

# Research Approach: Evolutionary genomics of maize

My research program applies population genomic approaches to investigate how plants evolve. Agricultural plants provide a particularly excellent system for basic discovery, with their rich history of basic genetic research, ample modern genomic resources, and the opportunity to sample or replicate genotypes across a wide array of environments. Among crop plants, maize provides perhaps the best opportunity for basic research and discovery. In addition to its status as the country's most valuable crop, maize is also one of the oldest model organisms, and its rich phenotypic and genetic diversity have enabled a number of basic scientific discoveries, from hybrid vigor [1] to the relationship between crossing over and recombination [2], meiotic drive [3], and transposable elements [4]. Maize continues to be an important model in the genomics era [5], and the publicly available resources are unparalleled among other crops: genome and whole-genome genotypes number in the tens of thousands, and association mapping studies now approach scales rarely seen outside of biomedical studies [e.g. 6, 7, 8, 9].

Maize is particularly well-suited as a model to study plant evolution. Domestication and modern breeding can be seen as experimental evolution, and both population genomic [10] and archaeological data [11] suggest that the rate and magnitude of adaptation in these situations is likely similar to that found in natural populations. While maize has a large genome compared to many other model plants — the maize reference is 2.3Gb [12] — it is in fact close to the median for most flowering plants [13], and thus an ideal organism in which to study transposable elements and other structural variation that abound in complex plant genomes. Finally, the wild relatives of maize, collectively referred to as teosintes, inhabit a diverse array of environments and include both endangered species and taxa with large, stable, randomly-mating populations [14]. Because all of these species are closely related, we can easily transfer genomic resources developed in maize to these wild taxa as well [e.g. 15, 16]. Though there are numerous efforts in maize to accelerate crop breeding, there are surprisingly few efforts to take advantage of these resources to study maize evolution. My group has taken the lead in evolutionary analyses of maize and its relatives; below I describe current and planned research towards understanding plant adaptation and genome evolution.

## Experimental Evolution

Plant domestication and modern breeding essentially represent examples of experimentally evolved populations. Studying these populations represents an opportunity to understand not only the genetic basis of evolutionary change (which alleles were selected), but also how the processes of evolution — particularly the interaction of selection and demographic change — shape modern genetic and phenotypic diversity.

We have worked to understand the process of selection during domestication, documenting its polygenic nature, the contributions of regulatory variation [10, 17], and the importance of processes such as convergent molecular evolution [18] and selection on standing genetic variation [18, 19, 20]. Our detailed analyses of inbred lines have revealed little evidence for strong selection during modern maize breeding, instead highlighting the effect of small population sizes in partitioning diversity into distinct populations of increasingly narrow ancestry [21].

