nature plants

Article

https://doi.org/10.1038/s41477-023-01370-8

# Synthetic maize centromeres transmit chromosomes across generations

Received: 9 October 2022

Accepted: 10 February 2023

Published online: 16 March 2023

Check for updates

R. Kelly Dawe[1,2,3] ✉, Jonathan I. Gent[2], Yibing Zeng[1], Han Zhang[1], Fang-Fang Fu[2,4], Kyle W. Swentowsky[2], Dong Won Kim[2], Na Wang[2,5], Jianing Liu[1] & Rebecca D. Piri[3]

Centromeres are long, often repetitive regions of genomes that bind kinetochore proteins and ensure normal chromosome segregation. Engineering centromeres that function in vivo has proven to be difficult. Here we describe a tethering approach that activates functional maize centromeres at synthetic sequence arrays. A LexA-CENH3 fusion protein was used to recruit native Centromeric Histone H3 (CENH3) to long arrays of LexO repeats on a chromosome arm. Newly recruited CENH3 was sufficient to organize functional kinetochores that caused chromosome breakage, releasing chromosome fragments that were passed through meiosis and into progeny. Several fragments formed independent neochromosomes with centromeres localized over the LexO repeat arrays. The new centromeres were self-sustaining and transmitted neochromosomes to subsequent generations in the absence of the LexA-CENH3 activator. Our results demonstrate the feasibility of using synthetic centromeres for karyotype engineering applications.

Decades of advances in genetics have led to a strong awareness that traits related to the most pressing problems in agriculture are polygenic. Information about the combined effects of hundreds of genes known to be involved in drought, heat and pathogen pressure[1] might be used to substantially improve plant performance, but we lack the methods necessary to genetically engineer crops at this scale. Many have suggested that synthetic biology approaches could be used to meet this challenge (for example, refs. [2–4]). In bacteria and yeast, it is possible to chemically synthesize megabase-scale sequences and shuttle them between species to alter gene order and structure[4–7]. Similar methods have been used to move DNA from yeast into human cells[4] and could be adapted to plants as well[8]. Engineering small chromosomes for practical applications in medicine and agriculture seems a reasonable near-term goal[8–10].

A major obstacle in whole chromosome design is the synthesis of functional centromeres—sites of kinetochore assembly that are required for chromosome segregation. Centromeres can span hundreds of kilobases of repetitive, ill-defined sequence. Centromere specification is also strongly epigenetic, such that the DNA sequence matters less than the centromere proteins that are stably bound to the sequence[11]. Early studies established that long tracts of native centromeric DNA can form artificial chromosomes in human cell lines[12], but success rates are low and rely on the presence of a primate-specific centromere protein (CENP-B)[13]. Many have suggested that centromere engineering might be better accomplished via protein tethering[13–20]. In this approach, a cell line, animal or plant is first transformed with an array of LacO DNA binding sites that provide a platform to build a new centromere. The line is then transformed with a gene that encodes a fusion between the corresponding DNA binding protein (LacI) and one of several proteins that associate with native centromeres[14,15,17,18,21]. Bound by the fusion proteins, the synthetic DNA recruits additional kinetochore proteins and shows features of a functional centromere[14,15,17,21]. The strengths of the tethering approach are that it relies on defined components and is broadly applicable to all species.

[1]Department of Genetics, University of Georgia, Athens, GA, USA. [2]Department of Plant Biology, University of Georgia, Athens, GA, USA. [3]Institute of Bioinformatics, University of Georgia, Athens, GA, USA. [4]Present address: Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China. [5]Present address: Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai, China. ✉e-mail: kdawe@uga.edu

https://doi.org/10.1038/s41477-023-01370-8

The most effective synthetic centromeres have been based on recruiting CENP-A/CENH3 to arrays of 256 LacO binding sites[22]. CENP-A/CENH3, a histone H3 variant, is the defining feature of centromere-specific nucleosomes[23]. In humans and *Drosophila*, CENP-A is deposited by specialized chaperones called HJURP and Cal1 (refs. [24,25]). When LacI-CENP-A, LacI-HJURP or LacI-Cal1 fusion proteins were introduced into cell lines carrying LacO arrays on chromosome arms, active centromeres formed over the LacO repeats[16,21,25]. The LacO sites (and often flanking sequences) retained kinetochore activity during mitosis even after the LacI fusion proteins were removed[19–21]. These data established that it is only necessary to seed centromere proteins at an ectopic site to activate the self-replication mechanism. An important test of synthetic centromeres in living organisms was recently carried out using *Drosophila*[20]. The authors inserted LacO arrays into several locations of the *Drosophila* genome and observed centromere activation by Cal1-CENP-A tethering at a high frequency in developing animals. However, the resulting chromosome instability led to apoptosis and loss of lineages containing the new centromeres[20], making it difficult to interpret whether the newly activated centromeres could independently control chromosome segregation or be transmitted through meiosis into progeny.

## Characterization of the synthetic centromere platform ABS4

We previously prepared long synthetic repeat arrays composed of 157 bp monomers that contain binding sites for several known DNA binding proteins[26]. Each monomer contains binding motifs for *Escherichia coli* LacI and LexA and budding yeast Gal4. The arrays were biolistically transformed into maize and recovered in three locations on different chromosome arms. The longest of these is Arrayed Binding Sites on chromosome 4 (ABS4), which is composed of a mixture of 157 bp ABS monomers, a marker plasmid called pAHC25, and possibly other re-arranged genomic sequences[26]. For this study, we sequenced a line containing ABS4 to determine the location of the array and copy number of ABS monomers within it. The data suggest that ABS4 contains ~2,400 copies of ABS (about 377 kb) and ~10.5 copies of pAHC25 (about 101 kb) (Fig. 1a). We detected junctions between the transformed sequences and chromosome 4 at two locations, one at position 183583798 and another at position 185020754. While biolistic transformation can create complex re-arrangements[27], most or all of the DNA between these insertion points is present (see below) and plants homozygous for ABS4 appear healthy, suggesting that few or no genes were damaged during transformation. A complete assembly of an ABS4 genome will be required to accurately interpret the structure of the ABS4 locus. This location (183583798–185020754) is on the long arm of chromosome 4, roughly 44 cM from the native centromere.

## LexA-CENH3 recruits native CENH3 to ABS4

In *Drosophila* and human, centromere protein tethering has made use of the LacI DNA binding domain[14–17,19–21]. We opted against using LacI because of its large size (366 amino acids (a.a.)), as data from *Arabidopsis* suggested that large N-terminal additions to CENH3 could interfere with function[28]. We prepared constructs to test the efficacy of the Gal4 and the LexA DNA binding domains when fused to the N-terminus of maize CENH3. In each case, we also replaced the native N-terminal tail of maize CENH3 with the N-terminal tail from oat CENH3 (Fig. 1b). This modification removed the epitope for a maize CENH3 antibody[29] and added the epitope for an oat CENH3 antibody[30]. Removing the maize epitope from the transgene allowed us to test whether the transgene could recruit native CENH3 to the ABS4 array (Fig. 1c). Our first-generation (v1) constructs were composed of a ~1.7 kb segment of the native *CENH3* promoter driving the chimeric proteins as intronless open reading frames. The constructs were transformed into maize, crossed to lines carrying ABS loci (either ABS4 or ABS3 on chromosome 3; ref. [26]), and the heterozygous progeny subjected to chromatin immunoprecipitation (ChIP) using the maize-specific CENH3 antibody. The results were assayed by both ChIP–quantitative real-time polymerase chain reaction (qPCR) and ChIP followed by sequencing (ChIP–seq) (Fig. 1d and Extended Data Fig. 1). *LexA-CENH3(v1)* but not *Gal4-CENH3(v1)* showed enrichment of ABS (Fig. 1d). These results demonstrate that *LexA-CENH3(v1)* interacts with ABS and recruits native CENH3 to ABS sequences.

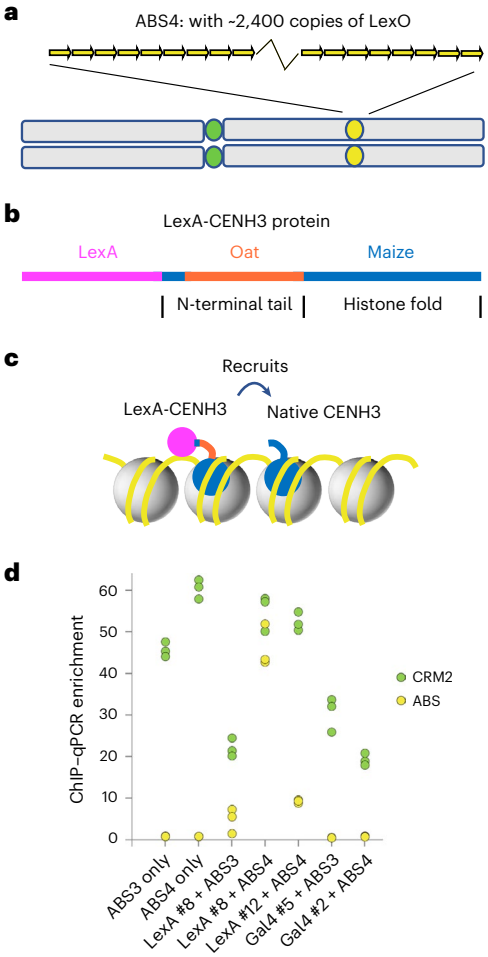



**Fig. 1 | *LexA-CENH3(v1)* recruits native CENH3 to ABS4. a**, Relative positions of centromere 4 (yellow) and ABS4 (green) on chromosome 4. **b**, Schematic of the LexA-CENH3 chimeric protein. **c**, Cartoon illustrating how LexA-CENH3 may interact with DNA and nucleosomes in ABS arrays. Not shown are other constitutive centromere proteins that bind to CENH3, some of which recruit additional CENH3 to maintain stable centromere positions. **d**, CENH3 ChIP enrichment relative to input, measured by qPCR in plants with *v1* constructs. *LexA-CENH3(v1)* is abbreviated to LexA and *Gal4-CENH3(v1)* is abbreviated to Gal4. Each data point represents one of three technical replicates. Two different transgenic events for each construct were tested (noted with #) along with ABS3 and ABS4. The CRM retroelement is common in native centromeres and is used as a positive control. *Y* axis indicates qPCR abundance using ABS- and CRM2-specific primers. The positive ChIP–PCR result for LexA #8 + ABS4 was confirmed by ChIP–seq (Extended Data Fig. 1).

