

GEPR: Genetic Architecture of Maize and Teosinte

Edward Buckler (PI-USDA-ARS, Cornell Univ.)

John Doebley (Univ. of Wisconsin)

Sherry Flint-Garcia (USDA-ARS, Univ. of Missouri)

Jeff Glaubitz (Project Manager)

Theresa Fulton (Outreach; Cornell Univ.)

James Holland (USDA-ARS, North Carolina St. Univ.)

Stephen Kresovich (Cornell Univ.)

Michael McMullen (USDA-ARS, Univ. of Missouri)

Qi Sun (Cornell Univ.)

Scientific Objectives and Approaches

Genetic architecture is the constellation of gene effects and interactions that underlie variation in a quantitative trait. Essentially, genetic architecture is the map between phenotype and genotype. Understanding variation in genetic architecture is key to understanding evolution, manipulating species for a sustainable agriculture, and preserving variation as species adapt. This project will strive to improve our understanding of the genetic basis of complex traits for maize and its wild relative, teosinte. Maize has a combination of life history, economic and societal value, and genetic tools that make it uniquely suited to studying genetic architecture.

This project will identify genes controlling domestication traits and the key agronomic traits, flowering, plant height, and kernel quality by employing linkage, association, and fine mapping approaches on the largest-most diverse set of mapping families publicly available for any species. A large series of isogenic lines will be used to characterize allelic series and epistatic interactions. We will compare and contrast genetic architecture for each of the traits and relate the functional diversity to recombination and domestication bottlenecks. Finally, we will evaluate the ability of genetic architecture-based models to predict phenotype in a range of germplasm including elite US hybrids.

Broader Impacts

Maize is the largest production crop in the world, and it plays a central role in all of US agriculture and food production. Maize also has the greatest molecular and phenotypic diversity among crop species, and this diversity enabled domestication and is key for future maize improvement. Understanding maize genetic architecture will aid in the selection and development of future crops. In addition, this project will generate valuable germplasm resources and develop genomic information regarding maize and teosinte diversity that will be used by many other groups to dissect numerous other traits and facilitate marker assisted breeding, allele mining, and genetics. These resources will be made available through a project website, integration with community websites, and stock centers.

Maize is also an excellent system for teaching about evolution, genetics, and agriculture. Our outreach will target four audiences: (1) The general public and students through a traveling museum exhibit on maize domestication and diversity; (2) High school teachers through an enrichment course with North Carolina Agriculture & Technical State University; (3) Collaborative science through African Scientist Fellowship through Cornell's Institute of Genomic Diversity, and (4) Undergraduate students through mentoring and research opportunities.

RESULTS UNDER PRIOR NSF SUPPORT:

Since the NSF Maize Diversity project started in 1999, our group of PIs has published 179 papers on maize diversity topics leveraging project data and concepts. These results have been cited 4152 times with 1108 citations in 2007. Most importantly, we think the best is yet to come.

NSF DBI-0321467: MOLECULAR AND FUNCTIONAL DIVERSITY IN THE MAIZE GENOME; JAN 2004-DEC 2008 (PIs: DOEBLEY, BUCKLER, GAUT, GOODMAN, HOLLAND, KRESOVICH, MCMULLEN, STEIN, AND WARE)

The main project goals were to understand the molecular diversity of maize and teosinte, develop resources for functional trait dissection, map QTL across maize and teosinte, and evaluate functional variation at candidate genes. Our project has succeeded in all of those areas, resulting in 61 publications in the last 4 years¹⁻⁶¹. The current proposal is essentially a renewal of this project (DBI-0321467), although the lead institution is changing.

Molecular Diversity: A major goal of the prior Maize Diversity Project was to describe the nucleotide variability in *Zea mays* and determine how the demographic history of maize, particularly artificial selection during domestication and subsequent maize breeding, has shaped the genetic diversity. Specific benchmarks were to characterize diversity in 3000 random and 1000 candidate genes, model demographic history, and identify genes that contributed to domestication and crop improvement. All goals were met or exceeded. By developing novel simulation approaches and measuring diversity in teosinte and maize, we demonstrated a two-class demographic model for maize²⁰, involving 1) a domestication bottleneck which reduced genetic variability of all genes by ~40% and 2) artificial selection which reduced variation to near zero in about 2% (1000) of maize genes. These results were extended by identifying additional selection candidates from genes with low diversity in maize inbreds²¹. We used stratified samples of teosinte, maize landraces and maize inbreds to pin-point the timing of selection of individual genes to domestication or crop improvement^{21, 45, 60}. By comparing expression levels of selected and neutral genes, we provided evidence that selection acted on genes expressed in the maize ear^{45, 60}. A product of these studies was an extensive SNP library that has been used for both QTL and association analysis^{36, 47, 48, 54, 58} (plus many in preparation).

Functional Diversity Resources:

Maize: Our goal was to develop 25 linkage populations that would capture a large proportion of maize diversity and be useful for both linkage and association mapping. These are collectively called the Nested Association Mapping (NAM) population, as its greatest strength is in association mapping across the populations⁵⁸ (see below). We exceeded the goal of developing 200 RILs (5000 RILs total) from each cross between B73 (the reference parent) and 25 diverse inbred lines. We are currently preparing bulked seed for distribution by Maize Stock Center in spring 2008 with problematic RILs to be deposited by the end of 2008.

We genotyped 6911 NAM RILs with 1536 SNPs (exceeding the goal of 300 markers), and quality control analysis resulted in eliminating 5% of the lines. A composite NAM map (combined across the 25 populations) was constructed for 4699 RILs and 1106 loci, resulting in a total map length of 1400 cM and an average density of one marker per 1.3 cM. Preliminary tests indicate the NAM map will anchor many currently unassigned BAC contigs and resolve numerous questionable contig orders of the current maize physical map (maizesequence.org). Although still under refinement, a preliminary version of the map and map-scores were released through PANZEA (www.panzea.org) in December 2007 and MAIZEGDB in January 2008.

We have conducted simulation studies that describe the power of joint linkage-association mapping in NAM⁵⁸ and evaluated the power to detect epistatic interactions⁴⁷. These

studies suggest that over 20 QTL per trait should be resolved to our LD decay limits (2000bp) given complete knowledge of founder genotypes (full sequencing of the parents).

Maize-Teosinte: Two maize-teosinte mapping populations that capture maize-teosinte domestication QTL alleles were created: (1) 1057 maize-teosinte NILs, each possessing a single teosinte chromosome segment in the W22 background (BC6S2). Genotyping is in progress (50% complete) and anticipated date of seed deposition at the Stock Center is June 2008; (2) ~825 maize-teosinte RILs (BC2S6) are under development and will be finished in Summer 2009.

Other: We developed three other resources: (1) twelve teosinte inbreds were created and deposited with NCRPIS in 2007; (2) seed of 94 teosinte populations was collected for use in association mapping and deposited with NCRPIS in 2007 in the form of a kit of the 94 populations; and (3) 27 diverse landrace inbreds (S5 or S6) are under development and will be deposited at the NCRPIS in May 2008.

	GERMPLASM	CURRENT PHENOTYPES	CURRENT GENOTYPING ^c	PROPOSED RESOURCE ^d
QTL Discovery	25 Maize NAM families; 5000 RILs IBM ^a ; 300 RILs	20+ traits	1106 SNPs on each RIL	5.2,5.6
	2 W22-teosinte BC1 families; 1749 & 1307 individuals	22 traits	300+ markers	
	W22-teosinte BC2S6 population; 825 RILs	22 traits	400+ markers	5.2
QTL Dissection	25DL Maize NILs; 7500 BC5S2 NILs in B73 ^b		128 SNPs	5.5
	W22-teosinte NILs; 1057 BC6S2 NILs		120 markers	5.2,5.5
	B73-teosinte NILs; 900 BC4S2 from 10 teo. in B73			5.2,5.5
Association Panels	Maize: 282 inbreds, 96 PVP, 26 LR inbreds	60+ traits	3500 SNPs	5.2,5.4
	Teosinte: 94 populations and 12 inbreds	13 + 33 traits	500+ SNPs	

a. IBM was developed by other groups and has been extensively genotyped⁶²; b. Donation of Syngenta (see letter), used the same 25 diverse line (DL) founders as NAM; c. SNP genotyping reported here, founders lines have been sequenced for up to 4500 amplicons and are now being NextGen sequenced. d. This germplasm will have additional development or genotyping in this proposed project (see Trait Dissection Resources Section 5).

Large-scale QTL mapping in maize and maize-teosinte: We conducted large scale QTL mapping experiments to discover the maize and teosinte regions harboring functional variation. Together these constitute the largest set of QTL mapping experiments ever conducted, and they provide an excellent starting point for the experiments proposed here to resolve many of these QTL to single genes and Quantitative Trait Nucleotide (QTN).

