RESEARCH ARTICLE

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

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

1 Department of Plant Sciences, University of California, Davis, Davis, California, United States of America, **2** Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany, **3** Division of Biological Sciences, University of Missouri, Columbia, Missouri, United States of America, **4** Center for Population Biology, University of California, Davis, Davis, California, United States of America, **5** Department of Evolution and Ecology, University of California, Davis, Davis, California, United States of America, **6** Department of Anthropology, University of California, Davis, Davis, California, United States of America, **7** Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America, **8** Genome Center, University of California, Davis, Davis, California, United States of America

* pbilinski@westshore.edu (PB); rossibarra@ucdavis.edu (JRI)

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 population structure alone cannot explain the observed variation, implying that clinal patterns of genome size are maintained by natural selection. Our modeling additionally provides evidence of selection on individual heterochromatic knob repeats, likely due to their large individual contribution to genome size. To better understand the phenotypes driving selection on genome size, we conducted a growth chamber experiment using a population of highland teosinte exhibiting extensive variation in genome size. We find weak support for a positive correlation between genome size and cell size, but stronger support for a negative correlation between genome size and the rate of cell production. Reanalyzing published data of cell counts in maize shoot apical meristems, we then identify 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 intraspecific variation in repetitive sequence to important differences in adaptive phenotypes.

Author summary

Genome size in plants can vary by orders of magnitude, but this variation has long been considered to be of little functional consequence. Studying three independent adaptations

to high altitude in *Zea mays*, we find that genome size experiences parallel pressures from natural selection, causing a reduction in genome size with increasing altitude. Though reductions in overall repetitive content are responsible for the genome size change, we find that only those individual loci contributing most to the variation in genome size are individually targeted by selection. To identify the phenotype influenced by genome size, we study how variation in genome size within a single wild population impacts leaf growth and cell division. We find that genome size variation correlates negatively with the rate of cell division, suggesting that individuals with larger genomes require longer to complete a mitotic cycle. Finally, we reanalyze data from maize inbreds to show that faster cell division is correlated with earlier flowering, connecting observed variation in genome size to an important adaptive phenotype.

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 and that genic content is relatively constant [3, 4]. 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 [5].

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 genetic drift permits maladaptive expansion [6] or contraction [7] of genomes across species. Other models propose that variation may be due to differences in the rates of insertions and deletions [8] or a consequence of changes in modes of reproduction [9, 10]. While each of these models find limited empirical support [11, 12], counterexamples are common [9, 10, 13, 14]. 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 [15], plant cell size [16], seed size [17], body size [18], and growth rate [19]. 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 [20]. In most of these models, however, the mechanistic link between genome size and phenotype remains unclear [21].

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 [22] or argued to be too small to have much evolutionary relevance [23]. Nonetheless, intraspecific variation in genome size has been documented in hundreds of plant species [23], including multiple examples of large-scale variation [24–26]. Correlations between intraspecific variation in genome size and other phenotypes or environmental factors have also been observed [24, 25, 27], suggesting the possibility that some of the observed variation may be adaptive.

Here we present an analysis of intraspecific genome size variation in the model system maize (*Zea mays* ssp. *mays*) and its wild relative highland teosinte (*Zea mays* ssp. *mexicana*). Genome size in *Zea* varies dramatically both within [28] and between [29] subsepecies, and previous work has also found substantial intraspecific variation in transposable element (TE) abundance [30, 31], the number of auxiliary B chromosomes [32], and the number and location of heterochromatic knobs [32]. Several authors have observed negative correlations between genome size and altitude [25, 28] and some repeats show similar clinal variation [33]. It remains unclear, however, whether these patterns can be explained by natural selection or how genome size might impact plant fitness.

We take advantage of parallel altitudinal clines in maize landraces from Mesoamerica and South America to investigate the evolutionary processes and sequence differences underlying genome size variation. Leveraging the intraspecific genome size variation in *Zea* taxa, we model genome size as a quantitative trait, using flow cytometry and genotyping to show that natural selection has reduced genome size in high elevation populations. In a similar analysis of repeat content from low coverage shotgun sequencing, we also identify evidence of selection directly on knob variants. 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 weaker support for a 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 77 diverse maize landraces from across a range of altitudes in Meso- and South America (S2 Table). Flow cytometry of these samples revealed a negative correlation with altitude on both continents (Fig 1A, r = -0.51 and -0.8, respectively, p-value <0.001). 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 (Fig 2 and S4 Fig; see Methods for details). We observed substantial variation among landraces in the abundance of individual transposable element families (S5 Fig), and both transposable elements as a whole and heterochromatic knobs showed clear decreases in abundance with increasing altitude in Meso- and South America (TE r = -0.57, -0.72; 180bp knob r = -0.48, -0.83; TR1 knob r = -0.66, -0.81; p-value <0.001), mirroring the pattern seen for overall genome size (Fig 1). In contrast, we found only a weak positive correlation between B chromosome abundance and altitude (p-value >0.05) (S6 Fig).

