Project Summary

Maximum of 1 page

Intellectual Merit

Broader Impacts

Project Description

Introduction

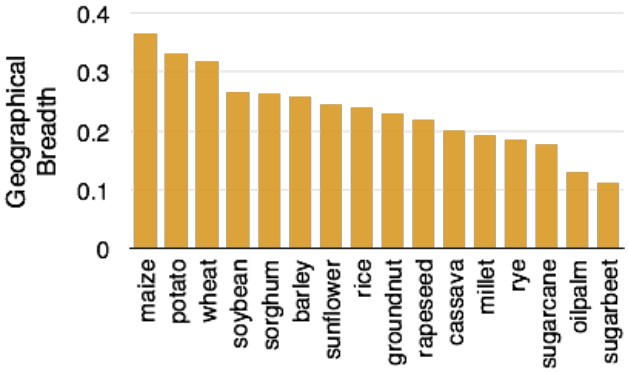
Local adaptation occurs when natural selection prevails over gene flow and organisms achieve maximum fitness at their native sites (Kawecki & Ebert 2004). An understanding of local adaptation is particularly pressing given current issues of climate change, conservation, and sustainable agricultural production (Savolainen et al. 2013). While the genetic basis of local adaptation is generally not well understood, the declining cost of next generation sequencing has enabled a handful of genome-wide studies across many populations of model species. For example, Fournier-Level and co-authors demonstrated that alleles associated with high fitness in Arabidopsis thaliana had a tendency to be both local and linked to climate (Fournier-Level et al. 2011). Likewise, a recent study across hundreds of accessions of Medicago truncatula identified candidate loci for local adaptation and found them to be predictive for growth rate under temperature and soil moisture treatments in a growth chamber (Yoder et al. 2014). Finally, our genome-wide study of 21 populations of teosinte (i.e., wild maize) revealed that inversion polymorphisms played a substantial role in adaptation across an elevational gradient (Pyhjrvi et al. 2013). Clearly, initial genomic studies are yielding valuable insights regarding local adaptation, yet much remains to be discovered.

Agricultural species represent particularly promising systems for ongoing research on local adaptation. Crops were, in most cases, domesticated in narrow geographic centers prior to global spread. During diffusion, crops encountered a wide range of novel environments and were forced to adapt. In many instances, putative traits underlying local adaptation in crops have already been identified. These systems therefore represent compelling opportunities for investigating the genetic architecture of local adaption. Moreover, insights gained regarding adaptive loci can feed back into modern crop improvement yielding valuable benefits in the face of modern rapid environmental change.

Here we propose an investigation of local, specifically highland, adaptation in maize (Zea mays ssp. mays). Maize was domesticated in the lowlands of southwest Mexico from the narrowly distributed teosinte Zea mays ssp. parviglumis (hereafter, parviglumis) (Matsuoka et al. 2002). Following domestication, maize spread to the highlands of the Central Plateau, a migration across more than 1000 meters of increasing elevation to a dramatically different environment (CITE). Maize landraces in the Central Plateau have distinct morphologies (e.g., highly pigmented and hairy leaves and stems) that are believed to confer adaptation to this cooler region (CITE). Interestingly, these morphologies are also found in the highland-adapted teosinte Zea may ssp. mexicana (hereafter, mexicana)(CITE). Our recent work has shown that these shared traits are likely the result of adaptive introgression from mexicana into maize (Hufford et al. 2013).

Colonization of the Central Plateau is not the only occasion where highland adaptation was required during the diffusion of maize. For example, landraces (i.e., local farmer varieties) of maize are commonly grown above 3000 meters in the Andes of South America. Andean landraces share adaptive highland traits with landraces from the Central Plateau of Mexico, yet these groups are highly diverged (Vigouroux et al. 2008). Moreover, Andean maize grows outside the distribution of mexicana and all other teosinte and must therefore have obtained highland adaptation de novo

Figure 1: Geographic breadth of the world's 16 most important crops. Geographic breadth is expressed in percent of land surface area in which each crop is cultiavated. Data from **?**.



or through standing variation. The genetic basis of highland adaptation may therefore be quite distinct between the Central Plateau and the Andes.

