

RESEARCH ARTICLE

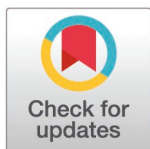
# Genetic and environmental influences on the distributions of three chromosomal drive haplotypes in maize

Meghan J. Brady<sup>1</sup>, R. Kelly Dawe<sup>1,2\*</sup>

**1** Department of Genetics, University of Georgia, Athens, Georgia, United States of America,

**2** Department of Plant Biology, University of Georgia, Athens, Georgia, United States of America

\* [kdawe@uga.edu](mailto:kdawe@uga.edu)



## OPEN ACCESS

**Citation:** Brady MJ, Dawe RK (2025) Genetic and environmental influences on the distributions of three chromosomal drive haplotypes in maize. PLoS Genet 21(7): e1011742. <https://doi.org/10.1371/journal.pgen.1011742>

**Editor:** Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, UNITED STATES OF AMERICA

**Received:** May 19, 2025

**Accepted:** July 6, 2025

**Published:** July 16, 2025

**Copyright:** © 2025 Brady, Dawe. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All code can be found at [https://github.com/dawelab/CDH\\_Distribution\\_In\\_Maize](https://github.com/dawelab/CDH_Distribution_In_Maize). GBS data can be accessed at the SRA under PRJNA1257855.

**Funding:** This work was supported by an NSF fellowship to MJB (2236869), an NSF grant

## Abstract

Meiotic drive elements are regions of the genome that are transmitted to progeny at frequencies that exceed Mendelian expectations, often to the detriment of the organism. In maize there are three prevalent chromosomal drive elements known as Abnormal chromosome 10 (Ab10), K10L2, and the B chromosome. There has been much speculation about how these drivers might interact with each other and the environment in traditional maize landraces and their teosinte ancestors. Here we used genotype-by-sequencing data to score more than 10,000 maize and teosinte lines for the presence or absence of each driver. Fewer than ~0.5% of modern inbred lines carry chromosomal drivers. In contrast, among individuals from 5331 open-pollinated landraces, 6.32% carried Ab10, 5.16% carried K10L2, and 12.28% carried at least one B chromosome. These frequencies are consistent with those reported in previous studies. Using a GWAS approach we identified unlinked loci that associate with the presence or absence of the selfish genetic elements. Many significant SNPs are positively associated with the drivers, suggesting that there may have been selection for alleles that ameliorate their negative fitness consequences. We then assessed the contributions of population structure, associated loci, and the environment on the distribution of each chromosomal driver. There was no significant relationship between any chromosomal driver and altitude, contrary to conclusions based on smaller studies. Our data suggest that the distribution of the major chromosomal drivers is primarily influenced by neutral processes and the deleterious fitness consequences of the drivers themselves. While each driver has a unique relationship to genetic background and the environment, they are largely unconstrained by either.

## Author summary

Meiotic drivers are selfish genetic elements that are transmitted to progeny at frequencies that exceed what is expected by chance. There are three such

to RKD (1925546) and an NIH training grant to MJB (T32GM007103). The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

meiotic drivers in maize known as Ab10, K10L2 and the B chromosome. Prior data indicate that each driver lowers plant fitness to some degree, partially explaining the fact that they are usually observed at low frequencies, but the impact of environment and genetic modifiers has not been investigated. Here we identified the presence or absence of each driver using low coverage sequencing data from over 10,000 maize inbreds, open pollinated landraces, and accessions of teosinte -- the grass-like ancestor of maize. We found that they exist at 5–12% frequencies in open pollinated varieties, but are rare in modern inbreds. We went on to use location data to assess the impact of environmental variables on their distribution, and applied genome-wide association methods to look for genetic modifiers that might contribute to their prevalence. We found that neither the environment nor genetic modifiers have strong influences on the distribution of the three meiotic drive elements, suggesting that their low but consistent frequencies are determined primarily by the fitness consequences of the drivers themselves.

## Introduction

While most genes in most species are transmitted in predictable Mendelian patterns, there are striking exceptions. Genes, gene complexes, sections of chromosomes, and entire chromosomes have evolved mechanisms that ensure they are transmitted at higher frequencies than would be expected based on chance [1–4]. These genetic elements generally confer no selective advantages to the species, and are often deleterious, but are nevertheless maintained in populations based on their selfish properties. They are often described with the catch-all term meiotic drive [5], though only a subset of meiotic drivers affect meiosis. Meiotic drivers that manipulate meiosis often do so in species where female meiosis results in only one functional egg cell. An example is the preferential transmission of larger centromeres towards the egg cell in some mice strains [6]. A more common class of meiotic driver interferes with the function of male gametes, often by setting up a dynamic where sperm or pollen are killed by a toxin unless they inherit the antidote present on the driving chromosome [7]. The term meiotic drive is also used to describe the maintenance of supernumerary B chromosomes [8], as well as many other varied phenomena, including biased gene conversion processes [9], mobile toxin-antidote systems [10], and engineered gene drive systems based on CRISPR [11].

The maize genome contains at least three meiotic drive elements that distort transmission of large chromosomal regions: Abnormal Chromosome 10 (Ab10), K10L2, and the B chromosome. Ab10 chromosomes contain large haplotypes at the end of normal chromosome 10 (N10) that acts as female meiotic drivers. Approximately 14% of the maize genome is composed of tandem repeat arrays called knobs [12]. They come in two classes defined by their repeat element: TR-1 and knob180. The Ab10 haplotypes contain large knobs predominated by repeats from both classes [13,14]. Ab10 haplotypes also encode two kinesin proteins, KINDR and TRKIN, which interact with knob180 and TR-1 knobs respectively [15, 16]. The kinesins pull the knobs ahead of the centromeres during meiotic anaphase resulting in their preferential transmission to

the egg cell [17,18]. By this mechanism the Ab10 haplotype as well as knobs throughout the genome are preferentially transmitted to ~60–80% of the next generation [18]. There are at least three major cytological types of Ab10 known as Ab10-I, Ab10-II and Ab10-III [19]. Ab10 is present in ~5% of maize landraces [20,21], but is prevented from going to fixation because it impairs fitness when homozygous [17,22]. Ab10 is recognized as an important driver of maize genome evolution [15,23].

The K10L2 variant is similar to Ab10 but is much smaller, having only two TR-1 rich knobs and the *Trkin* gene. It shows 51–52% meiotic drive when paired with normal chromosome 10 (N10) [24]. When K10L2 is paired with Ab10, Ab10 drive is severely suppressed, demonstrating K10L2 not only can drive itself but compete with the stronger Ab10 drive system. Recent data demonstrate that the region between the two TR-1-rich knobs is similar in sequence to a portion of at least one structural variant of Ab10 (Ab10 Type I), suggesting that Ab10 may have subsumed the K10L2 haplotype in recent evolutionary history [14]. K10L2 is present in about 5% of maize landraces [20, 21], but the fact that at least one traditional inbred line is homozygous for K10L2 suggests that it may not have severe effects on fitness when homozygous [24].

The B chromosome is a ~150Mb supernumerary chromosome composed primarily of transposable elements (TEs), organellar sequences, and a B-specific repeat element [25]. The B chromosome can accumulate to high copy numbers by a mechanism that takes advantage of the fact that there are two sperm in each pollen grain; one fertilizes the egg cell and the other fertilizes the central cell that gives rise to the starchy endosperm. The B chromosome normally non-disjoins at the second pollen mitosis and the sperm carrying two copies of the B chromosome preferentially fertilizes the egg [26]. There is known variation among lines for the efficiency of the second step. Most lines allow sperm carrying the B chromosome to preferentially fertilize the egg, but multiple lines do not [8], or even reverse it, such that B chromosome-carrying sperm preferentially fertilize the central cell [27]. In natural populations, B chromosomes are found in about 10% of landraces with copy numbers that are usually between 1–3 but may be as high as 14 [20,21,28]. When B chromosomes are present at higher than 15 copies the plants display reduced seed set and pollen viability [29].

The factors affecting the frequency and distribution of maize chromosomal drivers are not well understood. A prior modeling study showed that the known fitness defects associated with Ab10 can largely, but not completely, explain its low frequency in landraces and teosintes [22], suggesting that there are other variables that influence its distribution. Likely influences include the environment and genetic modifiers [2,30]. Prior evidence suggests that altitude may influence the distribution of both Ab10 and the B chromosome [23,28,31–34], which may be related to the fact that high-altitude maize lines tend to have smaller genomes [34–36]. Genetic variation outside of the drive haplotypes is also expected to alter their frequencies. Alleles that reduce the efficiency of drive (suppressors) should be selected for when the fitness consequences of the drive haplotypes are high [2,30,37–39]. It is possible that the resistance of some lines to the preferential fertilization of B chromosomes reflects ongoing selection against high B chromosome copy numbers [8]. The three maize drivers may also interact with each other in unpredictable ways. For instance, in some backgrounds, the B chromosome causes the breakage of chromosomes at knobs, including Ab10 [40,41].

