

# Project Summary

## Overview

The proposed project will investigate the genome-wide effects of hybridization and introgression in the genus *Zea* in two objectives. First, investigators will study the process of incipient speciation between the lowland-adapted subspecies *Zea mays* ssp. *parviglumis* (the wild progenitor of maize; hereafter, *parviglumis*) and the highland-adapted *Zea mays* ssp. *mexicana* (hereafter, *mexicana*). Through targeted population sampling, high-density genotyping, and common garden studies, the investigators will assess what fraction of the genome is porous to gene flow in hybrid zones, how fitness of these taxa varies across a hybrid zone, and how allele frequency clines are patterned at adaptive loci. Second, investigators will determine how hybridization and introgression have altered the *Zea* genus as domesticated maize (*Zea mays* ssp. *mays*) spread from its center of origin into sympatry with multiple wild *Zea* species. Population genomic analyses of sympatric collections of maize and its wild relatives will be used to assess whether maize served as a bridge for gene flow between otherwise allopatric *Zea* species and whether maize received gene flow from locally adapted wild relatives that facilitated its spread into new environments.

## Intellectual Merit

Much progress has been made in the study of hybridization and introgression through the development of theory, through field-based ecological research, and through genetic analyses based on a limited number of molecular markers. However, much remains to be discovered regarding how these evolutionary processes have shaped genomes. The research proposed here will leverage the genomic resources of the maize model system to investigate how hybridization and introgression have molded the genomes of both wild *Zea* species and domesticated maize as it spread from its center of origin in southwest Mexico. Through this work, the investigators will be able to assess the genome-wide effects of hybridization and introgression on two different timescales: 1) An evolutionary timescale covering 60,000 generations of divergence and gene flow between *parviglumis* and *mexicana*; and 2) An ecological timescale in which maize has spread across the Americas and adapted to local conditions, potentially through hybridization and introgression from endemic wild relatives. The analysis on an evolutionary timescale will generate basic knowledge on the process of incipient speciation and the porous nature of the genomes of diverging species, whereas the analysis on an ecological timescale can inform, for example, the study of biological invasions if maize is viewed as an invasive species aided in colonization by gene flow from locally-adapted native species.

## Broader Impacts

The investigators will achieve societally relevant outcomes in the proposed project on hybridization and introgression by providing STEM training opportunities for undergraduate and graduate students, participating in Iowa State University's GK12 Fellowship program, and establishing an exchange program between universities in the United States and Mexico. This project will provide ample training opportunities for both undergraduate and graduate students in laboratory, computational, and field-based research. The investigators have successfully recruited minority students

into their research programs in the past and will make every effort to do so as part of this proposed work. The graduate student funded to work at Iowa State University will participate in the University's GK12 program that serves the Des Moines public school system. Through this opportunity, the graduate student will bring their research on maize evolution into the classrooms of the most diverse public school system in the state of Iowa. The *Zea* study system provides an excellent opportunity to deliver evolution training to middle and high school students in Iowa. Corn is a highly important commodity in the state and dominates the landscape. Students will likely be intrigued to discover the rich evolutionary history of this crop and develop a greater appreciation of the processes of adaptation that have brought corn from the Pacific coast of Mexico to the Midwestern United States. Finally, the proposed exchange program would create an opportunity for students from the United States to conduct research internationally and also allow these students to interact with visiting students from Mexico. Through these interactions, students will be better prepared for modern STEM research, which is often highly collaborative and international in nature.

# Project Description

## Introduction

While the potential roles of hybridization and introgression as agents of evolution have long been appreciated (Anderson 1948; Anderson & Stebbins 1954; Stebbins 1959), only recently have technological innovations begun to allow their characterization on a genome-wide scale. Multiple genome-wide studies have now reported substantial inter-taxon introgression in both plant (Hufford et al. 2013; Renaut et al. 2013) and animal (Dasmahapatra et al. 2012; Staubach et al. 2012) species. In the investigation proposed here, we will build upon our previous work in maize (*Zea mays* ssp. *mays*) and its wild relatives (*Zea* spp.; collectively referred to as teosinte) in order to characterize the evolutionary role of hybridization and introgression in the genus *Zea*. This study system is ideally suited for such research due to its well-characterized history, large natural populations, and the availability of exceptional genomic resources (Hufford et al. 2012).