Aims

We will investigate the genetic basis of highland adaptation in maize by achieving three specific aims:

1. **Genetic architecture of highland traits in maize**
2. **Population genetics of highland adaptation in maize and teosinte**
3. **Functional characterization of adaptive quantitative trait loci**

Rationale and Significance

Maize is highly adaptive

Maize adaptation is important for food security important for foods
population expansion: need to grow more food in new places
climate change: climate will change in places maize is grown

We know almost nothing about genetics adaptation in maize we know lots about quant. traits in maize (temperate), but almost nothing about adaptive traits

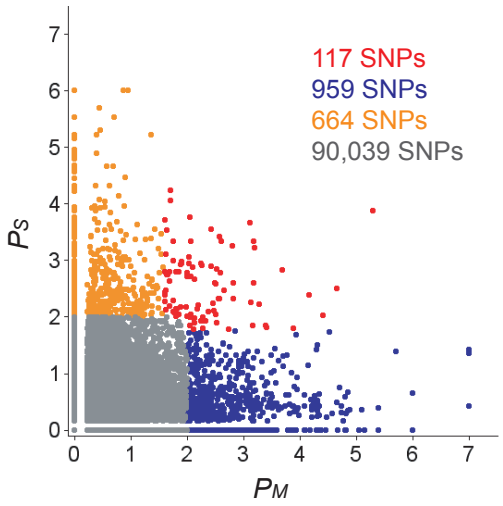
genetic arch. can differ. e.g. time to flowering in temp maize lots of genes small effect, but diff. between trop and temp is few photoperiod loci of large effect – so adaptation to temperate was predominantly a few loci of large effect even though the genetic arch. of flowering time among temperate is lots of loci of small effect (reference fisher?) cite brown et al. 2011

Preliminary Results

cite tanja for teosinte for maize, Huff 2013 shos stuff

describe current genome sequencing which will look at popgen and colonization

Figure 2: Scatter plot of $-\log_1 0 empirical P$-values of F_{ST} values in Mexico (P_M on x-axes) and South America (P_S on y-axes). Red, blue, orange and gray dots represents SNPs showing significance in both Mexico and South America, only in Mexico, or only in South America, respectively. The number of SNPs in each category is shown in the same color of dots.



but none of this looks at quantitative traits (cite why quant. traits are different – Rockman, lecorre & kramer)

blah blah F_ST **??**.

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 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?
- Are highland QTL/loci widespread in highland climes?

Aim 1.1 QTL mapping of highland adaptation

The first objective of this project is to identify the genomic regions controlling highland adaptation in maize, we will conduct QTL mapping studies of one Mexican and one South American population,

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

each derived by crossing an inbred landrace adapted to lowland conditions with a landrace adapted to highland conditions 1. We make use of specially-inbred landrace lines created by John Doebley (U. Wisconsin) and Seth Murray (Texas A&M), thus simplifying downstream mapping applications and allowing replication of alleles in the our functional studies (see Aim 3).

Because we plan to evaluate these populations in replicated trials over multiple years, we will self-pollinate the F2 plants to create 500 F2:3 families from each population. DNA will be extracted from each of the F2 plants and used for genotypic analysis. The parents of the population will be 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 in scale (tens millions of markers) to our previous work in maize (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 and allowing simple imputation of full-genome sequence. The genetic map will be created using standard methods (e.g. **?**).

The 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 (2). At each field location, best local practices will be used including fertilizers, and pest and weed control. In each site, the experiment will consist of two replications where the 500 entries and parental checks will be arranged in a randomized complete block design

Table 3: Phenotypes measured

Abbreviation	Phenotype	Lowland	Highland
MH	leaf sheath macrohairs		
FT	flowering time		
PH	plant height		
BM	total plant biomass		
EN	ear number		
FK	fifty kernel weight		
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		

or an augmented alpha lattice design. The parental checks will be used to control for field variation. We will collect a large number of phenotypes (3) using our in-house barcode-based data collection program. Germination at different planting depths and temperatures will be evaluated in controlled conditions 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, as expected. Each location will then be analyzed separately to derive least squares means to be used as the phenotypic data in QTL analyses. QTL analysis will be conducted using standard, publicly available 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. The QTL profiles will be compared across populations (Mexican vs South American) and across field sites (elevations) to determine how elevation affects putatively adaptive traits. We expect very different QTL profiles from the highland and lowland evaluation sites, and from the Mexican and South American populations. 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 number and effect size of each QTL and differences between crosses, and 2) Estimates of fitness differences (PH, BM, SM, and FK 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

Aim 2 Adaptive value of highland alleles

In aim **??** we will map loci corresponding to traits differing between highland and lowland maize and teosinte. In this section we will investigate the evolutionary consequence of the QTL identified in **??** in three natural introgression experiments: gene flow from *mexicana* into highland maize Aim 2.2 and admixture between *mexicana* and *parvgilumis* **??**.

Questions

- Are introgressed loci adaptive?
- Does evidence of introgression and natural selection correspond to QTL?

Aim 2.1 Global analysis of highland haplotypes

And though quantitative genetic theory suggests that adaptive phenotypic change may not correlate with strong evidence for selection on individual loci (**?**), recently developed methods from CoPI Coop (**?**) 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.

Aim 2.2 Population genetics of maize-teosinte introgression

Our previous work (**?**) we 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 revisit these populations with higher-density genotyping that will allow identification of ongoing gene flow in individual populations and ask whether introgressed regions have been targeted by natural selection.

