint-Garcia and 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 1). Back-crosses of the highland Mexican landrace Palomero Toluqueño into the reference genome inbred B73 have been made and selfed to generate a BC1S1 population that will be further developed in Aim 3.1 to dissect the function of the *Inv4m* polymorphism. Both SSRs and SNPs that identify introgressed *mexicana* alleles at *Inv4m* have also been developed.

Dr. Coop has developed analytical approaches for studying local adaptation, including methods that allow genome-wide association with environmental variables (Coop et al., 2010; Günther and Coop, 2013), detection of selection in introgressed populations (Brandvain et al., 2014), and powerful approaches to identify selection on quantitative traits Berg and Coop (2014). His group is currently working on methods for mapping and detecting selection in admixed populations.

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.
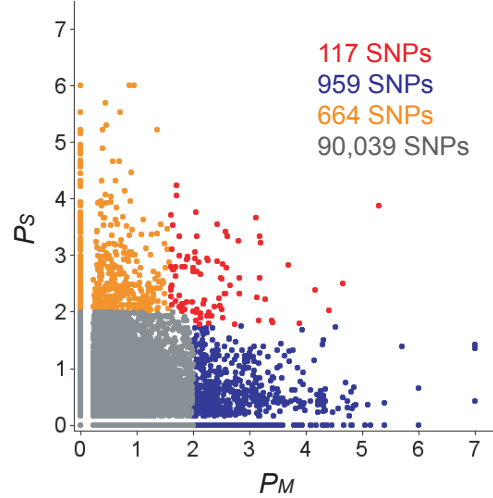
Specific Objectives

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 1). 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

Table 1: 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)	

Table 2: 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

genome-scale SNP data similar to our previous work (HapMap.v2; Chia et al., 2012). F2 plants will be genotyped using genotyping-by-sequencing (GBS; Elshire et al., 2011) and run through the standard maize GBS pipeline (Glaubitz et al., 2014) resulting in approximately ∼1M SNPs, allowing straightforward imputation of their full-genome sequence. The genetic map will be created using standard methods (Lander et al., 1987).

Populations will be phenotyped at three 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 2). 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. The entire field experiment will be replicated across two years. 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

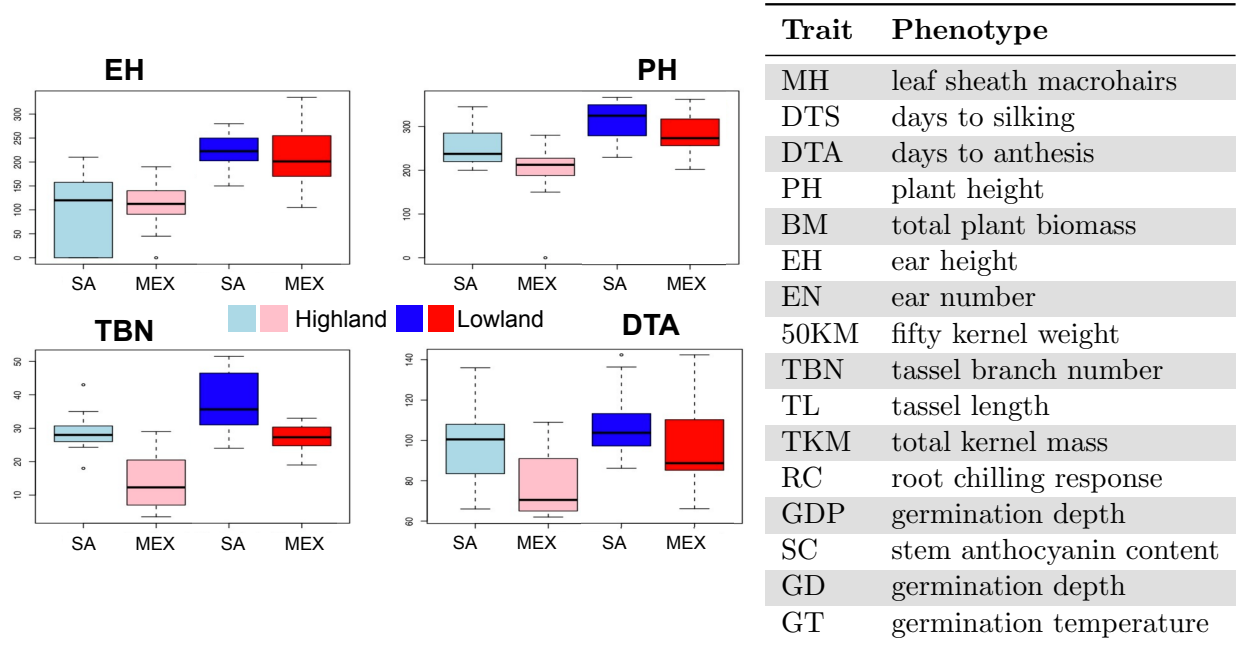| Trait	Phenotype
MH	leaf sheath macrohairs
DTS	days to silking
DTA	days to anthesis
PH	plant height
BM	total plant biomass
EH	ear height
EN	ear number
50KM	fifty kernel weight
TBN	tassel branch number
TL	tassel length
TKM	total kernel mass
RC	root chilling response
GDP	germination depth
SC	stem anthocyanin content
GD	germination depth
GT	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).