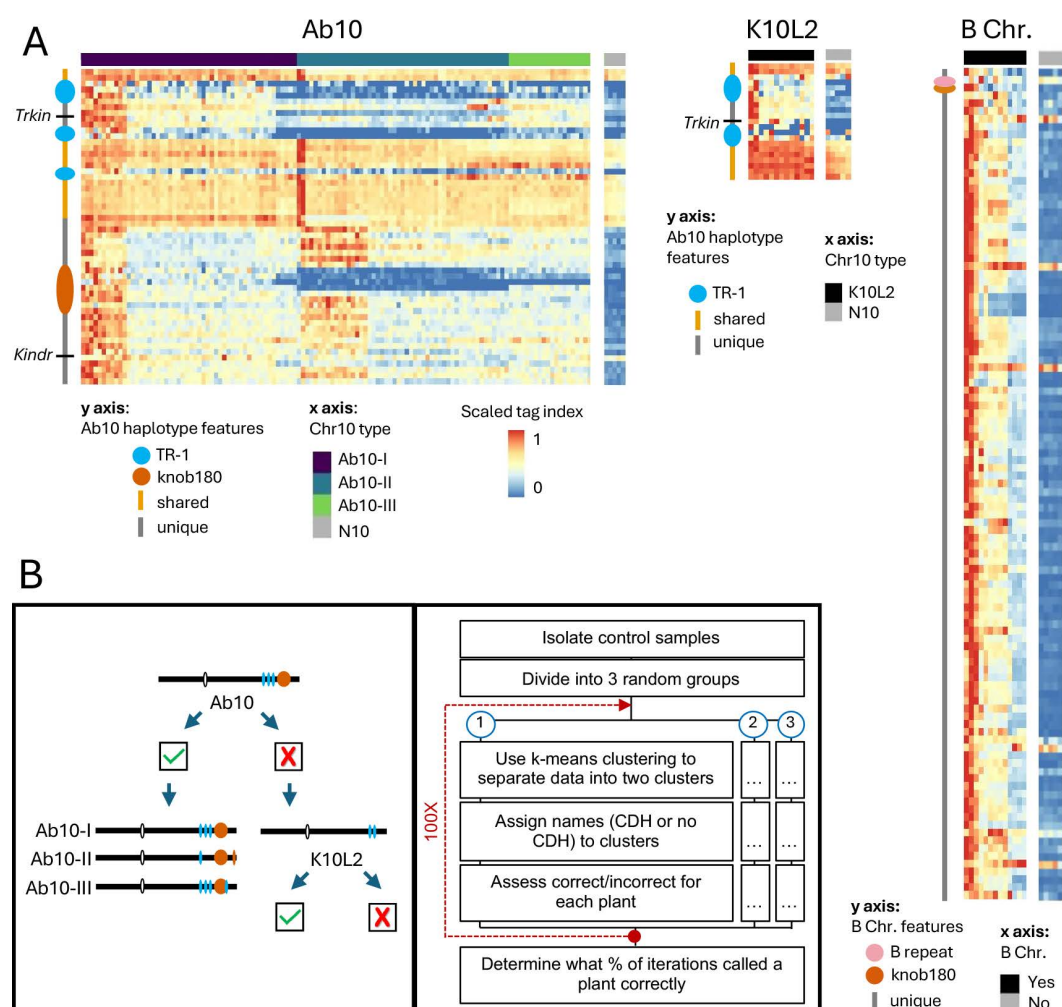
Here we developed methods to detect maize chromosomal drive haplotypes in genotype-by-sequencing (GBS) data, and scored their presence or absence in ~10,000 maize inbred, maize landrace, and teosinte individuals. We then determined how their distribution relates to genetic background, population structure, and the environment. In open pollinated landraces and teosintes, their distributions are significantly influenced by genetic modifiers and environmental factors, though the effects are small. The combined data, and the fact that all three drivers are present at very low frequencies in modern inbred lines, support the view that the major limitation to their spread is the negative fitness consequences of the drivers themselves.

## Results

### Genotype-by-sequencing data can reliably detect large structural variants

We speculated that the low coverage sequence data from GBS studies might be useful for identifying large chromosomal drive haplotypes (referred to as CDH in this study). To test the feasibility of this idea, we first generated GBS data from a collection of control lines carrying different isolates of Ab10, K10L2, and B chromosomes, as well as associated controls

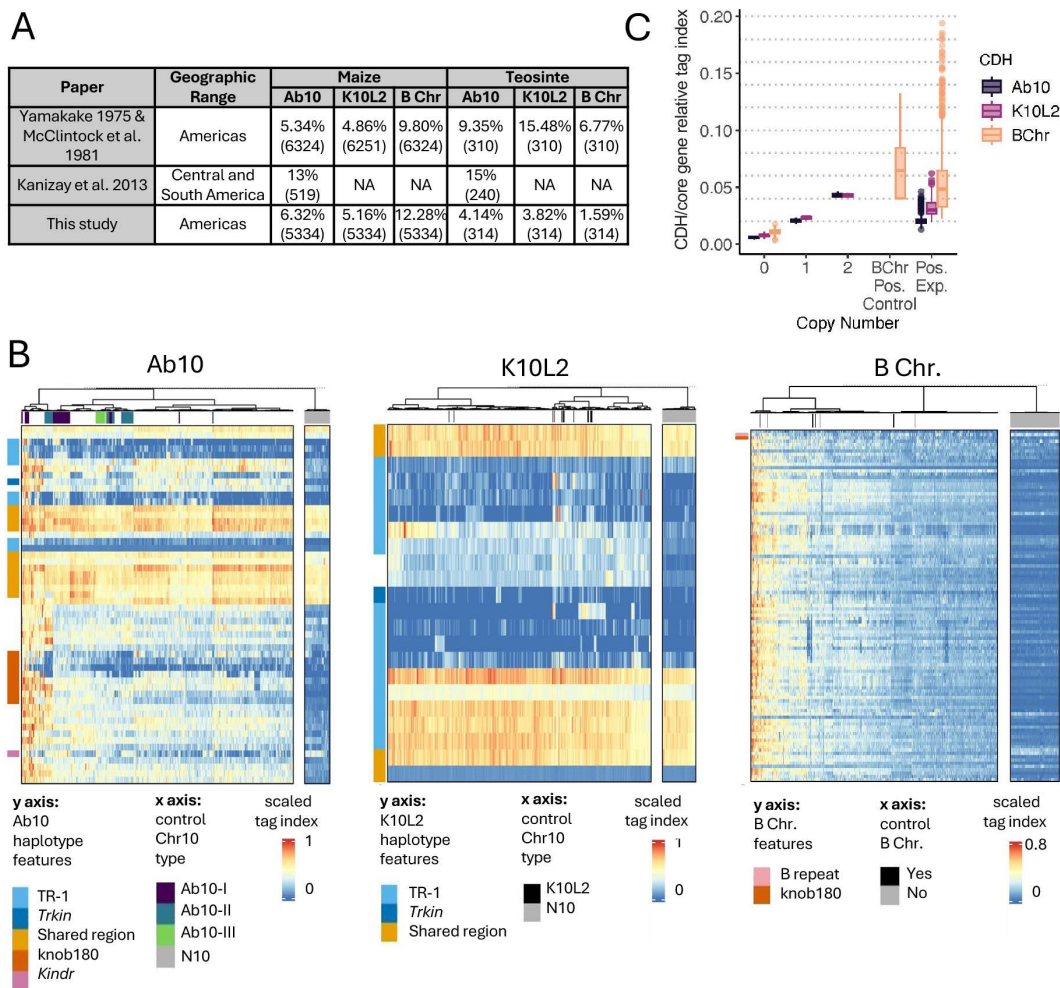
(S4 Table). The GBS data was then mapped to reference genomes carrying each drive haplotype: Ab10 [42], K10L2 [14], and the B chromosome [25]. The GBS protocol results in sequences that map to discrete regions in the genome that are called tag sites. To assess the read coverage over tag sites, we computed a tag index, which is a single value that describes both the number of tag sites and the read depth over tag sites in 1 Mb windows (S1A Fig). When the tag index was plotted as a heat map, the presence or absence of a CDH became visually apparent (Fig 1A). We then automated the scoring of CDHs using an iterative k-means clustering approach, where similar data points are grouped into a specified number of clusters (Fig 1A and 1B). Using this method we achieved 100% accurate discrimination of the presence and absence of each CDH in our control dataset. To estimate CDH copy number, we normalized CDH tag depth by the average tag depth across all single copy core genes [12] and correctly identified all Ab10 and K10L2 homozygous samples in our control set (Fig 2C). We applied the same method to estimate the copy numbers of B chromosomes, though in this case we did not know the copy number in our control samples.



**Fig 1. Detection of CDHs in GBS data.** A. Min/max scaled tag index for Ab10, K10L2, and the B chromosome. The CDH status of each control sample is shown on the x axis. Relevant features of each CDH are highlighted on the y axis. A scaled tag index was calculated for each 1 Mb bin; here the CDH haplotypes are scaled to their relative lengths. Ab10-1, Ab10-II, and Ab10-III are different cytological types of Ab10 that were included in our control set. There are multiple replicates of the same CDH chromosomes in each panel. B. Stepwise manner of differentiating Ab10 from K10L2 (left) and general workflow for identifying CDHs using scaled tag index bins (right). Check indicates present, x indicates absence.