Maize-Teosinte QTL: Large-scale maize-teosinte QTL mapping experiments were performed in two BC1 families of 1749 and 1307 plants, made by crossing W22 to two different teosintes. Large sample sizes and good marker coverage (over 300 per family) provided considerable power for QTL detection. To date, these experiments mapped ~406 QTL, many within 1-LOD support intervals of 5 cM or less³⁹. An outcome of this research was the development of a permanent set of RILs for fine-mapping maize-teosinte QTL to single genes (to be pursued in the current project). We also phenotyped 1082 F2 plants from the cross of the primitive, small-eared maize race Pollo and teosinte, segregating mainly for domestication differences. We anticipate completing analysis and submitting a paper by the end of 2008.

Maize QTL: We and collaborators collected phenotypic data on the NAM + IBM populations + Association Panel for more than 20 traits in up to 11 environments, with heritability for many traits at ~95%. With recent completion of the NAM map, we have initiated QTL mapping of these traits. We have mapped QTL for each family separately using composite interval mapping (CIM)⁶³ and inclusive CIM (ICIM)⁶⁴. For most complex traits, well over 40 QTL can be found across all families. If one were to examine each environment (11) and population (26) separately,

roughly 10 QTL per population are found, so by the end of 2008 we expect to have mapped ~50,000 QTL peaks. The focus is now on how those QTL are shared among populations.

We have developed modified ICIM and GLM-based approaches for joint linkage mapping across all 26 families, focusing on flowering time, and finding 50+ QTL across all the families. These QTL explain over 85% of the genetic variance. The current analysis suggests two very interesting findings: (1) Many of the QTL are shared among the families at roughly 20-30% frequency; however, (2) most of the QTL exhibit allelic series with both positive and negative effects relative to B73. This suggests that even for a complex trait like flowering time⁶⁵, a modest number of genes (<100) control most of the variation, but there are multiple functional polymorphisms per gene. Few of the key flowering genes from *Arabidopsis*^{66, 67} and rice^{65, 68} appear to be important in controlling maize variation (~2 of top 10 QTL).

Vgt1 is the only maize flowering time QTL discovered by fine mapping^{69, 70}, and represents a useful test of the power and resolution of NAM. *Vgt1* required over a decade to clone by very talented groups. Its analysis was complicated by close linkage to another flowering time QTL (*Vgt2*)^{69, 71} and by the 70kb distance between the functional polymorphism (*Vgt1*) and the coding gene that it regulates (*Rap2.7*). In the initial NAM analysis, we detected very strong QTL in *Vgt1-Vgt2* region in numerous populations, but the gene effects were conflated. When we scored additional markers on NAM lines with recombinations in this region, joint linkage mapping separated these QTL. We currently estimate that many QTL could be mapped to 0.3 cM (~10 genes) resolution given enough informative markers. Further, the joint linkage and association NAM analysis revealed a novel *Rap2.7* allele in tropical germplasm that confers later flowering (Fig. 1).

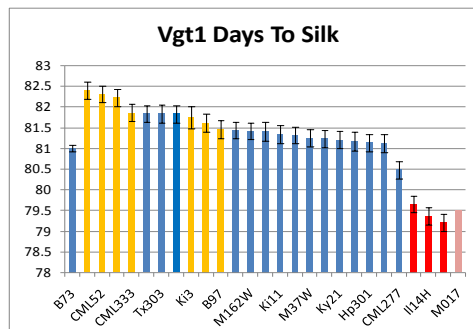


Figure 1. Mean silking dates of founder parent *Vgt1* (*Rap2.7*) alleles segregating in the NAM population. Red bars represent a northern-derived allele⁷⁰, blue bars are reference allele, and yellow bars are a novel tropical-derived allele. Scoring genotypes directly at *Vgt1* and *Rap2.7* on the founder lines resulted in very strong associations with the QTL.

Our results suggest that NAM has tremendous power to dissect complex traits; however, some of our initial assumptions about genetic architecture may be naïve. The high sharing of QTL among populations means that there should be more statistical power and resolution to fine map more QTL via NAM, while the complex allelic series at each QTL and gene suggest that haplotype-based statistical approaches are needed. We think this can be accomplished through genotyping NAM at 10,000 markers and then association mapping using haplotype-based approaches with both NAM and our diverse maize association panel.

Candidate gene association analyses: One of the greatest problems with association mapping has been preventing false positives produced by population structure^{72, 73}. We developed a practical mixed model statistical approach that controls for population and family relatedness using molecular markers^{36, 74}, which provides a robust method for controlling for nearly any type of relatedness (plant, animal⁵⁰, selfing⁶⁷ and outcrossing). This project made association analysis widely available for the maize community by establishing a diverse association panel, depositing seed in the stock center, molecularly characterizing the relatedness of lines in the panel, and estimating the heritabilities and phenotypes for over 60 traits¹¹ (PANZEA). This project used the panel for testing effects of candidate genes in several pathways:

- 1) Starch synthesis and starch quality⁹; Functional SNPs defining sweet maize throughout the Americas³⁴
- 2) Polymorphisms controlling provitamin A production; markers from the project are in use by international agriculture NGO for improving maize nutritional quality⁵⁴.
- 3) With Max Planck collaborators, we found tremendous variation and association for enzyme activity in central carbon metabolism and productivity traits⁵³

This panel is also being used by numerous groups (some in collaboration) for evaluating flowering, aluminum tolerance, drought, nitrogen use efficiency, micronutrients, telomere length, aflatoxin resistance, leaf blight resistance, drought tolerance, and inflorescence architecture.

Two candidate gene association studies in natural populations of teosinte were completed. One demonstrated that major regulatory genes in maize contribute to natural genetic variation in teosinte⁴⁸. In the second, larger-scale study (Weber *et al.* in preparation), the functions of domestication genes lacking variation in maize were determined in wild teosinte.

Bioinformatics: In conjunction with GRAMENE, the GDPDM database schema was developed to hold a wide range of genotypic and phenotypic data types (this schema has been adopted by rice, wheat, and sorghum communities). At PANZEA, we have provided several dynamic tools for users to access desired portions of our data, including Germplasm, Gene/Locus, Molecular Diversity and Phenotype searches. Our sequence alignments are displayed with a custom tool, LookAlign⁷⁵, which highlights potentially useful SNPs. Our sequences can be searched with a BLAST tool. Our genetic and physical maps are displayed via CMap. A custom search was developed to find SNPs or SSRs markers polymorphic between two lines. For germplasm with associated geographical coordinates (maize landraces and teosinte collections), a tool was developed to display their geographical locations as well as allele frequency pie graphs for SNP, SSR or isozyme markers. A Java-based advanced search Tool (GDPC⁷⁶) is available for streamlined retrieval of data with complex queries. For the advanced user, our entire MySQL database (Aztec) can be downloaded from our Datasets Downloads page; various custom datasets (genotypes, SNP marker context sequences, etc) are also available there. Project publications and maize diversity literature pages have been maintained, along with other informational pages. Our development of PANZEA has been highlighted in two publications^{37, 77}.

The statistical approaches developed by this project were rapidly implemented in open source software TASSEL (at sourceforge)³⁸, MTDFREML extensions⁵⁰, and also in SAS.

Outreach: Outreach efforts included the development of storyboard displays for use as classroom teaching aids. Four storyboards were developed, on Maize Domestication, the History of Maize, Barbara McClintock and Biofuels. Our outreach coordinator used these storyboards in visits to middle and high schools in WI, CA and MO, where she also engaged students with hands-on activities such as DNA extraction. She also used these materials in annual visits to Cornell of high school students from Newfield, NY. She made two visits to NC A&T, where she taught summer courses on Agricultural Genomics to high school teachers. PowerPoint versions of storyboards and their Spanish translations are accessible from PANZEA and have been used widely by educators. PIs have mentored over 30 undergraduates in their laboratories via this project, and many are continuing careers in science. Finally, PIs and the work of this project were featured in the NSF educational video “Secrets of Plant Genomes Revealed!”

NSF DBI-0638566: HIGH DENSITY SCOREABLE MARKERS FOR MAIZE TRAIT DISSECTION; NOV 2006-NOV 2008 (PI: BUCKLER & WARE)

The goal of this project is to evaluate next generation sequencing technologies for SNP scoring and library construction, and to score SNPs at high density across the founders of NAM. This

project has been successful at developing an *HpaII* based methyl-filtration approach for efficiently accessing the gene rich fraction of the maize genome. 454 sequencing of methyl-filtered B73 and Mo17 libraries revealed that they are 70% low copy and cover about 50-100 Mbp. of sequence. Even at this low coverage, over 100,000 SNPs could be called against the reference genome sequence (Gore, in review), and recent maize genome sequencing has facilitated differentiating paralogs from alleles⁷⁸. For production genotyping, we are using Solexa sequencing of these libraries. We have generated 9 billion bp of sequence across the 27 founders (roughly 3X per founder, 81X total). This will generate from 1-4 million SNPs by the end of 2008. Although not originally part of the goals of this project, we will paired-end sequence randomly sheared libraries from each NAM founder to about ~1X. This will allow copy number differences to be identified, which will be validated and genetically mapped by scoring CNV in NAM RILs with custom Nimblegen 2.1 million feature arrays. Additionally, we will obtain *HpaII* library sequence from 12 teosinte inbreds by the end of 2008 or early 2009.