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. 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 Missouri will account for day-length differences and inform about each group's agronomic value in the Midwest.

Expected outcomes: 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, TKM, and 50KM (Figure 3)) of highland and lowland plants, as well as F2 with various combinations of QTL, in both environments.

Aim 1.2 Admixture mapping in a teosinte hybrid zone

While *mexicana* and *parviglumis* are largely allopatric, the ecological niches of these subspecies overlap in two regions of Mexico (eastern Jalisco and the eastern Balsas River Basin (Hufford et al., 2012a)) and a number of hybrid populations have been documented in these regions (Fukunaga et al., 2005). We have previously documented near equal proportions of ancestry from the two sub-

species in one of these populations near the town of Ahuacatitlan in the eastern Balsas (Pyhajarvi et al., 2013). 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 that 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 Dr. Hufford's group 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 2). Phenotypes detailed in Figure 3 will be collected for admixture mapping. Many of these traits are known to differ considerably between *parviglumis* and *mexicana* (Wilkes, 1967). 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 (Winkler et al., 2010), but 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 genome-wide association studies but not in admixture mapping). We will implement novel methods currently under development by Dr. 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.

Expected outcomes: 1) A map of QTL underlying phenotypic differences between highland and lowland teosinte, detailing the effect size of each QTL, and 2) Estimates of fitness differences (PH, BM, SM, and FK (Figure 3)) in highland and lowland sites.

Aim 2   Adaptive value of highland alleles

In 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 highland-lowland phenotypic differences adaptive?
- Does natural selection favor introgression from adapted populations?

Aim 2.1   Develop 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 (Bjarnason, 1994). This population allows us to test

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.
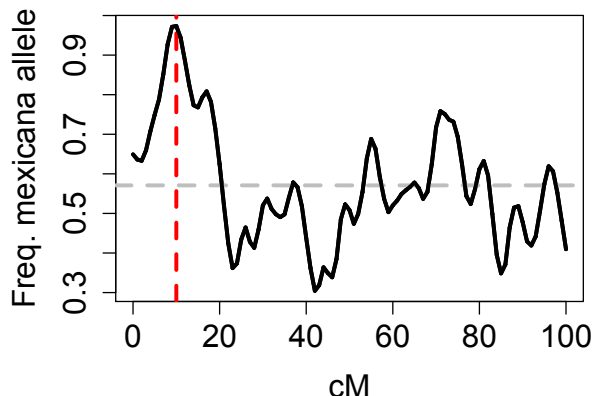
for similarities in the genetic architecture and adaptation to highland environments more broadly across all maize. This population will also highlight the diversity of highland adaptation alleles that can be drawn upon during maize breeding. Data from this worldwide sampling will be augmented by whole-genome-sequencing currently underway in the Hufford and Ross-Ibarra lab of a small set of highland maize lines from Guatemala and the Southwest U.S.