<https://doi.org/10.1371/journal.pgen.1011742.g001>





**Fig 2. Detection of CDHs in experimental samples.** A. Summary table for previous studies of CDH distribution as well as the results from the work presented here for comparison. Numbers in parentheses indicate the number of individuals surveyed. The data from Yamakake 1975 [20] were incorporated into the McClintock et al. 1981 [21] report. The 314 teosinte individuals assayed in this study were sampled from 79 accessions (see S1 Table). B. Scaled tag index for all CDH positive lines and N10 controls. CDH-positive controls and CDH-positive experimental samples are intermixed: those from controls are indicated by colors in the bars below the dendrograms (that show relatedness), those with no color are experimental samples. The N10 control samples are plotted separately. C. Estimated copy number of all 3 CDHs in control and experimental samples. B Chr Pos. Control refers to samples that are B chromosome positive via PCR (copy number was unknown). Pos. Exp. refers to all CDH positive experimental samples identified in public GBS data. Dotted lines show approximate one copy increments as determined by the relationship between 1 and 2 copies in control samples. The values on the y axis (CDH/core gene relative tag index) are not integers because much of the CDH sequence is not single copy and poorly conserved, resulting in relatively fewer tags aligning to the reference.

<https://doi.org/10.1371/journal.pgen.1011742.g002>

We then went on to identify CDH in previously published GBS data from landraces, teosintes and inbreds [43–45] using a similar stepwise k-means clustering approach (S1B Fig). Our Ab10 detection pipeline cannot detect K10L2, and our K10L2 pipeline cannot distinguish Ab10 from K10L2. Therefore we ran the Ab10 pipeline first, and then ran samples called as negative through the K10L2 detection pipeline (Fig 1B). Similarly, to accurately differentiate high copy B chromosome lines from low copy lines, we ran two different pipelines in sequence (see Methods). To ensure that clustering was driven by our control samples, we used roughly equal numbers (CDH positive and CDH negative) and randomly introduced experimental samples so that they were no more than 25% of the control samples (see Methods). Every sample was

assayed for each CDH 125x to gain an estimate of call confidence. We obtained confident calls (>95% calls) for more than 99% of samples assayed for each CDH (Figs 2B and S1B).

When developing our set of Ab10 control samples we had included isolates of the three major cytological types, known as Ab10-I, Ab10-II, and Ab10-III, that differ in the appearance of the major knobs on the haplotype [19]. We developed a random forest model to detect Ab10 type from this control data. When this model was applied to samples with unknown Ab10 types, it became clear that there is more diversity in natural Ab10 samples than is present in our three control types (Figs 1A and 2B). Using a confidence threshold that maintained the visual differences between types (see Methods), only 47 of 352 experimental Ab10-positive samples were classable (S2A Fig). To explore the variation in Ab10 type further, we performed a PCA on the scaled tag index of diagnostic regions. Plotting the first two principal components, accounting for 50% of the variation, revealed a nearly uniform distribution of the samples suggesting extensive genetic variation among Ab10 types (S2B Fig). Ab10-I and Ab10-II are known to recombine with each other in experimental populations [41], and some of the tag index patterns suggest recombinants are also present in natural populations (S2A Fig).

### Frequencies of three CDHs in maize

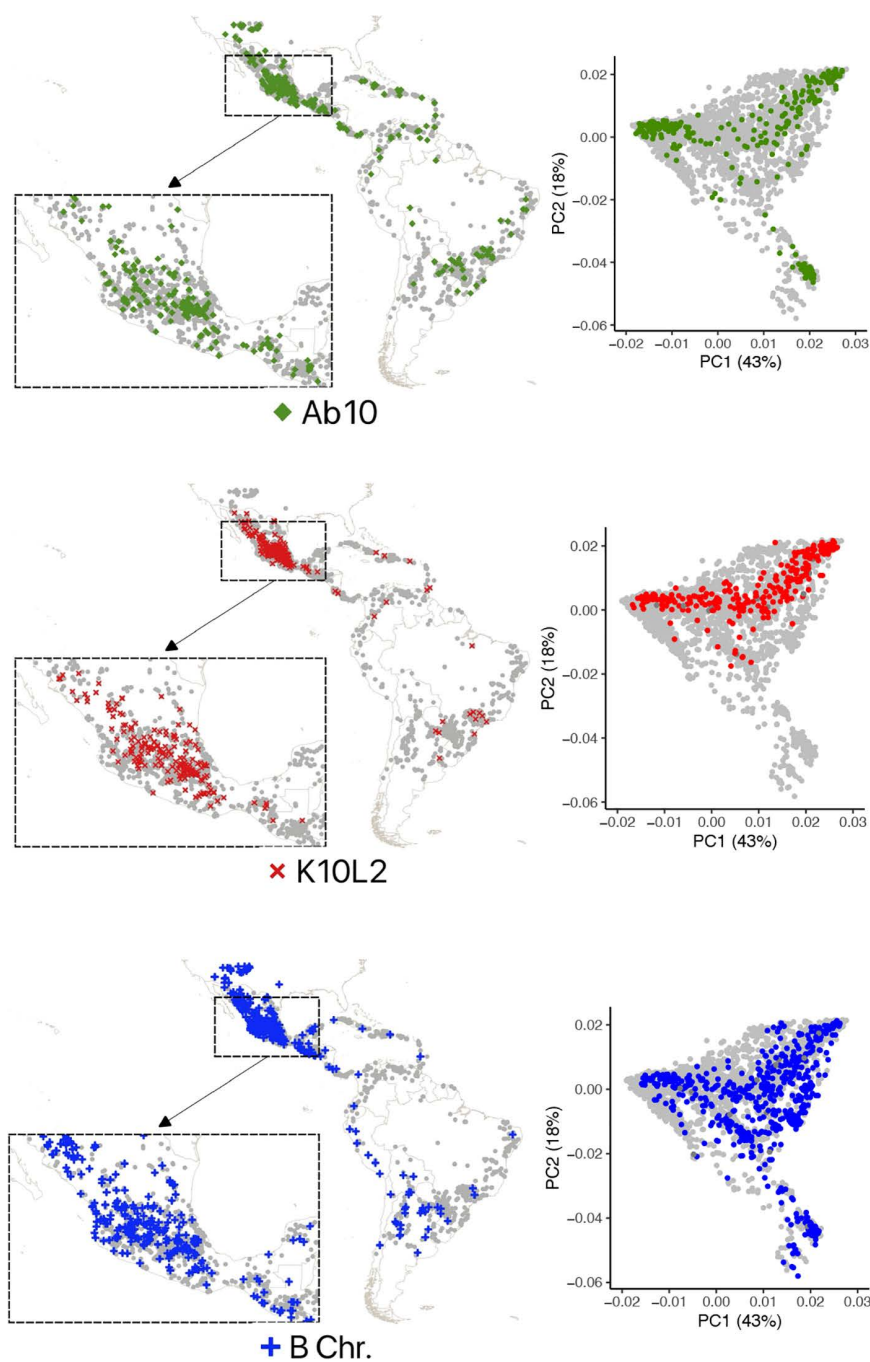
Our data on the frequencies of maize CDH in landraces are similar to what was previously reported by Kato, McClintock and Blumenschein based on their analysis of meiotic pachytene chromosomes ([20, 21], Fig 2A). We observed Ab10 at a frequency of 6.32%, K10L2 at 5.16%, and the B chromosome at 12.28% on a per-plant basis. Each CDH occurs across the entire range of the landrace accessions assayed (Fig 3). We observed substantially fewer CDH in teosintes than was previously reported [20, 21], which may reflect the fact that the 314 teosinte individuals were sampled from a smaller number of 79 accessions (S1 Table [43]). Ab10 and the B chromosome as well as K10L2 and the B chromosome occur together roughly as frequently as expected by chance (S1 Table).

We also identified all three CDHs in inbred maize lines [45], though at low frequencies (S1 and S2 Tables). There were 2 inbreds that scored positive for Ab10, 15 that scored positive for K10L2 and 12 that scored positive for the B chromosome (S2 Table). K10L2 had previously been detected in an inbred line [24], but Ab10 and the B chromosomes were thought to be absent from inbred lines [18, 29]. To confirm that the GBS genotyping was correct, we obtained seeds from six of the inbred lines – two that scored positive for each CDH. The seeds were from the Germplasm Research Information Network which maintains bulked samples derived from multiple ears. This is done to maintain any residual genetic diversity. We found that all six of the inbred lines contained individuals that scored positive for the CDH by PCR as well as individuals that scored negative by PCR (S2 Table). While these lines are presumed to be pure breeding for most of the genome, the CDH chromosomes are still segregating for presence or absence. The fact that very few inbreds carry CDH, and many of those that do are incompletely inbred, supports the view that all three CDHs have deleterious fitness consequences.

### Relationship of three CDH in maize to genetic variants

We went on to examine the relationship between each CDH and genetic variation in the genome. Using the GBS data from maize landraces, we identified high confidence SNPs that did not overlap any CDH or transposable element, resulting in ~50,000 usable SNPs (S3 Fig). These data allowed us to identify population structure within our accessions (Fig 3) and perform a genome wide association study (GWAS) on high confidence SNPs that associate with each CDH. Given our binary traits and relatively large number of individuals we chose to impose a stringent significance threshold of  $5 \times 10^{-8}$ .

We found that Ab10 was positively associated with seven SNPs on chromosomes 3, 4, 8, and 9 and negatively associated with four SNPs on chromosome 3, and 10. Linkage disequilibrium in maize decays over a span of 1–10 kb [46]), meaning that SNPs within a few kb are likely to be on one haplotype. As expected, several of the significantly associated loci are marked by two or more closely linked SNPs (S3 Table). A highly associated SNP on chromosome 9 is likely an alignment artifact, as it occurs within a sequence that has homology to the Ab10-I v1 reference [42] and the p value is similar to what we observed for SNPs that are tightly linked to a CDH (S4 Fig); it was not used in subsequent analysis.



