Parallel altitudinal clines reveal adaptive evolution of genome size in *Zea mays*

Paul Bilinski[1], Patrice S. Albert[2], Jeremy J Berg[3,4], James A Birchler[2], Mark Grote[5], Anne Lorant[1], Juvenal Quezada[1], Kelly Swarts[6], Jinliang Yang[1,7], Jeffrey Ross-Ibarra[1,3,8,*]

1 Dept. of Plant Sciences, University of California, Davis, CA, USA
2 Dept. of Biological Sciences, University of Missouri, Columbia, MO, USA
3 Center for Population Biology, University of California, Davis, CA, USA
4 Dept. of Evolution and Ecology, University of California, Davis, CA, USA
5 Dept. of Anthropology, University of California, Davis, CA, USA
6 Dept. of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, DE
7 Dept. of Horticulture and Agronomy, University of Nebraska, Lincoln, NE, USA
8 Genome Center, University of California, Davis, CA, USA
∗ **E-mail: rossibarra@ucdavis.edu**

Abstract

While the vast majority of genome size variation in plants is due to differences in repetitive sequence, we know little about how selection acts on repeat content in natural populations. Here we investigate parallel changes in intraspecific genome size and repeat content of domesticated maize (*Zea mays*) landraces and their wild relative teosinte across altitudinal gradients in Mesoamerica and South America. We combine genotyping, low coverage whole-genome sequence data, and flow cytometry to test for evidence of selection on genome size and individual repeat abundance. We find that genetic drift alone cannot explain the observed variation, implying that clinal patterns of genome size are likely maintained by natural selection. Our modeling provides little evidence of selection on individual repeat classes, suggesting that repetitive sequences are under selection primarily due to their contribution to genome size. To better understand the phenotypes driving selection on genome size, we conducted a growth chamber experiment using a highland teosinte from a single population varying more than 1Gb in 2C genome size. We find no evidence of a correlation between genome size and cell size, but do find statistical support for a negative correlation between genome size and cell production. Re-analysis of published shoot apical meristem data further identifies a negative correlation between cell production rate and flowering time. Together, our data suggest a model in which variation in genome size is driven by natural selection on flowering time across altitudinal clines, connecting repetitive sequence variation to important differences in adaptive phenotypes.

Introduction

Genome size varies many orders of magnitude across species, due to both changes in ploidy as well as haploid DNA content [1,2]. Early hypotheses for this variation proposed that genome size was linked to organismal complexity, as more complex organisms should require a larger number of genes. Empirical analyses, however, revealed instead that most variation in genome size is due to noncoding repetitive sequence [3]. While this discovery resolved the lack of correlation between genome size and complexity, we still know relatively little about the makeup of many eukaryote genomes, the impact of genome size on phenotype, or the processes that govern variation in repetitive DNA and genome size among taxa [4].

A number of hypotheses have been offered to explain variation in genome size among taxa. Across deep evolutionary time, genome size appears to correlate with estimates of effective population size, leading to suggestions that drift and ineffective selection permit maladaptive expansion [5] or contraction [6] of genomes across species. A recent evaluation of genome size and the strength of purifying selection among isopods finds evidence supporting this model [7], but broad-scale phylogenetic analysis across seed plants fails to find evidence of a correlation between effective population size and genome size, casting doubt on its generality [8]. Other models consider mutation rates, positing that genome sizes evolve to stable equilibria in which the loss of DNA through frequent small deletions is equal to the rate of DNA gain through large insertions. Evidence of the phylogenetic lability of genome size among plants in the family Brassicaceae [9], however, appears inconsistent with this model. Mating system has also been proposed as an explanation for difference in genome size [5], as lower effective population size in selfing or asexual species leads to a reduced ability to purge slightly deleterious novel insertions. Phylogenetic comparisons of repeat abundance and genome size across mating systems in *Oenothera*, however, find little support for this hypothesis [10]. In addition to these neutral models, many authors have proposed adaptive explanations for genome size variation. Numerous correlations between genome size and physiologically or ecologically relevant phenotypes have been observed, including nucleus size [11], plant cell size [12], seed size [13], body size [14], and growth rate [15]. Adaptive models of genome size evolution suggest that positive selection drives genome size towards an optimum due to selection on these or other traits, and that stabilizing selection prevents expansions and contractions away from the optimum [16]. In most of these models, however, the mechanistic link between genome size and phenotype remains unclear [17].

Much of the discussion about genome size variation has focused on variation among species, and intraspecific variation has often been downplayed as the result of experimental artifact [18] or argued to be too small to have much evolutionary relevance [19]. Nonetheless, intraspecific variation in genome size has been documented in hundreds of plant species [19] including multiple examples of large-scale variation [20–22]. Correlations between intraspecific variation in genome size and other phenotypes or environmental factors

have also been observed [20, 21, 23], suggesting the possibility that some of the observed variation may be adaptive.

Here we present an analysis of genome size variation in the model system maize (*Zea mays*) and its wild relatives teosinte. Genome size variation among individuals can reach up to 70% between maize and teosinte and 40% within maize, and genome size correlates with both altitude and latitude [21, 24]. Sequencing of the maize reference inbred B73 revealed that the vast majority (85%) of the genome is comprised of transposable elements (TEs) [25], and shotgun sequence comparison of maize and its relative *Zea luxurians* suggests that variation between species may be explained largely by differences in TE content [26]. Nonetheless, we know little about the factors that drive intraspecific genome size variation : while BAC sequencing has identified substantial TE polymorphism among individuals [27, 28], tandem satellite sequence from large heterochromatic knob repeats can expand to as much as 8% of the genome [29] and variation in knob content appears to explain a large proportion of variation in genome size among maize inbreds [30].

We take advantage of parallel altitudinal clines in maize landraces from Mesoamerica and South America to investigate both the sequence variation and evolutionary processes underlying genome size variation. Our comparison of flow cytometry data to genotyping reveals evidence that selection has shaped patterns of genome size variation, but similar analysis of repeat content from low coverage shotgun sequencing fails to identify widespread selection on individual repeats. We then perform growth chamber experiments to measure the effect of genome size variation on the developmental traits of cell production and leaf elongation in the related wild highland teosinte *Z. mays* ssp. *mexicana*. These experiments find modest support for slower cell production in larger genomes, but no correlation between genome size and cell size. Based on these results and reanalysis of published data, we propose a model in which variation in genome size is driven by natural selection on flowering time across altitudinal clines, connecting repetitive sequence variation to important differences in adaptive phenotypes.

Results

We sampled 83 diverse maize landraces from across a range of altitudes in Mesoamerica and South America (Supp. Table S4). Flow cytometry of these samples revealed a negative correlation with altitude on both continents (Figure 1A, r=-0.56 and -0.82, respectively). We used low-coverage whole-genome sequencing mapped to reference repeat libraries to estimate the abundance of repetitive sequences in each individual with estimated genome size, and validated this approach by comparing sequence-based estimates of heterochromatic knob abundance to fluorescence *in situ* hybridization (FISH) data from *mexicana* populations (Figure 2 and Supp. Fig. S1; see Methods for details). Consistent with previous work [26,30], transposable elements and heterochromatic knobs contributed most to variation in genome size across our maize samples. While there was substantial

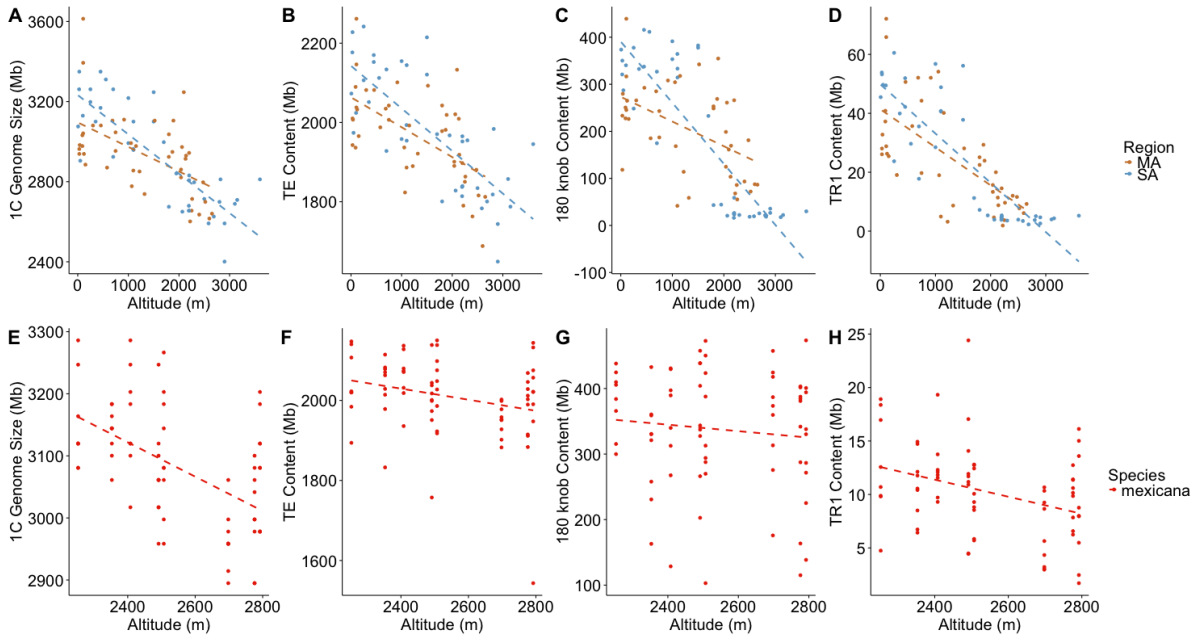

Figure 1. Genomic size and repeat content by altitude in *Zea* taxa. (A-D) Maize landraces from Mesoamerica (MA) or South America (SA). (E-H) Highland teosinte *Z. mays* ssp. *mexicana*. Only teosinte populations above 2000m that do not show admixture (see text) are included. (A,E) total genome size, (B,F) total transposable element content, (C,G) 180bp knob repeat content, (D,H) TR1 knob repeat content. Dashed lines represent the best fit linear regression.

variation among landraces in the abundance of individual transposable element families (Supp. Fig. S3), both transposable elements as a whole and heterochromatic knobs showed clear decreases in abundance with increasing altitude, mirroring the pattern seen for overall genome size (Figure 1).

We next sought to evaluate whether the observed clines in genome size and repeat abundance simply reflected genetic drift or could be better explained by natural selection along an altitudinal cline. We adopted an approach similar to Berg *et al* [32], modeling genome size as a quantitative trait that is a linear function of relatedness and altitude (see Methods, Equation 1). Across maize landraces, we rejected a neutral model in which genome size is unrelated to altitude, estimating a decrease of 100Kb in mean genome size per meter gain of altitude (Supp. Table S1). To evaluate whether selection has acted on individual repeats independent of their contribution to genome size, we treated abundance of each repeat class as a quantitative trait in a comparable model that included genome size as a covariate (Methods, Equation 2). After controlling for their contribution to genome size in this manner, we find no evidence of selection on individual repeats (Supp. Table S1).