## Re-engineering of the LexA-CENH3 transgene

Despite the strong ChIP enrichment of ABS in leaf tissue, the *LexA-CENH3(v1)* transgene was only weakly expressed in meiotic cells and did not rescue the embryo lethal phenotype of a homozygous *cenh3* null mutant (Extended Data Fig. 2). We had previously established that a ~5 kb region of the *CENH3* genomic sequence with a ~2.1 kb promoter and all introns could complement the *cenh3* null allele[31]. We modified this construct by inserting *LexA* and oat sequences into the *CENH3* coding sequence while retaining the order of and sequence of introns

https://doi.org/10.1038/s41477-023-01370-8

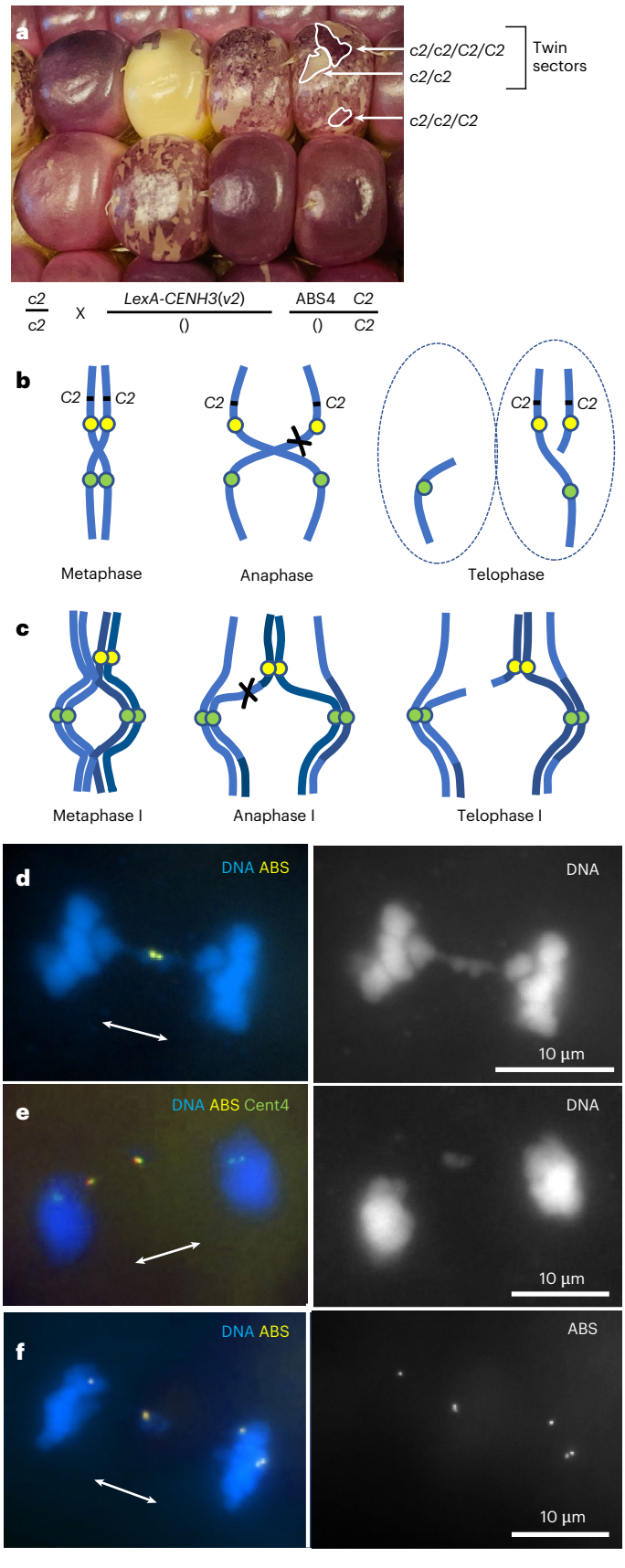



**Fig. 2 | Segregation errors in *LexA-CENH3(v2)* ABS4 plants. a**, Kernels from a cross where two copies of chromosome 4 (with *c2* null alleles) came from the female, and *LexA-CENH3(v2)* and one copy of chromosome 4 with ABS4 and *C2* came from the male (the cross is shown below the image). The *LexA-CENH3(v2)* transgene is unlinked to ABS4. Kernels that received ABS4 and *LexA-CENH3(v2)* are sectored, showing patches of colourless and deeply pigmented tissue, often in twin sectors. **b**, Dicentric chromosome segregation in mitosis. If only one chromatid breaks, the result will be a daughter cell with no ABS4 (or the closely linked *C2* locus) and a daughter cell with two copies of ABS4 (and two copies of *C2*). ABS4 is shown in yellow and Cent4 in green. This model is consistent with the sectors in **a** and cytological observations of roots grown from sectored kernels (Extended Data Fig. 4). **c**, Dicentric chromosome segregation in meiosis I. Sister centromere cohesion can restrain ABS4 to form a bridge. This can cause chromosome breakage if there is recombination between the native centromere and ABS. The parental chromosomes are in slightly different shades of blue. **d**, Anaphase I showing a bridge with ABS at the connecting point. Right image shows DNA. **e**, Telophase I cell showing apparent breakage and release of an ABS-containing fragment in the midzone. Cent4 is labelled green (one on each sister centromere) showing that ABS is separated from the native centromere. Right image shows DNA. **f**, Late anaphase I cell showing six ABS loci, where two are expected. The accumulation of ABS loci must have occurred before meiosis. Right image shows ABS loci.

(*LexA-CENH3(v2)*). Plants carrying the *LexA-CENH3(v2)* transgene and homozygous for the *cenh3* null allele were viable (Extended Data Fig. 2), although three of the four plants were thin and weak. The fourth plant was vigorous; it was self-crossed and proved to be homozygous for *LexA-CENH3(v2)* (Extended Data Fig. 3a). Protein immunoblot analysis of this plant demonstrated that native maize CENH3 protein was absent, while the larger LexA-CENH3(v2) protein, identified by the oat CENH3 epitope, was readily detectable (Extended Data Fig. 3b). These results show that *LexA-CENH3(v2)* can substitute for native *CENH3*, is properly expressed in all tissues and does not interfere with the assembly of functional kinetochores.

## Mitotic chromosome breakage in ABS4/LexA-CENH3 lines

Barbara Mclintock was the first to describe the cytological and genetic behaviour of dicentric chromosomes[32,33]. When located far apart on a chromosome, two centromeres are just as likely to move to opposite poles on the spindle as they are to move to the same pole. The opposite orientation results in criss-cross patterns with chromatin bridges that can rupture in the spindle midzone, creating double-stranded DNA breaks[34,35]. The breaks can then fuse by non-homologous end joining in different configurations, often reforming a dicentric chromosome that will break again in the next cell cycle[33]. Over time the breaks can heal by acquiring telomeres, separating the previously dicentric chromosome into two independent chromosomes[33,36].

A readily scorable colour gene, *colorless2* (*C2*), is located about 12 cM from ABS4 on the telomere side. *C2* confers purple pigmentation to the outer cell layers (aleurone) of the starchy part of the seed (endosperm). Unlike the embryo, the endosperm is triploid, where two complete genomes are contributed by the female and one from male. The intensity of the purple colour conferred by a wild-type *C2* allele is dosage dependent, such that the seeds are colourless in the absence of *C2*, one copy of *C2* gives a faint colour and two copies a darker colour[37]. In control crosses between a line homozygous for a recessive mutant *c2* allele (*c2*/*c2*) as a female and a wild-type (*C2*/*C2*) line as a male, the seeds are a uniformly faint purple colour. Likewise, the seeds were uniformly pigmented when we crossed *c2*/*c2* lines by *C2*/*C2* plants carrying ABS4 or *LexA-CENH3(v2)* alone.

When we crossed *c2*/*c2* females by *C2*/*C2* males heterozygous for both *LexA-CENH3(v2)* and ABS4, a subset of the kernels showed sectors of colourless, pale, and deeply pigmented aleurone (Fig. 2a). If ABS4 is activated to become a second functional centromere by *LexA-CENH3(v2)*, we expect the dicentric chromosomes to form criss-cross patterns that result in chromatid breakage. However, prior data from *Haemanthus* and wheat indicate that in some cases one centromere overcomes the other and pulls the entire chromatid to a single pole[34,38] (Fig. 2b). If one dicentric chromatid breaks and the other

does not, we expect that one of the two daughter cells will lack *C2* (and be colourless) and the other will contain two copies of *C2* (and be a dark colour). In fact we frequently observed twin sectors of colourless and dark aleurone juxtaposed on kernels (Fig. 2a). We also analysed roots from sectored seeds and observed that 4.2% of the nuclei lacked ABS and 5.7% had two copies of ABS (Extended Data Fig. 4a), further supporting this interpretation.

We also observed that plants heterozygous for *LexA-CENH3(v2)* and ABS4 tended to be short and have asymmetric leaves (Extended Data Fig. 5), suggesting that the loss and gain of genes on chromosome 4L negatively impacted plant development. It is also possible that the observed phenotypes are caused by the activation of cell cycle checkpoints that can delay or stop the division of heavily damaged cells[39]. Similar phenotypes were observed in pea plants carrying a dicentric chromosome[36].

### Meiotic chromosome breakage in ABS4/LexA-CENH3 lines

In meiosis I, criss-cross patterns are not expected because the sister centromeres remain attached throughout the first division[40]. Nevertheless, because ABS4 is ~44 cM from the native centromere, recombination will frequently place one copy of ABS4 on each chromosome (Fig. 2c). When this occurs, sister centromere cohesion will prevent complete disjunction, resulting in bridge structures where the arms remain attached specifically at the ABS locus. Bridges of this type were readily observed at mid-anaphase I (Fig. 2d). An analysis of 68 mid-anaphase I cells revealed that 63% of the recombinant chromosomes remained connected at ABS4 (Extended Data Fig. 4b). At late anaphase I, the bridges were resolved by pulling forces, either by breaking the cohesin bonds or by breaking one of the chromosome arms. Multiple examples of broken, ABS-containing fragments were observed in the midzone of late anaphase–telophase cells. A probe specific to centromere 4 (Cent4; ref. [41]) was used to demonstrate that ABS had separated from its native centromere in many cases (Fig. 2e). We also observed meiosis I cells with no ABS or multiple ABS loci caused by mitotic events that occurred during the preceding mitoses (Fig. 2f). However, few if any new fragments were produced during meiosis II, which appeared to proceed normally with no evidence of anaphase II bridges. These results suggest that LexA-CENH3-activated centromeres are of sufficient strength to break off fragments of chromosome 4L that are visible in meiosis.