We next sought to evaluate whether the observed clines in genome size and repeat abundance simply reflected underlying genetic differences due to population structure or could be better explained by natural selection. We adopted an approach similar to Berg and Coop [34], modeling genome size as a quantitative trait that is a linear function of relatedness and altitude (see Methods, Eq 1). Across maize landraces, we rejected a neutral model in which genome size is unrelated to altitude, estimating a decrease of 108Kb and 154Kb in mean genome size per meter gain of altitude in Meso- and South America, respectively (S8 Table). We then evaluated whether selection has acted on individual repeats, treating abundance of each repeat class as a quantitative trait in a comparable model that includes genome size as a covariate
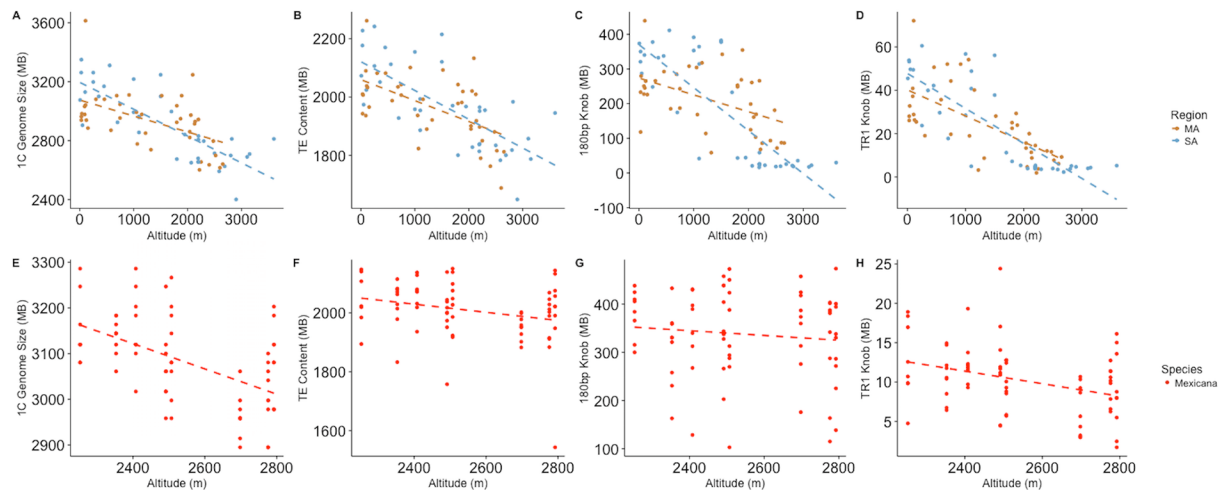
Fig 1. Genome size and repeat content by altitude in *Zea* taxa. (A-D) Maize landraces from Mesoamerica (MA) or South America (SA). (E-H) Highland teosinte *Z. mays* ssp. *mexicana*. Only teosinte populations above 2000m that do not show admixture (see text) are included. (A, E) total genome size, (B,F) total transposable element content, (C,G) 180bp knob repeat content, (D,H) TR1 knob repeat content. Dashed lines represent the best fit linear regression.

https://doi.org/10.1371/journal.pgen.1007162.g001

(Methods, Eq 2). In both Meso- and South America, TR1 knobs showed evidence of selection, while 180bp knobs also showed evidence of selection in South American landrace germplasm (S8 Table). Finally, our models for total transposable element content were not significant in either continent, and the number of individual TE families showing significant correlations with altitude was no greater than expected by chance (46/1156, binomial test p-value >0.05).

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* (herafter, *mexicana*), diverged from *parviglumis* ≈60,000 years ago [35] and has adapted to the higher altitudes of the Mexican central plateau [36]. 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) [37, 38]. Though both subspecies exhibit considerable variation, *parviglumis* samples have larger genomes than *mexicana* (S7 Fig; one tailed t-test p-value<0.05) but do not differ from lowland maize in Mexico, 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 (S4 Table). Genome size variation across these populations revealed no clear relationship with altitude (S8 Fig), but genotyping data [39] revealed consistent evidence of genetic separation (S9 Fig) and higher inbreeding coefficients (two-sided t-test p-value <0.001) in the three lowest altitude populations (see Methods). These three populations are also phenotypically distinct and relatively isolated from the rest of the distribution (A. O'Brien, pers. communication). We thus excluded these three populations, applying our linear model of altitude and relatedness to 70 individuals from the remaining 8 populations. After doing so, we find a negative relationship between genome size and altitude in *mexicana* (Fig 1E, p-value <0.001) of similar magnitude to that seen in maize (loss of 270Kb/m), suggesting parallel patterns of selection across *Zea*. In agreement with our results in maize, TR1 knob repeats showed evidence of selection after controlling for their contribution to genome size (S8 Table), though 180bp knob repeats did not. We found no evidence for selection on
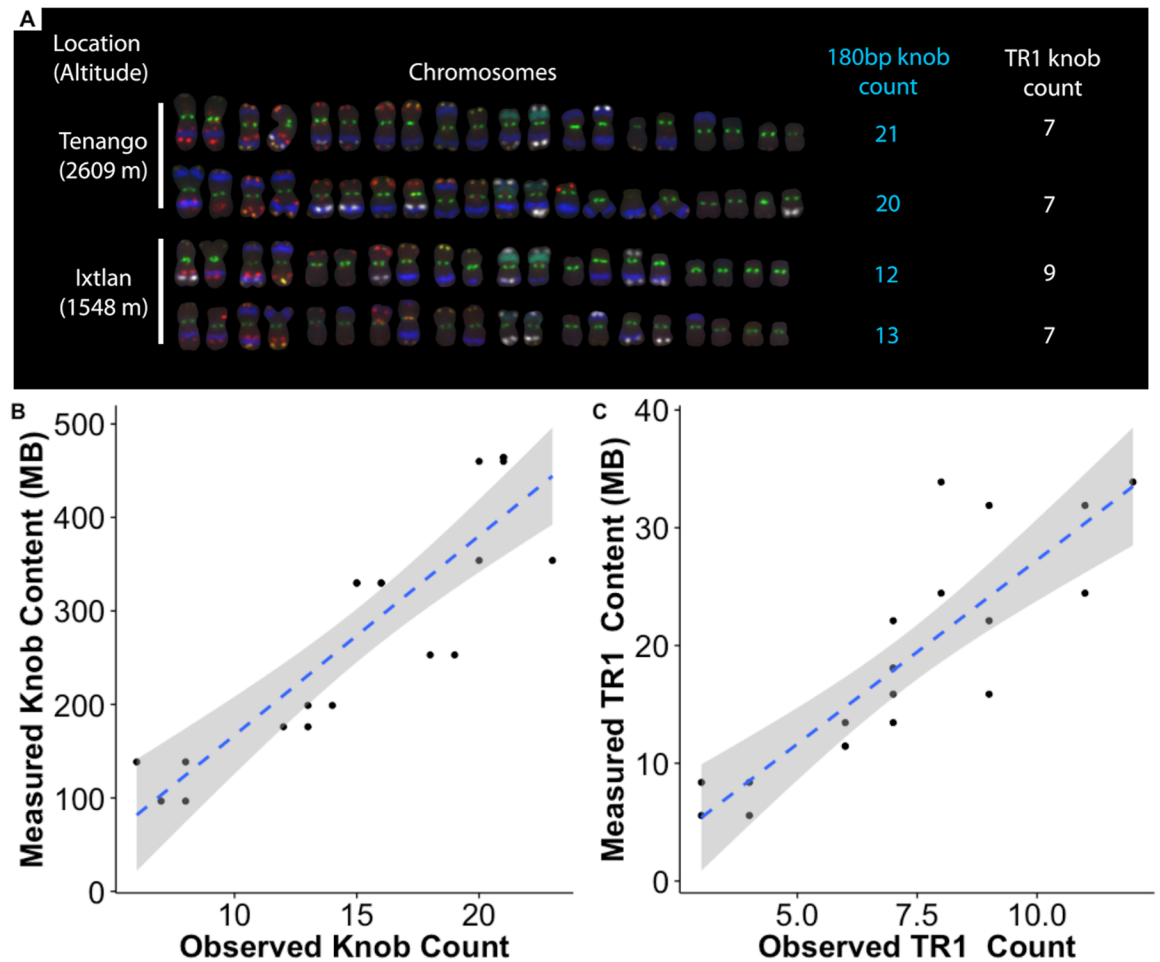
Fig 2. Knob content in highland teosinte estimated using FISH and low-coverage sequencing. (A) FISH from four *Z. mays* ssp. *mexicana* individuals, sampled from the highest and lowest altitude populations. Counts of cytological 180bp (blue) and TR1 (white) knobs are shown to the right of each individual. 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 [40]. (B) Plot of the population-level correlation between 180bp knob counts and sequence abundance for 20 *mexicana* individuals. 180bp knob r = 0.88, TR1 knob r = 0.86.