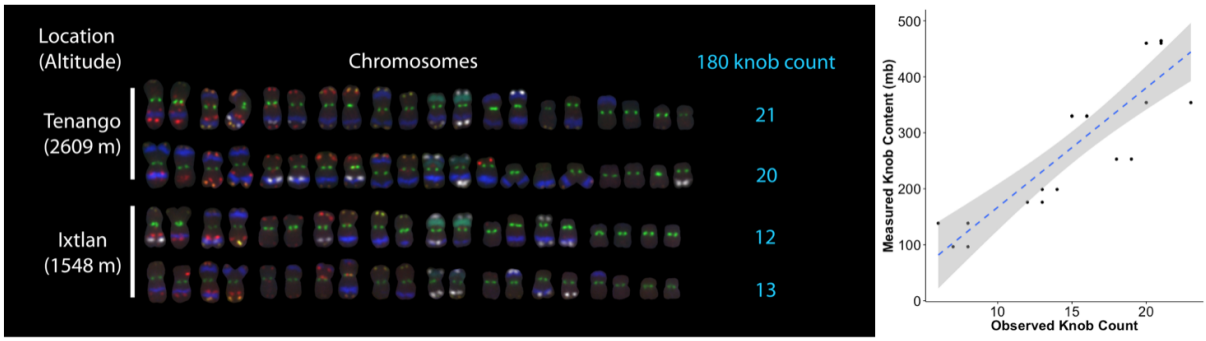

Figure 2. Knob content in highland teosinte estimated using FISH and low-coverage sequencing. (A) FISH from four *Z. mays* ssp. *mexicana* individuals, sampled from the highest and lowest altitude populations. Counts of cytological 180bp knobs (blue) are shown to the right of each individual. Other stained repeats are CentC and subtelomere 4-12-1 (green), 5S ribosomal gene (yellow), Cent4 (orange), NOR (blue-green), and TAG microsatellite 1-26-2 and subtelomere 1.1 (red). For further staining information, see [31]. (B) Plot of the population-level correlation between 180bp knob counts and sequence abundance for 20 *mexicana* individuals. Correlation with TR1 knob repeat abundance is shown in Supp. Fig S1. 180bp knob $R^2 = 0.78$, TR1 knob $R^2 = 0.74$

The wild ancestor of maize, *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*), grows on the lower slopes of the Sierra Madre in Mexico. A related wild teosinte, *Zea mays* ssp. *mexicana*, diverged from *parviglumis* more than 60,000 years ago [33] and has adapted to the higher altitudes of the Mexican central plateau [34]. We sampled leaves and measured genome size of two individuals each from previously collected populations of both subspecies (6 *parviglumis* populations and 10 *mexicana* populations) [35, 36]. Though both subspecies exhibit considerable variation, *mexicana* had a smaller average genome size than *parviglumis* (Supp. Fig. S4; t-test p-value=0.028), consistent with our observations of decreasing genome size along altitudinal clines in Mesoamerican and South American maize.

To evaluate clinal patterns across populations of highland teosinte in more detail, we sampled multiple individuals from each of an additional 11 populations of *mexicana* across its altitudinal range in Mexico (Supp. Table S5). Genome size variation across these populations revealed no clear relationship with altitude (Supp. Fig S5), but genotyping data [37] revealed consistent evidence of maize introgression (Supp. Fig S6) and higher inbreeding coefficients (2/3 significantly different from 0, t-test p-value <0.05; Supp. Fig. S7) in the three lowest altitude populations (see Methods). These three populations are also phenotypically distinct (A. O'Brien pers. communication) and relatively isolated from the rest of the distribution. We thus excluded these three populations, applying our linear model of altitude and relatedness to 70 individuals from the remaining 8 populations. We recovered a negative relationship between genome size and altitude in *mexicana*

(Fig. 1b, p-value <0.001) of similar magnitude to that seen in maize (loss of 260Kb/m), suggesting parallel patterns of selection across *Zea*. In contrast to our findings in maize, however, knob TR1 repeats still showed evidence of selection even after controlling for their contribution to genome size (Supp. Table S1).

Previous work has shown that selection for early flowering time in maize results in a concomitant reduction in genome size [38], and both highland maize and highland teosinte flower earlier than their lowland counterparts [39,40]. Extrapolating from this observation, we reasoned that genome size might be related to flowering time through its potential effect on the rate of cell production and consequently development. To test this hypothesis, we performed a growth chamber experiment to measure leaf elongation rate, cell size, and genome size using 201 individuals from 51 mothers of highland teosinte sampled from a single natural population (see Methods). Individual plants varied by as much as 1.13Gb in 2C genome size, with observed that leaf elongation rate (LER) varied from 1 to 8 cm/day (mean 4.56cm/day; Supp. Table S6). We designed a Bayesian model of leaf elongation as a function of cell size, cell production rate, and genome size (see Methods). Our posterior parameter estimates fail to find any relationship between genome size and cell size (γ_{GS}; Figure 3A), but we identified a significant negative relationship between genome size and cell production rate (β_{GS}; Figure 3B). We found that our model was sensitive to the prior density of LER (Supp. Fig. S8), but a negative relationship between genome size and cell production rate held across the range of plausible prior values found in the literature.

Recent work exploring shoot apical meristem (SAM) phenotypes across 14 maize inbred lines [41] allowed further exploration of our hypothesized connection between cell production and flowering time. Because SAM were sampled at equivalent growth stages, we interpreted variation in cell number as representative of differences in cell division rate among lines. We re-analyzed the Leiboff *et al* [41] data to investigate whether the cell number reported in each SAM was correlated with flowering time (Figure 3C). After controlling for population structure (see Methods) we found significant negative correlations between cell number in the SAM and adult flowering time in two of the three growth stages evaluated. Such negative correlation suggests that a lower number of cells in the seedling SAM — and thus a slower rate of cell division — is predictive of later flowering time.

Discussion

Genome size and repeat abundance

We report evidence of a negative correlation between genome size and altitude across clines in Mesoamerica and South America and in both maize and its wild relative teosinte (Figure 1). Genetic evidence suggests colonization of highland environments was independent in Mesoamerica and South America [42] and the populations exhibit little evidence of convergent evolution [43]. In addition to these altitidunal clines in maize, the highland

teosinte subspecies *mexicana* has occupied a high elevation niche [34], likely coinciding with its split from the lowland subspecies *parviglumis* long before maize domestication [33]. Previous investigations of genome size have also identified multiple altitudinal clines in maize and teosinte [21, 24] (but see [44] for a counterexample in the U.S. Southwest), suggesting that this observation is general and not an artifact of our sampling.

Our initial evaluation of genome size in highland teosinte found no significant correlation with elevation, due primarily to the very small genomes observed in the three lowest elevation populations (Supp. Figure S5). One explanation for this observation is the possibility of admixture with domesticated maize. All three low elevation *mexicana* populations showed evidence of such admixture (Supp. Fig. S6), and high elevation maize populations have smaller genomes than low elevation *mexicana* (Figure 1). These same three populations also showed higher levels of inbreeding than other *mexicana* populations. Because knobs are thought to have potentially deleterious effects on viability [45], increased homozygosity due to inbreeding may have facilitated selection against knobs and contributed to the smaller genome size in these populations.

A number of explanations are possible for the observed clines in genome size. Several authors have similarly identified ecological correlates of genome size and argued for adaptive explanations of such clines [21, 24, 46], but did not correct for relatedness among individuals or populations. We employ a modeling approach that uses SNP data to generate a null expectation for genome size variation among populations, allowing us to rule out these processes as driving forces, suggesting the action of selection in patterning clinal

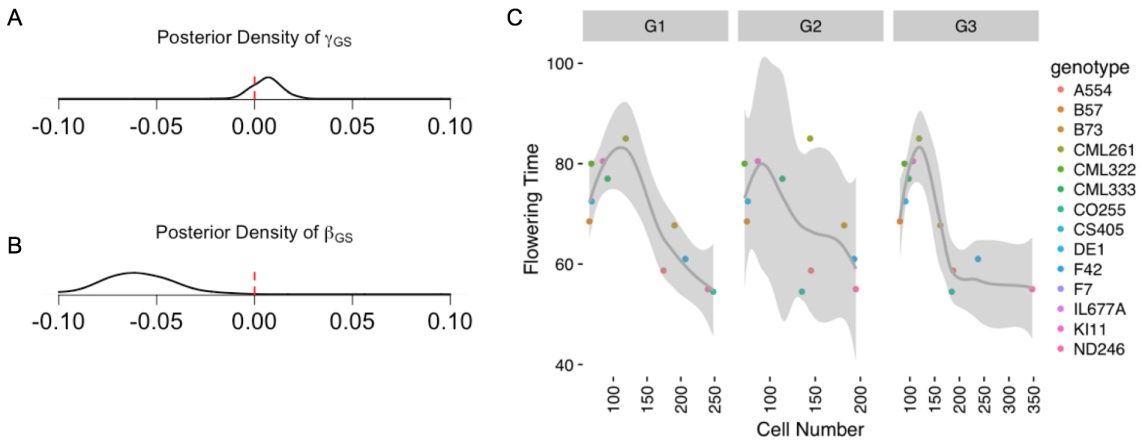

Figure 3. (A,B) Posterior densities of effects of genome size on cell size and cell production rate (γ_{GS} and β_{GS}, respectively) from a model with prior mean stomatal cell size of 30 microns and leaf elongation rate of 4cm/day. (C) Smoothed splines showing the relationship between flowering time and SAM cell number in inbred maize accessions. Measurements for cell number are shown for each of three growth phases (G1, G2, G3). Data from Leiboff *et al* [41].

differences in genome size. In addition to demography and selection, mutational biases may also play an important role. For example, plants grown at high altitudes are exposed to increased UV radiation and UV-mediated DNA damage may lead to higher rates of small deletions [47]. Alternatively, expansion of transposable elements in lowland populations could lead to increased rates of insertion. Neither scenario seems sufficient to explain our data, however. As UV damage causes small DNA deletions, it is unlikely to generate the gigabase-scale difference we see across altitudinal clines in the short time [35] since maize arrived in the highlands. And while transposable element expansion in lowland populations is not implausible, our analysis of reads mapping to individual TE families finds no evidence that this has occurred (Supp. Fig. S3).

Having concluded that natural selection is the most plausible explanation for decreasing genome size at higher elevations, we then asked whether these observations were the result of selection on genome size itself or merely a consequence of selection on specific repeat classes. After controlling for differences in genome size, our linear model shows no evidence of selection on individual repeat classes in maize (Supp. Table S1), suggesting that observed clines in other repeats are predominantly the result of selection on overall genome size. In the case of TEs, the model returned an unrealistic slope, suggesting colinearity between genome size and Mb of TEs may have lead to poor power to estimate parameters. This colinearity itself, however, suggests that TEs are not the sole contributor to genome size change — if so, we would expect TEs to show a significant result even after controlling for genome size as a covariate. Consistent with this, we find no more TE families showing significant change in abundance than expected by chance, and none of the families showing significant change could individually explain the overall changes in genome size (Supp. Fig. S3).

In contrast to results for maize, in teosinte the 350bp TR1 knob repeat shows more extreme changes in abundance across altitude than can be explained by genetic drift alone, even after accounting for changes in total genome size. Selection on genome size might be expected to act especially strongly on knobs because each locus may contain many megabases of repeats and thus represent a large contribution to variation in genome size. But while the meiotic drive exhibited by 180bp knob loci in the presence of abnormal chromosome 10 [45] might ameliorate selection against knobs, TR1 specific variants of abnormal chromosome 10 show extremely weak drive (51%; [48]), perhaps insufficient to counter selection for smaller genomes at high altitude. TR1 repeats constitute a comparably minor fraction of the genome, however, and the fact that both 180bp knob and transposable elements also show negative trends with altitude suggest that whatever the proximal cause of the striking decrease in TR1, selection on TR1 repeats is unlikely to explain overall trends in genome size.