### Identification of ABS-driven neochromosomes

Although many of the apparently broken chromosome fragments we observed in meiosis were probably lost or degraded, many more were probably included in telophase nuclei where they could be transmitted to the next generation. The fragments presumably also carried CENH3 or other chromatin marks that could enable ABS4 to serve as the centromere for a new chromosome (neochromosome). However, any neochromosome derived from chromosome 4L will lack hundreds of genes that are normally contributed by chromosome 4S. Since many of those genes are required for gametophyte development[42], such fragments cannot be transmitted to the next generation unless they are accompanied by a normal chromosome 4, which rescues the deficiency but also adds a second copy of 4L. The changes in gene dosage caused by adding an extra chromosome arm can likewise be detrimental to gametophytes, particularly through the male because pollen tubes compete with each other to reach the egg cell. Assuming the neochromosome can be successfully passed through gametophytes, the resulting progeny will be partially trisomic. Trisomic maize plants usually survive to maturity, but three chromosomes cannot pair properly in meiosis, causing errors in segregation that can further reduce the likelihood of passing on the chromosome[43]. These factors can make it difficult to recover and propagate neochromosomes. Nevertheless, we have identified independent ABS chromosomes using two different approaches.

The first was discovered in the progeny of a cross between a *c2*/*c2* female and a *C2*/*C2* male heterozygous for both *LexA-CENH3(v2)* and ABS4. We noticed that some kernels were sectored but lacked *LexA-CENH3(v2)*, suggesting that ABS had acquired independent centromere activity. One such sectored (C2/c2) kernel was planted and crossed to *c2*/*c2* in both directions, where it transmitted *C2* at a low frequency of 7% through the female and 11% through the male. However, most of the seedlings that grew from purple kernels did not have ABS when scored by PCR (14/17), suggesting that the ABS-containing chromosome was unstable. One plant that retained ABS was crossed again in both directions, where it showed improved transmission of *C2* (35% and 22% respectively). Root tips from the progeny of these crosses were analysed cytologically and found to be partially trisomic, with two copies of normal chromosome 4 and a fragment chromosome with ABS that we refer to as Neochromosome 4L number 1 (Neo4L-1) (Fig. 3a).

Three other neochromosomes were identified in a screen for seeds that had received two copies of chromosome 4L from a *LexA-CENH3(v2)* ABS4 parent. For this experiment, we obtained a fluorescent null allele of *c2* (*c2-GFP*) from a public reverse genetics resource[44]. The *c2-GFP* allele has a stable *Dissociation* transposable element containing GFP inserted in the second exon of the *c2* allele. Plants of the genotype ABS4 *C2*/*c2-GFP LexA-CENH3(v2)* were crossed as females to a *c2*/*c2* line (Fig. 3b) and screened for kernels that were both purple and fluorescent (Extended Data Fig. 6). We observed 181 kernels that were both purple and fluorescent in a sample of 16,757 kernels, indicating that as many as 1% of the kernels from ABS lines may have potential neochromosomes (Supplementary Table 1). To test if the screen was working as intended, we first planted a sample of 3 purple/fluorescent seeds and then another sample of 20 purple/fluorescent seeds (Supplementary Table 1). Of these 23 plants, 17 lacked ABS, suggesting that if neochromosomes were present they were unstable. Two of the six plants that retained ABS grew poorly and could not be crossed, and four others were crossed in both directions to *c2*/*c2*. One showed nearly Mendelian segregation of *C2* and did not have a visible neochromosome (this may be a translocation). Three others transmitted *C2* at low frequencies, and progeny contained partially trisomic individuals with two copies of normal chromosome 4 and ABS-containing neochromosomes that we named Neo4L-2, Neo4L-3 and Neo4L-4 (Fig. 3a).

Each of the four chromosomes displayed cytological constrictions at the ABS4 array (Fig. 3a), which is a classic indicator of centromere location. To confirm that centromeres had formed over ABS4, we carried out CENH3 CUT&Tag[45] on young ear tissue. The data show that CENH3 is localized over ABS and pAHC25 on all four neochromosomes (Fig. 3c). We also observed in all cases that CENH3 had spread into chromatin flanking the known insertion points of the ABS4 array (Fig. 3d). The least amount of spreading was observed on Neo4L-1, where CENH3 was observed over ~350 kb of chromatin flanking one insertion point, and the most spreading was observed on Neo4L-3, where there was a low level of CENH3 across ~1,750 kb of chromatin between the two insertions points (Fig. 3d).

### Genetic properties of neochromosomes

Transmission data are available for all four neochromosomes (Extended Data Fig. 7 and Extended Data Table 1). Neo4L-1, Neo4L-3 and Neo4L-4 were discovered in plants that lacked *LexA-CENH3(v2)*, indicating that the neochromosomes had acquired stable centromeres very quickly after meiosis. The first generation Neo4L-2 plant carried *LexA-CENH3(v2)*, but we chose plants that lacked *LexA-CENH3(v2)* for crossing in the second generation. The data are similar for all neochromosomes, showing 21–35% transmission through the female and variable frequencies through the male (0–22%, probably being influenced by how much pollen was applied, since excess pollen increases the amount of pollen competition and could select against partial trisomies (Extended Data Table 1)). These

https://doi.org/10.1038/s41477-023-01370-8

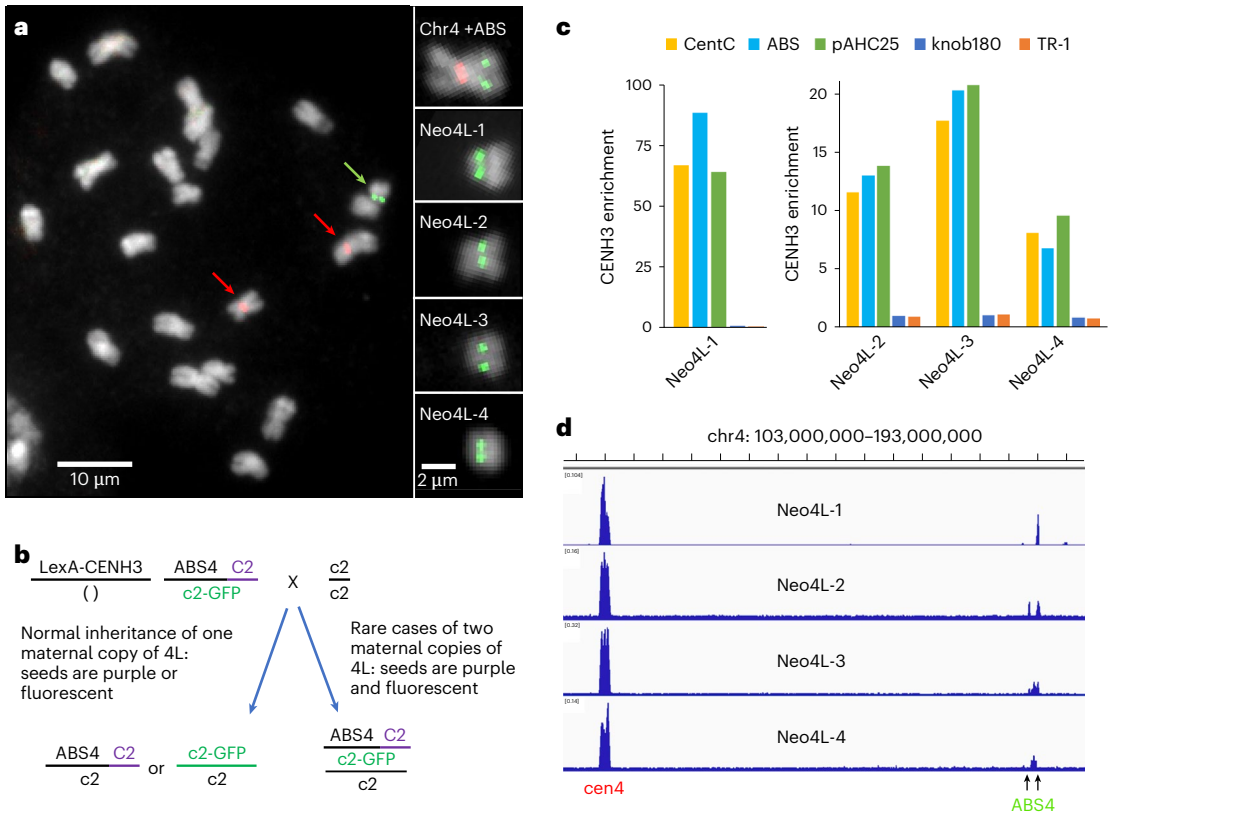



**Fig. 3 | Analysis of independent neochromosomes 4L. a**, Chromosome 4L neochromosomes. The main panel shows a full karyotype and typical partial trisomy of these lines, with two copies of normal chromosome 4 (red arrows) and Neo4L-1 (green arrow). To the right are images of the intact ABS4-containing chromosome and four neochromosomes. Neo4L-1 has a knob on the arm that makes it appear larger than the others. Cent4 is labelled in red and ABS in green. **b**, The genetic cross used to identify Neo4L-1, Neo4L-2 and Neo4L-3. Purple/fluorescent kernels should be partially trisomic. **c**, CENH3 CUT&Tag results from each neochromosome showing enrichment for ABS and pAHC25. Neo4L-1 is shown with a different *Y*-axis scale because this experiment was performed with a different batch of reagents than the other three and the enrichment values were higher. **d**, CENH3 peaks over the long arm of chromosome 4 (of the B73 reference genome). The data show enrichment of CENH3 over the native centromere (cen4) and regions flanking the insertion points of the ABS4 array. CENH3 CUT&Tag enrichment extends (roughly) over the following regions: Neo4L-1, 184,650,000-185,000,000; Neo4L-2, 183,250,000-183,450,000 and 184,750,000-185,150,000; Neo4L-3, 183,250,000-185,000,000; and Neo4L-4, 183,700,000-184600000. The two known insertion points of the ABS array are 183583798 and 185020754 (arrows).

results demonstrate that, after a synthetic centromere is activated and stabilized during plant growth, the *LexA-CENH3(v2)* transgene is not necessary for future transmission.