**Fig 3. Maps and PCA plots of maize landraces assayed as CDH negative or positive.** For each CDH, a location map is shown (left) and a plot of the first 2 principal components of population structure based on whole genome SNP profiling (right). Each grey dot indicates the location of a landrace that was assayed. Those with a CDH are highlighted in colors. The numbers in parentheses indicate the percent of variation that principal coordinates account for. Base maps from <https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries> were annotated using QGIS (<http://qgis.org>).

<https://doi.org/10.1371/journal.pgen.1011742.g003>

It is also possible that two other SNPs that are positively associated with very low p values (Ab10 Chr8 SNP1 and B chromosome Chr3 SNP1) are artifacts of alignment from CDH reads that differ significantly from our references; however in the absence of other evidence we have left them in the analysis. K10L2 was positively associated with five SNPs and negatively associated with two SNPs (Fig 4A). The B chromosome was positively associated with two SNPs and B chromosome copy number was not associated with any SNPs (Fig 4A). Of the total 20 associated SNPs, 13 overlap annotated genes (S3 Table). Henceforth we refer to loci associated with each CDH as putative genetic modifiers for ease of interpretation.

### Combined effects of genetic variants and environment on the distribution of CDHs

To test the impact of location and environment on CDH distribution, we chose maize landrace and teosinte lines that were confidently identified as CDH positive or negative and had GPS coordinates for their collection location [Ab10 = 3718, K10L2 = 3421, B Chr = 3718, B Chr Copy Number = 542] (S1 Table). We first assessed the effects of elevation while also accounting for population structure. In contrast to previous studies [19,28,32], we found no relationship between elevation and Ab10 or the B chromosome (presence/absence or copy number, S5 Fig). There was a weak positive relationship between K10L2 and elevation when considering population structure alone (S5 Fig).

We went on to develop models to test the effects of specific climatic variables [47] and soil conditions [48]. We began each model using the top 10 principal coordinates of population structure, as well as elevation, mean temperature of warmest quarter, precipitation of warmest quarter, temperature seasonality, precipitation seasonality, average annual solar radiation, average annual wind, average annual water vapor pressure, soil nutrient availability, soil rooting conditions, soil oxygen availability to roots, soil excess salts, and soil toxicity. For the B chromosome model, we also included Ab10 and K10L2 presence/absence to test whether they are independently distributed. For each CDH we generated a simplified model by removing variables that were not significantly associated (p value greater than 0.01) one at a time until all variables were significantly associated (p value less than 0.01). The presence or absence of Ab10 or K10L2 did not significantly affect the distribution of the B chromosome, consistent with expectations based on overlapping percentages (S1 Table). There was no significant association between elevation and Ab10, K10L2, or the B chromosome when considering both population structure and climatic variables (See S6 Fig for variables with significant associations).

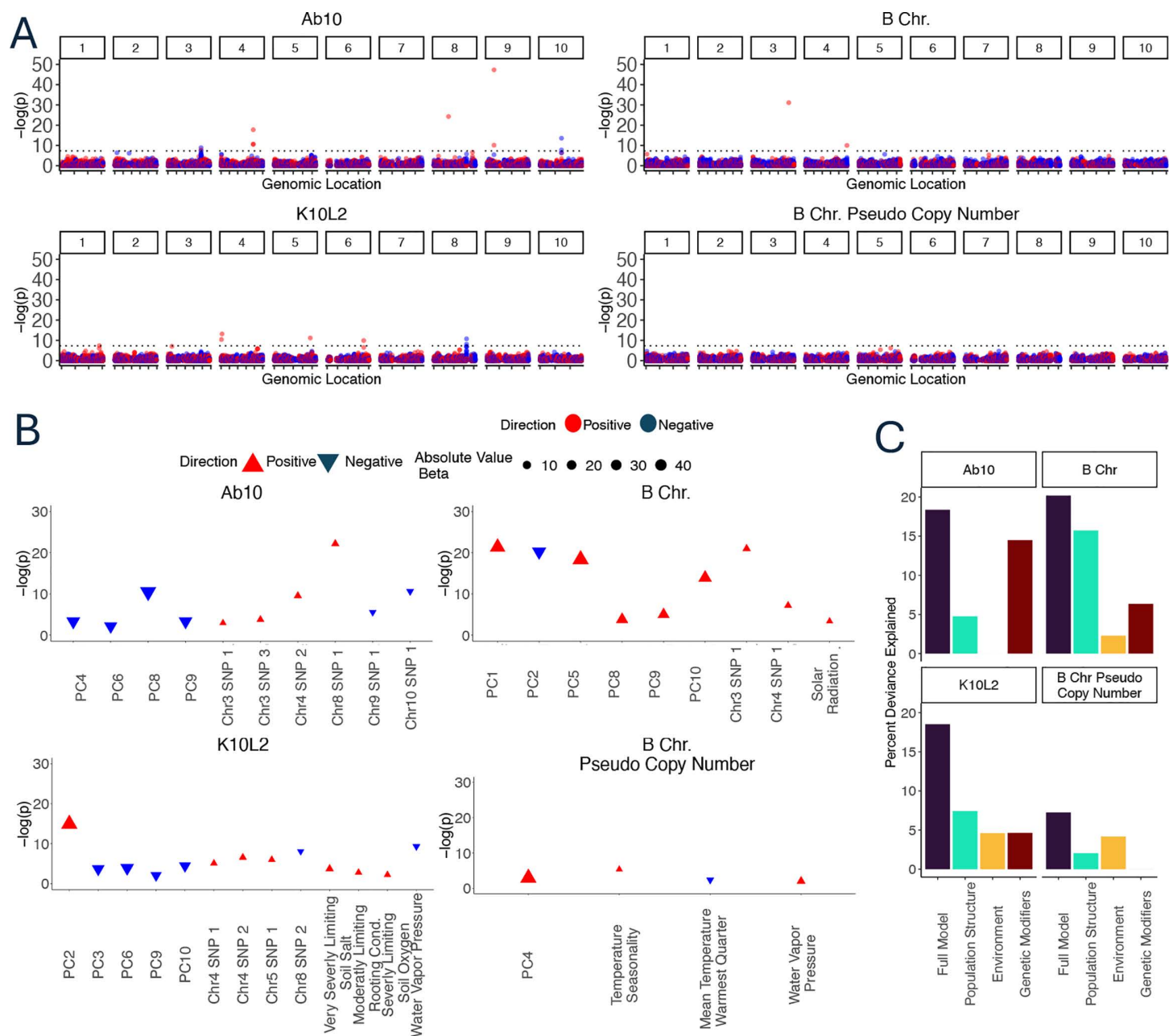
We then added genetic modifiers to the model, so as to include population structure, genetic modifiers and the environment (Fig 4). We calculated the amount of deviance explained by the full model and each of the variable classes [population structure, environment, and genetic modifiers] separately (Fig 4C). After accounting for all variables, only 12 of the original 20 SNPs showed significant associations. Due to interactions among the variables (which we did not pursue here), the deviance explained by each class of variables individually does not sum to the deviance explained by the full model.

**Ab10.** Ab10 significantly associates with six genetic modifiers and four principal components of population structure accounting for 18.4% of the deviance (Fig 4B and 4C). The environment seems to have no effect on the distribution of Ab10.

**K10L2.** K10L2 significantly associates with five principal components of population structure, four genetic modifiers, and four environmental factors accounting for 18.5% of the deviance (Fig 4B and 4C). It is associated with lower water vapor pressure. Further we found that K10L2 was overrepresented in poor quality soil specifically with respect to excess soil salts, soil rooting conditions, and soil oxygen (Fig 4B).

**B Chromosome.** The B chromosome significantly associates with six principal components of population structure, two genetic modifiers and one environmental variable accounting for 20.2% of the deviance. The B chromosome is more likely to occur in regions with higher solar radiation (Fig 4B and 4C).





**Fig 4. Relationship of CDH to the environment and genetic loci.** **A.** Manhattan plots of SNPs that associate with each CDH and B chromosome copy number. Dotted black line indicates a p value of  $5 \times 10^{-8}$ . The Ab10-associated SNP with very high p value ( $-\log(p)$  close to 50) on chromosome 9 may be an artifact of read alignment to the Ab10 haplotype itself. **B.** Plots of fully simplified generalized linear models for each CDH including population structure, genetic loci, and environmental variables. Numbers beginning with PC are principal components of population structure. Color and orientation of the triangle indicates whether a SNP is positively or negatively associated with the CDH. The size of the triangle represents the effect size. Very Severely and Severely Limiting Soil Salt refers to growth limiting excess soil salts. Moderately Limiting Rooting Cond. refers to moderately growth limiting rooting conditions. The direction of the relationship between Ab10 and Chr3 SNP1 and Chr9 SNP1 changes when considering each SNP in isolation (A) or with the other SNPs and environmental variables (B). **C.** Partitioned deviance of each model shown in B. The partitions do not sum to the full model due to shared variation among the partitions.

<https://doi.org/10.1371/journal.pgen.1011742.g004>

B chromosome copy number is associated with one principal component of population structure, and three environmental variables accounting for 7.25% of the deviance. Specifically, B chromosome copy number is higher in environments with higher temperature seasonality and water vapor pressure but a lower mean summer temperature.