RELEVANCE AND JUSTIFICATION:

Living organisms display tremendous natural genetic variation for development, chemical composition, and adaptations to specific environments. Understanding the nature of this genetic variation is key to understanding evolution⁷⁹, manipulating species for a sustainable agriculture, and preserving inter- and intra-specific variation through our current period of rapid environmental change. Although most detailed molecular genetic analyses have focused on genes of major effect, almost all standing genetic variation within natural and breeding populations is controlled by numerous genes that individually control less than 10% of the variation in species^{80, 81}. This project will strive to improve our understanding of the genetic basis of complex traits. Maize (*Zea mays* ssp. *mays*) and the teosintes (seven other named taxa in the genus *Zea*, including *Z. mays* ssp. *parviglumis*, the closest wild relative of maize) have a combination of life history, economic and societal value, and genetic tools that together make *Z. mays* uniquely suited to studying genetic architecture.

Genetic architecture is the constellation of gene effects and interactions that underlie variation in a quantitative trait⁸²⁻⁸⁵. Essentially, it is the map between phenotype and genotype. Using the powerful set of genetic tools developed over the past decade by the members of this project along with others in the maize community, the goal of this new project is a comprehensive “phenomic” analysis of four complex traits: domestication syndrome, flowering, plant height, and kernel quality. Specifically, we will identify genes that control these traits, determine the effects of the series of polymorphisms at these loci, examine epistatic interactions between these genes, and evaluate the interaction of the genetic architecture with environment. With an enhanced understanding of the genetic architecture for these four diverse traits in maize, we will compare and contrast the architectures of maize and teosinte and relate their functional variation to basic genome structure and to the population-genetic consequences of the domestication bottleneck. We will address the extent to which modern functional variation is a product of standing variation that evolved over a several million year time scale versus mutations arising during domestication over the last few thousand years². Finally, we will use this understanding of genetic architecture to evaluate our ability to predict phenotype from genotype, the fundamental goal of modern genetics⁸⁶.

Maize and the teosintes are an ideal system for understanding these aspects of genetic architecture. In coding regions, the average nucleotide diversity between any two maize lines ($\pi=1-1.4\%$)^{20, 87} is similar to the divergence between humans and chimps⁸⁸, and even greater

diversity (40% more) is harbored within *ssp. parviglumis*. It is not uncommon to find maize haplotypes that are 5% diverged from one another⁸⁹, indicating that the maize gene pool reaches back 2-4 million years or generations. In longer generation species (e.g., poplar⁹⁰), this would be the equivalent of tens of millions of years of evolution. *Essentially, population genetics in maize is equivalent in terms of diversity to deep evolution in many other clades of plants and vertebrates*. The advantage of maize and teosinte is that the entire range of diverse germplasm is interfertile, and seed stocks can be created, stored, and shared efficiently. Maize and teosinte are also predominantly outcrossing species, constantly producing new allelic combinations. The constant creation of new epistatic combinations, exposure to diverse environments, and relatively high migration rates between maize and some teosinte populations⁹¹ strongly influenced the genetic architecture of *Z. mays*. This type of genetic architecture is likely to be shared with most plants (80% are outcrossing)⁹² and animals, but it could be quite different from the genetic architecture of inbreeding species exemplified by *Arabidopsis* and rice.

The phenotypic variation of maize mirrors the molecular variation. Maize and its progenitor, teosinte, represent the greatest morphological divergence between any crop-ancestor pair, making them the most interesting and powerful system for the investigation of morphological change during domestication². Even within the domesticated subspecies, abundant phenotypic variation is observed for many traits⁹³. Through heritable changes in flowering time, responses to photoperiod and temperature, and plant architecture, maize has adapted to extremely diverse environments ranging from Northern Europe and Canada to the lowland tropics and to the high Andes. Maize varies in height from less than 1m to over 6m and ranges tremendously in biomass and carbon allocation. Similarly, maize kernels vary ~10 fold in their composition for oil and protein content⁹⁴. This heritable variation is the basis of worldwide maize breeding efforts to improve the maize grown on 144 million ha worldwide. In the US, maize is grown on 34 million ha, and is a base commodity crop supporting the \$1-2 trillion US food industry by providing protein, oil, and starch for food, animal feed, ethanol, and other bio-based products. The only way to sustainably increase yield without increasing inputs or area under cultivation is to continually improve the genetic adaptation of maize to modern farm agroecology and changing climate conditions⁹⁵. Understanding genetic architectures sufficiently to predict phenotypes will be key to increasing maize yield and sustainability.

Importantly, maize now has the some of the best genetic tools to conduct this research. Through our prior research efforts and recent contributions (e.g., a vast set of introgression lines donated by Syngenta to our project), we have nearly 15,000 genetic stocks that permit manipulation and isolation of the genetic variation throughout the entire species. With completion of the maize genome sequence in 2008 and the availability of next generation sequencing of maize diversity, it is now possible to use the full range of genomic tools to examine the genetic architecture of maize⁹⁶.

PROJECT GOALS:

The proposed project has two parts: First, research on genetic architecture (Sections 1-4); Second, development of germplasm, genetic, and bioinformatic resources so that any trait can be dissected (Section 5). Obviously, these resources are needed to accomplish our scientific aims, but they will also be of tremendous value to maize community for other projects. Our specific aims are:

1. Identify genes controlling domestication, flowering, plant height, and kernel quality.
2. Molecular characterization of QTL.
3. Analysis of genetic architecture with recombination and bottlenecks.

4. Evaluate the ability of the genetic architecture to predict phenotype.
5. Develop trait dissection resources:
 - 5.1 Project genotypic data from next generation sequencing onto mapping populations.
 - 5.2 Genotype core maize and teosinte germplasm with SNPs at high density.
 - 5.3 Develop analysis tools, bioinformatics displays integrated with GRAMENE.
 - 5.4 Expand core association germplasm to breeding lines, landraces, and teosinte.
 - 5.5 Develop a diverse indexed set of near isogenic lines.
 - 5.6 Grow NAM population each year for community access.

PLAN TO INTEGRATE RESEARCH AND EDUCATION: Our outreach will introduce evolution, genetics, and agriculture to four target audiences through a range of activities.

PROJECT PLAN

1. Identify the genes, alleles, and interactions controlling domestication, flowering, plant height, and kernel quality. Currently, we know the identity of three major and perhaps six minor QTN controlling these traits, but the goal is to identify as many genes and QTN as possible underlying these traits and synthesize the results into a species-wide understanding of trait architecture. Our ICIM QTL mapping to date has identified more than 50 QTL for some of these traits, and we will use joint linkage-association^{97,98} and fine mapping approaches to isolate the most important genes. These genes will be characterized in detail for allelic series and epistatic interactions. We have chosen to focus on these four traits because they evolved under differing selection pressures, are agriculturally important, and we have the capacity to evaluate them accurately on very large sample sizes. These traits will serve as models for future analyses of more phenotypically-difficult traits such as biotic or abiotic stress responses.

1.1 Joint population-association analysis: We have successfully developed materials and methods for joint population linkage mapping. Where we have association data (e.g. *Vgt1*), joint population association data clearly resolve complex QTL to the single gene level. By combining existing data across all resources and saturating the genome with markers, we will be able to implement genome-wide association analysis to understand species-wide allelic variation.

1.1.1 Map Integration: The current NAM genetic map is comprised of 1106 SNP loci, mapped on 4699 NAM RILs. The resolution of the NAM genetic map is 5 times greater than that of the IBM population currently used to anchor the physical map of maize. The NAM map allows an entirely independent physical map assembly by matching the original sequences used for SNP discovery to the sequenced BACs, permitting resolution of anchoring conflicts and detection of errors in contig alignments. Recently, we mapped 772 of the SNPs from the NAM map onto 286 IBM individuals to reconcile the genetic and physical maps. The NAM-anchored integrated genetic/physical map will serve as a framework for integrating the other genetic resources such as the maize-teosinte RIL and NIL populations and populations developed by other groups. Although marker order is consistent across populations, the recombination distances vary tremendously at the fine scale (up to 10X). To make the best quality joint map for QTL analysis, we will score our biparental populations with 10,000 SNPs (5.2) and rebuild the integrated map. Using sequence comparisons and functional gene classifications, we will incorporate gene annotations in the map. The final product will be a high resolution genetic map anchored to the maize genome sequence. This resource will contain diversity data from the 25 NAM founders and teosinte, plus genome annotation from maize and other species.

1.1.2 Joint Linkage QTL mapping: We have successfully used two QTL mapping models for NAM analysis. One approach is linear model selection with unique allele effects estimated

within each of the 26 populations. This model is robust and identifies up to 40 QTL per trait, although it does not use all the recombination data and may not resolve tightly-linked QTL. The second approach is modified Inclusive Composite Interval Mapping⁶⁴ for multiple populations (collaborator, Huihui Li), which has shown good performance, but exhibits some inconsistencies due to recombination differences between populations. The two approaches produce congruent results for many QTL (1-2 cM mapping resolution), and we will use these algorithms initially for joint population linkage mapping. When the intervals are nearly saturated with markers (5.2), the recombination events will be completely informative, differential recombination issues will disappear, and simpler algorithms can be used. We expect resolution of ~0.3-1 cM intervals (very roughly 10-30 genes). Multi-population QTL mapping is an active area of research, and we will use improved algorithms as they become available including Bayesian approaches⁹⁹⁻¹⁰⁸. Phenotypic data from up to 11 environments have already been collected for flowering, plant height, kernel quality (NIR underway), and domestication traits. The phenotype data set includes up to roughly 66,000 plot means with entry mean heritabilities of nearly 95%.