TE abundance after controlling for genome size, and none of the sequence from *mexicana* mapped to our B-repeat library.

To test whether genome size might be related to flowering time through its potential effect on the rate of cell production, we performed a growth chamber experiment to measure leaf elongation rate, cell size, and genome size using 201 *mexicana* individuals from 51 maternal families sampled from a single natural population (see Methods). Individual plants varied by as much as 1.13Gb in 2C genome size, with observed leaf elongation rate (LER) varying from 1 to 8 cm/day (mean 4.56cm/day; S9 Table). Without correcting for any relatedness, we see that genome size and leaf elongation rate have a negative correlation of -0.134 (p-value = 0.0576), while genome size and cell size are not correlated (p-value = 0.4452). To incorporate the family structure in our sample and directly connect genome size with cell division rate in a parametric fashion, 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 suggest a weak but positive relationship between genome size and cell size (γ_{GS}; Fig 3A) and a negative
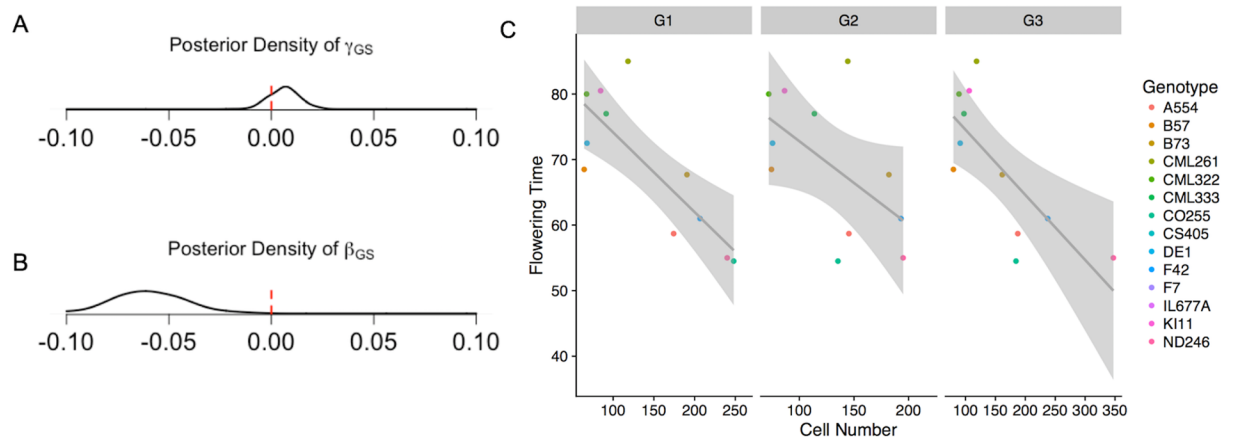
Fig 3. (A,B) Posterior densities of effects of genome size on cell size and cell production rate (γ_{GS} and β_{GS}, respectively) from a model with prior mean stomatal cell size of 30 microns and leaf elongation rate of 4cm/day. (C) Linear regression of flowering time and SAM cell number across inbred maize accessions. Measurements for cell number are shown for each of three growth phases (G1, G2, G3). Data from Leiboff *et al.* [41].

relationship between genome size and cell production rate (β_{GS}; Fig 3B). We found that our inferences were sensitive to prior specifications for leaf elongation rate and cell size (S3 Fig), but prior means $\geq$ 4cm/day for leaf elongation rate combined with prior means $\leq$ 0.003cm for CS, returned reliably negative relationships between genome size and cell production rate (see Methods).

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 Leiboff *et al.* sampled SAM at equivalent growth stages, we interpreted variation in cell number as representative of differences in cell production rate among lines. We re-analyzed these data to investigate whether the cell number reported in each SAM was correlated with flowering time (Fig 3C). After estimating genetic values for each inbred line used and correcting for population structure and the effects of two candidate genes (see Methods), we find a negative correlation between flowering time and cell production across all three developmental stages sampled (slopes of -0.11, -0.08, and -0.08 and p-values <0.01,<0.001, and 0.170, respectively).

Discussion

Genome size and repeat abundance

We report evidence of a negative correlation between genome size and altitude across clines in Meso- and South America in both maize and its wild relative highland teosinte (Fig 1). Maize was domesticated from *parviglumis*, suggesting that large genome size was likely ancestral, and we observed no difference in genome size between lowland Meso-American maize landraces and *parviglumis*. The subsequent colonization of highland environments occurred independently in Mesoamerica and South America [42], and while the populations share a number of adaptive phenotypes, they exhibit little evidence of convergent evolution at individual loci [43]. The teosinte subspecies *mexicana* is also found in the highlands of Mesoamerica [36], likely after its split from the lowland teosinte *parviglumis* $\approx$60,000 years ago [35]. Previous investigations of genome size have also identified negative altitudinal clines in maize and teosinte [25, 28] (but see Rayburn *et al.* [44] for a positive cline in the U.S. Southwest), suggesting that this observation is general and not an artifact of our sampling.