We propose to resample 18 individuals from each of the same 9 sympatric population pairs and two allopatric populations studied in **?**. Each individual will be genotyped by GBS **??** using greater than normal depth (48 plex) to improve genotyping heterozygous sites. These data will provide ¿1M SNPs across the genome (compared to 40K SNPs in **?**). We will use both haplotype **??** and heterozygosity-based **??** methods to identify introgressed segments in individual populations. Genomic regions showing evidence of introgression will be tested for selection using population genetic approaches which utilize evidence from the site frequency spectrum (**?**) and haplotype structure (**?**). Correlations between genetic differentiation and recombination will allow us to investigate selection against introgression (**?**), quantifying the "linkage drag" associated with introgression of potentially beneficial adaptive alleles. Finally, we will again apply the approach of **?** to evaluate selection on individual phenotypes across highland maize landrace populations.

The expected outcomes of this objective are 1) a fine-scale dissection of both ancient and ongoing introgression 2) identification of introgressed regions showing evidence of positive selection, identifying loci important for highland adaptation, 3) evidence for selection on specific phenotypic

traits, 4) quantification of the potential "linkage drag" or evidence against introgression across otehr regions of the genome.

Aim 2.3 Population genetics of hybridization in teosinte

make use of admixed pop sampled by matt Berg approach of phenotypes mapped there for selection in parentals and hybrid

GBS Additional 5 admixed parv/mex populations (50 inds. each) (JRI) also 4 new pops x (12 parv + 12 mex + 12 hybrids) = $8160 Ahuacatitlan and 3 more Introgression and adaptation in additional admix pops (JRI, GC) Selection for/against regions in parv/mex/admix Parallelism across pops in hybrid zone

Aim 3 Functional characterization of adaptive QTL

After mapping QTL for highland adaptation (**??**) and studying their adaptive significance (Aim 2), in this aim we will aim to better understand the functional genetic basis of adpative regions. First, we will fine map a chromosome 4 QTL and study its phenotypic effects introgressed into a maize background (Aim 3.2). Then we will evaluate the effects of an allelic series from highland and lowland maize and teosinte at the chromosome 4 QTL (Aim 3.3). Finally, we will use RNA sequencing data to investigate plasticity, differences in expression, and identify potential candidate loci within QTL (Aim 3.4).

Aim 3.1 Questions

- What do QTL/selected loci/introgressed loci do?

Aim 3.2 Fine mapping and phenotypic evaluation

In this objective we propose to functionally characterize the genetics of a large *mexicana*-maize introgression block located on chromosomes 4 (Chr4: 169-180Mb). Our previous analysis (**?**) indicates that this region is supported by a robust signature of introgression, shows broad distribution among highland races , and overlaps with a QTL identified in a *parviglumis* x *mexicana* cross (**?**) associated with leaf pigmentation and pubescence. Our analysis of linkage disequilibrium in teosinte suggests that the region represents an inversion polymorphism (**?**), suggesting the possibility of characterizing the region as a block by generation of introgression stocks.

We will generate heterogeneous inbred families (HIFS; **?**) from a cross of the landrace Palomero Toluqueño (PT) and the reference genome inbred B73. PT is 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 *mexicana* alleles at two SNPs shown previously to exhibit a fixed difference between *mexicana* and maize (**?**). We will screen an existing collection of 150 B73 x PT BC1S3 (three generations of selfing after 1 generation of back-cross) families to identify HIFs segregating for B73 and PT haplotypes using microsatellite markers we have shown distinguish B73 and PT alleles in this region. HIFs will be selfed to generate pairs of near-isogenic lines homozygous for the B73 or PT haplotype. In total, we will generate 12 test stocks, including 11 BC4S1 homozygous parents and the B73 reference line. These will be will

be genotyped using GBS **??** to confirm the extent of introgression at the Chr4 candidate locus and unlinked sites, and selfed to generate BC4S2 seed for phenotypic analysis. These lines will allow estimation of genotypic values at the candidate locus in genomic contexts that contains significant (
$approx$25%) unlinked PT content, capturing potential epistatic effects important to function of the candidate region.

Outcomes of this objective include XXX YYY ZZZ.

Z replicates B73 x PT stocks will be grown in our three field sites 2 and evaluated for phenotypes described in 3. Given the relatively small number of lines test

3.3 Phenotypic analysis The bulk of our phenotypic analysis will focus on B73 x PT stocks (described in 3.2.1, above) that will be available from early in the funding period. Diverse NILs (3.2.2) will not be available until year three, at which point they will play a key role in confirmation and dissection of effects identified in PT stocks. The results of targeted characterization of the Chr 4 candidate region may coincide with mapping experiments described in Section X. Given the relatively low number of test stocks, however, we envisage that a greater range and detail of phenotypic characterization will be possible, and that we may well identify effects not directly observed in mapping experiments. In addition to the activities described below, we anticipate that a number of collaborations will be established beyond the scope of the present proposal to allow precision characterization of highly specific phenotypic traits (e.g. cold tolerance; root architecture; biotic interactions).

