PLANT SCIENCES

Increased maize chromosome number by engineered chromosome fission

Yibing Zeng[1], Mingyu Wang[2], Jonathan I. Gent[3], R. Kelly Dawe[1,2,3]*

Activation of synthetic centromeres on chromosome 4 in maize leads to its breakage and formation of trisomic fragments called neochromosomes. A limitation of neochromosomes is their low and unpredictable transmission rates due to trisomy. Here, we report that selecting for dicentric recombinants through male crosses uncovers stabilized chromosome 4 fission events, which split it into 4a-4b complementary chromosome pairs, where 4a carries a native centromere and 4b carries a synthetic one. The cells rapidly stabilized chromosome ends by de novo telomere formation, and the new centromeres spread among genes without altering their expression. When both 4a and 4b chromosomes were made homozygous, they segregated through meiosis indistinguishably from wild type and gave rise to healthy plants with normal seed set, indicating that the synthetic centromere was fully functional. This work leverages synthetic centromeres to engineer chromosome fission, raising the diploid chromosome number of maize from 20 to 22.

INTRODUCTION

Karyotype engineering is a form of genome editing that involves altering large segments of a genome, joining chromosomes, or potentially increasing the number of chromosomes (*1*). Creating a new chromosome is particularly challenging because it requires engineering a new centromere. Centromeres are composed of DNA and a large number of kinetochore proteins that interact with DNA to facilitate microtubule binding, sense proper chromosome alignment, and initiate anaphase (*2*). However, in most cases, the interaction between centromeric DNA and kinetochore proteins is not sequence specific. Virtually any DNA sequence can serve as the structural foundation for a kinetochore (*3*, *4*), although sequences that are repetitive and gene-free are most common in nature (*5*, *6*). Once a centromere position is established, the replication process is epigenetic: Specialized deposition proteins mediate the recruitment of additional kinetochore proteins to the same site over subsequent cell cycles (*7*, *8*).

The lack of sequence specificity means that engineering a new centromere requires a method to recruit inner centromere proteins to a designated centromere platform. One method, called protein tethering, involves fusing centromere proteins to a DNA binding protein that recognizes sequence motifs in the designated centromere platform. Most or all of the known functions of the centromere can be recruited to a synthetic repeat structure by tethering the key centromere protein CENP-A/CENH3, chaperones that recruit CENH3, or several other proteins that interact closely with CENH3 (*9*–*14*). Synthetic centromeres formed through protein tethering have been shown to bind microtubules and confer mitotic chromosome segregation in multiple species (*10*, *14*–*20*). We recently demonstrated that synthetic centromeres can be induced in maize and transmitted through meiosis (*21*). We tethered a LexA-CENH3 fusion protein to an array of LexO binding sites on the long arm of maize chromosome 4 and showed that new centromeres cause the formation of dicentric chromosomes and chromosome breakage. Broken sections of chromosome 4 were recovered as trisomic neochromosomes (4b chromosomes), harboring new centromeres at the LexO array. The neochromosomes were propagated for several generations, but the transmission frequencies were low because of meiotic errors and gene dosage effects that accompany trisomic chromosomes.

Here, using synthetic centromere technology, we describe the fission and recovery of two chromosomes from a single ancestral chromosome. Starting with a trisomic 4b line, we demonstrate that crossing over between two centromere locations leads to dicentric derivatives that undergo breakage-fusion-bridge cycling, followed by de novo telomere formation and stabilization of a new 4a-4b karyotype with overlapping ends. We show that the expression of 22 genes embedded with the 4b centromere is not significantly different from the progenitor line and that DNA methylation is unchanged in the intergenic spaces, except for slight reductions in the CHG context immediately beneath CENH3. We further show that the synthetic centromere supports meiotic chromosome transmission that is indistinguishable from wild type.

RESULTS

Molecular structure of the arrayed binding site 4 binding platform

Our system for engineering synthetic centromeres is based on a binding platform called arrayed binding sites 4 (ABS4) that contains thousands of LexO binding sites on the long arm of chromosome 4. ABS4 was created by mixing polymerase chain reaction (PCR) products containing a repeating 157–base pair (bp) monomer with a marker plasmid (pAHC25) and transforming the mixture by biolistic transformation into the maize Hi-II line (which is related to the A188 inbred) (*22*, *23*). In a previous study (*21*), we used Illumina sequencing to estimate the copy number of the ABS monomer and identified two insertion locations on chromosome 4. Here, we used long-read sequencing to assemble a genome containing ABS4 so that we could more accurately interpret the epigenetic changes associated with newly formed centromeres.

The assembly shows that the two insertions are ~1.45 Mb from each other. The first insertion is only ~7 kb, while the second insertion is ~583 kb. The larger insertion contains 467 kb of the ABS

[1]Department of Genetics, University of Georgia, Athens, GA 30602, USA. [2]Institute of Bioinformatics, University of Georgia, Athens, GA 30602 USA. [3]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA.
*Corresponding author. Email: kdawe@uga.edu

monomer interspersed with 117 kb of fragmented copies of the pAHC25 plasmid (Fig. 1, A and B), although we do not have a complete picture of the internal structure because there remains an assembly gap in this region. Within the assembled sequence, there are approximately 2800 copies of LexO. The transformation process also created a rearrangement of the native genome. A comparison of the ABS4 assembly to the A188 reference assembly (*24*) indicates that a portion (~360 kb) of the genome next to the first insertion was inverted and inserted downstream of the second insertion (Fig. 1B).

Selection for 4b neochromosomes

Our method for inducing centromeres at ABS4 involves crossing it to a line expressing a LexA-CENH3 fusion protein (*21*). When LexA-CENH3 binds to the LexO binding sites in ABS4, it recruits native CENH3 and sufficient overlying kinetochore proteins to form a second active centromere on chromosome 4. Dicentric chromosomes are unstable because the two centromeres frequently move to opposite poles in mitosis, creating a bridge during anaphase that is broken. The break can be repaired either by de novo telomere formation or by fusing with the broken end from the other chromatid in the next cell cycle. If the break is repaired by fusing with another broken chromatid, then a dicentric chromosome is reformed, and the process can repeat itself in what McClintock referred to as the breakage-fusion-bridge (BFB) cycle (*25*). McClintock demonstrated that BFB can be maintained for multiple cell cycles before telomeres ultimately form (what she called "healing") and halt the process (*26*, *27*). The mechanism of de novo telomere formation is not well understood, although at least in vitro the maize telomerase enzyme can extend telomere repeats on a variety of DNA templates (*28*).

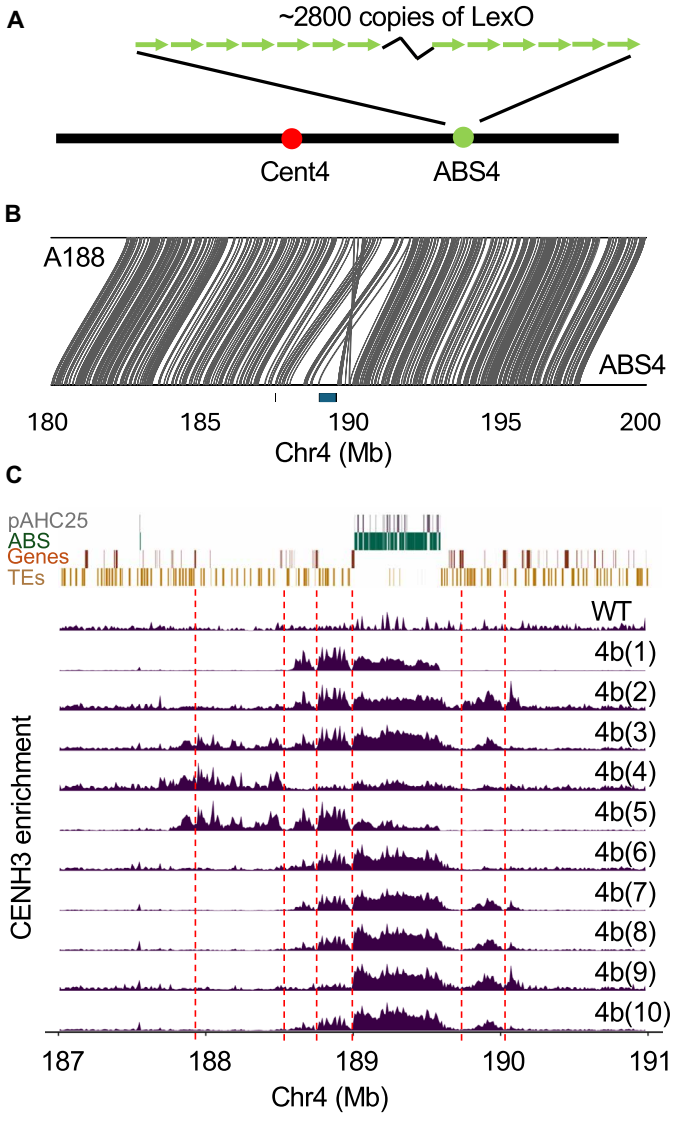

Fig. 1. Newly formed centromeres over ABS4. (**A**) Schematic of chromosome 4 showing the locations of normal centromere 4 (Cent4) and ABS4. (**B**) Assembled region of the ABS4 genome and comparison with the progenitor chromosome from the inbred A188. (**C**) Annotation features in the ABS4 region and CUT&Tag data from 10 neochromosomes. Red dashed lines show the locations of dips in CENH3 enrichment associated with genes. WT, wild type.