Our repeat analyses focus on known highly abundant repeats present in the reference B73 genome. But in addition to the standard 10 chromosomes present in all maize, maize and teosinte are well known to vary in the number of auxiliary B chromosomes [49]. Because B chromosomes approach the size of normal chromosomes, they could potentially

contribute substantially to differences in genome size. Previous cytological surveys of B chromosomes in maize find both positive and negative correlations with altitude [50, 51]. Abundance estimation using short read data is challenging, however, as much of the B chromosome sequence appears to have originated from other normal chromosomes [52] and no full-length sequence of a reference B chromosome exists. There are a few classes of repeats which appear to be unique to the B chromosomes, though they exist in lower abundance than other repeats [53]. The presence of these B-specific repeats allows a superficial evaluation of B chromosomes in our data. Mapping our maize sequences to these repeats in a manner analogous to our analyses for TEs and knob repeats (see Methods), we find no evidence of a correlation between B chromosome abundance and altitude (Supp. Fig. S2), and we thus conclude that B chromosomes are unlikely to explain our observed negative correlation between genome size and altitude.

Genome size and development rate

Several authors have hypothesized that genome size could be related to rates of cell production and thus developmental timing [45, 54]. We tested this hypothesis in a growth chamber experiment in which we measured leaf elongation rates across individuals from a single population of highland teosinte with wide variation in genome size. In order to estimate the effects of genome size on the rate of cell production, we applied a modeling approach similar in nature to models used previously to investigate constraints in the scaling relationship between genome size, cell size, and cell production [55] (see Methods). We observed no correlation between genome size and cell size within our single teosinte population. This result is in striking contrast to the findings of many authors who have reported a positive correlation between genome size and cell size across species [56,57]. We suggest that processes leading to genome size differences among species may differ from those generating variation among populations of a single species. While differences among species may be largely due to long-term expansion of transposable elements, much of the variation we see is likely due to deletions of transposable elements and presence/absence of tandem repeats. In particular, we believe our observation is consistent with recent findings in *Drosophila* showing that larger repeat arrays actually led to more compact heterochromatin despite the physical presence of more DNA [58]. We speculate that such an effect may ameliorate some of the physical increase in chromosome size due to the expansion of certain repeats, especially tandem arrays such as those found in dense heterochromatic knobs.

In support of the hypothesis that smaller genomes may enable more rapid development, our leaf elongation model indicates a negative correlation between genome size and cell production in our highland teosinte population. For prior values of LER greater than 4cm/day, our model returned a significant negative relationship (Supp. Fig. S8). Though these results did show strong prior sensitivity, the sign of the relationship between genome size and cell production did not change for values of leaf elongation within the range of

those published for maize (from 4.6 cm/day [59] to 12 cm/day [60]), all equal or larger to rates observed in our experiment. Tenaillon *et al* [61] also found a negative correlation between LER and genome size, though the effect was not significant after correcting for population structure.

We hypothesize that selection on flowering time is the driving force behind our observed differences in genome size. Larger genomes require more time to replicate [55], and slower rates of cell production in turn may lead to slower overall development or longer generation times [54]. Slower cell production is unlikely to be directly limiting to the cells that eventually become the inflorescence, as only relatively few cell divisions are required [62]. However, signals for flowering derive from plant leaves [63, 64], and slower cell production will result in a longer time until full maturity of all the organs necessary for the plant to flower. Highland populations of both maize and teosinte both flower earlier than lowland populations [39, 40], and previous experimental work in maize is suggestive that selection on flowering time alone could result in changes to genome size [38]. Consistent with this idea, maize plants with more cells in their SAM at a given developmental stage (and thus slower rates of cell production) appear to also exhibit later flowering [41]. Although space limitations prevented us from growing our teosinte population to maturity, future efforts to experimentally connect genome size to both cell production and flowering time within a single panel will be important to definitively establish a mechanistic connection between genome size and flowering time.

In addition to flowering time, the metabolic requirements of nucleotide synthesis could play a selective role in determining plant genome size variation. Nucleotide synthesis requires substantial nitrogen and phosphorous, and it has been argued that selection for rapid growth in nutrient-poor environments may as a result act to reduce genome size [65]. Indeed, phylogenetic comparisons find a significant correlation between nitrogen content (but not phosphorous) and genome size among *Primulina* growing in nutrient-limited karst soils [23]. We are unaware, however, of any meaningful correlations between nitrogen or phosphorus concentration and altitude in either Mexico or the Andes, suggesting that soil nutrients are unlikely to completely explain the patterns we observe.

Conclusion

The causes of genome size variation have been debated for decades, but these discussions have often ignored intraspecific variation or disregarded adaptive explanations. Our results suggest that differences in optimal flowering times across elevations are likely indirectly effecting clines in genome size due to a mechanistic relationship between genome size and cell production and developmental rate. We speculate that our observations on genome size and cell production may apply broadly across plant taxa. Intraspecific variation in genome size appears a common feature of many plant species, as is the need to adapt to a range of abiotic environments. Cell production is a fundamental process that retains many similar characteristics across plants, and genome size is likely to impact

cell division due to the limitations in replication kinetics that result from having a larger genome. Together, these considerations suggest that genome size itself may be a more important adaptive trait than has been previously believed.

Materials and Methods

Unless otherwise specified, raw data and code for all analyses are available on the project Github at (available upon publication) Table S2 shows the general relationship among samples and analyses; additional details are included below.

Genome Size

We sampled one seed from each of 83 maize landrace accessions collected across a range of altitudes in Mesoamerica and South America to quantify genome size (Supp. Table S4; [43]). For comparison to maize, we sampled two seeds from 6 and 10 previously collected populations of *parviglumis* and *mexicana*, respectively (Supp. Table S7; [36]). For our growth chamber experiment, we sampled 201 total seeds from 51 maternal plants collected from 11 populations of *mexicana* (Supp. Table S5 and S8). Finally, to assess the error associated with flow cytometry measures of genome size, we used 2 technical replicates of each of 35 maize inbred lines (Supp. Table S9). We germinated seeds and grew plants in standard greenhouse conditions. We collected samples of leaf tissue from each individual and sent material to Plant Cytometry Services (JG Schijndel, NL) for genome size analysis. *Vinca major* was used as an internal standard for flow cytometric measures. Replicated maize lines showed highly repeatable estimates (corr = 0.92), with an average difference of only 0.0346pg/1C between estimates.

Genotyping

We took advantage of published array genotyping [66] data for 35,468 SNPs from different individuals of the same maize landraces used for flow cytometry [43]. For the 11 *mexicana* populations used in our linear model, we used genotyping-by-sequencing [67] SNP data from [37] of the same individuals of *mexicana* used for flow cytometry. We filtered the imputed *mexicana* data with TASSEL (v5.20160408) [68] to remove samples with >45% missing data (da6, tx12) and retaining 93 individuals for kinship analyses. After filtering SNPs with minor allele frequency <0.01 and removing any sites with missing data, the resulting 29,526 SNPs were used to calculate a kinship matrix for *mexicana* in rrBLUP [69]. For admixture analyses, we used GBS data from a diverse set of 282 previously published maize inbreds (AllZeaGBSv2.7, [68, 70] ,`www.panzea.org`) and combined with our panel of 11 *mexicana* populations for a total of 374 individuals. We ran STRUCTURE [71] (burn-in=10000, numreps=20000) on the combined set of 374 individuals and the 8,474 polymorphic SNPs with no missing data. We visualized results from STRUCTURE

using CLUMPP [72]. Inbreeding statistics for individual *mexicana* plants were calculated following Ritland [73] from the filtered teosinte GBS data used to estimate the relatedness matrix.

Shotgun Sequencing

We used whole genome shotgun sequencing to estimate repeat abundance in the same 83 maize landrace accessions and 95 *mexicana* individuals for which we estimated genome size, as well as an additional set of *mexicana* individuals used to validate the approach cytologically (see below). In maize landrace accessions, repeat abundance and flow cytometry measures of genome size were performed on the same individual, while genotyping was performed on different individuals within the same accession. DNA was isolated from leaf tissue using the DNeasy plant extraction kit (Qiagen) according to the manufacturers instructions. Samples were quantified using a Qubit (Life Technologies) and 1ug of DNA was fragmented using a bioruptor (Diagenode) with cycles of 30 seconds on, 30 seconds off. DNA fragments were then repaired with the End-Repair enzyme mix (New England Biolabs), and a deoxyadenosine triphosphate was added at each 3'end with the Klenow fragment (New England Biolabs). Illumina Truseq adapters (Affymetrix) were then added with the Quick ligase kit (New England Biolabs). Between each enzymatic step, DNA was washed with sera-mags speed beads (Fisher Scientific). Samples were multiplexed using Illumina compatible adapters with inline barcodes and sequenced in 3 lanes of a Miseq (UC Davis Genome Center Sequencing Facility) for 150 paired-end base reads with an insert size of approximately 350 bases. The first lane included all maize landraces used for selection studies, the second had the *mexicana* populations used for FISH correlations, and the third included all *mexicana* samples used for selection studies.

Estimating Repeat Abundance

We gathered reference sequences for 180bp knob, TR1 knob, and rDNA repeats from NCBI and CentC repeats from [74]. Additionally, 23 repeats specific to the B chromosome were identified on NCBI and regions with greater than 80% homology to at least a 30bp alignment to the maize A genome were masked for abundance mapping (sequences and mapping tags available on github). Chloroplast DNA and mitochondrial DNA were taken from the maize reference genome (v2, www.maizesequence.org). For the transposable element database, we began with the TE database consensus sequences from [25, 75]. We then matched sequences against themselves using BLAST and masked any shared regions, retaining unique regions that were at least 70bp in length in our mapping reference. This process was repeated to eliminate regions of secondary homology and repeated using knobs and CentC to remove regions of homology to tandem repetitive sequences. For the 23 B-specific repeats on NCBI, we BLASTed sequences against the maize genome (v2, www.maizesequence.org), and masked any regions that had alignments of greater

than 30bp with 80% homology. Unmasked regions greater than 70bp were used as a mapping reference for B-repeat abundance. We mapped sequence reads to our repeat library using bwa-mem [76] with parameters -B 2 -k 11 -a to store all hit locations with an identity threshold of approximately 80%. We used a minimum seed length of 11 as it produced the most reads mapping against the full transposable element database, though the abundance was still lower (63-71%) than previous estimates [25]. To standardize comparisons of repetitive content across individuals, we first filtered out plastid sequences, then calculated Mb of sequence for each repeat class by multiplying its relative abundance in our sequencing data by genome size converted to base pair values. Previous simulations suggest that this estimate has good precision and accuracy in capturing relative differences across individuals [**?**].

Repeat Abundance Validation via FISH

We selected two individuals each from 10 previously collected populations of *mexicana* [35] for fluorescence *in situ* hybridization counts of knob content (FISH; Supp. Table S7). FISH probe and procedures closely followed [31]. Different individuals from these same populations were used in the initial subspecies comparison of genome size with *parviglumis*. From each population we also shotgun sequenced 9-12 individuals per population for the 10 *mexicana* used in the FISH to sequence comparisons (Fig. 2 and Supplementary Fig. S3).