Although all four neochromosomes were successfully passed through meiosis, none was completely stable in somatic tissues during the period of this study. When the neochromosomes (marked by *C2*) were passed through the male to a *c2*/*c2* line, there were usually kernels with colourless sectors, which we interpret as occasional chromosome loss. We also observed that about 31% of the seedlings grown from seeds with coloured (*C2*) aleurone lacked ABS4 when genotyped (Extended Data Table 1). This result is not a result of genetic recombination between ABS4 and *C2*. We assayed 87 colourless seeds from trisomic plants segregating neochromosomes and found only one with ABS, or about ~1% recombination (Extended Data Table 1). A likely explanation is that newly formed centromeres are particularly unstable in young embryos. Prior data demonstrate that when plants with small/defective centromeres are crossed to plants with larger/normal centromeres, the chromosomes from the small/defective centromere parent are lost at a high frequency during early development[28,46–48]. This process, called centromere-mediated genome elimination, has been demonstrated in *Arabidopsis*, barley, wheat and maize[31,46,48,49]. Centromere-mediated genome elimination occurs primarily in embryos (and to a far lesser extent in endosperm[31,50]), mirroring what we have observed with maize neochromosome loss in early generations.

## Discussion

Discussions of whole genome synthetic biology often leave out the issue of recreating centromeres, which are hundreds of kilobases long and generally do not function properly unless they are bound to CENP-A/CENH3 and/or other constitutive centromere proteins. Less appreciated is the flip side of this observation, which is that apparently any non-genic sequence has the potential to become centromeric[51,52]. Even a short array of bacterial LacO repeats can cause centromere movement when tethered to one of several constitutive centromere proteins[14,15,17,21]. Using a LexA-based system in maize, we have demonstrated that the basic principles behind targeting CENP-A/CENH3 to a synthetic repeat array to activate mitotically stable centromeres[19–21] are applicable to the plant kingdom. We have further demonstrated that centromeres engineered in this way can become permanent features of a stably modified genome. Newly created maize synthetic centromeres can drive independent chromosome segregation over several generations in the absence of the original activator protein.

On native maize centromeres, CENH3 is sparsely distributed over regions as long as 2 Mb (ref. [53]). Here we recruited CENH3 to a ~377 kb ABS array and observed CENH3 spreading into linked plasmid sequences and long tracts (up to ~1,750 kb) of flanking chromatin. Yet, as we show here, a large binding platform does not assure stable chromosome inheritance, particularly immediately after a centromere is formed. The limited data from our genetic screen suggest that only

~13% (3/23) of the ABS-containing fragments that are passed through meiosis will become neochromosomes in progeny. We also observed neochromosome loss in later generations, which may reflect the activity of a centromere size surveillance mechanism[31,46], or other structural abnormalities associated with centromeres or telomeres in these newly formed chromosomes.

Data from immunolocalization studies indicate that functional centromere size is positively correlated with genome size[54,55]. Our success engineering centromeres in a large genome plant such as maize bodes well for success in smaller genome plants such as *Setaria viridis* or *Arabidopsis thaliana*, where smaller LexO arrays may be sufficient to form more stable centromeres. It may be possible to improve the frequency of neochromosome formation by combining de novo centromere formation with precise cleavage of chromosome arms by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 instead of relying on random chromosome breakage as described here. By this approach it might be possible to split a chromosome into two separate chromosomes, dramatically altering the recombination landscape and potentially liberating genetic variation that was previously locked into low recombination regions. CRISPR/Cas9 has already been used to create inversions that alter recombination for this purpose[56–58]. It may also be possible to extend our general approach to create circular plant artificial chromosomes similar to the artificial chromosomes built by CENP-A tethering in animal cell lines[19]. When paired with continued efforts to improve large molecule transformation in plants, synthetic centromeres could become important tools for addressing global challenges related to the bioeconomy and food security.

## Methods

### Molecular characterization of ABS4

To create the ABS4 locus, amplified arrays of ABS monomers and the selectable marker plasmid pAHC25 (ref. [59]) were transformed into maize using biolistic transformation[26]. The location of ABS4 was previously known only by cytological position. To determine the exact location of the ABS4 array, we performed paired-end 150 Illumina sequencing at 16× coverage of an ABS4 heterozygous line. pAHC25 and ABS monomer sequences were concatenated with the B73 genome to create a reference for mapping. The reads were aligned with the Burrows–Wheeler Aligner[60] and the discordant and soft-clipped alignments in pAHC25 and ABS sequences were screened (by eye) using the Integrative Genomics Viewer. The coordinates of reads mapped to B73 were manually identified in the discordant read pair and clustered on the basis of their position. We then carried out de novo assembly with discordant read pairs from individual clusters using SPAdes (v3.10.0). Upon the generation of consensus sequences, we identified the precise junction points by blasting (v2.2.26) the assembled short segments against the reference sequence. By this means, we identified two junctions on chromosome 4: one between pAHC25 and Chr4 position at position 185020754 and another between ABS and Chr4 position 183583798 (in Zm-B73-REFERENCE-NAM-5.0 coordinates). We assume these represent independent insertions and there are at least four junctions, two on either side of each insertion, but we detected only two using this approach. Primers were developed to detect the second polymorphism for genotyping. One pair, abs4-p1 (TACCCTGGTTAGAGGGAGCC) and abs4-p4 (AGCCAGGCGGATAGAAGC) amplify the wild-type locus. Two pairs amplify the ABS insertion: abs4-p1 and abs4-p2 (ATGCAGTCGCCGAATACTGT), and abs4-p3 (TCCTCCGGAGTACCGTCT) and abs4-p4. To determine the copy number of ABS monomers, the read depth of both ABS and the plasmid pACH25 was calculated as the mean of per-base coverage using BEDtools[61] and multiplied by two because ABS was heterozygous. The data, with a MAPQ filter of 20, indicated an ABS copy number of 2,401, and a pACH25 copy number of 10.5.

The *C2* locus is at 198744424 on chromosome 4. The genetic distance between ABS4 and *C2* is about 12 cM (ref. [62]), as judged by the distance between *C2* and the *tunicate1* gene (which is close to ABS4 at position 183877026).

### Transgenic lines

The *LexA-CENH3(v1)* and *Gal4-CENH3(v1)* constructs were prepared as intronless open reading frames, driven by the maize *CENH3* promoter. The constructs encoded 72 a.a. from LexA[63], based on the crystal structure by ref. [64], or 74 a.a. from Gal4 (ref. [65]). Fused on the C-terminal side of DNA binding domains were 11 a.a. of maize sequence, 60 a.a. of oat CenH3 sequence, and the remaining histone fold domain of native maize CENH3. The sequences were codon optimized for maize, commercially synthesized by DNA2.0 (now ATUM), and recombined into pEarleygate vector pEG100 (ref. [66]). The 35S promoter from this vector was then replaced with 1,714 bp of the native CENH3 endogenous promoter (using the enzymes BstBI and XhoI) from the inbred HiII (this part of HIII genome is derived from the inbred A188) using the primers 5′-TTCGAAGGCAATTGCAGTAGTGCCT-3′ and 5′-CTCGAGCGCGGTGGGCGCCTCGCA-3′. For full annotated sequences, see Supplementary Data 1.

The *LexA-CENH3(v2)* construct was based on the *ImmuneCENH3* construct described in ref. [31]. The guide RNA and ZmPolIII promoter in *ImmuneCENH3* were removed. Then a portion of the gene between the XbaI and BamHI sites was resynthesized by GenScript (www.genscript.com) to include the LexA and oat sequences. The three introns that naturally occur in the sequence that encodes the *CENH3* N-terminal tail were distributed within the LexA and oat sequences at positions where splice site acceptor and donor sequences could be accommodated without altering the encoded amino acid sequence (Supplementary Data 1). All constructs were transformed into the maize line HiII by the Iowa State University Plant Transformation Facility (http://www.agron.iastate.edu/ptf/).

The *c2-gfp* allele, identified by ref. [44], was ordered as stock tdsgR64H07 from the Maize Genetics Cooperation Stock Center in Urbana, IL. It contains a *trDs** element in the second exon of the *c2* gene. Fluorescence was viewed using a blue light source and orange filter (Clare Chemical HL34) and imaged using an iPhone. Plants were grown in the University of Georgia Plant Biology greenhouses.

### ChIP and CUT&Tag

ChIP was carried out on individual whole plants as described previously[67] using anti-CENH3 (maize) antibodies[29]. Inputs for each ChIP sample were collected after micrococcal nuclease digestion but before addition of antibodies. qPCR was carried out in 96-well plates (Bio-Rad #HSP9601) in 15 µl volumes with 2 µl of DNA (0.25 ng µl$^{-1}$) as template, 1 µl primer (5 nM of each, forward and reverse), 4.5 µl water, and 7.5 µl SYBR Select Master Mix for CFX (Applied Biosystems #4472942). Real-time thermocycler conditions were as follows: An initial 50 °C for 2 min and 95 °C for 2 min, then 39 cycles of 95 °C for 15 s and 60 °C for 1 min. Primers JIG-261 (5′-CTAGGACTTGCTGGCTTCTATC-3′) and JIG-262 (5′-TCCCTTCTTCGTAAGCTCATTC-3′) amplified the centromeric retrotransposon CRM2, and primers syn3for_short (5′-TCCCTTCTTCGTAAGCTCATTC-3′) and syn3rev_short (5′-TCCGGAGGACAGTCCTCC-3′) amplified ABS. Enrichments were measured as Cq values of each primer pair relative to Cq values from ZmCopia primers (Diagenode #C17040005), which amplified a Copia retrotransposon that is not enriched in centromeres. Each ChIP and input sample was amplified in triplicate for technical replication. For ChIP–seq, a single ChIP and single input library was prepared from each of three plants using a KAPA HyperPrep kit (Roche #KK8502), amplified with five cycles of PCR, and Illumina sequenced single-end with 150 nt read lengths. Reads were trimmed of adapters using Cutadapt (v2.8)[68] with parameters -q 20 -a AGATCGGAAGAGC -e .05 -O 1 -m 50 and mapped to the Zm-B73-REFERENCE-NAM-5.0 assembly using the Burrows–Wheeler Aligner BWA-MEM (v7.17) with default parameters[60]. Reads were aligned to CentC, knob180, TR-1 and ABS

consensus sequence dimers[67,69] to identify matching reads using blastall with parameters as follows: -p blastn -e 1e-5 -W 7 -G 2 -E 1 -r 1 -q -1. The ABS sequence is CCAACGAAAATTTCATCCTACTAATTGTGAGCCGCTCACAATTCCAGTAGAAGGCTTACTGTATATATATACAGTATTCGGCGACTGCATCTCTATCACTGATAGGGACTAGCGGTACTCCGGAGGACAGTCCTCCGAGAAGGCATATGTCCCGTTT. For pAHC25 read alignments, BWA-MEM was used with default parameters.