Although we find altitudinal trends in genome size across all three clines, our initial evaluation of genome size in highland teosinte found no significant correlation with altitude, due primarily to the small genomes observed in the three lowest altitude populations (S8 Fig). We excluded these three *mexicana* populations because they showed higher levels of inbreeding than other *mexicana* populations as well as evidence of shared ancestry with *parviglumis* (S9 Fig). We speculate that the relationship between genome size and altitude may be more complex for low altitude *mexicana* due to the confounding effects of admixture impacting both adaptation and repeat evolution. These populations are nonetheless interesting and worthy of future investigation, as their genome size is smaller than either *parviglumis* or high altitude *mexicana* but their knob content does not differ from other *mexicana* populations, suggesting perhaps that inbreeding or admixture may have affected transposable element or other repeat abundance.

Our results suggest the best explanation for the observed clines in intraspecifc genome size variation is natural selection. Several authors have identified ecological correlates of variation in plant genome size and argued for adaptive explanations of such clines [25, 28, 45], but relatively few have corrected for relatedness among individuals or populations [27]. We employ a modeling approach that considers genome size as a quantitative trait and uses SNP data to generate a null expectation of variation among populations, allowing us to rule out stochastic processes and instead pointing to the action of selection in patterning clinal differences in genome size. Alternative explanations for our observations, including mutational biases and TE expansion, are unlikely. 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 [46]. But because UV damage causes small DNA deletions, it is unlikely to generate the gigabase-scale difference we see across altitudinal clines in the short time since maize arrived in the highlands [47]. And while expansion or replication of TEs in lowland populations could lead to increased rates of insertion and larger genome size, our analysis of reads mapping to individual TE families finds no evidence that this has occurred in a widespread manner, and genome size estimates from the direct wild ancestor of domesticated maize (the lowland teosinte *parviglumis*) suggest that smaller highland genomes are the derived state.

Having concluded that natural selection is the most plausible explanation for decreasing genome size at higher altitudes, 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. We find no evidence of selection on B repeats, consistent with the relatively mixed signals found in previous literature [48]. We also find little evidence of selection on TEs after controlling for genome size. Because individual TEs are relatively small, however, models of polygenic adaptation lead us to expect that such loci are unlikely to show a strong signal [49]. Nonetheless, TEs show the strongest overall correlation with genome size, suggesting that frequent small deletions of individual elements are likely a major contributor to genome size change across populations. In contrast to TEs, in both maize and teosinte the 350bp TR1 knob repeat shows greater differentiation in abundance across altitude than can be explained by population structure alone, even after accounting for changes in total genome size. The 180bp knob shows a similar strong decline in abundance in maize landraces, but is only statistically significant in the analysis of landraces in South America. Selection on genome size might be expected to act especially strongly on knobs, as each locus may contain many megabases of repeats and knob abundance is a large contributor to intraspecific genome size variation [50, 51]. These results are surprising, however, given the selfish nature of knobs and their ability to distort segregation ratios in female meiosis in the presence of a driving element known as abnormal chromosome 10 (Ab10) [52]. While our genotyping data do not include markers diagnostic of Ab10, previous analyses show that selection along altitudinal gradients has been sufficient to decrease

the frequency of at least one allele of the drive locus itself [53]. It is not entirely clear why we see more evidence of selection on the TR1 knob variant, which contributes nearly an order of magnitude fewer base pairs to the genome. The TR1 variant generally shows weaker drive, but has been shown to compete successfully against the 180bp variant [54]. It is thus possible that the weaker drive of TR1 makes it more susceptible to selection on overall genome size, and that the subsequent decrease in TR1 abundance may increase drive of the 180bp knob variant, potentially ameliorating the effects of selection against 180bp knobs at higher altitude. Finally, while we see decreasing abundance of both knob variants with increasing altitude, we note that knobs alone are not driving the overall signal: rerunning our model for genome size after removing base pairs attributable to both knob repeats still finds evidence of selection on genome size in all three clines (Mesoamerica p-value = 0.029; South America p-value = 0.04; *mexicana* p-value = 0.02).

Genome size and development rate

Several authors have hypothesized that genome size could be related to rates of cell production and thus developmental timing [52, 55]. 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 that exhibited wide variation in genome size. Our approach to characterizing the effect of genome size on the rate of cell production is consistent with scaling laws proposed in a recent study of the relationships between genome size, cell size, and cell production rate [56] (see Methods). We found only weak evidence for a positive correlation between genome size and cell size, a result that contrasts with the findings of many authors who have reported more definitive positive correlations between genome size and cell size across species [57, 58]. One potential explanation for this result may be found in recent work in *Drosophila* where larger repeat arrays were shown to lead to more compact heterochromatin despite the physical presence of more DNA [59]. 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 rate in our highland teosinte population. Though these results showed strong prior sensitivity, the sign of the relationship between genome size and cell production rate did not change for prior mean values of leaf elongation rate within the range of those published for maize (from 4.6 cm/day [60] to 12 cm/day [61]), all equal to or larger than the rates observed in our experiment. Tenaillon *et al.* [62] also find a negative correlation between the rate of leaf elongation and genome size among inbred lines, albeit one that does not survive statistical correction for population structure.