## Objectives

We will leverage the many positive attributes of the *Zea* study system to address two primary objectives:

### **Objective I) Assess the evolutionary significance of hybridization and introgression in locally-adapted, parapatric wild *Zea* taxa**

*Zea mays* ssp. *parviglumis* (the wild progenitor of maize; hereafter, *parviglumis*) and *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) diverged approximately 60,000 BP (Ross-Ibarra et al. 2009) and have parapatric distributions: while *parviglumis* occurs in the warm lowlands of southwest Mexico, *mexicana* is found in the cool highlands of the Mexican Central Plateau. Narrow regions of substantial admixture between these wild subspecies have been discovered at middle elevations (Fukunaga et al. 2005; Pyhjärvi et al. 2013). Through targeted collections, generation of high-density genotyping data, population genomic analyses, and common garden experiments, we will address the following research questions and work toward establishing this system as a model for the study of hybrid zones:

*Research Question IA) What fraction of the genome is porous to gene flow in hybrid zones?*

*Research Question IB) How do parviglumis, mexicana, and hybrid fitness vary across the hybrid zone?*

*Research Question IC) Are allele frequency clines at loci associated with fitness traits in mexicana and parviglumis steeper than those at non-associated loci?*

**Objective II) Determine the extent to which hybridization and introgression have altered the Zea genus during the post-domestication spread of maize**

Maize was domesticated in southwest Mexico from parviglumis 10,000 BP (Matsuoka et al. 2002) and quickly spread throughout North America, bringing this crop into sympatry with new species of teosinte (Vigouroux et al. 2008). Through dense genotyping of range-wide samples of maize and teosinte, we will assess two questions regarding the importance of introgression during the spread of maize:

*Research Question IIA) Did maize serve as a bridge for gene flow between previously isolated Zea taxa?*

*Research Question IIB) Was the spread of maize facilitated by gene flow from locally-adapted wild Zea?*

## Rationale and Significance

Pioneers in evolutionary biology including G. Ledyard Stebbins and Edgar Anderson recognized the important role hybridization and introgression could play in adaptation and speciation (Anderson 1948; Anderson & Stebbins 1954). These evolutionary forces were thought to be particularly influential when environmental conditions for a species were marginal, variable, or new (Stebbins 1959). More recently, theoretical and empirical investigations of hybridization have focused on hybrid zones, defined as regions where distinct taxa co-occur and mate, resulting in progeny of mixed ancestry (Harrison 1993). While hybrid zone theory has progressed and many compelling empirical examples have been identified based on hybrid morphology and limited genetic data (Delmore et al. 2013; Galindo et al. 2013; Parchman et al. 2013; Smith et al. 2013b), many outstanding questions remain. For example, additional study is needed to determine whether hybrid zones are primarily maintained as tension zones in which hybrids are selected against or as ecotones where hybrids have an advantage under certain environmental conditions (Kruuk et al. 1999; Rasmussen et al. 2012; Smith et al. 2013a). Additionally, genome-wide analysis of the fraction of the genome that is porous to gene flow in hybrid zones is rare and will likely offer considerable insight.

Both wild and domesticated Zea offer exciting opportunities to study hybridization. Parviglumis and mexicana are distributed across a steep altitudinal gradient and differ for putatively adaptive phenotypes such as the presence of macrohairs and stem pigmentation found in mexicana. Analysis of 100 microsatellite markers genotyped in a range-wide sample of parviglumis and mexicana identified clear regions of elevated admixture in two geographically-distinct, mid-elevation regions: eastern Jalisco state and the eastern Balsas River Basin (Fukunaga et al. 2005). Our genome-wide

analysis ( 40,000 SNPs) of a hybrid population in the eastern Balsas revealed extensive admixture (Pyhjärvi et al. 2013), and the relatively short length of shared haplotypes suggested continual gene flow between *parviglumis* and *mexicana* over a substantial period of time (Pyhjärvi et al. 2013). Furthermore, a reanalysis of a data set of 1000 SNPs from samples across the range of both taxa (Fang et al. 2012) reveals substantial admixture in multiple populations at middle elevation (Figure 1).