It appears that QTL identified within NAM are shared with the maize-teosinte crosses. Using the integrated maps, we will jointly map QTL in the NAM and teosinte populations (comprising nearly 8000 genotypes and 250,000 crossovers). While the domestication trait glume architecture controlled by *Tga* is not shared with NAM, many QTL in maize and teosinte are likely to be the same for other traits (e.g. *Tb1* appears to control tillering in NAM).

1.1.3 Association mapping across linkage populations (NAM) and association panels: Nested association mapping⁵⁸ will be used to resolve QTL positions to genes and potentially QTN, and preliminary evidence suggests it works well at the known gene *Vgt1* (see introduction). Because linkage disequilibrium decays within 2000bp for this diverse germplasm in genic regions¹⁰⁹, association mapping will require high marker density across the founder lines in recombinationally active regions. The Maize Marker Project (see prior support) will produce 0.5-5 million SNPs on the founders, roughly 10-100 SNPs per gene. This will be sufficient for association analysis across the linkage populations for many genes. In those QTL regions with poor coverage, we will use ABI sequencing of PCR products as we have done for 4500 other regions. Basic regression approaches will be used to association map across regions as there is no population structure beyond the well-defined 26 subpopulations introduced by their 26 founders⁵⁸. If multiple functionally different rare alleles are common¹¹⁰, mixed model approaches that relate phenotypic divergence to haplotypic divergence will be performed¹⁰⁵.

Once putative genes and regions have been identified within the segregating mapping populations, association mapping across diverse maize and teosinte germplasm panels will be used to confirm and refine the functional associations. Population structure will be controlled with mixed models³⁶, and results will be combined with the linkage results.

1.2 QTL Fine-mapping: Within the last few years, three maize QTL were positionally cloned^{70, 111} or fine-mapped¹¹². Each required enormous effort and a decade from initial QTL discovery. Such work is critical because it provides formal proof that a QTL corresponds to a specific gene. A full understanding of the inheritance of complex traits and nature of QTN requires we have a larger sample of fine-mapped QTL. We propose generating such a large sample by fine-mapping 22 additional QTL. This proposed fine-mapping plays a powerful complementary role to gene discovery through the high-throughput approaches in 1.1. Fine-mapping will be invaluable for closely coupled QTL, regions with high LD, and association results where the predicted gene ontology indicates a novel genetic factor.

The 22 QTL selected for fine-mapping represent four trait groups: domestication (4 QTL), and flowering time, plant height, and kernel quality (6 QTL each). The criteria for selecting QTL included very significant effects, narrow support intervals (1-3 cM), and consistency across multiple mapping populations. For the domestication traits, the QTL were identified in two large maize-teosinte populations³⁹. They are: KERN3.122f (kernel wt.), SPKLT2.23w (spikelet length), STAM3.66w (inflorescence sex), and TILL5.80w (tillering). For the other three trait groups, specific QTL will be chosen from the top 10 QTL identified by NAM mapping. The key genetic resources for fine-mapping are two sets of NILs (5.5). First, we have a set ~1000 maize-teosinte NILs, each possessing a single teosinte chromosome segment in W22 background. Second, we have a set of 7500 fully genotyped NILs carrying introgressions from the 25 NAM founders plus Mo17 in the B73 background provided to us by Syngenta.

The mapping strategy to generate the recombinants for fine mapping is as follows. NILs carrying the target allele introgression will be backcrossed to B73 or W22. F1s will be selfed to generate about 2000 F2 seeds per selected NIL. The endosperm of the F2s will be genotyped for flanking SNP markers to identify plants with recombinant chromosomes, and the following summer recombinant plants will be phenotyped and selfed to form F2:3 families. The F2:3 families will be planted in the following winter, individual F3s will be screened to identify plants homozygous for recombinant chromosomes and these plants will be selfed. The resulting recombinant chromosome NILs (RC-NILs) will be phenotyped the next summer in 4 reps at 4 locations to determine which QTL alleles they carry. The RC-NILs will be genotyped at a density of at least one informative SNP per predicted gene across the entire interval and QTL alleles (assayed by phenotype) will be mapped against RC breakpoints. Genotyping will be done by TAQman SNP assays¹¹³. Preliminary results indicate a confidence interval for a QTL of 1-3 cM for the target QTL in either the NAM or maize-teosinte populations before fine-mapping. For a 3 cM interval, we expect to recover 120 RC-NIL lines from 2000 initial F2s, and we expect the interval to contain ~90 genes requiring SNP tagging. Sequencing will be used to refine the resolution of critical breakpoints. We will start 8 QTL the first year, 8 QTL the second year, and 6 QTL the third year. While we expect to complete fine-mapping of many QTL during this grant period, we will encounter difficulties for some loci that may cause some delays.

1.3 Evaluate Allelic Series and Epistasis: Our initial analyses of NAM indicate that over 80% of the QTL have both positive and negative effects relative to B73, probably the result of allelic series (i.e. multiple QTN per locus) and epistasis as seen in *Drosophila*⁸⁴. Not surprisingly for such a diverse species, several association studies in maize have found evidence for allelic series⁵⁴. Epistasis for quantitative traits is difficult to study because of the requirement for very large population samples^{114, 115}. The NAM experiments conducted thus far provide some of the best data sets for evaluating the importance of epistasis in controlling complex traits because of their very large sample size and broad genetic inference space.

1.3.1 Evaluate Allelic Series by NILs: We will perform independent validation tests of the effects of alleles of diverse origin for 10 important QTL identified per trait from the NAM experiment. For each selected QTL, we will test a set of NILs carrying introgressions in the target QTL region from 10 different maize and 10 teosinte parents. Each NIL set will be evaluated along with the B73 control in 4 replications in 4 environments (QTL for different traits will be tested in different years). Within a NIL set representing alleles at a single QTL, we will estimate the minimum number of allele effects at a locus among the 20 sampled *Zea* founders by statistical tests for the number of phenotypically different NIL groups. In addition, we can test if the multiple allele effects predicted in the original NAM experiment are also observed in the NIL

series. Finally, we will be able to determine if the range of allele effects in maize is significantly smaller than the range observed from teosinte introgression NILs.

1.3.2 Evaluate Epistasis in Biparental Populations: We propose to test five hypotheses concerning epistasis using the NAM data on the four selected complex phenotypes. (1) The frequency and magnitude of epistatic effects are similar to that of additive effects. (2) Epistasis primarily occurs between QTL that exhibit significant additive effects. We will test these first two hypotheses by extending the additive NAM models to include digenic epistasis among QTL detected with additive effects. We will compare these results to models including both additive effects and digenic epistasis developed using two-dimensional searches among all possible pairs of loci and standard linear model selection techniques^{116, 117}. (3) The importance of higher-order epistasis is similar to that of digenic epistasis. We will test this by modeling three-way epistatic interactions⁴⁷ and compare the importance of 3-locus epistasis, 2-locus epistasis, and additive effects. (4) Epistasis can be adequately modeled with linear statistical models. This will be tested by using alternative analysis techniques, such as Bayesian epistasis mapping¹¹⁸, and recursive partitioning or multifactor dimensionality reduction¹¹⁹. Cross-validation will be used to test the validity of each model. (5) Higher-order epistasis causes the appearance of allelic series by shifting the additive value of alleles across populations. This hypothesis will be tested as part of section 1.3.1, in which the effects of the allelic series predicted by the original additive NAM analysis will be tested in NIL series representing the selected genes and alleles. If the allelic effects predicted from the additive NAM model are not validated, epistasis between the selected genes and the genetic background could be a cause. If this occurs, it will be followed up by an intensive search for digenic and higher-order interactions involving the selected loci.

Experimental validation of digenic epistasis will be conducted on 10 important two-locus epistatic interactions for each of the 3 agronomic traits. NILs will be crossed, selfed, and F2 seeds genotyped using an endosperm genotyping system (N. Coles lab methods). Once doubly homozygous seeds are found, seed will be bulked. Additive by additive epistasis will be tested by growing the double NILs, along with the single NILs and duplicate B73 entries, in replicated field trials (4 replications at each of 3 environments). This process of generating the stocks will take two years, and they will be evaluated in the final two years of the project.

1.4 Evaluate genotype-by-environment interactions: Genotype-by-environment (GxE) interactions represent another important complication of genetic architecture that critically impact both evolutionary biology¹²⁰ and plant breeding¹²¹. GxE observed in the NAM dataset, currently composed of 11 diverse testing environments, will be decomposed to specific gene-by-environment interactions¹²². We will determine if GxE is mainly due to a few genes with large interactions masking otherwise stable gene effects, or if they are mainly due to most genes having similar GxE responses. In addition, we will conduct a detailed analysis of GxE for flowering time, using environmental covariates¹²³, particularly photoperiod and temperature, to identify interactions between specific environmental factors and QTL alleles.