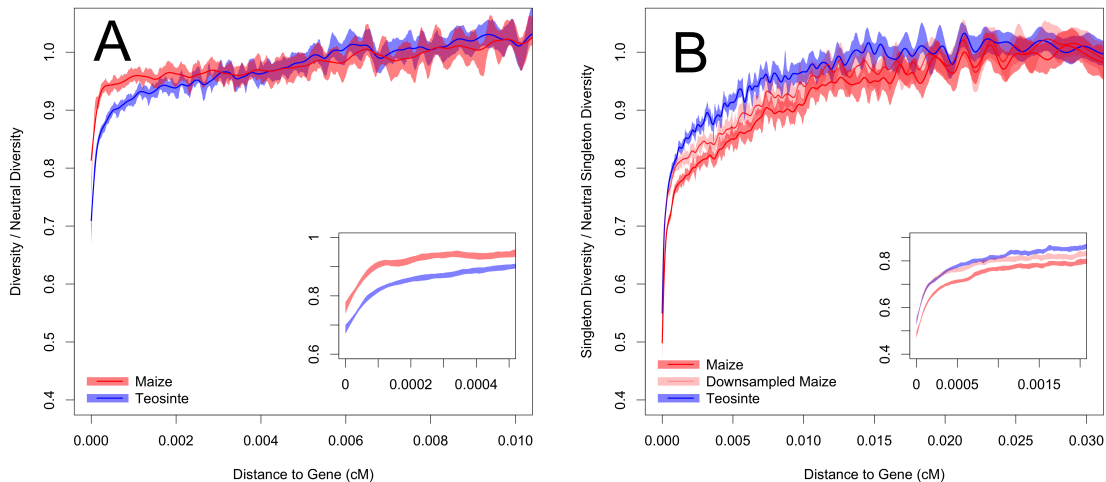


Figure 1: Relative pairwise nucleotide diversity (A) and singleton diversity (B) as a function of distance to the nearest gene.

Our current work on domestication focuses on how the processes of demographic change and adaptation have shaped genetic and phenotypic diversity. Maize underwent a demographic bottleneck during domestication, reducing its effective population size and thus the efficacy of purifying selection. Indeed, purifying selection in teosinte is stronger due to its larger effective population size, resulting in both a deeper and wider valley of diversity around conserved genes (Figure 1A). But maize population grew quickly after domestication, eventually exceeding that of teosinte. New mutations in maize are thus subject to stronger selection than in teosinte, a shift reflected in patterns of variation in recent low frequency variants such as singleton polymorphisms (Figure 1B). Population demographic change can also impact the size, number, and dominance of loci underlying phenotypic traits [22, 23]. Our future work will utilize data from large-scale association mapping studies, coupled with population genomic inference of demographic change and selection, to study how the process of domestication may have shaped the genetic architecture of phenotypic traits.

Building on our work highlighting the importance of deleterious alleles in phenotypic variation [24], we have used experimentally evolved populations to track haplotype frequencies over time and assess the genetic basis of gain in hybrid yield. Our recent analysis finds little overlap in selected haplotypes between two populations bred for increasing hybrid yield, consistent with a model for complementation of deleterious variants brought to high frequency by hitchhiking [25]. We are now working to extend these analyses has combined genotype and historical data to build a pedigree of more than 4,000 maize lines with which to extend these haplotype analyses across many populations.

## Local adaptation

After domestication, maize spread rapidly, adapting to a wide range of environments. Today maize is grown across a broader geographic breadth than any of the world’s other staple crops [26], from sea level to altitudes of > 4,000m and from deserts to near-flooded conditions. The wild relatives

of maize have also adapted to environments varying widely in elevation, temperature, and moisture gradients. Our previous work has shown that adaptation in teosinte is often restricted to discrete local populations and has often made use of regulatory variation [15]. We also find evidence that inversion polymorphisms are common and associated with environmental gradients and phenotypic variation [15, 16]. In contrast to classical expectations of underdominance, even large inversion polymorphisms appear able to circumvent problems of underdominance in heterozygotes without negatively affecting normal chromosomal pairing [27]. We suspect such structural variation is common in complex plant genomes, and work is underway both to characterize the genomic and phenotypic effects of individual inversions and to more broadly characterize structural variation within natural populations.

In many instances of local adaptation, independent populations have converged on similar phenotypic adaptations. We have worked with maize populations adapted to high elevation environments in Mexico and South America, seeking to understand whether convergent phenotypic evolution is associated with convergent evolution at the molecular level. Our previous efforts pointed to a key role for adaptive introgression from wild teosinte in enabling maize to adapt to the mountains of central Mexico [28], but in spite of similar phenotypes our recent work shows virtually no overlap between Mexican and Andean maize populations in the genes most affected by selection [29]. Our population genetic analyses suggest that most local adaptation in these populations has not been mutation limited, but that instead stochastic differences in local founding populations, each with abundant standing genetic variation, likely explains the lack of convergent genetics.