Clinal Models of Genome Size and Repeat Abundance

We model genome size as a phenotype whose value is a linear function of altitude and kinship (Equation 1). We assume genome size has a narrow sense heritability $h^2 = 1$, as it is the sum of the base pairs inherited from both parents. In our model P is our vector of phenotypes, μ is a grand mean, A is a vector of altitudes included as a fixed effect, g is the effect of neutral relatedness, which is modeled as a random effect of with covariance structure given by the kinship matrix ($\mathbf{K}$), and ε captures an uncorrelated error term. The coefficient β_{alt} of altitude then represents selection along altitude, while the additive genetic (V_A) and error (V_ϵ) variances are nuisance parameters.

$$\begin{aligned} P &= \mu + \beta_{alt} * A + g + \varepsilon \\ g &\sim MVN\left(0, V_A \mathbf{K}\right) \\ \varepsilon &\sim N\left(0, V_\epsilon\right) \end{aligned} \tag{1}$$

We implemented our linear model in EMMA [77], and used it to test for selection on genome size. In supplementary table S1, Equation 1 was only used to calculate β values for genome size. To test for selection on components of genome size, we include genome size (GS) as a predictor in order to learn whether the abundance of a specific

repeat class (conditional on GS) may be under selection (Equation 2). Equation 2 was used to estimate β for the components of genome size (TEs, 180bp knob, and TR1 knob) in Supplementary Table S1.

$$P = \mu + \beta_{alt} * A + \beta_{GS} * GS + g + \varepsilon \quad (2)$$
$$g \sim MVN\left(0, V_A \mathbf{K}\right) \quad (3)$$
$$\varepsilon \sim N\left(0, V_\epsilon\right) \quad (4)$$

Growth Chamber Experiment

We conducted a growth chamber experiment to investigate whether genome size variation has an impact on cell production and leaf elongation. We sampled 202 individuals from a single high elevation (2408m) *mexicana* population at Tenango Del Aire selected for its wide intrapopulation variation in genome size as seen in our altitudinal transect of *mexicana* populations with GBS data. We soaked each fruit in water for 24 hours, then manually removed most of the fruitcase and placed the seed in a Petri dishes inside a growth chamber (23°C, 16h Light / 8h dark) with cotton balls and water to prevent drying.

Germinated seedlings were transferred to soil pots and into a growth chamber ($23^{\circ C}$, 16h Light / 8h dark). Individually potted seedlings were randomly placed in trays, given fertilized water via bottom watering, and monitored for third adult leaf emergence. We measured leaf elongation rate from the first visible emergence of the third leaf. Leaf length was measured daily from the base of the plant for 3 days post emergence. We clipped the first 8cm of leaf material from the tip of the measured leaf, then extracted a 1cm section which was dipped in propidium iodide (.01mg/ul) for fluorescent imaging (10x magnification, emission laser 600-650, excitation 635 at laser power 6). A minimum of 5 non-overlapping images were taken per leaf sample, horizontally across the leaf segment if possible. Cell length was measured for multiple features, including stomatal aperture size and rows adjacent to stomata. Lengths across different features were highly correlated, so stomatal aperture size was used as the repeated measure of cell lengths in the growth model.

Modeling the Effect of Genome Size on Cell Production

We model leaf elongation (LER) as the product of the rate of cell production (CP) and cell size (CS). We impose constraints on the model stating that CP and CS both have a linear effect on leaf elongation, and that genome size (GS) has no direct effect. The function for leaf elongation is thus expressed as:

$$CP * CS = LER \quad (5)$$

The multiplicative expression in Equation 5 is linearized by taking the natural logarithm on both sides of the equation, and model-fitting is performed on the log scale. The models connecting GS to CS, CP, and LER are then:

$$log(CS) = \gamma_0 + \gamma_{GS} * log(GS) \tag{6}$$

$$log(CP) = \beta_0 + \beta_{GS} * log(GS) \tag{7}$$

$$log(LER) = \tau_0 + \tau_{GS} * log(GS) \tag{8}$$

We hypothesize that genome size affects LER only through its effects on CS and CP; thus the effect of genome size on LER in equation 6 is marginal to CS and CP. Our growth chamber experiment produces measures of leaf elongation, cell size, and genome size, but cell production is not directly measured. We instead estimate the effect of genome size on cell production by subtraction:

$$\beta_{GS} = \tau_{GS} - \gamma_{GS} \tag{9}$$

The estimation strategy is illustrated in the path diagrams shown in Supp. Figure S9a,b. We adopt a computational Bayesian strategy for parameter estimation that incorporates individual and maternal level random intercepts in equations 6-8. Our posterior parameter densities, and therefore conclusions, are sensitive to different specifications of prior information. We identified previous averages for stomatal cell size and daily leaf elongation (CS=0.003 cm, LER=4.0-4.8cm/day or 2mm/hr) [59,78–80]. We incorporated these informative priors into specification of the individual level random effects, the level at which there is the most experimental replication. A prior density for LER is required, as LER is derived from daily observations of leaf length. Because our model shows prior sensitivity, we also identify prior LERs for which the sign of the relationship between genome size and cell production (β_{GS}) changes (Supp. Fig S8). Changes to the assumption of linear growth had no effect, while changes in our prior CS and LER densities shifted values of the coefficient. We generated posterior samples using JAGS, a general purpose Gibbs sampler invoked from the R statistical language using the library rjags [81]. We allowed for a burn in of 200,000 iterations and recorded 1,000 posterior estimates by thinning 500,000 iterations at an interval of 500.

Analysis of Maize SAM Cell Number and Flowering Time

To evaluate evidence for a relationship between cell production and flowering time, we used flowering time and meristem cell number data for 14 maize inbred lines from [41]. Because meristems were sampled at identical growth stage and time point, differences in cell number should reflect differences in the rate of cell division. We fitted a mixed linear

model to estimate the best linear unbiased estimates (BLUEs) of the cell counts for each growth period separately:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon \tag{10}$$

In this model, Y_{ij} is the cell count value of the i^{th} genotype evaluated in the j^{th} replicate; μ, the overall mean; α_i, the fixed effect of the i^{th} genotype; β_j, the random effect of the j^{th} block; and ε, the model residuals.

Each line's genotype at trait-associated SNPs for the candidate genes BAK1 and SDA1 [41] was considered as a fixed effect and replication as a random effect. We then fitted mixed linear models to study the relationship of flowering time and cell counts by controlling the population structure and known trait associated SNPs:

$$\begin{aligned} Y_i &= \mu + \alpha_i G_i + \beta_{BAK1} + \beta_{SDA1} + g + \varepsilon \\ g &\sim MVN\left(0, V_A \mathbf{K}\right) \\ \varepsilon &\sim N\left(0, V_E\right) \end{aligned} \tag{11}$$

Here Y_i is the flowering time (days to anthesis) of the i^{th} genotype; μ, the overall mean; α_i, the fixed effect of the i^{th} Genotype; β_{BAK1} and β_{SDA1} the fixed effects of the BAK1 and SDA1 loci; g a random effect modeled with a covariance structure given by the kinship matrix $\mathbf{K}$; and ε an uncorrelated error. The additive genetic (V_A) and environmental (V_E) variances are nuisance parameters.

Cell counts were included as fixed effects and the standardized genetic relatedness matrix was fitted as a random effect to control for the population structure [82]. The genetic relatedness matrix was calculated using GEMMA [83] using publicly available GBS genotyping for these lines (AllZeaGBSv2.7 at `www.panzea.org`, [68]). In the calculation, we used 349,167 biallelic SNPs after removing SNPs with minor allele frequency <0.01 and missing rate >0.6 using PLINK [84].

Acknowledgments

We would like to thank Michelle Stitzer and Graham Coop as well as members of the Ross-Ibarra, Coop, and Burbano labs for helpful discussion. We thank Anna O'Brien for providing seed from her *mexicana* collections and for early access to her SNP data. We acknowledge financial support from NSF grants IOS-0922703 and IOS-1238014 and the USDA Hatch project CA-D-PLS-2066-H. P.B. would like to thank the UC MEXUS Dissertation Grant, DuPont Pioneer, and the UC Davis Department of Plant Sciences for funding and support.

References

1. Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131: 452–462.

2. Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49–63.

3. Pagel M, Johnstone RA (1992) Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. Proceedings of the Royal Society of London B: Biological Sciences 249: 119–124.

4. Gregory T (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biological Reviews 76: 65–101.

5. Lynch M, Conery JS (2003) The origins of genome complexity. Science 302: 1401–1404.

6. Kuo CH, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. Genome Research 19: 1450–1454.

7. Lefébure T, Morvan C, Malard F, François C, Konecny-Dupré L, et al. (2017) Less effective selection leads to larger genomes. Genome Research : gr–212589.

8. Whitney KD, Baack EJ, Hamrick JL, Godt MJW, Barringer BC, et al. (2010) A role for nonadaptive processes in plant genome size evolution? Evolution 64: 2097–2109.

9. Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, et al. (2005) Evolution of genome size in Brassicaceae. Annals of Botany 95: 229–235.

10. Ågren JA, Greiner S, Johnson MT, Wright SI (2015) No evidence that sex and transposable elements drive genome size variation in evening primroses. Evolution 69: 1053–1062.

11. Baetcke K, Sparrow A, Nauman C, Schwemmer SS (1967) The relationship of DNA content to nuclear and chromosome volumes and to radiosensitivity (LD50). Proceedings of the National Academy of Sciences 58: 533–540.

12. Pegington C, Rees H, et al. (1970) Chromosome weights and measures in the Triticinae. Heredity 25: 195–205.

13. Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, et al. (2007) Correlated evolution of genome size and seed mass. New Phytologist 173: 422–437.

14. Gregory TR, Hebert PD, Kolasa J (2000) Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. Heredity 84: 201–208.

15. Cavalier-Smith T (1978) Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. Journal of Cell Science 34: 247–278.

16. Gregory TR, Hebert PD (1999) The modulation of DNA content: proximate causes and ultimate consequences. Genome Research 9: 317–324.

17. Knight CA, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology and phenotype. Annals of Botany 95: 177–190.

18. Greilhuber J (1998) Intraspecific variation in genome size: a critical reassessment. Annals of Botany 82: 27–35.

19. Šmarda P, Bureš P, et al. (2010) Understanding intraspecific variation in genome size in plants. Preslia 82: 41–61.

20. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, et al. (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. Nature Genetics 45: 884–890.

21. Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, et al. (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. New Phytologist 199: 264–276.

22. Zaitlin D, Pierce AJ (2010) Nuclear DNA content in *Sinningia* (Gesneriaceae); intraspecific genome size variation and genome characterization in *S. speciosa*. Genome 53: 1066–1082.

23. Kang M, Wang J, Huang H (2015) Nitrogen limitation as a driver of genome size evolution in a group of karst plants. Scientific Reports 5: 11636.

24. Poggio L, Rosato M, Chiavarino AM, Naranjo CA (1998) Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). Annals of Botany 82: 107–115.

25. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

26. Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J (2011) Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. Genome biology and evolution 3: 219–229.

27. Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *BZ* locus. Proceedings of the National Academy of Sciences 103: 17644–17649.

28. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. The Plant Cell 17: 343–360.

29. Dennis E, Peacock W (1984) Knob heterochromatin homology in maize and its relatives. Journal of Molecular Evolution 20: 341–350.

30. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, et al. (2012) Maize hapmap2 identifies extant variation from a genome in flux. Nature Genetics 44: 803–807.

31. Albert P, Gao Z, Danilova T, Birchler J (2010) Diversity of chromosomal karyotypes in maize and its relatives. Cytogenetic and Genome Research 129: 6–16.

32. Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genetics 10: e1004412.

33. Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. Genetics 181: 1399–1413.

34. Hufford MB, Martínez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. PLoS One 7: e47659.

35. Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, et al. (2013) The genomic signature of crop-wild introgression in maize. PLoS Genetics 9: e1003477.

36. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. Genome Biology and Evolution 5: 1594–1609.

37. O'Brien AM, Ross-Ibarra J. Teosinte genotype-by-sequencing: central highland populations. URL `http://dx.doi.org/10.6084/m9.figshare.4714030`.

38. Rayburn AL, Dudley J, Biradar D (1994) Selection for early flowering results in simultaneous selection for reduced nuclear DNA content in maize. Plant Breeding 112: 318–322.

39. Jiang C, Edmeades G, Armstead I, Lafitte H, Hayward M, et al. (1999) Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. Theoretical and Applied Genetics 99: 1106–1119.

40. Rodriguez F J, Sanchez G J, Baltazar M B, de la Cruz L L, Santacruz-Ruvalcaba F, et al. (2006) Characterization of floral morphology and synchrony among *Zea* species in Mexico. Maydica 51: 383–398.

41. Leiboff S, Li X, Hu HC, Todt N, Yang J, et al. (2015) Genetic control of morphometric diversity in the maize shoot apical meristem. Nature Communications 6.

42. van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, et al. (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proceedings of the National Academy of Sciences 108: 1088–1092.

43. Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, et al. (2015) Independent molecular basis of convergent highland adaptation in maize. Genetics 200: 1297–1312.

44. Rayburn AL, Auger J (1990) Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. Theoretical and Applied Genetics 79: 470–474.

45. Buckler ES, Phelps-Durr TL, Buckler CSK, Dawe RK, Doebley JF, et al. (1999) Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153: 415–426.

46. Bennett MD (1987) Variation in genomic form in plants and its ecological implications. New Phytologist 106: 177–200.

47. Sinha RP, Häder DP (2002) UV-induced DNA damage and repair: a review. Photochemical & Photobiological Sciences 1: 225–236.

48. Kanizay LB, Albert PS, Birchler JA, Dawe RK (2013) Intragenomic conflict between the two major knob repeats of maize. Genetics 194: 81–89.

49. Kato Yamakake TA, et al. (1976) Cytological studies of maize [*Zea mays* L.] and teosinte [*Zea mexicana* Schrader Kuntze] in relation to their origin and evolution .

50. Rosato M, Chiavarino A, Naranjo C, Hernandez J, Poggio L (1998) Genome size and numerical polymorphism for the B chromosome in races of maize (*Zea mays* ssp. *mays*, Poaceae). American Journal of Botany 85: 168–168.

51. Bretting P, Goodman M (1989) Karyotypic variation in mesoamerican races of maize and its systematic significance. Economic Botany 43: 107–124.

52. Alfenito MR, Birchler JA (1993) Molecular characterization of a maize B chromosome centric sequence. Genetics 135: 589–597.

53. Stark EA, Connerton I, Bennett ST, Barnes SR, Parker JS, et al. (1996) Molecular analysis of the structure of the maize B-chromosome. Chromosome Research 4: 15–23.

54. Bennett M (1972) Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences 181: 109–135.

55. Šímová I, Herben T (2012) Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences 279: 867–875.

56. Gregory TR (2001) The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood Cells, Molecules, and Diseases 27: 830–843.

57. Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA (2008) Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist 179: 975–986.

58. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, et al. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529: 418–422.

59. Van Volkenburgh E, Boyer JS (1985) Inhibitory effects of water deficit on maize leaf elongation. Plant Physiology 77: 190–194.

60. Salah HBH, Tardieu F (1996) Quantitative analysis of the combined effects of temperature, evaporative demand and light on leaf elongation rate in well-watered field and laboratory-grown maize plants. Journal of Experimental Botany 47: 1689–1698.

61. Tenaillon MI, Manicacci D, Nicolas SD, Tardieu F, Welcker C (2016) Testing the link between genome size and growth rate in maize. Technical report, PeerJ Preprints.

62. Watson JM, Platzer A, Kazda A, Akimcheva S, Valuchova S, et al. (2016) Germline replications and somatic mutation accumulation are independent of vegetative life span in *Arabidopsis*. Proceedings of the National Academy of Sciences : 201609686.

63. Huang T, Böhlenius H, Eriksson S, Parcy F, Nilsson O (2005) The mRNA of the *Arabidopsis* gene *FT* moves from leaf to shoot apex and induces flowering. Science 309: 1694–1696.

64. Lin MK, Belanger H, Lee YJ, Varkonyi-Gasic E, Taoka KI, et al. (2007) FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the Cucurbits. The Plant Cell 19: 1488–1506.

65. Hessen DO, Jeyasingh PD, Neiman M, Weider LJ (2010) Genome streamlining and the elemental costs of growth. Trends in Ecology & Evolution 25: 75–80.

66. Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, et al. (2011) A large maize (*Zea mays* L.) snp genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6: e28334.

67. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. PLoS One 6: e19379.

68. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, et al. (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9: e90346.

69. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4: 250-255.

70. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, et al. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal 44: 1054–1064.

71. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

72. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806.

73. Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetical Research 67: 175–185.

74. Bilinski P, Distor K, Gutierrez-Lopez J, Mendoza GM, Shi J, et al. (2015) Diversity and evolution of centromere repeats in the maize genome. Chromosoma 124: 57–65.

75. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genetics 5: e1000732.

76. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25: 1754–1760.

77. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709–1723.

78. Orcen N, Nazarian G, Barlas T, Irget E (2013) Variation in stomatal traits based on plant growth parameters in corn (*Zea mays* L.). Annals of Biological Research 4: 25–29.

79. Ben-Haj-Salah H, Tardieu F (1995) Temperature affects expansion rate of maize leaves without change in spatial distribution of cell length (analysis of the coordination between cell division and cell expansion). Plant Physiology 109: 861–870.

80. Bos H, Tijani-Eniola H, Struik P (2000) Morphological analysis of leaf growth of maize: responses to temperature and light intensity. NJAS-Wageningen Journal of Life Sciences 48: 181–198.

81. Plummer M (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

82. Therneau T (2012) COXME: mixed effects Cox models. R package version 2.2-3. Vienna: R Foundation for Statistical Computing .

83. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44: 821–824.

84. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 1.

Supplemental Materials

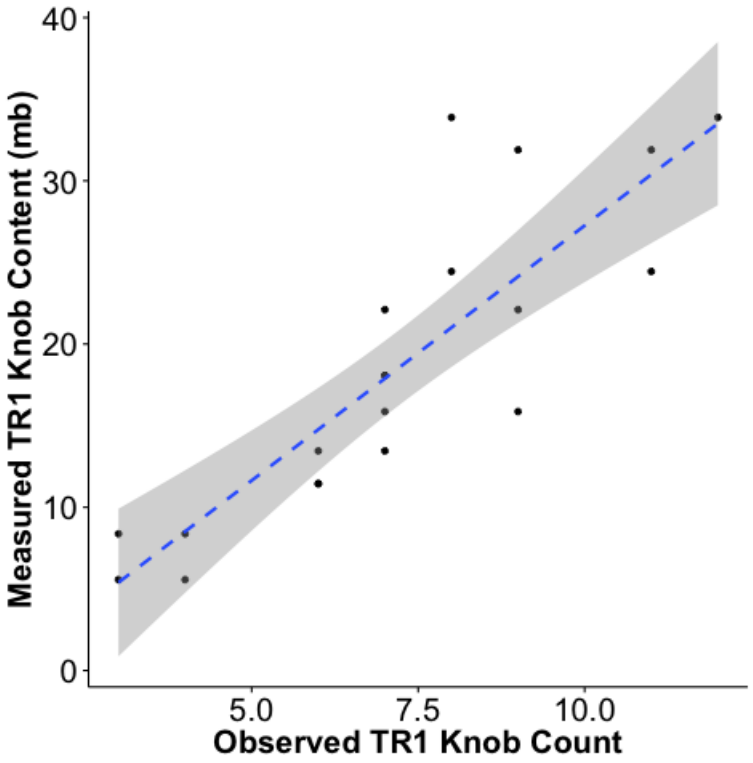

Figure S1. Plot of the population-level correlation between TR1 knob counts and TR1 knob sequence abundance for 20 *mexicana* individuals.

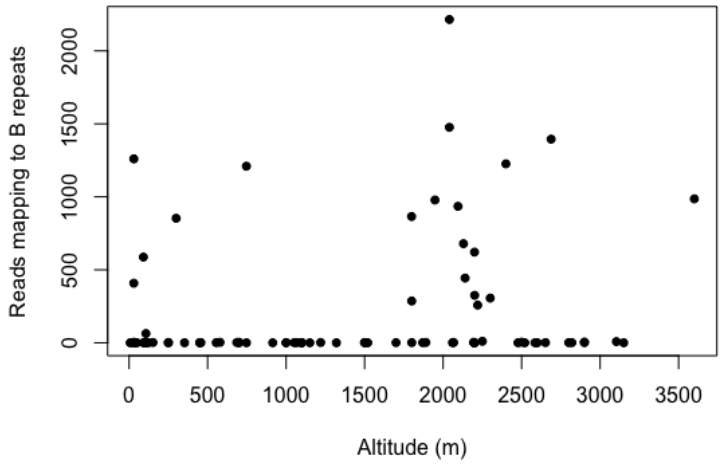

Figure S2. Plot of reads mapping to B chromosome specific repeats in maize landraces.

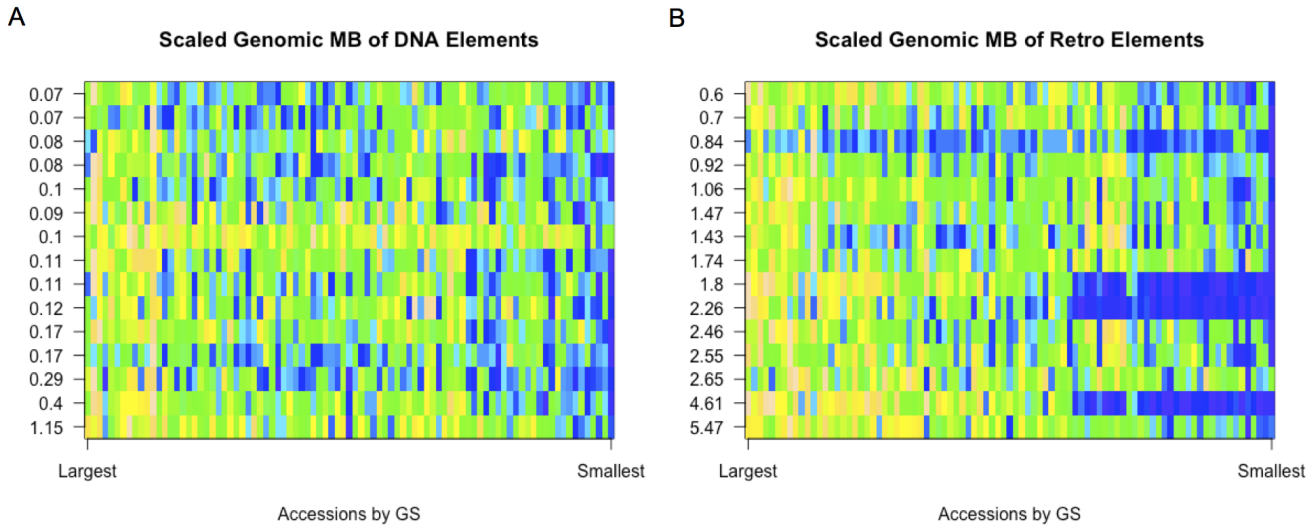

Figure S3. Variation in (A) DNA and (B) RNA transposable element abundance in maize landraces. The y-axis indicates the average abundance in Mb of a given TE subfamily. The fifteen highest abundance subfamilies are shown. The x-axis are maize landraces accessions ordered by genome size, with the largest genome size accessions on the left. Values plotted are bp measures scaled from 0 (blue) to 1 (yellow) per row.