Highland Mexican landraces Hufford et al. (2013) documented extensive introgression between *mexicana* teosinte and highland maize landraces, demonstrating an overlap with teosinte QTL for macrohairs and stem pigmentation (Lauter et al., 2004). 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 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 potentially beneficial adaptive alleles.

Admixed teosinte populations Here, we will complement the Ahuacatitlan population from Aim 1.2 with samples from four additional admixed populations that have been identified using small-scale SNP data from across the range of each taxon (van Heerwaarden et al., 2011). We will revisit each of these populations to sample seed, collecting 50 individuals per population. Because these admixture events appear to be ancient (Pyhajarvi 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 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 error at heterozygous sites. 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., 2014) 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. Quantitative genetic theory suggests, however, that adaptive phenotypic change can frequently occur without strong selection on individual loci (Le Corre and Kremer, 2012). To search for evidence of selection on adaptive phenotypes, we will employ recently developed methods from Dr. Coop (Berg and Coop, 2014) that provide a powerful statistical framework to identify coordinated shifts in allele frequencies at causative QTL (from Aim 1.1,Aim 1.2) 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 (Ralph and Coop, 2010).

   Expected outcomes: 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 introgressed into highland maize (Aim 3.1). Then we will use RNA sequencing to find differences in expression, plasticity, and identify potential candidate loci within QTL (Aim 3.2).

Questions

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

Aim 3.1   Functional evaluation of *Inv4m*

In this objective we propose to characterize the genetics of *Inv4m*, an inversion polymorphism located on chromosome 4 (169-180Mb). Our previous work (Hufford et al., 2013; Pyhajarvi et al., 2013) 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*
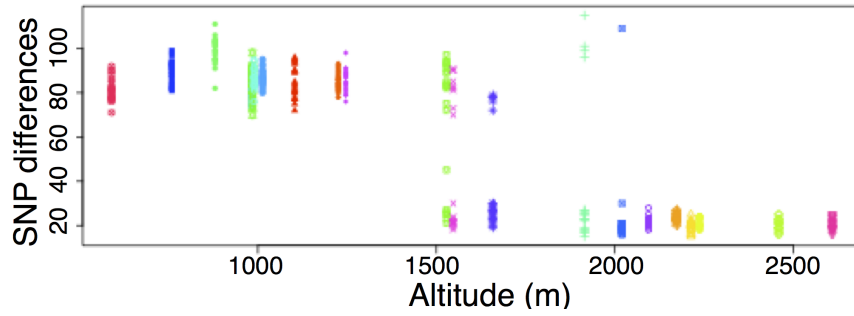
Figure 5: Clinal variation at the Chr4 inversion. Genetic distance (# of SNPs) from the canonical highland haplotype is plotted against elevation for 20 teosinte populations (shown as different colors). Low elevation (<1500m) populations lack the inversion completely, while it is fixed in populations abouve 2000m. Data from Pyhajarvi et al. (2013).

x *mexicana* cross (Lauter et al., 2004) associated with leaf pigmentation and pubescence (Figure 5).

We will generate heterogeneous inbred families (HIFs; Tuinstra et al., 1997) 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 (Reif et al., 2006); it also exhibits the highest level of *mexicana* introgression among characterized material (Matsuoka et al., 2002). Furthermore, inspection of the PT genome sequence (Vielle-Calzada et al., 2009) shows that PT carries the *mexicana* allele at *Inv4m* (Hufford et al., 2013). We will screen an existing collection of ∼150 B73 x PT BC1S3 families (three generations of selfing after one 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 six HIF derived NIL pairs (i.e. 12 lines), will be characterized at our three field sites (Table 2) and evaluated for phenotypes described in Figure 3. At each site, we will plant 3 replicate rows of our NILs and B73 and PT parents. Data will be analyzed similar to Aim 1.1, treating the introgression region either as a single block or considering individual markers.