2. Molecular Characterization of QTL

The molecular basis of QTL action can be broadly classified as either effects due to regulatory variation (gene expression) or protein functional variation^{124, 125}. Quantitative variation in plants has been associated with both mechanisms⁸², but their relative importance to quantitative variation remains unclear.

2.1 Sequence Analysis: Once specific gene-QTL association has been identified by joint linkage-association mapping (1.1) or fine-mapping (1.2), we will examine sequence alignments in detail to identify potential causative sites that control the phenotypic differences. Candidate causative

sites will be assigned as either coding (causing an amino acid substitution) or non-coding, and sequence analysis will be used to identify putative QTN that modify conserved non-coding sequences^{126, 127}, protein structural changes^{128, 129}, codon bias^{130, 131}, splicing effects^{132, 133}, or could be involved in small RNA pathways¹³⁴⁻¹³⁶. We will also examine copy number variation for these key genes using the Nextgen sequencing, Southern, and linkage analysis.

2.2 Expression: We will assay relative gene expression levels for the 22 fine-mapped QTL (1.2) and the ~80 genes identified by joint linkage-association mapping (1.1), whether the putative causative site is coding or non-coding. We will assay expression in 25 maize × teosinte or 25 NAM founder diverse line (DL) × B73 F₁ hybrids (including reciprocal F₁s) for multiple tissues and time points specific to each trait. In F₁ hybrids, the contrasted alleles are in the same *trans*-acting ‘soup’, and thus differences in expression must result from *cis*-acting variants. We will use allele-specific expression assays to identify abrupt shifts in expression between maize and teosinte alleles or B73 and DL alleles. Allele-specific expression assays will be designed in collaboration with Sequenom (San Diego) using Sequenom MassARRAY technology. Three biological reps of the cDNAs for each F₁ will be assayed. F₁ DNAs and 1:1, 1:2, and 2:1 mixes of parent DNAs will be used as standards¹³⁷.

2.3 Mutagenesis: In conducting the joint linkage-association analysis and fine-mapping, we will undoubtedly identify candidate genes with no known function based on gene ontology. In order to investigate function to these novel genes, we will initiate studies with knockout lines from the Maize TILLING Project, the Maize Targeted Mutagenesis Project, and/or the Uniform Mu Maize Project. The gene knockouts will be backcrossed into B73 during this project, but final molecular and biochemical characterization will be conducted in subsequent projects.

2.4 Distribution of functional polymorphisms: We anticipate identifying the putative functional polymorphisms in up to 100 genes. We will score the distribution of these polymorphisms on diverse stocks to understand their geographic distribution among maize inbreds (282 public + 92 commercial lines, see 4.2 below), landraces (92) and teosinte (92) throughout Mexico. This analysis will help determine when the key polymorphisms arose. For example, is the polymorphism old and segregating among all of *Zea*, restricted to the maize ancestor (*Z. mays* ssp. *parviglumis*), or is it a recent polymorphism only found in modern inbreds?

3. Analysis of genetic architecture with recombination and bottlenecks

The above studies will provide a comprehensive understanding of complex traits; here we propose to synthesize the results and relate them to key genetic, molecular, and evolutionary processes. We will compare genetic architectures, relate functional diversity to recombination and genome structure, and evaluate the effects of bottlenecks on functional diversity.

3.1 Contrast the genetic architectures: We will compare genetic architectures of different traits, evaluating whether number of effective loci and allele effects distributions are similar. With our gene-level resolution of QTL, we can ask how often the same genes underlie diverse traits (pleiotropy). Has the recent directional and diversifying selection of maize worked mostly on novel mutations or on standing variation? Our large sample of genes and traits will go beyond the few current examples to provide a much clearer understanding of the patterns and trends of genetic architecture.

3.2 Genomic constraints on molecular and functional diversity: One puzzling aspect of molecular diversity is its relationship with recombination^{42, 138, 139}. Like many species, maize exhibits higher recombination rates/physical distance near the ends of the chromosomes²². We previously found that SNP diversity was positively related to gene level estimates of recombination, but not to chromosome level differences in recombination observed

cytogenetically^{87, 140}. With the high density diversity measurements and detailed recombination maps provided by NAM, we will have a substantially improved data set to re-evaluate how molecular diversity and linkage disequilibrium relate to recombination rates and gene density along the chromosomes. All diversity estimates and QTL from the prior project and this project will be placed onto the diversity-annotated genetic/physical map (1.1.1). This will allow us to test if functional variation (i.e. the number of QTL alleles) is related to molecular diversity or differential recombination rates, and if domestication and improvement QTL regions are diversity limited in modern inbreds.

3.3 Evaluate bottlenecks on functional diversity: The demographic effect of the domestication/crop improvement bottlenecks was a 40% loss of molecular diversity from teosinte to maize inbreds²⁰. A loss of >95% of diversity occurred in selected genes. We will examine whether there has been a corresponding loss in functional diversity in inbreds relative to teosinte, if the loss of functional diversity is reflected differently in genes that experienced artificial selection, and if teosinte harbors alleles conferring more extreme phenotypes at most genes or only at those that were selected during domestication and/or later improvement.

4. Evaluate the ability of genetic architecture estimates to predict phenotypes.

We will test whether the allele effect predictions from the NAM population can be applied beyond the scope of the original population - a novel approach that is needed for large-scale genetic architecture studies to have broad applicability for crop improvement⁴⁴. We will test three unique sets of maize germplasm: related RIL populations, unrelated commercial and landrace germplasm, and hybrids. For each trait, we will identify a set of QTL that have been resolved to functional haplotypes which demonstrate strong internal LD across the species and compare four different methods to predict phenotypes for each germplasm set: (1) Functional haplotype effects estimated from the NAM experiment; (2) Functional haplotypes identified in NAM and re-estimated in the test data set; (3) Phenotype modeling using random markers via genome-wide selection¹⁴¹ and pair-wise relatedness measures³⁶; and (4) Functional haplotype predictions from parts 1 and 2 combined with the standard approach in part 3. The reliability of the predictions will be assayed with jackknife techniques in each case.

4.1. Test functional haplotype effect predictions in related mapping populations: Are functional haplotypes identified in the NAM population predictive in other crosses involving B73? To test this, we will first jackknife the NAM dataset over populations – in each step, we will exclude one population from the NAM dataset, re-estimate haplotype effects, and use them to predict values in the excluded population. Second, we will evaluate the predictions in additional RIL populations from crosses between B73 and CML254 or Ki14, two lines that exhibit strong photoperiod responses and are not part of NAM. These RIL populations have been phenotyped previously for flowering time and ear/plant height in four to six environments and will be genotyped with the same 1106 SNPs as the NAM population, plus the target functional haplotype-defining SNPs.

4.2. Test functional haplotype effect predictions in public, private, and landrace inbreds: We will score target functional haplotypes in 282 inbred lines representing the global diversity of public maize. Phenotypic data (flowering and height) for this set are available from the same environments as the NAM population. We will broaden the scope of germplasm further by also testing 92 private-sector inbred lines representing the genetic base of modern U.S. hybrids whose plant variety protection (PVP) certificates have recently expired¹⁴² and 27 inbreds representing maize landraces. These 119 lines will be phenotyped along with the 27 founders of the NAM population in four replications in three environments.

4.3 Test functional haplotype effect predictions in hybrids: We will test the predictive value of allele effects from the NAM genotypes using two distinct hybrid data sets to determine the value of functional haplotype predictions in hybrids and to estimate dominance interactions between alleles. This will provide unique insights into heterosis in maize. The first data set includes testcrosses of 281 of the diverse inbred lines described in section 4.2 to the common parent B73. The second set of germplasm will include about 300 F1s made between lines representing different heterotic groups in the commercial expired-PVP set. These hybrids will be phenotyped in replicated trials in at least three environments.

5. Develop trait dissection resources

The research aims above will guide our analysis of the nature of genetic architectures for four important traits. There are numerous other traits and studies to be conducted, and the resources we develop will directly support the research aims of many research groups.

5.1. Project genotypic data from next generation sequencing projects onto mapping populations:

In order to unite our diverse germplasm sets, the parental lines need to be genotyped at the resolution of the LD decay rate, which is about 2000bp in genic regions (much longer in the recombination-limited retrotransposon regions). These data are being generated by a collaborative grant [NSF DBI-0638566; see prior results above]. In the current proposal, we will project these SNPs and existing SNPs from founders onto the NAM and maize-teosinte RILs (analogous to imputation approaches¹⁴³, but simpler since all ancestral alleles are known). High density genotyping (5.2) will be used to verify that these SNPs are being projected to the correct location. We will project data to distinguish paralogs from alleles and to score copy number variation as these genotyping capabilities develop. The curated data will be made available at PANZEA, GRAMENE and MaizeGDB.