We hypothesize that selection on flowering time is the driving force behind our observed differences in genome size. Common garden experiments show that highland populations of both maize and teosinte flower earlier than their lowland counterparts [63, 64], and an artificial selection experiment in maize found the traits to be genetically correlated ($r \approx 0.14$; data from Rayburn *et al.* [65] assuming heritabilities of $h^2 = 0.8$ for flowering time and $h^2 = 1$ for genome size). Larger genomes require more time to replicate [45], and slower rates of cell production in turn may lead to slower overall development or longer generation times [55], though our data cannot tease apart an S-phase effect from a general cell cycle effect. 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 [66]. However, signals for flowering derive from plant leaves [67, 68], and slower cell production will result in a longer time until

full maturity of all the organs necessary for the plant to flower. Consistent with our hypothesis, reanalysis of published data from SAM of maize inbred lines suggests that plants with more cells in their SAM at a given developmental stage (and thus faster rates of cell production) also exhibit earlier flowering [41]. Further evidence supporting this idea comes from Jain *et al.* [69], who observe the predicted negative correlation between genome size and flowering time among a diverse panel of maize inbreds (although the relationship is not significant after correcting for kinship). Finally, though additional environmental factors have been hypothesized to elicit adaptive changes in genome size *e.g.* [27, 70], we are unaware of alternative selective explanations for the genome size correlations seen in our Mesoamerican or South American altitudinal clines, and we suggest that future efforts should focus on experimental validation of the mechanistic connection between genome size and both cell production and flowering time suggested by our results.

Conclusion

The causes of genome size variation have been debated for decades, but these discussions have often ignored intraspecific variation. Our results suggest that differences in optimal flowering times across altitudes are likely indirectly effecting clines in genome size due to a mechanistic relationship between genome size and cell production and developmental rate. We also show that selection on genome size has driven changes in repeat abundance across the genome, including significant reductions in individual repeats such as knobs that contribute substantially to intraspecific variation in genome size. 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 similar characteristics across plants, and genome size is likely to impact cell production due to the constraints on 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, and that the phenotypic effects of genome size may have consequences for the evolution of individual repeats.

Materials and methods

Unless otherwise specified, raw data and code for all analyses are available on the project Github at https://github.com/paulbilinski/GenomeSizeAnalysis and S1 Table shows the general relationship among samples and analyses; additional details are included below.

Genome size

We quantified genome size in 77 maize landraces (S2 Table; [43]) and two samples from previously collected populations of *parviglumis* (n = 6) and *mexicana* (n = 10) (S3 Table; [38]). For our growth chamber experiment, we sampled 201 total seeds from 51 maternal plants collected from 11 populations of *mexicana* (S4 and S5 Tables). To assess the error associated with flow cytometry measures of genome size, we used 2 technical replicates of each of 35 maize inbred lines (S6 Table). We germinated seeds and grew plants in standard greenhouse conditions and sent leaf tissue from each individual to Plant Cytometry Services (JG Schijndel, NL) for genome size analysis. *Vinca major*, with a genome size of 2.1pg/1C, was used as an internal standard for flow cytometric measures, and standard and unknowns were co-prepared and co-stained. Replicated maize lines showed highly repeatable estimates (corr = 0.92), with an average difference of 0.0346pg/1C between estimates.

Genotyping

We used genotyping-by-sequencing (GBS) [71] data from Takuno *et al.* [43] for maize accessions along altitudinal clines in Mesoamerica and South America. For the 11 *mexicana* populations used in our linear model, we used GBS SNP data from O'Brien *et al.* [39]. All samples were filtered with TASSEL (V5.2.37) [72] to remove sites with >40% missing data and individuals with >90% missing data, resulting in 170 total individuals with genotyping data for 223,657 sites. We elected to use this per-site cut off as it did not qualitatively change the site frequency spectrum (S1 Fig).

Kinship and admixture

Kinship matrix calculation was performed using centered identity-by-state (IBS) as implemented in the software TASSEL [73]. We elected to use random imputation in our kinship calculations, as mean imputation biases the estimate of inbreeding within individual [73]. However, we tested both mean and KNN [74] imputation, and our results were robust to both methods. Inbreeding statistics for individual *mexicana* plants were calculated from the diagonal of the randomly imputed kinship matrix.

Admixture analyses were performed using Admixture v1.23 [75]. For admixture analyses we also included additional GBS data from diverse maize inbred lines [76], landraces and teosintes [77] (S2 and S4 Tables), for a total of 611 individuals before filtering. We filtered individuals and sites as above, but additionally removed one individual (the sample with lowest sequencing depth) of each pair with an IBS distance closer than 0.07. A Hardy-Weinberg filter was then applied using only outbred genotypes with a read depth between 9-300 using a chi-squared goodness of fit test, p-value <0.05. We then thinned sites by linkage disequilibrium, removing lower coverage sites within physical distance less than 1000bp and sites with r^2 >0.8 and significant at p-value <0.05. Only sites with at least 12 high depth genotypes were tested. After filtering, 526 individuals and 18,716 sites remained.

Shotgun sequencing

We used whole genome shotgun sequencing to estimate repeat abundance in the same 77 maize landrace accessions and 93 *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, data available on Figshare at DOI 10.6084/m9.figshare.5117827). DNA was isolated from leaf tissue using the DNeasy plant extraction kit (Qiagen) according to the manufacturer's instructions. Samples were multiplexed 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 to a depth of <0.5X coverage per sample. 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 analysis of clinal variation.

Estimating repeat abundance

We gathered reference sequences for 180bp knob, TR1 knob, B chromosome, and rDNA repeats from NCBI. CentC repeats were taken from Bilinski *et al.* [78], and chloroplast DNA and mitochondrial DNA were taken from the maize reference genome (v2, www.maizesequence.org). B chromosomes repeats [79] were matched against the maize genome (v2, www.maizesequence.org) using BLAST, and any regions within the repeats that had alignments of greater than 30bp with 80% homology were masked. The remaining unmasked

regions with length greater than 70bp were used as a mapping reference for B-repeat abundance. For the transposable element database, we began with the TE database consensus sequences [80, 81]. Using BLAST, we masked shared regions and retained unique regions of 70bp or greater as our reference, repeating this process to additionally mask tandem repeats. We mapped sequence reads to our repeat library using bwa-mem [82] with parameters -B 2 -k 11 -a to store all hit locations with an identity threshold of approximately 80%. We filtered out plastid sequence and calculated Mb of repeat by multiplying our estimated abundance by genome size. The correlation between the abundance of each repeat and genome size were as follows: TE = 0.95; 180bp knob = 0.81; TR1 = 0.86. Previous simulations suggest that this estimate has good precision and accuracy in capturing relative differences across individuals [78].