When broken fragments formed during ABS/LexA-CENH3–activated BFB are healed by telomere formation, the outcome is a new chromosome (neochromosome) with an engineered centromere at ABS4. Our objective was to recover newly formed neochromosomes in progeny and study their behavior in subsequent generations (*21*). As described previously, we designed a screen to pass neochromosomes through the female gametophyte, along with a complete chromosome 4 (Fig. 2A). Instead of tracking ABS4 directly, we used alleles of a seed pigment gene called *colorless2* (*c2*) that is linked to ABS4 by about 15 cM. The female parent was heterozygous for ABS4 and the dominant *C2* allele that pigments seeds (to a purple color), and a second chromosome containing the *c2-gfp* allele that makes the seeds yellow (not pigmented) and fluorescent under blue light. Plants of this genotype (ABS4 *C2*/*c2-gfp*) and carrying LexA-CENH3 were crossed as females to a *c2*/*c2* line that has colorless kernels. About 1% of the kernels from this cross are both pigmented and fluorescent, and often contain newly formed neochromosomes. Because we chose kernels that had received both *c2* alleles from the female parent, this screen recovers neochromosomes as partially trisomic lines (Fig. 2A). We originally described four neochromosomes that we named Neo4L-1, Neo4L-2, Neo4L-3, and Neo4L-4 (*21*). To simplify the nomenclature and to accommodate new results, we have renamed the original four as 4b(1), 4b(2), 4b(3), and 4b(4). Using kernels from the same genetic screen, we have now identified and characterized six additional neochromosomes, 4b(5), 4b(6), 4b(7), 4b(8), 4b(9), and 4b(10). In all cases, LexA-CENH3 was only used to activate the ABS4 centromeres; after the neochromosomes were identified, we chose individuals for further crossing that lacked LexA-CENH3 (which is unlinked to ABS4).

We carried out whole-genome Illumina sequencing to interpret the structure of the neochromosomes (Fig. 3, A to C, and fig. S1). All 10 appeared to contain the entirety of the original sequence to the right of ABS4 (that is, between ABS4 and the telomere of 4L). However, because of the effects of BFB and the random nature of chromosome breakage, each chromosome differs with respect to the sequence to the left of ABS4. The shortest is 4b(6), which appears to have broken right next to ABS4, leaving very little of a short arm on the neochromosome. The longest is 4b(10), which contains most of the sequence between centromere 4 (Cent4) and ABS4. Two of the neochromosomes contain evidence of duplications [4b(2) and 4b(10)] that are likely a result of the BFB cycle that preceded the formation of the stable neochromosomes (fig. S1).

We also carried out CENH3 CUT&Tag to interpret the location and size of the newly formed centromeres. In most cases, CENH3

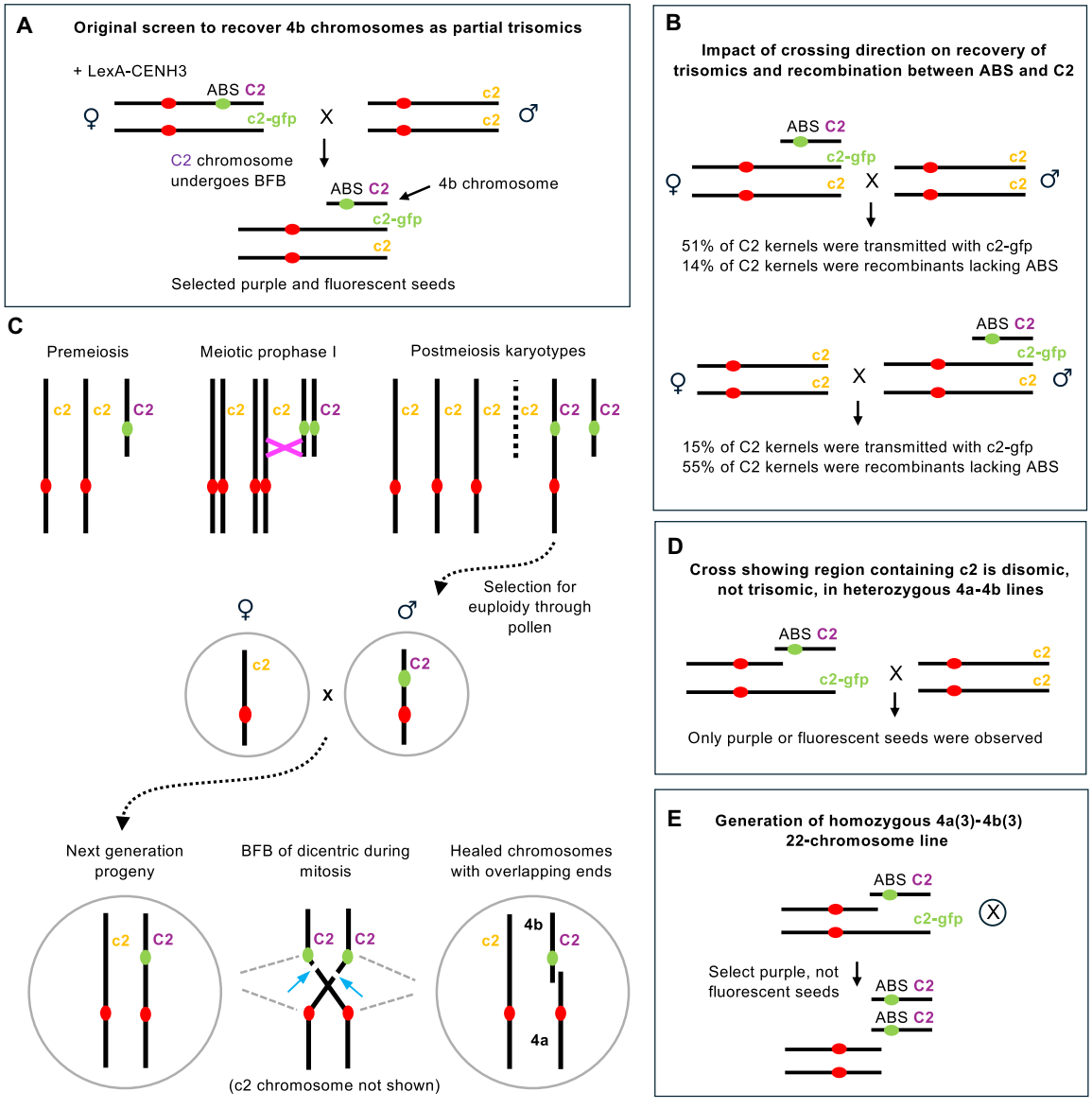

Fig. 2. Genetic explanations. (**A**) Crossing scheme used to identify trisomic 4b chromosomes. (**B**) How female and male crosses of partially trisomic 4b lines differ in the likelihood of passing on the trisomic condition (measured by how frequently the ABS *C2* chromosome is transmitted with a chromosome carrying *c2-gfp*) and apparent recombination between *C2* and ABS. We believe that recombination is elevated in the male cross because euploid chromosomes carrying *C2* are more likely to be transmitted. (**C**) Model for the formation of 4a chromosomes from partially trisomic lines. Trisomic chromosomes pair irregularly but undergo normal recombination. Recombination between normal chromosome 4 and 4b (magenta X) can create a dicentric chromosome that is transmissible through pollen [as well as a nontransmissible acentric fragment (dotted line)]. In the next generation, the dicentric chromosome will undergo BFB, and two newly broken 4a and 4b chromosomes, often with overlapping ends, can result. The BFB panel shows mitotic anaphase, where thin gray dashed lines indicate spindle microtubules and light blue arrows indicate the positions of new breaks. It is also possible for the breaks to occur during the second pollen mitosis (not shown) or as a direct outcome of meiosis (fig. S4). (**D**) Demonstration that a 4a chromosome had formed, making the region with the *c2* locus disomic, not trisomic. (**E**) Self-cross that led to the recovery of the homozygous 22-chromosome line.

was localized over the ABS4 insertion as well as flanking sequences to the left and right of ABS4, occupying regions of approximately 1.5 to 2.3 Mb, similar in size to native centromeres (*29*–*31*). In flanking genomic sequences, CENH3 is enriched in intergenic spaces and dips to low levels directly over genes, similar to cases where genes have been observed in native centromeres (Fig. 1C) (*32*–*34*). There were also smaller, unexpected peaks of CENH3 several megabases away from ABS4 in 4b(1), 4b(5), and 4b(6) (fig. S2). These peaks may represent cases of spontaneous accumulation of CENH3 that sometimes occur on acentric fragments (*35*), or structural rearrangements of the 4b chromosomes such as deletions or duplications that bring these apparently distal sequences on the reference in close proximity to the ABS4 array on the neochromosome (we have not assembled the genomes of these neochromosomes).