## Discussion

In this work we used GBS data to identify chromosomal drive haplotypes in over 10,000 maize and teosinte accessions with the aim of better understanding how drive haplotypes interact with each other and the environment. While GBS was originally developed as a method to score SNPs [49], our approach using k-means clustering illustrates that the low coverage sequence data can also be used to identify large haplotypes such as CDHs. The method can accurately identify Ab10, K10L2 and B chromosomes, as well as variants that differ from the reference haplotypes that may be worthy of further study (S2 Fig).

Although all three CDH can be transmitted at super-Mendelian levels, they are only found in natural populations at low frequencies, indicating that their distributions are limited by genetic or environmental factors. There is little doubt that the fitness consequences of the drivers themselves are major limiting factors. Lines carrying Ab10 have reduced seed number and weight, and the B chromosome causes sterility when present at high copy numbers [17, 29]. However, at least in the case of Ab10, modeling shows that the known fitness defects are not sufficient to explain the low observed frequencies of Ab10 [22]. Here we assessed the importance of population structure, genetic modifiers, and the environment to the distribution of Ab10, K10L2, and B chromosomes, as well as their interactions with each other.

The environment appears to have little effect on the distribution of any of the maize chromosomal drivers. Prior data suggested that the B chromosomes and Ab10 might occupy different altitudinal clines [19, 28, 32]. However, after controlling for population structure, we found no correlation with elevation for either driver, and that Ab10 and the B chromosome occur together as frequently as expected by chance (S1 Table). There has been an assumption in the prior literature that since both the B chromosome and Ab10 increase genome size, they should be selected against at higher altitudes where smaller-genome lines are more fit [34]. While this may be true, the level of selection may be weaker than is commonly assumed. Recent results suggest that in the large-genome maize plant, a gain of 14 Mb results in a 0.1% reduction in yield [50]. By this reasoning, a single ~30 Mb knob [13], the ~85 Mb Ab10 haplotype [14], or the ~150 Mb B chromosome would be expected to result in an ~<1% drop in yield, which may not be sufficient to counteract the selfish properties of these powerful drivers.

We used a GWAS approach to identify potential genetic modifiers that impact the distribution of the maize CDH. At the outset, we anticipated that most of the modifiers would be negatively associated, and represent potential suppressors that reduce the efficiency of the drivers. Extensive prior literature suggests that such suppressors are likely to evolve when the fitness burden is high [2, 30, 37, 38]. For instance, we anticipated identifying SNPs that are linked to alleles that suppress the preferential fertilization of the egg by B chromosome-carrying sperm [8]. However, only 3 of the 12 associated loci showed such a negative relationship (Fig 3B, significant SNPs after accounting for environment). The remaining 9 SNPs were positively associated. While enhancers of meiotic drive are known to exist, they are not expected to evolve in positions that are unlinked to a driver [51]. This surprising outcome suggests that the maize genome may be adapting to the presence of chromosomal drivers with unlinked alleles that reduce their negative fitness consequences. This would be similar, for example, to genetic modifiers that reduce disease severity in humans [52], or mutations that bypass the phenotype of otherwise lethal mutations in yeast [53].

Taken together, our data suggest that the major limit to the spread of Ab10, K10L2 and the B chromosome are the fitness defects associated with the drivers themselves at high copy numbers. The fact that modeling based on the known fitness defects associated with Ab10 (reduced seed number, reduced seed weight and mildly reduced pollen viability [17]) does not predict the observed low frequencies suggests that there are additional fitness defects that have not yet been accounted for [22]. One way to better assess the fitness defects would be to start with the few (near)-inbred lines that we

show are still segregating for the major chromosomal drivers ([S2 Table](#)). Further self crossing should make it possible to identify sibling inbreds that either do or do not carry the CDH, and measure a broader array of fitness and fertility variables, including, for instance, seed germination and survival from seedling to reproductive stages [[22](#)].

## Methods

### GBS sequencing controls

Our control GBS data were obtained from two different sources, Cornell and CD Genomics. For the Cornell dataset, plants known to be heterozygous for Ab10-I and Ab10-II (the variants known as Ab10-I-MMR or Ab10-II-MMR [[17](#)]) were self crossed to create populations segregating for either Ab10 structural variant. Ab10 was marked by an allele of the *colored1* gene (*R1*) which makes the kernels purple ([S4 Table](#)). There is an approximately 2% chance of recombination between Ab10 and *R1* [[54](#)]. We extracted genomic DNA from plants grown from purple seeds (likely Ab10 positive) and colorless seeds (likely N10 homozygotes) using a CTAB extraction protocol [[55](#)] (Ab10-I-MMR=41, Ab10-II-MMR=37, N10=16). Using these DNA samples, GBS libraries were prepared in accordance with [[49](#)] and sequenced on an Illumina HiSeq 2000 with 100bp single end reads by the Genomic Diversity Facility, Cornell University (this facility is no longer in operation).

For the CD genomics dataset, we grew plants from 49 Ab10 controls from 11 genetic backgrounds, 13 K10L2 controls from 2 genetic backgrounds, 18 B chromosome controls from 5 genetic backgrounds, and 18 no CDH controls from 7 genetic backgrounds ([S4 Table](#)). The collection of Ab10s included nine different previously collected Ab10 variants (Ab10-I-Jal, Ab10-I-MMR, Ab10-II-Tel, Ab10-II-MMR, Ab10-II-Sal, Ab10-III-Caq, Ab10-III-Hui, Ab10-III-Gua, and Ab10-III-Oax, see [S4 Table](#), [[17](#)]). We first verified that the controls were CDH positive or negative by extracting DNA using CTAB extraction [[55](#)] and performing PCR for *Kindr*, *Trkin*, or the B repeat ([S5 Table](#)). We then sent leaf tissue to CD Genomics (Shirley, NY) who extracted DNA using QIAgen DNeasy Plant Kits. They prepared GBS libraries as described in [[49](#)] with minimal modification. Basically this involved digesting DNA with ApeKI (New England Biolabs, Ipswich, MA), adding barcoded adapters, and sequencing the libraries on an Illumina NovaSeq6000 using a 150 × 2 paired-end sequencing protocol. For compatibility with the other single end sequencing data we only used read1.

After receiving the data we identified several lines that appeared to be misclassified based on scaled tag index k-means clustering (W23\_AB10-I.11.DC1, W23\_AB10-I.13.DC1, W23\_AB10-II.36.DC1, W23\_N10.14.DC1, NSL-2833\_B-Chrom.2.DC2, B542C\_L289\_B-Chrom.1.DC2). W23\_AB10-I.11.DC1, W23\_AB10-I.13.DC1, W23\_AB10-II.36.DC1, and W23\_N10.14.DC1 were likely recombinants, but this could not be verified by PCR as we no longer had the samples and were excluded from further analysis. NSL-2833\_B-Chrom.2.DC2 and B542C\_L289\_B-Chrom.1.DC2 were re-genotyped and reclassified as having no B chromosomes.

### Obtaining GBS Data

We obtained GBS sequence reads from the authors of three prior publications [[43–45](#)]. These data were generated following the protocol of [[49](#)] to produce 100bp single end reads. The data from [[44](#)] were in the format of demultiplexed qualified reads; we converted them to a format usable for TASSEL using custom R v4.3.1 code and barcode faker [[56](#)]. The data from each plant described in [[44](#)] was split into approximately 4 libraries as technical replicates, and these were summed during analysis (see below).

### K-means clustering of controls

We first established that it was possible to differentiate Ab10, K10L2, and N10 from each other as well as B chromosome presence/absence using GBS data. We began by mapping the full set of control GBS data to the B73-Ab10 v2 [[14](#)] genome with the B chromosome appended [[25](#)] and the CI66 inbred genome carrying the K10L2 haplotype [[14](#)] using

TASSEL v5.2.44 [56] and BWA v0.7.17 [57]. Using TASSEL v5.2.44 [56], we obtained the coordinates of each tag and the number of associated reads in each sample for both the B73-Ab10/B-Chromosome assembly and the CI66-K10L2 assembly (TagByTaxa Table). We converted the alignments to a bed file using samtools v0.1.20 [58], and bedtools v2.29.2 [59] to locate the position of each tag in the TagByTaxa table. For each assembly, we summed the tag counts for all technical replicates per biological individual for [44]. Unless otherwise noted all further steps were carried out using custom R v4.3.1 code. In order to normalize across libraries of varying size, we calculated reads per million for each tag in each individual sample. We calculated the minimum proportion of missing data for blank samples (where no genomic DNA was added; this represents sequencing background), and subtracted 0.001. We then removed any sample with more missing data than this cut off, as well as any tag with a BWA mapping quality of less than 20. We verified that all datasets were affected similarly by these filters and extracted all tags on each CDH. We then calculated the tag index in non overlapping 1 Mb bins across all CDHs ( $\sqrt{c} + d$ ), where  $c$  is the count of tags mapped to that bin and  $d$  is the sum of the read depth of all tags in that bin (S1A Fig). Then we visualized control samples with known CDH status in a min/max scaled heat map of the tag index. We found that the CDH positive and negative lines were visually very distinct (Fig 1A). We did not observe any visual distinction between our two sets of control data (the Cornell and CD genomics datasets), indicating that this method is robust to differences in sequencing. This is important as the experimental data set is pooled from multiple data sources.