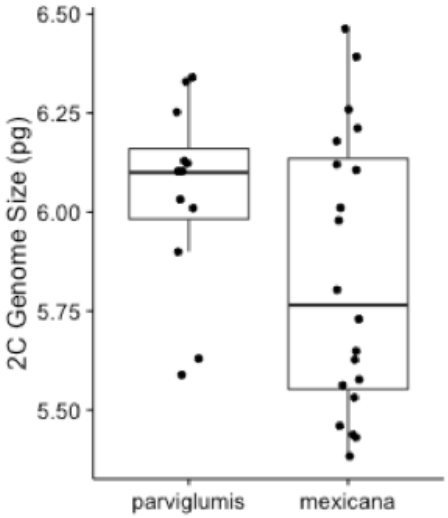

Figure S4. Genome size for two individuals per teosinte population sampled in [35, 36]. Points indicating individual genome size estimates are jittered around the center.

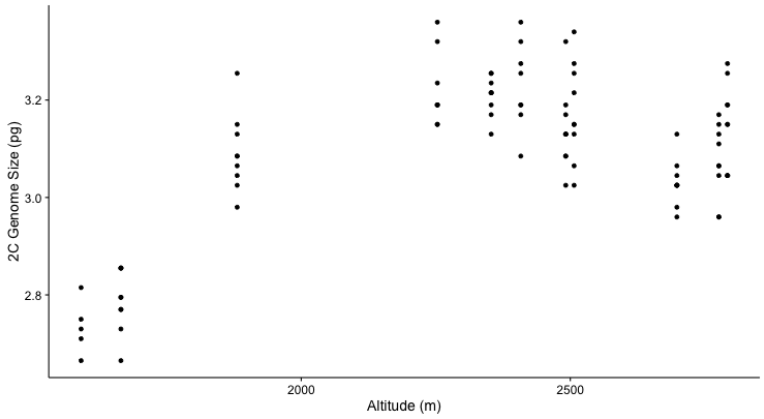

Figure S5. Genome size by altitude of *mexicana*. All samples, including low elevation admixed populations are shown.

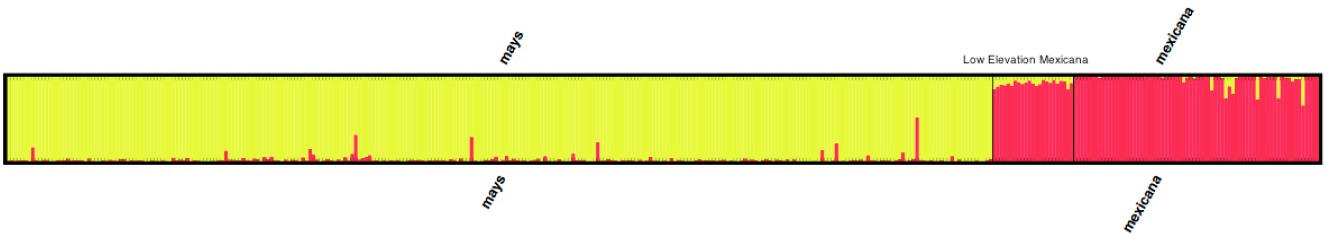

Figure S6. Structure plot of maize and *mexicana* populations. Maize populations consist of the 282 panel (see [70]). Teosinte populations are arranged in altitudinal order.

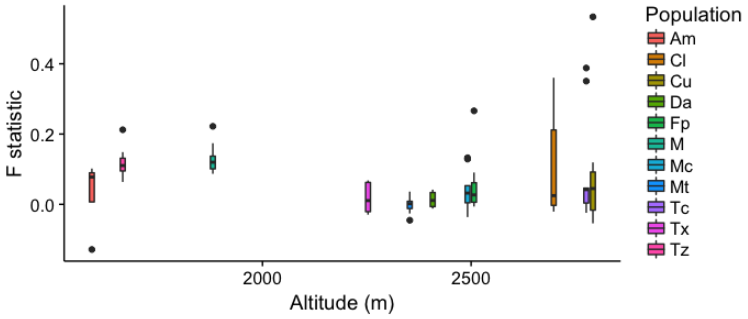

Figure S7. Boxplot of F statistic by *Mexicana* population.

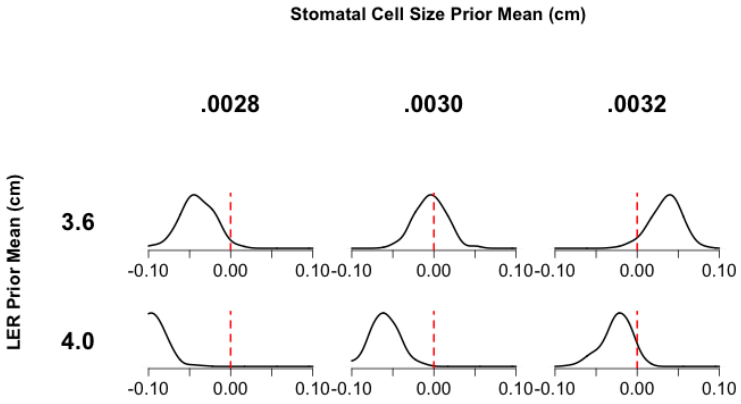

Figure S8. Effect of LER and Stomatal Cell Size Priors on Posterior Density of the Cell Production Coefficient β_{GS}.

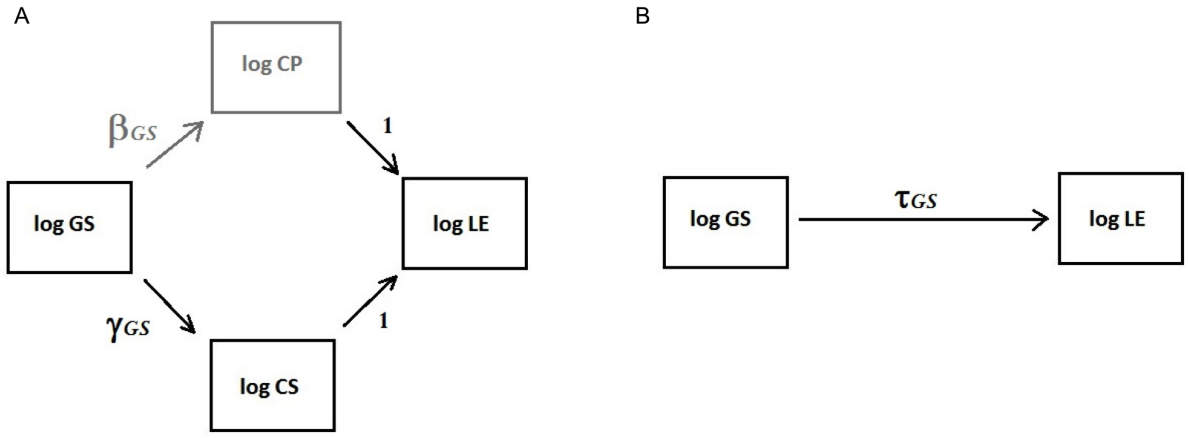

Figure S9. Path models for estimation of genome size effects. Arrows indicate predictor-outcome relationships and are annotated with model coefficients (slopes) from equations 3-6. A: Genome Size (GS) predicts Leaf Elongation rate (LE) through the mediators Cell Size (CS) and Cell Production rate (CP). CP, shown in grey, is not directly observed. The unit coefficients connecting log LE with log CP and log CS reflect the assumption CP*CS = LE (equation 3). B: Marginal model for the effect of GS on LE (equation 6). The path diagrams, along with equations 4-6, produce two distinct expressions for log LE. When equated, these yield: $\log(LE) = \gamma_0 + \gamma_{GS}$ *log(GS) + $\beta_0 + \beta_{GS}$ * log(GS) = $\tau_0 + \tau_{GS}$ * log(GS). Collecting terms, the coefficient of log CP is obtained by subtraction as $\beta_{GS} = \tau_{GS}$ - γ_{GS}.

Table S1. Altitudinal coefficients from selection models using maize landraces and highland teosinte

Taxa	Maize	Mexicana
Genome Size	-0.110**	-0.26**
180bp knob	-0.0117	0.0032
TR1 Knob	-0.0037	-0.0093**
TE	-1.9E6	0.07

Calculated altitudinal coefficients (β) from the models testing for altitudinal selection. β values are given in units of megabases per meter. *=p-value<0.05; **=p-value<0.005

Table S2. Description of data sets used in each analysis.

Data Group	Genome Size	WGS sequencing Individual	Genotyping Type	Genotyping Individual	FISH performed?	Used in
83 Landraces	Per Individual	Same individual as GS	55K CHIP	Different individual than GS but within accession	No	Selection study
95 Mexicana individuals from 11 Populations	Per Individual	Same individual as GS	GBS	Same individual as GS	No	Selection study
6 Parviglumis Populations	2 Individuals per population	NA	NA	NA	No	Teosinte pilot study
10 Mexicana populations	2 Individuals per population	9-12 Individuals per population	NA	NA	Yes, different individuals than GS	Teosinte pilot study

Table S3. *Mexicana* Population IDs and number of individuals used for FISH analyses

Population ID	**Number of individuals**
RIMME0021	12
RIMME0026	12
RIMME0028	12
RIMME0029	12
RIMME0030	12
RIMME0031	12
RIMME0032	12
RIMME0033	12
RIMME0034	9
RIMME0035	12