Selection for 4a chromosomes

The biology of plant reproduction has a major effect on the inheritance of aneuploid gametes. Both the female and male gametophytes

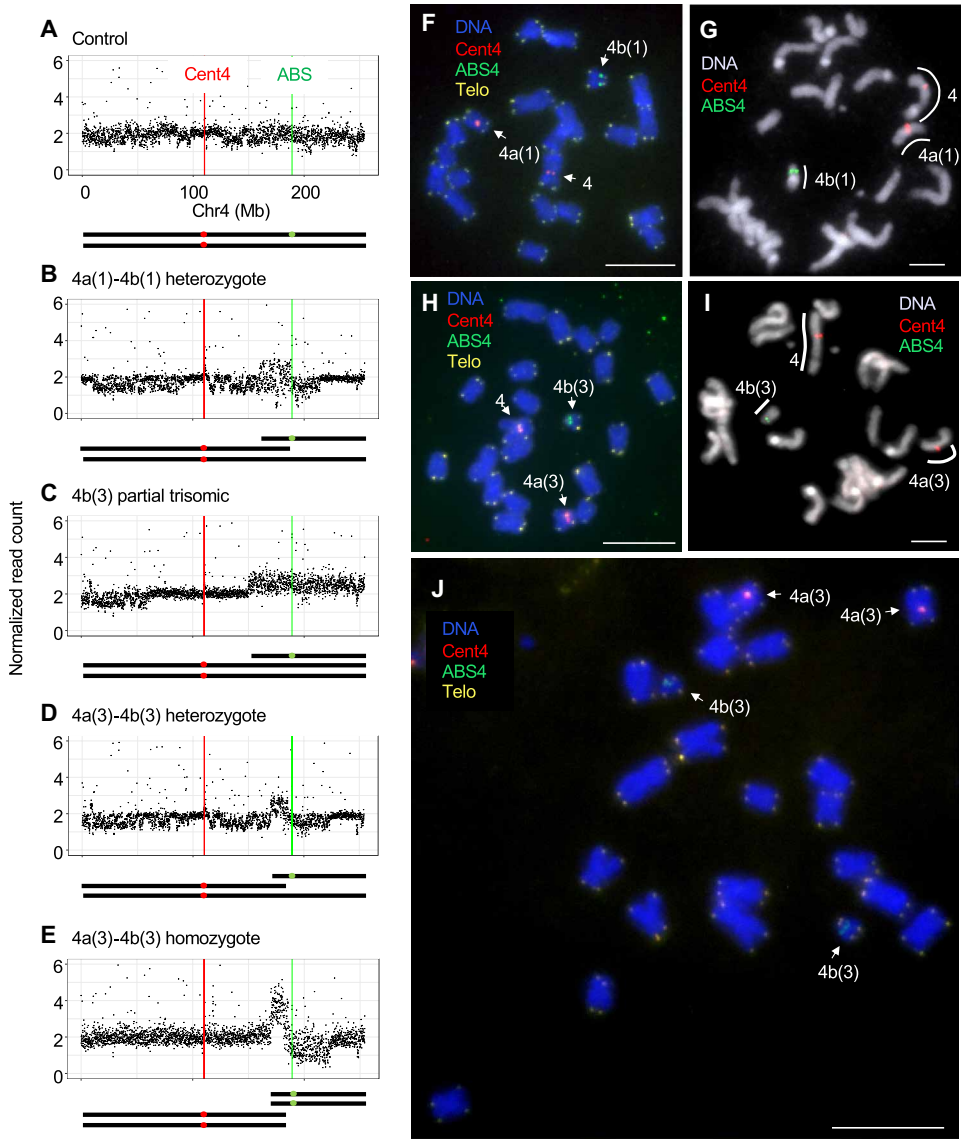

Fig. 3. Karyotypes of 4a-4b pairs. (**A** to **E**) Illumina sequencing of plants carrying 4a(1)-4b(1) and 4a(3)-4b(3) pairs. The *y* axis indicates the relative ploidy, expressed as normalized read count on 100-kb intervals. The control is heterozygous for the karyotypically normal ABS4 (related to A188 inbred) chromosome and the *c2* tester chromosome (related to the W22 inbred). Sequence reads were aligned to the W22 reference. The 4a-4b lines have been backcrossed to the *c2* tester to various degrees and show some areas of homozygosity, evidenced by narrower distributions of read counts. Panel (B) shows the 4b(1) chromosome, as it was originally found, as a heterozygous 4a(1)-4b(1) pair with a single normal chromosome 4 (not a partial trisomic), although we did not know that when it was first reported (*21*). On (E), the homozygous 4a(3)-4b(3) line, the region to the right of ABS4 (diffuse distribution), is homozygous for A188, while closer to the telomere (narrower distribution) it is heterozygous for W22. When the same sequence data were aligned to a third reference (Mo17), the pattern was uniformly diffuse, confirming that there is no ploidy difference (fig. S7A). Cartoon representations of the karyotypes are shown below each genotype. (**F**) FISH image of a 4a(1) 4b(1) heterozygote in late prophase. (**G**) FISH image of a 4a(1)-4b(1) heterozygote in early prophase. (**H**) FISH image of a 4a(3)-4b(3) heterozygote in late prophase. (**I**) FISH image of a 4a(3)-4b(3) heterozygote in early prophase. (**J**) FISH image of a 4a(3)-4b(3) homozygote in late prophase. Native Cent4 was detected with a probe to a pericentromeric repeat called Cent4 (*67*). The Telo label within panels indicates telomeres. Scale bars, 10 μm.

(pollen) are multicellular haploid structures that express genes that support their function. Because pollen tubes compete with each other to reach the egg cell, normal euploid pollen regularly outcompetes aneuploid pollen with altered gene dosage (*36*). In contrast, female gametophytes do not compete with each other and are more tolerant of altered gene dosage. The effect is that extra chromosomes are transmitted at higher frequencies through the female than the male (*37*).

Most of the neochromosomes were recovered as trisomics carrying three alleles, *C2*, *c2-gfp*, and *c2*, where *C2* was on the neochromosome. When plants of this genotype are crossed as a female, we observed that ~20 to 30% of the kernels carried *C2*, and when they were crossed as a male, we observed that about ~8 to 12% of the kernels carried *C2* (table S1). We initially thought that most of the *C2* kernels from both female and male crosses were trisomic (*21*).

Whether the progeny are trisomic can be estimated by the frequency of *C2* kernels that carry *c2-gfp* (both pigmented and fluorescent). Assuming random assortment, about half of the *C2*-carrying trisomic individuals will have *c2-gfp* and half will have *c2*. We observed this to be true in female crosses, where, on average, 51% of the progeny were fluorescent. However, this was not true in male crosses, where only 15% carried *c2-gfp* (Fig. 2B and table S1). These results suggested that most of the *C2* progeny from male crosses were not trisomic and likely did not have the original neochromosome. To test whether recombination on the neochromosome might be higher than expected in male crosses, we designed PCR primers over a 26-bp insertion/deletion polymorphism that differentiates the *C2* from *c2* alleles used in our crosses. We found that recombination between *C2* and ABS4 was 14% in female crosses and 55% in male crosses (Fig. 2B and table S2, $P < 0.01$ based on a logistic regression test). It is unlikely that recombination frequencies are elevated in males by threefold genome wide (*38*). Rather, we believe that recombination rates are similar in males and females, but that recombinants with *C2* but not ABS4 are selectively recovered because they are transmitted in euploid pollen that outcompetes aneuploid pollen. In other words, by choosing *C2* kernels, we biased the sample toward recombinants (Fig. 2C).

The realization that trisomic neochromosome progeny is rarely recovered from male crosses led us to reinterpret our 4b(1) data. Unlike all other neochromosomes, which were identified through female crosses, 4b(1) was found fortuitously in the progeny of a male cross (*21*). Our Illumina sequencing of this line seemed to show that the region to the right of ABS4 was disomic, not trisomic, as was the case for all others (Fig. 3B). We further noted that the transmission of 4b(1) was consistently higher than what we observed in other lineages, exceeding 40% through female crosses (table S1). We then analyzed the chromosomes in mitotic prophase cells from root tips and noted that Cent4 was present on two clearly different sized chromosomes (Fig. 3, F and G). This size difference is apparent in the kernels from the first generation. As confirmation, we crossed 4b(1) (carrying *C2*) to *c2-gfp* and crossed the resulting *C2*/*c2-gfp* heterozygote to a *c2*/*c2* line (Fig. 2D). The progeny of these crosses segregated for only pigmented and fluorescent kernels (*C2*/*c2* and *c2-gfp*/*c2*), with neither nonpigmented and nonfluorescent kernels nor pigmented and fluorescent kernels (*c2*/*c2* or *C2*/*c2-gfp*), demonstrating that the region to the right of ABS4 is disomic. If 4b(1) had been trisomic, then the expectation would be the segregation of all combinations of the three alleles. These data indicate that two broken chromosomes were recovered together in the original isolate, both 4b(1) and a truncated form of chromosome 4 that we will refer to as 4a(1).

The discovery of the 4a(1)-4b(1) pair led us to reassess the pedigree and transmission data for 4b(3). In the second generation, we propagated 4b(3) from a male cross. When one of the seeds from the male cross was planted and crossed again, the transmission was unusually high, showing 47% transmission of *C2* through the female and 35% transmission of *C2* through the male (table S1, crosses involving KD4310-2). Progeny from this lineage continued to show female *C2* transmission in excess of 40%, suggesting that the 4b(3) chromosome was segregating in a disomic manner and that a 4a(3) chromosome may be present. Analysis of mitotic prophase chromosomes demonstrated that this was the case (Fig. 3, H and I). We had Illumina sequenced the first generation plant, and at this point, the line was trisomic, and the break point of the 4b chromosome was at position ~150 Mb on the ABS4 reference genome (Fig. 3C). After crossing as a male and the progeny propagated, there were two new break points: the break point on the 4a chromosome at position ~182 Mb and the break point on a now-shorter 4b chromosome at position ~170 Mb (Fig. 3D). On all four broken chromosomes [4a(1), 4b(1), 4a(3), and 4b(3)], telomeres were visible at the newly formed ends by fluorescence in situ hybridization (FISH) (Fig. 3, F, H, and J).

These data suggest that when a trisomic 4b line is crossed as a male, a subset of the 4b chromosomes has undergone recombination on the short arm (between the break point and ABS4) to reconstitute a dicentric chromosome (Fig. 2C). When the newly formed dicentric chromosomes begin mitotic divisions, they are expected to undergo BFB. Supporting this view is the fact that about 8% of the *C2* kernels from male crosses have nonpigmented sectors that are suggestive of BFB [see table S1; though we initially thought sectors reflected centromere instability (*21*)]. As a test of this hypothesis, we carried out FISH on a sibling of the plant that gave rise to 4b(3). The data show that this lineage contains a different 4b chromosome (with a much shorter arm, appearing to be telocentric) and a 4a chromosome [see fig. S3A and table S1 for the 4b(3) lineage]. We also grew eight more kernels from the original male cross and analyzed root tip chromosomes by FISH. Only one seedling contained ABS4, and in this case, Cent4 and ABS4 were on the same chromosome and nearly juxtaposed, demonstrating that recombination had occurred on the short arm and that BFB had shortened the distance between Cent4 and ABS4 (fig. S3, B and C).