Because of its recent domestication and rapid spread, hybridization between maize and other *Zea* must be quite young from an evolutionary perspective. Our recent work has provided convincing evidence of introgression from *mexicana* into maize at QTL for phenotypes (e.g., pigment and macrohairs) that distinguish highland *mexicana* from lowland *parviglumis* (Hufford et al. 2013). Our interpretation of this result is that maize received adaptive introgression from *mexicana* that allowed the crop to spread into the highlands of Mexico. Subsequent to its diffusion across the Mexican highlands, maize spread into sympatry with two additional teosinte taxa in Guatemala, *Zea mays* ssp. *huehuetenangensis* (hereafter, *huehuetenangensis*) and *Zea luxurians* (hereafter, *luxurians*). Hybrids between maize and the Guatemalan teosinte taxa have been observed in the field (Wilkes 1977) and evidence has been found of *mexicana* haplotypes segregating in *luxurians* (Ross-Ibarra et al. 2009). *Mexicana* and *luxurians* are entirely allopatric in their distributions, which suggests maize has served as a bridge for gene flow between these two taxa.

## Preliminary Results

### Specific Objectives

### Broader Impacts

Our efforts to broaden the impact of the research proposed here will begin within our groups through our commitment to effectively mentor volunteer undergraduate interns as well as graduate students and/or postdoctoral scholars funded by the project. Students and postdocs will receive one-on-one training from the investigators and senior personnel on laboratory, computational, and field research methods. Mentees will also be encouraged and funded to present their work at scientific conferences. Our groups have an excellent mentoring track record with four undergraduate students in the last five years publishing their work in scholarly journals and multiple underrepresented minorities participating in our research. In addition to the student and postdoc mentoring that will occur within our groups, as part of our broader impact activities one of our graduate students will participate in Iowa State University's GK12 Fellowship program. The selected graduate student will spend one full day each week in a science middle or high school classroom for the entire academic year in the Des Moines Public School District. This is the largest and most diverse school district in Iowa. The graduate student will introduce the K12 students to the scientific process through inquiry-based activities, relate the students science curriculum to real world examples, work with students on their science fair projects, and serve as a role model in a STEM profession. Furthermore, the graduate student will introduce students to his/her research project on hybridization and introgression in *Zea*, a topic that is particularly well suited for teaching evolution in Iowa given the important role that maize plays in the Iowan economy. In introducing his/her dissertation research, the graduate student will engage Des Moines students in how re-

search is conducted and provide STEM content professional training to his/her partner teacher. Finally, we will establish a student exchange program between the Eguiarte Laboratory at UNAM in Mexico and the Hufford and Ross-Ibarra Laboratories in the United States. The Ross-Ibarra Laboratory has run an NSF-supported, US-Mexico exchange program for the last three years with two students earning authorship on forthcoming papers and we will build upon the success of this program. A student from the Eguiarte group will spend 2-3 months in either the Hufford or Ross-Ibarra Laboratory learning the GBS methodology and/or honing his/her skills in population genomic analysis, whereas a student from the Hufford and/or Ross-Ibarra Laboratories will travel to Mexico to participate in sample collection trips and to obtain expertise in common garden field experiments. This exchange will build capacity in all groups involved and will provide a valuable international research experience for a graduate student supported by the grant.

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

**Broader impacts** Co-PI Ross-Ibarra has established an international student exchange program as part of this grant. Data and result of this project have been disseminated via publications and presentations as well as deposited in the maize genetics community database [www.maizegdb.org](http://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. (2012a,b, 2013); Melters et al. (2013); Kanizay et al. (2013); Pyhajarvi et al. (2013)