For CUT&Tag, nuclei were extracted from 3 ml or less finely ground, frozen unfertilized ears, ranging from 3.5 to 6.5 cm in length. A sucrose cushion method of nuclei extraction was used to separate nuclei from abundant amyloplasts[70]. Rather than diluting the nuclei suspension from the Percoll/sucrose cushion interface in Nucleus Extraction Buffer, the nuclei suspension was diluted in a 1× volume of 1× Wash Buffer from a CUT&Tag-IT Assay Kit (Active Motif 53160). To estimate the yield of nuclei, 50 µl was combined with 50 µl of water and sonicated with a Diagenode Bioruptor for 10 min on high setting with 30 s on–off intervals. DNA concentration was measured using 1 µl of sonicated nuclei with a Qubit dsDNA HS Assay Kit (Invitrogen Q32854). Intact nuclei (not sonicated) containing 150 ng of DNA for leaf or 350 ng for ear were then diluted to a total volume of 1,500 µl in 1× Wash Buffer and used as input for the CUT&Tag-IT Assay Kit with both anti-CENH3 (maize) antibodies[29] and IgG for each sample (Epicypher 13-0042). Libraries were amplified with 15 or 16 cycles of PCR, Illumina sequenced paired-end 150 nt, and analysed in the same way as the ChIP–seq libraries (except in paired end mode), and a different adapter sequence was trimmed (CTGTCTCTTATACACATCT).

### Protein extraction and immunoblot analysis

Developing ears approximately 3–4 cm in length were flash-frozen and ground in liquid nitrogen. Total protein was extracted by adding 1 ml of 2× Laemmli buffer (4% SDS, 20% glycerol and 120 mM Tris–Cl, pH 6.8), incubating for 5 min at 4 °C with gentle rotation, and centrifuging at 15,000*g* for 5 min at 4 °C. The supernatant was retained and the centrifugation step was repeated. The total protein concentration for each sample was measured using the Qubit Protein Assay Kit (Thermo cat. no. Q33211). Approximately 20 µg of each sample were loaded on a 4–20% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad cat. no. 4561094). Proteins were transferred to a nitrocellulose membrane (Bio-Rad cat. no. 1620145), and the membranes were blocked using TBST containing 5% powdered milk. Rabbit antibodies against maize CENH3 (ref. [29]) or oat CENH3 (ref. [30]) were added to a final dilution of 1:1,000 and membranes incubated in primary antibody solution overnight at 4 °C with gentle mixing. The maize CENH3 antibodies were previously validated by protein blot analysis and immunolocalization[29]. The oat CENH3 antibodies were previously validated by immunolocalization[30]. Membranes were washed three times with TBST and incubated with blocking buffer containing Rabbit IgG HRP Linked Whole Ab (Sigma cat no. GENA934-1ML) at 1:5000 for two hours with gentle mixing. After three final washes in TBST, SuperSignal West Dura Extended Duration Substrate (Thermo Fisher cat. no. 34075) was applied to membranes and chemiluminescence was visualized by exposing to X-Ray Film (Research Products International cat. no. 248300) in a dark room. Duplicate samples run on the same gels were used for total protein staining with QC Colloidal Coomassie Stain (Bio-Rad cat. no. 1610803).

### Fluorescent in situ hybridization

Male meiocytes were prepared for cytological analysis as described in ref. [71]. Tassels were staged such that anthers containing cells undergoing meiosis were isolated, and anthers were incubated in fixative solution (4% paraformaldehyde, 1% Triton X-100 diluted in PHEMS (30 mM PIPES, 12.5 mM HEPES, 5 mM EGTA, 1 mM $MgCl_2$ and 175 mM D-sorbitol, pH 6.8)) for 1 h. Anthers were washed three times in PBS following fixation. Meiocytes were extruded from anthers and immobilized onto polylysine-coated coverslips by centrifuging at 100*g* for 1 min then washed three times with PBS. Coverslips containing fixed meiocytes were incubated with 2× SSC (300 mM NaCl and 30 mM sodium citrate, pH 7.0) containing 30% formamide, then 2× SSC containing 50% formamide for 10 min each. Two different sets of labelled oligonucleotides were used for ABS, either four Alexa488-linked oligos (5′-TACCGCTAGTCCCTATCAGT-3′, 5′-CACAATTAGGATGAAAT-3′, 5′-TGTATATATATACAGTAAGC-3′, and 5′-CTTCTCGGAGGACTGTCCTC-3′) or eight Texas-Red-labelled oligos (5′-TACCGCTAGTCCCTATCA-3′, 5′-TGATAGAGATGCAGTCGC-3′, 5′-AATACTGTATATATATAC-3′, 5′-TAAGCCTTCTACTGGAAT-3′, 5′-TGAGCGGCTCACAATTAG-3′, 5′-AGGATGAAATTTTCGTTG-3′, 5′-AAACGGGACATATGCCTT-3′, 5′-CGGAGGACTGTCCTCCGG-3′). To detect the Cent4 repeat[41], a collection of six Cy5-labelled oligos were used (5′-ACCCTATGTATCGAAGGA-3′, 5′-CCACTAAAGAACCAAGAT-3′, 5′-CCAAGTAATAGTAAAATA-3′, 5′-GTACAATTTATCCAAACC-3′, 5′-TTAATAAATGTCTAGAGA-3′, 5′-ATGTGATTTGTGTCCAAC-3′). Probe hybridization solution (2× SSC, 50% formamide and each oligo at 1 µM concentration) was administered to meiocytes by suspending coverslips above slides using corners of broken coverslips. Sides of coverslip were sealed with rubber cement, and meiocytes were incubated at 95 °C for 5 min, then at room temperature overnight to facilitate probe hybridization. Coverslips were removed from slides and washed for 10 min each in four subsequent washing solutions: 2× SSC, 20% formamide, 0.01% Tween-20; 1× SSC, 10% formamide, 0.001% Tween-20; 1× SSC, 1× PBS; 1× PBS. Cells were stained with 0.1 µg $ml^{-1}$ DAPI and mounted in 1× PBS containing 100 mg $ml^{-1}$ of 1,4 diazobicyclo (2,2,2) octane.

Mitotic chromosomes were prepared for cytological analysis as described in ref. [72]. Roots from freshly germinated seeds were incubated in $N_2O$ gas for 3.5 h and then placed on ice. About 1 mm of tissue from the tips of roots was incubated in 2% cellulase Onozuka R-10 (Research Products International) and 1% pectolyase Y-23 (MP Biomedicals) 37 °C for 20–50 min to digest cell walls. The cells were rinsed twice in 100% ethanol, leaving ~4 µl of ethanol in the tube, which was then mixed with 20–30 µl of 100% acetic acid. A dissecting needle was used to break up the tissue, and 10 µl of cell suspension was added to slides and allowed to dry. The chromatin was fixed to the slides by UV crosslinking using a total energy of 200 mJ per square cm. Ten microlitres of probe hybridization solution (5 µl salmon sperm DNA (85 ng $µl^{-1}$ salmon sperm DNA in 2× SSC, 1× Tris–EDTA buffer), 4 µl 1× PBS, 0.5 µl of 10 µM Alexa488-labelled ABS oligos, and 0.5 µl of 10 µM Cy5-labelled Cent4 oligos) was added to slides and covered with a plastic coverslip. The slides were incubated at 95 °C in a metal incubator for 5 min, rinsed with PBS, and mounted with Drop-n-Stain EverBrite Mounting Medium with DAPI (BIOTIUM: 23009). All specimens were imaged using Zeiss Axio Imager.M1 fluorescence microscope with a 63× plan-apo Chromat oil objective. Data were analysed using Slidebook software (Intelligent Imaging Innovations).

### Statistics and reproducibility

We did not apply statistical tests in this study. The reproducibility of the key results was established as follows. To obtain the data shown in Fig. 2, meiotic cells from four plants with good representation of anaphase–telophase I were analysed in detail. The anaphase I segregation pattern shown in Fig. 2d, with bridges formed by the association of ABS at the midzone, was documented in three of these plants (the fourth had very few cells in mid-anaphase). Quantification of these data are shown in Extended Data Table 1. Images of telophase I cells of the type shown in Fig. 2e, with apparently broken ABS-positive fragments in the midzone, were collected from all four plants. Cells of this type were commonly observed. However, we did not count the total numbers of cells with and without fragments at telophase I. Images of meiosis I cells with more than two ABS signals (Fig. 2f) were collected from all four plants. Cells with this staining pattern were relatively rare, but when observed were often in clusters as would be expected if the causal

events occurred in preceding mitoses. The neocentromere karyotypes shown in Fig. 3a were confirmed by analysing at least five high-quality metaphase spreads.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA874319. For description of files, see Supplementary Table 2.