Repeat abundance validation via FISH

We selected two individuals each from 10 previously collected populations of *mexicana* [37] for fluorescence *in situ* hybridization counts of knob content (FISH; S3 Table). FISH probe and procedures closely followed Albert *et al.* [40].

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 (Eq 1). We assume genome size has a narrow sense heritability $h^2 = 1$, as it is simply 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 represents an additive genetic component modeled as a random effect 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.

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

We implemented our linear model in EMMA [83] to test for selection on genome size. In a second model, we then include genome size (GS) as a fixed effect in order to test for correlations between specific repeat classes and altitude conditional on genome size (Eq 2). Controlling for genome size allows us to test whether we see evidence of selection on a repeat beyond its contribution to the total base pairs it contributes toward genome size. Because TEs make up 85% of the genome, for example, without such a correction any selection on genome size will appear to be selection on TEs.

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

Growth chamber experiment

We sampled 202 seeds from multiple maternal plants collected in a single high altitude (2408m) *mexicana* population at Tenango Del Aire, chosen because it exhibited the most variation in genome size in our altitudinal transect of *mexicana*. Germinated seedlings were

transferred to soil pots and into a growth chamber (23°C, 16h L/8h D). We measured leaf length daily for 3 days after the first visible emergence of the third leaf. 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). 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 rate (LER) as the product of cell size (CS) and the rate of cell production (CP):

$$LER = CS * CP. \tag{3}$$

The general approach is described as a model with two "mediators" in Mackinnon [84] and a full explanation can be seen in S1 Appendix. The multiplicative expression in Eq 3 is linearized by taking the natural logarithm on both sides of the equation, and model-fitting is performed on the log scale. We hypothesize that genome size affects LER only through its effects on CS and CP. The strategy for estimation of genome-size effects is illustrated by path diagrams shown in S2 Fig, where additional details are given. We adopt a computational Bayesian approach for parameter estimation, incorporating seedling and maternal random effects in models that make use of the hierarchical dataset structure (cells and days of growth within seedlings, seedlings within maternal parents). The signs and magnitudes of our estimated effects, and therefore our conclusions, are sensitive to different specifications of prior information. We identified previous averages for maize stomatal cell size and daily leaf elongation rate (CS = 0.003 cm, LER = 4.0-4.8cm/day or 2mm/hr) [60, 85–87], and incorporated these into informative priors for the random effects. Because our model shows prior sensitivity, we also identify prior means for which the sign of the relationship between genome size and cell production rate (β_{GS}), or cell size (γ_{GS}) changes (S3 Fig). We generated posterior samples of model parameters using JAGS, a general-purpose Gibbs sampler invoked from the R statistical language using the library rjags [88]. 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 rate and flowering time, we used flowering time and meristem cell number data for 14 maize inbred lines from Leiboff *et al.* [41]. Because meristems were sampled at an identical growth stage and time point, differences in cell number should reflect differences in the rate of cell production. We fitted a mixed linear model to estimate the best linear unbiased estimates (BLUEs) of the cell counts for each growth period separately:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon$$
$$\varepsilon \sim N(0, V_E) \tag{4}$$

In this model, Y_{ij} is the cell count value of the ith genotype evaluated in the jth replicate; μ, the overall mean; α_i, the fixed effect of the ith genotype; β_j, the random effect of the jth 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 fit mixed linear models to study the relationship of flowering time and cell counts by controlling for population structure and known trait-associated SNPs:

$$Y = \mu + \alpha G + \beta_{BAK1} + \beta_{SDA1} + g + \varepsilon$$

$$g \sim MVN(0, V_A \mathbf{K}) \tag{5}$$

$$\varepsilon \sim N(0, V_E)$$

Here Y is the flowering time (days to anthesis); μ, the overall mean; α, the fixed effect; 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 [89]. The genetic relatedness matrix was calculated using GEMMA [90] from publicly available GBS genotyping for these lines (AllZeaGBSv2.7 at www.panzea.org, [72]). 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 [91].

Supporting information

S1 Fig. Plot of reads mapping to B chromosome specific repeats in maize landraces. Key indicates percent of site coverage, ranging from unfiltered (0.0) to a requirement of full data presence across individuals (1.0). The spectrum begins to shift after a 60% site coverage requirement.
(TIFF)

S2 Fig. Path models for estimation of genome size effects. Arrows indicate predictor-outcome relationships and are annotated with model coefficients (slopes) from equations 9-11. (A) Genome Size (GS) predicts Leaf Elongation Rate (LER) through the mediators Cell Size (CS) and Cell Production rate (CP). CP, shown in grey, is not directly observed. The unit coefficients connecting log LER with log CP and log CS reflect the assumption LER = CS * CP (Eq 3). (B) Marginal model for the effect of GS on LER (equation 10).
(TIFF)

S3 Fig. Effect of LER and stomatal cell size priors on posterior density of the cell production coefficient β_{GS}.
(TIFF)

S4 Fig. Knob content in highland teosinte estimated using FISH and low-coverage sequencing, showing all sampled individuals referenced in Fig 2. Counts of cytological 180bp knobs (blue) and TR1 knobs (white) 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 [40].
(TIFF)

S5 Fig. Variation in (A) RNA and (B) DNA 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.
(TIFF)

S6 Fig. Plot of reads mapping to B chromosome specific repeats in maize landraces.
(TIFF)

S7 Fig. Genome size for two individuals per teosinte population sampled in [37, 38]. Points indicating individual genome size estimates are jittered around the center. With a parametric t-test of unequal variance, the one sided p-value is 0.03. Using a non-parametric Wilcoxon test, the one tailed p-value is 0.06.
(TIFF)

S8 Fig. Genome size by altitude of *mexicana*. All samples, including low altitude populations, are shown.
(TIFF)

S9 Fig. Population structure of maize and *mexicana* populations. (A) Admixture plots for K = 6, with altitude of accessions shown above. *Mexicana* populations and maize landraces are those used in this study. We include *parviglumis* [77] and maize inbreds [76]. (B) and (C) Multi-dimensional scaling analyses showing clustering of whole genome SNPs and those used to generate the admixture plot. Points are color coded based on the label underneath the admixture plot. (D) 5-fold cross validated error as estimated by Admixture, indicating the best estimate of number of populations, K.
(TIFF)

S1 Appendix. Path model.
(PDF)

S1 Table. Description of data sets used in each analysis.
(PDF)

S2 Table. Geographic information for maize landrace accessions.
(PDF)

S3 Table. Measures of genome size from two individuals from each of the 10 populations used in FISH to sequence correlation (Fig 2).
(PDF)

S4 Table. Geographic information for teosinte populations used in selection studies.
(PDF)

S5 Table. Genome size estimates and altitudinal information for mexicana populations.
(PDF)

S6 Table. Repeated measures of genome size from maize inbreds lines.
(PDF)

S7 Table. *Mexicana* population IDs and number of individuals used for FISH analyses.
(PDF)

S8 Table. Altitudinal coefficients from selection models using maize landraces and highland teosinte. 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.
(PDF)

S9 Table. 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.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Paul Bilinski, Jeffrey Ross-Ibarra.