We then established that we could correctly and automatically detect CDH presence or absence in our control data set. We chose to use an iterative k-means clustering method on the scaled tag index. The entire pipeline outlined below was performed using custom R v4.3.1 code unless otherwise stated. We selected only the CDH-specific portions, which were the regions that showed stark differences between CDH positive and negative controls (Fig 1A). For each CDH we had high and low copy number controls. For Ab10 and K10L2, the high copy number controls were homozygous plants with two copies of the CDH, and low copy number controls were heterozygous plants with one copy of the CDH. For the B chromosome, the copy number was unknown and they were divided into high and low copy number controls by visual comparison of the min/max scaled tag index heat maps (Fig 1A). We analyzed the high and low copy number controls separately in order to ensure that clustering was based on the distinction between presence and absence rather than copy number. First, we split our group of control samples into three randomly selected groups. On each subsample we performed k-means clustering ( $k=2$ ). If a cluster was composed of at least 80% CDH (Ab10, K10L2, B chromosome) or non CDH (N10, no B chromosome) samples it was assigned as such (this is the naming step in Fig 1B). The k-means cluster assignment was then compared to the true CDH status of that sample in order to determine if the k-means clustering assigned the sample correctly. This was repeated 100 times for each sample, where each iteration involved a different, randomly selected set of control individuals (Fig 1B). For the B chromosome low copy number model we didn't have adequate samples to break them into three subsamples so they were clustered as a single group. Using this method we were able to correctly identify the CDH status of all of the control samples 100% of the time, regardless of where the GBS data were acquired (either from Cornell or CD genomics).

### Use of K-means clustering on experimental samples

Having established that the method correctly identifies each CDH in control data, we then extended it to our experimental samples [43–45]. We identified the chromosome 10 CDHs and the B chromosomes in two separate workflows before finally estimating their copy number. The entire pipeline outlined below was performed using custom R v4.3.1 code unless otherwise stated.

The general approach was to select roughly equal numbers of the appropriate controls (positive and negative) for each CDH and then randomly add a small number of experimental samples. For Ab10 and K10L2, the number of experimental samples added was 25% of the number of controls, for the B chromosome, the number of experimental samples added was 10% of the number of controls. Then we performed k-means clustering (S1B Fig). If a cluster was composed of at



least 80% CDH (Ab10, K10L2, B chromosome) or non CDH (N10, no B chromosome) samples it was assigned as such. We verified that all control samples were correctly identified. If they were not, we repeated the k-means clustering until all controls were correctly identified. We then assigned all experimental samples to the class of their k-means cluster. We repeated this workflow until all experimental samples had undergone one round of k-means clustering. Then we repeated the entire process 125 times to obtain 125 independent calls for the CDH class per experimental sample. To make the final CDH class calls, we required that the experimental sample be called the same class 95% of the time. All other samples were labeled ambiguous.

Our Ab10 pipeline cannot distinguish K10L2 from N10, while our K10L2 pipeline cannot distinguish Ab10 from K10L2. Therefore we employed them one after the other. We ran the Ab10 pipeline first and identified 394 Ab10 positive samples. We then isolated the samples called as N10 and ran the K10L2 pipeline, identifying 310 K10L2 positive samples. We plotted all the Ab10 and K10L2 positive samples in single heat maps with ward.D clustering ([Fig 2B](#)).

The variability in B chromosome copy number in experimental samples sometimes caused our k-means clustering pipeline to fail (lines with many copies of the B chromosome sometimes formed their own cluster). Therefore we used a two-step process. First we extracted all high copy number experimental samples using high copy number controls. Then we took all samples not identified as B chromosome positive in the high copy number iteration and ran them through the same pipeline using the low copy number B chromosome controls. In this way we were able to extract all B chromosome positive samples without introducing unnecessary variation in the k-means clustering. We plotted all B chromosome positive samples in a single heat map with ward.D clustering ([Fig 2B](#)).

### Random forest modeling

We then attempted to differentiate the Ab10 types within our experimental classes. We first trained a random forest model on 70% of the Ab10 control data with known types [[17](#)]. We checked the random forest model's performance using the remaining 30% of the Ab10 control data. It correctly predicted type 100% of the time. We then applied the same random forest model to all of our experimental samples. We required that 65% of decision trees call the same Ab10 type; if less than 65% called the same type, the haplotype was classified as ambiguous ([S2A Fig](#)). We selected this confidence threshold as it preserved the visually apparent differences between types when plotted as a heat map ([Fig 1A](#)). However, only 11.9% of Ab10 samples were classable in this manner. To better explore Ab10 types we extracted the bins with the highest mean decreasing Gini in the random forest model, meaning the model suffered the most when these variables were excluded, and performed a principal coordinate analysis ([S2B Fig](#)).

### Estimating the copy number of CDHs

To estimate the copy number for each CDH we needed an estimate of what the tag index of a single copy gene was. To this end we lifted over annotations from the B73 v5 reference [[12](#)] onto both the Ab10 v2 and CI66 K10L2 assemblies [[14](#)] using liftoff v1.6.3 [[60](#)]. We then extracted all single copy core genes, and calculated their tag index in 1 Mb bins ([S1A Fig](#)). The 1 Mb bins were composed of 1 Mb of single copy core gene sequence and not genomic coordinates. Then we calculated the average tag index across all single copy core gene bins for each sample. We divided the average CDH specific tag index value by that sample's average single copy core gene tag index value.

### PCR verification of CDHs in inbred lines

The frequency of CDHs varies in open pollinated populations like landraces and teosinte and are rarely if ever fixed [[19,24,26](#)]. Thus while we may have scored one plant from a landrace as positive for a CDH, it is unlikely that the next plant we scored from the same population would have the CDH. However, we identified several maize inbred lines containing Ab10, K10L2, and B chromosomes ([S1 Table](#)). We ordered two inbred lines called as positive for each CDH from

the Germplasm Resources Information Network (GRIN), Ames IA. We extracted DNA using a CTAB extraction [55] and performed PCR for *Kindr*, *Trkin*, or the B repeat to verify the CDHs presence (S2 and S5 Tables).

## GWAS

We generated artificial reference genomes with Mo17 [13] chromosome 1–10 and the Ab10 v2 haplotype [14], K10L2 haplotype [14], or the B chromosome [25]. For the Mo17 Ab10 and K10L2 reference genomes we used samtools v1.18 [58] to truncate Mo17 chr10 at the beginning of the *colored1* gene, which traditionally defines the beginning of the Ab10 and K10L2 haplotypes [18]. We then isolated the CDHs beginning at the *colored1* gene using samtools v1.18 [58] and SeqKit v0.16.1 [61] and appended them to the modified Mo17 genomes. For the B chromosome we left all Mo17 chromosomes intact and added the B chromosome [25]. We modified the key for all samples such that all technical replicates from [44] were read into a single biological sample. We then used TASSEL v5.2.44 [56] and BWA v0.7.17 [57] to align reads from all samples to the Mo17 + CDH references. We filtered mapped tags to a mapping quality of 20 using samtools v1.18 [58] and called SNPs using TASSEL v5.2.44 [56]. We then extracted SNPs on chromosomes 1–10 and the relevant CDH using bcftools v1.15.1 [58]. We isolated only maize landraces using bcftools v1.15.1 [58] because we suspected genetic modifiers might be less frequent in inbred lines due to the low frequency of CDHs (S1 Table). Then we applied the following filters: Read depth >3 and <20, minor allele frequency >= 0.05, genotype quality >60, and a per sample missingness of 75% or less. We used BEAGLE v5.4 to impute missing data based on haplotypes found in our data [62]. We did not use a reference panel due to concerns about the maintenance of genetic modifiers for CDHs in inbred lines. Using PLINK v1.9 [63] we removed plants with more than 10% missing data. We were left with ~50,000 SNPs. We carried out a principal component analysis on the whole genome non-CDH SNPs to identify population structure in the data, and included the top 10 principal components in a genome wide association test as covariates using PLINK v1.9 [63]. For Ab10, K10L2, and B chromosome presence/absence we used a logistic regression and for B chromosome copy number we used a linear regression. We plotted the output using custom R v4.3.1 code.

GBS tags are just 64 bp [56]. We know that some regions of each CDH are homologous to chromosomes 1–10 [14,18,24]. It seemed possible that a GBS tag originating from a CDH could map to chromosomes 1–10 and create an erroneous association. We identified genes orthologous between the CDH and the Mo17 genome using OrthoFinder v2.5.5 [64] and removed any associated loci overlapping them using bedtools v2.31.0 [59]. Additionally we removed any SNP overlapping an annotated transposable element [13] using bedtools v2.31.0 [59]. Finally, we extracted 64 bp upstream and downstream of each SNP using Samtools v1.18 [58] and used BLAST v2.13.0 to compare the region to the CDH in the B73-Ab10 v2 [14], K10L2 [14], or B chromosome [25,65] references. We removed any SNP that had homology to a CDH of at least 62 bp with a percent identity of 80% or greater. The sequence surrounding Chr9 SNP 2 (29026173) had 83% percent identity to the Ab10 v1 reference [42], though there was no significant homology to the B73-Ab10 v2 reference [14]. We suspect this associated SNP is an artifact and excluded it from further analysis.

We repeated the above procedure using all the same criteria on SNPs on the CDH as a control for loci linked to the CDH (S4 Fig). We removed SNPs that did not occur in at least 75% of the samples, so any locus specific to the CDH should have been removed. Loci present in the inverted shared region of Ab10 could exist in individuals carrying Ab10 or N10 allowing them to pass the missing data filters. While K10L2 also has a shared region it is not inverted [14] and is known to recombine with N10 [24] so while many SNPs passed the filtering we would expect fewer of them to strongly associate with K10L2 presence. The B chromosome does not have a shared region thus most SNPs on it occur in less than 75% of samples and were removed by the missing data filter.