## References


1. Mittler, R. & Blumwald, E. Genetic engineering for modern agriculture: challenges and perspectives. *Annu. Rev. Plant Biol.* **61**, 443–462 (2010).
2. Baltes, N. J. & Voytas, D. F. Enabling plant synthetic biology through genome engineering. *Trends Biotechnol.* **33**, 120–131 (2015).
3. Wurtzel, E. T. et al. Revolutionizing agriculture with synthetic biology. *Nat. Plants* **5**, 1207–1210 (2019).
4. Venter, J. C., Glass, J. I., Hutchison, C. A. 3rd & Vashee, S. Synthetic chromosomes, genomes, viruses, and cells. *Cell* **185**, 2708–2724 (2022).
5. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
6. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
7. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
8. Dawe, R. K. Charting the path to fully synthetic plant chromosomes. *Exp. Cell. Res.* **390**, 111951 (2020).
9. Boeke, J. D. et al. GENOME ENGINEERING. The Genome Project-Write. *Science* **353**, 126–127 (2016).
10. Zhou, J. et al. Centromeres: from chromosome biology to biotechnology applications and synthetic genomes in plants. *Plant Biotechnol. J.* **20**, 2051–2063 (2022).
11. Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* **9**, 923–937 (2008).
12. Harrington, J. J., Van Bokkelen, G., Mays, R. W., Gustashaw, K. & Willard, H. F. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* **15**, 345–355 (1997).
13. Gambogi, C. W., Dawicki-McKenna, J. M., Logsdon, G. A. & Black, B. E. The unique kind of human artificial chromosome: bypassing the requirement for repetitive centromere DNA. *Exp. Cell. Res.* **391**, 111978 (2020).
14. Gascoigne, K. E. et al. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410–422 (2011).
15. Gascoigne, K. E. & Cheeseman, I. M. Induced dicentric chromosome formation promotes genomic rearrangements and tumorigenesis. *Chromosome Res.* **21**, 407–418 (2013).
16. Barnhart, M. C. et al. HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* **194**, 229–243 (2011).
17. Hori, T., Shang, W.-H., Takeuchi, K. & Fukagawa, T. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *J. Cell Biol.* **200**, 45–60 (2013).
18. Teo, C. H., Lermontova, I., Houben, A., Mette, M. F. & Schubert, I. De novo generation of plant centromeres at tandem repeats. *Chromosoma* **122**, 233–241 (2013).
19. Logsdon, G. A. et al. Human artificial chromosomes that bypass centromeric DNA. *Cell* **178**, 624–639.e19 (2019).
20. Palladino, J., Chavan, A., Sposato, A., Mason, T. D. & Mellone, B. G. Targeted de novo centromere formation in drosophila reveals plasticity and maintenance potential of CENP-A chromatin. *Dev. Cell* **52**, 379–394.e7 (2020).
21. Mendiburo, M. J., Padeken, J., Fülöp, S., Schepers, A. & Heun, P. *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**, 686–690 (2011).
22. Robinett, C. C. et al. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
23. Black, B. E. & Bassett, E. A. The histone variant CENP-A and centromere specification. *Curr. Opin. Cell Biol.* **20**, 91–100 (2008).
24. Dunleavy, E. M. et al. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497 (2009).
25. Chen, C.-C. et al. CAL1 is the *Drosophila* CENP-A assembly factor. *J. Cell Biol.* **204**, 313–329 (2014).
26. Zhang, H. et al. Stable integration of an engineered megabase repeat array into the maize genome. *Plant J.* **70**, 357–365 (2012).
27. Liu, J. et al. Genome-scale sequence disruption following biolistic transformation in rice and maize. *Plant Cell* **31**, 368–383 (2019).
28. Ravi, M. & Chan, S. W. L. Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618 (2010).
29. Zhong, C. X. et al. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* **14**, 2825–2836 (2002).
30. Topp, C. N. et al. Identification of a maize neocentromere in an oat-maize addition line. *Cytogenet. Genome Res.* **124**, 228–238 (2009).
31. Wang, N., Gent, J. I. & Kelly Dawe, R. Haploid induction by a maize cenh3 null mutant. *Sci. Adv.* **7**, eabe2299 (2021).
32. Mcclintock, B. The fusion of broken ends of sister half-chromatids following chromatid breakage at meiotic anaphases. *MO Agric. Exp. Station Res. Bull.* **290**, 1–48 (1938).
33. McClintock, B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
34. Bajer, A. Observations on dicentrics in living cells. *Chromosoma* **14**, 18–30 (1963).
35. Zheng, Y. Z., Roseman, R. R. & Carlson, W. R. Time course study of the chromosome-type breakage–fusion–bridge cycle in maize. *Genetics* **153**, 1435–1444 (1999).
36. Saccardo, F. Behaviour of dicentric chromosomes in peas. *Caryologia* **24**, 71–84 (1971).
37. Coe, E. H. Location and effects of c2. *Maize Genet. Coop. N. Lett.* **36**, 60 (1958).
38. Sears, E. R. & Câmara, A. A transmissible dicentric chromosome. *Genetics* **37**, 125–135 (1952).
39. Cools, T. & De Veylder, L. DNA stress checkpoint control and plant development. *Curr. Opin. Plant Biol.* **12**, 23–28 (2009).
40. Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N. & Grelon, M. The molecular biology of meiosis in plants. *Annu. Rev. Plant Biol.* **66**, 297–327 (2015).
41. Page, B. T., Wanous, M. K. & Birchler, J. A. Characterization of a maize chromosome 4 centromeric sequence: evidence for an evolutionary relationship with the B chromosome centromere. *Genetics* **159**, 291–302 (2001).
42. Nelms, B. & Walbot, V. Gametophyte genome activation occurs at pollen mitosis I in maize. *Science* **375**, 424–429 (2022).
43. Einset, J. Chromosome length in relation to transmission frequency of maize trisomes. *Genetics* **28**, 349–364 (1943).

44. Li, Y., Segal, G., Wang, Q. & Dooner, H. K. Gene tagging with engineered Ds elements in maize. *Methods Mol. Biol.* **1057**, 83–99 (2013).
45. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).
46. Marimuthu, M. P. A. et al. Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. *Sci. Adv.* **7**, eabk1151 (2021).
47. Ishii, T., Karimi-Ashtiyani, R. & Houben, A. Haploidization via chromosome elimination: means and mechanisms. *Annu. Rev. Plant Biol.* **67**, 421–438 (2016).
48. Wang, N. & Dawe, R. K. Centromere size and its relationship to haploid formation in plants. *Mol. Plant* **11**, 398–406 (2018).
49. Lv, J. et al. Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nat. Biotechnol.* **38**, 1397–1401 (2020).
50. Ravi, M. et al. A haploid genetics toolbox for *Arabidopsis thaliana*. *Nat. Commun.* **5**, 5334 (2014).
51. Lo, A. W. et al. A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J.* **20**, 2087–2096 (2001).
52. Liu, Y. et al. Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. *Proc. Natl Acad. Sci. USA* **112**, E1263–E1271 (2015).
53. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
54. Zhang, H. & Dawe, R. K. Total centromere size and genome size are strongly correlated in ten grass species. *Chromosome Res.* **20**, 403–412 (2012).
55. Plačková, K., Bureš, P. & Zedek, F. Centromere size scales with genome size across eukaryotes. *Sci. Rep.* **11**, 19811 (2021).
56. Schwartz, C. et al. CRISPR–Cas9-mediated 75.5-Mb inversion in maize. *Nat. Plants* **6**, 1427–1431 (2020).
57. Rönspies, M., Dorn, A., Schindele, P. & Puchta, H. CRISPR–Cas-mediated chromosome engineering for crop improvement and synthetic biology. *Nat. Plants* **7**, 566–573 (2021).
58. Schmidt, C. et al. Changing local recombination patterns in *Arabidopsis* by CRISPR/Cas mediated chromosome engineering. *Nat. Commun.* **11**, 4418 (2020).
59. Taylor, M. G., Vasil, V. & Vasil, I. K. Enhanced GUS gene expression in cereal/grass cell suspensions and immature embryos using the maize uhiquitin-based plasmid pAHC25. *Plant Cell Rep.* **12**, 491–495 (1993).
60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
61. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
62. Coe, E. H. Genetic maps 2007. *Maize Genet. Coop. N. Lett.* **82**, 87–102 (2008).
63. Zuo, J., Niu, Q.-W. & Chua, N.-H. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**, 265–273 (2000).
64. Fogh, R. H. et al. Solution structure of the LexA repressor DNA binding domain determined by $^{1}$H NMR spectroscopy. *EMBO J.* **13**, 3936–3944 (1994).
65. Padidam, M., Gore, M., Lu, D. L. & Smirnova, O. Chemical-inducible, ecdysone receptor-based gene expression system for plants. *Transgenic Res.* **12**, 101–109 (2003).
66. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
67. Gent, J. I., Wang, N. & Dawe, R. K. Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. *Genome Biol.* **18**, 121 (2017).
68. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
69. Liu, J. et al. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol.* **21**, 121 (2020).
70. Lee, Y.-J., Chang, P., Lu, J.-H., Chen, P.-Y. & Wang, C.-J. R. Assessing chromatin accessibility in maize using ATAC-seq. Preprint at *bioRxiv* https://doi.org/10.1101/526079 (2019).
71. Yu, H. G., Hiatt, E. N., Chan, A., Sweeney, M. & Dawe, R. K. Neocentromere-mediated chromosome movement in maize. *J. Cell Biol.* **139**, 831–840 (1997).
72. Kato, A., Lamb, J. C. & Birchler, J. A. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl Acad. Sci. USA* **101**, 13554–13559 (2004).


## Acknowledgements

We thank the Maize Genetics Cooperation Stock Center for providing maize stocks and M. Tindall Smith for genotyping stocks. This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, as well as grants from the National Science Foundation (IOS-1444514 and IOS-2040218).

## Author contributions

R.K.D., J.I.G. and H.Z. conceived and designed experiments. R.K.D., J.I.G., Y.Z., H.Z., F.-F.F., K.W.S., D.W.K., N.W., J.L. and R.D.P. performed experiments. R.K.D., J.I.G., Y.Z., F.-F.F., K.W.S., N.W., J.L. and R.D.P. analysed the data. R.K.D. and J.I.G. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

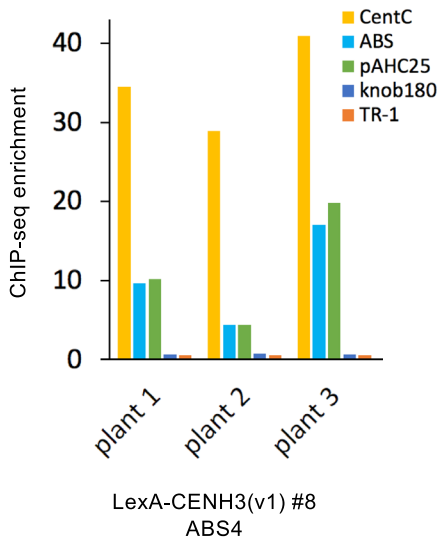
**Extended data** is available for this paper at https://doi.org/10.1038/s41477-023-01370-8.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41477-023-01370-8.

**Correspondence and requests for materials** should be addressed to R. Kelly Dawe.

**Peer review information** *Nature Plants* thanks Anne Britt, Ian Henderson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41477-023-01370-8