While the probable lack of recombination between haplotypes of *Inv4m* (Pyhajarvi et al., 2013) 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: three lowland haplotypes represented by the two lowland parents of our mapping populations (Table 1) and an inbred *parviglumis*; and four highland haplotypes represented by the two highland parents of our mapping populations (Table 1), an inbred *mexicana* and the Palomero Toluqueño haplotype segregating in our HIFs. Two of the highland maize varieties are predicted to carry the *mexicana* inverted haplotype at the *Inv4m*

region, but our preliminary data suggest that highland South American maize should not (Takuno *et al.* In Prep). Each of these parents either have resequenced genomes (Vielle-Calzada et al., 2009; Chia et al., 2012) 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 3.2   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 eight inbred lines that serve as parents of our allelic series analysis in Aim 3.1 at our highland and lowland field sites (Table 2) 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 three biological replicates for each tissue/time/environment combination. We will use edgeR (Robinson et al., 2010) 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 3.1. 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. Rosas et al., 2013) 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 3.1. Using NILs generated from each of the same seven 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) in both our highland and lowland field sites (Table 2). 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. Swanson-Wagner et al., 2012) 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

Our proposal seeks to broaden educational opportunities for students and other academics through an exchange program and phenotyping workshop. In addition to normal means of research dissemination (publications, conferences), we will also proactively distribute code for teaching and research. 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. 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 team members will participate, sending students to Mexico and/or accepting students from Mexico for internships. Dr. 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 six 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 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. 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 team members, 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, Missouri 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 user accounts with a Palm handheld, as well as setting up locations, traits, and projects, assigning plots 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 student's understanding of genetics and evolution, but these subjects are often conceptually quite difficult. A basic understanding of genetic variation is important for all citizens, due to the rise of personal genomics and genomic medicine (e.g. Redfield, 2012). 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 F2:3 generation 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 the 1990s, and is one of the areas most vulnerable to climate change (Bellon et al., 2011); germplasm

developed here could be an important contribution to furthering such breeding 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, 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 maize evolution has been revised and published online.
Publications Peiffer et al. (2013); Romay et al. (2013); Wills et al. (2013); Mezmouk and Ross-Ibarra (2014); Peiffer et al. (2014); Sood et al. (2014)

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 Dr. Ross-Ibarra's work on centromere evolution 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 protein mutants, assess the phylogenetic evidence for coevolution, and study patterns of centromere and genome size variation in natural populations.
Broader impacts Co-PI Ross-Ibarra has established an international student exchange program as part of this grant. Data and 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 Shi et al. (2010); Chia et al. (2012); Fang et al. (2012); Hufford et al. (2012b,c, 2013); Melters et al. (2013); Kanizay et al. (2013); Pyhajarvi et al. (2013)

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 Bradburd et al. (2013)

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 pioneered the characterization of population genetic and evolutionary parameters of maize diversity, developed resources to connect genetic diversity to phenotype through association and joint linkage-association mapping, and conducted fine scale analysis of domestication and agronomic QTL. The project successfully released and analyzed the maize Nested Association Mapping (NAM) population, collaborated on making haplotype maps for maize, resolved domestication traits, developed a range of novel statistical approaches for association mapping, and dissected a number of complex traits.

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), online Teacher Friendly Guide to the Evolution of Maize, seven Genotyping-By-Sequencing (GBS) workshops, and training of postdocs, graduate students and undergraduates, the vast majority of which have continued in scientific careers. Former trainees on this grant include Drs. Flint-Garcia and Ross-Ibarra; only their publications are shown below.

Publications Buckler et al. (2009b); Flint-Garcia et al. (2009a,b,c); Gore et al. (2009); McMullen et al. (2009); Ross-Ibarra et al. (2009); Bottoms et al. (2010); Dubois et al. (2010); Zhang et al. (2010); Van Heerwaarden et al. (2010); van Heerwaarden et al. (2010); Brown et al. (2011); Morrell et al. (2011); Studer et al. (2011); van Heerwaarden et al. (2011); Tian et al. (2011); Chia et al. (2012); Cook et al. (2012); Fang et al. (2012); Hufford et al. (2012c); Hung et al. (2012b,a); Romay et al. (2013)

Ross-Ibarra, Hufford: USDA #2009-65300-05668: Scanning for Climate Change: High-throughput Discovery of Loci for Advanced Breeding in Maize

$300,000. 09/01/013-08/31/15. 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. Data and result of this project have been disseminated via publications and presentations.

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