Together, the genetic data suggest that most of the *C2* kernels from male crosses of trisomic 4b lines are derived from euploid recombinants. Many have undergone recombination on the short arm of 4b and are transmitted as dicentrics that undergo BFB and form 4a-4b chromosome pairs (Fig. 2C). Under this model, breakage events occur during mitosis to form a 4a chromosome and a new 4b chromosome with overlapping ends. The breaks could occur in the gametophyte (for instance, the second pollen mitosis) or during the growth of the plant (as illustrated in Fig. 2C). An alternative mechanism is that breakage to form a 4a chromosome occurs in meiosis II (fig. S4). While breakage during meiosis is likely to occur, this mechanism preserves the structure of the original 4b chromosome and would not easily explain the origin of the 4a(3)-4b(3) pair. Under both models, breakage in mitosis or breakage in meiosis, telomere formation must occur to stabilize the ends of the breaks.

Phenotypes of the 22-chromosome lines

To obtain pure breeding 22-chromosome lines, we self-crossed plants heterozygous for 4a-4b pairs and normal *c2-gfp* chromosomes (Fig. 2E). From 4a(1)-4b(1) lines, we were unable to obtain 4a(1)-4b(1) homozygotes, suggesting that there is a mutation(s) and/or deletion(s) in the 4a(1) or 4b(1) chromosome that makes the homozygote inviable. In contrast, 4a(3)-4b(3) homozygotes were readily obtained from self-crosses [homozygote, in this context, refers only to the centromeres and telomeres that define the 4a(3) and 4b(3) chromosomes]. As expected, 22 chromosomes were visible in root tips (Fig. 3J). The resulting plants appeared normal, although they varied in vigor and stature, as is typical of self-crossed lines from mixed genetic backgrounds.

To assess the accuracy of meiosis, we analyzed meiotic prophase stages in three plants homozygous for the 4a(3)-4b(3) pair. At pachytene, where homologous chromosomes synapse and recombine, we observed 237 cells with normal tight pairing in the centromere regions (that is, the Cent4 and ABS FISH signals appeared as paired

dots; we did not attempt to trace the entire length of each chromosome). At diakinesis and diplotene, the chromosomes condense and shorten, making it possible to differentiate the entirety of the two chromosomes. All 87 cells showed the homologous 4a(3) and homologous 4b(3) chromosomes paired separately and associated by what appeared to be chiasmata (Fig. 4A and table S3). We observed very few cells in the later stages of meiosis but noted no obvious errors. As a second measure of the accuracy of meiosis, we collected fresh pollen and stained the grains with a vital stain. The data indicate that pollen from homozygous 4a(3)-4b(3) plants was morphologically indistinguishable from control plants (Fig. 4B and fig. S5). Together, the data suggest that plants homozygous for the 4a(3)-4b(3) pair can proceed through meiosis in a normal manner.

We obtained seven full ears from self- or sib-crossed homozygous 4a(3)-4b(3) lines (table S4 and fig. S6). We also made both female and male crosses to a line that is homozygous for both the recessive *c2* allele and an allele of the *R1* gene (*R1-scm2*) that promotes strong anthocyanin formation in the scutellum (a visible part of the seed

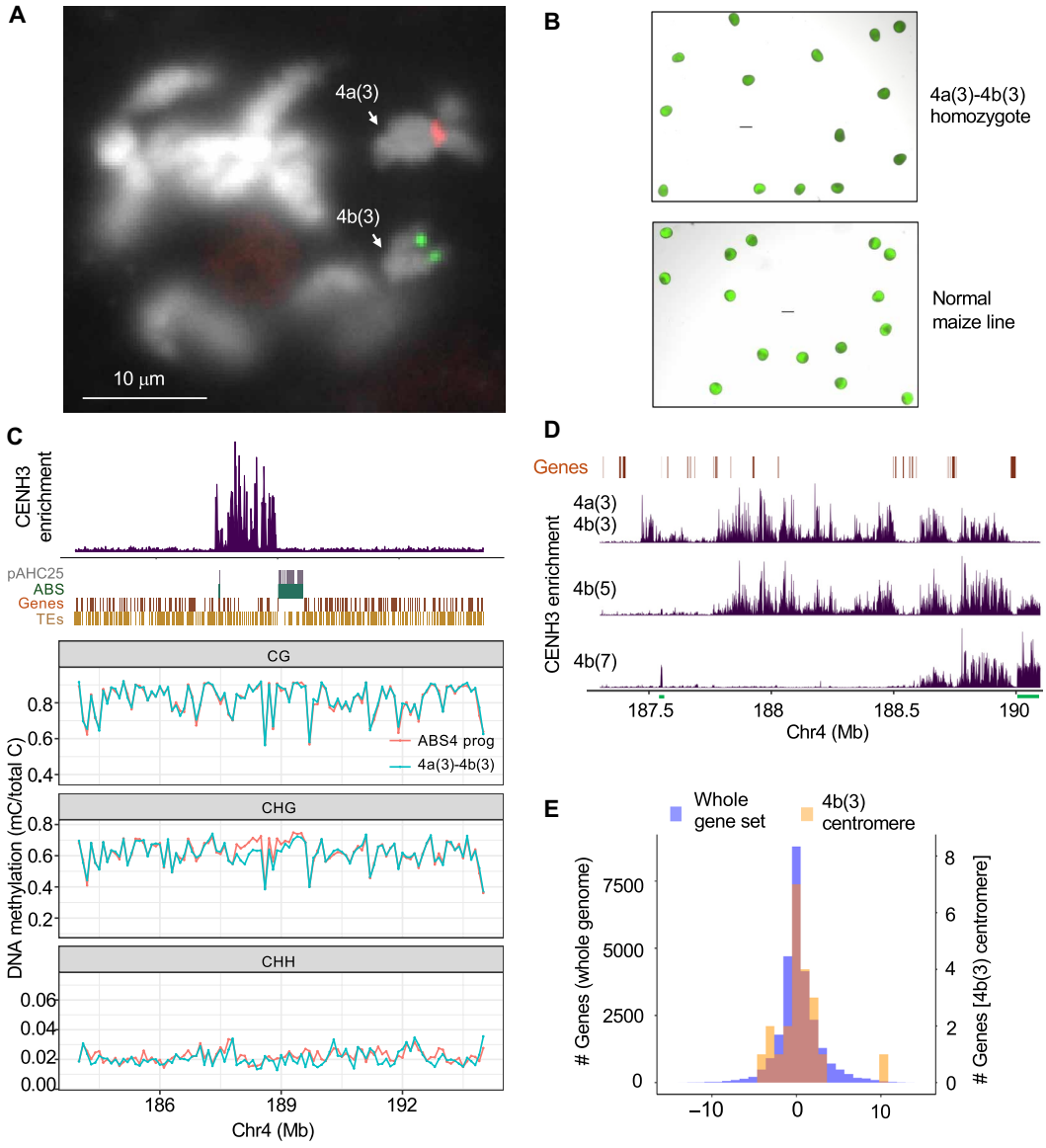

Fig. 4. Analysis of homozygous 4a(3)-4b(3) lines. (**A**) Meiotic cell in late diplotene showing the pairing of the homozygous 4a(3) and 4b(3) chromosomes. Arrows point to chiasmata. DNA is in white, Cent4 is in red, and ABS4 is in green. (**B**) Pollen from a homozygous 4a(3)-4b(3) plant (top) and normal karyotype *c2*/*c2* plant (bottom) stained with fluorescein diacetate (FDA). Scale bars (suspended in the images), 100 μm. (**C**) CENH3 CUT&Tag and DNA methylation profiles in the homozygous 4a(3)-4b(3) line. DNA methylation profiles in each of the CG, CHG, and CHH contexts in the same plants used for RNA expression analysis in (E). Methylation is expressed as methylated cytosines/total cytosines (mC/total C). (**D**) Zoomed-in view of the CENH3 CUT&Tag profile of the homozygous 4a(3)-4b(3) line, showing dips over genic regions. The CENH3 profiles from the 4b(5) and 4b(7) partially trisomic lines (Fig. 1C) are also included for comparison. Green bars show ABS sequences. (**E**) Relative gene expression between 4a(3)-4b(3) homozygotes and the ABS4 progenitor line. The proportion of genes with increased or decreased expression in the 4b(3) centromere was not significantly different from the whole-genome average (chi-square test, $P = 0.2071$ for increased and 1 for decreased). Figure S7D shows the expression of all 22 genes in the four replicates of each genotype.

derived from the embryo). These crosses were designed to assess whether 4b(3) might be subject to centromere erasure and chromosome loss in the embryo, similar to the centromere-mediated haploid induction process (*39*, *40*). Of 432 kernels from female crosses and 889 kernels from male crosses, we observed no kernels with colorless scutella (table S4). We did, however, observe a low frequency of sectored kernels in male crosses (1%). Sectored kernels likely represent cases of recombination between the short overlapping regions of 4a(3) and 4b(3), leading to dicentric chromosomes and BFB.