Table S4. Geographic information for maize landrace accessions

Accession ID	Region	Altitude (m)	Latitude	Longitude
RIMMA0388.1	SAL	1500	6.85	-75.28333333
RIMMA0389.1	SAL	7	10.38333333	-74.88333333
RIMMA0390.1	SAL	353	4.516666667	-75.63333333
RIMMA0391.1B	SAL	700	NA	NA
RIMMA0392.1	SAL	555	1.75	-75.58333333
RIMMA0393.1	SAL	100	8.316666667	-75.15
RIMMA0394.2	SAL	1000	4.783333333	-74.68333333
RIMMA0395.2	SAL	30	8.5	-77.26666667
RIMMA0396.1	SAL	1100	2.583333333	-75.3
RIMMA0397.1	SAL	50	11.55	-72.91666667
RIMMA0398.1	SAL	27	9.433333333	-75.7
RIMMA0399.1	SAL	250	10.18333333	-74.05
RIMMA0403.2	SAL	1000	1.25	-77.51666667
RIMMA0404.1	SAL	1500	7.3	-72.51666667
RIMMA0405.1B	SAL	1100	NA	NA
RIMMA0406.1	SAL	450	4.966666667	-74.9
RIMMA0407.1	SAL	450	NA	NA
RIMMA0409.1	ML	107	15.43333333	-92.9
RIMMA0410.1B	ML	107	NA	NA
RIMMA0416.1	MH	2140	29.35	-107.75
RIMMA0417.1	MH	2040	28.55	-107.4833333
RIMMA0418.1	MH	2040	28.56666667	-107.4833333
RIMMA0421.1	MH	2250	19.85	-97.98333333
RIMMA0422.1	MH	2200	19.1	-98.3
RIMMA0423.1	MH	2095	29.21666667	-108.1333333
RIMMA0424.1	MH	2400	27.98333333	-107.5833333
RIMMA0425.1	MH	2130	26.81666667	-107.0666667
RIMMA0426.1	SAH	2500	-9.066666667	-77.81666667
RIMMA0428.1	SAL	700	-9.3	-76
RIMMA0430.1	SAH	2585	-9.166666667	-77.73333333
RIMMA0431.1	SAH	2900	-8.65	-77.08333333
RIMMA0433.1	ML	457	14.71666667	-89.5
RIMMA0436.1	SAH	2200	-6.15	-77.91666667
RIMMA0437.1	SAH	2688	-9.266666667	-77.63333333
RIMMA0438.1	SAH	2820	-8.7	-77.38333333
RIMMA0439.1	SAH	2900	NA	NA
RIMMA0441.1	ML	1320	19.16666667	-96.96666667

RIMMA0462.1	SAL	1700	1.6	-77.15
RIMMA0464.1	SAH	1800	-12.33333333	-74.7
RIMMA0465.1	SAH	2300	-5.566666667	-79.53333333
RIMMA0466.1	SAH	3600	-14.31666667	-72.91666667
RIMMA0467.1	SAH	2800	-13.58333333	-72.91666667
RIMMA0468.1	SAH	3150	-9.383333333	-77.16666667
RIMMA0473.1	SAH	3104	1.083333333	-77.61666667
RIMMA0614.1	MH	2060	19.95	-103.7666667
RIMMA0615.1	ML	152	20.13333333	-97.2
RIMMA0616.1	MH	1800	20.81666667	-102.7666667
RIMMA0619.1	ML	747	18.35	-99.53333333
RIMMA0620.1	MH	1799	20.21666667	-100.8833333
RIMMA0621.1	MH	1870	21.11666667	-101.6833333
RIMMA0623.1	MH	2520	20.03333333	-103.6833333
RIMMA0625.1	MH	2600	19	-97.38333333
RIMMA0626.2	MH	2260	19.86666667	97.58333333
RIMMA0628.1	ML	300	23.31666667	-99.01666667
RIMMA0630.1	MH	2220	19.8	-97.25
RIMMA0657.1	SAH	2201	-17.5	-65.66666667
RIMMA0658.1	SAH	1948	-21.83333333	-64.13333333
RIMMA0661.1	SAH	2195	-2.85	-78.66666667
RIMMA0662.1	SAH	2195	0	-78
RIMMA0663.1	SAH	2067	0.433333333	-78.2
RIMMA0667.1	SAH	2201	-21.83333333	-64.13333333
RIMMA0671.1	MH	2477	14.76666667	-91.25
RIMMA0674.1	MH	2652	19.28333333	-99.66666667
RIMMA0680.1	MH	1890	20.36666667	-102.1666667
RIMMA0690.1	SAL	250	8.316666667	-73.63333333
RIMMA0691.1	SAL	1098	6.55	-73.13333333
RIMMA0696.1	ML	30	16.51666667	-90.16666667
RIMMA0700.1	ML	579	16.75	-93.16666667
RIMMA0701.1	ML	686	16.6	-92.71666667
RIMMA0702.1	ML	1052	14.46666667	-90.75
RIMMA0703.1	ML	30	20.83333333	-88.51666667
RIMMA0708.1	SAL	1098	-3.5	-78.6
RIMMA0709.1	ML	747	16.5	-92.5
RIMMA0710.1	ML	91	15.33333333	-92.63333333
RIMMA0712.1	ML	1220	15.28333333	-90.25
RIMMA0716.1	ML	91	15.31666667	-92.66666667
RIMMA0720.1	ML	39	15.46666667	-88.85
RIMMA0721.1	ML	915	14.61666667	-90.08333333

RIMMA0727.1	ML	1151	14.4	-90.46666667
RIMMA0729.1	ML	122	15.4	-89.66666667
RIMMA0730.1	ML	1067	14.48333333	-90.8
RIMMA0731.1	ML	1520	16.78333333	-96.66666667
RIMMA0733.1	ML	107	16.56666667	-94.61666667

Table S5. Geographic information for teosinte populations used in selection studies.

Population	Latitude	Longitutde	Altitude (M)	Locality
TZ	18.975744992	-99.069713429	1665.116699	Tepoztln
FP	19.211739153	-99.126956714	2506.984619	San Francisco Pedregal
MT	19.211638905	-98.808725439	2352.93457	San Mateo Tezoquipan
DA	19.145727055	-98.862849986	2408.450439	Tenango del Aire
MC	19.076366471	-98.84329861	2501.216797	San Matas Cuijingo
M	18.953779545	-99.501451766	1882.853516	Malinalco
TC	19.260426778	-99.722122969	2776.151855	Toluca
TX	19.504571417	-98.922480522	2252.718018	Texcoco de Mora
CL	19.151357012	-99.616249725	2697.564941	Calimaya Lower
CU	19.160612	-99.632908	2792	Calimaya Upper
AM	18.97155	0-99.036917	1591	Amatln

Table S6. Measured growth rate for Tenango del Aire population in the growth chamber experiment. In plant ID's, the first digit indicates the mother, while the second is a unique identifier for each individual.

ID	Genome Size (pg/2C)	Initial	24hr (cm)	48hr (cm)	72hr (cm)
1-1	6.11	1.7	6.6	11.3	17.6
1-11	5.98	1	3.7	7.6	12.7
1-12	5.94	3	6.8	11.1	16.2
1-14	5.86	7	11.3	16.3	20.5
1-4	6.29	5.3	9.6	18	NA
1-5	5.92	9.4	10.8	12	12.5
1-6	5.72	3.5	7	11.5	NA
1-7	5.91	5.8	8	10.2	11.9
1-8	6.14	4.9	7.7	11.3	14.6
11-2	5.92	7.6	9	12.6	17.9
11-3	5.8	2.5	6.9	11.6	16.8
11-4	5.8	6.4	11.3	17	21.2
11-6	5.84	2.9	6.1	9.6	13.5
3-2	6.26	7.7	9.4	14.8	17
3-4	6.05	7.4	12.7	16.9	20.9
3-6	6.09	2.9	3.9	12.5	17
3-7	6.22	8.9	14.8	21	26
4-1	5.75	4.5	10	15.6	22.5
4-2	5.92	7.5	11.3	13.4	18.6
4-4	6.34	4.5	8.6	13.5	18.3
4-6	6.13	5.7	8.9	10	14.6
4-7	5.96	6	11.3	17.6	24.3
8-1	5.96	6.1	9.1	13.1	16.5
8-10	6.01	4.1	7.5	11.5	12.7
8-11	5.92	3.5	7.3	13	19.4
8-12	5.84	3.5	6.8	11.2	16.4
8-13	6.05	8	13.1	17.3	21.8
8-14	5.75	3.5	7.3	10.9	15.5
8-15	6.09	2.1	6	11.3	16.6
8-16	6.13	5.1	10.3	16.5	23
8-2	6.17	6	9.5	14	17.5
8-3	6.55	4.7	9.2	12.5	18.2
8-4	6.34	4.2	7	10.6	16
8-5	6.43	3	5.1	7.9	11.2

8-7	6.01	8.6	13.6	18.9	NA
8-8a	6.51	6.3	8	NA	10.7
8-8b	6.09	2.5	4.1	7.8	12.2
8-9	6.22	8.1	12	16.8	22.1
a-1	5.92	2	5.9	9.9	14.7
a-10	6.01	3.5	4.8	7.2	10.5
a-12	5.84	3.2	6.5	10	13.2
a-2	5.71	7.5	12.9	18.5	23
a-3	5.96	14.1	16.2	19	20.1
a-4	6.09	13	18	22.6	26
a-5	5.96	5.5	9.2	14.5	19.3
a-6	5.92	5.3	7.7	12.2	16
a-7	5.8	5.4	7.2	10.5	15
a-8	6.22	4.2	8.6	14	18.9
a-9	6.05	8	11.4	15	18.1
b-1	6.01	3.3	5.5	9.5	15
b-2	6.55	2	5.6	9.6	13.9
b-3	5.8	6.2	11.3	16.8	19.2
b-4	6.3	8.3	12.5	14	17.5
b-6	5.96	9.4	12.9	18.3	24.6
b-7	5.92	2.4	5.2	12	17.4
bulk-1	6.01	8.6	12.1	15.9	20.5
bulk-10	6.43	2	5.5	10.5	16
bulk-11	6.17	5.5	9.2	14.8	18.5
bulk-12	5.71	8.4	13.3	16.2	22.1
bulk-13	5.96	5.1	8.9	12	17
bulk-14	6.13	5.3	8.1	12.7	17.3
bulk-16	5.96	1.6	4.5	8.2	14
bulk-17	6.05	2.7	6.5	11.4	15.2
bulk-18	5.92	1.2	4.5	7.5	12.5
bulk-19	6.07	4	6	10	14.6
bulk-20	6.1	3.7	6.3	11.1	14.7
bulk-21	5.98	1.8	4.8	6.9	13.7
bulk-22	6.01	6.1	9	12.8	18.1
bulk-23	5.96	0.9	6	8	13.6
bulk-24	5.98	3	4.4	8	13.6
bulk-25	5.88	8.6	13.9	16.6	22
bulk-26	5.85	3	6.9	10.9	15.1
bulk-28	5.85	2	6	10.5	15.9
bulk-29	5.85	7.7	11.8	16.2	20.5
bulk-30	6.11	5.6	8.3	13	18