5.2. Genotype core maize and teosinte germplasm with high density SNP genotyping: Our NAM RIL populations and maize-teosinte RIL and NIL populations together capture almost 250,000 recombination events (~5/gene). Although NAM and some of the teosinte RILs have been genotyped with 1106 and 420 markers, respectively, we are not able to fully take advantage of all the recombination information. An additional complication with the current level of genotyping is that the majority of SNPs are not informative in all NAM populations so that, while mean interval length is 1.3 cM in the composite NAM map, there are many regions where markers are 5 cM apart within individual populations. Additionally, since recombination rates vary considerably across populations, the populations can only be accurately united with shared markers. Currently, we are saturating our maps in selected regions by sequencing recombinants, but this would be very inefficient if numerous researchers are going to use these resources.

We propose to genotype the NAM (5000 RILs), key maize inbreds, the IBM population, and maize-teosinte RILs and NILs with a 10,640 feature Illumina Infinium array. From the maize genome sequence and SNP data from the NextGen sequencing project, we will design a SNP marker in roughly every 5 genes. At this density, a SNP will be informative for any given population in roughly every 10th gene, and it should be possible to refine linkage-based QTL intervals down to a 5-10 gene resolution. This marker density will also be very useful for refining genome structure and recombination estimates. Since the Infinium assay is semi-quantitative¹⁴⁴, it may be possible to detect and map copy number variation and large deletions.

5.3. Bioinformatics: This project, in conjunction with the Gramene project (Buckler, co-PI & see letter) and USDA-ARS quantitative genetics group (see letter), has developed a robust bioinformatic platform. We will continue to expand its capabilities as outlined below.

5.3.1. Curation: Our groups have established pipelines for genotypic and phenotypic data curation. The GDPDM database schema, developed under a previous grant, has proven to be an accurate and flexible representation of plant genetic diversity. In the next phase, we will expand the GDPDM schema to accommodate new data types from the new next generation sequencing and genotyping. The amount of genotypic data will increase by several orders of magnitude and optimized storage and retrieval approaches will need to be developed. A designated programmer from this project will work together with the GRAMENE team to develop pipelines for periodic data deposit to the GRAMENE diversity database. We will also help curate genetic maps and functional annotations in MAIZEGDB (with USDA-ARS support in Missouri).

5.3.2 Accessibility: We will expand the functionality of the project website PANZEA (www.panzea.org). The GRAMENE project will be responsible for QTL and SNP displays leveraging existing and emerging Ensembl and QTL resources at Gramene, providing integrative views with QTL results from rice and *Arabidopsis*. We will link to these resources so researchers can seamlessly move back and forth between GRAMENE, PANZEA, and MAIZEGDB. PANZEA will provide access to data sets in well-organized flat files, web pages, GDPC, and full GDPDM schema dumps. The database will be updated as large datasets are released (at least twice per year). With future advances in next generation sequencing, we anticipate that the maize community will have information on 25 million common SNPs and indels for key maize and teosinte lines. This project will produce genotypic data on nearly 15,000 RILs, NILs, and association panels. Projection of all SNPs onto these lines will produce 375 billion data points. We will create bioinformatic tools and displays to allow researchers to identify specific germplasm with recombination events or desired allelic combinations in targeted genome regions. Phenotypic data will be displayed with relevant germplasm. We will also explore methods to display genetic architecture (genomic locations, allelic series, epistasis, and GxE).

The PANZEA software will remain open source, and will be implemented on a MySQL database and Linux Apache server. After the project is finished, the database and the PANZEA web site will be maintained by both the Buckler group and the Computational Biology Service Unit of Cornell University (co-PI Sun, Senior Scientist).

5.3.3. Analysis: We will explore a wide range of statistical approaches for dissecting genetic architecture and making predictions based on the data generated by this project. These approaches will be initially tested in SAS and custom coded software (all will be posted). Once useful algorithms have been settled upon, this project will collaborate with the USDA-ARS group at Ithaca (see letter) to implement the approaches in the open source TASSEL³⁸. One of the advantages of the TASSEL software is that it efficiently combines linkage and association data in memory efficient ways and is multi-threading. Statistical models for NAM analysis to be implemented in TASSEL will include the 26 level GLMselect, ICIM, and association projection. In collaboration with Jannink and Zhang, prediction approaches will be developed and implemented in TASSEL.

5.4. Expand maize association panel to include breeding lines and landraces: Recently, ~120 PVP inbred lines have been released to the public. These inbreds represent the elite germplasm developed by the private sector during the 1980s. We have SNP data on most of these inbreds, and will select 92 for addition to our core samples. Our project also created 27 inbreds from landraces, which otherwise exist only as outbred populations; these will be added to our maize association panel.

5.5. Develop a diverse indexed set of NILs: We will develop an indexed set of NILs that capture maize inbred and teosinte diversity. NIL sets will be created for each of the 25 NAM founders (+

Mo17, see Syngenta letter) and 10 teosinte accessions in the B73 background, as well as teosinte in the W22 background. The end product will be a complete introgression library of diverse *Zea* alleles, and perhaps the largest collection of Mendelized quantitative variation in the world.

5.6. Provide NAM each year for community access and additional trait measurement: Growing the maize NAM (~6000 plots including checks) is a large endeavor for most public maize groups. To make NAM more accessible, our group will grow the full set of 5000 RILs in one location each year of the grant for community access in addition to trait measurements for our own project.

PLAN TO INTEGRATE RESEARCH AND EDUCATION: Our outreach will teach evolution, genetics, and diversity to four target audiences: (1) The general public, including K-12 students; (2) High school teachers; (3) International scientists, and (4) University students.

A. The General Public. Expanding on the strengths of maize as an inter-disciplinary educational tool that we exploited in our previous grant, we intend to engage larger audiences representing the general public by creating a traveling exhibit that uses maize evolution in considering the important role that evolution, has played in society. The exhibit will be compact and modular (about 100 sq ft), developed with venues in rural communities in mind (e.g., small museums, public space at land grant universities, etc.). We will undertake this outreach in collaboration with the Museum of the Earth (MotE, <http://www.museumoftheearth.org/>), which specializes in making natural history accessible to audiences of all ages and backgrounds. The project will benefit from the MotE's extensive experience. Dr. Fulton will work with MotE staff to translate information from our project into material appropriate for the exhibit, develop content, and assist in the design, production and traveling display of the exhibit. Associated with the exhibit will be a kit of related activities, including hands-on materials and suggested activities for a range of educators. An online version of the exhibit, a teacher-friendly guide, and information about ongoing research will be added to the PANZEA website. Hosting the exhibit first at MotE will allow for evaluations before the exhibit travels. After the 1.5 year interval of development, we expect the exhibit to travel to 3-5 venues per year over the following 2.5 years. At the end of the grant, the exhibit may be made self-sustainable by charging a small fee for shipping and renting.

B. High School Teachers. During the previous project, Dr. Fulton established a relationship with Dr. Mulumebet Worku, an Ethiopian scientist now on faculty at North Carolina Agriculture & Technical State University, one of the Historically Black Colleges and Universities. Dr. Fulton has visited annually for the last several years to teach in Dr. Worku's summer course for high school teachers and in her undergraduate classes. We will continue this collaboration to further engage minority high school teachers and college students in the project.

C. International Scientists. Dr. Fulton coordinates the African Scientist Fellowship program at Cornell University, which supports African scientists to work at the Institute for Genomic Diversity for 3-6 months. Almost 100 African scientists have applied, and awardees have included scientists from Niger, Egypt, and Nigeria. The most recent African scientist, from Ghana, is working as a student on our current project. We will continue to integrate scientists from Africa and elsewhere into the new maize project. This project will support two African scientists at a time to work with this project (total 8 scientists over the 4 years).

D. University Students. Collectively our labs host and mentor a large number of undergraduate students each year, and produce many displays and activities for the general public, including the well-known living maize map garden in Ithaca, NY. This mentoring will be continued throughout the new project.

References Cited

The first 61 references were papers produced in the last four years by the Maize Diversity Project.