## GLM models on All CDHs

We obtained climatic data from WorldClim2 [47] and soil quality data from the FOA harmonized world soil database [48]. We chose to begin each model with environmental variables known to be associated with either maize or the CDHs: elevation, mean temperature of warmest quarter, precipitation of warmest quarter, temperature seasonality, precipitation

seasonality, average annual solar radiation, average annual wind, average annual water vapor pressure [47], soil nutrient availability, soil rooting conditions, soil oxygen availability to roots, soil excess salts, and soil toxicity [48,66]. For Ab10 we included B chromosome presence/absence (we could not include K10L2 because we cannot detect K10L2 in lines where Ab10 is present). For the B chromosome we included Ab10 and K10L2 presence/absence. We extracted environmental data for each collected sample using the raster package in R v4.3.1 [67]. For solar radiation, average annual wind, and average annual water pressure we used custom R v4.3.1 code to generate the average value for each location from monthly data [47]. We tested for collinearity between the environmental variables using custom R v4.3.1 code and found that elevation and soil nutrient availability had a greater than 70% correlation to other variables (S7 Fig). We excluded soil nutrient availability from all starting models due to its collinearity. We chose to include elevation in the starting models because it has been investigated previously [19,33]. We also included the first 10 principal components of population structure generated as part of the GWAS. For the CDH presence/absence we used binomial family models. For B chromosome copy number we log transformed the variable to make it normally distributed and used a gaussian family model. From the largest model we performed stepwise model simplification using custom R v4.3.1 code. We ensured all models were robust to variable order. We tested model fit using the DHARMA R package before and after simplification [68]. We determined model fit was acceptable in all cases. Our models represented a relatively low proportion of the total deviance, thus we chose to apply an alpha value of 0.01.

To assess the relative contributions of population structure, genetic loci, and the environment we selected all significant loci from the GWAS using custom R v4.3.1 code. Because some loci had more than one alternate allele, we began by coding these as factors to check if any of the minor alleles had a significant relationship to any CDH. We found that they did not and removed them, coding the remaining minor alleles additively. We then performed stepwise model simplification. We plotted the results using custom R v4.3.1 code. To partition the variance from the entire model into population structure, genetic loci, and environmental variables we first used custom R v4.3.1 code to remove all missing data to ensure the null model was identical between all runs. Then we ran individual models with only the remaining population structure, genetic loci, or environmental variables. We calculated the amount of deviance explained by each of these models and plotted them using custom R v4.3.1 code.

## Supporting information

**S1 Table. Accessions and individuals assayed, and lines that show both B chromosome and Ab10 or B chromosome and K10L2.**

(XLSX)

**S2 Table. Inbred lines that scored positive for CDHs, and PCR data from a subset.**

(XLSX)

**S3 Table. All CDH associated SNPs and genes they overlap.**

(XLSX)

**S4 Table. Accession information for control lines used for GBS.**

(XLSX)

**S5 Table. Primers used for genotyping.**

(XLSX)

**S1 Fig. Workflow diagrams. A.** Workflow diagram for the generation of the tag index for CDHs and single copy core genes. In brief, we first identified tags, which are identical 64 bp sequences. We then mapped those tags to the relevant reference genome. Then we calculated reads per million for each mapped tag and filtered to remove unreliable tags. To

calculate tag index, we isolated the CDH and divided it into 1 Mb bins. On each bin we calculated the tag index. **B.** Diagram of the workflow for detecting CDHs in experimental samples. Check indicates passing, x indicates failing.  
(PDF)

**S2 Fig. Identification of Ab10 type.** **A.** Results of a random forest model trained on control data and applied to all experimental Ab10 positive samples. Ab10 types classified by the random forest model are plotted separately and each group is ward.D clustered. The x axis shows individual samples, the y axis shows features of the Ab10 haplotype and the importance of each 1 Mb bin in determining Ab10 type in the random forest model (Mean Decrease Gini). The RF confidence value indicates the proportion of decision trees that are called the predominant Ab10 type. **B.** A PCA of all the Ab10 positive samples scaled tag index with controls and their type indicated. Control samples are shown in large shapes colored by their Ab10 class, experimental samples are shown in small shapes. Experimental samples that were confidently classed by the random forest model are shown in their respective color.  
(PDF)

**S3 Fig. Location of SNPs used for GWAS.** Numbers below the CDH name indicate the total number of SNPs.  
(PDF)

**S4 Fig. Manhattan plots of SNPs that passed through filtering but lie within or close to a CDH.** Numbers below the CDH names indicate the number of SNPs in the plot. Dotted grey line indicates a p value of  $5 \times 10^{-8}$ . Negatively associated loci are the result of alleles linked to the non-CDH homologs (the N10 shared regions).  
(PDF)

**S5 Fig. Relationship of CDHs to elevation.** Each point represents a sample either with or without a CDH and the elevation at which they were originally collected. The Ab10, K10L2, and the B Chr. data were analyzed using a generalized linear model with a binomial family and the following model: CDH presence/absence ~ all 10 whole genome SNP principal components + elevation of collection. For B chromosome copy number the same formula was used with a linear family. N.S. indicates not significant.  
(PDF)

**S6 Fig. Plots of simplified generalized linear models for each CDH including population structure and environmental variables (but not genetic modifiers).** Shape color and orientation indicate the direction of the relationship to the CDH. Shape size represents the effect size.  
(PDF)

**S7 Fig. Correlation matrix for selected environmental variables.**  
(PDF)

## Acknowledgments

We thank Doua Bensasson for guidance on modeling, Brad Nelms for help developing the CDH detection method, and Robert Unckless and Jonathan Gent for helpful comments on the manuscript. We also thank Cinta Romy, Edward Buckler, and the International Maize and Wheat Improvement Center (CIMMYT) for providing the raw GBS reads. We are grateful to the Georgia Advanced Computing Resource Center for providing technical expertise.

## Author contributions

**Conceptualization:** Meghan J Brady, R Kelly Dawe.

**Data curation:** Meghan J Brady.

**Formal analysis:** Meghan J Brady.



**Funding acquisition:** Meghan J Brady, R Kelly Dawe.

**Investigation:** Meghan J Brady.

**Methodology:** Meghan J Brady, R Kelly Dawe.

**Project administration:** R Kelly Dawe.

**Resources:** R Kelly Dawe.

**Software:** Meghan J Brady.

**Supervision:** R Kelly Dawe.

**Validation:** Meghan J Brady.

**Visualization:** Meghan J Brady, R Kelly Dawe.

**Writing – original draft:** Meghan J Brady, R Kelly Dawe.

**Writing – review & editing:** Meghan J Brady, R Kelly Dawe.

## References

1. Werren JH. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 2(Suppl 2):10863–70. <https://doi.org/10.1073/pnas.1102343108> PMID: [21690392](#)
2. Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution*. 2016;31:315–26.
3. Ågren JA, Clark AG. Selfish genetic elements. *PLoS Genet*. 2018;14(11):e1007700. <https://doi.org/10.1371/journal.pgen.1007700> PMID: [30439939](#)
4. Clark FE, Akera T. Unravelling the mystery of female meiotic drive: where we are. *Open Biol*. 2021;11(9):210074. <https://doi.org/10.1098/rsob.210074> PMID: [34465214](#)
5. Sandler L, Novitski E. Meiotic drive as an evolutionary force. *Am Nat*. 1957;91:105–10.
6. Akera T, Trimm E, Lampson MA. Molecular Strategies of Meiotic Cheating by Selfish Centromeres. *Cell*. 2019;178(5):1132–1144.e10. <https://doi.org/10.1016/j.cell.2019.07.001> PMID: [31402175](#)
7. Courret C, Wei X, Larracunte AM. New perspectives on the causes and consequences of male meiotic drive. *Curr Opin Genet Dev*. 2023;83:102111. <https://doi.org/10.1016/j.gde.2023.102111> PMID: [37704518](#)
8. Douglas RN, Birchler JA. B chromosomes. *Chromosome structure and aberrations*. New Delhi: Springer India. 2017;13–39.
9. Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure*. 2011;19(1):7–15. <https://doi.org/10.1016/j.str.2010.12.003> PMID: [21220111](#)
10. Saupe SJ, Johannesson H. On the Mechanistic Basis of Killer Meiotic Drive in Fungi. *Annu Rev Microbiol*. 2022;76:305–23. <https://doi.org/10.1146/annurev-micro-041320-113730> PMID: [36075094](#)
11. Bier E. Gene drives gaining speed. *Nat Rev Genet*. 2022;23(1):5–22. <https://doi.org/10.1038/s41576-021-00386-0> PMID: [34363067](#)
12. Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*. 2021;373(6555):655–62. <https://doi.org/10.1126/science.abg5289> PMID: [34353948](#)
13. Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, et al. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet*. 2023;55(7):1221–31. <https://doi.org/10.1038/s41588-023-01419-6> PMID: [37322109](#)
14. Brady MJ, Gupta A, Gent JI, Swentowsky KW, Unckless RL, Dawe RK. Conflicting Kinesin-14s in a single chromosomal drive haplotype. *Genetics*. 2025;230(3):iyaf091. <https://doi.org/10.1093/genetics/iyaf091> PMID: [40365704](#)
15. Dawe RK, Lowry EG, Gent JI, Stitzer MC, Swentowsky KW, Higgins DM. A kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*. 2018;173:839–850.e18.
16. Swentowsky KW, Gent JI, Lowry EG, Schubert V, Ran X, Tseng K-F, et al. Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev*. 2020;34(17–18):1239–51. <https://doi.org/10.1101/gad.340679.120> PMID: [32820038](#)
17. Higgins DM, Lowry EG, Kanizay LB, Becraft PW, Hall DW, Dawe RK. Fitness Costs and Variation in Transmission Distortion Associated with the Abnormal Chromosome 10 Meiotic Drive System in Maize. *Genetics*. 2018;208(1):297–305. <https://doi.org/10.1534/genetics.117.300060> PMID: [29122827](#)
18. Dawe RK. The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Res*. 2022;30(2–3):205–16. <https://doi.org/10.1007/s10577-022-09693-6> PMID: [35652970](#)