Effect of the 4b(3) centromere on gene expression and DNA methylation

Having 4b(3) in a homozygous state allowed us to assess DNA methylation and gene expression in the CENH3-occupied area. To this end, we first performed CENH3 CUT&Tag on the homozygous line and observed that, in the three generations since the first CUT&Tag was carried out (*21*), the centromere location had shifted entirely into genomic sequence to the left of ABS4 and occupied a region of ~1.5 Mb (Fig. 4C). A zoomed-in comparison of the CENH3 profile to gene annotations shows that CENH3 is enriched between genes and depleted within genes (Fig. 4D), similar to what we observed for the other 4b centromeres (Fig. 1C). We also looked at the position of the 4a(3) centromere and found that its position had not noticeably changed (fig. S7B).

Prior analyses of native plant centromeres have shown that CENH3-enriched regions and pericentromeric regions have high levels of DNA methylation in the CG context, while the CENH3-enriched regions have somewhat reduced CHG methylation (*6*, *41*–*43*). To test whether the same trends were observed in the 4b(3) centromere, we grew four plants from the ABS4 inbred (no centromere at ABS4) and four plants that were homozygous for the 4a(3)-4b(3) pair and carried out enzymatic methyl sequencing (EM-seq) of leaves. Analysis of the methylation patterns revealed no visible changes in CG methylation throughout the ABS4 region (Fig. 4C and fig. S7C). However, consistent with prior data from native centromeres, there was a slight reduction in CHG methylation in the CENH3-enriched area (from ~70 to 60%). DNA methylation within and around genes was largely unaffected by the formation of the 4b(3) centromere.

When a centromere shifts into genomic regions, the expression of genes in the affected area might be altered. To test this, we used tissue from the same eight plants [four ABS4 control and four 4a(3)-4b(3) homozygotes] for comparative RNA sequencing (RNA-seq) analysis. Overall, 8222 of 23,204 genes with detectable expression were differentially expressed (Wald test, Bonferroni-adjusted $P < 0.01$). A total of 4493 genes showed increased expression in the homozygous 4a(3)-4b(3) plants, and 3729 genes showed decreased expression relative to the ABS4 control. Of the 22 genes in the CENH3-enriched region with detectable expression, four had increased expression in 4a(3)-4b(3) plants, and one had decreased expression (Fig. 4E and fig. S7D). The proportion of genes with increased or decreased expression in the CENH3-enriched region was not significantly different from the whole-genome average (chi-square test, $P = 0.2071$ for increased and 1 for decreased).

DISCUSSION

Methods for routine centromere engineering could be used to markedly alter karyotypes, restructure recombination landscapes, and speed the development of artificial chromosomes (*1*, *44*, *45*). In this study, we describe a pathway for the formation of stable new karyotypes from lines with newly formed centromeres. The key criterion is that the new centromere be on the same chromosome as the original centromere and far enough away that meiotic recombination between the two centromeres is possible. When starting with a partially trisomic line, recombinants of this type are favored when the plants are crossed as a male (Fig. 2C). The dicentric chromosomes formed by recombination undergo BFB and can be converted to chromosome pairs with overlapping ends, which, as we show here, can be recovered as new 22-chromosome karyotypes. Lines with 22 chromosomes have been previously produced by breakage of an existing centromere into two parts (centromere misdivision) (*46*–*48*) and by combining translocations so that a native centromere is duplicated but the arms are on separate chromosomes (*49*, *50*). In our approach, the additional chromosome is derived from a chromosome arm fragment and is stably transmitted by a synthetic centromere.

The rapid formation of telomeres on newly broken ends is critical for the formation of new chromosomes. McClintock demonstrated that BFB cycles usually terminate early in development by a process of healing (de novo telomere formation) (*26*, *27*). Later studies in wheat showed that dicentric chromosomes underwent BFB until about 4 weeks after germination, at which point most of the broken ends were healed with telomeres visible by in situ hybridization (*51*). Four telomere addition sites were sequenced, and the sites of addition had 2 to 4 nucleotides of homology to the telomere repeat (*52*), consistent with biochemical studies showing that plant telomerases can initiate new telomeres on a variety of templates (*28*). In our study, we did not visualize telomeres until the neochromosomes had passed through at least one generation of crossing, but in the lines where we looked, telomeres were visible by in situ hybridization on the recently broken ends.

The ten 4b centromeres described here cover a minimum of 1.5 Mb of sequence, similar to native centromeres (*30*, *53*), and in all cases, CENH3 spread into the flanking genic areas. For the homozygous 4a(3)-4b(3) line, we found no evidence for reduced gene expression within the CENH3 domain relative to the progenitor ABS4 chromosome beyond normal variation. The fact that centromeres can form among genes has also been documented in native centromeres and spontaneous neocentromeres in both plants and animals (*32*, *35*, *54*, *55*). The forces that drive and limit CENH3 spreading and the mechanisms that exclude CENH3 from genes are not yet understood. Newly formed centromeres also lack the long spans of flanking pericentromeric heterochromatin that typically surround native centromeres. On the 4b(3) centromere, we observed no increase in DNA methylation that might indicate an elevated or altered heterochromatin environment (Fig. 4C). These data support the view that pericentromeric heterochromatin is the result of a gradual buildup of repetitive sequences in low-recombination regions within and surrounding centromeres and is not required for centromere function (*55*–*58*).

Perhaps the most compelling case that the new 4b(3) centromere and telomeres are functional, and not impairing gene expression or other essential processes, comes from the fact that the 22-chromosome maize line grew and reproduced as well as any standard laboratory strain. However, despite its normal phenotype, the 22-chromosome line will no longer be compatible with breeding programs involving standard (20-chromosome) maize lines. Whenever a 22-chromosome line is crossed to a standard maize line, recombination between the

centromeres on normal chromosome 4 and neochromosome 4b should lead to dicentric chromosomes and further structural rearrangements (Fig. 2C), reducing fertility and severely limiting the transfer of alleles from one chromosome 4 karyotype to the other. For this reason, major karyotype alterations are thought to be one form of genetic isolation that leads to speciation (*59*).

The fact that 22-chromosome maize lines are not compatible with standard maize breeding lines could be seen as an advantage. It may be possible to initiate new breeding programs based on maize lines with altered karyotypes. In this way, proprietary maize lines could be durably marked by their unique karyotypes. Engineered gene clusters could be placed on (for instance) the 4b chromosome, perhaps close to the centromere, to ensure that they are not inadvertently introduced into other standard backgrounds. Further, the fission of one chromosome into two separate chromosomes should markedly alter the recombination landscape, perhaps releasing genetic variation that has the potential to improve maize in new and unpredictable ways (*1*).

MATERIALS AND METHODS

Plant materials

The ABS4 inbred was developed by self-crossing the original transgenic line in a Hi-II genetic background five times (*23*). The *c2-gfp* allele was obtained as stock tdsgR64H07, and the *c2 R1-scm2* strain was obtained as stock X25B from the Maize Genetics Cooperation Stock Center in Urbana, IL. Plants were grown and crossed in the University of Georgia Botany greenhouses.

Genotyping

The presence of the ABS4 insertion was assayed using primers abs4-p3 (5′-TCCTCCGGAGTACCGTCT-3′) and abs4-p4 (5′-AGCCAGGCGGATAGAAGC-3′), whereas the wild-type locus was amplified using primers abs4-p1 (TACCCTGGTTAGAGGGAGCC) and abs4-p4 (*21*). To test for the presence of the dominant *C2* allele by PCR, we developed primers (C2-F1, 5′-CACAGCGTCCCCATCACC-3′; C2-R1, 5′-GAACGAGACGACGACGAATTG-3′) that amplify a sequence over 26-bp insertion-deletion polymorphism in the 3′ untranslated region of the *c2* gene (the *C2* allele in the ABS4 inbred has the insertion, whereas our standard *c2*/*c2* line has the deletion). Fluorescence from the *c2-gfp* allele was visualized with a blue light source and orange filter (Clare Chemical, HL34). In cases where both the *C2* and *c2-gfp* alleles were present, it was often necessary to sand off (with sandpaper) a small portion of the aleurone to visualize the fluorescence.

HiFi sequencing and assembly of the ABS4 inbred

The two top leaves of a single ABS4 inbred plant were sent to the Arizona Genomics Institute in Tucson, Arizona. DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (*60*), and library preparation was performed with the SMRTbell Express Template Prep Kit 3.0. The final library preparation included size selection for fragments ranging from 10 to 25 kb using the BluePippin system (Sage Science). Sequencing was conducted on a PacBio Revio system in Circular Consensus Sequence mode. The read length N50 was 22,840 bp. Genome assembly was carried out using Hifiasm v0.19.4 (*61*) with the "-t 256" and "-l 0" parameters to enable phasing, sequence error correction, and assembly. The optional parameter -u was initially tested but was disabled because of increased assembly errors. The final assembly produced 1136 contigs with a total length of ~2.3 Gb and a contig N50 of 34,924,731 bp.

RagTag v2.1.0 (*62*) was used to scaffold the contigs using the A188 reference genome (*24*). The sequenced plant had a low level of heterozygosity, particularly on chromosome 5, where heterozygous contigs were shorter and had lower coverage. RagTag v2.1.0 does not distinguish heterozygous contigs and can place genome duplicates into the final reference. A custom algorithm was developed to identify and filter duplicate contigs by calculating the percent overlap between consecutive contigs using the RagTag v2.1.0–generated assembly genome path (AGP) file. Contigs with over 99% overlap with the preceding contig were removed. Contig ptg000071l was misassembled because of a repetitive area with TAC trinucleotide repeats. The RagTag v2.1.0 "correct" command was used to split this contig and rescaffold with the scaffold command.

We identified four primary contigs containing ABS and pAHC25 sequences. Two of these contigs were placed by RagTag, as they contained maize genome sequences, while the other two unplaced contigs overlapped with 100% sequence identity to portions of the placed contigs. We manually scaffolded these two unplaced contigs into the final assembly by joining the overlapping regions. To plot the synteny between the A188 and the ABS4 genomes, syntenic genes were paired by gene name and the data were visualized using KaryoploteR (*63*).