3.3.2 Growth chamber characterization MATT? PRECISE PHENOTYPING PIGMENTS AND HAIRS (CONTROLLED LIGHT; METHODS FROM LAUTER; SEM?) OTHER SEEDLING TRAITS? FOCUS ON 12 HIF FAMILIES?

Aim 3.3 Allelic series for QTL of interest

Introgression of multiple candidate haplotypes into the B73 reference background To investigate further functional diversity in the Chr4 candidate region, we will generate a series of NILs by marker-assisted recurrent backcross to B73 using a collection of diverse donor varieties (parental lines of mapping populations described in section X; a reference inbred ssp. mexicana; two additional lowland races; two additional highland races). Exotic donors have been selected on the basis of diverse origin, availability of genomic information and their use in other components of the project. To select for exotic haplotypes during introgression, we will use the SSR markers described above coupled with GBS. Our exotic donors include partially inbred and open-pollinated stocks. Where prior characterization has revealed multiple haplotypes to be present in a single stock, we will select the most common haplotype for introgression by analyzing multiple F1 plants and then selecting a single individual for the first backcross. BC4S2 families are projected to be available by the end of 2017 (Table X).

Aim 3.4 Gene expression

We will first assess the effects of high and low elevation environments on genome-wide expression differences to identify genes responsive to these environments. We will grow the 8 inbred lines which serve as parents of our allelic series analysis in Aim 3.3 in a highland and lowland field site 2. 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 assess differences in expression across environments and identify overlap between differentially expressed (DE) genes and QTL from **??**, loci showing selection identified in Aim 2, and introgressed regions showing phenotypic differences in **??**. These results will help narrow down potential candidate genes in QTL and serve as functional validation of loci showing population genetic evidence of selection in introgressed and admixed populations. 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 to ask whether genes showing a plastic response in unadapted material (lowland landraces, *parviglumis* B73) show constitutive response in adapted lines (*mexicana*,highland landraces).

Our second approach will be a targeted analysis of transcriptomic changes in the chromosome 4 NIL lines from Aim 3.3. 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 to the donor transcriptomes will allow us to differentiate between cis and trans regulation of expression within the chromosome 4 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.

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-Garcias 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 additional use of this mapping population to study additional phenotypes of interest (e.g. root morphology and growth). Seed from our NIL populations will allowing 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. 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 ??????????????????????????

Table 4: Proposed timeline of activities and responsibilities

Year	2015	2016	2017	2018	2019
Objective **??** Allelic series	2015	2016	2017	2018	2019
Objective **??** Fine mapping	–	RS, AC	AC,JRI	AC,JRI	–
Objective **??** RNA-seq	2015	2016	2017	2018	2019
Objective Aim 2.2 Maize/mexicana introgression	2015	2016	2017	2018	2019
Objective **??** Admix mapping	2015	2016	2017	2018	2019
Objective **??** QTL mapping	2015	2016	2017	2018	2019
Objective **??** Highland haplotypes	2015	2016	2017	2018	2019
Objective Aim 2.3 Admixture population genetics	2015	2016	2017	2018	2019

Biographical Sketch: Your Name

Maximum of 2 pages

Biographical Sketch

Professional Preparation

Undergraduate Institution(s)	Major	Degree	Year
Graduate Institution(s)	Major	Degree	Year
Postdoctoral Institution(s)	Area		Year
Year-present	Position, Department, Institution		
Year(s)	Position, Department, Institution		

Publications

Five Publications Most Closely Related to the Proposed Project

1. Author(s): Article Title, *Journal Title* **Volume Number**, Page Numbers, Year of Publication.
2.
3.
4.
5.

Ten Other Significant Publications

1.
2.
3.
4.
5.
6.
7.
8.
9.

10.

1.

2.

3.

4.

5.

Collaborators & Other Affiliations

Collaborators:

Graduate and Postdoctoral Advisors:

Thesis Advisor and Postgraduate-Scholar Sponsor:

Budget Justification

Maximum of 3 pages

Personnel

Other Personnel

Fringe Benefits

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

- Faculty - % in FYs 2012, 2013, and 2014
- Postdocs - % in FYs 2012, 2013, and 2014
- Graduate students - % in FYs 2012, 2013, and 2014
- Undergraduate students - % in FYs 2012, 2013, and 2014
- Staff - % in FYs 2012, 2013, and 2014
- Part time staff - % in FYs 2012, 2013, and 2014

Additionally, the university applies a risk management charge of 1% on all personnel salaries

Travel

Other Direct Costs

Materials and Supplies:
Consultant Services:
Graduate Student Tuition

Indirect Costs

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.