19. Kanizay LB, Pyhäjärvi T, Lowry EG, Hufford MB, Peterson DG, Ross-Ibarra J, et al. Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *Zea mays*. *Heredity* (Edinb). 2013;110(6):570–7. <https://doi.org/10.1038/hdy.2013.2> PMID: [23443059](#)
20. Yamakake Kato ÁT. Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana* (Schraeder) Kuntze) in relation to their origin and evolution. University of Massachusetts. 1975.
21. McClintock B, Yamakake T, Blumenschein A. Chromosome constitution of races of maize: its significance in the interpretation of relationships between races and varieties in the Americas. Chapingo, Edo. Mexico: Colegio de Pos graduados, Escuela Nacional de Agricultura. 1981.
22. Hall DW, Dawe RK. Modeling the Evolution of Female Meiotic Drive in Maize. *G3* (Bethesda). 2018;8(1):123–30. <https://doi.org/10.1534/g3.117.300073> PMID: [29122849](#)
23. Buckler ES 4th, Phelps-Durr TL, Buckler CS, Dawe RK, Doebley JF, Holtsford TP. Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics*. 1999;153(1):415–26. <https://doi.org/10.1093/genetics/153.1.415> PMID: [10471723](#)
24. Kanizay LB, Albert PS, Birchler JA, Dawe RK. Intragenomic conflict between the two major knob repeats of maize. *Genetics*. 2013;194(1):81–9. <https://doi.org/10.1534/genetics.112.148882> PMID: [23457233](#)
25. Blavet N, Yang H, Su H, Solanský P, Douglas RN, Karafiátová M, et al. Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. *Proc Natl Acad Sci U S A*. 2021;118(23):e2104254118. <https://doi.org/10.1073/pnas.2104254118> PMID: [34088847](#)
26. Birchler JA, Yang H. The supernumerary B chromosome of maize: drive and genomic conflict. *Open Biol*. 2021;11(11):210197. <https://doi.org/10.1098/rsob.210197> PMID: [34727722](#)
27. Carlson WR. Reversal of preferential fertilization. *Maize Genet Coop News Lett*. 1999;73:39.
28. Rosato M, Chiavarino A, Naranjo C, Hernandez J, Poggio L. Genome size and numerical polymorphism for the B chromosome in races of maize (*Zea mays* ssp. *mays*, Poaceae). *Am J Bot*. 1998;85(2):168. PMID: [21684902](#)
29. Randolph LF. Genetic Characteristics of the B Chromosomes in Maize. *Genetics*. 1941;26(6):608–31. <https://doi.org/10.1093/genetics/26.6.608> PMID: [17247025](#)
30. Price TAR, Windbichler N, Unckless RL, Sutter A, Runge JN, Ross PA. Resistance to natural and synthetic gene drive systems. *J Evol Biol*. 2020;33:1345–60.
31. Porter HL, Rayburn AL. B-chromosome and C-band heterochromatin variation in Arizona maize populations adapted to different altitudes. *Genome*. 1990;33:659–62.
32. Poggio L, Rosato M, Chiavarino AM, Naranjo CA. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). *Ann Bot*. 1998;82:107–15.
33. Lia VV, Confalonieri VA, Poggio LB. B chromosome polymorphism maize landraces: adaptive vs. demographic hypothesis clinal variation. *Genetics*. 2007;177:895–904.
34. Bilinski P, Albert PS, Berg JJ, Birchler JA, Grote MN, Lorant A, et al. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genet*. 2018;14(5):e1007162. <https://doi.org/10.1371/journal.pgen.1007162> PMID: [29746459](#)
35. Rayburn AL, Auger JA. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theor Appl Genet*. 1990;79(4):470–4. <https://doi.org/10.1007/BF00226155> PMID: [24226450](#)
36. Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte LE. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*. 2013;199:264–76.
37. Lyttle TW. Experimental Population Genetics of Meiotic Drive Systems II. Accumulation of Genetic Modifiers of Segregation Distorter (SD) in Laboratory Populations. *Genetics*. 1979;91(2):339–57. <https://doi.org/10.1093/genetics/91.2.339> PMID: [17248888](#)
38. Cazemajor M, Landré C, Montchamp-Moreau C. The sex-ratio trait in *Drosophila simulans*: genetic analysis of distortion and suppression. *Genetics*. 1997;147(2):635–42. <https://doi.org/10.1093/genetics/147.2.635> PMID: [9335600](#)
39. Courret C, Chang C-H, Wei KH-C, Montchamp-Moreau C, Larracuente AM. Meiotic drive mechanisms: lessons from *Drosophila*. *Proc Biol Sci*. 2019;286(1913):20191430. <https://doi.org/10.1098/rspb.2019.1430> PMID: [31640520](#)
40. Rhoades MM, Dempsey E. On the mechanism of chromatin loss induced by the B chromosome of maize. *Genetics*. 1972;71:73–96.
41. Rhoades MM, Dempsey E. Structural heterogeneity of chromosome 10 in races of maize and teosinte. In: Freeling M. *Plant Genetics*. New York: Alan R. Liss. 1985;1–18.
42. Liu J, Seetharam AS, Chougule K, Ou S, Swentowsky KW, Gent JI, et al. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol*. 2020;21(1):121. <https://doi.org/10.1186/s13059-020-02029-9> PMID: [32434565](#)
43. Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science*. 2017;357:512–5.
44. Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet*. 2017;49(3):476–80. <https://doi.org/10.1038/ng.3784> PMID: [28166212](#)
45. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. 2013;14(6):R55. <https://doi.org/10.1186/gb-2013-14-6-r55> PMID: [23759205](#)

46. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*. 2009;4(12):e8451. <https://doi.org/10.1371/journal.pone.0008451> PMID: 20041112
47. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;37:4302–15.
48. FAO, IIASA. Harmonized World Soil Database Version 2.0. Rome, Italy; Laxenburg, Austria: FAO; International Institute for Applied Systems Analysis (IIASA). 2023.
49. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
50. Stitzer MC, Khaipho-Burch MB, Hudson AI, Song B, Valdez-Franco JA, Ramstein G. Transposable element abundance subtly contributes to lower fitness in maize. *Evolutionary Biology*. 2023.
51. Crow JF. Why is Mendelian segregation so exact?. *Bioessays*. 1991;13(6):305–12. <https://doi.org/10.1002/bies.950130609> PMID: 1909864
52. Riordan JD, Nadeau JH. From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health. *Am J Hum Genet*. 2017;101(2):177–91. <https://doi.org/10.1016/j.ajhg.2017.06.004> PMID: 28777930
53. van Leeuwen J, Pons C, Tan G, Wang ZY, Hou J, Weile J, et al. Systematic analysis of bypass suppression of essential genes. *Mol Syst Biol*. 2020;16(9):e9828. <https://doi.org/10.1525/msb.20209828> PMID: 32939983
54. Rhoades MM. Preferential segregation in maize. *Genetics*. 1942;27:395–407.
55. Doyle JJ, Doyle JL. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull*. 1987;19:5.
56. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 2014;9(2):e90346. <https://doi.org/10.1371/journal.pone.0090346> PMID: 24587335
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
58. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008> PMID: 33590861
59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
60. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021;37(12):1639–43. <https://doi.org/10.1093/bioinformatics/btaa1016> PMID: 33320174
61. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta*. 2024;3(3):e191. <https://doi.org/10.1002/imt2.191> PMID: 38898985
62. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet*. 2021;108(10):1880–90. <https://doi.org/10.1016/j.ajhg.2021.08.005> PMID: 34478634
63. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795> PMID: 17701901
64. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y> PMID: 31727128
65. Liu Q, Liu Y, Yi C, Gao Z, Zhang Z, Zhu C, et al. Genome assembly of the maize B chromosome provides insight into its epigenetic characteristics and effects on the host genome. *Genome Biol*. 2025;26(1):47. <https://doi.org/10.1186/s13059-025-03517-6> PMID: 40050975
66. Ali S, Umair M, Makanda TA, Shi S, Hussain SA, Ni J. Modeling current and future potential land distribution dynamics of wheat, rice, and maize under climate change scenarios using MaxEnt. *Land (Basel)*. 2024;13:1156.
67. Hijmans R. Raster: Geographic Data Analysis and Modeling. 2025. <https://github.com/rsbath/raster>
68. Hartig F. DHARMA: Residual diagnostics for hierarchical (multi-level/ mixed) regression models. <https://CRAN.R-project.org/package=DHARMA>. 2024.