## References Cited

- JM Chia, C Song, PJ Bradbury, D Costich, N de Leon, J Doebley, RJ Elshire, B Gaut, L Geller, JC Glaubitz, M Gore, KE Guill, J Holland, MB Hufford, J Lai, M Li, X Liu, Y Lu, R McCombie, R Nelson, J Poland, BM Prasanna, T Pyhajarvi, T Rong, RS Sekhon, Q Sun, MI Tenaillon, F Tian, J Wang, X Xu, Z Zhang, SM Kaeppler, J Ross-Ibarra, MD McMullen, ES Buckler, G Zhang, Y Xu, and D Ware. Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet*, 44(7):803–807, 2012.
- Z Fang, T Pyhajarvi, AL Weber, RK Dawe, JC Glaubitz, J Gonzalez Jde, C Ross-Ibarra, J Doebley, PL Morrell, and J Ross-Ibarra. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, 191(3):883–894, 2012.
- MB Hufford, P Bilinski, T Pyhajarvi, and J Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends Genet*, 28(12):606–615, 2012a.
- MB Hufford, X Xu, J van Heerwaarden, T Pyhajarvi, JM Chia, RA Cartwright, RJ Elshire, JC Glaubitz, KE Guill, SM Kaeppler, J Lai, PL Morrell, LM Shannon, C Song, NM Springer, RA Swanson-Wagner, P Tiffin, J Wang, G Zhang, J Doebley, MD McMullen, D Ware, ES Buckler, S Yang, and J Ross-Ibarra. Comparative population genomics of maize domestication and improvement. *Nat Genet*, 44(7):808–811, 2012b.
- MB Hufford, P Lubinsky, T Pyhajarvi, MT Devengenzo, NC Ellstrand, and J Ross-Ibarra. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477, 2013.
- LB Kanizay, T Pyhajarvi, EG Lowry, MB Hufford, DG Peterson, J Ross-Ibarra, and RK Dawe. Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *zea mays*. *Heredity (Edinb)*, 110(6):570–577, 2013.
- DP Melters, KR Bradnam, HA Young, N Telis, MR May, JG Ruby, R Sebra, P Peluso, J Eid, D Rank, JF Garcia, JL Derisi, T Smith, C Tobias, J Ross-Ibarra, I Korf, and SW Chan. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*, 14(1):R10, 2013.
- S Mezmouk and J Ross-Ibarra. The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda)*, 4(1):163–171, 2014.
- JA Peiffer, SA Flint-Garcia, N De Leon, MD McMullen, SM Kaeppler, and ES Buckler. The genetic architecture of maize stalk strength. *PloS one*, 8(6):e67066, 2013.
- JA Peiffer, MC Roday, MA Gore, SA Flint-Garcia, Z Zhang, MJ Millard, CA Gardner, MD McMullen, JB Holland, PJ Bradbury, and ES Buckler. The genetic architecture of maize height. *Genetics*, 2014.
- T Pyhajarvi, MB Hufford, S Mezmouk, and J Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome Biol Evol*, 5(9):1594–1609, 2013.
- MC Roday, MJ Millard, JC Glaubitz, JA Peiffer, KL Swarts, TM Casstevens, RJ Elshire, CB Acharya, SE Mitchell, SA Flint-Garcia, MD McMullen, JB Holland, ES Buckler, and



- CA Gardner. Comprehensive genotyping of the usa national maize inbred seed bank. *Genome Biol*, 14(6):R55, 2013.
- J Shi, SE Wolf, JM Burke, GG Presting, J Ross-Ibarra, and RK Dawe. Widespread gene conversion in centromere cores. *PLoS Biol*, 8(3):e1000327, 2010.
- S Sood, S Flint-Garcia, MC Willcox, and JB Holland. Mining natural variation for maize improvement: Selection on phenotypes and genes. In *Genomics of Plant Genetic Resources*, pages 615–649. Springer, 2014.
- DM Wills, CJ Whipple, S Takuno, LE Kursel, LM Shannon, J Ross-Ibarra, and JF Doebley. From many, one: Genetic control of prolificacy during maize domestication. *PLoS Genetics*, 9(6):e1003604, 2013.

## Budget Justification

### Personnel

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

### Other Personnel

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

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

### Fringe Benefits

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

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

### Equipment

No equipment funds are requested.

### Travel

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

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

### Participant Support

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

## Other Direct Costs

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

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

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

**RNA sequencing** In total, RNA sequencing will be performed on 384 individuals (8 inbreds x 2 stages x 2 tissues x 2 environments x 3 replicates + 8 NILs x 2 genotypes x 2 stages x 2 environs x 3 replicates). Cost to prepare RNA libraries in our lab are approximately \$100 per library, and sequencing costs for single-end 50bp reads at the UCD Genome Center are approximately \$1,000 per lane. Multiplexing 12 barcodes per lane, this comes out to 32 lanes of sequence and a total cost of \$70,400.