Data curation: Paul Bilinski, Patrice S. Albert, Anne Lorant, Juvenal Quezada.

Formal analysis: Paul Bilinski, Patrice S. Albert, Jeremy J. Berg, James A. Birchler, Mark N. Grote, Anne Lorant, Kelly Swarts, Jinliang Yang, Jeffrey Ross-Ibarra.

Funding acquisition: Paul Bilinski, Jeffrey Ross-Ibarra.

Investigation: Paul Bilinski.

Methodology: Paul Bilinski, Patrice S. Albert, Jeremy J. Berg, Mark N. Grote, Anne Lorant, Kelly Swarts.

Project administration: Paul Bilinski.

Resources: Paul Bilinski, James A. Birchler.

Software: Jeremy J. Berg, Mark N. Grote, Jeffrey Ross-Ibarra.

Supervision: Paul Bilinski, Jeffrey Ross-Ibarra.

Validation: Patrice S. Albert, Jeremy J. Berg.

Visualization: Paul Bilinski, Mark N. Grote, Jeffrey Ross-Ibarra.

Writing – original draft: Paul Bilinski, Mark N. Grote, Jeffrey Ross-Ibarra.

Writing – review & editing: Paul Bilinski, Jeffrey Ross-Ibarra.

References

1. Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131: 452–462. https://doi.org/10.1016/j.cell.2007.10.022 PMID: 17981114

2. Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49–63. https://doi.org/10.1023/A:1016072014259 PMID: 12188048

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. https://doi.org/10.1098/rspb.1992.0093

4. Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. Genome biology 17: 37. https://doi.org/10.1186/s13059-016-0908-1 PMID: 26926526

5. Gregory T (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biological Reviews 76: 65–101. https://doi.org/10.1111/j.1469-185X.2000.tb00059.x PMID: 11325054

6. Lynch M, Conery JS (2003) The origins of genome complexity. Science 302: 1401–1404. https://doi.org/10.1126/science.1089370 PMID: 14631042

7. Kuo CH, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. Genome Research 19: 1450–1454. https://doi.org/10.1101/gr.091785.109 PMID: 19502381

8. Petrov DA (2002) Mutational equilibrium model of genome size evolution. Theoretical population biology 61, 531–544. https://doi.org/10.1006/tpbi.2002.1605 PMID: 12167373

9. 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. https://doi.org/10.1111/evo.12627 PMID: 25690700

10. Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, Ahearne TE, Cutter AD, Phillips PC (2015) Reproductive mode and the evolution of genome size and structure in Caenorhabditis nematodes. PLoS genetics 11: e1005323. https://doi.org/10.1371/journal.pgen.1005323 PMID: 26114425

11. 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. https://doi.org/10.1101/gr.212589.116 PMID: 28424354

12. Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. Science 11: 1060–2. https://doi.org/10.1126/science.287.5455.1060

13. Whitney KD, Garland T Jr. Did genetic drift drive increases in genome complexity? PLoS genetics 6(8): e1001080. https://doi.org/10.1371/journal.pgen.1001080 PMID: 20865118

14. 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. https://doi.org/10.1093/aob/mci016 PMID: 15596470

15. 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. https://doi.org/10.1073/pnas.58.2.533

16. Pegington C, Rees H, et al. (1970) Chromosome weights and measures in the Triticinae. Heredity 25: 195–205. https://doi.org/10.1038/hdy.1970.24

17. 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. https://doi.org/10.1111/j.1469-8137.2006.01919.x PMID: 17204088

18. 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. https://doi.org/10.1046/j.1365-2540.2000.00661.x PMID: 10762390

19. 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. PMID: 372199

20. Gregory TR, Hebert PD (1999) The modulation of DNA content: proximate causes and ultimate consequences. Genome Research 9: 317–324. PMID: 10207154

21. Knight CA, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology and phenotype. Annals of Botany 95: 177–190. https://doi.org/10.1093/aob/mci011 PMID: 15596465

22. Greilhuber J (1998) Intraspecific variation in genome size: a critical reassessment. Annals of Botany 82: 27–35. https://doi.org/10.1006/anbo.1998.0725

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

24. 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. https://doi.org/10.1038/ng.2678 PMID: 23793030

25. 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. https://doi.org/10.1111/nph.12247 PMID: 23550586

26. 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. https://doi.org/10.1139/G10-077 PMID: 21164539

27. 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. https://doi.org/10.1038/srep11636 PMID: 26109237

28. 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. https://doi.org/10.1006/anbo.1998.0757

29. Laurie DA and Bennett MD (1985) Nuclear DNA content in the genera Zea and Sorghum. Intergeneric, interspecific and intraspecific variation Heredity 55: 307–313.

30. 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. https://doi.org/10.1073/pnas.0603080103

31. 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. https://doi.org/10.1105/tpc.104.025627 PMID: 15659640

32. 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.