Annotation of the ABS4 inbred genome

Gene annotation was performed with Liftoff v.1.6.3 (*64*) using the A188 reference and gene annotation as input (*24*). Transposable elements were annotated using EDTA v2.1.0 with "--species Maize --step anno --overwrite 1 --cds --u 1.3e-8 --threads 18" parameters (*65*).

To estimate the number of LexO sites in the ABS4 array, the large (second) insertion region was extracted using bedtools v2.31.0 (*66*) and a nucleotide basic local alignment search tool (BLAST) database built using the "makeblastdb" command. The LexO motif (TACTGTATATATATACAGTA) was queried against the database using the "blastn-short" command and only exact matches to all 20-bp counted. By this measure, the estimated number of LexO binding sites is 2783.

FISH of mitotic and meiotic chromosomes

The preparation of metaphase spreads from root tip chromosomes, and subsequent FISH was carried out as previously described (*21*), using fluorescein isothiocyanate (FITC)–labeled oligos for ABS, Cy5-labeled oligos for Cent4 (*67*), and a Cy3-labeled oligo (5′-TTTAGGGTTTAGGGTTTAGGG-3′) for telomere repeats. The preparation of meiotic samples and subsequent FISH was carried out as previously described (*21*), using FITC-labeled oligos for ABS and Cy5-labeled oligos for Cent4. Cells were imaged using a Zeiss Axio Imager M1 fluorescence microscope with 63× or 100× Plan-Apochromat oil objectives. Data were analyzed using SlideBook software (Intelligent Imaging Innovations).

Whole-genome Illumina sequencing and analysis

Whole-genome sequencing libraries were prepared with KAPA HyperPrep Kits (KK8502) using 50 ng of sonicated input DNA with NEXTflex DNA Barcodes (Bioo Scientific, NOVA-520996), reducing all reaction volumes by half, and amplifying them with four or five cycles of PCR using the NEXTflex primer mix. Libraries were sequenced paired-end with 150-nt reads. Adapter sequences were trimmed using Trim Galore v0.6.7 (*68*) with the --fastqc, --gzip, and --paired -a AGATCGGAAGAGC parameters. Trimmed reads were

then aligned to the reference genome using BWA-mem v0.11.3 with the -M parameter (*69*). Three reference genomes—ABS4 assembly, Mo17 (*31*), and W22 (*70*)—were tested, with the W22 genome yielding the most consistent results. A 100-kb window file for the W22 genome was generated using bedtools v2.31.0 (*66*) with the makewindows tool. Then, the intersect tool with "-c" was used to calculate read counts for each window. The output file was processed in R, where the normalized read count was calculated as the read count of each 100-kb window divided by the total mapped read count times the total number of 100-kb windows times two.

CENH3 CUT&Tag assays and analysis

Unfertilized ears, varying from 3 to 8 cm in length, were used for CUT&Tag with CENH3 antibodies (*71*). Nuclei were purified using a sucrose/Percoll cushion method (*72*), and intact nuclei containing 350 ng of DNA were used as input for each CUT&Tag reaction with Active Motif CUT&Tag-IT kits and amplified with 15 cycles of PCR, as described previously (*21*). The nuclei purification method was modified slightly in that two layers of miracloth were used instead of four, nuclei suspensions were filtered only once through cell strainers instead of twice, and 100- and 40-μm cell strainers were used instead of 20- and 10-μm cell strainers. CUT&Tag libraries were Illumina sequenced paired-end 150 nt. Adapter sequences were trimmed using Trim Galore v0.6.7 --fastqc, --gzip, --paired -a CTGTCTCTTATACACATCT (*68*). Trimmed reads were aligned to the ABS4 assembly using BWA-mem v0.11.3 with the "-M" parameter (*69*). A 10-kb window file for ABS4 assembly was generated using bedtools v2.31.0 (*66*) with the makewindows tool. Then, the intersect tool with -c was used to calculate read counts for each window. The results were visualized using ggplot2 with the geom_area function plotting the CENH3 enrichment peaks and the geom_rect function plotting different genetic elements (genes, transposable elements, ABS4, and pAHC25).

Pollen morphology

Freshly released pollen from three plants of each genotype was immersed in an aqueous solution of fluorescein diacetate (FDA; 5 ng/ml; Chem Cruz, #Sc-294598) in 1.4 M sucrose. The FDA assay tests for the integrity of the plasmalemma of the vegetative cell, which is an indirect measure of cell viability (*73*). At least 100 pollen grains were counted per plant using images taken with a Nikon ECLIPSE Ti2 inverted microscope, and any that had partially collapsed cytoplasm or irregular shape were counted as abnormal.

Gene expression assays

Strand-specific RNA-seq libraries were prepared from leaves from four individual plants from the ABS4 inbred and four individual plants homozygous for 4a(3) and 4b(3) using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina following the manufacturer's instructions. The individuals from the homozygous 4a(3)-4b(3) line had a complex genetic background but were identical to the ABS4 inbred in the ABS4 region. Briefly, the RNAs were fragmented for 8 min at 94°C. The first-strand and second-strand cDNAs were subsequently synthesized. The second strand of cDNA was marked by incorporating deoxyuridine triphosphate (dUTP) during the synthesis. cDNA fragments were adenylated at 3′ ends, and an indexed adapter was ligated to cDNA fragments. Limited-cycle PCR was used for library enrichment. The incorporated dUTP in second-strand cDNA quenched second-strand amplification, which preserved strand specificity. Libraries were Illumina sequenced paired-end 150 nt. Adapter sequences were AGATCGGAAGAGCACACGTCTGAACTCCAGTCA (read 1) and AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT (read 2).

After trimming the adapter with Trim Galore v.6.7 (*68*), mRNA reads were mapped to the ABS4 inbred using STAR v.2.7 with "--outSAMunmapped Within, --outSAMattributes Standard" command (*74*). Cufflinks v.2.2.1 (*75*) transformed the Liftoff-generated gene annotation gff file to gtf format. featureCounts v.1.6 (*76*) was used to calculate read count for each exon. featureCounts output files were then channeled into R to calculate the read count for each gene.

DNA methylation analyses

EM-seq libraries were prepared from the same leaves that were used for RNA-seq using a NEBNext Enzymatic Methyl-seq Kit (New England Biolabs, #E7120S). The input for each library consisted of 200 ng of DNA that had been combined with 1 pg of control pUC19 DNA and 20 pg of control lambda DNA and sonicated to fragments averaging ~600 bp in length using a Diagenode Bioruptor. The protocol for large insert libraries was followed with formamide as the denaturing agent, and libraries were amplified with four PCR cycles and Illumina sequenced using paired-end 150-nt reads. EM-seq reads were trimmed of adapter sequence using cutadapt (*77*), parameters -q 20 -a AGATCGGAAGAGC -A AGATCGGAAGAGC -O. Reads were aligned to each genome, and methylation values were called using BS-Seeker2 v.2.1.5, parameters -m 1 –aligner=bowtie2 -X 1000 (*78*). The resulting files in CGmap format were processed using CGmapTools v.1.2 (*79*). The replicates from the ABS4 inbred and 4a(3)-4b(3) homozygous lines were merged two by two using the merge2 tool and the 100-kb window methylation calculation done with the mbin tool "-c 1, -B 100000," and the results were plotted with gpplot2 wrapped in tidyverse.

Statistical analyses

To compare recombination between *C2* and ABS4 in female and male crosses (table S2, $n \geq 9$ from three families for both female and male crosses), we conducted a logistic regression with a custom R script with the model fit Recombination ~ Cross Direction. Logistic regression showed that the log odds of recombination increased by 1.9994 (95% confidence interval, 1.3374 to 2.6566; $P = 0.002325$) in male crosses compared to female crosses. We excluded two samples from the counts that were missing both ABS4 and *C2*.

To assess the effect of the new centromere in the 4a(3)-4b(3) line on gene expression, we first determined that 22 of 26 genes in the CENH3-enriched region had nonzero expression in at least one sample [$n = 4$ biological replicates for both the ABS4 inbred and homozygous 4a(3)-4b(3) lines]. To compare the expression of these 22 genes, DESeq2 (*80*) was used with default parameters to calculate relative log expression and perform differential expression analysis. Genes with Bonferroni-adjusted *P* values of less than 0.01 were defined as differentially expressed. A chi-square test was used to compare the frequency of differentially expressed genes among the 22 genes versus the whole genome, producing a *P* value of 0.21 for increased gene expression and a *P* value of 1 for decreased gene expression.