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

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

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

### **Total Direct Costs**

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

### **Indirect Costs**

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

# Facilities, Equipment, and Other Resources

## Facilities, Equipment & Other Resources

### UC Davis

Dr. Ross-Ibarra has four standard laboratory benches as part of a shared lab space at UCD. The shared space is the single largest lab space on campus, and provides for seamless interaction between the labs housed there. The space currently houses three other PIs, all working on the genetics and genomics of economically important plant taxa (Dubcovsky, Neale, Dandekar). The lab is equipped with standard equipment and tools for molecular biology, including freezers and refrigeration, a shared liquid handling robot, thermal cyclers, centrifuges, gel rigs, balances, and standard molecular biology supplies. A dedicated low-humidity refrigerator for seed storage is available through the university, and low-humidity storage cabinets for tissues and temporary seed storage are in the laboratory. Dr. Ross-Ibarra occupies half of a large office suite that includes a conference room and cubicle space for 25 people. Both Macintosh and PC workstations are available for student and postdoc employees. 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 bioanalyzer and library preparation robots. As a member of the Genome Center, Dr. Ross-Ibarra also has access to their additional computational facilities. UC Davis has also entered into a recent partnership with BGI (formerly the Beijing Genomics Institute) to provide additional high-throughput sequencing services via a new Sacramento-based sequencing facility.

Dr. Coop's dry space is located on the 3rd floor of the Storer building, which houses the Department of Evolution and Ecology. The space is newly renovated 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.

#### **USDA-ARS, Missouri**

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

Langebio's mandate is to conduct top-ranked research while promoting genomic knowledge for the protection and sustainable use of Mexican biodiversity. Its unique location in the agricultural center of Mexico facilitates field sampling and field experimentation. 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.

SEE APPENDIX A-1 UPLOADED AS A SUPPLEMENTARY DOCUMENT

## Supplementary Documentation

### 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 NSF-funded 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 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. All of our institutions have seminar series specifically for postdoctoral and graduate students to practice presentation skills; members of our labs will be encouraged to attend these.

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

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

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

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.



# Supplementary Documentation

## Sharing of Results and Management of Intellectual Property

### Data Types

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

### Data Access, Sharing

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

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

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

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

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

All 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.

## **Data Archiving**

All data, code, presentations, and publications will be made publicly available online (see above). Prior to public release, all data will be hosted locally. Dr. Ross-Ibarra will maintain a backup of all raw genotyping, sequence, and phenotyping data. His lab maintains a DROBO distributed backup server (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.

Both our F2:3 families and our near isogenic lines will require multiple generations of development until they are mature resources for mapping traits related to highland adaptation. We will archive a sample from each generation of population development in temperature- and humidity-controlled facilities at Iowa State University and Langebio. Sample accession data will be securely stored in a MySQL server hosted at the University of California, Davis and backed up on a weekly basis offsite. International agreements prohibit some of the maize and teosinte germplasm collected in Mexico from being stored and distributed by USDA. We will, however, deposit small quantities of seed from all our collections with the CIMMYT germplasm bank in Mexico, and deposit samples of our mapping populations (F2:3 seed) in the USDA-ARS Maize Stock Center at the University of Illinois. Both centers provide public access to seed.

# Supplementary Documentation

## Management Plan

### Communication

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

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

### Outreach

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

Dr. Flint-Garcia will coordinate the annual phenotyping workshop, held each year in Columbia. The workshop will be timed to coincide with data collection at the end of the field season each year. The workshop will be advertised broadly (evoldir list-serv, maizegdb, etc.). Attendees will be expected to pay their own travel and purchase a handheld device

### Research

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

- Angelica Cibrian Jaramillo: 5%
- Graham Coop: 5%
- Sherry Flint-Garcia: 10%
- Matthew Hufford: 15%
- Jeffrey Ross-Ibarra: 10%
- Ruairidh Sawers: 15%