Our current work on local adaptation builds on these results, focusing on detailed characterization of individual populations of both maize and teosinte. We are making use of deep sequencing of the genome and methylome of a number of individuals, combined with common garden evaluation of progeny, to identify the loci and processes involved in adaptation at an extremely local scale and ask how these differ from processes operating species-wide.

## Genome evolution

In addition to discerning the genetic basis of phenotypic evolution, we are interested in understanding the processes that shape evolution of the genome itself. We have shown that differences in deletion bias can effect large changes in genome size over phylogenetic time scales [30], and documented extensive variation in copy number across diverse maize and teosinte [31, 32]. Current work characterizing copy number variation in a single wild population of teosinte has revealed problems with population genetic methods that assume data is missing due to stochastic sampling. Tajima's D, for example, shows a strong positive correlation with CNV frequency, potentially leading to misinterpretation of selection or demographic change if biologically missing data was not accounted for.

The single largest component of the maize genome — and of the majority of flowering plants — is transposable elements (TEs). We have previously shown important functional consequences of individual TE insertions on phenotype and gene expression [19, 33], we are just beginning to understand their genome-wide significance. We have recently re-annotated all of the retroelements in the maize genome (> 70% of the genome), and are now using high-throughput sequencing and our annotation to identify new insertions in nonreference genomes. Analysis of the first deeply sequenced inbred, for example, revealed 350,000 high-confidence nonreference insertions, of which nearly 3,000 were into CDs of annotated genes. We are currently evaluating the impacts of these

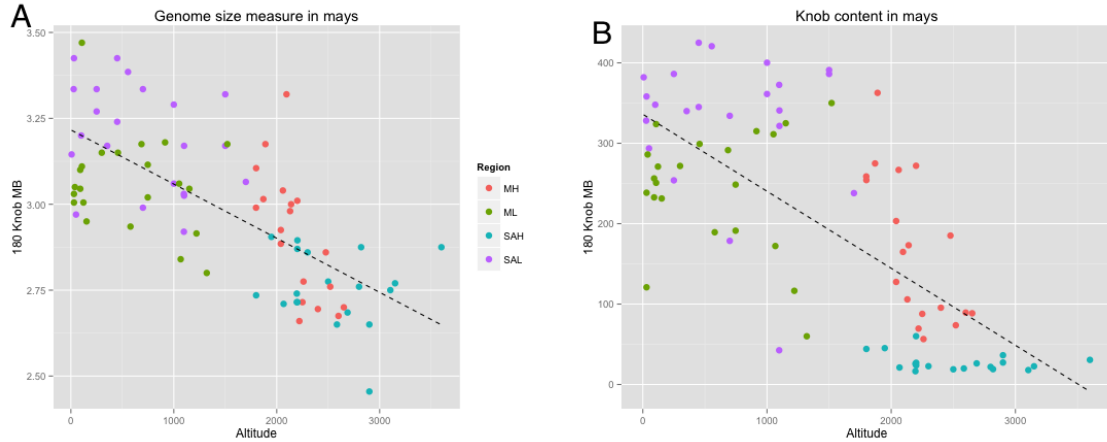


Figure 2: Heterochromatic knob abundance (A) and 1C genome size (B) as a function of elevation. Data are shown from Mexican (M) and South American (SA) highland (H) and lowland (L) populations.

insertions on patterns of methylation and gene expression across a panel of inbred lines, and in the future will apply these methods to natural populations.

Finally, in one project underway we are modeling the genome as a phenotype to understand the evolutionary processes shaping genome size change. Building on prior observations of genome size variation, we have developed methods to quantify the abundance of repetitive fractions of the genome and test for selection on repeat abundance after controlling for relatedness. We find that both overall genome size (Figure 2A) and, in spite of their potential for meiotic drive, heterochromatic knob abundance (Figure 2B) both show evidence of selection for smaller genomes across altitudinal gradients, perhaps as a means of accelerating development and flowering.