Supplementary Materials

This PDF file includes:

Figs. S1 to S7

Tables S1 to S5

REFERENCES AND NOTES

1. H. Puchta, A. Houben, Plant chromosome engineering—Past, present and future. *New Phytol.* **241**, 541–552 (2024).
2. A. Musacchio, A. Desai, A molecular view of kinetochore assembly and function. *Biology* **6**, (2017).
3. P. E. Warburton, Chromosomal dynamics of human neocentromere formation. *Chromosome Res.* **12**, 617–626 (2004).
4. J. Zhou, Y. Liu, X. Guo, J. A. Birchler, F. Han, H. Su, Centromeres: From chromosome biology to biotechnology applications and synthetic genomes in plants. *Plant Biotechnol. J.* **20**, 2051–2063 (2022).
5. N. Altemose, G. A. Logsdon, A. V. Bzikadze, P. Sidhwani, S. A. Langley, G. V. Caldas, S. J. Hoyt, L. Uralsky, F. D. Ryabov, C. J. Shew, M. E. G. Sauria, M. Borchers, A. Gershman, A. Mikheenko, V. A. Shepelev, T. Dvorkina, O. Kunyavskaya, M. R. Vollger, A. Rhie, A. M. McCartney, M. Asri, R. Lorig-Roach, K. Shafin, J. K. Lucas, S. Aganezov, D. Olson, L. G. de Lima, T. Potapova, G. A. Hartley, M. Haukness, P. Kerpedjiev, F. Gusev, K. Tigyi, S. Brooks, A. Young, S. Nurk, S. Koren, S. R. Salama, B. Paten, E. I. Rogaev, A. Streets, G. H. Karpen, A. F. Dernburg, B. A. Sullivan, A. F. Straight, T. J. Wheeler, J. L. Gerton, E. E. Eichler, A. M. Phillippy, W. Timp, M. Y. Dennis, R. J. O'Neill, J. M. Zook, M. C. Schatz, P. A. Pevzner, M. Diekhans, C. H. Langley, I. A. Alexandrov, K. H. Miga, Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).
6. M. Naish, M. Alonge, P. Wlodzimierz, A. J. Tock, B. W. Abramson, A. Schmücker, T. Mandáková, B. Jamge, C. Lambing, P. Kuo, N. Yelina, N. Hartwick, K. Colt, L. M. Smith, J. Ton, T. Kakutani, R. A. Martienssen, K. Schneeberger, M. A. Lysak, F. Berger, A. Bousios, T. P. Michael, M. C. Schatz, I. R. Henderson, The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science* **374**, eabi7489 (2021).
7. K. L. McKinley, I. M. Cheeseman, The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17**, 16–29 (2016).
8. P. Parashara, B. Medina-Pritchard, M. A. Abad, P. Sotelo-Parrilla, R. Thamkachy, D. Grundei, J. Zou, C. Spanos, C. N. Kumar, C. Basquin, V. Das, Z. Yan, A. A. Al-Murtadha, D. A. Kelly, T. McHugh, A. Imhof, J. Rappsilber, A. A. Jeyaprakash, PLK1-mediated phosphorylation cascade activates Mis18 complex to ensure centromere inheritance. *Science* **385**, 1098–1104 (2024).
9. K. E. Gascoigne, K. Takeuchi, A. Suzuki, T. Hori, T. Fukagawa, I. M. Cheeseman, Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410–422 (2011).
10. M. J. Mendiburo, J. Padeken, S. Fülöp, A. Schepers, P. Heun, *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**, 686–690 (2011).
11. K. E. Gascoigne, I. M. Cheeseman, Induced dicentric chromosome formation promotes genomic rearrangements and tumorigenesis. *Chromosome Res.* **21**, 407–418 (2013).
12. T. Hori, W.-H. Shang, K. Takeuchi, T. Fukagawa, The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *J. Cell Biol.* **200**, 45–60 (2013).
13. C. H. Teo, I. Lermontova, A. Houben, M. F. Mette, I. Schubert, De novo generation of plant centromeres at tandem repeats. *Chromosoma* **122**, 233–241 (2013).
14. J. Cao, T. Hori, M. Ariyoshi, T. Fukagawa, Artificial tethering of constitutive centromere-associated network proteins induces CENP-A deposition without Knl2 in DT40 cells. *J. Cell Sci.* **137**, jcs261639 (2024).
15. S. Lacefield, D. T. C. Lau, A. W. Murray, Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. *Nat. Cell Biol.* **11**, 1116–1120 (2009).
16. M. C. Barnhart, P. H. J. L. Kuich, M. E. Stellfox, J. A. Ward, E. A. Bassett, B. E. Black, D. R. Foltz, HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* **194**, 229–243 (2011).
17. C.-C. Chen, M. L. Dechassa, E. Bettini, M. B. Ledoux, C. Belisario, P. Heun, K. Luger, B. G. Mellone, CAL1 is the *Drosophila* CENP-A assembly factor. *J. Cell Biol.* **204**, 313–329 (2014).
18. G. A. Logsdon, C. W. Gambogi, M. A. Liskovykh, E. J. Barrey, V. Larionov, K. H. Miga, P. Heun, B. E. Black, Human artificial chromosomes that bypass centromeric DNA. *Cell* **178**, 624–639.e19 (2019).
19. J. Palladino, A. Chavan, A. Sposato, T. D. Mason, B. G. Mellone, Targeted *de novo* centromere formation in *Drosophila* reveals plasticity and maintenance potential of CENP-A chromatin. *Dev. Cell* **52**, 379–394.e7 (2020).
20. C. W. Gambogi, G. J. Birchak, E. Mer, D. M. Brown, G. Yankson, K. Kixmoeller, J. N. Gavade, J. L. Espinoza, P. Kashyap, C. L. Dupont, G. A. Logsdon, P. Heun, J. I. Glass, B. E. Black, Efficient formation of single-copy human artificial chromosomes. *Science* **383**, 1344–1349 (2024).
21. R. K. Dawe, J. I. Gent, Y. Zeng, H. Zhang, F.-F. Fu, K. W. Swentowsky, D. W. Kim, N. Wang, J. Liu, R. D. Piri, Synthetic maize centromeres transmit chromosomes across generations. *Nat. Plants* **9**, 433–441 (2023).
22. C. L. Armstrong, C. E. Green, R. L. Phillips, Development and availability of germplasm with high type II culture formation response. *Maize Genet. Coop. News Lett.*, 92–93 (1991).
23. H. Zhang, B. H. Phan, K. Wang, B. J. Artelt, J. Jiang, W. A. Parrott, R. K. Dawe, Stable integration of an engineered megabase repeat array into the maize genome. *Plant J.* **70**, 357–365 (2012).
24. G. Lin, C. He, J. Zheng, D.-H. Koo, H. Le, H. Zheng, T. M. Tamang, J. Lin, Y. Liu, M. Zhao, Y. Hao, F. McFraland, B. Wang, Y. Qin, H. Tang, D. R. McCarty, H. Wei, M.-J. Cho, S. Park, H. Kaeppler, S. M. Kaeppler, Y. Liu, N. Springer, P. S. Schnable, G. Wang, F. F. White, S. Liu, Chromosome-level genome assembly of a regenerable maize inbred line A188. *Genome Biol.* **22**, 175 (2021).
25. B. McClintock, The fusion of broken ends of sister half-chromatids following chromatid breakage at meiotic anaphases. *Missouri Agr. Exp. Sta. Res. Bull.* **290**, 1–48 (1938).
26. B. McClintock, The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl. Acad. Sci. U.S.A.* **25**, 405–416 (1939).
27. B. McClintock, The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
28. M. S. Fitzgerald, E. V. Shakirov, E. E. Hood, T. D. McKnight, D. E. Shippen, Different modes of de novo telomere formation by plant telomerases. *Plant J.* **26**, 77–87 (2001).
29. J. I. Gent, N. Wang, R. K. Dawe, Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. *Genome Biol.* **18**, 121 (2017).
30. N. Wang, J. Liu, W. A. Ricci, J. I. Gent, R. K. Dawe, Maize centromeric chromatin scales with changes in genome size. *Genetics* **217**, iyab020 (2021).
31. J. Chen, Z. Wang, K. Tan, W. Huang, J. Shi, T. Li, J. Hu, K. Wang, C. Wang, B. Xin, H. Zhao, W. Song, M. B. Hufford, J. C. Schnable, W. Jin, J. Lai, A complete telomere-to-telomere assembly of the maize genome. *Nat. Genet.* **55**, 1221–1231 (2023).
32. K. Nagaki, Z. Cheng, S. Ouyang, P. B. Talbert, M. Kim, K. M. Jones, S. Henikoff, C. R. Buell, J. Jiang, Sequencing of a rice centromere uncovers active genes. *Nat. Genet.* **36**, 138–145 (2004).
33. K. Wang, Y. Wu, W. Zhang, R. K. Dawe, J. Jiang, Maize centromeres expand and adopt a uniform size in the genetic background of oat. *Genome Res.* **24**, 107–116 (2014).
34. H. Zhao, X. Zhu, K. Wang, J. I. Gent, W. Zhang, R. K. Dawe, J. Jiang, Gene expression and chromatin modifications associated with maize centromeres. *G3* **6**, 183–192 (2016).
35. H. Su, Y. Liu, Y.-X. Liu, Z. Lv, H. Li, S. Xie, Z. Gao, J. Pang, X.-J. Wang, J. Lai, J. A. Birchler, F. Han, Dynamic chromatin changes associated with de novo centromere formation in maize euchromatin. *Plant J.* **88**, 854–866 (2016).
36. D. L. Auger, J. A. Birchler, Maize tertiary trisomic stocks derived from B-A translocations. *J. Hered.* **93**, 42–47 (2002).
37. J. Einset, Chromosome length in relation to transmission frequency of maize trisomes. *Genetics* **28**, 349–364 (1943).
38. P. M. A. Kianian, M. Wang, K. Simons, F. Ghavami, Y. He, S. Dukowic-Schulze, A. Sundararajan, Q. Sun, J. Pillardy, J. Mudge, C. Chen, S. F. Kianian, W. P. Pawlowski, High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. *Nat. Commun.* **9**, 2370 (2018).
39. N. Wang, J. I. Gent, R. Kelly Dawe, Haploid induction by a maize *cenh3* null mutant. *Sci. Adv.* **7**, eabe2299 (2021).
40. M. P. A. Marimuthu, R. Maruthachalam, R. Bondada, S. Kuppu, E. H. Tan, A. Britt, S. W. L. Chan, L. Comai, Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. *Sci. Adv.* **7**, eabk1151 (2021).
41. J. I. Gent, Y. Dong, J. Jiang, R. K. Dawe, Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res.* **40**, 1550–1560 (2012).
42. P. G. Hofstatter, G. Thangavel, T. Lux, P. Neumann, T. Vondrak, P. Novak, M. Zhang, L. Costa, M. Castellani, A. Scott, H. Toegelová, J. Fuchs, Y. Mata-Sucre, Y. Dias, A. L. L. Vanzela, B. Huettel, C. C. S. Almeida, H. Šimková, G. Souza, A. Pedrosa-Harand, J. Macas, K. F. X. Mayer, A. Houben, A. Marques, Repeat-based holocentromeres influence genome architecture and karyotype evolution. *Cell* **185**, 3153–3168.e18 (2022).
43. Y.-T. Kuo, A. S. Câmara, V. Schubert, P. Neumann, J. Macas, M. Melzer, J. Chen, J. Fuchs, S. Abel, E. Klocke, B. Huettel, A. Himmelbach, D. Demidov, F. Dunemann, M. Mascher, T. Ishii, A. Marques, A. Houben, Holocentromeres can consist of merely a few megabase-sized satellite arrays. *Nat. Commun.* **14**, 3502 (2023).
44. J. Luo, X. Sun, B. P. Cormack, J. D. Boeke, Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018).
45. R. K. Dawe, Engineering better artificial chromosomes. *Science* **383**, 1292–1293 (2024).
46. K. L. Rose, R. W. Staub, Centromeric breakage and the recovery of both telocentric fragments. *Maize Genet. Coop. News Lett.*, 94–95 (1990).
47. T. A. L'Heureux, D. A. Muzumdar, M. C. Schneerman, D. F. Weber, Haploids and monosomics are produced by maize plants containing telocentrics for both arms of chromosome 3 and B chromosomes. *Maize Genet. Coop. News Lett.*, 66–67 (1997).
48. A. J. Lukaszewski, "Misdivision of centromeres" in *Plant Centromere Biology*, J. Jiang, J. A. Birchler, Eds. (Wiley-Blackwell, 2013), pp. 111–128.
49. W. R. Carlson, C. Curtis, A new method for producing homozygous duplications in maize. *Can. J. Genet. Cytol.* **28**, 1034–1040 (1986).