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

Year	1	2	3	4	5
?? QTL mapping	SFG, RS, JRI	SFG, MBH, RS, JRI	SFG, MBH, JRI	SFG, JRI	SFG
?? Admix mapping	MBH	MBH, GC	MBH, RS, GC	MBH, GC	–
?? Population genetics	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC	JRI, MBH, GC
?? Functional analyses	RS	RS, SFG	RS, SFG, MBH	RS, SFG	RS, SFG
?? RNA-seq	–	ACJ	JRI, ACJ, RS	JRI, ACJ	JRI, ACJ

Table 1: Proposed timeline of activities showing which team members will be responsible for each objective. Team member names are abbreviated: MBH, Matthew Hufford; JRI, Jeffrey Ross-Ibarra; SFG, Sherry Flint-Garcia; GC, Graham Coop; RS, Ruairidh Sawers; ACJ, Angelica Cibrian Jaramillo

## Year 1

?? SFG will generate seed of F2:3 for Mexico and F2 for S. American cross. JRI will sequence parents of both crosses. RS will choose highland site.

?? MBH will collect seed from Ahuacatitlan. GC will focus on developing methods for admix mapping.

?? MBH will collect seed from additional admixed populations. JRI will genotype samples from highland Mexico maize. AJC and MBH will put together dataset of global highland maize. GC will work on methods for selection in admix populations.

?? RS will screen HIFs and advance the population. Initial crosses for allelic series will be performed.

## Year 2

?? SFG and RS will grow mapping population at each of 3 locales. SFG, RS, and MBH will phenotype populations in field. SFG will genotype F2 plants. JRI will phenotype root chilling.

?? MBH will genotype samples. RS and MBH will grow samples at two locations. GC will begin data analysis.

?? JRI will genotype seed from additional admix populations. MBH will genotype global highland maize collection. JRI and GC will begin data analysis of introgressed highland maize.

?? RS will advance the HIF and NIL populations.

?? ACJ will choose NILs for RNAseq analysis

### **Year 3**

---

?? SFG and RS will grow second replicate of mapping population at each of 3 locales. SFG, RS, and MBH will phenotype populations in field. SFG and JRI will build map and begin QTL analysis.

?? GC and MBH will complete data analysis and begin writing.

?? JRI and GC will work on data analysis of admixed teosinte and highland Mexico maize. MBH will begin data analysis of global highland maize.

?? RS, SFG, and MBH will grow and phenotype HIFs and NILs. RS will advance both populations.

?? ACJ and RS will grow NILs and donors in highland and lowland environment. ACJ will extract RNA. JRI will perform RNAseq library prep and sequencing.

### **Year 4**

---

?? SFG and JRI will perform QTL analysis

?? GC and MBH will write paper.

?? JRI and GC will finish data analysis and begin papers for admixed teosinte and highland Mexico maize. MBH will finish analysis of global highland maize.

?? RS will advance both populations. SFG and RS will analyze data.

?? JRI and ACJ will analyze RNAseq data.

### **Year 5**

---

?? SFG will write paper.

?? MBH, JRI, and GC will write papers.

?? RS will advance both populations. SFG and RS will write paper.

?? JRI and ACJ will write paper.

# Supplementary Documentation

## Plans for Undergraduate and Graduate Student Mentoring

### Undergraduate Students

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

### Graduate Students

The current proposal requests funding for graduate students only at UC Davis, although it is hoped that additional students will participate in this grant through other funding mechanisms (institutional support, competitive fellowships, etc.). Students will be trained in order to prepare them for research careers (academic or otherwise). All students will be expected to take part in internal lab meetings (the Coop and Ross-Ibarra labs at UC Davis hold joint lab meetings) at which they will be given numerous opportunities to hone their presentation skills. The Coop, Ross-Ibarra and Hufford labs currently host weekly journal clubs in which students gain additional training in reading, presenting, and dissecting scientific literature. All 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. 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. Graduate students on the grant will be expected to produce first-author papers for peer-review as part of their project, and encouraged to contribute to additional papers as middle author. Students will be expected to attend and present a poster or talk at a scientific conference each year; UC Davis provides several opportunities for travel funds to support students in this manner. Finally, issues of ethics and organization will be included in training. These will include authorship, reproducibility, and basic scientific ethics. For example students will be encouraged to pursue open science, including the submission of preprints and pre-publication data release and students will be required to maintain Github repositories of their computational work to ensure reproducibility and transparency.