- *1. Clark, R. M., Linton, E., Messing, J. & Doebley, J. F. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci U S A* 101, 700-707 (2004).*
- *2. Doebley, J. The genetics of maize evolution. *Annual Review Of Genetics* 38, 37-59 (2004).
- *3. Gallavotti, A. et al. The role of *barren stalk1* in the architecture of maize. *Nature* 432, 630-5 (2004).
- *4. Goodman, M. M. Developing temperate inbreds using tropical maize germplasm: Rationale, results, conclusions. *Maydica* 49, 209-219 (2004).
- *5. Goodman MM. Plant breeding requirements for applied molecular biology. *Crop Sci* 44, 1913-1914 (2004).
- *6. Tarter, J. A., Goodman, M. M. & Holland, J. B. Recovery of exotic alleles in semiexotic maize inbreds derived from crosses between Latin American accessions and a temperate line. *Theor App Genet* 109, 609-617 (2004).
- *7. Tenaillon, M. I., U'Ren, J., Tenaillon, O. & Gaut, B. S. Selection versus demography: A multilocus investigation of the domestication process in maize. *Mol Biol Evol* 21, 1214-1225 (2004).
- *8. Tiffin, P., Hacker, R. & Gaut, B. S. Population genetic evidence for rapid changes in intraspecific diversity and allelic cycling of a specialist defense gene in *Zea*. *Genetics* 168, 425-434 (2004).
- *9. Wilson, L. M. et al. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16, 2719-2733 (2004).
- *10. Clark, R. M., Tavare, S. & Doebley, J. Estimating a Nucleotide Substitution Rate for Maize from Polymorphism at a Major Domestication Locus. *Mol Biol Evol* (2005).
- *11. Flint-Garcia, S. A. et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* 44, 1054-1064 (2005).
- *12. Fukunaga, K. et al. Genetic diversity and population structure of teosinte. *Genetics* 169, 2241-2254 (2005).
- *13. Liu, K. J. & Muse, S. V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128-2129 (2005).
- *14. Lockton, S. & Gaut, B. S. Plant conserved non-coding sequences and paralogue evolution. *Trends Genet.* 21, 60-65 (2005).
- *15. Matsuoka Y. Origin matters: Lessons from the search for the wild ancestor of maize. *Breeding Sci*, 383-390 (2005).
- *16. Szalma, S. J., Buckler, E. S., Snook, M. E. & McMullen, M. D. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *Theor App Genet* 110, 1324-1333 (2005).
- *17. Vigouroux, Y. et al. An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169, 1617-1630 (2005).
- *18. Vollbrecht, E., Springer, P. S., Goh, L., Buckler, E. S. & Martienssen, R. Architecture of floral branch systems in maize and related grasses. *Nature* 436, 1119-26 (2005).
- *19. Wright, S. I. & Gaut, B. S. Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol* 22, 506-519 (2005).

- *20. Wright, S. I. et al. The effects of artificial selection on the maize genome. *Science* 308, 1310-4 (2005).
- *21. Yamasaki, M. et al. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17, 2859-2872 (2005).
- *22. Anderson, L. K., Lai, A., Stack, S. M., Rizzon, C. & Gaut, B. S. Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. *Genome Res* 16, 115-122 (2006).
- *23. Buckler ES & Stevens NM. Maize domestication. In *Darwin's Harvest: New Approaches to the Origins, Evolution, and Conservation of Crops* edited by T. Motley, N. Zerega and H. Cross. Columbia University Press, (2006).
- *24. Buckler, E. S., Gaut, B. S. & McMullen, M. D. Molecular and functional diversity of maize. *Current Opinion in Plant Biology* 9, 172-176 (2006).
- *25. Buckler, E. S., Goodman, M. M., Holtsford, T. P., Doebley, J. F. & Sanchez, J. Phylogeography of the wild subspecies of *Zea mays*. *Maydica* 51, 123-134 (2006).
- *26. Canaran, P., Stein, L. & Ware, D. Look-Align: an interactive web-based multiple sequence alignment viewer with polymorphism analysis support. *Bioinformatics* 22, 885-886 (2006).
- *27. Doebley, J. Plant science - Unfallen grains: How ancient farmers turned weeds into crops. *Science* 312, 1318-1319 (2006).
- *28. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* 127, 1309-1321 (2006).
- *29. Holland JB. Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. *Crop Sci*, 642-654. (2006).
- *30. Kirst, M. et al. Genetic diversity contribution to errors in short oligonucleotide microarray analysis. *Plant Biotechnology Journal* 4, 489-498 (2006).
- *31. Morton, B. R., Bi, I. V., McMullen, M. D. & Gaut, B. S. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. *Genetics* 172, 569-577 (2006).
- *32. Robertson, L. A. et al. Heritabilities and correlations of fusarium ear rot resistance and fumonisin contamination resistance in two maize populations. *Crop Sci.* 46, 353-361 (2006).
- *33. Robertson-Hoyt, L. A. et al. QTL mapping for fusarium ear rot and fumonisin contamination resistance in two maize populations. *Crop Sci.* 46, 1734-1743 (2006).
- *34. Tracy, W. F., Whitt, S. R. & Buckler, E. S. Recurrent mutation and genome evolution: Example of Sugary 1 and the origin of sweet maize. *Crop Sci.* 46, S49-S54 (2006).
- *35. Yu, J. & Buckler, E. S. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17, 155-60 (2006).
- *36. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet.* 38, 203-208 (2006).
- *37. Zhao, W. et al. Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucl. Acids Res.* 34, D752-757 (2006).
- *38. Bradbury, P. J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633-2635 (2007).

- *39. Briggs, W. H., McMullen, M. D., Gaut, B. S. & Doebley, J. Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* 177, 1915-1928 (2007).
- *40. Ersoz, E. S., Yu, J. & Buckler, E. S. in *Genomics Assisted Crop Improvement Vol. 1: Genomics Approaches and Platforms* (eds. Varshney, R. K. & Tuberosa, R.) 97–119 (Springer, 2007).
- *41. Esch, E., Szymaniak, J. M., Yates, H., Pawlowski, W. P. & Buckler, E. S. Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. *Genetics* 177, 1851-1858 (2007).
- *42. Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J. & Anderson, L. K. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet* 8, 77-84 (2007).
- *43. Gore, M. et al. Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. *Crop Sci.* 47, S135-S148 (2007).
- *44. Holland, J. B. Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology* 10, 156 (2007).
- *45. Hufford, K. M., Canaran, P., Ware, D. H., McMullen, M. D. & Gaut, B. S. Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. *Plant Physiol.* 144, 1642-1653 (2007).
- *46. Ross-Ibarra J, Morrell PL & Gaut BS. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci U S A*, 8641-8648. (2007).
- *47. Stich, B. et al. Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* 176, 563-70 (2007).
- *48. Weber, A. et al. Major Regulatory Genes in Maize Contribute to Standing Variation in Teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 177, 2349-2359 (2007).
- *49. Yamasaki, M., Wright, S. I. & McMullen, M. D. Genomic screening for artificial selection during domestication and improvement in maize. *Annals Of Botany* 100, 967-973 (2007).
- *50. Zhang, Z., Todhunter, R. J., Buckler, E. S. & Vleck, L. D. V. Use of Marker Based Relationships with MTDFREML. *J Anim Sci* 85, 881-885 (2007).
- *51. Canaran, P. et al. Panzea: an update on new content and features. *Nucl. Acids Res.*, gkm1022. (in review).
- *52. Ersoz, E. S., Yu, J. & Buckler, E. S. Applications of linkage disequilibrium and association mapping in maize. *Molecular Genetic Approaches to Maize Improvement*. (2008(In press)).
- *53. Gur, A. et al. Central carbon metabolism genes are important underexploited targets for crop improvement. (In review).
- *54. Harjes, C. E. et al. Natural genetic variation in lycopene epsilon cyclase tapped for biofortification of maize. *Science* 319, 330 - 333 (2008).
- *55. Sanchez-Villeda, H. et al. DNAAAlignEditor: DNA Alignment Editor Tool. *BMC Bioinformatics* (Under review).
- *56. Stich, B., Maohringy, J., Piephoy, H.-P., Buckler, E. S. & Melchinger, A. E. Comparison of mixed-model approaches for association mapping. *Genetics* (Accepted).
- *57. Thuillet AC et al. A weak effect of background selection on trinucleotide microsatellites in maize. *Advance Access* (2007).
- *58. Yu, J., Holland, J. B., McMullen, M. D. & Buckler, E. S. Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics* 178, 539-551 (2008).

- *59. Yu, J. et al. On the adequacy of numbers of background markers for relationship estimation and sample sizes for association mapping. ((In Review)).
- *60. Zhao, Q. et al. The role of regulatory genes during maize domestication: evidence from nucleotide polymorphism and gene expression. *Genetics Accepted* (2008).
- *61. Yamasaki, M., Schroeder, S., Sanchez-Villeda, H., Gaut, B. & McMullen, M. D. Empirical analysis of selection screens for domestication and improvement loci in maize by extended DNA sequencing. *The Plant Genome In press* (2008).

*-----

- 62. Lee, M. et al. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Mol Biol* 48, 453-61 (2002).
- 63. Zeng, Z.-B. Precision mapping of quantitative trait loci. *Genetics* 136, 1457-1468 (1994).
- 64. Li, H. H., Ye, G. Y. & Wang, J. K. A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361-374 (2007).
- 65. Chardon, F. et al. Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168, 2169-2185 (2004).
- 66. Baurle, I. & Dean, C. The timing of developmental transitions in plants. *Cell* 125, 655-64 (2006).
- 67. Zhao, K. et al. An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS Genet* 3, e4 (2007).
- 68. Izawa, T. Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice. *J. Exp. Bot.* 58, 3091-3097 (2007).
- 69. Salvi, S. et al. Toward positional cloning of *Vgt1*, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol.* 48, 601-613 (2002).
- 70. Salvi, S. et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci U S A* 104, 11376-81 (2007).
- 71. Chardon, F., Hourcade, D., Combes, V. & Charcosset, A. Mapping of a spontaneous mutation for early flowering time in maize highlights contrasting allelic series at two-linked QTL on chromosome 8. *Theor Appl Genet* 112, 1-11 (2005).
- 72. Pritchard, J. K. & Rosenberg, N. A. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65, 220-228 (1999).
- 73. Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. Association mapping in structured populations. *Am J Hum Genet* 67, 170-181 (2000).
- 74. Stich, B., Maohringy, J., Piephoy, H.-P., Buckler, E. S. & Melchinger, A. E. Comparison of mixed-model approaches for association mapping. *Genetics In Press* (2008).
- 75. Canaran, P., Stein, L. & Ware, D. Look-Align: an interactive web-based multiple sequence alignment viewer with polymorphism analysis support. *Bioinformatics* 22, 885-6 (2006).
- 76. Casstevens, T. M. & Buckler, E. S. GDPC: connecting researchers with multiple integrated data sources. *Bioinformatics* 20, 2839-2840 (2004).
- 77. Canaran, P. et al. Panzea: an update on new content and features. *Nucl. Acids Res.*, gkm1022 (2007).
- 78. Emrich, S. J. et al. Nearly identical paralogs: implications for maize (*Zea mays* L.) genome evolution. *Genetics* 175, 429-39 (2007).
- 79. Mitchell-Olds, T. & Schmitt, J. Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* 441, 947-952 (2006).