## References

- [1] George H Shull. The composition of a field of maize. *Journal of Heredity*, (1):296–301, 1908.
- [2] Harriet B Creighton and Barbara McClintock. A correlation of cytological and genetical crossing-over in zeo mays. *Proceedings of the National Academy of Sciences of the United States of America*, 17(8):492–497, 1931.
- [3] MM Rhoades. Preferential segregation in maize. *Genetics*, 27(4):395–407, 1942.
- [4] Barbara McClintock. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6):344–355, 1950.
- [5] Natalie J Nannas and R Kelly Dawe. Genetic and genomic toolbox of zeo mays. *Genetics*, 199(3):655–669, 2015.
- [6] ES Buckler, JB Holland, PJ Bradbury, CB Acharya, PJ Brown, C Browne, E Ersoz, S Flint-Garcia, A Garcia, JC Glaubitz, MM Goodman, C Harjes, K Guill, DE Kroon, S Larsson, NK Lepak, H Li, SE Mitchell, G Pressoir, JA Peiffer, MO Rosas, TR Rocheford, MC Romay, S Romero, S Salvo, H Sanchez Villeda, HS da Silva, Q Sun, F Tian, N Upadyayula, D Ware, H Yates, J Yu, Z Zhang, S Kresovich, and MD McMullen. The genetic architecture of maize flowering time.\*. *Science*, 325(5941):714–718, 2009.
- [7] Jason A Peiffer, Maria C Romay, Michael A Gore, Sherry A Flint-Garcia, Zhiwu Zhang, Mark J Millard, Candice AC Gardner, Michael D McMullen, James B Holland, Peter J Bradbury, et al. The genetic architecture of maize height. *Genetics*, 196(4):1337–1356, 2014.
- [8] MC Romay, 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.
- [9] Sarah Hearne, Charles Chen, Ed Buckler, and Sharon Mitchell. Unimputed gbs derived snps for maize landrace accessions represented in the seed-maize gwas panel. <http://hdl.handle.net/11529/10034>, 2015. Accessed: 2015-02-16.
- [10] Matthew B Hufford, Xun Xu, Joost Van Heerwaarden, Tanja Pyhäjärvi, Jer-Ming Chia, Reed A Cartwright, Robert J Elshire, Jeffrey C Glaubitz, Kate E Guill, Shawn M Kaeppler, et al. Comparative population genomics of maize domestication and improvement. *Nature Genetics*, 44(7):808–811, 2012.
- [11] Michael D Purugganan and Dorian Q Fuller. Archaeological data reveal slow rates of evolution during plant domestication. *Evolution*, 65(1):171–183, 2011.
- [12] Patrick S Schnable, Doreen Ware, Robert S Fulton, Joshua C Stein, Fusheng Wei, Shiran Pasternak, Chengzhi Liang, Jianwei Zhang, Lucinda Fulton, Tina A Graves, et al. The b73 maize genome: complexity, diversity, and dynamics. *science*, 326(5956):1112–1115, 2009.
- [13] Ilia J Leitch and Andrew R Leitch. Genome size diversity and evolution in land plants. *Plant Genome Diversity Volume 2*, pages 307–322, 2013.
- [14] MB Hufford, P Bilinski, T Pyhäjärvi, and J Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends in Genetics*, 12:606–615, 2012.