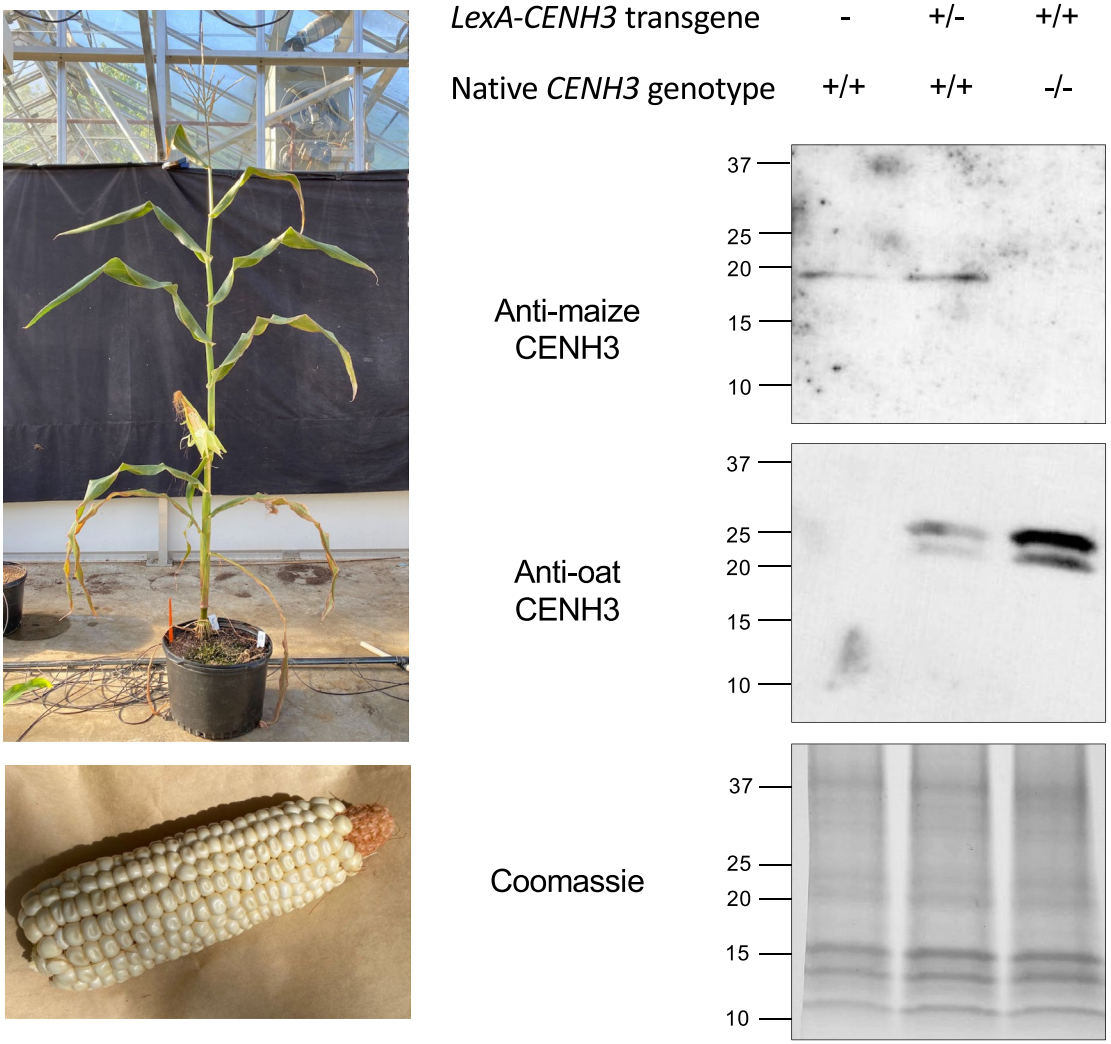

**Extended Data Fig. 1 | Confirmation that *LexA-CENH3(v1)* recruits native CENH3 by a ChIP-seq assay.** The data show CENH3 ChIP enrichment relative to input, measured by Illumina sequencing. Each ChIP is from an individual plant that was heterozygous for ABS4 and the *LexA-CENH3(v1)* transgene. CentC is a native centromere repeat whereas knob180 and TR-1 are non-centromeric repeats. The Y-axis indicates the numbers of reads in the ChIP sample divided by numbers in input and normalized by the total number of reads.

https://doi.org/10.1038/s41477-023-01370-8

| Cross | Transgene genotypes | *cenh3* genotypes |
|---|---|---|
| *-/LexA-CENH3(v1) +/cenh3* ⊗ | *-/LexA-CENH3(v1) or LexA-CENH3(v1)/LexA-CENH3(v1)* (**67**) | +/+ (33)<br>+/*cenh3* (34)<br>*cenh3*/*cenh3* (**0**) |
| | -/- (11) | |
| *-/LexA-CENH3(v2) +/cenh3* ⊗ | *-/LexA-CENH3(v2) or LexA-CENH3(v2)/LexA-CENH3(v2)* (**20**) | +/+ (7)<br>+/*cenh3* (9)<br>*cenh3*/*cenh3* (**4**) |
| | -/- (4) | |

**Extended Data Fig. 2 | Complementation of the *cenh3* null by *LexA-CENH3* transgenes.** Complementation of the *cenh3* null by *LexA-CENH3* transgenes. A minus sign indicates the lack of *LexA-CENH3* transgene. A plus sign indicates the wild-type *CENH3* allele. Plants heterozygous for either *LexA-CENH3(v1)* or *LexA-CENH3(v2)* were crossed to lines carrying the *cenh3* null and heterozygotes self-crossed. The genotypes of progeny are shown to the right. Note that of 67 plants carrying *LexA-CENH3(v1)*, none were homozygous for *cenh3*, indicating that the transgene cannot complement the null. In contrast, of 20 plants carrying *LexA-CENH3(v2)*, four were homozygous for *cenh3*. One of these four plants, which proved to be homozygous for *LexA-CENH3(v2)*, looked normal (see Extended Data Fig. 3).

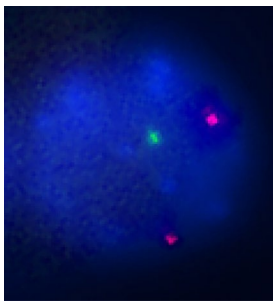
https://doi.org/10.1038/s41477-023-01370-8

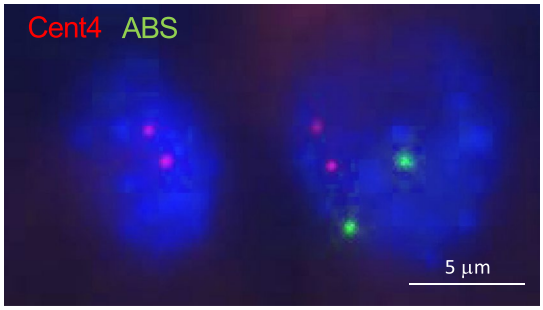



**Extended Data Fig. 3 | Demonstration that *LexA-CENH3(v2)* is sufficient for cell division and plant growth. a**). A plant homozygous for the *LexA-CENH3(v2)* transgene and homozygous for a *cenh3* null mutant. The plant was self-crossed to reveal a full ear (below), indicating that the transgene was homozygous (one quarter of the seeds would have been dead if the plant was heterozygous for *LexA-CENH3(v2)*). The second ear of this plant was used for the third lane of the protein blot in B. **b**) Protein blot analysis of CENH3 in plants of different genotypes. A minus sign indicates the lack of *LexA-CENH3(v2)* transgene. A plus sign indicates wild-type *CENH3* allele. Native maize CENH3 is 17 kDa and recognized by a maize-specific antibody. The oat CENH3 antibody recognizes the LexA-CENH3 fusion protein only. The predicted ~27 kDa LexA-CENH3 band is observed along with a smaller band of unknown cause/origin. Importantly, native CENH3 was not detectable in the *LexA-CENH3(v2)/LexA-CENH3(v2), cenh3/cenh3* null plant (third lane). The lower panel (Coomassie) shows that the amounts of protein loaded into each lane were similar. These staining patterns were observed in two different protein blots using the same protein samples.

https://doi.org/10.1038/s41477-023-01370-8

A






| | 1 ABS | 0 ABS | 2 ABS |
|---|---|---|---|
| Plant #1 | 56 | 8 | 6 |
| Plant #2 | 97 | 3 | 8 |
| Plant #3 | 138 | 5 | 3 |
| Plant #4 | 166 | 5 | 11 |
| Plant #5 | 84 | 3 | 5 |
| Total | 541 | 24 (4.2%) | 33 (5.7%) |

B

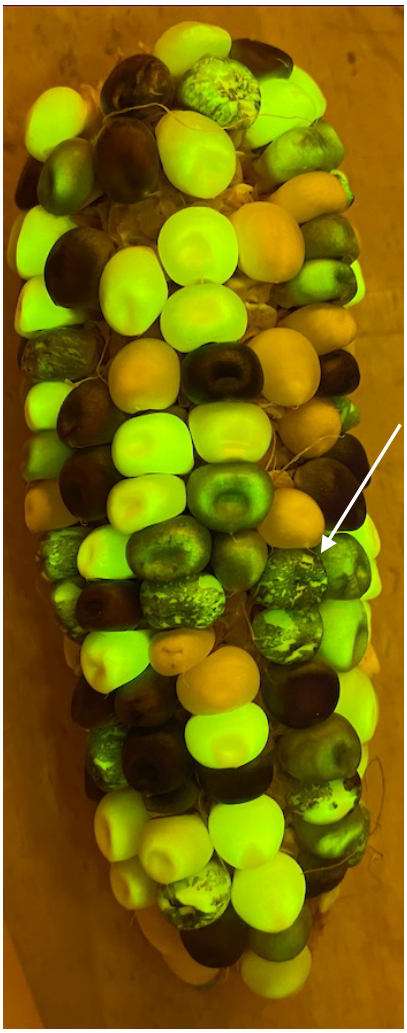
| | No recombination | Yes recombination but no bridge | Presumed recombination, bridge present and ABS staining in midzone |
|---|---|---|---|
| Plant #2 | 17 | 10 | 17 |
| Plant #12 | 5 | 3 | 2 |
| Plant #16 | 8 | 1 | 5 |
| Total | 30 | 14 | 24 |

**Extended Data Fig. 4 | Segregation errors in *LexA-CENH3(v2)* ABS4 plants. a**) Mitotic errors. Root tips from five sectored seeds were processed for FISH and the number of Cent4 and ABS spots counted in interphase nuclei. Two Cent4 (red) spots were observed in all nuclei. ABS (green), which is on one copy of chromosome 4, showed high frequencies of mis-segregation consistent with the model in Fig. 2a. The images shown above are examples. The two cells in the right image are daughter cells of a single division, where the left cell has no ABS and the right cell as two ABS loci. **b**) Meiotic errors. Three plants heterozygous for *LexA-CENH3(v2)* and ABS4 were analyzed at meiosis (siblings from a single ear). Only cells in mid-anaphase I, identified by the short distance between segregating chromosomes masses, were tallied. Figures with no recombination between the centromere and ABS4 were identified by having ABS dots only on one side (first column). Cases where recombination occurred and ABS4 loci segregated freely to one pole were identified by having one ABS dot on both sides (second column). Figures with bridges invariably had one or two ABS dots in the midzone, suggesting there had been recombination between centromere and ABS4, and that centromere cohesion at ABS4 restrained movement. At late anaphase I and telophase I, no intact bridges were observed.

https://doi.org/10.1038/s41477-023-01370-8


**Extended Data Fig. 5 | Leaf defects observed in plants with both *LexA-CENH3(v2)* and ABS4.** Leaf defects observed in plants with both *LexA-CENH3(v2)* and ABS4. The leaves of nearly all *LexA-CENH3(v2)* ABS4 plants looked and felt uneven, with ridges or crinkles. Occasionally the leaves had missing pieces, such as shown here.