50. Y. Z. Zhang, W. R. Carlson, Further construction of proximal duplication stocks. *Maize Genet. Coop. News Lett.*, 37–39 (1997).
51. B. Friebe, R. G. Kynast, P. Zhang, L. Qi, M. Dhar, B. S. Gill, Chromosome healing by addition of telomeric repeats in wheat occurs during the first mitotic divisions of the sporophyte and is a gradual process. *Chromosome Res.* **9**, 137–146 (2001).
52. H. Tsujimoto, N. Usami, K. Hasegawa, T. Yamada, K. Nagaki, T. Sasakuma, De novo synthesis of telomere sequences at the healed breakpoints of wheat deletion chromosomes. *Mol. Gen. Genet.* **262**, 851–856 (1999).
53. M. B. Hufford, A. S. Seetharam, M. R. Woodhouse, K. M. Chougule, S. Ou, J. Liu, W. A. Ricci, T. Guo, A. Olson, Y. Qiu, R. Della Coletta, S. Tittes, A. I. Hudson, A. P. Marand, S. Wei, Z. Lu, B. Wang, M. K. Tello-Ruiz, R. D. Piri, N. Wang, D. W. Kim, Y. Zeng, C. H. O'Connor, X. Li, A. M. Gilbert, E. Baggs, K. V. Krasileva, J. L. Portwood II, E. K. S. Cannon, C. M. Andorf, N. Manchanda, S. J. Snodgrass, D. E. Hufnagel, Q. Jiang, S. Pedersen, M. L. Syring, D. A. Kudrna, V. Llaca, K. Fengler, R. J. Schmitz, J. Ross-Ibarra, J. Yu, J. I. Gent, C. N. Hirsch, D. Ware, R. K. Dawe, De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
54. O. J. Marshall, A. C. Chueh, L. H. Wong, K. H. A. Choo, Neocentromeres: New insights into centromere structure, disease development, and karyotype evolution. *Am. J. Hum. Genet.* **82**, 261–282 (2008).
55. W.-H. Shang, T. Hori, N. M. C. Martins, A. Toyoda, S. Misu, N. Monma, I. Hiratani, K. Maeshima, K. Ikeo, A. Fujiyama, H. Kimura, W. C. Earnshaw, T. Fukagawa, Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev. Cell* **24**, 635–648 (2013).
56. A. Alonso, D. Hasson, F. Cheung, P. E. Warburton, A paucity of heterochromatin at functional human neocentromeres. *Epigenetics Chromatin* **3**, 6 (2010).
57. K. W. Y. Yuen, K. Nabeshima, K. Oegema, A. Desai, Rapid de novo centromere formation occurs independently of heterochromatin protein 1 in *C. elegans* embryos. *Curr. Biol.* **21**, 1800–1807 (2011).
58. P. Kalitsis, K. H. A. Choo, The evolutionary life cycle of the resilient centromere. *Chromosoma* **121**, 327–340 (2012).
59. L. H. Rieseberg, Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).
60. J. J. Doyle, J. L. Doyle, A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull.* **19**, 11–15 (1987).
61. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
62. M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, X. Wang, Z. B. Lippman, M. C. Schatz, S. Soyk, Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
63. B. Gel, E. Serra, karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).
64. A. Shumate, S. Salzberg, LiftoffTools: A toolkit for comparing gene annotations mapped between genome assemblies. *F1000Res.* **11**, 1230 (2022).
65. S. Ou, W. Su, Y. Liao, K. Chougule, J. R. A. Agda, A. J. Hellinga, C. S. B. Lugo, T. A. Elliott, D. Ware, T. Peterson, N. Jiang, C. N. Hirsch, M. B. Hufford, Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
66. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
67. B. T. Page, M. K. Wanous, J. A. Birchler, Characterization of a maize chromosome 4 centromeric sequence: Evidence for an evolutionary relationship with the B chromosome centromere. *Genetics* **159**, 291–302 (2001).
68. F. Krueger, Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham Institute (2015). https://github.com/FelixKrueger/TrimGalore.
69. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. N. M. Springer, S. N. Anderson, C. M. Andorf, K. R. Ahern, F. Bai, O. Barad, W. B. Barbazuk, H. W. Bass, K. Baruch, G. Ben-Zvi, E. S. Buckler, R. Bukowski, M. S. Campbell, E. K. S. Cannon, P. Chomet, R. K. Dawe, R. Davenport, H. K. Dooner, L. H. Du, C. Du, K. A. Easterling, C. Gault, J.-C. Guan, C. T. Hunter, G. Jander, Y. Jiao, K. E. Koch, G. Kol, T. G. Köllner, T. Kudo, Q. Li, F. Lu, D. Mayfield-Jones, W. Mei, D. R. McCarty, J. M. Noshay, J. L. Portwood II, G. Ronen, A. M. Settles, D. Shem-Tov, J. Shi, I. Soifer, J. C. Stein, M. C. Stitzer, M. Suzuki, D. L. Vera, E. Vollbrecht, J. T. Vrebalov, D. Ware, S. Wei, K. Wimalanathan, M. R. Woodhouse, W. Xiong, T. P. Brutnell, The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* **50**, 1282–1288 (2018).
71. C. X. Zhong, J. B. Marshall, C. Topp, R. Mroczek, A. Kato, K. Nagaki, J. A. Birchler, J. Jiang, R. K. Dawe, Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* **14**, 2825–2836 (2002).
72. J.-W. A. Hsieh, P.-Y. Lin, C.-T. Wang, Y.-J. Lee, P. Chang, R. J.-H. Lu, P.-Y. Chen, C.-J. R. Wang, Establishing an optimized ATAC-seq protocol for the maize. *Front. Plant Sci.* **15**, 1370618 (2024).
73. J. Heslop-Harrison, Y. Heslop-Harrison, Evaluation of pollen viability by enzymatically induced fluorescence; Intracellular hydrolysis of fluorescein diacetate. *Stain Technol.* **45**, 115–120 (1970).
74. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
76. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
77. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
78. W. Guo, P. Fiziev, W. Yan, S. Cokus, X. Sun, M. Q. Zhang, P.-Y. Chen, M. Pellegrini, BS-Seeker2: A versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774 (2013).
79. W. Guo, P. Zhu, M. Pellegrini, M. Q. Zhang, X. Wang, Z. Ni, CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**, 381–387 (2018).
80. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgments: We thank D. w. Kim for technical contributions throughout this study, including genotyping, CUT&Tag assays, and library preparation. We thank M. Brady for statistical help with recombination data. **Funding:** This work was funded by National Science Foundation grant no. 2040218 (R.K.D.). **Author contributions:** Writing—original draft: R.K.D., J.I.G., and Y.Z. Conceptualization: R.K.D., Y.Z., and J.I.G. Investigation: Y.Z., M.W., J.I.G., and R.K.D. Writing—review and editing: R.K.D., J.I.G., Y.Z., and M.W. Methodology: Y.Z., J.I.G., and R.K.D. Resources: R.K.D. Funding acquisition: R.K.D. Data curation: Y.Z., J.I.G., and M.W. Validation: Y.Z., M.W., and R.K.D. Formal analysis: Y.Z. and M.W. Software: Y.Z. and M.W. Supervision: R.K.D. and J.I.G. Project administration: R.K.D. and J.I.G. Visualization: Y.Z., J.I.G., and R.K.D. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The sequencing data generated in this study are available at the NCBI BioProject database (www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA874319. The SRA accession numbers used for each experiment are listed in table S5. The ABS4 reference genome and annotation are available at https://zenodo.org/records/14501890. All code is available at https://github.com/dawelab/Centromere_Engineering and archived at https://zenodo.org/records/15098108. All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 28 January 2025
Accepted 17 April 2025
Published 21 May 2025
10.1126/sciadv.adw3433