- [15] Tanja Pyhäjärvi, Matthew B Hufford, Sofiane Mezmouk, and Jeffrey Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome biology and evolution*, 5(9):1594–1609, 2013.
- [16] Zhou Fang, Tanja Pyhäjärvi, Allison L Weber, R Kelly Dawe, Jeffrey C Glaubitz, José de Jesus Sánchez González, Claudia Ross-Ibarra, John Doebley, Peter L Morrell, and Jeffrey Ross-Ibarra. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, 191(3):883–894, 2012.
- [17] Ruth Swanson-Wagner, Roman Briskine, Robert Schaefer, Matthew B Hufford, Jeffrey Ross-Ibarra, Chad L Myers, Peter Tiffin, and Nathan M Springer. Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences*, 109(29):11878–11883, 2012.
- [18] David M Wills, Clinton J Whipple, Shohei Takuno, Lisa E Kursel, Laura M Shannon, Jeffrey Ross-Ibarra, and John F Doebley. From many, one: genetic control of prolificacy during maize domestication. *PLoS Genetics*, 9(6):e1003604, 2013.
- [19] Anthony Studer, Qiong Zhao, Jeffrey Ross-Ibarra, and John Doebley. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, 43(11):1160–1163, 2011.
- [20] Laura Vann, Thomas Kono, Tanja Pyhäjärvi, Matthew B Hufford, and Jeffrey Ross-Ibarra. Natural variation in teosinte at the domestication locus *teosinte branched1* (*tb1*). *PeerJ*, 3:e900, 2015.
- [21] Joost van Heerwaarden, Matthew B Hufford, and Jeffrey Ross-Ibarra. Historical genomics of north american maize. *Proceedings of the National Academy of Sciences*, 109(31):12420–12425, 2012.
- [22] Kirk E Lohmueller. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genetics*, 10(5), 2014.
- [23] Elodie Gazave, Diana Chang, Andrew G Clark, and Alon Keinan. Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics*, 195(3):969–978, 2013.
- [24] Sofiane Mezmouk and Jeffrey Ross-Ibarra. The pattern and distribution of deleterious mutations in maize. *G3: Genes— Genomes— Genetics*, 4(1):163–171, 2014.
- [25] Justin P Gerke, Jode W Edwards, Katherine E Guill, Jeffrey Ross-Ibarra, and Michael D McMullen. The genomic impacts of drift and selection for hybrid performance in maize. *arXiv preprint arXiv:1307.7313*, 2013.
- [26] Sarah Hake and Jeffrey Ross-Ibarra. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*, 4:e05861, 2015.
- [27] Marjorie P Maguire. The relationship of crossing over to chromosome synapsis in a short paracentric inversion. *Genetics*, 53(6):1071, 1966.
- [28] MB Hufford, P Lubinsky, T Pyhäjärvi, MT Devengenzo, NC Ellstrand, and J Ross-Ibarra. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477–e1003477, 2013.

- [29] Shohei Takuno, Peter Ralph, Kelly Swarts, Rob J. Elshire, Jeffrey C. Glaubitz, Edward S. Buckler, Matthew B. Hufford, and Jeffrey Ross-Ibarra. Independent molecular basis of convergent highland adaptation in maize. *Genetics*, 2015. doi: 10.1534/genetics.115.178327. URL <http://www.genetics.org/content/early/2015/06/15/genetics.115.178327.abstract>.
- [30] Maud I Tenaillon, Matthew B Hufford, Brandon S Gaut, and Jeffrey Ross-Ibarra. Genome size and transposable element content as determined by high-throughput sequencing in maize and *zea luxurians*. *Genome biology and evolution*, 3:219–229, 2011.
- [31] Michael A Gore, Jer-Ming Chia, Robert J Elshire, Qi Sun, Elhan S Ersoz, Bonnie L Hurwitz, Jason A Peiffer, Michael D McMullen, George S Grills, Jeffrey Ross-Ibarra, et al. A first-generation haplotype map of maize. *Science*, 326(5956):1115–1117, 2009.
- [32] Jer-Ming Chia, Chi Song, Peter J Bradbury, Denise Costich, Natalia de Leon, John Doebley, Robert J Elshire, Brandon Gaut, Laura Geller, Jeffrey C Glaubitz, et al. Maize hapmap2 identifies extant variation from a genome in flux. *Nature Genetics*, 44(7):803–807, 2012.
- [33] Irina Makarevitch, Amanda J Waters, Patrick T West, Michelle Stitzer, Candice N Hirsch, Jeffrey Ross-Ibarra, and Nathan M Springer. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genetics*, 11(1):e1004915, 2015.