33. Bretting P, Goodman M (1989) Karyotypic variation in mesoamerican races of maize and its systematic significance. Economic Botany 43: 107–124. https://doi.org/10.1007/BF02859330

34. Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genetics 10: e1004412. https://doi.org/10.1371/journal.pgen.1004412 PMID: 25102153

35. Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. Genetics 181: 1399–1413. https://doi.org/10.1534/genetics.108.097238 PMID: 19153259

36. 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. https://doi.org/10.1371/journal.pone.0047659 PMID: 23155371

37. 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. https://doi.org/10.1371/journal.pgen.1003477 PMID: 23671421

38. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. Genome Biology and Evolution 5: 1594–1609. https://doi.org/10.1093/gbe/evt109 PMID: 23902747

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

40. 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. https://doi.org/10.1159/000314342 PMID: 20551613

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. https://doi.org/10.1038/ncomms9974 PMID: 26584889

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. https://doi.org/10.1073/pnas.1013011108

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. https://doi.org/10.1534/genetics.115.178327 PMID: 26078279

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. https://doi.org/10.1007/BF00226155 PMID: 24226450

45. Bennett MD (1987) Variation in genomic form in plants and its ecological implications. New Phytologist 106: 177–200. https://doi.org/10.1111/j.1469-8137.1987.tb04689.x

46. Sinha RP, Häder DP (2002) UV-induced DNA damage and repair: a review. Photochemical & Photobiological Sciences 1: 225–236. https://doi.org/10.1039/b201230h

47. Piperno DR, Flannery KV. The earliest archaeological maize (Zea mays L.) from highland Mexico: new accelerator mass spectrometry dates and their implications (2001) Proceedings of the National Academy of Sciences 98(4):2101–3. https://doi.org/10.1073/pnas.98.4.2101

48. 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. https://doi.org/10.2307/2446305 PMID: 21684902

49. Chevin LM, Hospital F (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics, 180(3), 1645–1660. https://doi.org/10.1534/genetics.108.093351 PMID: 18832353

50. 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. https://doi.org/10.1038/ng.2313 PMID: 22660545

51. Dennis E, Peacock W (1984) Knob heterochromatin homology in maize and its relatives. Journal of Molecular Evolution 20: 341–350. https://doi.org/10.1007/BF02104740 PMID: 6439888

52. 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. PMID: 10471723

53. Kanizay LB, Pyhäjärvi T, Lowry EG, Hufford MB, Peterson DG, Ross-Ibarra J, Dawe RK (2013) Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *Zea mays* Genetics 153: 415–426.

54. Kanizay LB, Pyhäjärvi T, Lowry EG, Hufford MB, Peterson DG, Ross-Ibarra J, and Dawe RK (2013) Intragenomic conflict between the two major knob repeats of maize. Genetics 194: 81–89. https://doi.org/10.1534/genetics.112.148882 PMID: 23457233

55. 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. https://doi.org/10.1098/rspb.1972.0042 PMID: 4403285

56. 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. https://doi.org/10.1098/rspb.2011.1284

57. 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. https://doi.org/10.1006/bcmd.2001.0457 PMID: 11783946

58. 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. https://doi.org/10.1111/j.1469-8137.2008.02528.x PMID: 18564303

59. 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. https://doi.org/10.1038/nature16496 PMID: 26760202

60. Van Volkenburgh E, Boyer JS (1985) Inhibitory effects of water deficit on maize leaf elongation. Plant Physiology 77: 190–194. https://doi.org/10.1104/pp.77.1.190 PMID: 16664006

61. 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. https://doi.org/10.1093/jxb/47.11.1689

62. 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.

63. 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. https://doi.org/10.1007/s001220051315

64. Rodriguez FJ, Sanchez GJ, Baltazar MB, 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.

65. 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. https://doi.org/10.1111/j.1439-0523.1994.tb00690.x

66. 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. https://doi.org/10.1073/pnas.1609686113

67. 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. https://doi.org/10.1126/science.1117768 PMID: 16099949

68. 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. https://doi.org/10.1105/tpc.107.051920 PMID: 17540715

69. Jian Yinqiao and Xu Cheng and Guo Zifeng and Wang Shanhong and Xu Yunbi and Zou Cheng (2017) Maize (Zea mays L.) genome size indicated by 180-bp knob abundance is associated with flowering time. Scientific Reports 7: 11636. https://doi.org/10.1038/s41598-017-06153-8

70. Hessen DO, Jeyasingh PD, Neiman M, Weider LJ (2010) Genome streamlining and the elemental costs of growth. Trends in Ecology & Evolution 25: 75–80. https://doi.org/10.1016/j.tree.2009.08.004

71. 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. https://doi.org/10.1371/journal.pone.0019379 PMID: 21573248

72. 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. https://doi.org/10.1371/journal.pone.0090346 PMID: 24587335

73. Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. G3: Genes—Genomes—Genetics 2: 1405–1413. https://doi.org/10.1534/g3.112.004259 PMID: 23173092

74. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, et al. (2015) Linkimpute: Fast and accurate genotype imputation for nonmodel organisms. G3: Genes—Genomes—Genetics 5: 2383–2390. https://doi.org/10.1534/g3.115.021667 PMID: 26377960

75. Alexander David H and Lange Kenneth (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC bioinformatics 12: 246. https://doi.org/10.1186/1471-2105-12-246 PMID: 21682921

76. 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. https://doi.org/10.1111/j.1365-313X.2005.02591.x PMID: 16359397

77. Swarts K., Gutaker R.M., Benz B., Blake M., Bukowski R., Holland, et al. (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America Science 357:512–5. https://doi.org/10.1126/science.aam9425 PMID: 28774930

78. 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. https://doi.org/10.1007/s00412-014-0483-8 PMID: 25190528

79. 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. https://doi.org/10.1007/BF02254939 PMID: 8653263

80. 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. https://doi.org/10.1126/science.1178534 PMID: 19965430

81. 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. https://doi.org/10.1371/journal.pgen.1000732 PMID: 19936065

82. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows—wheeler transform. Bioinformatics 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

83. 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. https://doi.org/10.1534/genetics.107.080101 PMID: 18385116

84. MacKinnon David P and Rose JS and Chassin L and Presson CC and Sherman SJ (2000) Contrasts in multiple mediator models Multivariate applications in substance use research: New methods for new questions: 141–160.

85. 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.

86. 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. https://doi.org/10.1104/pp.109.3.861 PMID: 12228638

87. 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. https://doi.org/10.1016/S1573-5214(00)80013-5

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

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

90. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44: 821–824. https://doi.org/10.1038/ng.2310 PMID: 22706312

91. 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. https://doi.org/10.1186/s13742-015-0047-8