bulk-31	6.01	2.5	5.5	7.9	12
bulk-32	5.92	3.5	8.4	13.7	19.2
bulk-35	6.01	6	11.4	18	23.6
bulk-36	5.96	3.9	10.1	16.6	24.2
bulk-37	6.01	2.6	5.5	9.2	16.8
bulk-38	5.9	0.5	4.6	10	16.8
bulk-4	6.05	6.4	9.7	16	19
bulk-40	6.1	6.7	12.2	19.1	25
bulk-42	6.13	5.5	10.1	15.2	20.4
bulk-5	5.96	1.6	4.8	9.8	14.5
bulk-6	6.13	6	9	13.5	16.6
bulk-7	6.17	3.5	6.8	13	16.7
bulk-8	6.43	5.2	8.4	12.6	17
bulk-9	6.05	5.4	8.4	13.1	17.6
c-1	5.71	6.4	8.7	14.3	20
c-10	5.88	1.9	6.6	12	18.6
c-11	5.84	2.8	6.1	11.1	15.7
c-2	5.59	4.4	7.3	12	17.5
c-3	5.67	5	11	17.1	24.3
c-4	5.63	3.5	5	9	12.8
c-5	5.71	5.5	8.8	13.5	17.6
c-6	5.84	5.5	10.3	16.5	22.3
c-7	5.71	NA	22.8	28.8	34.3
c-8	5.84	11.5	18	24.1	30.5
c-9	5.63	6	10.6	16.1	22.1
d-1	5.96	6.2	11.5	16.6	22
d-10	6.09	1.9	6.4	11.8	18
d-12	6.05	2.1	6	11.6	17.8
d-13	6.13	1.6	6	11	17.4
d-14	6.13	3.2	7.2	13.1	20.3
d-15	5.84	2.6	7.8	13	NA
d-16	6.01	4.6	9.8	15.6	22
d-17	5.75	2.1	6	10.5	14.6
d-18	5.84	8	13.4	19.5	26.5
d-2	5.84	NA	20.5	27.8	34.5
d-20	6.05	3.6	9.7	16.6	24
d-3	5.88	7	12.1	18	23.5
d-4	5.84	1.7	5.3	12.3	14.2
d-6	5.75	3.5	9.4	13.2	19.3
d-7	5.84	9.5	15	18.2	22
d-8	5.75	3.1	6.8	10.1	14.7

d-9	6.09	7.5	10.1	14.8	18
e-1	5.84	1.7	2	8.5	13.4
e-10	6.26	3.5	7.8	13.1	19
e-11	5.88	3.5	8.6	15.5	22.3
e-12	5.92	4.3	9.5	16.6	24.1
e-13	5.63	5	10.5	18	25
e-14	5.96	2.7	7.3	13.5	19.9
e-15	5.8	2.3	7	11.6	17.2
e-16	6.3	5.6	10	15.8	21.6
e-17	6.01	9.1	14.3	19.2	23.5
e-18	5.84	1.5	4.5	8.4	12.5
e-19	5.92	5.4	9.5	14.5	18.6
e-3	5.96	5	11.9	12.2	15.6
e-4	5.92	1.5	4.3	8.8	14.7
e-6	6.05	3.3	6.4	11.5	14.1
e-8	6.01	4.5	8	13.2	21.2
e-9	5.92	2.5	6	13.8	21.5
f-1	5.88	7.5	11	16.2	22.2
f-2	5.42	6.2	8.3	12.2	16.5
f-3	5.8	3.3	6.8	11.2	16.2
f-4	5.59	1.1	4	8.2	12.8
f-6	5.88	2.6	4.1	9.1	11.5
f-7	5.63	1.2	5.1	10	16
f-8	5.84	0.3	2.6	6.8	12.2
g-1	5.54	8.5	13.5	19.6	25.2
g-10	5.75	3.6	7.3	12.6	18.6
g-2	5.71	8.5	13.4	18.5	22.6
g-3	5.75	5.3	8.8	13.4	17.4
g-4	5.84	8	11.6	15.2	21
g-5	5.67	3.2	6.2	9.6	15.5
g-6	5.71	3.5	7.3	12.7	18.6
g-7	6.01	4.6	8	11.4	16
g-8	6.05	3.5	6.9	11.5	17
g-9	5.84	10.1	15.3	21.3	30.5
h-10	5.96	4.6	8	13.1	17
h-13	6.09	3.5	8.5	13.4	18.3
h-2	5.96	4	8	12.3	16
h-4	6.01	3	5.6	8.6	12.5
h-5	5.84	2.4	5.3	10	15.5
h-6	6.13	3	7	11.5	16.2
h-7	6.09	7	11.5	16	20.3

h-8	6.43	8.4	13.5	21	27.1
h-9	5.92	1.8	5.3	10.8	17.1
i-1	5.88	4.5	7.2	12.2	17
i-10	5.96	5	8.4	12.6	18.3
i-12	5.75	3.1	6.4	9.6	12.2
i-13	5.88	3.6	7.3	11.6	16.6
i-2	6.01	6	10.5	15.5	20.6
i-3	5.84	7.5	10.1	14.5	19.5
i-4	5.88	5.2	14.5	20.3	26
i-7	5.8	3.5	6.9	12.1	14.5
i-9	5.71	4.3	8.9	13.1	17.6
j-1	6.09	2.1	3.8	9.4	12
j-10	6.13	3.5	8.5	13.5	19.4
j-11	5.88	0.9	5	11	NA
j-2	6.38	7.4	10.5	14.8	18.3
j-3	6.3	2	6	10.5	14.6
j-4	5.88	3.5	8.2	11.7	16.7
j-5	5.96	4.7	8.4	12.6	17.3
j-7	5.92	1.5	5.5	9	11.7
j-8	6.34	2.4	5.6	10.5	15.6
j-9	6.09	3.5	7.8	13	18.5
k-10	5.74	1.5	6	11.5	18.2
k-11	5.5	2.5	6.5	11.8	17.9
k-14	5.94	2.1	5.6	9.8	14.1
k-15	5.77	1	4.5	9	13.6
k-16	5.82	1.1	4	8.1	12.5
k-17	5.93	5.6	10.4	15.9	21.5
k-3	5.88	1.4	3	8.7	13.5
k-5	5.74	4.8	10.5	18	30.1
k-6	5.9	2.1	6.1	11.6	18.1
k-7	5.66	7.7	13.3	19.5	25.1
k-8	6.01	4.2	9.6	16.6	23.6
k-9	5.77	3.4	6	9.8	11.5
l-10	5.75	7.4	11	16	21.5
l-11	5.8	10.6	15.7	21.2	25.5
l-12	5.71	4.6	9	14	19.5
l-13	6.05	7.4	7.9	16.9	22.4
l-15	6.22	0.3	4.5	9.9	15.5
l-16	5.92	13.2	18.5	23.6	27.2
l-3	5.88	4.5	8.9	14.1	18.6
l-6	5.96	6.9	11.5	16.5	21.1

l-7	6.01	8	14.3	22.9	29
l-8	6.17	3.4	6.7	11.6	14.5
l-9	5.96	4.2	8.9	15	21

Table S7. Measures of genome size from two individuals from each of the 10 populations used in FISH to sequence correlation (Fig. 2).

Accession	Subspecies	Ind1	Ind2	Altitude(m)
RIMME0021	mexicana	6.12	6.01	2094
RIMME0026	mexicana	6.21	6.11	2214
RIMME0028	mexicana	5.65	5.53	1916
RIMME0029	mexicana	5.46	5.58	1547
RIMME0030	mexicana	5.98	5.8	2458
RIMME0031	mexicana	6.26	6.39	2609
RIMME0032	mexicana	5.63	5.73	2016
RIMME0033	mexicana	5.43	5.44	1657
RIMME0034	mexicana	5.56	5.38	2173
RIMME0035	mexicana	6.18	6.46	2237
RIMPA0071	parviglumis	6.1	6.1	985
RIMPA0086	parviglumis	6.01	5.9	982
RIMPA0087	parviglumis	5.59	5.63	590
RIMPA0096	parviglumis	6.12	6.03	1528
RIMPA0135	parviglumis	6.25	6.34	880
RIMPA0142	parviglumis	6.33	6.13	1103

Table S8. Genome size estimates and altitudinal information for mexicana populations

Population	Accession	DNA (pg/2C)	Altitude (m)
Tz	13	5.59	1665
Tz	2	5.46	1665
Tz	3	5.33	1665
Tz	4	5.71	1665
Tz	19	5.71	1665
Tz	23	5.54	1665
Tz	9	5.59	1665
Tz	10	5.54	1665
Tz	12	5.71	1665
Fp	1	6.51	2507
Fp	2	6.26	2507
Fp	3	6.3	2507
Fp	4	6.43	2507
Fp	5	6.3	2507
Fp	E	6.05	2507
Fp	8	6.68	2507
Fp	9	6.13	2507
Fp	12	6.55	2507
Mt	1	6.47	2353
Mt	4	6.34	2353
Mt	5	6.51	2353
Mt	7	6.26	2353
Mt	8	6.38	2353
Mt	9	6.43	2353
Mt	10	6.51	2353
Mt	11	6.43	2353
Mt	12	6.43	2353
Da	1	6.72	2408
Da	2	6.64	2408
Da	3	6.38	2408
Da	5	6.51	2408
Da	6	6.47	2408
Da	8	6.17	2408
Da	9	6.55	2408
Da	10	6.34	2408
Da	11	6.38	2408
Mc	1	6.34	2491.5

Mc	2	6.05	2491.5
Mc	3	6.26	2491.5
Mc	5	6.26	2491.5
Mc	6	6.38	2491.5
Mc	9	6.17	2491.5
Mc	10	6.17	2491.5
Mc	11	6.26	2491.5
Mc	12	6.64	2491.5
M	2	6.26	1881
M	4	5.96	1881
M	5	6.05	1881
M	6	6.13	1881
M	7	6.17	1881
M	8	6.09	1881
M	9	6.17	1881
M	10	6.51	1881
M	11	6.3	1881
Tc	1	6.09	2776
Tc	2	6.13	2776
Tc	3	6.22	2776
Tc	4	5.92	2776
Tc	5	6.34	2776
Tc	E/1	5.92	2776
Tc	7	6.13	2776
Tc	8	6.26	2776
Tc	12	6.3	2776
Tx	2	6.64	2253
Tx	3	6.47	2253
Tx	4	6.38	2253
Tx	7	6.38	2253
Tx	8	6.38	2253
Tx	9	6.3	2253
Tx	10	6.3	2253
Tx	11	6.72	2253
Tx	12	6.72	2253
Cl	1	5.96	2698
Cl	2	6.13	2698
Cl	4	6.05	2698
Cl	5	6.05	2698
Cl	7	6.09	2698
Cl	8	6.05	2698

Cl	9	6.05	2698
Cl	11	6.26	2698
Cl	12	5.92	2698
Cu	1	6.38	2792
Cu	2	6.55	2792
Cu	4	6.3	2792
Cu	5	6.3	2792
Cu	7	6.38	2792
Cu	9	6.51	2792
Cu	10	6.09	2792
Cu	11	6.09	2792
Cu	12	6.09	2792
Am	A	5.42	1591
Am	B	5.46	1591
Am	C	5.5	1591
Am	D	5.33	1591
Am	E	5.63	1591

Table S9. Repeated measures of genome size from maize inbreds lines

SampleID	Measure1	Measure2
Ki3	5.96	6.01
Ky21	5.50	5.63
NC358	5.92	5.88
B73	5.42	5.46
CML247	6.05	6.05
CML52	6.22	6.13
P39	5.50	5.50
H95	5.84	5.80
A188	5.71	5.63
K55	5.80	5.84
K64	5.71	5.80
NC33	5.75	5.80
Pa762	5.88	5.80
K4	5.54	5.50
M14	5.84	5.80
B64	5.92	5.71
T8	5.59	5.59
B84	5.63	5.50
IDS28	5.46	5.50
CH9	5.67	5.71
CML5	6.13	6.13
CML10	5.96	6.05
CML220	6.13	6.30
CML331	6.05	6.05
CML332	6.05	6.09
NC310	5.67	5.80
NC318	5.84	5.88
NC336	6.26	6.13
NC344	5.67	5.67
SA24	5.84	5.71
SA55	6.05	5.96
Mo46	5.96	5.88
CML264	6.01	5.96
DE1	5.80	5.63
RIMMA0806	5.71	5.67