https://doi.org/10.1038/s41477-023-01370-8



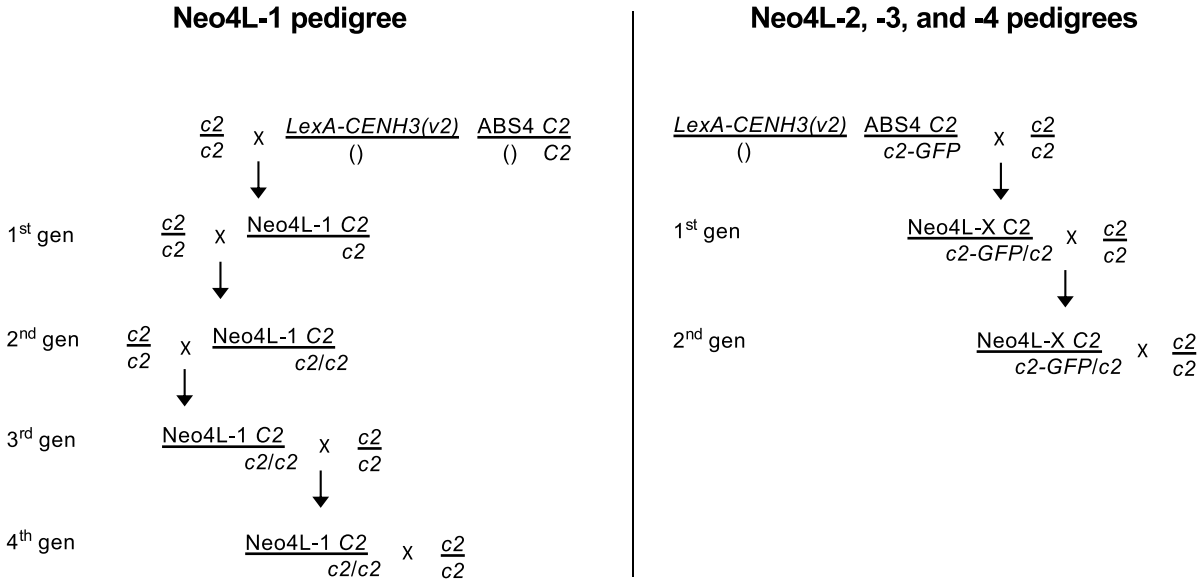


LexA-CENH3 () c2-GFP C2 C1 c1 x ABS4 C2 ABS4 C2 c1 c1 r1 r1

LexA-CENH3 () ABS4 C2 c2-GFP C1 R1 c1 r1 x c2 c2 C1 C1 R1 R1

**Extended Data Fig. 6 | Crosses involving *c2-GFP*.** **a**) The cross used to generate material for the neochromosome screen. *C2* confers purple pigmentation to the outer layers of the endosperm (purple appears black under the blue light used here). The binding of LexA-CENH3 to ABS4 causes errors in chromosome segregation and loss of the linked *C2* gene in sectors (arrow). Two other recessive color alleles (*c1* and *r1*) are also segregating in these transgenic lines, which caused many kernels to be colorless regardless of the presence of *C2* (in this case only *c1* was heterozygous on the female side). **b**) Fluorescent sectored kernels from A were planted and crossed as females to a *c2* tester. Kernels that received both *C2* (purple aleurone) and *c2-GFP* (fluorescent endosperm) were candidates for having inherited a neochromosome. This ear shows such a kernel. Filing off the outer layer of cells removes the purple pigment and makes the underlying fluorescence more obvious (arrow).

https://doi.org/10.1038/s41477-023-01370-8




**Extended Data Fig. 7 | Pedigrees of Neo4L chromosomes.** Females are shown on the left side of the crosses. Neo4L-2, Neo4L-3, Neo4L-4 are denoted as Neo4L-X in the pedigree.

https://doi.org/10.1038/s41477-023-01370-8

**Extended Data Table. 1 | Segregation data for Neo4L chromosomes**

| | Crossed as female | C2 | total | % C2 | # plants from C2 seed that did not have ABS[1,3] | Crossed as male | C2 | total | % C2 | # C2 seed that were sectored | # plants from C2 seed that did not have ABS[2,3] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Neo4L-1** | | | | | | | | | | | |
| 1st gen | KD4205-5 X 4202 | 10 | 140 | **7%** | | KD4202 X 4205-5 | 26 | 237 | **11%** | 0/26 | 7/10 |
| 2nd gen | KD4212-6 X 4211 | 33 | 93 | **35%** | 2/8 | KD4211 X 4212-6 | 41 | 184 | **22%** | 10/41 | |
| 3rd gen | KD4246-4 X 4244 | 70 | 266 | **26%** | 1/9 | KD4245 X 4246-4 | 29 | 180 | **16%** | 6/29 | |
| | KD4246-5 X 4244[4] | 60 | 240 | **25%** | 3/10 | KD4245 X 4246-5 | 22 | 189 | **12%** | 1/22 | 5/10 |
| | KD4246-6 X self [5] | 135 | 219 | 62% | | KD4244 X 4246-6 | 73 | 348 | **21%** | 12/73 | 9/18 |
| 4th gen | KD4301-3 X 4277 | 49 | 231 | **21%** | 1/9 | KD4277 X 4301-3 | 21 | 287 | **7%** | 4/21 | 8/10 |
| | KD4302-3 X 4277 | 56 | 207 | **27%** | | KD4277 X 4302-3 | 3 | 51 | **6%** | 0/3 | |
| | KD4302-4 X 4277 | 62 | 199 | **31%** | | KD4277 X 4302-4 | 5 | 41 | **12%** | 0/5 | |
| **Neo4L-2** | | | | | | | | | | | |
| 1st gen[6] | KD4249-1 X 4244 | 59 | 214 | **27%** | 0/24 | KD4245 X 4249-1 | 0 | 214 | **0%** | | |
| 2nd gen[7] | KD4300-4 X 4277 | 48 | 171 | **28%** | | N/A[8] | | | | | |
| | KD4300-5 X 4277 | 85 | 318 | **27%** | 4/18 | N/A | | | | | |
| | KD4300-7 X 4277[9] | 61 | 230 | **27%** | 3/20 | KD4277 X 4300-7 | 3 | 35 | **9%** | 0/3 | |
| | KD4300-8 X 4277 | 49 | 200 | **25%** | | N/A | | | | | |
| **Neo4L-3** | | | | | | | | | | | |
| 1st gen | KD4303-2 X 4277 | 31 | 124 | **25%** | 2/6 | KD4277 X 4303-2 | 53 | 392 | **14%** | 5/53 | 6/10 |
| **Neo4L-4** | | | | | | | | | | | |
| 1st gen | KD4304-1 X 4277 | 65 | 303 | **21%** | 0/9 | KD4277 X 4304-1 | 27 | 261 | **10%** | 6/22 | 5/10 |

[1]Among progeny from female crosses, 16/113 (14%) of the seeds with C2 endosperm gave rise to seedlings that did not have ABS.
[2]Among progeny from male crosses, 40/68 (59%) of the seeds with C2 endosperm gave rise to seedlings that did not have ABS.
[3]ABS4 was scored by PCR from DNA extracted from one leaf tip. Absence by this assay does not imply ABS4 was absent in all tissues.
[4]48 colorless seeds were planted and 1 had ABS.
[5]62% C2 on self cross is not consistent with ~20% transmission through male & female. The plant may have had two copies of Neo4L-1.
[6]This plant carried LexA-CENH3(v2).
[7]These plants lacked LexA-CENH3(v2).
[8]All plants carrying Neo4L-2 in this generation exerted few anthers and had little pollen.
[9]39 colorless seeds were planted and none carried ABS.

nature portfolio

Corresponding author(s): R. Kelly Dawe

Last updated by author(s): Feb 6, 2023

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | SlideBook 6.022 |
| Data analysis | Cutadapt (2.8), Burrows-Wheeler Aligner BWA-MEM (7.17) |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA874319. Sequence data were aligned to the B73-REFERENCE-NAM-5.0 genome (https://maizegdb.org/genome/assembly/Zm-B73-REFERENCE-NAM-5.0).

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | N/A |
| Population characteristics | N/A |
| Recruitment | N/A |
| Ethics oversight | N/A |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | The experiments were exploratory and sample sizes were not predetermined. We endeavored here to show that tethered CENH3 can recruit native CENH3 using multiple independent tests, including ChIP-seq CUT&Tag, chromosome segregation assays, and ultimately the derivation of independent neochromosomes. Sample sizes were chosen at each stage to ensure the observations were reproducible. |
| Data exclusions | No datasets were excluded. |
| Replication | Experiments were replicated at least twice except the CUT&Tag analyses of individual neochromosomes. For the CUT&tag data, the four neochromosomes serve as replication since the experiment was designed to show CENH3 localization to ABS monomers and pAHC25 plasmids (which was true in all cases). |
| Randomization | Randomization was not used because statistical tests that benefit from randomization were not applied. |
| Blinding | The data measured - ChIP-PCR, short read alignment to a reference genome after ChIP and CUT&Tag, sectored kernels, meiotic chromosome segregation, FISH staining and karyotype analysis - were not subjective. A blinded design was not necessary. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

### Methods

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | Maize and oat CENH3 antibodies were generated in our laboratory (Zhong et al. 2002, Topp et al. 2009). |

| | |
|---|---|
| Validation | The maize CENH3 antibodies were validated by protein blot analysis and immunolocalization (Zhong et al. 2002). The oat CENH3 antibodies were validated by immunolocalization (Topp et al. 2009) and protein blot analysis (this study). |

## ChIP-seq

### Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| | |
|---|---|
| Data access links<br>*May remain private before publication.* | https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA874319. Note that we HAVE NOT deposited graph files for called peaks. This is because we did not call peaks. CUT&Tag enrichment is described in general terms so as to convey the degree of CENH3 spreading. |
| Files in database submission | 15 files of genomic, ChIP-seq and CUT&Tag Illumina data. |
| Genome browser session<br>(e.g. UCSC) | No longer applicable. |

### Methodology

| | |
|---|---|
| Replicates | ChIP-seq was performed with three technical replicates. |
| Sequencing depth | 0.1 to 5x sequence depth depending on the experiment. |
| Antibodies | Custom primary antibodies and standard commercial secondary antibodies. |
| Peak calling parameters | We did not call peaks because our intent was to describe enrichment over repeat arrays (CentC, ABS, pAHC25, knob180 and TR-1), or to describe in general terms the degree of CENH3 spreading into flanking chromatin. |
| Data quality | We did not call peaks, and because of this, did not assess peak quality. |
| Software | Standard sequence alignment software BWA-MEM. |