80. Laurie, C. C. et al. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168, 2141-2155 (2004).
81. Schon, C. C. et al. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167, 485-498 (2004).
82. Alonso-Blanco, C., Mendez-Vigo, B. & Koornneef, M. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int J Dev Biol* 49, 717-32 (2005).
83. Hansen, T. F. The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics* 37, 123-157 (2006).
84. Mackay, T. F. C. The genetic architecture of quantitative traits: lessons from *Drosophila*. *Current Opinion in Genetics & Development* 14, 253-257 (2004).
85. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Mass., 1998).
86. Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* 298, 2345-9 (2002).
87. Tenaillon, M. I. et al. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci U S A* 98, 9161-9166 (2001).
88. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69-87 (2005).
89. Henry, A. M. & Damerval, C. High rates of polymorphism and recombination at the Opaque-2 locus in cultivated maize. *Mol Gen Genet* 256, 147-157 (1997).
90. Tuskan, G. A. et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596-604 (2006).
91. Moeller, D. A., Tenaillon, M. I. & Tiffin, P. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 176, 1799-809 (2007).
92. Takebayashi, N. & Morrell, P. L. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am. J. Bot.* 88, 1143-1150 (2001).
93. Goodman, M. M. in *Maize Breeding and Genetics* (ed. Walden, D. B.) 143-158 (John Wiley & Sons, New York, NY, 1978).
94. Dudley, J. W., Dijkhuizen, A., Paul, C., Coates, S. T. & Rocheford, T. R. Effects of random mating on marker-QTL associations in the cross of the Illinois High Protein X Illinois Low Protein maize strains. *Crop Sci.* 44, 1419-1428 (2004).
95. Cassman, K. G. Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *Proc Natl Acad Sci U S A* 96, 5952-9 (1999).
96. Phillips, R. L. Genetic Tools from Nature and the Nature of Genetic Tools. *Crop Sci* 46, 2245-2252 (2006).
97. Wu, R., Ma, C.-X. & Casella, G. Joint Linkage and Linkage Disequilibrium Mapping of Quantitative Trait Loci in Natural Populations. *Genetics* 160, 779-792 (2002).
98. Wu, R. & Zeng, Z. B. Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157, 899-909. (2001).
99. ter Braak, C. J., Boer, M. P. & Bink, M. C. Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. *Genetics* 170, 1435-8 (2005).

100. Xu, S. Estimating polygenic effects using markers of the entire genome. *Genetics* 163, 789-801 (2003).
101. Meuwissen, T. H. E., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M. E. Fine Mapping of a Quantitative Trait Locus for Twinning Rate Using Combined Linkage and Linkage Disequilibrium Mapping. *Genetics* 161, 373-379 (2002).
102. Blanc, G., Charcosset, A., Mangin, B., Gallais, A. & Moreau, L. Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor Appl Genet* 113, 206-24 (2006).
103. Crepieux, S., Lebreton, C., Servin, B. & Charmet, G. Quantitative Trait Loci (QTL) Detection in Multicross Inbred Designs: Recovering QTL Identical-by-Descent Status Information From Marker Data. *Genetics* 168, 1737-1749 (2004).
104. Verhoeven, K. J., Jannink, J. L. & McIntyre, L. M. Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* 96, 139-49 (2006).
105. Blott, S. et al. Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. *Genetics* 163, 253-266 (2003).
106. Zhang, Y. M. et al. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). *Genetics* 169, 2267-2275 (2005).
107. Jansen, R. C. Studying complex biological systems using multifactorial perturbation. *Nature Reviews Genetics* 4, 145-151 (2003).
108. Li, M., Boehnke, M. & Abecasis, G. R. Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. *Am J Hum Genet* 76, 934-49 (2005).
109. Remington, D. L. et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A* 98, 11479-11484 (2001).
110. Cohen, J. C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869-72 (2004).
111. Wang, H. et al. The origin of the naked grains of maize. *Nature* 436, 714-9 (2005).
112. Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genet.* 38, 594-597 (2006).
113. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 14, 143-9 (1999).
114. Phillips, P. C. The Language of Gene Interaction. *Genetics* 149, 1167-1171 (1998).
115. Malmberg, R. L. & Mauricio, R. QTL-based evidence for the role of epistasis in evolution. *Genet Res* 86, 89-95 (2005).
116. Holland, J. B. EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. *J Hered* 89, 374-375 (1998).
117. Holland, J. B., Portyanko, V. A., Hoffman, D. L. & Lee, M. Genomic regions controlling vernalization and photoperiod responses in oat. *Theor App Genet* 105, 113-126 (2002).
118. Xu, S. An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics* 63, 513-21 (2007).
119. Briollais, L. et al. Methodological issues in detecting gene-gene interactions in breast cancer susceptibility: a population-based study in Ontario. *BMC Med* 5, 22 (2007).

120. Mitchell-Olds, T., Willis, J. H. & Goldstein, D. B. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet* 8, 845-56 (2007).
121. Kang, M. S. in *Quantitative Genetics, Genomics and Plant Breeding* (ed. Kang, M. S.) 221-243 (CABI Publishing, New York, 2002).
122. Piepho, H. P. Statistical tests for QTL and QTL-by-environment effects in segregating populations derived from line crosses. *Theor Appl Genet* 110, 561-6 (2005).
123. Vargas, M., van Eeuwijk, F. A., Crossa, J. & Ribaut, J. M. Mapping QTLs and QTL x environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. *Theor Appl Genet* 112, 1009-23 (2006).
124. Morgante, M. & Salamini, F. From plant genomics to breeding practice. *Curr Opin Biotechnol* 14, 214-9 (2003).
125. Paran, I. & Zamir, D. Quantitative traits in plants: beyond the QTL. *Trends Genet* 19, 303-6 (2003).
126. Thomas, B. C., Rapaka, L., Lyons, E., Pedersen, B. & Freeling, M. Arabidopsis intragenomic conserved noncoding sequence. *Proceedings of the National Academy of Sciences* 104, 3348-3353 (2007).
127. Inada, D. C. et al. Conserved Noncoding Sequences in the Grasses. *Genome Res.* 13, 2030-2041 (2003).
128. Ng, P. C. & Henikoff, S. Predicting Deleterious Amino Acid Substitutions. *Genome Res.* 11, 863-874 (2001).
129. Chakrabarti, S. & Lanczycki, C. J. Analysis and prediction of functionally important sites in proteins. *Protein Sci* 16, 4-13 (2007).
130. Yang, Z. & Nielsen, R. Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage. *Mol Biol Evol*, msm284 (2008).
131. Carbone, A., Zinovyev, A. & Kepes, F. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 19, 2005-2015 (2003).
132. Brendel, V., Xing, L. & Zhu, W. Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. *Bioinformatics* 20, 1157-69 (2004).
133. Sparks, M. E. & Brendel, V. Incorporation of splice site probability models for non-canonical introns improves gene structure prediction in plants. *Bioinformatics* 21 Suppl 3, iii20-30 (2005).
134. Zhang, Y. miRU: an automated plant miRNA target prediction server. *Nucl. Acids Res.* 33, W701-704 (2005).
135. Rhoades, M. W. et al. Prediction of Plant MicroRNA Targets. *Cell* 110, 513 (2002).
136. Johnson, C., Bowman, L., Adai, A. T., Vance, V. & Sundaesan, V. CSRDB: a small RNA integrated database and browser resource for cereals. *Nucl. Acids Res.* 35, D829-833 (2007).
137. Stupar, R. M. & Springer, N. M. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. *Genetics* 173, 2199-210 (2006).
138. Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S. & Przeworski, M. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am J Hum Genet* 72, 1527-35 (2003).
139. Nachman, M. W. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev* 12, 657-63 (2002).

140. Tenaillon, M. I. et al. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp *mays* L.). *Genetics* 162, 1401-1413 (2002).
141. Bernardo, R. & Yu, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Sci* 47, 1082-1090 (2007).
142. Mikel, M. A. Availability and Analysis of Proprietary Dent Corn Inbred Lines with Expired U.S. Plant Variety Protection. *Crop Sci* 46, 2555-2560 (2006).
143. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39, 906 (2007).
144. Peiffer, D. A. et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 16, 1136-1148 (2006).